



THE PRONUNCIATION OF
ENGLISH BY SPEAKERS
OF OTHER LANGUAGES

Edited by
Jan Volín and Radek Skarnitzl

The Pronunciation of English by Speakers of Other Languages

The Pronunciation of English by Speakers of Other Languages

Edited by

Jan Volín and Radek Skarnitzl

Cambridge
Scholars
Publishing



The Pronunciation of English by Speakers of Other Languages

Edited by Jan Volín and Radek Skarnitzl

This book first published 2018

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2018 by Jan Volín, Radek Skarnitzl and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-0390-9

ISBN (13): 978-1-5275-0390-8

TABLE OF CONTENTS

Introduction	1
Foreign Accents and English in International Contexts Jan Volín and Radek Skarnitzl	

Part I: Perspectives on 'Accented' Speech

Chapter One.....	4
Foreign Accents and Responsible Research Jan Volín	

Chapter Two	19
English Pronunciation Instruction through the Accent Lens Ewa Waniek-Klimczak	

Chapter Three	46
Short- and Long-term Effects of Pronunciation Teaching: EFL Learners' Views Pekka Lintunen and Aleksi Mäkilähde	

Chapter Four.....	73
Perceptual Impact of Foreign-Accented Speech Jan Volín, Radek Skarnitzl and Alice Henderson	

Part II: Mapping the Differences in Segments

Chapter Five	96
Character of Vowel Reduction in Czech English Kristýna Poesová and Lenka Weingartová	

Chapter Six	117
Are Word-final Consonants Codas? Evidence from Brazilian Portuguese ESL/EFL Learners Paul John and Walcir Cardoso	

Chapter Seven.....	139
Passing for a Native Speaker: Production and Perception	
Ksenia Gnevsheva	

Part III: Mapping the Differences in Prosody

Chapter Eight.....	160
Phonetic Connectedness in Non-native Speech	
Šárka Šimáčková and Václav Jonáš Podlipský	

Chapter Nine.....	181
Weak Structural Words in British and Czech English	
Jan Volín and Terezie Johaníková	

Chapter Ten	196
Acoustic Correlates of Temporal Structure in North-Vietnamese English	
Ondřej Slówik and Jan Volín	

Chapter Eleven	210
Patterns of Articulation Rate in English-French Tandem Interactions	
Barbara Kühnert and Tanja Kocjančič Antolík	

Part IV: Methodology of Testing and Teaching

Chapter Twelve	228
Evaluating the Essentials: The Place of Prosody in Oral Production	
Dan Frost and Jean O'Donnell	

Chapter Thirteen.....	260
L2 Pronunciation Feedback in English-French Tandem Conversations	
Céline Horgues and Sylwia Scheuer	

Chapter Fourteen	287
Phonetic Inspirations in Authentic Materials: Stimulating Students' Phonetic Awareness	
Marta Nowacka	

INTRODUCTION

FOREIGN ACCENTS AND ENGLISH IN INTERNATIONAL CONTEXTS

JAN VOLÍN AND RADEK SKARNITZL

The title of the book in which you are presently taking interest comprises notions of “pronunciation”, “English”, and “speakers of other languages”. All three concepts deserve a little comment that can make it easier to understand the contents and the general objective of the book.

The narrow meaning of the word “pronunciation” refers to the articulation of speech sounds like [i:, v, s, g, m]. This popular use of the word can be a bit misleading in the field of our research. The small speech segments are actually only descriptive tools reflecting what we recognize when we consciously observe and analyze spoken texts. They are most probably neither the true building blocks of the phrases that we utter, nor the templates we use for decoding spoken messages. Our understanding of the word pronunciation encompasses the production of stronger and weaker syllables (including their segmental features), melodies and temporal or amplitudinal attributes that make speech real in the psychological and neurophysiological sense.

An interesting justification of this stance is the term *accent* itself. It is primarily motivated by the prosodic phenomenon referring to the manifestation and distribution of prominences in the speech continuum. Groups of people share certain specific speech production features that are recognized by other groups of people as typical of the observed group. The fact that these features fall under an umbrella term of *accent* and not *vowelism or *phonemia acknowledge the importance of the wider approach to speech.

In the same vein, the terms phonetics or the adjective phonetic will refer to the entire complex sound structure of speech. We find expressions like, for instance, *phonetics and intonation* ridiculous (a simple Google search for this exact phrase returned 41,600 results). For a thoughtful

phonetician, intonation belongs to the domain of phonetics. (Coordination of the two terms is analogous to *food and apples* or *animals and rabbits*.)

English is currently the language of international communication and there are various theories why this has happened. Instead of speculating about languages that might take over, we build on the fact that hundreds of millions of people learn English as something serviceable, something they would like to master. To many, English is not the mother tongue. Those are the *speakers of other languages* in our title. The sound of their English is influenced by the sound structures of languages they have learned beforehand. We find these variations in the sound of English fascinating and for many practical reasons beneficial to explore and exhibit.

The work on this book started during the final stages of the 4th international conference *English Pronunciation – Issues and Practices*, which we organized in Prague in May 2015. More than seventy participants from four continents with 52 presentations manifested unusual dedication to research in the field. We realized that besides the proceedings on a CD we should invite some of the most dedicated researchers to expand on their topics and write a book chapter that would allow for sharing their findings with wider audiences. The peer-review process eliminated a few contributions and helped to improve the rest of them. The result is enclosed in this book.

It consists of four parts. In Part 1 broader, more general considerations of foreign-accented speech are exposed together with analyses of learner beliefs and attitudes to pronunciation instruction. Polish and Finnish learning environment is used to demonstrate certain issues. The second part brings several accounts of consonantal and vocalic phenomena demonstrated on Czech, German, Korean and Portuguese accents of English. Part 3 complements the preceding chapters with questions of speech prosody and adds Vietnamese and French-accented English. The fourth and final part considers methodological aspects of English pronunciation learning and offers inspiring tips for classroom management, testing pronunciation skills and enhancement of the learning process.

We would like to thank all the authors for their disciplined and responsible attitude to the book preparation and the staff of the publishing house for their professional, friendly and helpful approach.

Jan Volin & Radek Skarnitzl

PART I:
PERSPECTIVES ON 'ACCENTED' SPEECH

CHAPTER ONE

FOREIGN ACCENTS AND RESPONSIBLE RESEARCH

JAN VOLÍN¹

1.1 Introduction

In this chapter I am going to argue that the impact of foreign accents is far from trivial and when dealing with them, whether in discussions or in actions, we should do our utmost to avoid approaches based on ideologies or wishful thinking. Current science offers an alternative to these futile approaches. It provides adequate equipment to allow for a thorough exploration of the true nature of psychological and social consequences of accented speech. We are capable of searching for information that will substantiate useful practices both in classrooms and outside school settings. It is still prudent to acknowledge, however, that without genuine interest and impartial stances we will hardly succeed.

1.2 Controversies of variation and standard

It is a well-known fact that no two people in the world speak exactly the same way. Just as every individual exhibits a unique appearance, he or she also displays a unique manner of speaking. However, once we start studying these unique speech production patterns systematically, we realize that it is not a disarray of unpredictable idiosyncrasies. The pronunciation patterns can be grouped according to various similarities into accents. (Pronunciation is understood in its wider sense, i.e., including prosodic features – see Introduction to this book.)

¹ This study was supported by the Metropolitan University Prague internal research scheme IGS D45-06 financed from the programme “Institutional Support of Long-Term Projected Advancement of Research Organizations in 2017”.

In the past, people would recognize speech production peculiarities of the neighbouring villagers and they would commonly express such awareness in teasing, which often took form of fixed mocking phrases. The differing element would be exaggerated in these phrases or, metaphorically, magnified so as to create a caricature. The ancient idea that “we” speak correctly and whoever differs is incorrect is a deep-rooted one. Even open-minded individuals with the gift of tolerance, who embrace variation in the world and do not link strange to bad, still need to belong somewhere. The psychological need to be part of a community is innately human (e.g., Maslow, 1943), and speech provides quite a handy way to manifest belonging to or distancing oneself from various human communities (more about this in the following section).

The Anglophone world is quite rich in accents – but this is not necessarily due to its geographical vastness. Australia with its 7 692 000 square kilometres is more than thirty times larger than the United Kingdom, yet the accent variation is much richer in England, Scotland, Wales and Northern Ireland. For accents to develop, specific suitable social conditions must exist, and history does not provide those quickly and easily. The current increase in the mobility of human population makes geographical factors relatively weaker, but other factors of variation such as socio-economic status, ethnicity, gender or age still lead to differentiation among speech patterns of specific social groups.

The very special situation of English among other languages stems from its status of a widespread international language. It is often pointed out that the Earth now accommodates more speakers of English whose mother tongues are different from English than those who were born into English speaking families, i.e., native speakers of English. A natural consequence of this is that there are more speakers who produce foreign-accented English than those who display native pronunciation patterns. Obviously, this division builds on the standpoint of those who might want to claim the language, i.e., people who heard and learnt their first words from their English speaking parents. They somehow feel a stronger bond to the language and some of them even the right to fashion its fate. Although at times this feeling is labelled as undemocratic, it might prove difficult to change from the orators’ pulpits.

Another current controversy that emerges relatively often when people discuss pronunciation variants of a language is the concept of the *standard*. In a wider sense the adjective *standard* can mean *typical* or *generally accepted* and the noun usually stands for a set of rules that describe some sort of a typical or generally accepted effect of human activities. (The gold standard in economy differs, but as a system for

setting the value of currency, it too has to be generally valid to be functional.)

The idea of the *standard form of a language* can be quite easily used to harass users of forms that are classified as non-standard. Lippi-Green, for instance, speaks of standard language ideology (SLI) and complains that the set of rules for the standard language use is based on the speech of upper-middle classes (Lippi-Green, 2012: 67). She is understandably bitter about speakers being bullied due to the use of various subordinated language forms, yet she seems to be a little too passionate in rejecting the whole concept of standard. While it is possible that she is correct in suggesting that the language standard is a collective delusion and should be abandoned altogether, it might be worthwhile to spend some time thinking of how standards are generated and what roles they fulfil. It is clearly one thing to disagree with social injustice, but a completely different thing to refuse to discuss it impartially so that all its important aspects can be considered. (It will be stressed repeatedly in this chapter that impartiality provides better chances to discover useful facts than partisanship.)

It can be universally observed that communities of speakers have appreciation for certain ways of speaking, or with regard to the narrower focus of the present book, certain pronunciation patterns. People cannot be prevented from evaluating the sound of someone's speech. Whenever something is said, a process of assessment is triggered on the part of the listener not only as to the contents of the utterance, but also as to its form. This process cannot be disabled at will. We are constantly trained to pay attention to the form, as it may contain important signals which modify and sometimes even invert the representative meanings of the words spoken. Furthermore, phonetic forms also signal the membership of the speaker in a group and, potentially, his or her attitude towards this membership (the desire to enhance or subdue it). It should be remembered that there are as many standards as there are groups of speakers, and one speaker normally belongs to more than one group. Countless groupings of speakers unconsciously produce the norms (very often tacit ones) that are observed and served. Norms are an inseparable part of the human social conduct.

However, if a standard form of an entire language is considered, what is usually meant is a widely understood and accepted variant that tends to signal some sort of detachment. Such a variant may offer the speaker the option to stay non-aligned, not linked to factions, but appear somehow above ordinary groupings that one enters when solving problems of daily life. And just as the entire language community could be a complex

dynamic aggregate of numerous social groups, so is the “language standard” an ever-changing construct that cannot be described in its entirety. Educators may be unhappy at times not to have a complete and reliable description of the standard, but it has to be stressed that the language community as a self-organizing system always has enough (if not a vast majority of) mature members who do not have to be guided as to what social and language norms match individual situations.

When speakers opt for standard forms they do not express their admiration or allegiances to upper-middle classes. In many cultures the alignment is with education, politeness, carefulness or formality. (In addition to that, Milroy and Milroy (1992) discuss the link between the standard and prestige.) Various social norms exist that guide the speaker as to when to use the standard. If a school teacher instructs the children to use standard forms *all the time*, then he or she is obviously denying the existence of social norms. However, if a propagandist declares a war on the concept of the standard, then he or she makes the same blunder as the afore-mentioned school teacher.

It is clear from the propositions above that the standard is a concept based on relatively vaguely shared ideas. (The vagueness refers to the difficulty to capture the system through conscious analytical thinking and reach consensus with others.) When language teachers want to offer their students the option of generally accepted pronunciation norms, they may either use a published prescription and hope it is in reasonable harmony with reality, or rely on their own intuitions. From a researcher’s point of view, the attempts to capture the standard in English lead up and down slippery paths. Hypothetically, the guidance could be provided by a consensus about the level of oddity that the listeners perceive when they encounter a given form. It could be argued that if a certain pronunciation form passes unnoticed by an absolute majority of the speakers, then it is accepted in terms of standard pronunciation. Empirical validation of such consensus, however, would definitely require many well-designed projects.

1.3 Neurophysiological and psychological background

In its psychological nature speech is behaviour. In an individual instantiation it comprises a set of actions with a communicative purpose. The previous section mentions a strong psychological need to belong. Groups of people who belong together regulate behaviour of their members by social norms, and speech behaviour is no exception to that. Perhaps, social norms should not be viewed separately from the speech-

related norms. Even a very cautious view should include the notion that speech decoding norms are a subset of social norms, and a bold view would argue that both sets of norms are very tightly intertwined and often inseparable. This holds not only for speech production but also for speech perception norms, i.e., those that regulate the way speakers decipher spoken messages. Implicit learning of how to understand certain phonetic features of an utterance takes place with great intensity in the first years of our lives and continues throughout the lifespan.

Another strong psychological need of an individual is the need to be respected or relevant. Unless seriously affected by an autistic disorder, a human being talks primarily not to pass information, but passes information to be relevant to others. To achieve this, not only the surface contents of the utterance must resonate with the listener's interests, but the form must be such that the listener understands. It must be emphasised especially in the context of foreign-accented speech that messages which are barely intelligible will not lead to the desired outcome. We might go even further and suggest that to be respected or relevant the speaker needs to be comfortably intelligible. It is a common experience of many people that without comfortable intelligibility frustration or irritation takes place. The listener (sometimes without knowing why) may grow increasingly uneasy or impatient, and may wish to reduce social contact with the speaker whose speech is taxing. Below I am offering a neurophysiological explanation of this.

The third psychological need of healthy humans is the need to be free. However, social foundations of our lives make freedom a complicated concept. For centuries, philosophers have been analysing the links between freedom and accountability, and even without philosophical training many people will ask: "Free – at whose expense?" This question is strongly phonetically pertinent. If we decide to free ourselves from pronunciation norms, then it is the listener who will have to pay the dues. It is a well-known principle observable at various levels of speech behaviour. The speaker's economy of effort has to be counter-balanced by the labour on the part of the listener. This principle is reflected in many cultures in the link between careful pronunciation and politeness. In other words, to display a polite stance the speaker may choose to exert greater effort so that their speech is clear. Conversely, to signal disregard one may merely choose to reduce the articulatory gestures including the pitch range and loudness.

From the discussion of the three selected psychological needs in the preceding paragraphs it is clear that attention to appropriate norms is strongly recommended. The consequences of not doing so can be quite

damaging. Yet, there is another severely damaging phenomenon that has to be mentioned in connection with accented speech and its psychological aspects. It is an unfounded negative attitude to a group of people called a *prejudice*. As Gordon Allport, the icon in the field of prejudice research, showed more than half a century ago, it is an extremely powerful phenomenon affecting at some point virtually anyone's life (Allport, 1954).

Why is prejudice – an attitude formed without much (sometimes without any) experience – so prevalent in our lives? It is a product of a cognitive process called stereotyping. Human lives require constant decision making and under certain conditions the decisions are a crucial factor in terms of survival. If the decisions have to be quick (in dangerous situations), there is no time to gather arguments for this or that choice. The individual has to draw on the past experience even if it is miniscule or indirect, i.e., based on what someone else reported. It seems that it is evolutionarily advantageous to decide, even if the decision is wrong, rather than stay passive. In the greater scale of events not doing anything means a smaller probability to succeed than acting, albeit sometimes erroneously. The ability of stereotyping developed to help humans survive.

Unfortunately, as the dangerous situations requiring fast decisions become scarce with the economic progress of human society, the positive aspect of stereotyping loses its relative dominance, while the negative aspect stays and grows. Our everyday lives are nowadays organized in such a manner that we typically do have time to collect evidence and contemplate the substance of problems. Yet the old cognitive mechanism seems to be better disposed to perform. Thus, hundreds of studies show that foreigners or minor ethnic groups are perceived with suspicion or worse. Allport's classic experiment with accommodation quest, when the same male was sometimes welcomed, sometimes refused over the telephone, based only on the surname he used to introduce himself (typical English names like Jefferson or Whitney meant success, Jewish-sounding ones like Rosenfield or Silverstein led to refusal), has been adapted to the phonetic domain repeatedly (e.g., Lambert et al., 1965; Aronovitch, 1976; Brennan & Brennan, 1981; Purnell, Idsardi & Baugh, 1999; Campbell-Kibler, 2007).

Prejudice should not be confused with the assessments that are produced by our fast adaptive brain mechanisms. Ambady and Rosenthal (1993) report that after just a thirty-second observation of a videoed lecturer (with the sound track off), students were able to produce estimates of their teaching quality that significantly correlated with evaluations of other students who were actually taught by those lecturers for a semester.

In other words, a very brief and incomplete observation may not differ from a long-term and a relatively thorough one. The quick unconscious processes may produce outcomes that are useful or misleading. Listening to low-pass filtered speech (i.e., speech where words cannot be recognized but rhythm and intonation can be perceived) for less than a minute produced assessments of medical doctors that significantly correlated with the number of law suits filed against them. Again, some unknown detail in the tone of voice leads to non-random estimates of the individual's personality (Ambady et al., 2002).

In a sense related to both prejudice and fast adaptive thinking are the results of Rubin (1992) and Niedzielski (1999). The former showed that even if two groups of respondents listened to the same speech sample, their memory retention and comprehensibility judgements differed. The only manipulated element in the experimental situation was a photograph of the alleged speaker that the respondents watched while listening. The latter achieved a similar effect by just orally informing respondents about the origins of the speaker. (Niedzielski did not measure memory retention, though – she was interested in the evaluation of differences between vowels.)

To explain the neurophysiological foundation of the effects of foreign-accented speech, it is useful to consider a few long known facts about speech perception. First of all, as Liberman and his colleagues already proposed in their Motor Theory of Speech Perception, the listener is not passive when a chain of acoustic events is unfolding in front of his ears (Liberman et al., 1967 or a revised account in Liberman & Mattingly, 1985). There is evidence that quite intense cerebral processes are going on and they comprise much more than just observation of the incoming acoustic signal and decoding the meaning from it. Our brain performs some sort of hidden mimicking of the articulatory gestures that produced the speech which is coming in through the ears. While listening to someone, we, to some extent, imitate that person's articulatory gestures by our neural production circuitry, except the outcome is not sent to the muscles so no external movements are executed.

Apparently, our brain can easily imitate only familiar gestures. If unusual muscular manoeuvres are used, the listeners struggle with the incoming speech signal. The term unusual pertains to timing and temporal alignment as well. Huggins (1979) reported his perceptual experiments with temporal patterns in speech and their impact on intelligibility. Standard rhythmic configurations of syllables positively influenced recognition of words in comparison with non-standard ones, which generated a considerably high word error rate. (Huggins himself, however,

speculated that this effect could be explained by malfunctions of short term memory: *ibid.* p. 283). Similar experiments a few years later already explicitly refer to the rhythm of speech, i.e., distribution of prominence contrasts in time (Buxton, 1983). Reaction time measurements revealed that ordinary rhythmic patterns lead to easier cerebral processing while the distorted ones require greater cerebral effort. Since then, numerous modifications or replications of these experiments confirmed that unusual acoustic patterns in speech activate additional cognitive resources, which may take its toll in areas such as attention or working memory (recently, e.g., Van Engen & Peelle, 2014).

The physiological framework of the phenomena observed in the experiments cited above is outlined in the Adaptive Resonance Theory (Pollen, 1999; Carpenter & Grossberg, 2002; Grossberg, 2003; Amis & Carpenter, 2009). Adaptive resonance is a general cognitive mechanism concerning recognition of visual or auditory objects and learning. An object in our field of interest could be a spoken word, morpheme, syllable etc. The mechanism builds on powerful predictive activities performed by our neurons. In perceiving speech, the incoming acoustic signal pre-processed by the auditory cortex meets with the expectational neural representations. Those are generated by the predictive activities and they are based on our experience. Our brain knows the language and various contexts well enough to produce quite accurate expectational representations. When these meet with the incoming signal and they reasonably match with it, an act of neural resonance takes place. It is a moment of synchronized activities of the neurons involved at the “meeting point”. That is also the instant of object recognition, i.e., the moment when a word (morpheme, syllable) is recognized.

A lay person might wonder why neurons waste energy on prediction if they could just wait for what is coming in and compare it with some stored templates. The answer is simple. There is no extra energy spent on predictions. They happen due to the very nature of neurons. These little cells can only switch on and off, but the fascinating thing is that they do it in dependence on each other. They form associations and these help them map the structure of the language (or any system) and compute probabilities that, for instance, with this sound that one is most probable to occur, and with this word that one is likely to collocate, and after this idea that one should follow. So even if we do not know the topic that the speaker is discussing, based on our knowledge of the language and the world we still manage to be slightly ahead with our perception. Just confirming the predictions appears to be a more economical (and practical) way of perception than waiting with an empty plate.

How does this mechanism enter the debate on foreign accents? There is one crucial condition for smooth perceptual flow (i.e., for shorter reaction times, fewer errors – see above). It is the synchronization of the two streams of neural activity. The incoming and the expectational signals must meet at the right time. The brain is sending the expectational representations so that the timing of their encounter with the incoming imprints of reality is optimal. However, the timing in foreign-accented speech is not necessarily predictable and the neurons that are involved may have to repeat their activities to engage the reciprocal assembly. Without neural resonance an object is not recognized and repair mechanisms make the process of speech perception less economical. It is quite possible that irritation or unease on the part of the listener unfamiliar with the given type of non-standard patterning in speech is the consequence of extra labour his or her neural assemblies have to perform. In connection with this it is also interesting to consider the results of the study of Volín, Poesová and Skarnitzl (2014 – see below in Section 1.5).

1.4 Social significance and current approaches

Allusions or explicit references to accented speech can be found in very old scripts. The Old Testament (specifically The Book of Judges, Chapter 12) comprises a testimony of undercover invading soldiers being recognized and punished thanks to non-native pronunciation. It was the confusion of alveolar and post-alveolar voiceless fricative – a feature found regularly in Dutch, Finnish or Spanish accents of English – that allegedly cost thousands of lives. The importance of clear speech and its impact on listeners is mentioned by many ancient philosophers, orators, politicians. References to coaches giving paid lessons in “good speaking” and to the legendary self-taught Demosthenes (4th cent. BC) are illustrations of the awareness of the value attributed to the sound of speech. Until today, certain speaking styles are in many cultures linked to education, which is viewed as a prestigious commodity.

The social significance of pronunciation is also reflected in the modern empirical approach to speech, which dates back several centuries (e.g., von Kempelen, 1791). A prominent signpost in the history of social attention to speech is the foundation of the International Phonetic Association in 1886, in which teachers of foreign languages were heavily involved. It is useful to remember that the teachers themselves demanded scientific foundation for their work.

Yet unfortunately, in today’s social debate on foreign accented speech, two extreme stances can be heard more often than anything. On the one

hand, there is the authoritative prescriptionism, and on the other hand, there is the naïve liberalism. The prescriptionists argue that there are “ways thing should be” and learners of languages are obliged to sound “proper”. The trouble with identifying the proper is mentioned in Section 1.1 above. A compelling claim voiced by the prescriptionist camp concerns the social advantage that competent speakers possess. This claim, however, easily translates into not succeeding in life without decent pronunciation, and leads to anxiety of failure in some learners or the feeling of exclusion from the society (Gluszek & Dovidio, 2010; Gluszek, Newheiser & Dovidio, 2011). It would be very helpful to be able to measure objectively to what extent such feelings really motivate the individuals, or, conversely, impede their true potentials.

The other extreme is represented by claims that the importance of one’s sound is grossly inflated and if we recognize that, we will be liberated. The underlying belief could be expressed by an infamous quote of the wife of a former top U.S. politician. When asked about drug addiction, she advised the affected subjects: “Just say no”. A similar solution to the impact of accented speech is proposed by naïve liberals. They offer a simplistic impractical stance arguing by an individual’s right to choose. Although very appealing in the political sense, the claims of naïve liberals are not supported by any rigorous, representative research. They, too, are most probably hugely demotivating in the learning environment.

These two camps are similar to football fans or to fervent supporters of competing political parties. Although their members are mostly educated people, they too have a strong, and in some cases even blinding desire to win. Unfortunately, they do not realize that the ability to see clearly and objectively is not under the conscious control of an individual. We do not see objectively at will. Once we accept a certain opinion as correct, we find it very difficult not to overlook counter-evidence, even if we do not want to cheat. Our mental capacities cannot be ordered to see the entirety clearly and accurately. In the field of academic research, however, we can help them by refusing to decide ahead what we want to find (or in its softer version – what the truth most probably is).

1.5 Responsible research

The number of researchers who are ready to carry out respectable experiments and observations in the field of foreign accents is growing. An indirect indicator could be seen in the number of conferences and journals dedicated to the topic. Although they do not necessarily testify

about the reputable scientific methods used in the field, they hint that societies are willing to spend money on this research, i.e., they consider it useful.

More direct proofs of the scientific prowess of researchers in the field can be found inside the dedicated journals and other publications. The methods used are compatible (or even identical) with those used in psychology, sociology or experimental linguistics – more traditional disciplines with an acknowledged impact on the development of science as a whole.

Apart from constantly scrutinized methodology, one of the features of current research work is prudence when it comes to strong claims. The availability of information which is typical of the present era (connected with the internet and relatively low costs of printing) makes scientists realize more than ever before that one single individual can hardly apprehend the complex phenomena of the world. The capacity of human cognitive mechanisms is quite limited, and without the collective effort we cannot hope to appreciate the complexities of natural or social structures. In today's setting, strong claims are linked to immaturity rather than boldness. If, for instance, the pressure to standardize language forms across language communities is called unjust (or malicious, evil, etc.), then we can use the *personal social network analysis* (PSNA) to find out what it actually means in a less abstract, but more expedient manner.

PSNA was introduced by social anthropologists (see, e.g., Mitchell, 1986), but soon attracted the attention of sociolinguists since it allowed to explain specific choices of speaking individuals rather than cumulate metrics of large groups. One of the important findings was the link between the denseness and tightness of the social networks and the willingness to accept influences from outside. Hence, communities with looser and not so numerous interpersonal ties may adopt standard features more easily than communities with closely knit ties. However, what is also clear is that the freedom not to standardize is traded for strict obligations to the group. An individual *must* rather than *can* observe the norms of the community and may be constrained in multiple aspects of personal life. Apparently, gang behaviour is not a matter of simple choice. It follows that fighting too fiercely for or against standardization means a risk of great harm, while thorough research of the problem can only help to avoid unpleasant (and sometimes even disastrous) consequences. (Too many people acquired bitter experience due to promises that they will all live as one nice happy family if they accept “historical inevitability”. Ignoring psychological and social dispositions of humans has brought immense suffering in the name of “justice and order”.)

A popular belief in the domain of foreign accents links all the perceptual difficulties with prejudice. As I have demonstrated in Section 1.2 above, there is growing evidence that most of the effects are not prejudice induced. One more example deserves to be mentioned. Lev-Ari and Keysar tested the impact of accentedness on the perceived truthfulness of statements (Lev-Ari & Keysar, 2010). They found that people attribute lower truthfulness scores to propositions that are spoken with a foreign accent. Although it would be quite easy to blame the results on prejudice, they carried out a follow-up study in which they explicitly urged the respondents to guard themselves against the foreign accent effects. The results suggest that, indeed, it is not necessarily prejudice, but rather “fluency effects”. (Fluency is used in a metaphorical sense as the potential ease of processing – see, e.g., Oppenheimer, 2008.)

Results that point in the similar direction were obtained in the study by Volín, Poesová & Skarnitzl (2014). They extracted a number of longer statements from various English radio programmes and for each of them created a duplicate in which stressed vowels were artificially shortened and unstressed reduced vowels lengthened. The change was barely noticeable. Adult listeners were asked to try to judge the personality of the speaker from his or her voice. All guiding questions were linked to the emotional stability of the speaker and the estimates were implemented on a seven-point scale. A statistically significant outcome indicated that speakers with “less usual” rhythmic flow of vowels are perceived as more neurotic, irascible or less emotionally stable. It has to be emphasised that the durational manipulations did not model any specific foreign accent so, again, it would be difficult to argue that the listeners were prejudiced. Yet equalizing durations of stressed and unstressed syllables has been reported from many foreign accents of English.

Even in the domain of speech production, where all effects tend to be habitually attributed to the influence of the mother tongue, examples can be found which invite a re-evaluation of our approaches. For instance, Volín, Poesová & Weingartová (2015) found that certain features of F0 tracks (correlates of intonation) in Czech-accented English do not lie midway between Czech and English. They were found outside the space dividing Czech and English speech melodies. Rather than to the interference of the two prosodic systems the results seem to point toward certain “dysfluency” in the target language. (The use of “dysfluency” is again in the sense of Oppenheimer, 2008.) It is not to say that speakers of Czech English stuttered, stopped abruptly or re-started their sentences – they read them from a script with which they familiarized themselves ahead of reading. Most probably it was their implicit uncertainty about the

language, some sort of subliminal hesitation or hidden feeling of lower competence that prevented them from producing more vivid forms (i.e., those that exhibit ease of speaking). The intonations disclosed this even though the affective charge of the situation was not explicit.

All the examples above suggest that rather than being overly concerned with correctness, our teaching methods should pay more attention to fluency. And again, the research has to show to what extent. It is only too easy to become a newly established educationalist and start a “reform” that introduces a new extreme against an old one. In my lengthy career in schools, I have witnessed many such expensive exercises of newly gained power by officials in education, but the outcome has always been meager if any. Top-down reforms in state run schools principally offer a good opportunity to spend money. Apparently, there are individuals who believe that their reforms will improve the educational effectiveness, but since they do not test the new methods rigorously on impartially selected samples of students, and they do not care to disseminate the research outcomes among teachers in a comprehensible and trustworthy manner, the reforms make hardly any impact.

Responsible researchers are not biased, they are patient, diligent, and spend time and energy thinking of how to speak to teachers at schools, how to make them interested. This is because only interested teachers will care to implement the new training procedures. Without that they may travel to special seminars dedicated to the new method implementation and yet come back home unconvinced and unwilling to divert from their routines. For an extra bonus they may be willing to pretend they are supporters of the novelties, but this all usually fades away only to recur when a new cohort of attention seeking reformers is installed in the offices.

It is not easy to withstand the pressure from laymen on researchers to provide an unequivocal answer as to whether accents matter or not. Nevertheless, not knowing does not necessarily mean people should stay inactive. Everyone is invited to take a stand against injustice in individual cases (particularly those that one understands well). However, sweeping political measures should be discouraged for the time being, and a greater support for responsible research should be made available.

References

- Allport, G. W. (1954). *The Nature of Prejudice*. Cambridge, Mass.: Addison-Wesley Publ. Co.
- Ambady, N. & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64, pp. 256–274.
- Ambady, N., Laplante, D., Nguyen, T., Rosenthal, R., Chaumeton, N. & Levinson, W. (2002). Surgeons' tone of voice: A clue to malpractice history. *Surgery*, 132/1, pp. 5–9.
- Amis, G. & Carpenter, G. A. (2009). Self-supervised ARTMAP. *Neural Networks*, 23, pp. 265–282.
- Aronovitch, Ch. D. (1976). The voice of personality: Stereotyped judgements and their relation to voice quality and sex of speaker. *Journal of Social Psychology*, 99, pp. 207–220.
- Brennan, E. M. & Brennan, J. S. (1981). Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research*, 10/5, pp. 487–501.
- Buxton, H. (1983). Temporal predictability in the perception of English speech. In: Cutler, A. & Ladd, D. R. (Eds.), *Prosody: Models and Measurements*, pp. 111–121. Berlin: Springer-Verlag.
- Campbell-Kibler, K. (2007). Accent, (ing) and the social logic of the listeners perception. *American Speech*, 82/1, pp. 32–64.
- Carpenter G. A. & Grossberg, S. (2002). Adaptive resonance theory. In: Arbib, M. A. (Ed.), *The Handbook of Brain Theory and Neural Networks*, 2nd edition, 87–90. Cambridge, Mass.: MIT Press.
- Gluszek, A. & Dovidio, J. F. (2010). Speaking with a non-native accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of Language and Social Psychology*, 29, pp. 224–234.
- Gluszek, A., Newheiser, A. K. & Dovidio, J. F. (2011). Social psychological orientations and accent strength. *Journal of Language and Social Psychology*, 30, pp. 28–45.
- Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31, pp. 423–445.
- Huggins, A. W. F. (1979). Some effects on intelligibility of inappropriate temporal relations within speech units. In: *Proceedings of 9th ICPHS*, 283–289. Copenhagen: University of Copenhagen.
- Kempelen, J. W. von (1791). *Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine*. Wien: J. B. Degen.

- Lambert, W., Anisfeld, M. & Yeni-Komshian, G. (1965). Evaluational reactions of Jewish and Arab adolescents to dialect and language variations. *Journal of Personality and Social Psychology*, 2/1, pp. 84–90.
- Lev-Ari, S. & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology* 46, pp. 1093–1096.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74/6, pp. 431–461.
- Lieberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21/1, pp. 1–36.
- Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States*. London, UK: Routledge.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50/4, pp. 370–396.
- Milroy, L. & Milroy, J. (1992). Social network and social class: Toward an integrated sociolinguistic model. *Language in Society*, 21, pp. 1–26.
- Mitchell, J. C. (1986). Network procedures. In: D. Frick & H.-W. Hoefert (Eds.), *The Quality of Urban Life*, 73–92. Berlin: de Gruyter.
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18/1, pp. 62–85.
- Oppenheimer, D. N. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, 12/6, pp. 237–241.
- Pollen, D. A. (1999). On the neural correlates of visual perception. *Cerebral Cortex*, 9, pp. 4–19.
- Purnell, T., Idsardi, W. & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, 18/1, pp. 10–30.
- Rubin, D. L. (1992). Non-language factors affecting undergraduate's judgments of non-native English speaking teaching assistants. *Research in Higher Education*, 33/4, pp. 511–531.
- Van Engen, K. J. & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers In Human Neuroscience*, 8, pp. 1–4.
- Volín J., Poesová K. & Skarmitzl R. (2014). The impact of rhythmic distortions in speech on personality assessment. *Research in Language*, 12/3, pp. 209–216.
- Volín, J., Poesová, K. & Weingartová, L. (2015). Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language*, 13/1, pp. 107–123.

CHAPTER TWO

ENGLISH PRONUNCIATION INSTRUCTION THROUGH THE ACCENT LENS

EWA WANIEK-KLIMCZAK

2.1 Introduction

Accent is adopted in this chapter as a working perspective for the discussion of problems and possible solutions in pronunciation instruction. Referring to accents, the chapter explores the role of native and non-native accents in specifying the model and the target in teaching and learning English. The main aim is to show that while accents, not only native but also non-native, need to be accepted as an obvious reality of natural language use, the instructional setting requires clear guidelines as to which pronunciations used by the learners need to be corrected. Such guidelines can be offered by a linguistic phonetics approach which focuses on linguistically relevant categories as the basis for sound system formation and usage (Ladefoged, 1997). Extending the original view to non-native systems, pronunciation instruction can be seen as aiming at the formation of sound categories recognized as linguistically relevant across accents of English, or in a specific accent of English, depending on the choice and needs of the learners. However, the realization of these categories, i.e. the degree of 'native-like' pronunciation is expected to vary and to be accepted as characteristic for a given non-native accent. The specificity of this approach is in the use of phonetic parameters for the description and analysis of sound categories. Thus, unlike a traditional phoneme/allophone approach as represented by e.g., Gimson's *Pronunciation of English* (Gimson, 1962), rather than talking about abstract categories and their physical implementation, the linguistic phonetics based approach advocates the use of physically-real sound description as the basis of the model for learners. The model is defined then in terms of the range of values typical for specific native accent or accents; the range used by non-

native speakers is expected to vary, with the target defined as the production of parameter values sufficient to mark the contrast at the category level.

There are two major reasons for adopting an accent perspective: firstly, accents are of major importance in pronunciation instruction, and secondly, more importantly, all of us, native and non-native speakers of English, speak with an accent. In this sense, an accent, defined as a ‘loose bundle of prosodic and segmental features distributed over geographic and/or social space’ (Lippi-Green, 1997: 42) is seen as a natural consequence of language experience. The key difference between native and non-native accents results from the nature of previous language experience. Thus, a native language speaker’s accent reflects their experience of the variability within what is defined as one language system (intra-language experience) while a non-native speaker’s accent reflects the experience of the variability in more than one language (inter-language). The inter-language has been defined by Major (2001: 1) as ‘An adult second language learner’s linguistic system (...) or simply the language of a non-native speaker’, which ‘is a product of a combination of parts of the L1, parts of the L2, and universals’ (ibid.: 4). Notice, however, that discussing accents we refer to the accents of a particular language. Consequently, while we can talk about the inter-language experience of a non-native speaker, the elements of language transfer and language universals, it is the system of the language the accent of which we describe or discuss that functions as the point of reference. Consequently, extending Lippi-Green’s definition to non-native accents, in the analysis of the elements of those loose bundles of segmental and prosodic features found in the speech of non-native English speakers, we define and describe non-native accents from the perspective of the system of English adopted as a point of reference, typically RP or GA, not unlike Wells in his *Accents of English* (1982) or Collins and Mees (2008) in their pedagogical description of accents for learners. Notice also that all accents can be analyzed at different levels: segmental or suprasegmental, and what can be called sub-segmental, i.e., the level of phonetic parameters, or the level of sound-categories and their phonetic realization¹. However, when we talk about accents, what is foregrounded is the use of language and a specific user or groups of users of the language, with the focus on their language experience and other characteristics, including language attitudes and individual differences.

¹ Presenting the most comprehensive description of English accents, Wells (1982) refers to systemic and realizational differences, supplemented by lexical-incidental and phonotactic.

Thus, at a descriptive level, accent is something we all have and something we can describe. It is an inevitable (expected, natural) element of language use. However, when an accent goes to a language school, it becomes not only ‘good’ or ‘bad’ but also ‘right’ or ‘wrong’, ‘correct’ or ‘incorrect’. This is not surprising: in parallel to other elements of the language system, grammar and vocabulary, pronunciation instruction relies on a clear identification of what is expected from the learner and ‘success’ is measured by the degree to which correctness has been reached. This is true for native and non-native accents alike, however, in the case of non-native accents, it is not only the question of ‘good’ or ‘bad’ in terms of socially desirable or not, but also, or - as many teachers will agree, more importantly - whether it is intelligible or not. Consequently, it is for a long time now that intelligibility rather than correctness has been stressed in formulating aims for pronunciation instruction. In the discussion of pronunciation priorities for intelligibility, the issue of an accent seems to have been lost, or at least moved to the background; this paper hopes to bring it back to the foreground in order to demonstrate that a non-native accent perspective offers possible solutions to many problems and challenges that we face in pronunciation instruction.

The chapter divides into two major parts: in the first one, problems and contexts for pronunciation instruction and accent issues are explored in comparison to other elements of the system of English (Section 2.2), and then from the perspective of pronunciation teachers, pronunciation learners and finally, fully proficient pronunciation learners who become users of English (Section 2.3). Finally, a linguistic phonetic approach to non-native accents is proposed as a possible solution to the problem of accents in pronunciation instruction (Section 2.4). The data used for the discussion come predominantly from the Polish context, which is treated here as representative for problems and challenges shared across different places where English is taught as an additional language.

2.2 Problems

Over the years, while teaching the pronunciation of English to non-native speakers – in my case Polish learners and users of English – I have begun to realize that there are two major problems that I need to solve: putting it in simple words, the first one is what to teach, and the second one – what to reach. These two problems may seem trivial for anyone involved in a sister-system element of the language, i.e. grammar. From the perspective of someone trained in teaching English as a Foreign Language (but not a specialist in grammar instruction, who will have further insights into the

problem), it seems that when we teach grammar, we know that we need to teach grammatical structures, starting from simpler / more frequent to the more complex / less frequent ones (most of the time we follow the order of grammatical elements suggested by the textbook we happen to be using). And unless we start looking into language variability, we accept the descriptions of the grammatical system without much controversy about the need to use e.g. the verb ‘to be’ in progressive forms, ‘s’ in third person singular verbs or single negation. In other words, we have a clearly defined point of reference with respect to ‘what to teach’; but more than that, we seldom have any doubts as to ‘what to reach’. As teachers then, whenever we hear a learner say ‘she ride’, ‘I talking’, or ‘I didn’t take nothing’ we react to these forms as ‘wrong’ and tend to correct them at an appropriate point during the lesson. However, if the learner uses ‘t’ rather than ‘d’ in ‘ride’, no aspiration in ‘talking’ or ‘take’, or a sound closer to ‘f’ than ‘th’ in the word ‘nothing’, our decision as to categorizing these as ‘wrong’ and correcting the learner may not be equally straightforward and obvious. In other words, even when we accept the description of Standard English as the reference point in pronunciation instruction, the decision as to the degree to which this model will be used in practice as the basis for correction is far from obvious. Thus, although there are many similarities between grammar and pronunciation instruction when discussed with reference to when to correct, the decision as to what to correct differs. There are many other parallels between grammar and pronunciation based on the crucial distinction between accuracy, system-based instruction and skill practice in teaching speaking. However, the questions ‘what to teach’ and ‘what to reach’ seem to be specifically true for pronunciation: it is in pronunciation instruction that the debate continues as to the model itself and the degree to which this model determines priorities and/or success in pronunciation learning. Problems and challenges related to these two crucial questions will be briefly considered in the following sections.

2.2.1 What to teach?

First the problem is ‘what to teach’. On the face of it, the answer is easy: when teaching pronunciation, we teach the sound system of English. Well, if only we could stop there, we could accept a description of the system of English and teach our learners – or rather help them learn – this system. The native-speaker model used in pronunciation classes would correspond to the description and the learner would be expected to recognize sound categories and contrasts and to (re)produce them in her (re)production. The ‘re’ element proves to be crucial here. Historically speaking, it was

with emphasis on communication and production rather than reproduction that this straightforward, well-structured approach to pronunciation started to be amended so as to incorporate the communicative needs of the learners. This led to a major shift, with ‘comfortable intelligibility’ (e.g. Kenworthy, 1987), and not native-like (or near-native) pronunciation treated as the goal in pronunciation instruction. Notice that unlike the previous approach, with the variability introduced by the term ‘comfortable’ as well as ‘intelligibility’ (with the question of intelligible for whom?), the communicative approach has not provided clear answers as to the ‘what to teach’ question, as it is clearly assumed that not all elements of the system are equally important and not all of them need to be taught or learned. However, which ones they are has not been experimentally determined and remains obscure. The only set of pronunciation priorities for non-native speakers formulated to-date was proposed by Jennifer Jenkins (2000, 2002) in her intuition and observation rather than research-based² Lingua Franca Core (LFC). One of the problems with LFC is that it sanctions, or indeed advocates ‘mispronunciations’, such as the lack of a dental fricative or weak forms, with the latter observed to hinder communication between non-native speakers (ibid.). The suggestion that certain elements of native English pronunciation can be abandoned in pronunciation teaching has divided teachers of English pronunciation into two camps: those who adhere to it and try to put it into practice (e.g. Walker, 2005) and those who oppose it on linguistic (e.g. Szypra-Kozłowska, 2005) or logical (e.g. Sobkowiak, 2005) grounds. Not surprisingly, the lack of solid research and a somewhat patronizing approach behind LFC (interpreted as ‘let them use bad English’) has provoked a strong negative reaction among many non-native teachers of pronunciation, who feel that (1) LFC does not address problems of their learners (as voiced by e.g. Szypra-Kozłowska, 2005) and, importantly, find it a frustrating way of rejecting their own experience of being able to overcome many pronunciation problems. Teachers of English pronunciation who are not native speakers of English tend to share with their learners not only L1, but also the learning experience; they may rightly feel that they know best how to teach pronunciation to the specific L1 learners.

As a compromise between ELF and a traditional native-speaker model, Szypra-Kozłowska in her recent book (2015) advocates the idea of NELF, i.e. Native English as a Lingua Franca, originally proposed in a slightly

² I am grateful to an anonymous reviewer for noticing and bringing to my attention a possible misinterpretation of LFC for an experimentally determined set of priorities, which it definitely is not.

jocular manner³ by van den Doel (2008). What they suggest is going back to a native speaker model, RP or GA (with preference for RP), while accepting the fact that non-native speakers may not reach the native-like pronunciation level. Thus the question ‘what to teach’ receives the answer ‘native model’ once again. However, the novelty of NELF is in the outcomes of pronunciation instruction, with learners not expected to master all aspects of a native phonetic model.

The NELF approach is intuitively adopted by an overwhelming majority of EFL teachers who, being predominantly non-native speakers of English and speaking it with some degree of a foreign accent, in their pronunciation instruction take a native phonetic model as a reference point, but focus on selected features of English which they consider particularly important for successful communication. They do not require their students to master every phonetic detail as they know very well that in this respect perfection is impossible to achieve. (Szpyra-Kozłowska, 2015: 27).

The same sentiment seems to guide Trudgill, when he voices his cautious support for native English models:

(...) the sensible, pragmatic course is to continue, as before, employing ENL models [...] with an understanding that in most cases phonetic accuracy is unlikely to be achieved. (Trudgill, 2005: 93).

The above mentioned understanding that ‘accuracy’ or ‘perfection’ is unlikely or even impossible to be achieved, sensible and practice-based as it may be, does not seem to help in deciding what it really is that phonetic instruction aims to achieve. It seems that adopting a native English model and then accepting (understanding) the fact that it will not be mastered leaves us with a huge grey area, where deciding what to teach, what to correct, and what to accept as ‘correct’ is left for teachers to decide. In other words, even if we solve the problem of ‘what to teach’, we are left with ‘what to reach’.

2.2.2 What to reach?

Once again, the problem is true for pronunciation rather than grammar, where it is difficult to imagine a controversy as to whether to treat ‘She ride’ as less or more acceptable than ‘I talking’ or ‘I didn’t take nothing’. In the case of pronunciation, on the other hand, the reaction of the teacher

³ van den Doel, personal communication.

to different mispronunciations may vary in relation to (1) the degree to which they may agree with LFC, (2) the degree to which they are familiar with their learners' pronunciations, and finally (3) their own language and pronunciation beliefs. The pronunciation of 'th' is a good example here: with the pronunciation of 'th' not included in the LFC, i.e. claimed not to be relevant for communication between non-native speakers. Thus, the teachers' lack of correction in the case of the substitution of 'th' for 'f' may be expected to follow the LFC. At the same time, however, as the substitution is typical for Polish learners of English, so it may either go unnoticed on familiarity grounds, or conversely, a Polish teacher may find it annoying and correct it even though it is not in any way detrimental to intelligibility. Let us also notice that a different mispronunciation of 'th' may provoke a different reaction of the same teacher, e.g. a Polish teacher is much more likely to correct the 't' or 's' version of 'th'. Thus, not only is it possible for different mispronunciations of the same sound to be evaluated differently, but also to predict a different reaction to the same mispronunciation, with none of them likely to cause intelligibility problems. As already mentioned, the reaction of the teacher is likely to be largely determined by her language / accent attitude and teaching / learning experience (the teacher may pronounce 'th' as 'f' herself, she may be aware of the fact that 'f' is characteristic for a Polish accent and/or she may have the experience with 'f' going unnoticed in communication more easily than 's' or she may assume that 'f' is 'better' than 't' or 's' because this is the pronunciation she hears most often in the Polish context). The same is true in the case of other potential pronunciation difficulties in the examples quoted above: the pronunciation of 't' in 'talking' or 'take' without aspiration, 'ng' in 'talking' with a velar nasal, as 'ing', 'in' or 'ink'; and finally, a more problematic case of final consonant devoicing in 'ride', with a potential intelligibility problem ('ride' vs. 'write').

Clearly, when considering possible answers to the 'what to reach' question, we face problems with whether to aim towards a native-like pronunciation or not, and, more importantly again, if not – which (mis)pronunciations can be treated as acceptable. The suggestions formulated by Trudgill (2005) or Szypra-Kozłowska (2015) are too vague to be of real use. However, one aspect that they mention can be taken as a link-up to the possible solution to the problem suggested in this paper: the aspect of phonetic detail. More specifically, we will look at the ways in which phonetic detail or phonetic accuracy can be defined and applied to pronunciation instruction (see Section 2.4). Before the solution proposed to major problems with accents in L2 is put forward, however, we will briefly explore the attitudes to the issue of accent among teachers, learners

and those advanced language learners who have become language users in their life experience of immigration to English-speaking countries.

2.3 Accent in teaching, learning and using English

The key issue explored in this section is the understanding of the term accent and the attitude to it among teachers, learners and users of English. This brief presentation of the results from several surveys conducted in Poland and across Europe aims to show that while the term accent is associated with native speech in instructed pronunciation learning, non-native accents become an accepted reality in the actual use of English. This does not mean, however, that all pronunciations are accepted; on the contrary, with teachers commenting on the need for non-native accents to be clearer, learners accepting L1 features only when they reach a high level of language proficiency and advanced users of English realizing the importance of cross-accent experience for reaching ‘comfort’ in communication in English, there is a need for a clear reference-point for what a clear non-native accent may mean.

2.3.1 Teachers’ views

The attitude towards accents and the role of accents in the process of pronunciation instruction as seen by the teachers seems a natural starting point for the discussion of accents in the educational setting – after all, although the learners’ views are of paramount importance for the learning process, it is the teacher who tends to set the original directions in the process. The data discussed below come from the survey conducted among teachers in Europe between 2012 and 2013 as the result of collaborative work of researchers from 7 countries: Finland, France, Germany, Macedonia, Poland, Spain and Switzerland. Results obtained from the European Pronunciation Teaching in Europe Survey (EPTiES) have been reported in a number of publications (e.g. Henderson et al., 2012, 2013; Kirkova-Naskova et al., 2013; Waniek-Klimczak, 2013). The survey explored numerous issues related to pronunciation teaching; for the present discussion, it is only the issue of accent as referred to in teachers’ comments and their answers to the questions regarding the use of accents in the classroom that will be taken into account.

Respondents to the survey come from the European countries mentioned above (843 responses), with an average teacher aged 43 and having 12 years of teaching experience (median-based results). The majority of respondents are non-native speakers of English (91%), they

teach teenagers (56%), young adults (18%) and children (16%); they believe pronunciation to be fairly important (3.77 on the 1 to 5 ‘very important’ scale), and relatively easy to teach (3.15, scale 1–5 ‘extremely easy’). What seems essential is that almost all of them say they do teach pronunciation in their courses (96.6%). When commenting on the amount of time devoted to pronunciation teaching, however, some of them stress that it does not need to be taught independently, e.g.

[475]: I think pronunciation is just a part of the skill of speaking.

[164]: (...) whenever we talk we teach pronunciation.

The issue of accent appears in the comments regarding the importance of pronunciation and the aims for pronunciation teaching, when teachers say that accents and occasional errors in production are not a problem as long as a learner is able to communicate (e.g. respondent [56]); however, as noticed by respondent [447], a strong accent, especially when unfamiliar to the interlocutor, may be detrimental to intelligibility. This leads to an interesting formulation of the main goal for pronunciation instruction: a clearer non-native accent.

[56]: (...) although pronunciation can be important, I believe that building the students self-confidence and fluency is more important. As long as they are able to communicate and to be understood, I do not believe accents and occasional production errors are a particular problem.

[447]: I believe that more emphasis should be put on pronunciation for understanding, that is not native-like pronunciation, but clearer non-native pronunciation. I have realized that it is easier to understand a person who makes grammatical errors than it is to understand a person with a strong unfamiliar accent.

This final comment ([447]) includes two elements crucial for the line of reasoning developed here: firstly, accent familiarity is stressed as the basis of understanding, and secondly, a (clearer) non-native accent is accepted as the aim of instruction. What this comment leaves unspecified are the criteria for clarity, another vague concept often used in connection with pronunciation teaching – the problem that re-appears throughout the discussion of what to teach and what to reach (see Section 2.2).

The distinction between native vs. non-native accents has been partly explored in EPTiES in the section concerned with the explicit accent choice for productive and receptive work in the case of the teachers and their perception of their learners’ preferences. This part of the survey asked for the choice of possible options, ranging from well-established

reference accents (British ‘RP’ English or American English) through other national varieties of English to a ‘Type of International English’, ‘Another Variety’ and ‘No Preference’. Given this choice, a vast majority of respondents marked standard reference accents (with a clear preference for British ‘RP’ English in the case of the teachers’ choices – 53% for receptive, 50% for productive work, compared to 45% and 36% in for American English – and a slight preference for American English in the expected learner choices, with 41 and 40% for RP and 41.3 and 39% for American English, respectively) as crucial for both receptive and productive work, with the figure regularly higher for receptive than productive work. The responses prove that other national varieties are also often used by the teachers, mostly for receptive work (67% declare using them for receptive work), as is a rather vague ‘A type of International English’ (16% for receptive vs. 9.4% for productive work). Interestingly, it is this category that receives similar ratings from the teachers regarding their own choices and the choices they believe their students make – about 9.4% believe ‘International English’ is chosen not only by them, but also by their students. ‘Another variety’ which could possibly stand for a specific non-native accent (e.g. ‘Polish English’), receives a negligible support, with slightly higher figures in the teachers’ preferences than their expectations of their students’ choices (2.3% for receptive and 0.7% for productive work for teachers, as compared to 0.95% and 1.2% respectively for the learners).

From the perspective of a non-native accent, the EPTiES results reveal very little: on the one hand, teachers voice their preference for a communicative, intelligibility-based approach to pronunciation, on the other hand, however, they seem to stick to native-based models not only for receptive but also for productive work with their students. With ‘International English’ used in EPTiES as a cover term for a possibly ‘clear’ non-native variety receiving a weak support and ‘Other varieties’ (separate from ‘Other national varieties’) not used at all, non-native accents seem not to have entered the scene yet. In other words, work on a clearer version of a specific L2 accent, as manifested by its fully proficient speakers, is not what teachers seem to believe worth doing.

2.3.2 Learners’ views

Not surprisingly, perhaps, when asked directly about their preferences, learners do not seem to differ much from their teachers’ perception of their needs. More specifically, when asked about their goals in pronunciation instruction, they mention fluency, intelligibility, and self-confidence in

speech (Waniek-Klimczak, 1997, 2011a). Students also regularly declare pronunciation to be more important than accent (Waniek-Klimczak, *ibid.*), indirectly supporting the implicit understanding of pronunciation as a way of using speech regardless of an accent, the concept not far from the ‘imperfect’ but ‘correct’ pronunciation advocated by Trudgill (2005) or Szypra-Kozłowska (2015). At the same time a native speaker model remains valid, at least at the level of ‘wishful thinking’ – when asked whether they would like to speak like native speakers of English, Polish students say ‘yes’ (e.g. Janicka, Kul & Weckwerth, 2005; Waniek-Klimczak & Klimczak, 2005); however, answering the question about the degree of confidence in the possibility of reaching such a goal, they differ, with students majoring in English more optimistic (and determined) than other advanced learners (Waniek-Klimczak & Klimczak, 2005). The goals for pronunciation instruction voiced in one of the studies (Waniek-Klimczak, 2011a) include a frequently repeated ‘*I want to be easily understood*’, the sentiment not unlike the one found in the teachers. Interestingly, one of the comments also includes an explicit mention of an L2 accent, with a student saying ‘*I don’t want to have a strong Polish accent*’.

Negative attitudes towards their L1-based accent in English, i.e. the accent typical for the L1 background that respondents share, have been found in a number of studies (e.g. Dalton-Puffer, Kaltenboek and Smit, 1997; Waniek-Klimczak, Porzuczek and Rojczyk, 2013, 2015). The view that it is an accent that decides about the way a specialist in English (a teacher or a translator) will be perceived with respect to the level of professional expertise re-appears regularly in the study of English majors (Waniek-Klimczak et al., 2015), supporting earlier findings suggesting that the more advanced the learners, the stronger the desire to speak like native speakers (Dalton-Puffer, 1997; Smojver and Stanojević, 2013). However, the observed tendency to strive towards native-like pronunciation in parallel to an increased level of language proficiency has been found to give way to a less extreme position with a growing experience: in Waniek-Klimczak et al. (2015), attitudes towards Polish pronunciation features in English proved sensitive to the level of education (BA vs. MA) and gender, with MA students caring less about their English pronunciation not having features characteristic for Polish than their BA colleagues ($n = 507$; BA = 393, MA = 114; $F(1, 505) = 4.762, p < 0.05$) and females more critical of their pronunciation (i.e. claiming they have more Polish features in their English pronunciation, with n for females = 403, males = 98, $F(1, 499) = 5.693, p < 0.05$). While this observation supports the tendency for females to be more concerned with correctness,

the overall tendency for more advanced (but also older) students to be more cautious in claiming that their pronunciation does not contain Polish features interacts with the tendency to voice smaller concern about correctness and care for not having Polish features in their English pronunciation.

2.3.3 Language users' views

Students majoring in English form a specific group of language learners: highly motivated, they have selected this language as their main tool and subject of enquiry. Naturally, then, it is in this group that we expect to find the strongest support for native-like pronunciation as the goal of pronunciation instruction. What seems interesting, however, is the effect of experience and maturation, which seems to make both the goals and the assessment of one's English pronunciation more lenient towards the acceptance of L1-based accent. In the study cited in 2.3.2. (Waniek-Klimczak et al., 2015) the effect was observed in the difference between BA and MA students investigated through a large-scale, cross sectional design. Continuing the discussion of the attitudes found among experienced language learners, we move now to a specific group of respondents: former students enrolled at BA or MA levels in English in Poland who decided to move for permanent residence to UK and continue their education there. A brief account of their attitudes and accent experience in a naturalistic environment presented below comes from a larger study, with some of the results reported in Waniek-Klimczak (2011b). The participants meet the following criteria: they are all English majors, graduated at the BA or MA level and immigrated to the UK after 2004; they continued tertiary level education in the UK (MA or PhD programmes), they worked in an English-speaking environment and at the time of the study they had the minimum of 6–7 years of experience in the UK. They were approached through a network of former students. The data were collected using a semi-structured interview conducted via Skype.

The key question of interest in the study was the clash between expectations and language skills on arrival with the communicative needs in a specific English-speaking environment, the experience of language in general and more specifically, the experience related to accents. As reported by one of the respondents, the beginning was often tough:

[M]: Well, there is a story about that actually, because when we first came to England there was quite a shock I must say, because that local accent is quite distinguishable and hard to understand. What we studied in Poland it

was mainly at school, very RP and formal, so when I started working in a call centre, well, obviously I could speak English but my friend said that I couldn't and what he meant was that my English was not natural, it was what I call 'bookish' English, so I lacked this fluency and this natural flow and now it has improved so much.

As the respondent came to the UK having graduated in English from a teacher training college (at the BA level), she can be assumed to have had a relatively high level of expectations with regards to her own ability to communicate in English; having been faced with local accents and the need to use informal style, she needed time to adjust. Similar experience is reported by another respondent, who stresses the fact that it was not only the local accent but also the L1 accent she had that made communication difficult.

[Mr]: I could experience the stages where after a couple of months it started to become easy to understand people and then after a couple of years you sort of get to pick up the local vocabulary, accent and it is just easier to speak, to listen and to be understood, and it is important because many people when we came couldn't understand us because our accents were so strong.

Reflecting on initial difficulties, respondents also mention the effect of experience in the UK on their use of English. In most cases, what they find most important is a high level of proficiency that they have achieved in all skills; additionally, however, some respondents also mention the intricate connection between how they sound and who they are:

[S1]: Regardless of the degree I had in Poland my listening comprehension was really bad, the phonetics and vocabulary that I had..... It was shocking, for the first couple of months I could barely understand anyone and English I used seemed unreal, with proper RP and proper intonation I found it strange and after a while I had to embrace the fact that I'm not an English person and I will always speak with an accent. (...) However long we live here, I'm just the girl with wavy hair and with a funny accent and I will always be a foreigner here, however good I feel in here.

Speaking with an accent as an inevitable effect of our language experience in the process of language acquisition may seem obvious, but in the context of L2 accents it gains socio-cultural consequences that go far beyond intelligibility. In the case of the respondent quoted above, an accent is a part of her identity, a part of her image, the way she looks and

the way she speaks are naturally interconnected, making her different from those around her.

Believed inevitable by fully proficient, highly experienced learners of English functioning in a natural context, a foreign accent seems to be a reality that all learners of English need to embrace. The fact that it is the most experienced learners who tend to realize that an L1 based accent in their L2 is there to stay, suggests that a higher level of language awareness as well as life experience are needed to accept the inevitability of accentedness. Notice that it does not mean that other priorities, including the wish to be easily understood, change in any way; much rather, we can talk about the effect of maturation, which leads to the acceptance of certain features of L1 in L2 as long as they do not interfere with easy communication. Which features these are remains an open question. As will be argued in Section 2.4, what might be hidden behind an often-mentioned ‘clear’ foreign-accented speech could be the pronunciation based on the recognition of all sound categories used in the target (foreign/second) language (L2), with phonetic realizations not matching the expected ones and thus forming a foreign accent.

2.4 Possible solution: Linguistic phonetics approach

Describing accents at the level of phonemic contrasts and details of their phonetic realization has a long history, with the idea of Trubetzkoy taken up and employed to the classification of English accents by Wells (1982). Gimson (1962) refers to the two levels defining performance targets for the learners of English. Distinguishing between two extremes, he talks about ‘minimum general intelligibility’ as the lowest requirement for approximating to native speech and defines it as possessing “a set of distinctive elements which correspond in some measure to the inventory of the RP phonemic system” (p. 316). At the other extreme Gimson (*ibid.*) puts ‘high acceptability’, defined as

(...) a form of speech which the native listener may not identify as non-native, which conveys information as readily as would a native’s and which arrives at this result through precision in the phonetic (allophonic) realization of phonemes and by confident handling of accentual and intonational patterns. (Gimson, 1962: 316).

The difference in the targets corresponds then to the degree to which the phonemic inventory and its phonetic realization couched in ‘confidently used’ organization of speech at the suprasegmental level matches the native speakers’, with intelligibility related to the phonemic level and

native-like pronunciation with the phonetic one. It is interesting to notice that for Gimson (*ibid.*) intelligibility and high acceptability are elements of a continuum formed by different degrees to which a learner approximates to native English speech; the contrast is made between making this attempt *vs.* not making it at all, and employing the phonological and phonetic system of the learner's own language. From the perspective of further research in the field (e.g. nativeness *vs.* the intelligibility principle formulated by Levis, 2005), this view has the advantage of stressing an obvious relationship between nativeness and intelligibility, with the latter defined as a limited degree of the former. A different degree will be claimed here to correspond to the level of approximation, i.e. the extent to which native-English sound categories are employed by the learner. However, rather than talking about abstract categories (phonemes), and their phonetic realizations (allophones), we propose to refer to the phonetic reality of speech and to re-define the targets for pronunciation learning and teaching in terms of the values of phonetic parameters needed in implementation of language-specific contrasts.

Approaching learner pronunciation from the perspective of phonetic parameter values we follow linguistic phonetics (Ladefoged, 1971, 1997), which focuses on contrasts observable at the systematic phonetic level. Adopting this perspective, we concentrate on language-specific contrasts (language categories) at the level of phonetic detail, i.e. phonetic parameters used in realization (implementation) of the categories in speech. The phonetic parameters are articulatory gestures (and their representation in the acoustic signal) used in signalling the categories by language users. As parameters are physical in nature and gradient, their values form a continuum, the range of which is limited by the category boundary on the one hand, and a specific language experience / community preference on the other. Within the range of values, we can talk about the expected ones in a given context, i.e. a certain range of values will be interpreted as fitting a given context best. The 'goodness of fit' criterion, often mentioned in perception studies (e.g. Ganong III, 1980; Bohn & Best, 2012) is used here for the production data interpreted in correspondence to the native speaker range of values for individual contexts. Thus, adopting a linguistic phonetics approach to non-native speech, we assume that statistically significant contrast in the values used for a given category is sufficient to signal the category; however, the actual values of the parameter are assessed for the goodness-of-fit on the basis of native-speaker parameter values (see Waniek-Klimczak, 2005 for an application of 'goodness-of-fit' to production).

The modal value interpretation makes it possible to re-interpret the model *vs.* target dilemma by referring to the values of phonetic parameters, with the model defined in terms of contrasts observable at the systematic phonetic level in the reference system and the range of values of phonetic parameters used in the implementation of the contrasts as specified and measured for the model. The target, on the other hand, is defined as a set of values for parameters used in the implementation in the reference system with a wider range of values within each category. The target then relies on the implementation of the contrasts, not the use of values specified for the model.

The approach advocated here hopes to offer a direction for solving the problem of the aims for pronunciation instruction and the teaching and learning practice, which requires clear guidelines as to what to accept for the learner and teacher alike. The terms that need to be defined for practical usage include not only model and target, but also ‘clear’, ‘good’ and ‘bad’, and, most importantly for the school context, ‘correct’. In the approach proposed here, ‘clear’ and ‘correct’ is interpreted with reference to system-specific categories, which means that when a learner employs English categories and maintains contrasts between them, his or her speech can be treated as correct. The degree to which the learner uses phonetic values expected as best-fitting in the realization of these contrasts decides about accentedness of his or her speech, and as long as the contrasts are used, does not affect ‘correctness’ as defined above. ‘Clear’, ‘good’ and ‘bad’, on the other hand, need to remain open to negotiation between teacher and learner (and interlocutors) as to the degree of approximation that is attempted in an instructed context and the degree of compatibility of values used by interlocutors in a natural context.

The practical application of the above assumptions rests on the availability of the data that could be used as the point of reference for teachers and learners alike. As phonetic parameters do not lend themselves easily to the school practice, what is needed are studies the results of which could be interpreted as guidelines for the teaching practice. The following sections attempt to illustrate the proposal formulated above with the data coming from an example of such a study, conducted among two groups of Polish-English bilingual with a different type of language experience (Early *vs.* Late Bilinguals) and reported in Waniek-Klimczak (2005). Two aspects of consonantal voicing: aspiration and vowel length are explored from the perspective of the realization of the phonetic category with the use of the phonetic parameter of Voice Onset Time (VOT) and vowel length.

2.4.1 Data analysis 1: Aspiration

A brief description of the sound systems of English and Polish shows that within the system of plosives, it is the difference in the implementation of voicing in stressed syllable initial voiceless, fortis plosives /p,t,k/ that requires the formation of a new sound category [voiceless aspirated]. As Polish uses two categories: [voiced] vs. [voiceless unaspirated]⁴, and English apart from the above two includes [voiceless aspirated] (Keating et al., 1981), it is this category that will be explored here.

The phonetic parameter commonly used for the measure of the category [aspirated] is Voice Onset Time (Lisker and Abramson, 1964); its mean values as found for Polish and English in earlier studies are shown in Table 2-1. With the boundary for category [aspirated] set at around 30 ms for bilabials and alveolars and 50 ms for velars, the data suggest that it is only in the case of /k/ that Polish speakers may get close to aspiration. Even when longer VOT is used, as for /k/, a systematic lengthening of the VOT needs to be learned in order to implement the contrast in voicing typical for English⁵. However, the degree of this lengthening may vary, with the tendency for mid-values, i.e. values intermediate between the ones most typical for L1 and L2, attested in many studies (e.g. Flege and Eefting, 1988; Flege, 1991; Waniek-Klimczak, 2005).

	/p/	/b/	/t/	/d/	/k/	/g/
Polish	22	-88	28	-90	52	-66
English	59	12	67	19	84	25

Table 2-1. Voice Onset Time in Polish and English. Data in milliseconds, based on Keating et al. (1981); Lisker & Abramson (1964); Waniek-Klimczak (2005).

Several experiments conducted on Polish speakers of English report success with respect to the tendency to use the VOT significantly differently to mark the difference in the category (e.g. Waniek-Klimczak, 2009), however, as predicted, the studies also report the use of shorter values than expected in the implementation of [voiceless aspirated] (e.g. Waniek-Klimczak, 2005, 2009, 2013c). While the VOT values vary in correspondence to language experience of the learners, they also exhibit regularity with respect to relative lengthening dependent on the place of

⁴ Aspiration has been found to have an emphatic function in Polish (Waniek-Klimczak, 2011d).

⁵ This expectation is based on the use of aspiration in reference accents, i.e. RP or GA.

articulation of the plosive and the following vowel⁶. Fig. 2-1 illustrates the variability in the mean values for the VOT produced by native speakers of English (NS), highly proficient and experienced speakers of English (EB) and less proficient/experienced Polish users of English (LB, data from Waniek-Klimczak, 2005).

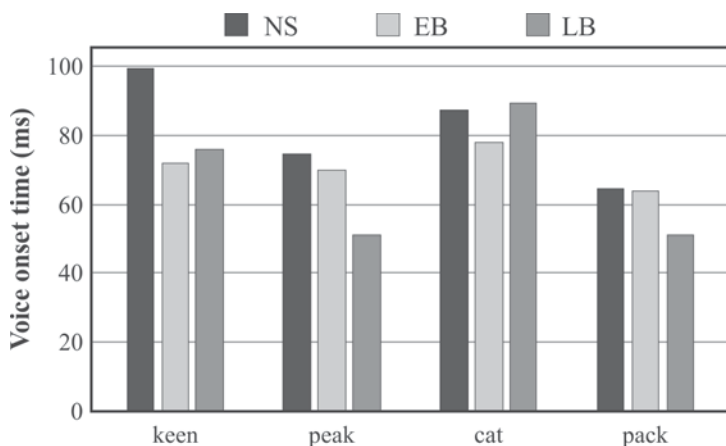


Fig. 2-1. Mean VOT values (in milliseconds) for three groups of respondents: Native Speakers of English (NS=5), Early Bilinguals (EB=10), Late Bilinguals (LB=28).

The four words, *keen*, *peak*, *cat*, *pack* have been produced in carrier sentences; the predicted length of the VOT values, as illustrated by the results for the NS group, is expected to be longer in the case of velars than bilabials and longer in the case of the following of high than non-high vowel. The results show that while the values produced by native speakers of English follow a predicted pattern, the ones produced by non-native speakers vary, with the less proficient LB speakers producing values closer to the expected ones only in the case of a word *cat*. Clearly, then, there are other factors than the place of articulation and the quality of the following vowel that affect the degree to which a given category is implemented with the values close to those produced by naive speakers. Word familiarity and the perceptual salience of aspiration at a particular word level may affect the modal values. What is crucial for the present

⁶ The difference in the VOT value corresponding to the place of articulation and the height of the following vowel has been noticed across languages and included into the inventory of phonetic universals (Maddieson, 1997).

model, however, is the fact that the category [voiceless aspirated] is regularly implemented, so the level of ‘correctness’ has been reached. Accentedness is another matter – it varies across respondents and will be decided on with reference to the degree to which this category is implemented in the way corresponding to the modal values expected in particular contexts.

	<i>KEEN</i>			<i>PEAK</i>			<i>CAT</i>			<i>PACK</i>		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
NS	100	38	75 – 166	75	17	59 – 101	88	7.8	75 – 94	65	11	51 – 77
EB	72	14	51 – 96	70	18	43 – 111	78	14	58 – 102	64	17	46 – 109
LB	76	28	36 – 107	51	25	16 – 119	90	40	52 – 270	51	27	15 – 127

Table 2-2. VOT data (mean values, standard deviation (SD) and the range of values from the minimum to maximum) for three groups of respondents: Native Speakers of English (NS=5), Early Bilinguals (EB=10), Late Bilinguals (LB=28).

The comparison of the data with respect to standard deviation and the range of values produced within each group (see Table 2-2) reveals considerable differences across the respondents with respect to the range of the values used in the implementation of the category in each word and the degree of variability (as illustrated by standard deviation). With the exception of the word *keen*, where the group of native speakers proves to be most variable, there seems to be a regular tendency for the less proficient group of Late Bilinguals to exhibit the highest degree of variability; the observed variability results from a wide range of values produced in this group and the lack of consistency, characteristic for a less stable system. Consequently, the degree of accentedness varies, with the shortest modal values used in this group not reaching the minimal VOT level for aspiration (i.e. 30 ms). Across words, it is the word *cat* that is consistently pronounced with the longest VOT.

Overall, the data suggest that while a number of factors affect the values of the phonetic parameter in individual words, it is the tendency to

implement the category [voiceless aspirated] across the respondents that is shared as an element of their pronunciation of English. The degree to which the values used approximate the expected ones differ, and it is the more proficient group of Early Bilinguals that regularly produce less ‘accented’, i.e. closer to the modals values produced by the native speakers. For the purpose of a practical application of the reasoning developed here, it is only the cases of the shortest VOT that require teachers’ intervention at the level of correctness of pronunciation.

2.4.2 Data analysis 2: Vowel duration

Vowel duration represents a special case in the system of English, as it is used in the realization of vowel as well as consonant categories, as in *heed* vs. *hid* and *heat* vs. *hit* for vowel category difference and *heed* vs. *heat* and *hid* vs. *hit* for consonantal voicing (see e.g. Raphael, 1972; Luce & Charles-Luce, 1985).

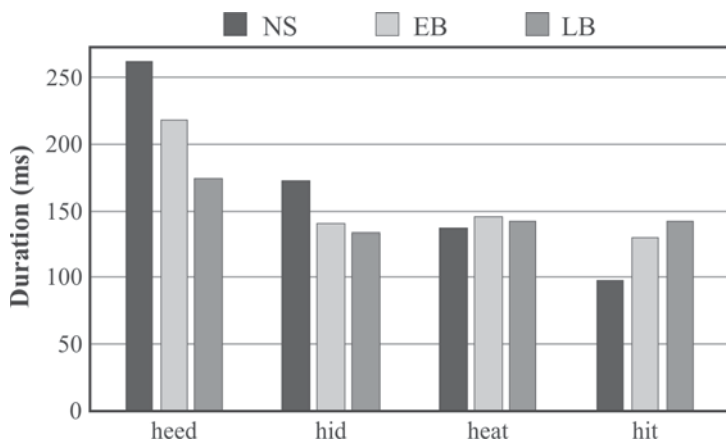


Fig. 2-2. Mean vowel duration values for three groups of respondents: Native Speakers of English (NS=5), Early Bilinguals (EB=10), Late Bilinguals (LB=28).

When the four words are put in the order of vowel duration difference on the basis of the data obtained from native speakers of English (see Waniek-Klimczak, 2005 for a full report on the study), it is the effect of the following consonant that overrides the inherent length of the vowel, i.e. the vowel is expectedly the longest as *heed* and the shortest in *hit*, with the mid values longer for a lax vowel in a voiced consonant context in *hid* than for a tense vowel followed by a voiceless consonant in *heat*.

	<i>HEED</i>			<i>HID</i>			<i>HEAT</i>			<i>HIT</i>		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
NS	263	55	221 – 356	173	56	115 – 251	138	37	90 – 191	100	35	70 – 145
EB	220	50	150 – 322	141	28	112 – 207	147	32	101 – 185	131	39	98 – 216
LB	176	39	108 – 251	134	43	55 – 206	147	46	63 – 232	143	48	62 – 252

Table 2-3. Vowel duration data (mean values, standard deviation (SD) and the range of values from the minimum to maximum) for three groups of respondents: Native Speakers of English (NS=5), Early Bilinguals (EB=10), Late Bilinguals (LB=28).

As can be seen in Fig. 2-2, the pattern seems not to be followed by Polish-English bilinguals, who tend to rely on inherent length rather than context-dependent one, and shorten the lax vowel more than a tense one across contexts. Moreover, it is the lax vowel that seems not to be regularly sensitive to the following context of consonantal voicing: the mean values for *hid* and *hit* differ very little. This general observation is supported by statistical tests: t-tests for paired observations on the bilingual data show that it is the tenseness of the vowel that is realized regularly differently in respondents from both groups of bilingual speakers; the effect of the following voicing, although also present, is irregular, with some the less proficient bilinguals reversing the tendency, i.e. lengthening rather than shortening the vowel in the voiceless context for *hit* (see Waniek-Klimczak, 2005 for a full report, including tempo-of-speech adjustments). While there are many possible reasons for this irregularity, including transfer from Polish, where final consonants are always devoiced and consequently, no effect of vowel length difference is used (see e.g. Slowiaczek & Dinnsen, 1985; Jassem & Richter, 1989) and a word-familiarity effect, this observation suggests that the contrast is easier to implement in the case of longer durations, where its perceptual salience increases.

The overview of a range of values (Table 2-3) further supports the observation that it is a shorter vowel duration that causes greater difficulty, leading to the use of vowel duration as cue for consonantal voicing unsuccessful in the case of many less proficient speakers (LB). It is also in this group that the variability level increases in the expected shorter duration contexts, for *heat* and *hit*. This leads to problems with the realization of consonantal voicing, with unexpectedly longer vowel in *hit* than either *heat* or *hid* for some speakers from the LB group. From the correctness perspective, then, it is the use of shorter vowel durations that requires practice, with particular attention needed in the case of lax vowels in the context of the following voiced *vs.* voiceless consonants. This means the need for re-setting some of the priorities, with the lengthening of the vowels rather than their shortening receiving more attention in the teaching practice at the moment.

2.5 Conclusion

The line of reasoning proposed in this paper begins with a rather obvious observation that an accent is a natural consequence of language experience; thus, although non-native accents differ from native accents due to their users' multi-system (inter-language) experience, they need to be accepted as a reality by language teachers and language learners. The acceptance, however, does not mean abandoning the task of teaching and learning pronunciation; on the contrary, it is when we accept non-native accents as a natural consequence of language experience that pronunciation becomes teachable. Rather than formulating lists of priorities along the lines of LFC, choosing between nativeness *vs.* intelligibility, or talking about a mysterious level of inaccuracy or imperfection as Trudgill (2005) and Szpyra-Kozłowska (2015) do, what we propose here is a firm focus on the categories of the phonetic linguistic system of English. The need to realize all phonetic categories that have a linguistic function in the target language seems to be the best guarantee of communication; the ease of this communication, an often mentioned goal for pronunciation teaching and learning, will be further related to the values of phonetic parameters used in the realization of phonetic categories. In other words, what we propose is to organize pronunciation teaching (and learning) around phonetic categories, as they need to be produced to ensure intelligibility. The production of the categories relies on the use of phonetic parameters in the way that ensures contrasts between the categories. The degree to which this contrast is implemented, the range of values for individual phonetic parameters used in the phonetic

category realization depend on numerous factors; the range of values used by native-speakers using a given accent of English determine the range accepted as best-fitting for this accent in a given context. In other words, as the range of values used by non-native speakers may differ from the one used by native speakers, it is the requirement of implementing all categories that is the priority. The use of values closer or further away from the ones accepted as best-fitting by a given English-speaking community will determine the ease of communication with this community. What seems crucial here is the awareness of the model understood as the range of modal values as specified for the reference system; the target, on the other hand, is understood as the use of phonetic parameters that ensures realization of all phonetic categories, i.e. the use of phonetic parameters in a contrastive way. The implementation of contrasts does not mean that native-like range of values for phonetic parameters is required or indeed desired – whether a learner decides s/he wants to modify the values, by e.g. lengthening the VOT beyond 30 ms for /p/ to have stronger aspiration, or extending vowel length difference between *heat* and *heed* remains their choice. The nativeness – intelligibility dilemma becomes a matter of degree to which the modal values match; the ease of communication will be re-defined depending on the target community, and it is again in terms of the categories and the values of phonetic parameters that the target is best defined. With respect to a particular non-native accent, some contrasts may be more difficult, as e.g. vowel length contrasts between *hid* and *hit* in Polish learners of English, and it is important to identify them and to include them in pronunciation teaching. This identification requires further studies into individual accents of English, as the implementation of the proposal put forward here rests on a good understanding of the nature of phonetic parameter usage in each case. What we hope to have shown is that it is possible to use the values of phonetic parameters as an independent measure of ‘success’. And while individual teachers invariably take the decision of what to teach and what their students need to reach, having a firm basis for these decisions might offer help in this difficult task.

References

- Bohn, O.-S. & Best, C. T. (2012). Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. *Journal of Phonetics*, 40(1), pp. 109–128.

- Collins, B. & Mees, I. (2008, 2nd Edition). *Practical Phonetics and Phonology: A Resource Book for Students*. London: Routledge.
- Dalton-Puffer, Ch., Kaltenboeck, G. & Smit, U. (1997). Learner attitudes and L2 pronunciation in Austria. *World Englishes*, 16, pp. 115–127.
- Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, 89, pp. 395–411.
- Flege, J. E. & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America*, 83, pp. 729–740.
- Ganong III, W. F. (1980). Phonetic categorization in auditory perception. *Journal of Experimental Psychology*, 6, pp. 110–125.
- Gimson, A. C. (1962). *An Introduction to the Pronunciation of English*. London: Edward Arnold.
- Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Levey, D., Waniek-Klimczak, E., Cunningham, U., Curnick, L. (2012). The English pronunciation teaching in Europe survey: selected results. *Research in Language*, 10(1), pp. 5–28.
- Henderson, A., Tergujeff, E., Kirkova-Naskova, A., Kautzsch, A., Waniek-Klimczak, E., Levey, D., Frost, D. & Curnick, L. (2013). English Pronunciation Teaching in Europe Survey: Factors inside and outside the classroom. In: Mompeán, J. & Fouz, J. (Eds.), *Proceedings of the Third International Conference on English Pronunciation: Issues and Practices (EPIP3), May 8–10, 2013*, pp. 36–40. Murcia: University of Murcia.
- Janicka, K., Kul, M. & Weckwerth, J. (2005). Polish students' attitudes to native English accents as models for EFL pronunciation. In: Dziubalska-Kołodziej, K & Przedlacka, J. (Eds.), *English Pronunciation Models: A Changing Scene*, pp. 251–292. Frankfurt: Peter Lang.
- Jassem, W. & Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, 17, pp. 205–212.
- Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics*, 23, pp. 83–103.
- Keating, P., Mikoś, M. & Ganong III, W. (1981). A cross-language study of range of voice onset time in the perception of stop consonant voicing. *Journal of the Acoustical Society of America*, 70, pp. 1260–1271.

- Kenworthy, J. (1987). *Teaching English Pronunciation*. London: Longman.
- Kirkova-Naskova, A., Tergujeff, E., Frost, D., Henderson, A., Kautzsch, A., Levey, D., Murphy, D. & Waniek-Klimczak, E. (2013). Teachers' views on their professional training and assessment practices: Selected results from the English Pronunciation Teaching in Europe Survey. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference, August 2012*, pp. 29–42. Ames, IA: Iowa State University.
- Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustic measurements. *Word*, 20, pp. 384–422.
- Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 36, pp. 369–377.
- Lippi-Green, R. (1997). *English with an Accent*. London: Routledge.
- Ladefoged, P. (1971). *Preliminaries to Linguistics Phonetics*. Chicago: The University of Chicago Press.
- . (1997). Linguistic phonetic description. In: Hardcastle, W. & Laver, J. (Eds.), *The Handbook of Phonetic Sciences*, pp. 589–618. Oxford: Blackwell.
- Luce, P. A. & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America*, 78, pp. 1949–1957.
- Maddieson, I. (1997). Phonetic Universals. In: Hardcastle, W. J. & Laver, J. (Eds.), *The Handbook of Phonetic Sciences*, pp. 613–639. Oxford: Blackwell.
- Major, R. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaum.
- Raphael, L. J. (1972). Preceding vowel duration as a cue to the voicing characteristics of word-final consonants in English. *Journal of the Acoustical Society of America*, 51, pp. 1296–1303.
- Slowiaczek, L. & Dinnsen, D. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics*, 13, pp. 325–341.
- Smojver, V. J. & Stanojewić, M.-M. (2013). Stratification as a Lingua Franca: Identity construction of learners and speakers. In: Waniek-Klimczak, E. & Shockey, L. (Eds.), *Teaching and Researching English Accents in Native and Non-native Speakers*, pp. 191–206. Berlin: Springer.
- Sobkowiak, W. (2005). Why not LFC? In: Dziubalska-Kołodczyk, K. & Przedlacka, J. (Eds.), *English Pronunciation Models: A Changing Scene*, pp. 131–149. Bern: Peter Lang.

- Szpyra-Kozłowska, J. (2005). LFC, phonetic universals and the Polish context. In: Dziubalska-Kołodziej, K. & Przedlacka, J. (Eds.), *English Pronunciation Models: A Changing Scene*, pp. 151–176. Bern: Peter Lang.
- (2015). *Pronunciation in EFL Instruction. A Research-Based Approach*. Clarendon: Multilingual Matters.
- Trudgill, P. (2005). Native-speaker segmental phonological models and the English Lingua Franca Core. In: Dziubalska-Kołodziej, K. & Przedlacka, J. (Eds.), *English Pronunciation Models: A Changing Scene*, pp. 77–98. Bern: Peter Lang.
- Van den Doel, R. (2008). The blind spots of Jenkins's Lingua Franca. In: Waniek-Klimczak (Ed.), *Issues in Accents of English*, pp. 140–149. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Waniek-Klimczak, E. (1997). Context for teaching English phonetics and phonology at Polish Universities and Colleges: A survey. In: Waniek-Klimczak, E. (Ed.), *Teaching English Phonetics and Phonology II: Accents '97*, pp. 5–17. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- (2005). *Temporal parameters in second language speech: An applied linguistic phonetics approach*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- (2009). Category implementation in non-native speech: The case of aspirated plosives in Polish-English bilinguals. In: Oleksy, W. & Stalmaszczyk, P. (Eds.), *Festschrift for Barbara Lewandowska-Tomaszczyk*, pp. 497–512. Frankfurt: Peter Lang.
- (2011a). 'I am good at speaking, but I failed my phonetics class': Pronunciation and speaking in advanced learners of English. In: Pawlak, M., Waniek-Klimczak, E. & Majer, J. (Eds.), *Speaking in Contexts of Instructed Foreign Language Acquisition*, pp. 117–130. Clarendon: Multilingual Matters.
- (2011b). Acculturation strategy and language experience in expert SLA speakers: An exploratory study. *Studies in Second Language Learning and Teaching*, 1(2), pp. 227–246.
- (2011c). Aspiration and style: A sociophonetic study of the VOT in Polish learners of English. In: Dziubalska-Kołodziej, K., Wrembel, M. & Kul, K. (Eds.), *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*, pp. 301–314. Frankfurt: Peter Lang.
- (2011d). Aspiration in Polish: A sound change in progress? In: Pawlak, M. & Bielak, J. (Eds.), *New Perspectives in Language, Discourse and Translation Studies*, pp. 3–11. Berlin: Springer-Verlag.

- . (2013). It's all in teachers' hands: The English Pronunciation Teaching in Europe Survey from a Polish perspective. In: Gabryś-Barker, D., Piechurska-Kuciel, E. & Zybert, J. (Eds.), *Investigations in Teaching and Learning Languages*, pp. 245–260. Heidelberg: Springer Verlag.
- Waniek-Klimczak, E. & Klimczak, K. (2005). Target in speech development: Learners' views. In: Dziubalska-Kolaczyk, K. & Przedlacka, J. (Eds.), *English Pronunciation Models: A Changing Scene*, pp. 229–250. Bern: Peter Lang.
- Waniek-Klimczak, E., Porzuczek, A. & Rojczyk, A. (2013). Affective dimensions in SL pronunciation: A large-scale attitude survey. In: Gabryś-Barker, D. & Bielska, J. (Eds.), *The Affective Dimension in Second Language Acquisition*, pp. 124–137. Bristol: Multilingual Matters.
- Waniek-Klimczak, E., Rojczyk, A. & Porzuczek, A. (2015). 'Polglish' in Polish eyes: What English studies majors think about their pronunciation in English. In: Waniek-Klimczak, E. & Pawlak, M. (Eds.), *Teaching and Researching the Pronunciation of English*, pp. 23–34. Berlin: Springer-Verlag.
- Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

CHAPTER THREE

SHORT- AND LONG-TERM EFFECTS OF PRONUNCIATION TEACHING: EFL LEARNERS' VIEWS

PEKKA LINTUNEN AND ALEKSI MÄKILÄHDE

3.1 Introduction

Mastery of the pronunciation of a second language (L2) is an integral part of successful L2 learning. For decades, researchers have been interested in contrastive phonetics and in how language systems differ. Similarly, the production of the target language (TL) sounds by non-native learners has been the focus of many studies. Considering the role of English around the world, it is hardly surprising that the TL in these studies has often been English. As a result, there is a lot of information available on the learning processes and learning difficulties of non-native learners of English with varied background factors, such as the native language (L1). Earlier studies have also shown that formal teaching facilitates learning. Despite the research results available, the concern often raised by phoneticians and pronunciation teachers is that pronunciation skills are a neglected element at schools and in L2 teaching in general (e.g., Derwing, 2009). Studies on the effectiveness of formal instruction often focus on skills and may be interested in short-term or long-term effects (e.g., Couper, 2006; Saito, 2012; Thomson & Derwing, 2015). In this study, we were also interested in the short- and long-term effects, but instead of skills we focused on affective factors.

For a comprehensive understanding of the learning process of L2 pronunciation, we must consider the learner as an individual and learners' own views that have an effect on the learning outcomes. Learner beliefs play an important role in foreign language learning, and their importance has been acknowledged in Second Language Acquisition (SLA) studies

(see, e.g., Kalaja & Barcelos, 2013).¹ Furthermore, SLA researchers have suggested that there is a need for studies that take into account the dynamic nature, development or transformation of beliefs. The purpose of this study was to focus on learner beliefs about L2 pronunciation and how formal teaching affects these beliefs both directly after teaching and after a longer interval of time. We were interested in the effects that pronunciation teaching has on TL awareness, general attitudes towards TL pronunciation and its importance, learner confidence and goals, and the development of TL pronunciation skills as experienced by advanced university learners of English.

3.2 Pronunciation instruction and learner beliefs

3.2.1 The effectiveness and goals of pronunciation instruction

SLA studies have been interested in the effects of instruction on L2 learning. The learning of a new L2 is a complex developmental process in which learners vary in their rate and success, whereas the route or stages of development remain fairly similar between learners. In the case of foreign languages, it is common that learning takes place both naturalistically outside classrooms and formally in teacher-led classroom contexts. Naturalistic learning is possible if the learner receives TL input or is otherwise able to use the TL outside classrooms. Earlier research has shown that formal instruction has a positive effect on the rate and ultimate success of learning (see, e.g., Ortega, 2009), while the effects of instruction can be short-term or long-term (Couper, 2006). Short-term effects are immediate, whereas long-term effects can be understood as either more lasting or gradual. In a similar manner, studies on L2 pronunciation skills have also shown that instruction improves them (e.g., Lintunen, 2004; Derwing & Munro, 2005: 387–388; Couper, 2006; Saito, 2012; Lee, Jang & Plonsky, 2015). Opposite results are rare (e.g., Purcell & Suter, 1980), and their methodologies have been criticized (Munro & Derwing, 2015: 16–17). Moreover, empirical studies have shown that non-native speakers are even able to achieve the native level in L2 pronunciation skills (Moyer, 2014).

Despite this overwhelming evidence that instruction improves L2 pronunciation skills, there are also concerns that pronunciation skills are,

¹ In this chapter, we use second language acquisition as an umbrella term that includes foreign language learning and will not draw a sharp distinction between the concepts, following the trend in contemporary L2 studies (Ortega, 2009: 5–6).

nevertheless, often neglected in formal education. At university level, pronunciation practice is common in many countries, but at lower levels of education pronunciation teaching seems to be unsystematic. There may be many reasons for this: for instance the washback effect of language exams that are still mostly based on written language (Szypra-Kozłowska, 2015: 4–6). Another reason may be limited teacher education (e.g., Henderson et al., 2012). The learners themselves, however, have been found to have positive attitudes towards pronunciation teaching and to want more practice (e.g., Tergujeff, 2013; Pawlak, Mystkowska-Wiertelak & Bielak, 2015).

In addition to formal instruction in general, the awareness-raising effect that teaching has seems to improve pronunciation skills in particular (Couper, 2006). It seems that learners learn best by noticing differences between their pronunciation and the TL sound system (Schmidt, 1990). Therefore, L2 pronunciation teaching does not merely influence pronunciation skills: teaching also increases learners' metaphonetic awareness (Wrembel, 2011) of the TL. The effect of teaching may also be negative because increased awareness may lead to feelings of unsatisfactory pronunciation self-image, self-efficacy and self-assessment as shown in studies on phonetics learning anxiety (e.g., Baran-Łucarz, 2013, 2014; see also Horwitz, Horwitz & Cope, 1986). Pronunciation is part of the language learner's identity (Rindal, 2010), and L2 pronunciation success may also depend on many affective factors, such as learner orientation towards TL speakers (Gatbonton, Trofimovich & Magid, 2005), which makes pronunciation teaching a complex activity combining fine motor skill training with various affective factors that need to be considered.

Recently, another challenge for pronunciation teaching has been widely discussed: which model to follow and, accordingly, what is the goal for non-native learners? Learners are no longer expected to reach the level of a monolingual native speaker, and this is not even the desired goal for most learners (Dalton & Seidlhofer, 1994: 6–8; Jenkins, 2000). However, the goal is not synonymous with the model. Hewings (2004: 13) notes that the model might not be the target *per se* but used as a point of reference, needed not as a goal but as a guide that also allows variation. As mentioned, earlier studies have provided evidence that reaching the level of a native speaker is possible for some learners, but, for instance, Derwing and Munro (2005: 384) emphasize that native level is not a realistic goal as achieving it is rare and context-dependent. In addition, the concept of native level is difficult to define. Naturally the goal always depends on the learner's communicative needs: some may want to aim at a

native level, whereas for many, fluent TL communicative use is the most important goal, although it might be lower than the native level. Based on this, the learner is at the centre of the learning process, as their goals and ambitions guide the learning. Therefore, it is important to focus on the learners and their beliefs about the goals and importance of L2 pronunciation.

3.2.2 Learner beliefs

Learner beliefs have been found to play an important role in the learning process and to affect learning outcomes (Kalaja & Barcelos, 2013). Earlier studies have used various methodologies to investigate the learner perspective and learner opinions (see, e.g., Nowacka, 2012). The traditional approaches towards beliefs have been complemented by studies following the contextual approach, which does not rely on questionnaires alone but rather uses many kinds of data and also observes learner actions. Within this approach especially, learner beliefs have been found dynamic, socially constructed and context-dependent (Barcelos, 2008; Barcelos & Kalaja, 2011). Therefore, studies should also focus on the development, change or transformation of beliefs. Peng (2011) suggests that L2 beliefs often undergo changes especially when learners move from secondary to tertiary education.

Earlier studies focusing on the learner perspective on L2 pronunciation have often been based on questionnaires and interviews, in which learners have been asked about their pronunciation preferences, goals and attitudes. Usually learners find L2 pronunciation skills very important. According to Waniek-Klimczak and Klimczak (2005) and Waniek-Klimczak (2011), for instance, Polish learners considered vocabulary, pronunciation and grammar, in this order, as the most important aspects of speech. In formal education, however, the emphasis is often only on vocabulary and grammar.

When learner attitudes towards pronunciation goals have been studied, many studies have found that learners want to sound like native speakers and have the native level as their goal (e.g., Janicka, Kul & Weckwerth, 2005; Waniek-Klimczak & Klimczak, 2005; Simon & Taverniers, 2011; Nowacka, 2012; Waniek-Klimczak, Porzuczek & Rojczyk, 2013; Levis 2015). Moreover, Dalton-Puffer, Kaltenboeck & Smit (1997) showed that Austrian university students of English had negative attitudes towards their own non-native accent of English. Other studies have shown that the native level is not necessarily the learners' goal, at least when university learners are not focused on, and instead they emphasize intelligibility (e.g.,

Tergujeff, 2013). One explanation for the different results may be that university students are more likely to aim at native levels and to have higher learning goals than other language learners, while at lower levels of education goals may be more varied. University students are a special group, and studies on advanced learners should not be generalized too far: only a fraction of language learners pursue their studies onto the university level (*cf.* Waniek-Klimczak et al., 2013; Pawlak et al., 2015: 19). Waniek-Klimczak (2011) noted that while university students mostly aim at the native level in pronunciation, they nevertheless prioritize fluency and ease of communication over accuracy. Research results may, of course, depend on the number and specificity of options given in questionnaires.

Studies on learner beliefs about L2 pronunciation often focus on beliefs at a certain time (e.g., Waniek-Klimczak & Klimczak, 2005; Simon & Taverniers, 2011; Tergujeff, 2013). Few studies have considered the transformation of learners' pronunciation beliefs (see, however, e.g., Waniek-Klimczak, Rojczyk & Porzuczek, 2015), which was our main focus in the present study. Levis (2015) suggests that changes in beliefs are often needed for L2 pronunciation success as a learner commonly has contradictory beliefs, which can prevent pronunciation improvement. Beliefs and attitudes are also dynamic and may change as learners gain more experience (e.g., Shizuka, 2008). For example, the educational level has been shown to affect the attitudes of Polish university students of English, as the more experienced students were less concerned with their non-native accents and valued pronunciation skills more for their usefulness in future employment (Waniek-Klimczak et al., 2013; Waniek-Klimczak et al. 2015). During university education, beliefs about the importance of native and non-native accents seem to transform from idealistic to more realistic.

To investigate the long- and short-term effects of pronunciation instruction on learner beliefs further, we focused on four questions: 1. How do learners evaluate their earlier pronunciation instruction? 2. How important do learners consider pronunciation in comparison to other language skills? 3. How do learners evaluate their skills and the goals they set for themselves? 4. How do learner beliefs transform due to instruction, considering both short-term and long-term effects? On the basis of earlier studies, we hypothesized that learners believed that they had not received enough formal training in pronunciation skills but considered pronunciation a very important TL skill. We also expected university learners to have high learning goals and express criticism towards their own skills. Finally, we hypothesized that formal teaching has an effect on

the transformation of learner beliefs, but as there is very little earlier research on this, our final research question was exploratory in nature.

3.3 Method

To answer our research questions, we collected data from 161 L1 Finnish university students of English. Our main method was a questionnaire with 24 statements. The participants used a 5-step Likert scale (1 = strongly agree; 5 = strongly disagree) to respond to statements on the pronunciation of English. Some participants left some statements unanswered. The choices 1 and 2 agreed with the statement, and 4 and 5 meant that the participants disagreed. We did not use gender or age as variables in our study, but most participants were women in their early 20s.

We collected data cross-sectionally, and our participants were divided into three groups: first year students who had not taken a practical pronunciation course (group “before teaching”, $n=41$), first year students who had taken a practical pronunciation course (group “after teaching”, $n=74$) and more experienced post-BA students who had taken a pronunciation course at least three years earlier (group “after >3 years”, $n=46$). In this sense, the decisive criterion separating these groups was their phonetic awareness and experience as students of English. The first group had received the least formal instruction, whereas the third group had not only received formal instruction but this had been reinforced several times during their studies as the learners gained more experience and spoke more English. The first year data were collected at the end of a compulsory lecture course. All participants had taken at least a theoretical lecture course on phonetics (12 x 90-min lecture) at the time of testing. Group “after teaching” had also taken a practical pronunciation course (12 x 90-min of pronunciation exercises) concurrently with the phonetics course, whereas group “before teaching” would take it later due to curricular reasons. The more advanced student data were collected during MA courses, and only the answers of those students who had taken the pronunciation course at least 3 years earlier were included in the analyses.

To see if taking an explicit pronunciation course and further English studies affect student beliefs, we compared the beliefs of the three groups using the nonparametric independent-samples Kruskal-Wallis test with pairwise comparisons to compare individual groups to each other. The subject group’s answers to different statements were compared with the Wilcoxon signed ranks test. To compare three statements, we used the nonparametric Friedman test (Larson-Hall, 2010: 276–285).

The questionnaire data were complemented by semi-structured interviews (n=10) in Finnish at the end of the theoretical course on phonetics. The interviews ranged from 19 to 54 minutes, and the students were asked to elaborate on the same questions that were posed in the questionnaire. The interviews were carried out by the teacher of the course, which may have affected the results as students may have overemphasized the positive effect of teaching or left some negative thoughts unmentioned. However, the interview responses seemed to concur with the questionnaire data. The more advanced students (“after >3 years”) were asked to answer some additional open-ended questions in the questionnaire, which enabled them to elaborate on their beliefs about formal pronunciation instruction and how their beliefs had developed or changed after the course.

3.4 Results

3.4.1 Earlier education

There were four statements that directly focused on the learners’ opinions on their earlier education. The statements and average numbers are included in Table 3-1. An average closer to 1 means that the group agreed with the statement more strongly.

Statement	All groups	Before teaching	After teaching	After >3 years	Kruskal-Wallis (<i>p</i>)
I was taught enough grammar at school	2.52 (1.22)	2.61 (1.30)	2.80 (1.27)	1.98 (0.86)	0.002*
I was taught enough pronunciation at school	4.09 (1.03)	3.78 (1.17)	4.31 (0.94)	4.00 (0.97)	0.018*
I have had pronunciation training before university	4.13 (1.15)	4.22 (1.08)	4.20 (1.10)	3.93 (1.27)	0.403
Schools concentrate too much on spoken skills	4.46 (0.75)	4.41 (0.77)	4.45 (0.73)	4.52 (0.78)	0.559

Table 3-1. Beliefs about earlier education (mean, standard deviation).

The students often agreed that they had been taught enough grammar at school (57.8% agreed, 24.2% disagreed; mean 2.52 for all groups combined), but quite clearly thought that pronunciation had not been

taught enough (9.9% vs. 78.9%; mean 4.09). The results reveal a clear difference in student beliefs about these two levels of language (see also Fig. 3-1). The fairly negative views on earlier pronunciation or spoken language training were further reinforced as the learners also quite strongly disagreed with the statements that they had had pronunciation training before (12.4% vs. 78.3%; mean 4.13) and that schools focus too much on spoken skills (1.2% vs. 88.2%; mean 4.46). Especially these last two statements were strongly disagreed with. The difference between pronunciation and grammar teaching was statistically significant (Wilcoxon signed ranks test $p < 0.001$). The results confirmed our hypothesis that earlier education is considered insufficient.

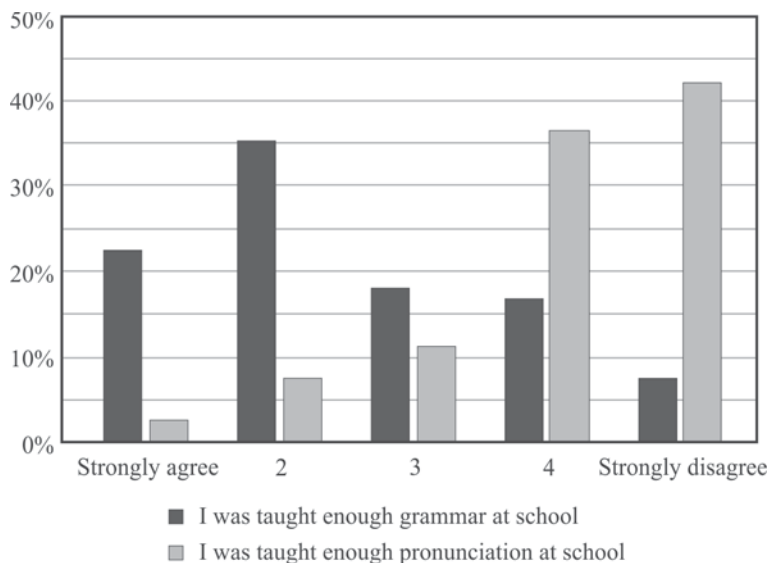


Fig. 3-1. The sufficiency of earlier grammar and pronunciation instruction according to all groups combined.

When we focused on differences between groups, a significant difference was found in the first two statements. The more advanced students were more of the opinion that enough grammar had been taught at school. The reason may be that most beginning students were also taking a grammar course at the same time, which may have revealed some gaps in their knowledge of grammar. In contrast, when the more experienced students thought about their knowledge of grammar more retrospectively, they were more content with their earlier education. In pairwise comparisons,

statistical significance ($p < 0.05$) was found between groups “after teaching” and “after >3 years”. Naturally, one also has to remember that university students are interested and active learners who may require a lot of training before using the word “enough”.

Another significant difference was discovered in the belief about earlier pronunciation teaching: the students who had attended a pronunciation course were more of the opinion that pronunciation had not been taught enough before. The distribution of the responses to this statement can be seen in Fig. 3-2.

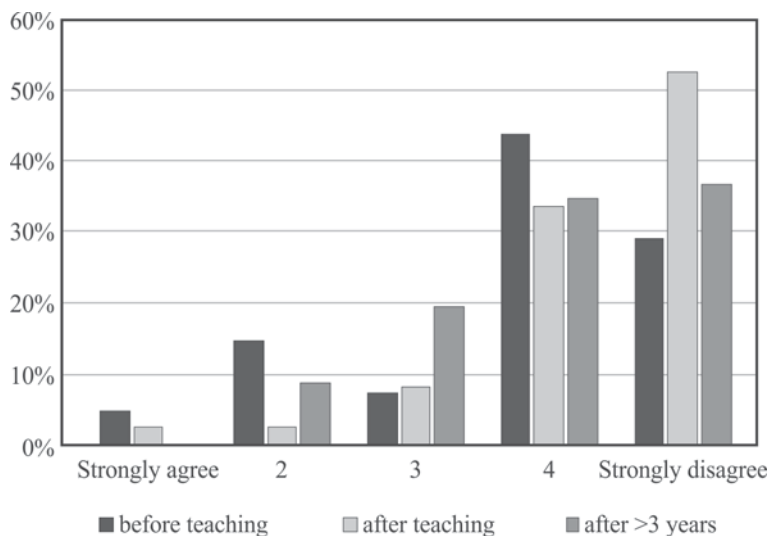


Fig. 3-2. Responses to “I was taught enough pronunciation at school” per group.

As also revealed by Fig. 3-2, learners fairly strongly disagreed with this statement directly after the pronunciation course (“after teaching”), meaning that formal instruction seemed to intensify their belief. The more advanced students’ responses (“after >3 years”) did not reveal this effect although all groups had in general quite negative responses to this statement. In pairwise comparisons, statistical significance was found between groups “before” and “after” teaching. Formal instruction did not affect students’ beliefs about having pronunciation training before, but about the sufficiency of such training. Thus it seems that explicit pronunciation teaching made students realize what they had missed in their earlier education. This finding was confirmed by the semi-structured interviews with the group “after teaching”, as can be seen from the

following comments (Examples 1–2). Similar responses could be gleaned from the group “after >3 years” (Examples 3–4). The examples have been coded to indicate the type of data (I=interview, group “after teaching”; Q=questionnaire, group “after >3 years”) and the number of the subject in either category:

- (1) Nothing like that [how individual sounds are produced] was taught at school. It’s a pity that it wasn’t. (I2)
- (2) Grammar doesn’t feel that difficult anymore because we have studied it for several years already, [...] but since I haven’t learnt pronunciation anywhere before, it has been the most difficult aspect for me. (I8)
- (3) [My pronunciation] definitely did [improve], I hadn’t considered individual phonemes in English before and knowing how each sound was made really helped me. (Q49)
- (4) Yes it has [improved awareness] because I haven’t had any previous experience on pronunciation teaching. (Q43)

Based on the answers, it seems that grammar has been taught more explicitly and with the aid of special terms and categories while pronunciation has not.

3.4.2 The role of pronunciation

Ten statements concentrated further on the role of pronunciation as opposed to other levels of language. The statements and average numbers are included in Table 3-2, which lists them from the most agreed with to the least agreed with statements. As opposed to the statements on earlier education, none of the statements in this category was very strongly disagreed with.

According to the students, they often pay attention to their pronunciation (85.1% vs. 5.6%; mean 1.83) and grammar (78.9% vs. 8.1%; mean 2.01). There was a significant difference in the responses to these statements (Wilcoxon signed ranks test $p < 0.01$). There was a slight change in paying attention to pronunciation between students who had not received (“before teaching”) and had just received (“after teaching”) explicit pronunciation instruction so that training seems to make students notice their own pronunciation more. This difference was not statistically significant, though. There seemed to be no such difference when it comes to grammar. The results also revealed that before pronunciation training the students paid slightly more attention to their grammar than pronunciation, but after formal instruction this order was reversed. The difference in responses between the groups was not great and therefore not

significant. Taken together, the students were active language learners who paid attention to their language skills, including pronunciation.

Statement	All groups	Before teaching	After teaching	After >3 years	Kruskal-Wallis (<i>p</i>)
I often pay attention to my pronunciation of English	1.83 (0.82)	2.10 (1.00)	1.70 (0.68)	1.80 (0.81)	0.131
I often pay attention to my grammar of English	2.01 (0.94)	2.05 (1.09)	1.99 (0.91)	2.02 (0.86)	0.907
To use words accurately is important for communication	2.05 (0.89)	2.02 (0.72)	1.96 (0.90)	2.22 (1.01)	0.296
To be accurate in pronunciation is important for communication	2.28 (0.87)	2.32 (0.82)	2.18 (0.83)	2.43 (0.96)	0.449
I believe that the grammatical rules of English are easy to follow	2.53 (0.97)	2.44 (0.98)	2.74 (1.00)	2.28 (0.86)	0.035*
I'm more interested in the pronunciation than the grammar of English	2.71 (1.14)	3.15 (1.04)	2.53 (1.11)	2.61 (1.18)	0.020*
To be accurate in grammar is important for communication	2.75 (0.96)	2.73 (0.78)	2.77 (1.09)	2.74 (0.91)	0.994
I believe that English is easy to pronounce	2.77 (1.03)	2.93 (1.19)	2.81 (0.97)	2.57 (0.96)	0.315
English has clear grammatical rules	2.93 (0.98)	2.80 (1.14)	3.11 (0.92)	2.74 (0.91)	0.172
English has clear pronunciation rules	3.16 (1.04)	3.29 (0.93)	3.19 (1.07)	2.98 (1.09)	0.442

Table 3-2. Beliefs about the role of pronunciation (mean, standard deviation).

According to the results, the learners also thought that to use words accurately is important for communication (78.3% vs. 7.5%; mean 2.05), whereas pronunciation (65.2% vs. 8.7%; mean 2.29) and grammar (44.7%

vs. 23.6%; mean 2.75) were considered slightly less important; this agrees with Waniek-Klimczak and Klimczak's (2005) and Waniek-Klimczak's (2011) results with Polish learners of English. The difference between words, pronunciation and grammar was statistically significant (Friedman $p < 0.001$). This order remained constant between all learner groups. Thus, pronunciation seems to be an important part of TL skills, but not the most important skill. The students who had just received explicit pronunciation instruction ("after teaching") responded marginally more strongly towards the importance of pronunciation for communication, but this did not change the order of importance in their beliefs (see also Fig. 3-3).

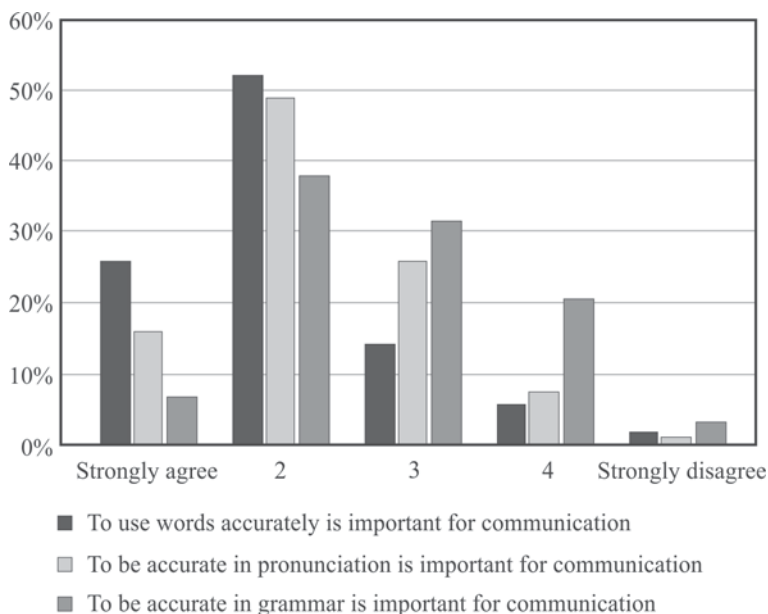


Fig. 3-3. The importance of words, pronunciation and grammar for communication according to all groups combined.

In a similar manner, more learners believed that English has clear grammar rules than clear pronunciation rules, and this tendency remained constant between the groups. The difference between the responses to these statements was statistically significant (Wilcoxon signed ranks test $p < 0.01$). Naturally, the earlier focus on grammar may have affected the result. It may also be that pronunciation had not been thoroughly discussed before university level, and prescriptive rules about language are more

likely to be used at lower than higher levels of education. The learners also often believed that the grammatical rules of English are easy to follow (54.7% vs. 16.8%; mean 2.53) and that English is easy to pronounce (42.9% vs. 23.6%; mean 2.77). The difference between the responses to these statements was statistically significant (Wilcoxon signed ranks test $p < 0.02$). Especially the more advanced students (“after >3 years”) believed that the grammatical rules are easy to follow. In pairwise comparisons, the group “after >3 years” differed significantly ($p < 0.04$) from the group “after teaching”.

When analyzing the effect of teaching, those students who had attended pronunciation classes agreed more with the statement “I’m more interested in the pronunciation than the grammar of English”. There was a statistically significant difference between the groups: formal instruction made students respond more positively (see also Fig. 3-4). We can again observe that explicit pronunciation instruction affected learner beliefs as, in pairwise comparisons, the “before teaching” group differed significantly ($p < 0.02$) from the “after teaching” group.

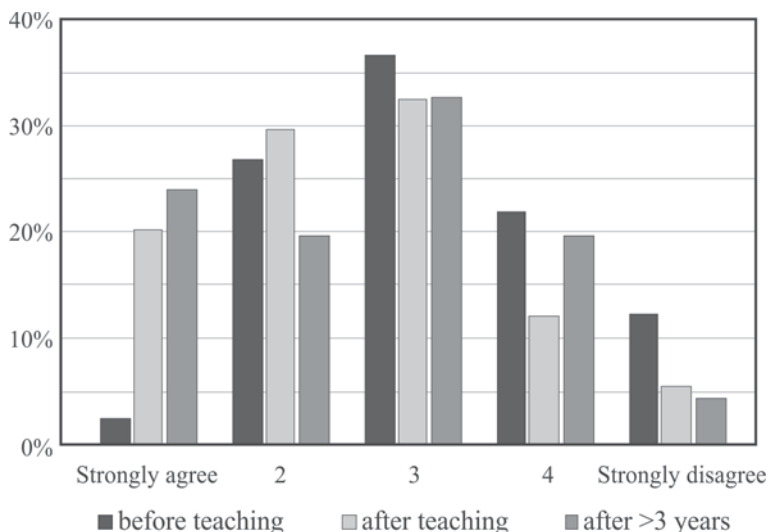


Fig. 3-4. Responses to “I’m more interested in the pronunciation than the grammar of English” per group.

Although the results did not reveal a significant change in students’ paying more attention to pronunciation after taking the practical course, the results from the semi-structured interviews suggest that a change in beliefs

had, indeed, taken place at least for some participants (Examples 5–8). Similar answers were provided by the more advanced students (Examples 9–10), and it is notable that when they were asked if pronunciation teaching had increased their awareness of English pronunciation, all but one answered affirmatively (the remaining student answered “maybe”).

(5) Of course [I’ve begun to pay considerably more attention to my pronunciation after the course]. (I2)

(6) Yes [I pay more attention to my pronunciation], because now certain things have names and their own categories, and when listening to people I may start to pick out things. (I6)

(7) Yes, I now pay much more attention to my own pronunciation, but also to other speakers’ pronunciation. (I10)

(8) Yes, nowadays I analyze what I say a bit too much [...] I’ve become too conscious of how I speak. (I8)

(9) My pronunciation improved during the course because after that I was able to pay attention to the subtle differences in pronunciation. (Q9)

(10) [My pronunciation has improved after the course] because I’ve paid attention to my pronunciation after taking the course. (Q24)

It seems that students pay more attention to their pronunciation simply because they are more aware of it and are able to discern specific features of pronunciation. Moreover, their newly acquired analytical attitude is not limited to their own skills but expands to those of others as well. In addition, Example (8) shows that students could also view awareness as a negative outcome of formal instruction.

3.4.3 Pronunciation skills and goals

As our third research question, we were interested in student beliefs about their own pronunciation skills and goals. There were 10 questionnaire statements related to this theme. Table 3-3 lists the statements from the most agreed to the least agreed ones.

The results suggest that the participants wish to sound like native speakers of English (88.8% vs. 2.1%; mean 1.43) and also try to avoid a Finnish accent of English (86.3% vs. 2.5%; mean 1.57). As the subjects were university students of English, it was quite expected that they set high goals for their English skills. In addition, the first statement could refer to an idealized situation. On the other hand, the statement that received mostly negative responses was that English spoken with a Finnish accent sounds good (5.0% vs. 80.8%; mean 4.26). Here we can see negative attitudes towards the learners’ own non-native accent and positive attitudes towards native accents (*cf.* Dalton-Puffer et al. 1997).

Students' experience seemed to affect their beliefs with this last statement, as the more advanced students ("after >3 years") had somewhat less strongly negative responses although the majority of responses in all groups was negative (see also Fig. 3-5). The difference between the "after >3 years" group and the other groups was statistically significant ($p < 0.06$). In a similar manner, the "after >3 years" group differed significantly from the "after teaching" group in the statement "I try my best not to have a Finnish accent of English" as the experienced students seemed to accept their non-native accent more often ($p < 0.05$), which agrees with Waniek-Klimczak et al. (2015).

Statement	All groups	Before teaching	After teaching	After >3 years	Kruskal-Wallis (p)
I would like to sound like a native speaker of English	1.43 (0.74)	1.39 (0.72)	1.35 (0.69)	1.61 (0.83)	0.100
I try my best not to have a Finnish accent of English	1.57 (0.81)	1.49 (0.87)	1.45 (0.69)	1.83 (0.90)	0.030*
I still have a lot to learn in my English skills	1.78 (0.89)	1.78 (1.04)	1.59 (0.72)	2.09 (0.95)	0.012*
I pronounce English well	2.29 (0.88)	2.54 (0.87)	2.31 (0.91)	2.04 (0.79)	0.023*
I feel confident when I speak English	2.50 (1.16)	2.32 (1.15)	2.75 (1.18)	2.26 (1.08)	0.026*
I believe I can achieve the competence of a native speaker	2.80 (1.19)	2.78 (1.17)	2.77 (1.17)	2.87 (1.28)	0.891
Intonation is the most important aspect of pronunciation	2.89 (0.76)	2.95 (0.84)	2.95 (0.72)	2.76 (0.77)	0.423
Consonants of English are difficult to pronounce	3.15 (1.17)	3.22 (1.13)	3.10 (1.18)	3.17 (1.20)	0.862
Vowels of English are difficult to pronounce	3.61 (0.95)	3.41 (1.02)	3.66 (0.87)	3.72 (1.00)	0.249
English spoken with a Finnish accent sounds good	4.26 (0.93)	4.41 (0.87)	4.45 (0.83)	3.83 (1.00)	0.000*

Table 3-3. Beliefs about pronunciation skills and goals (mean, standard deviation).

When the learning goal was worded differently ("I believe I can achieve the competence of a native speaker"), not being restricted to pronunciation alone, the students were more divided (44.1% vs. 30.4%; mean 2.80) with more than 25% choosing the middle value on the Likert scale, implying

their uncertainty. On the other hand, based on the responses, the students believed that, in general, they still had a lot to learn in their English skills (82.0% vs. 5.0%; mean 1.78). The most experienced group differed from the other groups as they did not agree as strongly with this statement, which could imply that they were more confident with their general language skills than first year students. The difference between the groups “after teaching” and “after >3 years” was statistically significant ($p < 0.01$) as especially the group that had recently received explicit pronunciation training (“after teaching”) was of the opinion that they still had a lot to learn.

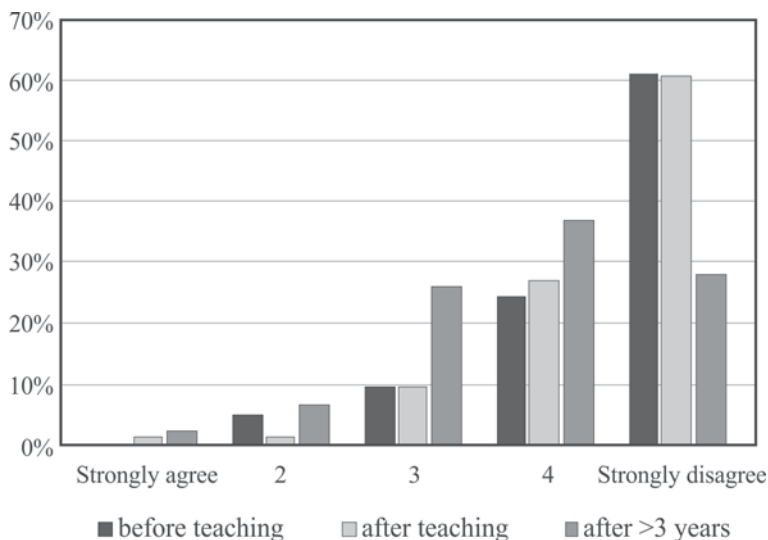


Fig. 3-5. Responses to “English spoken with a Finnish accent sounds good” per group.

The semi-structured interviews confirmed the finding that the students preferred native accents, but as also Examples 11–14 show, their attitudes may have been more flexible than what the questionnaire data indicate.

(11) It would be great to be able to speak like a native. I don’t know if that’s possible, but I suppose it’s at least worth aiming at. It’s not my priority as long as I can pronounce well. (I3)

(12) I don’t think I need to sound like a native speaker. After all, I’m a Finn and my native language is Finnish and that is not going to change. It’s always going to affect the way I speak English. (I4)

- (13) I'm more critical towards English spoken with a Finnish accent. [...] I suppose that it is so close to Finnish, while other accents have a foreign tone so that they already sound more different. (I5)
- (14) The accent doesn't matter if you understand the message. (I6)

The answers suggest that, although the students preferred native models, they were still realists with regard to their practical goals. Furthermore, successful communication seems to be more important than perfect pronunciation (*cf.* Waniek-Klimczak, 2011). The more advanced students also commented on how their beliefs and attitudes regarding these aspects had transformed after formal pronunciation instruction (Examples 15–17).

- (15) I realized I will never sound like a native speaker so I relaxed. (Q10)
- (16) I have realized that the most of the situations that require me to use English happen when I communicate with people whose first language is not English. It's more important to be understood than speak perfect English in these situations. (Q17)
- (17) After the course I felt less like there was a "correct" way to pronounce English or at least there was lots of different correct ways. (Q49)

Presumably such attitudes would strengthen as the students accumulated more experience and communicated with people with various kinds of accents. However, the more advanced students could still aim for a high level while at the same time being more tolerant of others' accents.

The subject groups thought that consonants were more difficult to pronounce than vowels (Wilcoxon signed ranks test $p < 0.001$), which corresponds to earlier studies on the most common pronunciation problems of Finnish learners in English (Lintunen, 2004). This might also reflect the fact that the English vowel system varies and allows more freedom than the consonant system between accents (see also Examples 18–19). The importance of intonation was mostly uncertain as more than half of the students chose the middle value on the Likert scale with this statement.

- (18) [Vowels are easier because] you're already more used to vowels in Finnish, the same vowels. (I3)
- (19) Consonants are clearly more difficult [...] because there are consonants which we don't have in Finnish and I also hadn't realized that some of them, like the sibilants, are different, but vowels are more similar to the Finnish ones, so they are quicker to learn. (I10)

When asked whether the students felt confident when they speak English, the majority of responses were positive (54.0% vs. 18.0%; mean

2.50). The distribution of positive and negative responses also reveals that the number of uncertain answers was remarkable. When the different groups were compared, a statistically significant difference between the groups was found: recently received pronunciation instruction made the students more uncertain (see Fig. 3-6). Therefore, explicit instruction may increase learners' doubts towards their own use of the TL. However, the more experienced students ("after >3 years") mostly responded as students without formal instruction, which implies that possible increased insecurity was not a lasting phase, but instead we can see the transformation of beliefs for a short period of time after (and possibly during) explicit pronunciation teaching.

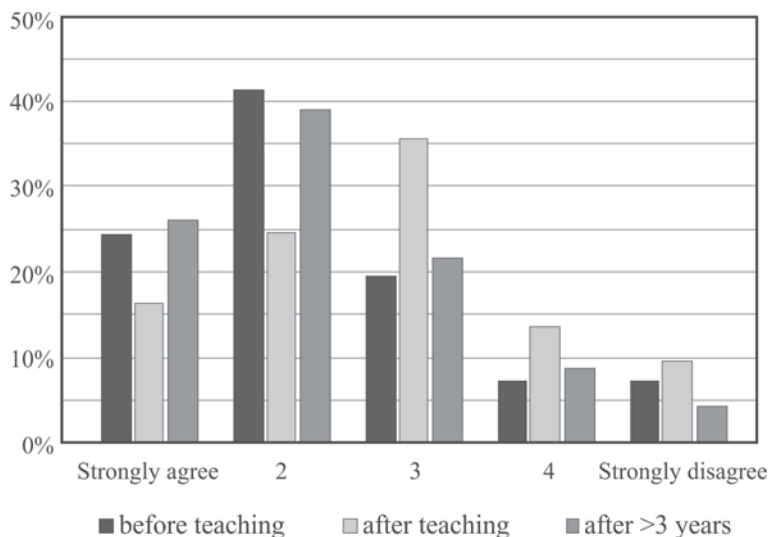


Fig. 3-6. Responses to "I feel confident when I speak English" per group.

Although it seems that many students had slightly negative beliefs about their confidence of speaking English, they, nevertheless, mostly believed that they pronounce English well (65.2% vs. 9.3%; mean 2.29). There seemed to be no negative period directly after formal teaching as with the previous statement. Instead, the students who had not received explicit teaching were more unsure of their skills, as demonstrated by the number of middle answers on the Likert scale (see Fig. 3-7). Student beliefs seemed to transform into more positive ones after explicit teaching and as the students gained more experience. This development was clear: the

difference between the groups “before teaching” and “after >3 years” was statistically significant ($p < 0.02$).

Corrective individual feedback during a specialized course may have affected student confidence (*cf.* Baran-Łucarz, 2013). Yet, the students still trust their abilities in general, which could mean that, in their opinion, formal classroom situations where language skills are being evaluated differ from casual conversations in everyday use of English.

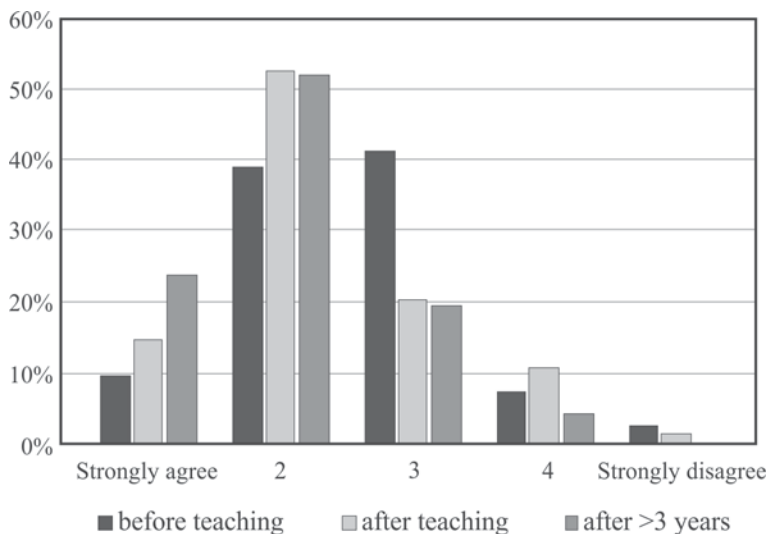


Fig. 3-7. Responses to “I pronounce English well” per group.

Results from the interviews (Examples 20–23) also shed some light on the changes in self-confidence. In particular, some of the students made a connection between increased awareness and the loss of confidence (Examples 21–23; *cf.* also Example 8 above):

(20) Halfway through the course I was maybe less confident, [...] but now after the course I am perhaps a bit more confident, and now I know what the most difficult things are and which things I need to practise more. (15)

(21) Now I realize what goes wrong and I pay so much attention to it, at least right now, that I wonder if I am able to say anything at all anymore. (12)

(22) Perhaps I’m now more confident, but sometimes I notice that I’m not doing so well with my pronunciation, since there are so many of those rules, and it sets me back a bit. (13)

(23) I used to be confident, but having taken the course I'm not anymore. I pay much more attention to my speech, and try to be more careful, and try to articulate correctly. But maybe it's a good thing because I didn't pay much attention to my pronunciation before. [...] I am more aware of my own problems. (I10)

These answers suggest that explicit instruction had transformed the students' beliefs by making them at the same time more aware of their problems or mistakes in pronunciation and more confident about their knowledge of pronunciation.

3.4.4 Short-term and long-term effects of formal instruction

As our fourth research question, we focused more closely on the short-term and long-term effects of formal instruction on learner beliefs. The results above have already demonstrated that many student beliefs develop and transform during their studies. Some changes take place during or immediately after explicit teaching, whereas other changes may take a longer period of time. It is likely that the students' pronunciation skills have improved during the pronunciation course and their later studies as they have had more opportunities to use English actively. We wanted to analyze how their beliefs transformed and possibly reflected the improvement in their skills.

When we focused on the short- and long-term effects, we noticed that learner beliefs about the sufficiency of earlier pronunciation, and to some extent grammar teaching, became more critical when they took similar courses at university level. As mentioned, increased awareness may have made students realize that their earlier education could have been more comprehensive. The belief that pronunciation skills were neglected in earlier education seemed to remain fairly constant during studies, but the belief that enough grammar had been taught before university seemed to become stronger when long-term effects are considered: the more experienced students were more satisfied with their earlier grammar teaching than beginning university students. In addition, students seemed to prefer pronunciation to grammar more after formal instruction on pronunciation, and this effect also remained constant. There also seemed to be a marginal development of beliefs to consider pronunciation more important for communication skills than before, but this was not a lasting change.

Pronunciation teaching clearly increased the learners' awareness and seemed to make them pay more attention to their skills. This effect appeared to remain constant. A development that seemed to take a longer

time was student beliefs about the native level goals and the attractiveness of a non-native accent in English. However, similar ideas were already presented in student interviews shortly after the pronunciation course although not revealed by the questionnaire data. As learners became users of English with more background and education in the language, native level goals were less common, although high learning goals were found in all learner groups. In other words, beliefs about accents seem to change from idealistic to more realistic. The results also suggested that student beliefs about their own pronunciation become steadily more positive with experience. The negative effect of explicit pronunciation teaching on learner confidence was revealed as a temporary effect and not shown in the beliefs of more advanced students. Therefore, it seems that formal instruction does not only have short- and long-term effects, but also positive and negative effects.

The results from the interviews and the open-ended questions suggest that the change in awareness, in particular, was connected to both positive and negative effects (*cf.* Examples 8, 21–23 above). When the more advanced students were asked why their pronunciation did or did not improve during the course or after it, many of them mentioned awareness as a positive factor (Examples 24–27):

(24) It did improve [during the course]. I became aware of different sounds and learned how to produce those sounds. (Q10)

(25) It improved [during the course], because it made me aware of the mistakes I made in pronunciation before. (Q43)

(26) I do think it improved [during the course] because I became much more aware of issues concerning pronunciation. (Q46)

(27) I would say [that it has improved after the course]. By being aware of the desired pronunciation features, I tend to think about them more often. (Q4)

Although over half of the more advanced students thought that teaching had increased their self-confidence, their answers often showed uncertainty. Many of them also found that the increased awareness had had negative effects, either as a drop in their self-confidence or as a distraction in communication (Examples 28–32):

(28) Maybe [pronunciation teaching] makes you even more aware of your pronunciation, which might actually increase your nervousness. (Q38)

(29) I'm not able to speak freely because I constantly control what comes out of my mouth. (Q43)

(30) I know when I sound right, which makes me feel good, but when I say something accidentally wrong, my confidence begins to gradually collapse.

(Q48)

(31) On one hand [it has increased my self-confidence], because now I know how I should be pronouncing, but on the other hand, I'm self aware (maybe too much!) of how I am pronouncing. (Q24)

(32) After [the pronunciation course] I was really conscious of my way of speaking and I felt I was constantly analyzing myself when speaking. However, this changed after my pronunciation stabilized. (Q10)

As shown by Examples (24) and (32) as well as (25) and (29), the same student could comment on both the positive and the negative effects of awareness-raising. In summary, increased awareness seems to be an immediate change which contributes to the gradual increase in the students' positive evaluation of their own pronunciation, while at the same time contributing to a temporary loss of self-confidence. The quantitative analysis of the questionnaire results suggested that the negative effects on student beliefs were temporary (short-term) effects, whereas the positive effects were lasting (long-term) effects (*cf.* Example 32). However, language learners are also individuals, and as the qualitative analysis of the interviews and open questions showed, some learners may also experience negative long-term effects of pronunciation instruction on their self-confidence.

3.5 Conclusion

According to our results, most students believed that their earlier education had lacked formal pronunciation instruction (*cf.* Derwing, 2009). Moreover, pronunciation was considered an important TL skill: more important for successful communication than grammar but less important than the correct use of words (*cf.* Waniek-Klimczak & Klimczak, 2005; Waniek-Klimczak, 2011). The study showed that university students of English are active learners who have high learning goals, as they often aimed at the native level, evaluated their non-native accents negatively (*cf.* Dalton-Puffer et al., 1997), and thought that they still had a lot to learn. However, they were also mostly of the opinion that they pronounced English well. Our hypotheses, based on the results from earlier studies, seem to have been generally confirmed by these findings.

The study also revealed that formal pronunciation instruction has both short-term and long-term effects on learners' beliefs. Short-term effects refer to immediate effects that may be lasting or temporary. Long-term effects can refer to lasting effects or effects that are revealed later. In a

sense, this may be a difference caused by the duration of the effect or the moment of the beginning of the effect. For clarity, we refer to the immediate but temporary effects as short-term. When long-term effects are concerned, if the change is gradual, it may not be possible to show when the change has begun. This also suggests that beliefs are dynamic (*cf.* Shizuka, 2008; Barcelos & Kalaja, 2011). Explicit teaching seemed to make the students more critical towards the sufficiency of their previous pronunciation instruction and to prefer pronunciation over grammar, both of which were also long-term changes. At the same time, the more experienced students seemed to be less critical towards earlier grammar teaching than first year students. A marginal change was that the students seemed to consider pronunciation more important for communication than before, but this was only a short-term change. Another long-term change which began immediately after (or during) instruction was an increased awareness of TL pronunciation (*cf.* Couper, 2006). This may also have been connected to a short-term loss of self-confidence (*cf.* Baran-Łucarz, 2013, 2014). The results also suggested that with experience the students became more tolerant of non-native accents and set their goals below the native level (*cf.* Dalton-Puffer et al., 1997; Waniek-Klimczak, 2011; Waniek-Klimczak et al., 2013; Waniek-Klimczak et al., 2015). Furthermore, explicit instruction made the students evaluate their own pronunciation skills more positively, and this effect became steadily stronger.

L2 teachers and learners should be aware of the affective factors related to language learning and teaching. Skills and beliefs are intertwined and form a complex system of cause and effect relations linked to learners' personalities, learning histories and other background factors. Affective factors need to be considered as an integral part of the learning process, and they could even be directly addressed in L2 classrooms. The development of beliefs that facilitate learning can be included in the learning goals of any L2 pronunciation course. Negative beliefs should be addressed and acknowledged, as Levis (2015) suggested that changes in beliefs are sometimes a prerequisite to learning. Negative beliefs may also be caused by formal pronunciation instruction, but this may be only a temporary (i.e., short-term) effect. Increased awareness can lead to feelings of inadequate skills. In a way, knowledge is a double-edged sword: it can help the language learner, but for critical thinkers it might also raise the bar.

This study has certain limitations that need to be acknowledged. Naturally, the cross-sectional approach cannot reveal the development of individual learners' beliefs, and individual learner differences can affect the results. Questionnaire as a data elicitation method has its inherent

limitations related to the number and wording of the alternatives. To overcome some of the limitations, we complemented our study with thematic semi-structured interviews and open-ended questions. Our study focused on university students of English, who form a special group of learners. They are very motivated and advanced learners who may have less negative emotions towards the TL and learning in general. Naturally, as mentioned, university students are also high-achievers who are critical towards their own skills and, therefore, may have more negative, critical emotions towards their skills and L2 competence than most language learners.

A longitudinal approach could complement the results of this study: to see when and why beliefs develop and transform, individual learners should be followed more closely with questionnaires, interviews and journals where learners can reflect on their own thoughts during formal instruction. Wider sampling could also limit the effect of the teacher. To gain more information on learners' L2 pronunciation beliefs, studies should not only focus on university learners, but include different learner groups so that the skill level could be used as a variable. By following individual learners closely during formal instruction, we can investigate how the change in skill level affects pronunciation beliefs and how various changes in beliefs can be linked to learner and teacher actions in class. All studies of L2 pronunciation beliefs could benefit from pronunciation skill testing (*cf.* Pawlak et al., 2015). The link between beliefs and skills is essential as the ultimate goal is to support learners' pronunciation success by investigating the affective factors to limit beliefs that may hinder development and to reinforce beliefs that facilitate the learning process. The purpose of increased TL awareness is to raise self-confidence and trust in one's TL skills. Moreover, it may be that formal instruction has more effect on beliefs than skills, or that the effects on beliefs are more immediate than the effects on skills. Increased TL awareness and beliefs that support the autonomous learning of the TL may be the most important outcomes of formal pronunciation instruction. They are the prerequisites of better pronunciation skill learning.

References

- Baran-Łucarz, M. (2013). Phonetics learning anxiety – results of a preliminary study. *Research in Language*, 11(1), pp. 57–79.
- . (2014). Researching Phonetics Learning Anxiety: A qualitative approach. In: D. Gabryś-Barker & A. Wojtaszek (Eds.), *Studying*

- Second Language Acquisition from a Qualitative Perspective*, pp. 160–174. Cham: Springer International Publishing Switzerland.
- Barcelos, A. M. F. (2008). Researching beliefs about SLA: A critical review. In: P. Kalaja & A. M. F. Barcelos (Eds.), *Beliefs about SLA: New Research Approaches*, pp. 7–33. Dordrecht: Kluwer Academic Publishers.
- Barcelos, A. M. F. & Kalaja, P. (2011). Introduction to beliefs about SLA revisited. *System*, 39, pp. 281–289.
- Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, 21(1), pp. 46–66.
- Dalton-Puffer, C., Kaltenboeck, G. & Smit, U. (1997). Learner attitudes and L2 pronunciation in Austria. *World Englishes*, 16(1), pp. 115–128.
- Dalton, C. & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.
- Derwing, T. (2009). Utopian goals for pronunciation teaching. In: Levis, J. & LeVelle, K. (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, pp. 24–37. Ames: Iowa State University.
- Derwing, T. & Munro, M. (2005). Second language accent and pronunciation teaching: a research-based approach. *TESOL Quarterly*, 39(3), pp. 379–397.
- Gatbonton, E., Trofimovich, P. & Magid, M. (2005). Learners' ethnic group affiliation and L2 pronunciation accuracy: A sociolinguistic investigation. *TESOL Quarterly*, 39(3), pp. 489–511.
- Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U. & Curnick, L. (2012). The English pronunciation teaching in Europe survey: selected results. *Research in Language*, 1(1), pp. 5–27.
- Hewings, M. (2004). *Pronunciation Practice Activities*. Cambridge: Cambridge University Press.
- Horwitz, E. K., Horwitz, M. B. & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), pp. 125–132.
- Janicka, K., Kul, M. & Weckwerth, J. (2005). Polish students' attitudes to native English accents as models for ESL pronunciation. In: Dziubalska-Kořaczyk, K. & J. Przedlacka (Eds.), *English Pronunciation Models*, pp. 251–292. Bern: Peter Lang.
- Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.

- Kalaja, P. & Barcelos, A. M. F. (2013). Beliefs in second language acquisition: learner. In: C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.
- Larson-Hall, J. (2010). *A Guide to Doing Statistics in Second Language Research with SPSS*. London: Routledge.
- Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), pp. 345–366.
- Levis, J. M. (2015). Learners' views of social issues in pronunciation learning. *Journal of Academic Language & Learning*, 9(1), pp. 42–55.
- Lintunen, P. (2004). *Pronunciation and Phonemic Transcription: A Study of Advanced Finnish Learners of English*. Turku: University of Turku.
- Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics*, 35(4), pp. 418–440.
- Munro, M. J. & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation*, 1(1), pp. 11–42.
- Nowacka, M. (2012). Questionnaire-based pronunciation studies: Italian, Spanish and Polish students' views on their English pronunciation. *Research in Language*, 10(1), pp. 43–61.
- Ortega, L. (2009). *Understanding Second Language Acquisition*. London: Hodder Education.
- Pawlak, M., Mystkowska-Wiertelak, A. & Bielak, J. (2015). Exploring advanced learners' beliefs about pronunciation instruction and their relationship with attainment. In: Waniek-Klimczak, E. & M. Pawlak (Eds.), *Teaching and Researching the Pronunciation of English*, pp. 3–22. Cham: Springer.
- Peng, J.-E. (2011). Changes in language learning beliefs during a transition to tertiary study: the mediation of classroom affordances. *System*, 39, pp. 314–324.
- Purcell, E. T. & Suter, R. W. (1980). Predictors of pronunciation accuracy: a reexamination. *Language Learning*, 30(2), pp. 271–284.
- Rindal, U. (2010). Constructing identity with L2: Pronunciation and attitudes among Norwegian learners of English. *Journal of Sociolinguistics*, 14(2), pp. 240–261.
- Saito, K. (2012). Effects of instruction on L2 pronunciation development: a synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, 46(4), pp. 842–854.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, pp. 129–158.

- Shizuka, T. (2008). The effects of a 24-session EFL pronunciation course as reflected in learners' self-reports. *JACET Journal*, 47, pp. 67–80.
- Simon, E. & Taverniers, M. (2011). Advanced EFL learners' beliefs about language learning and teaching: a comparison between grammar, pronunciation and vocabulary. *English Studies*, 92(8), pp. 896–922.
- Szpyra-Kozłowska, J. (2015). *Pronunciation in EFL Instruction: A Research-Based Approach*. Bristol: Multilingual Matters.
- Tergujeff, E. (2013). Learner perspective on English pronunciation teaching in an EFL context. *Research in Language*, 11(1), pp. 81–95.
- Thomson, R. I. & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: a narrative review. *Applied Linguistics*, 36(3), pp. 326–344.
- Waniek-Klimczak, E. (2011). “I am good at speaking, but I failed my phonetics class” – Pronunciation and speaking in advanced learners of English. In: Pawlak, M., Waniek-Klimczak, E. & J. Majer (Eds.), *Speaking and Instructed Foreign Language Acquisition*, pp. 117–130. Bristol: Multilingual Matters.
- Waniek-Klimczak, E. & Klimczak, K. (2005). Target in speech development: learners' views. In: K. Dziubalska-Kołodziejczyk & J. Przedlacka (Eds.), *English Pronunciation Models*, pp. 229–249. Bern: Peter Lang.
- Waniek-Klimczak, E., Porzuczek, A. & Rojczyk, A. (2013). Affective dimensions in SL pronunciation: A large-scale attitude survey. In: Gabryś-Barker, D. & Bielska, J. (Eds.), *The Affective Dimension in Second Language Acquisition*, pp. 124–137. Bristol: Multilingual Matters.
- Waniek-Klimczak, E., Rojczyk, A. & Porzuczek, A. (2015). ‘Polglish’ in Polish eyes: What English studies majors think about their pronunciation in English. In: Waniek-Klimczak, E. & Pawlak, M. (Eds.), *Teaching and Researching the Pronunciation of English*, pp. 23–34. Cham: Springer.
- Wrembel, M. (2011). Metaphonetic awareness in the production of speech. In: Pawlak, M., Waniek-Klimczak, E. & Majer, J. (Eds.), *Speaking and Instructed Foreign Language Acquisition*, pp. 169–182. Bristol: Multilingual Matters.

CHAPTER FOUR

PERCEPTUAL IMPACT OF FOREIGN-ACCENTED SPEECH

JAN VOLÍN, RADEK SKARNITZL
AND ALICE HENDERSON¹

4.1 Introduction

Foreign-accented speech has been shown to affect human behaviour relatively strongly at the level of subconscious processes. Subconscious language-based discrimination, in the form of latent, intra-individual, perceptual stereotypes and prejudice, is a well-documented phenomenon, in spite of the fact that most people would subscribe to overt tolerance. Negative reactions occur quite probably because our accent is one of the most salient indications of in-group or out-group status, allowing other individuals to

... immediately identify us as either familiar or foreign, young or old, male or female, and so on. It is also the basis for intelligibility, affecting the extent to which others understand what we are trying to say (Moyer, 2013: 9).

We define *accent* in line with Moyer, who uses the term to refer to both native and non-native speech, defining it as “a set of dynamic segmental and suprasegmental habits that convey linguistic meaning along with social and situation affiliation” (2013: 11). For our purposes, non-native speech can be understood as synonymous with foreign-accented speech (FAS) and will be used to refer to speech exhibiting acoustic features from the speaker’s mother tongue.

¹ The work on this chapter was supported by the Czech Science Foundation, project GAČR 14-08084S.

Listeners' perception of FAS and their ability to understand it have been extensively studied, which is of particular interest here in relation to the correlated yet independent constructs of accentedness (or accent strength), intelligibility (how much a listener understands the speaker's message), and comprehensibility (the effort a listener has to exert to understand the speaker's message) (see, e.g. Munro & Derwing, 1995). In addition to listeners' perceptual abilities and cognitive processing ease, socio-psychological factors also influence the way our speech is received because the symbolic value of our accent extends beyond its specific acoustic features (Brown & Levinson, 1979). An underlying assumption for this chapter is that, for the increasing numbers of European university students receiving English-medium instruction (EMI), the ease with which they can accept and process accented speech is a key issue in their academic and professional development. Therefore, the study described here represents a small yet essential contribution to evidence-based decision-making in European higher education, drawing on research from applied linguistics, sociolinguistics and social psychology.

4.1.1 Foreign-accented speech and listeners

Gluszek and Dovidio survey over four decades of research in social psychology, showing that more negative perceptions are attributed to speakers with non-native accents (2010: 217), and the stronger the accent the more negative the evaluations. It follows that accent can have a major, quantifiable impact on people's lives, in the form of employment and housing discrimination, asylum refusal, harassment, racist abuse and other human rights issues (see, for example, Munro, 2003). In the "safer" world of experimental research, studies have demonstrated that foreign-accented speakers tend to be associated with lower status ratings, including lower intelligence or competence. Miller and Hewgill (1964) examined the correlations between dysfluent speech and credibility ratings, and found an inverse relationship: the more dysfluencies there are in speech, the lower the credibility ratings. Brennan and Brennan (1981) explored correlations between accent and attitude, in relation to Mexican-American speech, and found that several of the examined features of foreignness provoked the perception of lower social status of the speakers. Bresnahan et al. (2002) added identity and intelligibility to correlation studies of accent, showing that FAS elicited more positive attitude and affective response if it was judged to be more intelligible. Another interesting finding was that people with a strong ethnic identity preferred American English and those with a weak ethnic identity were "more accepting" of FAS. Continuing in the

ethnic vein, Lindemann (2003) examined attitudes towards Korean speakers of English in the Midwestern United States compared to native speakers of English; she found that the Koreans were evaluated more negatively than the natives on different aspects of the status dimension (most notably, competence, intelligence, and education), while rating on the solidarity dimension did not differ between the two speaker groups. Lev-Ari and Keysar (2010) explored the perception of traits of truthfulness and accentedness in their study, finding that statements read by non-native speakers were rated by native listeners as “less truthful” than the same statements delivered by native speakers. Their study is pertinent here because this finding held even in a second stage of the experiment, where listeners were made aware of the bias and consciously tried to avoid it.

Similar research has been carried out using undergraduates at American universities. For example, Kavas and Kavas (2008) collected questionnaire data from undergraduate students at one university. In contrast to experimental designs and correlational analyses, a questionnaire taps into participants’ conscious judgments. The results show that 82% of these students felt instructors’ “knowledge of subject” is very important in classroom learning; in other words, they expect a teacher to be competent. In relation to speech qualities, 43% felt accent is very important in classroom learning, and 48% felt pronunciation is extremely important in that context. However, 30% agreed that “Foreign accent of a faculty member does not affect my ability to learn” (Kavas & Kavas, 2008: 886). While their results are not conclusive, it is possible to argue that FAS potentially has an impact on learning motivation and ability.

In an earlier study at another American university, Rubin and Smith (1990) studied the correlations between students’ perceptions of teaching ability and the accent, ethnicity and lecture topic among non-native speaker teaching assistants (NNSTAs). Previously they had found that undergraduates have stereotyped perceptions of NNSTAs, with 40% of participants admitting to having dropped a class taught by an NNSTA. Therefore, Rubin and Smith designed an experiment whereby, as undergraduate participants did the listening-rating experiment, a picture of the “said” speaker was projected onto the wall, either a Caucasian or a Chinese face (see also Rubin, 1992). The results were analyzed to distinguish between language-mediated stereotypes and other cultural stereotypes. The key finding was that higher levels of perceived accentedness correlated with lower ratings of expected teaching ability.

The role of such perceptions in spoken interactions has been extensively studied with regard to correlations between accentedness, comprehensibility and intelligibility. The origins of this approach can be

traced back more than a century, when Sweet (1900) argued for a distinction between intelligibility and native-like performance. In the last 20 years, the work of Munro, Derwing, Rossiter and others has been central to disentangling these three constructs via experimental work and to relocating the emphasis in L2 pronunciation instruction from the nativeness principle onto intelligibility:

... there is now empirical evidence, first, that few adult learners ever achieve native-like pronunciation in the L2 (Flege, Munro & MacKay, 1995) and, second, that intelligibility and accentedness are partially independent (Munro & Derwing, 1995). (Derwing & Munro, 2015: 6–7).

The “empirical evidence” referred to above demonstrated that foreign-accented speech is not necessarily unintelligible. Listeners may be quite successful when asked to respond to foreign-accented speech, for example by writing the words they heard or by answering comprehension questions. However, if the focus is on the listeners’ feelings when exposed to foreign-accented speech or, simply, on comprehensibility and accentedness, the results do not necessarily correlate. Unconscious stereotyping or similar mechanisms are clearly at play in such cases.

Cognitive processing demands are a key influence on listeners’ judgments of accentedness and comprehensibility, and on their ability to decipher the speech signal. The socio-psychological concept of “processing fluency” can be defined as the way speech stimuli are judged as a function of the cognitive load involved in their processing (see Reber & Schwarz, 1999; Oppenheimer, 2008); the more difficult it is to decipher a person’s speech, the less favourably that person may be judged. Mai and Hoffman refer to the *vampire effect*, where “the accent distracts the receiver from processing the central message” (2014: 149) because cognitive resources are being devoted to speech decoding, “sucked into” the effort of listening to the accent.

Fortunately, it is possible for listeners to perceptually adapt to FAS, and examining the ability of listeners to adapt to both novel speakers and accents is also a rich area of research (see Bradlow & Bent, 2008; Baese-Berk, Bradlow & Wright, 2013). The latter study found that training based on systematic variability of accented speech can facilitate foreign-accent adaptation, as well as the generalization of that adaptation to both novel talkers and novel accents. Participants simply had to listen and write down the sentences they heard. These were then scored, based on the proportion of words correctly recognized. Given that such word recognition is necessary for effective note-taking, this research has important implications for EMI in higher education contexts.

4.1.2 European universities, English-medium instruction and lectures

Beginning with the Bologna process in 1999, universities across Europe have made a concerted effort to promote the internationalization of their student body. Concomitantly, English-medium instruction (EMI) has become more and more common, especially at Master's level and usually in northern Europe – but not exclusively. In spring 2012, the Politecnico di Milano, Italy's equivalent to American MIT, announced that from 2014 its degree courses would be taught in English. At the time, the University Rector, Giovanni Azzone, justified the decision as follows:

We strongly believe that our classes should be international classes and the only way to have international classes is to use the English language. ... Universities are in a more competitive world. If you want to stay with the other global universities, you have no other choice.²

The Politecnico is currently appealing an Italian regional court's ruling against the switch, but the debates which accompany such changes reveal an institution's underlying principles and goals, as well as wider societal concerns. For example, many Politecnico faculty members signed a petition stating that not only was obligatory English-medium instruction unconstitutional, it was also a threat to Italian culture (Helm and Guarda, 2015).

Instruction in a non-maternal language can be seen as a form of CLIL (Content and Language Integrated Learning). The European Commission states that CLIL “involves teaching a curricular subject through the medium of a language other than that normally used”.³ The trend towards CLIL has been greatly aided by the support of European bodies, which see it as yet another means to facilitate the European ideal of integration and plurilingualism (see de Zarobe, 2008; Bonnet, 2012). According to Van den Craen (2002: 1), since the European Commission's 1995 White Book on Education, “CLIL suits European aspirations of educating citizens capable of speaking, apart from their mother tongue, two community languages”. Promoting CLIL is therefore a policy choice of both symbolic

² <http://www.guardian.co.uk/education/2012/jun/19/elt-diary-june-technology-innovation>

³ European Commission (EC) Web site: http://ec.europa.eu/languages/language-teaching/content-and-language-integrated-learning_en.htm.

and pedagogical importance, especially as it most often takes the form of classes taught in English rather than in other languages⁴.

The present study investigates the status ratings of foreign-accented speakers, focusing specifically on the evaluation of lecture-style speech. Lectures are a common element of university pedagogy and, as such, are of great interest to applied linguists, especially with the rise of corpus linguistics. For example, two major corpora have been compiled, both of which include lecture-style speech in English. The Michigan Corpus of Academic Spoken English⁵ contains almost 200 hours of recordings and was compiled between 1997 and 2002 at the University of Michigan. It was followed by the British Academic Spoken English (BASE) corpus of 160 lectures and 40 seminars, compiled between 2000–2005 at the Universities of Warwick and Reading⁶. Lecture-style speech is of particular interest to the present paper because of its widespread use in European universities, as a common means for teachers to communicate information to students. Lectures can be described as monologic or interactive, and of varying complexity or educational level, but one underlying assumption common to all lectures is that the listeners expect the lecturer to be competent. The perceived competence may, however, be threatened if students have difficulty processing a lecturer's foreign-accented speech or do not perceive it favourably.

Such difficulties are becoming more probable, as the Bologna process has increased staff and student mobility. At an institutional level, universities derive status and prestige from student numbers and success rates, and statistics related to these can influence funding possibilities, both internal and external: attracting high percentages of foreign students is assessed positively, and the speed with which graduates find employment after graduation is a common performance indicator. Mastery of English is widely thought to improve one's employment prospects, providing an argument in favour of EMI. And even when EMI or another form of CLIL has not been adopted, ERASMUS and other exchange programmes have led many European universities to achieve 10–15% of foreign intake each year. As a consequence, the number of interactions

⁴ However, this may be due to the fact that more and more frequently European universities are competing to attract Asian, not European students (see <http://asiancorrespondent.com/50398/top-25-european-universities-for-asian-students/>).

⁵ See the on-line searchable part of the corpus at: <http://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase>

⁶ Publications related to the BASE corpus listed at : <http://www2.warwick.ac.uk/fac/soc/al/research/collect/base/research/>

between non-native speakers of English has risen, be they student-to-student, teacher-to-student or teacher-to-teacher. The impact of foreign-accented speech can have very real consequences, especially in formal contexts, where standard accents are preferred and receive higher ratings on the traits of success and self-confidence (Creber & Giles, 1983; Côté & Clément, 1994). It is evident that the stereotyped scenario of ERASMUS students laughing off misunderstandings in a bar (an informal setting) does not involve the same stakes as in formal situations of unequal status, such as in teacher-to-student interactions. This begs several questions. How do European students perceive NNS teachers' field expertise if their speech is accented? If a student fails a course and thus fails to receive credit, are they apt to blame the teacher's way of speaking? It is not uncommon for American undergraduates to try to switch groups, leaving their international teaching assistant's class for one taught by a native speaker. European students may not have this option, due to smaller university populations and correspondingly fewer parallel groups for a course.

These questions are too vast to be answered here. Nonetheless, this chapter looks at a specific research question in relation to this evolving, European context: How does foreign-accented speech affect the perception of a speaker's status, and more specifically, their competence? Our most general hypothesis in this pilot experiment is that native speakers will receive higher ratings for competence. However, since we are examining French- and Czech-accented speakers, we are also interested in the potential discrepancy in the evaluation of these two speaker groups.

4.2 Method and Participants

Several lecture-style passages were recorded by both male and female, native and non-native speakers of English, specifically: 8 native English speakers, 5 native French speakers and 5 native Czech speakers. Their speech will hence be referred to as NaE, FrE, and CzE, respectively. The lecture-style passages included topics from economics, political science, natural sciences, and linguistics. Several shorter (ca. 15 words) and longer (ca. 40 words) phrases were selected from each speaker. Care was taken to select phrases which contained no disturbing dysfluencies or hesitations.

The phrases were compiled into a two-stage perception test which was administered to 21 Czech and 19 French listeners, as well as to 12 native English-speaking, predominantly American, listeners. These respondents will henceforward be referred to as CzR, FrR, NER, respectively. Both the CzR and FrR were first-year English majors, the NER were mostly undergraduates studying for a semester in France.

The first stage of the perception test, which featured the shorter phrases, contained 20 items; the second stage with the longer phrases contained 9 items. The items in both test stages were ordered pseudo-randomly. Two versions, ordered differently, were generated for each stage, and respondents were randomly assigned to one of the two versions. The tests were delivered through PRAAT (no pause, no rewind was allowed) after the experimenter had slowly read through the instructions aloud for each part in English. Three training items and three distractor items were created from various other voices and used in the tests.

In the first stage, the respondents were asked to indicate their preference for one of two speakers, based on the shorter phrases, with the instructions as follows:

You will play the role of a personnel manager who has to choose the better candidate to represent your company. You are looking for a person who will have to acquire a lot of nontrivial knowledge about the company and who will be competent and trustworthy when representing the company in negotiations. Obviously, personnel managers normally have a CV and other information about the candidates. Your task in this game is to make the decision based only on the candidates' voice.

The pairs of speakers that the respondents heard for each test item were producing identical spoken texts.

In the second stage, respondents evaluated the longer phrases on three seven-point Likert scales, where the statements concerned three dimensions of competence (see McCrae & Costa, 1987): the speakers' reliability and punctuality, their memory and ability to see connections between things or concepts, and their job effectiveness. The full description was given in the instructions as follows:

- a) The speaker can be relied upon, keeps his or her promises, comes on time for meetings.
- b) The speaker likes to read and remembers a lot of information, sees connections between things.
- c) The speaker solves problems effectively, does not avoid them, is respected by his or her colleagues.

The instructions informed participants that "Here you will **not** be comparing two candidates. We would like you to make judgements relating to the speakers' personality, based on his or her voice." Evaluations for each of the dimensions were marked on a 7-point Likert scale.

After completing the two stages of the perception test, participants also answered questions about their general language experience. In general, all participants can be described as relatively experienced language learners but with varying degrees of proficiency. One question asked “If you have studied other languages, please name those and estimate your level”. The average number of languages mentioned for all participants was 2.4. Their average self-assessed level for the foreign language (FL) they believed they could speak best was 4.06 (on a scale from 0 to 6), as shown in Table 4-1.

Respondents	n	Proficiency in a FL	Familiarity with FAS
French (FrR)	19	4.32	4.11
Czech (CzR)	21	4.01	4.70
Native Engl. (NER)	12	3.75	5.17
All	52	4.06	4.59

Table 4-1. Mean self-assessed level of proficiency in a foreign language of the test participants (respondents) and their familiarity with foreign-accented speech (on a seven-point scale of 0–6).

Table 4-1 also shows that in terms of FL proficiency in their best mastered language, the native-English speakers rated themselves least favourably, whereas the French students rated their own proficiency the most favourably. Another question asked about familiarity with “speech that sounds foreign” (again on a scale from 0–6). In general, participants rated themselves as highly familiar with foreign-sounding speech, although the French self-rated lower than both the Czech and NER participants (4.11 compared to 4.7 and 5.17, respectively).

Participants were also asked to estimate the frequency with which they hear or speak a foreign language, assessed on a scale from 0 (never) to 5 (every day). Table 4-2 shows that, on average, all participants hear a foreign language more often than they speak it. The French participants seem to speak and hear an FL the most often of all three groups, whereas the Czech speakers lag behind for both activities. The differences, however, are insignificant.

Respondents	n	Speaking FL	Hearing FL
French (FrR)	19	4.53	4.84
Czech (CzR)	21	4.05	4.76
Native Engl. (NER)	12	4.50	4.83
All	52	4.33	4.81

Table 4-2. Mean self-assessed frequency of speaking and hearing a foreign language (on a scale of 0–5; see text).

The questions asking about the context of FL exposure revealed that the FL was most frequently encountered by listening to music, and then (in descending order) by watching videos or TV shows, reading, “other”, face-to-face contact, on-line contact and, finally, gaming.

It is also interesting to see which languages are studied by our listeners, the assumption being that familiarity with French or Czech could conceivably influence the participants’ judgments of recorded extracts of English. However, the knowledge of French or Czech does not necessarily mean that listeners are familiar with English spoken with a French or Czech accent. All but one of the 19 French respondents had studied another Romance language, yet only one had studied a Slavic language. Similarly, 15 of the 21 Czech listeners had studied a Romance language (French for 9 of them) but only two had studied another Slavic one. The most commonly studied language family among the 12 NERs was Romance (Latin, French, or Spanish) and three had studied Russian or Czech, one Korean, one Mandarin, and one German. In summary, it can be said that the Czech respondents are potentially more familiar with a French accent or other Romance-language accents than the FrR or NER are with Slavic accents.

4.3 Results

Results of the study are presented in two parts: preferences between pairs of speakers based on short extracts (Stage 1) and evaluations of three dimensions of competence based on longer extracts (Stage 2).

4.3.1 Stage 1: Preference between Pairs of Speakers

The results for the test where participants were asked to choose between two speakers of short extracts are presented in Fig. 4-1, where the total number of responses appears on the vertical axis and the types of paired

speakers on the horizontal axis. It has to be stressed again that both speakers in a pair were saying the same words in identical phrases.

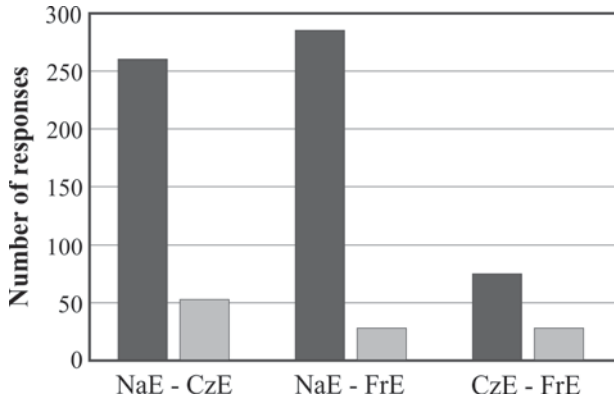


Fig. 4-1. Preferences between speakers of different accents in pairs. The prevailing choice is in dark grey, the minority choice is in light grey. All participating listeners' responses are combined. NaE – speech of native speakers of English, CzE – Czech-accented English, FrE – French-accented English.

A clear preference for native English emerges compared to foreign-accented English, and Czech-accented speech was preferred more than French-accented speech. However, since there were unequal numbers of participants in our three groups of listeners (FrR, CzR, NER), a further breakdown of the data was necessary. Fig. 4-2 shows differences between the three sets of participants, with the numbers of responses now expressed as percentages of the whole set of responses by the given group of listeners concerning the given comparison. (It should be noted that in the actual testing, the order of compared accents was counterbalanced and randomized.)

In both the NaE-CzE pairings and the NaE-FrE pairings, native English was preferred by all three groups of listeners, although the strength of the preference was not equal. The top panel shows that Czech participants rejected Czech accented-English more often than the other two groups, while native English participants disfavoured CzE in only about one quarter of the cases. A similar picture occurred for the NaE-FrE comparison (middle panel) in that the French listeners rejected their own accent most often and native English participants were more tolerant than the other groups. As to the CzE-FrE pairings (bottom panel), Czech-accented English was preferred by all three groups of respondents albeit,

again, to different extents. French respondents rejected FrE more often than the other two groups and native English respondents displayed relatively balanced choices. To summarize, there seems to be clear out-group preference in both non-native groups, and it is stronger in the French group.

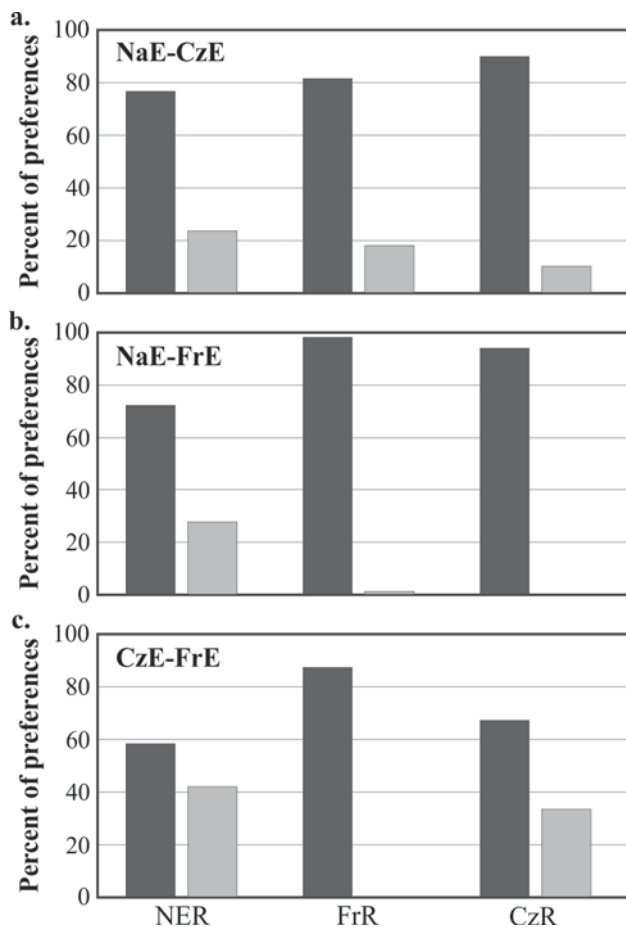


Fig. 4-2. Preferences in percentages of choices by three groups of participants: **a.** preferences in NaE-CzE pairings; **b.** preferences in NaE-FrE pairings; **c.** preferences in CzE-FrE pairings.

4.3.2 Stage 2: Dimensions of Competence

Respondents evaluated the longer phrases on three 7-point Likert scales (-3 to 0 to +3), where the statements concerned three dimensions of competence: the speakers' reliability and punctuality, their memory for facts and ability to see connections between events (insight), and their job effectiveness. Fig. 4-3 shows the mean ratings by all respondents for the three types of accented speech, where each bar represents one dimension of competence.

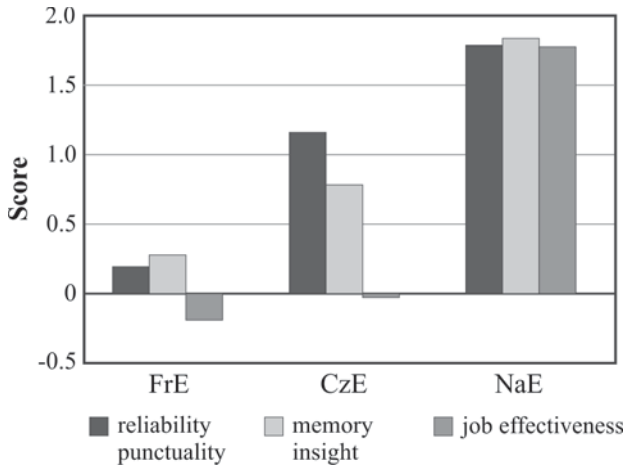


Fig. 4-3. Competence evaluation in three competence-related dimensions on a 7-point scale from -3 to 0 to +3. All listeners combined.

Native-speaker English (NaE) was preferred for all dimensions by all listeners, and Czech-accented speech was rated higher than French-accented speech. On the “job effectiveness” dimension, negative ratings were given to French-accented speech, and to a lesser extent to Czech-accented speech.

Generally, in the non-native speakers of English, the reliability/punctuality dimension was awarded the highest scores, while job effectiveness was the lowest. Contrary to that, the evaluations of the three dimensions in the NaE group were quite balanced.

A similar preference for native-speaker English is evident in the results broken down by the mother tongue of the respondents (Fig. 4-4).

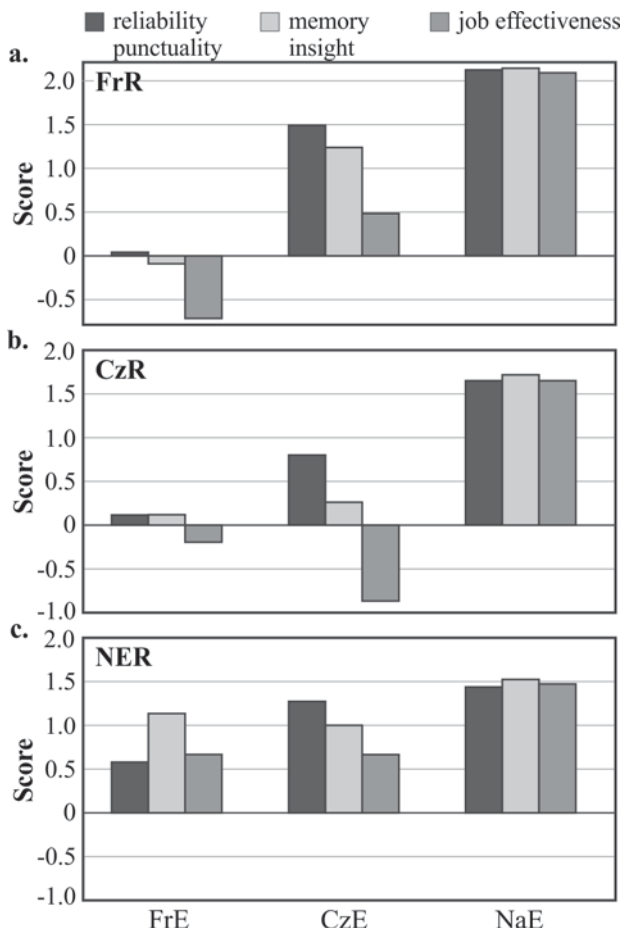


Fig. 4-4. Competence evaluation in three competence-related dimensions on a 7-point scale from -3 to +3. **a.** French respondents (FrR); **b.** Czech respondents (CzR); **c.** respondents who are native speakers of English (NER).

The groups of listeners differ in several respects. First, French respondents rated native speakers of English higher on all dimensions than both the Czech and native respondents did. An out-group preference is also most visible among French listeners, who rated Czech-accented speech higher than French-accented speech on all three dimensions. The Czech participants also rated Czech-accented speech positively on the first two dimensions, but lower in comparison with the French listeners' ratings. It

seems that FrR perceive Czech speakers of English as more reliable and competent than speakers with a French accent. Both FrR and CzR gave negative ratings to their “own-accented” speech on the third dimension, job effectiveness. This could be interpreted as a lower in-group acceptance.

The native English respondents attributed higher ratings to the native-speakers of English on all dimensions, but gave no negative ratings to either of the non-native speaker groups. While they awarded similar scores on the memory/insight and job effectiveness dimensions to both foreign-accented speech groups, they rated CzE more favourably than FrE on the first dimension, reliability/punctuality. However, they perceive both foreign-accented speech groups as more reliable than the accent bearers see their own kin. Similarly, it should be noted that the NER gave positive ratings on the job effectiveness dimension to both foreign-accented speech samples, unlike the negative ratings for “own-accented” speech delivered by the French and Czech listeners.

4.4 Discussion

The results suggest that, in general, foreign-accentedness does influence the perception of speaker status in the expected direction, with native speakers of English receiving superior evaluations. Our initial hypothesis is thus validated.

The results of Stage 1 (where listeners had a pair of speakers to choose from) show a clear preference for native speakers of English over the non-native ones (*cf.* Fig. 4-1); in addition, Czech speakers were selected as superior more often than French speakers, regardless of listener group (Fig. 4-2). On the side of French listeners, this out-group preference for Czech speakers of English was unexpected, especially as the French participants had not reported extensive familiarity with learning a Slavic language, and also because it is known that people tend to favour speakers who sound similar to themselves (see, e.g., Preston, 1999). The Czech listeners did not show a very strong out-group preference but they did choose French- over Czech-accented speech more often than the other groups. One explanation for this asymmetrical finding could be that the status of a native-speaker model is stronger in France than in the Czech Republic: “if you can’t say it properly, don’t say anything at all”, as one French participant put it. If the opposite holds true, or if the position is less extreme in the Czech Republic, this would presumably lead Czech participants to be more tolerant of foreign-accented speech. As harsh judges of a French accent, French participants would be expected to prefer

speech with a lesser-known accent to their own, identifiably French-accented speech.

Interestingly, the native English listeners did not show as categorical an in-group preference as might be expected. In both the English–Czech and English–French pairings, they preferred Czech-accented speech and French-accented speech proportionately more often than the Czech and French native respondents did. It might therefore be useful to speak of the degree of accent tolerance or acceptance, instead of simply preference or rejection. It could be argued that because native speakers of English tend to hear a greater variety of accented Englishes (see Table 4-1, last column), they are more likely to be tolerant of such varieties. However, the inverse explains neither the French out-group preference nor the Czech participants’ relative “leniency” when judging French-accented speech. A possible clarification is perhaps hinted at by a small informal post-hoc test in which we asked 14 young Vietnamese listeners to assess accentedness in our samples. The test was carried out in Hanoi with listeners who were familiar with neither Czech- nor French-accented English. These listeners perceived our FrE speakers as “more accented” than the CzE speakers.

In section 4.1.2, several studies were mentioned which examined the attitudes toward foreign-accented teachers in America. One should keep in mind, however, that the context of our study cannot be directly compared to research using American undergraduate students. In America, students would usually pay much higher tuition fees and they may feel entitled to a “return on investment”; they might therefore attribute more direct, monetary value to receiving intelligible and accent-free speech from their instructors. They might feel less tolerant of deviations from a norm when such deviations endanger their return on investment. In other words, the stakes are high when tuition fees are more substantial than in Europe. The English-speaking participants in our study, though they were native speakers, can also hardly be compared with American students in the US: since they were studying abroad, they may represent a self-selecting population which is more tolerant of foreign-accented speech.

In Stage 2 of our experiment, in which listeners were asked to intuitively assess speakers on several dimensions related to competence, the native speakers of English received much better evaluations by all three groups of listeners. Again, however, the two non-native groups differ, with Czechs being evaluated more positively than the French on all three dimensions (Fig. 4-3). A more detailed analysis reveals strong out-group preferences on the part of the French respondents and a less categorical preference for “own-accented” speech by the Czech respondents. Own-group ratings were harshest (reaching negative scores)

on the job effectiveness dimension for both Czech and French listeners. The native-speaker listeners distinguished themselves by providing no negative rankings; this may be related to the above-mentioned acquaintance with and, therefore, greater tolerance of foreign-accented speech in general. Since native speakers of English would have engaged in professional encounters with foreigners, their intuitive judgements of foreigners' competence may be less biased based on accent alone.

At least four factors could have significantly affected the results and could be controlled for in future. First, the field of study of the listeners could have influenced their judgments; English majors may be more aware of or sensitive to accented speech than students in other fields. Second, a more representative sample of native-speaker participants would be those studying in their own country, instead of students studying abroad. Third, respondents were told to imagine they were choosing employees for a company, which means that a specific type of product or service may have influenced their choice. For example, a French accent might be rated more preferably for selling certain foods or luxury products, clothing, or perfumes. Lastly, the scenario applied in this study could be modified, as there may be other dimensions of competence relevant for the academic context, apart from those investigated here – reliability, memory/insight, and effectiveness.

Our study suggests that the implications for English-medium instruction in European universities are not as pessimistic as one might expect. A large body of research has examined training programs specifically aimed at improving the intelligibility of international teaching assistants (ITAs) in the United States and Canada. The research findings have influenced these programs, for example by providing evidence of the key role played by lexical and phrasal stress (Hahn, 2004) or intonational groupings (Slater et al., 2015). These findings could in turn be exploited by European teachers who wish to improve their intelligibility for lecturing in a non-native language to international student bodies. For example, De Meo's (2012) findings provide concrete advice for non-native speakers of Italian, whose spoken Italian was evaluated by native speakers:

... the correlation between foreign accent and credibility is delivered by comprehensibility: poor comprehensibility generally lowers the credibility level of an utterance. When comprehensibility is high, a reduced tonal range and longer silent pauses, i.e. the suprasegmental features of the utterance, determine a significant increase of trustworthiness. (3)

An innovative tack to the issue was also expressed by Rubin and Smith (1990), where they argue the problem is “owned” by both the teaching

assistants and the students. Given that it is unlikely that training will eliminate the level of accentedness which is an ethnic marker and which triggers expectations about teaching ability, Rubin and Smith concluded that it is American students who need to be trained “to listen to accented English and to distinguish levels of accent” (350). To our knowledge, such perception training is rarely put in place; however, Derwing and Munro (2014) provided an excellent summary of research studies which explore the extent to which processing effects and subconscious bias can be overcome through explicit training. The next question would not only revolve around whom to train – teachers, administrators and/or students – but also what type of training to provide. One effective approach would consist in implementing training similar to that used by Baese-Berk, Bradlow & Wright (2013), where listeners were trained to generalize their perceptual gains to novel speakers and accents. Derwing and Munro concur that “NS reactions to L2 accented speech can be mitigated through training, perspective-taking exercises, and carefully managed contact activities” (2015: 152). Similarly, Weyant (2007) encourages us to direct efforts not at accent reduction but at NS-listener adaptation. Not only is such adaptation feasible, it can also be rapid (Clarke & Garret, 2004).

A more controversial top-down approach could be seen in 2013 at Milan’s Politecnico, where all professors and support staff were attending weekly English classes – against their will, for some. This raises other issues: one blog cites an Argentinian student who finds it ironic that “he had to pass a stringent English exam to get in, when many of his professors would flunk it. ‘You have two kinds of teachers here,’ says Hualpa. ‘The ones who have done a PhD outside Italy – they speak clear English; and the Italian ones who learned English locally with an Italian cadence. Even for the international students we say, *You don’t speak well.*’⁷ The exact meaning of “clear”, “Italian cadence” and “well” is open to debate but one thing is certain: it is an unavoidable fact that accented speech does influence how speakers are perceived and this fact must not be ignored in situations of unequal status with real stakes.

4.5 Conclusion

The present study points to the complexity of the perception of socio-indexical characteristics, especially when it is based uniquely on auditory

⁷ <https://patrickcox.wordpress.com/2013/06/12/elite-italian-university-meets-resistance-as-it-tries-to-go-all-english/>

cues. Two things are clear. First, in studies which examine the evaluation of speakers' personality – whether in their native language or when speaking a foreign language – great emphasis must be put on experimental design, specifically on eliminating any undesirable variables. Second, speaker-perception studies can be productive in relation to European higher education trends. This type of language attitude research offers a concrete means of obtaining valuable evidence about the potential impact of foreign-accented speakers teaching in English – or any other non-native language. Therefore, if evidence-based decisions are to become more widespread in institutional language policy, more studies of this type should be undertaken. Furthermore, foreign-accented speech concerns both speakers and listeners. European universities which implement CLIL should at the very least experiment with listener training, to improve their students' ability to process foreign-accented speech and to encourage greater acceptance of diversity. This could be an excellent opportunity for European higher education to promote ingenious solutions to a global issue.

References

- Baese-Berk, M. M., Bradlow, A. R. & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America*, 133/3, pp. 174–180.
- Bonnet, A. (2012). Towards an evidence base for CLIL: How to integrate qualitative and quantitative as well as process, product and participant perspectives in CLIL research. *International CLIL Research Journal*, 4/1, pp. 66–78.
- Bradlow, A. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106/2, pp. 707–729.
- Brennan, E. M. & Brennan, J. S. (1981). Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research*, 10, pp. 487–501.
- Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y. & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. *Language and Communication*, 22, pp. 171–185.
- Brown, P. & Levinson, S. (1979). Social structure, groups and interaction. In: K. Scherer & H. Giles (Eds.), *Social Markers in Speech*, pp. 291–342. Cambridge: Cambridge University Press.
- Clarke, C. M. & Garrett, M. F. (2004). Rapid adaptation to foreign accented English. *Journal of the Acoustical Society of America*, 116, pp. 3647–3658.

- Costa, F. & Coleman, J. A. (2010). Integrating content and language in higher education in Italy: Ongoing Research. *International CLIL Research Journal*, 1/3, pp. 19–29.
- Côté, P. & Clément, R. (1994). Language attitudes: An interactive situated approach. *Language and Communication*, 14, pp. 237–251.
- Creber, C. & Giles, H. (1983). Social context and language attitudes: The role of formality-informality of the setting. *Language Sciences*, 5, pp. 155–161.
- De Meo, A. (2012). How credible is a non-native speaker? Prosody and surroundings. In: M. G. Busà & A. Stella (Eds.), *Methodological Perspectives on Second Language Prosody Papers from ML2P 2012*, pp. 3–9. Padova: CLEUP.
- Derwing, T. M. & Munro M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42/2, pp. 476–490.
- Derwing, T. M. & Munro, M. J. (2014). Training native speakers to listen to L2 speech. In: J. M. Levis & A. Moyer (Eds.), *Social Dynamics in Second Language Accent*, pp. 219–236. Boston: Mouton de Gruyter.
- Derwing, T. M. & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*. Amsterdam: John Benjamins Publishing Co.
- de Zarobe, Y.R. (2008). CLIL and foreign language learning: A longitudinal study in the Basque Country. *International CLIL Research Journal*, 1/1, pp. 60–73.
- Gluszek, A. & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, 14/2, pp. 214–237.
- Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38, pp. 201–223.
- Helm, F. & Guarda, M. (2015). Improvisation is not allowed in a second language: A survey of Italian lecturers' concerns about teaching their subjects through English. *Language Learning in Higher Education*, 5/2, pp. 353–373.
- Henderson, A. (2014). Takes two to tango: Research into accent, comprehensibility and intelligibility and implications for CLIL. In: W. Szubko-Sitarek et al. (Eds.), *Second Language Learning and Teaching*, pp. 63–79. Lausanne: Springer International Publishing.
- Hroch, M. (1999). The Slavic World. In: J. Fishman (Ed.), *Handbook of Language and Ethnic Identity*, pp. 319–333. Oxford: Oxford University Press.

- Kavas, A. & Kavas, A. (2008). An exploratory study of undergraduate college students' perceptions and attitudes toward foreign accented faculty. *College Student Journal*, 42, pp. 879–890.
- Lev-Ari, S. & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46, pp. 1093–1096.
- Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about nonnative English speakers in the United States. *Journal of Sociolinguistics*, 7, pp. 348–364.
- Mai, R. & Hoffmann, S. (2014). Accents in Business Communication: An integrative model and propositions for future research. *Journal of Consumer Psychology*, 24/1, pp. 137–158.
- McCrae, R. R. & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52/1, pp. 81–90.
- Miller, G. R. & Hewgill, M. A. (1964). The effect of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech*, 50/1, pp. 36–44.
- Moyer, A. (2013). *Foreign Accent: The Phenomenon of Non-Native Speech*. Cambridge: Cambridge University Press.
- Munro, M. J. (2003). A primer on accent discrimination in the Canadian Context. *TESL Canada Journal*, 20, pp. 38–51.
- Munro, M. J. & Derwing T. M. (1995). Foreign accent, comprehensibility, and intelligibility in speech of second language learners. *Language Learning*, 45/1, pp. 73–97.
- Oppenheimer, D.M. (2008). The secret life of fluency. *Trends in Cognitive Science*, 12, pp. 237–241.
- Preston, D. (1999). *Handbook of Perceptual Dialectology*. Amsterdam: John Benjamins.
- Reber, R. & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8, pp. 338–342.
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33, pp. 511–531.
- Rubin, D. L. & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, 14, pp. 337–353.
- Slater, T., Levis, J. & Levis Muller, G. (2015). Spoken parentheticals in instructional discourse in STEM and Non-STEM disciplines: The interaction of the prosodic, ideational, and interpersonal resources in

- signaling information structure. In G. Gorsuch (Ed.), *Talking Matters: Research on Talk and Communication of International Teaching Assistants*, pp. 3–31. Stillwater, OK: New Forums Press.
- Sweet, H. (1900). *The Practical Study of Languages: A Guide for Teachers and Learners*. New York: Henry Holt & Co.
- Van de Craen, P. (2002). Content and language integrated learning, culture of education and learning theories. Retrieved from http://www.see-educoop.net/education_in/pdf/cont_lang_integ_learning-oth-enl-t06.pdf.
- Weyant, J. M. (2007). Perspective taking as a means of reducing negative stereotyping of individuals who speak English as a second language. *Journal of Applied Social Psychology*, 37, pp. 703–716.

PART II:

MAPPING THE DIFFERENCES IN SEGMENTS

CHAPTER FIVE

CHARACTER OF VOWEL REDUCTION IN CZECH ENGLISH

KRISTÝNA ČERVINKOVÁ POESOVÁ
AND LENKA WEINGARTOVÁ¹

5.1 Introduction

In his approach to pronunciation teaching and learning, professor Allan James has ingeniously come to terms with the changing linguistic landscape dominated by the transformation of English into the world's principal lingua franca. In comparison with the widely discussed and researched pronunciation programme Lingua Franca Core (Jenkins, 2000; Walker, 2010; Cogo, 2012) which stems from analysing misunderstandings among non-native users of English, James' concept of socially meaningful anglophony, largely embedded in social semiotics, offers different implications for pronunciation pedagogy. While the proponents of English as a Lingua Franca (ELF) tend to dismiss native accents from their investigations and predominantly focus on the changes English undergoes in various non-native interactions, the underlying principle of James' approach consists in revealing what the existing anglophone codes have in common and in identifying salient features that capture the nature of sounding English (James, 2013). Instead of the ELF segmental-oriented pronunciation syllabus supplemented by the development of accommodation skills, James proposes a balanced repertoire of both segmental and prosodic aspects that aim at sounding English, from which L2 users make informed choices in order to express social meanings. As the encounters with native and highly proficient non-native speakers is not likely to

¹ The first author was supported by the grant provided by the Czech Science Foundation, GAČR 14-08084S. The authors would like to thank all the recorded participants for their cooperation. Also, our sincere thanks go to Jan Volín for all the inspirational and helpful discussions about this paper.

decrease in the current European context (Berns, 2009), the pursuit of Englishness as the common denominator seems to prepare L2 speakers effectively for successful communication across a huge variety of interlocutors and settings.

Sharing the above outlined framework, we are thoroughly convinced that the process of vowel reduction, which involves the neutralization of vocalic qualities in unstressed positions (Laver, 1994: 516) and is often materialized as a neutral mid-central lax vowel schwa, significantly contributes to the specific sound of native English. Unlike the ELF approach, which tends to ignore vowel reduction, we suggest it should occupy a central place in pronunciation syllabi, particularly if designed for students coming from schwaless linguistic backgrounds. On both word and sentence level, the acoustically and perceptually inconspicuous character of schwa helps stressed parts stand out in the speech signal and thus create clear prominence contrasts essential for the natural flow of English rhythm.² Rhythmical distortions, on the other hand, caused either by stress misplacement or insufficient vowel reduction, may delay word recognition due to mismatches between the deviated surface forms and listeners' expectations and therefore hinder smooth message decoding. A growing body of literature points out that the increased listening effort, or in Munro and Derwing's (1995) terms the reduced comprehensibility, resulting from more demanding cerebral processing of accented speech (Van Engen & Peelle, 2014) may possibly activate dormant prejudices or trigger off largely subconscious social and behavioural reactions to non-native manifestations, such as various kinds of biases, negative judgements or discriminatory acts (Rubin, 1992; Lev-Ari & Keysar, 2010; Gluszek & Dovidio, 2010; Lippi-Green, 2012).

Although the change of vowel quality is by no means the only perceptual cue for identifying stressed and unstressed syllables, the others being length, loudness and pitch change (Cruttenden, 2014: 234–235), several studies emphasize its undisputed role in perceiving prominence patterning. For instance, the findings of Fear et al.'s (1994) research revealed that unstressed unreduced vowels create a clear-cut, intermediate category between full and reduced vowels. Nevertheless, in the subsequent verification task, in which perceptual ratings of naturalness of manipulated vowel tokens were compared with selected acoustic measurements, it was

² The term 'stress' refers to a phonological characteristic of a vowel or syllable, while 'prominence' denotes its phonetic manifestation (we are following here the terminology set by Bolinger, 1958). A stressed syllable therefore may be more or less prominent, depending on the phonetic characteristics of itself and its neighbourhood.

found out that spectral characteristics seem to outweigh duration and intensity, as listeners grouped unstressed unreduced vowels more consistently with stressed than reduced vowels. Similarly, in a study investigating the factors that hamper speech perception of Dutch-accented English (Braun et al., 2011), vowel quality turned out to be vitally important for the perceptual distinction between stressed and unstressed syllables. Moreover, it was not overridden by other examined cues, duration and spectral tilt in particular. The function of vowel reduction as an extra cue for disambiguating certain sets of English words was confirmed by Van Bergem (1995), who also suggested that certain “impoverishment” of the speech, to which schwa hugely contributes, might be necessary for achieving natural and intelligible productions.

5.1.1 Prominence patterns in Czech and English

Unlike English, in which the process of vowel reduction plays a crucial role and schwa represents the most frequent sound, the vocalic inventory of standard Czech contains only six distinct vowel qualities /i, ɪ, ε, a, o, u/ and lacks schwa. Despite this fact, vowel reduction is not a completely unfamiliar phenomenon for Czech speakers, although the process is mainly associated with lower and more colloquial speech styles (Volín, 2010). Both vowel and consonant reduction may be employed for rhythmical reasons in affective spontaneous speech, possibly with the aim of increasing prominence contrasts, too. However, more research is needed to confirm this informal observation.

It has been repeatedly pointed out that vowel reduction is a systematic correlate of word stress in English together with length, loudness and pitch. A typical English stressed vowel is longer, louder, higher and less reduced than the neighbouring unstressed ones. This is not the case in Czech. This language has a fixed stress with a delimitative function of marking word, or more precisely stress group boundaries. Its position is always tied to the first syllable of a stress group, and stress misplacement in Czech may lead to the perception of foreign accent. On the other hand, English features a variable stress that can be in virtually any position in a word and can even change its meaning (re'cord vs. 'record), which presents an unknown concept for a naïve monolingual Czech speaker. A stressed syllabic nucleus in Czech is neither systematically longer, louder, nor higher than its surroundings and the vowels require full qualities in stressed, as well as unstressed positions (Palková, 1994: 170). Moreover, Czech has highly regular letter-to-sound correspondences, which makes the speakers increasingly sensitive to the written form of words. The

Czechs have no experience with vowel sounds without a corresponding letter and are used to predicting both the quality and quantity of a vowel from its graphemic representation. Obviously, these clashes create considerable demands on speakers learning English as a foreign language.

5.1.2 Current research in Czech English

In the past decade the issue of prominence and related aspects has enjoyed a keen interest among researchers of Czech English in both perception and production domains. In the former, Czech listeners demonstrate a weakened perceptual sensitivity to English vowel reductions (Poesová, 2012), presumably as a result of a speech illusion called mutation (Sebastián-Gallés, 2005), during which the schwa sound present in the speech signal is filtered through the sieve of Czech short vowels and often perceptually transmuted into /ɛ, a, o, u/ often in accordance with the spelling of the given word. Furthermore, Czech listeners face greater difficulties identifying strong beats in English words when individual acoustic attributes indicating syllable prominence are in conflict, which is likely to occur in polysyllabic words carrying both primary and secondary stresses (Skarnitzl, 2005). It is exactly in these situations in which schwa may remain the last resort in perceived prominence, although this claim requires experimental verification for Czech English.

As far as production is concerned, Czech-accented English can be generally characterised as lacking clear temporal contrasts, that is to say stressed syllables tend to be shorter and unstressed syllables longer than in native speech. The strong inclination of Czech speakers of English towards equalizing vowel lengths affects the remaining acoustic correlates of stress, too (Volín et al., 2013; Weingartová et al., 2014). At a closer inspection, however, individual words bear witness to various strategies for signalling prominence in Czech English. These largely depend on a number of factors, such as phonological structure of words, their textual frequency and resemblance to Czech counterparts (Volín, 2005). Furthermore, the level of speaker exposure to the target language proved to play a certain role: more advanced speakers show a more native-like treatment of prominence patterns (Weingartová et al., 2014).

The ability of Czech speakers of English to produce local prominence contrasts was scrutinized in two studies in which the ratios and differences of acoustic parameters in stressed and adjacent unstressed vowels were compared. The context of the first research was the effectiveness of pronunciation teaching in the area of vowel reduction. The subjects were 12 to 13-year-old pupils of a Czech lower-secondary school at a pre-

intermediate level. The findings revealed substantial improvement only in the temporal domain; vowel obscuration was identified to a lesser degree and loudness or pitch were not utilized by the speakers at all (Poesová, 2012). Interestingly, these findings accord with Gralińska-Brawata's study in which the effect of instructed learning on Polish speakers' ability to better reduce in English was explored (2015). Similar conclusions were also drawn in the second experiment comparing Czech-accented productions at two distinct stages of phonological acquisition (beginners and intermediates) with a native performance (Weingartová et al., 2014). The speech behaviour of Czech speakers neatly illustrated certain fragmentation and low interconnectedness of the acoustic features for signalling (non-)prominence as opposed to their native English counterparts who exploited them in a more cohesive way. Both studies provide ample evidence that temporal contrasts seem to be the easiest to acquire for Czech learners of English, which is suggested to be connected with the use of duration in phonological contrasts between Czech long and short vowels.

Apart from Czech, there are naturally many other languages in which the process of vowel reduction is largely absent (Polish, French), restricted to certain contexts (German) or functions differently (Portuguese). Consequently, non-native speakers are likely to encounter partly similar and partly specific interferences in contrasting English prominence, based on their mother tongue phonologies. In this respect, Czech and Polish Englishes, both with L1 of Slavic origin, seem to share the insufficient vowel reduction encompassing both vowel quality and quantity (Gralińska-Brawata, 2015; Rojczyk, 2012). Another fact that needs to be accounted for in cross-linguistic comparison is a group of factors that may exert a discernible impact on the scope and occurrence of vowel reduction. Among the most investigated ones belong word stress, word class, frequency of words, speaking style, vowel type or position of the vowel in the word. All of these criteria except for vowel type were investigated, for example, in French conversational speech (Meunier & Espesser, 2011). Reduction in weak forms of function words was analysed both acoustically and perceptually in intermediate Brazilian English (Fragozo, 2011) and only acoustically in advanced German English (Sönning, 2014). High versus low frequency words were examined by Swerts et al. (2003) in spontaneous Dutch. To conclude, negative transfer from L1 sound systems tends to determine and shape the character of vowel reduction mainly through a combination of lexical and prosodic factors.

5.1.3 Research aims

The purpose of this chapter presenting the findings of a qualitatively-oriented study is to map the acoustic characteristics and occurrence of Czech schwa using the British standard as a reference accent for all measurements. Specifically, we aim at providing a detailed examination of Czech users' (in)ability to reduce English unstressed vowels and unveiling the possible reasons for their anticipated failure to do so. Last but not least, potential implications for L2 pedagogy will be drawn and practical recommendations for English as a Foreign Language (EFL) classes suggested. In the previously described studies, the neutral vowel served for other purposes, for example, as a means of assessing phonological acquisition, the degree of foreign-accentedness or the benefits of pronunciation teaching. In comparison, the current text focuses solely on schwa and attempts to provide a comprehensive analysis of its nature and occurrence in Czech English. In addition, the sample of schwa realizations analyzed in this study is substantially high, which should make our results more generalizable.

5.2 Schwa in Czech English

5.2.1 Data and method

Eight speakers were recorded in the sound-treated studio of the Institute of Phonetics in Prague, with an electret microphone IMG ECM 2000, soundcard SB Audigy 2 ZS, 32-kHz sampling frequency and 16-bit resolution. Four of the speakers were native Czech speakers of English (two males and two females) at an intermediate level. They learned English as their second or third language in a formal setting and did not undergo any phonetic training. However, the participants were fluent, intelligible and confident users of English with a discernible Czech accent. The reference accent was represented by four British speakers (two males and two females again) living in Prague, but retaining a non-accented Southern Standard British English.

All eight participants were recorded reading a list of 112 sentences from the NonCol corpus built at the Institute of Phonetics in Prague. They familiarized themselves with the text prior to recording and were instructed to read the sentences in a fluent, natural manner (i.e., not acting or news-reading) which was monitored by the experimenters. The material was manually labelled by the authors in the Praat software (Boersma &

Weenink, 2014) at the word, phoneme and phone level, with a special focus on vowel qualities.

For further analyses, all positions with a canonical schwa occurrence (according to Wells, 2000) were investigated. This comprised all unstressed syllables in lexical words and weak forms of grammatical words (such as *a*, *the*, *to*, *was*, *at*, *from*, etc.). The total amount of these ‘potential’ schwas in all recordings was 4613.

Apart from the actual pronounced vowel quality, which was perceptually determined by the experimenters (schwa, rhotic schwa, full vowel or other phones), the following acoustic measurements were taken from every item:

- duration (in ms; normalized by the speaker’s average articulation rate over all sentences),
- intensity (in dB; normalized by the average of the utterance),
- formants (in Hz and ERB; F1, F2, F3),
- formant bandwidths (in Hz and ERB; F1, F2, F3).

The formants and formant bandwidths were measured in Praat, with the Burg method, maximum of 4 formants and 4000 Hz. We measured each of the formant’s mean value in the middle third of the vowel duration. The formants represent the acoustic correlates of vowel quality; duration and intensity of a vowel correspond to its length and loudness, respectively.

5.2.2 Acoustic parameters of schwa

In this first set of analyses, we wanted to describe and compare the realizations of English and Czech English schwa, therefore only the items where the speaker actually pronounced a schwa were taken into account. The total amount of cases for the British speakers was 2064 and only about a half, 1046, for the Czech speakers.

The schwas in Czech English proved to be significantly longer than those of the native speakers (see Fig. 5-1). While native English schwa is on average 46 ms long, the non-native counterpart is distinctly longer, 53 ms on average. A one-way ANOVA proves the difference to be highly statistically significant: $F(1, 3109) = 107.4; p < 0.001$. Note that the values are normalized, so the differences in speaking tempo between the individual speakers should be controlled for.

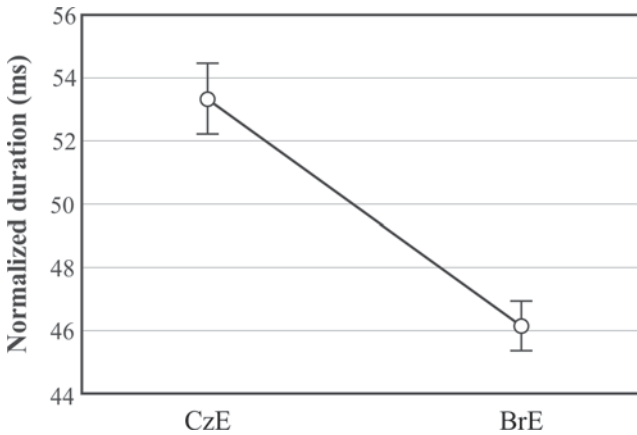


Fig. 5-1. Normalized duration (in ms) of schwa realizations for Czech English (CzE) and British English (BrE) speakers. Whiskers indicate the 95% confidence interval.

In the vowel formant analyses, only non-rhotic schwas were included, because rhoticity influences F2 and F3 considerably (see, for example, Heselwood & Plug, 2011). This filtered material yielded 2893 items, from which 2048 were pronounced by the British speakers and 845 by the Czech. The measurements in Hz and in ERB produced very similar results, so only ERBs are presented further, since they better represent the nonlinearity of human hearing. In accordance with Volín et al. (2013), there seems to be a significant difference in the bandwidth of F1, which is smaller in Czech speakers than in the British: $F(1, 2891) = 159.7$; $p < 0.001$. This means that the first formant of Czech English schwa has a “sharper” peak in the frequency spectrum than the British schwa.

The difference in bandwidths of F2 was also statistically significant, but surprisingly enough, in the other direction – Czech speakers showed a larger F2 bandwidth: $F(1, 2891) = 8.5$; $p = 0.004$. A two-way ANOVA with factors LANGUAGE and GENDER reveals that this difference is caused by male speakers only. The result is shown in Fig. 5-2.

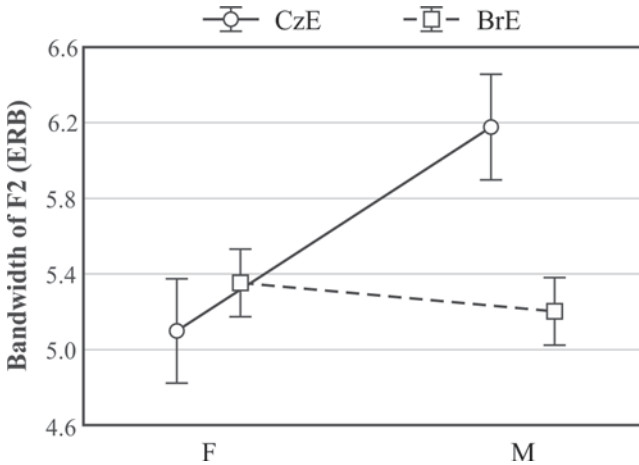


Fig. 5-2. Values of bandwidth of F2 (in ERB) of schwa realizations for Czech English (CzE) and British English (BrE) speakers, broken down by gender (F – female, M – male). Whiskers indicate the 95% confidence interval.

Even though the female speakers show the tendency for Czech schwas to have smaller bandwidths, as it was for F1 bandwidth, they are overridden by the male speakers. Since the number of speakers is not large enough (two for each language and gender), it cannot be hypothesized at this point whether this is indeed a general difference between male and female speakers or just an idiosyncrasy of our speaker sample. The results for F3 bandwidth were not significant.

The frequency of the first formant was similar for both groups and the difference was not statistically significant. The second formant, on the other hand, was on average significantly lower for the Czech speakers: $F(1, 2891) = 9.1$; $p = 0.003$, see Fig. 5-3.

A closer look reveals that this difference is again caused by the male speakers only – Fig. 5-4 shows that F2 of native and non-native female speakers is virtually the same.

At first sight, this result seems to differ from the above mentioned study (Volín et al., 2013), where no difference in F2 was found. However, note that the authors of the earlier study employed female participants only. Lowering of F2 suggests that the male Czech speakers are pronouncing their schwas more in the back part of the vocalic space, in the direction of /o/, but not enough to be perceived as such – the Czech /o/ has an average F2 value of 15.66 ERB (Skarnitzl & Volín, 2012). The third formant showed again no difference between British and Czech speakers.

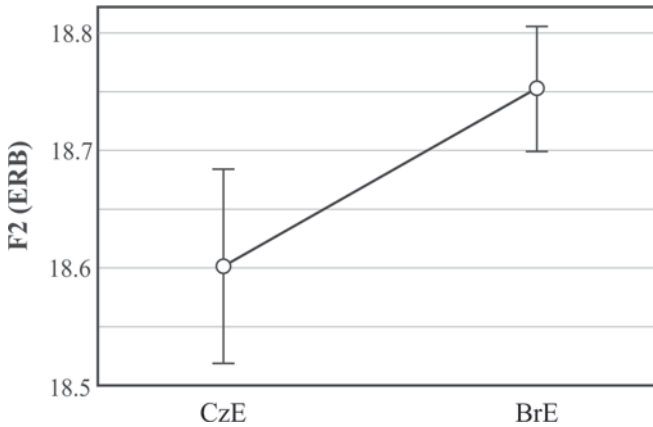


Fig. 5-3. Values of F2 (in ERB) of schwa realizations for Czech English (CzE) and British English (BrE) speakers. Whiskers indicate the 95% confidence interval.

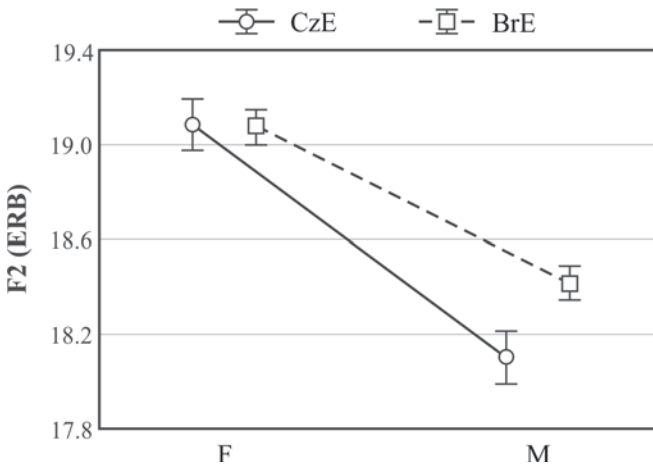


Fig. 5-4. Values of F2 (in ERB) of schwa realizations for Czech English (CzE) and British English (BrE) speakers, broken down by gender (F – female, M – male). Whiskers indicate the 95% confidence interval.

If we overlay our Czech English schwa values with the Czech reference vowel diagram from Skarnitzl and Volín (2012), we can see that schwa indeed occupies a separate position in the Czech vowel space, see Fig. 5-5.

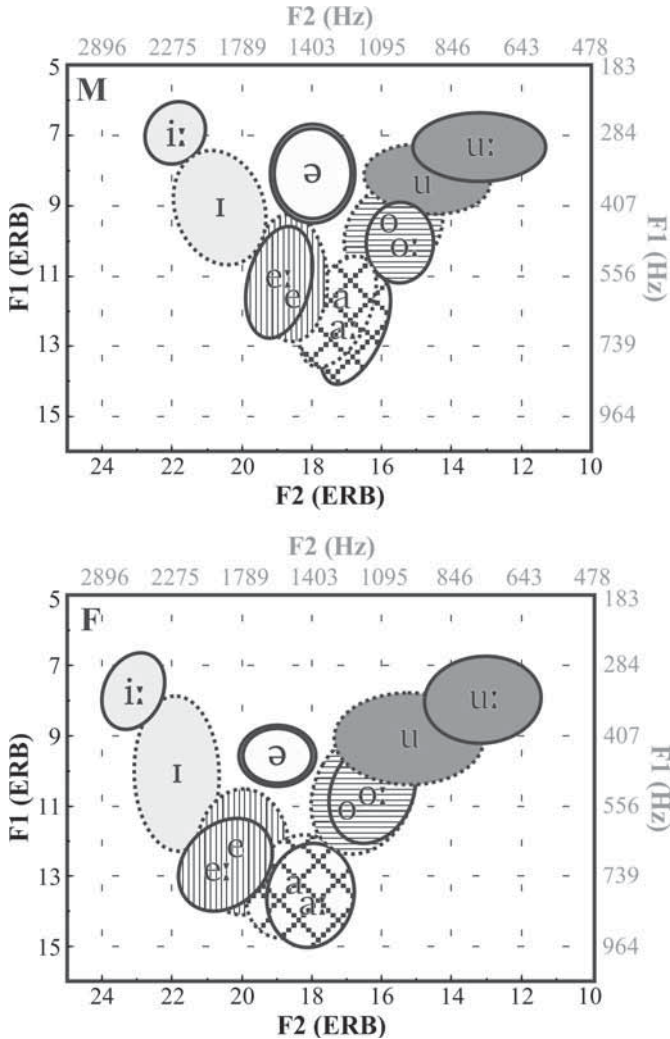


Fig. 5-5. Vowel space of Czech monophthongs (Skarnitzl & Volín, 2012), with overlaid schwa values from this study. Male values (M) are at the top, female (F) at the bottom. Ellipses indicate one standard deviation.

Table 5-1 shows mean values of the first two formants in Hz and ERB, for male and female speakers individually. The ratio of F1:F2 in Hz, which is traditionally thought to be 1:3 following Fant's single tube resonance

model (Fant, 1960), is actually closer to 1:4 in our sample. This tendency is mainly caused by the values of F1, which are below the traditional 500 Hz, even for the females. This is true for our native British speakers as well, their average F1 values are 380 Hz (male) and 434 Hz (female). This is in line with the results of Flemming & Johnson (2007), who found out that the average F1 values for non-final schwas are significantly lower than for final schwas. Our data consist of non-final schwas only.

	Hz	std. dev.	ERB	std. dev.
F1 male	362	125	8.2	1.4
F1 female	449	65	9.6	0.9
F2 male	1437	225	18.1	1.3
F2 female	1615	211	19.1	1.1

Table 5-1: Average values and standard deviations (in Hz and ERB) of Czech English schwa realizations.

The last measurement focused on the intensity parameter. In this analysis, all schwa realizations were taken into account. It is clear from Fig. 5-6 that intensity is significantly higher for Czech speakers than for the British: $F(1, 3109) = 103.2; p < 0.001$; the mean difference is 1.5 dB.

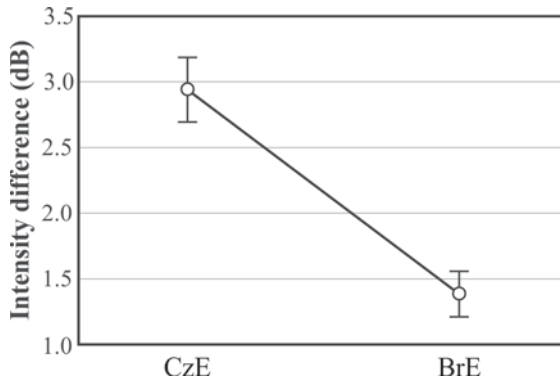


Fig. 5-6. Values of intensity difference from utterance average (in dB) of schwa realizations for Czech English (CzE) and British English (BrE) speakers. Whiskers indicate the 95% confidence interval.

Values in Fig. 5-6 are not absolute numbers, but a relative difference from utterance average. For example, a value of 1 dB means that the schwa's intensity is 1 dB higher than the utterance average. Measuring "absolute" intensity is impossible in this context, because many other variables would interfere, such as microphone setting, distance of the speaker's mouth from the microphone or different overall vocal effort of each speaker. Therefore any intensity measurements need to be normalized in some way.

5.2.3 The nature of schwa

The nature of schwa in Czech vs. British English has been determined on the basis of in-depth acoustic analysis. Schwa produced by Czech speakers proved to be distinctly longer than its British counterpart, by 7 ms on average. This may not seem a lot, however, the value lies above the just noticeable difference threshold for vowel duration, which Nootboom and Doodeman (1979) set to 5 ms for segments of 90 ms duration. Perceptual verification involving a comparison of two identical words taken from the Czech and British English data confirms audible length differences of the target sound. Furthermore, the durational values were normalized in order to control for potential fluctuations in the overall speaking rates caused, for instance, by Czech respondents' slower tempo resulting from a higher cognitive load during the experimental task.

Similarly, the intensity of schwa turned out to be markedly higher in Czech as opposed to British English, by 1.5 dB. Again, this may not appear to be a strong contrast, but it lies well over the just noticeable difference for intensity of sounds, which is thought to be around 0.5 dB, depending on the frequency of a sound (Johnson et al., 1993). As for the spectral characteristics, the non-rhotic schwas from the Czech English corpus display a smaller bandwidth of F1 than the British schwas and in case of the male speakers a larger bandwidth of F2. Also, our results show that Czech male speakers realize their schwas with a lower F2 than the British counterparts. Overall, our findings indicate that despite achieving the desired vowel, differences in duration, intensity, as well as vowel quality between Czech and British English schwas prevail. These minor, yet clear distinctions may allow space for perception uncertainty about the sound identity on the part of listeners. First, Czech English schwa might be perceived as a full vowel by native English listeners due to its less reduced acoustic characteristics, and second, it can disrupt the expected rhythmic patterns.

5.2.4 Occurrence of schwa

In the second series of analyses, the set of examined items was extended in order to cover all positions in which schwa should have occurred according to the rules of British standard pronunciation (Wells, 2000). The occurrences of schwa, full vowel, syllabic consonant and other elements expressed in percentages are summarized in Table 5-2 below.

	BrE speakers		CzE speakers	
	count	perc.	count	perc.
Schwa	2064	89.2%	1047	45.5%
Full vowel	78	3.4%	1023	44.4%
Syllabic cons.	156	6.7%	182	7.9%
Other	15	0.7%	50	2.2%
<i>Total</i>	<i>2313</i>	<i>100%</i>	<i>2302</i>	<i>100%</i>

Table 5-2: Counts and percentages of schwa realizations in canonical schwa positions for both groups of speakers, English native (BrE) and Czech native (CzE).

From the total number of 2313 instances in the text, which should be canonically realized as schwa, the British speakers indeed pronounced almost 90% of them as schwa. The other realizations included most often syllabic consonants (6.7%; in the words ending with a sonorant like *bicycle*, *example*, *location*, *migration*, *purple* or *some*) and slightly less often full vowels, which were typically unstressed, but not reduced (3.4%; they mainly appeared in the words *some*, *to*, *and* and *interest*). The last category involves cases of complete elision of the vowel.

The total number of canonical schwas for the Czech speakers of English was 2300. The inequality of the total counts between the two groups is due to a few cases of disfluencies, slips of the tongue or changes in words that had to be discarded. The Czech speakers realized almost half of the analyzed items (45.5%) as schwas (rhotic schwas comprised 8.8% of the sample and 19.3% of all Czech English schwas) and the other half (44.4%) retained full vowel qualities. The number of syllabic consonants was similar to the native speakers and also covered similar words. The Other category contained elisions, substitutions with a diphthong (such as *curtains* pronounced as [kʊt̪eɪns]) or a monophthong and a consonant (such as *her* pronounced as [hɛr]).

The correctly produced items in Czech English speech were identified mainly in the articles *the* and *a*, and in words ending with *-er* or *-re* (such

as *never*, *covers*, *metre* or *offered*). These realizations were very often rhotic (in 63%). Other weak forms (such as *was*, *to*, *some* or *from*) were in 87% of the cases pronounced with a full vowel. The weak forms of *her* and *were* created an exception in this data set with an increased number of schwa manifestations, probably because of the endings *-er*. Most schwas in our corpus were represented by the fairly frequent articles *the* and *a*, a total of 455 realizations. If not a schwa, a full vowel was employed, in case of *the* it was most commonly [i:], in case of *a* the vowel [ɛ] always replaced schwa.

Interestingly, a strong tendency was identified in content words with and without Czech counterparts. The words without a similar Czech equivalent were pronounced more often with schwa (in 56% of cases, full vowel was realized only in 26%), than the words with a Czech counterpart (e.g., “America” vs. “Amerika”). Those tended to be pronounced with a full vowel instead of schwa (schwa was realized in 37% of these words, full vowel in 44%). This difference proved to be statistically highly significant: $\chi^2(1; n=732) = 39.7; p < 0.001$.

5.2.5 Schwaful or schwaless?

The sound system of standard Czech can be described as schwaless, whereas native English accents abound with the reduced vowel schwa. Both linguistic backgrounds have participated in shaping the interlanguage of the investigated Czech speakers of English. Our data suggest a systematic usage of schwa in certain contexts, for examples in articles and content words ending in *-er* or *-re*. Nevertheless, in other contexts the respondents did not succeed in minimizing the mother tongue interference, which is by no means the only force behind the substitution of schwa with full vowel qualities in Czech English. Our last analysis encompassed primarily the words lacking vowel reduction, and the findings indicate that the speech behaviour of Czech speakers of English seems to be affected by a combination of one or more factors, some of which had been observed in the previous studies (see part 1.2). As the extent to which individual aspects contribute to the overall tendency towards insufficient reduction in Czech English seems to be hard to assess, no significance should be attached to the selected order in the following list of factors.

Spelling definitely enhances full vowel realizations, e.g., *velocity* pronounced as [ˈvelosɪtɪ] or *possessed* as [ˈpɒsɛst]. This factor is often accompanied by stress shift to the first syllable, which is a very common feature in Czech English (Volín & Weingartová, 2014).

Transfer from the mother tongue, particularly in words with Czech near-equivalents, e.g., *technology* pronounced as [ˈtɛknoloʒɪ] (the Czech word is [ˈtɛxnologie]) or *positive* as [ˈpozɪtɪf] (in Czech [ˈpozɪtɪvɲiː]). This influence cannot be detached from the spelling factor mentioned above because of the regular letter-to-sound correspondences typical of Czech.

Overgeneralization of pronunciation rules may explain the tendency of our subjects not to choose the vowel that corresponds with the orthographic symbol in certain words, for example, *about* pronounced as [ˈɛbaut] and not [ˈabout], *constructing* sometimes produced as [ˈkanstraktɪŋk] not [ˈkonstraktɪŋk] or *earlier* as [ˈiːrlɪɜː] and not [ˈɛːrlɪɜː]. It seems that at the intermediate level of proficiency, the speakers have already noticed or learnt that some graphemes often correspond to different vowel qualities. For instance, they are aware of the fact that the letter *a* in basic words like *cat*, *black* or *care* requires an *e*-like quality and they are likely to transfer this knowledge to other words such as *about*, *America*, *attack* or *admit*. In words like *constructing*, the [a] sound in the first syllable may be the result of overgeneralization from words like *come* or *sound*. The case of *earlier* seems to point to an obvious culprit *ear*.

The phenomenon of **perceptual illusion** (Sebastián-Gallés, 2005) strengthens or otherwise interacts with all the previously discussed factors. Czech speakers tend to map phonetically similar English vowels onto the existing L1 perception categories which may subsequently constrain production accuracy (Flege, 1999). For example, before a new category for the English sound /æ/ is formed, it tends to be attracted by the Czech vowel /ɛ/ resulting in *cat* being perceived and produced as [kɛt]. Similarly, English /ɑː/ and /ʌ/ are drawn to the /a/ category. The perceptually fluid character of schwa often succumbs to one of the following Czech vowels: /ɛ, a, o, u/.

Ignorance of reduction in weak forms also plays a significant role. Many Czech speakers may be wholly unaware of the existence of weak forms and their function in creating natural rhythmical patterns. Their repertoire is thus limited only to strong forms of grammatical words. In our corpus, numerous examples were found, for example, *was* pronounced as [wos], *can* as [ken], *to* as [tu], *some* as [sam] or *as* as [ɛs].

The last factor of **word frequency** is very likely to account for successful productions of the items *a* and *the* in the majority of investigated cases. Both articles represent words with extremely high frequency. Exemplars of frequent words are expected to be internalized first and their pronunciation is then transferred to less familiar or unknown words. For instance, the inclination towards overgeneralization seems to

stem from the usual rather than irregular letter-to-sound correspondences and from more rather than less frequent words.

In this second section we have examined the factors that may prevent Czech intermediate users of English from attaining vowel reduction in canonical positions. One needs to bear in mind that the individual factors are hugely interconnected and therefore it is difficult to determine which of them exerts the greatest influence. Even without knowing which factor plays a dominant role, the proposed categorization is believed to be didactically beneficial. Firstly, it may help teachers better understand the difficulties their students face in the target area. Secondly, the classification establishes the basis for devising activities that might help learners develop effective strategies necessary for overcoming the discussed obstacles (see section 3).

5.3 Summary and teaching implications

This chapter has looked into the nature and occurrence of schwa, the main representative of vowel reduction processes, in Czech English. In approximately half of the analysed speech material the respondents produced full vowel qualities instead of schwa. In the other half the target sound was pronounced, however, the measurements of the acoustic parameters confirmed its specific nature: Czech English schwa sounds longer (duration), louder (intensity) and its quality can be described as sharper and less reduced (F1 bandwidth) compared with the British English sample. Besides, Czech male speakers' schwa has a lower second formant which indicates a more backed position of their tongue.

As far as the order of acquisition is concerned, the respondents succeeded in reducing the vowel quality in unstressed syllables of lexical words to a larger extent than in weak forms of grammatical words, which accords with Sönning (2014) and Porzuczek (2010). Their research studies offer convincing evidence from German and Polish, respectively, about advanced students' preference for strong forms or, in other words, about their inability to reduce vowels in function words. Turning back to the Czech English corpus, the analysed items bear witness to the interplay of several factors contributing to incomplete or non-existent vowel reduction: spelling, transfer from the mother tongue, overgeneralization, perceptual illusion, ignorance of pronunciation rules and word frequency.

The current findings can help ELT practitioners set priorities in pronunciation teaching and approach the investigated area more effectively. Our suggestions involve more systematic work on suppressing the extensive reliance of learners on the written form using, i.e., the

technique of partial transcription Brə'ziliən 'childrən cən 'draw ə'mazing 'parrəts (Volín, 2002), exploit parallels between word and sentence level, i.e., *a 'tree* → *a'gree*, *Bra'zil* → *for 'Jill* (Lane, 2010), employ L1 equivalents to illustrate differences between English and Czech prominence patterning and change of vowel quality, i.e., *photographer* /fə'tɒgrəfə/ × *fotograf* /'fotograf/ or focus on weak form practice, particularly in the area of perception, i.e., drawing students' attention to /jɔ:/ and /jə/ that do not present two different words. In production-based activities aimed at weak forms, rhythmical minimal pairs can be employed, e.g., *four* /fɔ:/ *generations* × *for* /fə/ *generations*.

The last pedagogical recommendation relates to raising students' awareness of the process of vowel reduction, which is largely rule-governed. The potential advantage is speeding up the acquisition of the target phenomenon and/or overcoming certain fossilized forms which may be associated with foreign accentedness. The key features contributing to perceiving and producing prominence contrasts in English should not be kept apart in pronunciation classes and schwa should always be treated in the context of the neighbouring syllables or words (Poesová, 2015). Developing phonetic awareness may consist of simple tasks such as engaging students in observing, noticing and identifying prominence patterns in various language material and helping them see the hidden powers of schwa as a (non-)prominence indicator. Vowel reduction represents one of the building blocks co-creating the specific English sound and its gradual mastery may boost the confidence of non-native users of English.

References

- Berns, M. (2009). English as lingua franca and English in Europe. *World Englishes*, 28, pp. 192–199.
- Boersma, P. & Weenink, D. (2014). *Praat: doing phonetics by computer* (Version 5.4.09). Retrieved from <http://www.praat.org>.
- Bolinger, D. L. (1958). A theory of pitch accent in English. *Word*, 14, pp. 109–149.
- Braun, B., Lemhöfer, K., Mani, N. (2011). Perceiving unstressed vowels in foreign-accented English. *Journal of the Acoustical Society of America*, 129, pp. 376–387.
- Cogo, A. (2012). English as a Lingua Franca: Concepts, use, and implications. *ELT Journal*, 66/1, pp. 97–105.
- Cruttenden, A. (2014). *Gimson's Pronunciation of English*. London: Edward Arnold.

- Fant, G. (1960). *Acoustic Theory of Speech Production*. Mouton: The Hague.
- Fear, B. D., Cutler, A. & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97, pp. 1893–1904.
- Flege, J. E. (1999). Age of Learning and Second-Language Speech. In: Birdsong, D. (Ed.), *Second Language Acquisition and the Critical Period Hypothesis*, pp. 133–159. Mahwah, NJ: Lawrence Erlbaum.
- Flemming, E. & Johnson, S. (2007). Rosa's roses: Reduced vowels in American English. *Journal of the International Phonetic Association*, 37, pp. 83–96.
- Fragozo, C. S. (2011). Acoustic and perceptual characteristics of reduced vowels produced by speakers of English as a foreign language. In: *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: City University of Hong Kong.
- Gluszek, A. & Dovidio, J. F. (2010). Speaking with a non-native accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of Language and Social Psychology*, 29, pp. 224–234.
- Gralińska-Brawata, A. (2015). The acquisition of vowel reduction by Polish learners of English. In: *Proceedings of EPIP4 – 4th International Conference on English Pronunciation: Issues & Practices*, pp. 43–44. Prague: Institute of Phonetics.
- Heselwood, B. & Plug, L. (2011). The role of F2 and F3 in the perception of rhoticity: Evidence from listening experiments. In: *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: City University of Hong Kong.
- James, A. (2013). *Pronunciation as socially meaningful performance: beyond target codes and understandability*. Talk presented at 47th IATEFL Conference in Liverpool.
- Jenkins, J. (2000) *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- Johnson, J. H., Turner, C. W., Zwislocki, J. J. & Margolis, R. H. (1993). Just noticeable differences for intensity and their relation to loudness. *Journal of the Acoustical Society of America*, 93/2, pp. 983–991.
- Lane, L. (2010). *Tips for Teaching Pronunciation: A Practical Approach*. New York: Pearson Education.
- Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

- Lev-Ari, S. & B. Keysar. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46, pp. 1093–1096.
- Lippi-Green, R. (2012). *English with an Accent*. Oxford: Routledge.
- Meunier C. & Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, 39/3, pp. 271–278.
- Munro, M. & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, pp. 73–97.
- Nooteboom, S. G. & Doodeman, G. J. N. (1979). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America*, 67, pp. 276–287.
- Palková, Z. (1994). *Fonetika a fonologie češtiny*. Praha: Karolinum.
- Poesová, K. (2012). *Vliv systematického používání vybraných metod výuky výslovnosti anglického jazyka na percepci a produkci hlásky schwa u žáků ZŠ*. Dissertation thesis. Prague: Charles University in Prague, Faculty of Education.
- (2015). Under the baton of schwa. *Speak Out!*, 42, pp. 30–39.
- Porzuczek, A. (2010). The weak forms of TO in the pronunciation of Polish learners of English. In: E. Waniek-Klimczak (Ed.), *Issues in Accents of English 2: Variability and norm*, pp. 309–322. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Rojczyk, A. (2012). Vowel Quality and Duration as a Cue to Word Stress for Non-native Listeners: Polish Listeners' Perception of Stress in English. In: E. Waniek-Klimeczak & L. Shockey (Eds.), *Teaching and researching English accents in native and non-native speakers*, pp. 59–72. Berlin/Heidelberg: Springer Verlag.
- Rubin, D. L. (1992). Non language factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33, pp. 511–531.
- Sebastián-Gallés, N. (2005). Cross-Language Speech Perception. In: D. B. Pisoni, & R. E. Remez (Eds.), *The Handbook of Speech Perception*, pp. 546–566. UK: Blackwell Publishing.
- Skarnitzl, R. (2005). English word stress in the perception of Czech listeners. In: J. Čermák, A. Klégr, M. Malá & P. Šaldová (Eds.), *Patterns, A Festschrift for Libuše Dušková*, pp. 183–194. Praha: Kruh moderních filologů.
- Skarnitzl, R. & Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické listy*, 18/1, pp. 7–11.

- Sønning, L. (2014). Unstressed vowels in German learner English: An instrumental study. *Research in Language*, 12/2, pp. 163–173.
- Swerts, M., Kloots, H., Gillis, S. & De Schutter, G. (2003). Vowel reduction in spontaneous spoken Dutch. In: *Proceedings ISCA & IEEE Workshop on SSPR*, Tokyo, Japan, pp. 31–34.
- Van Bergem, D. R. (1995). *Acoustic and lexical vowel reduction*. Dissertation thesis. University of Amsterdam, Studies in Language and Language Use.
- Van Engen, K. J. & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, 8, pp. 1–4.
- Volín, J. (2002). Anglická střední středová nenapjatá samohláska. *Cizí jazyky*, 1/46, pp. 9–12.
- (2005). Rhythmical Properties of Polysyllabic Words in British and Czech English. In: J. Čermák, A. Klégr, M. Malá & P. Šaldová (Eds.), *Patterns, A Festschrift for Libuše Dušková*, pp. 279–292. Praha: Kruh moderních filologů.
- (2010). Fonetika a fonologie. In: V. Cvrček a kol., *Mluvnice současné češtiny*, pp. 35–64. Praha: Karolinum.
- Volín, J. & Weingartová, L. (2014). Acoustic Correlates of Word Stress as a Cue to Accent Strength. *Research in Language*, 12/2, pp. 175–183.
- Volín, J., Weingartová, L. & Skarnitzl, R. (2013). Spectral Characteristics of Schwa in Czech Accented English. *Research in Language*, 11/1, pp. 31–39.
- Walker, R. (2010). *Teaching the Pronunciation of English as a Lingua Franca*. Oxford: Oxford University Press.
- Weingartová, L., Poesová, K. & Volín, J. (2014). Prominence Contrasts in Czech English as a Predictor of Learner's Proficiency. In: N. Campbell, D. Gibbon & D. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody*, pp. 236–240. Dublin: TCD.
- Wells, J. C. (2000). *Longman Pronunciation Dictionary*. Harlow: Pearson Education Limited.

CHAPTER SIX

ARE WORD-FINAL CONSONANTS CODAS? EVIDENCE FROM BRAZILIAN PORTUGUESE ESL/EFL LEARNERS

PAUL JOHN AND WALCIR CARDOSO¹

6.1 Introduction

While first language (L1) acquisition is generally distinguished by across-the-board success, quite the opposite applies in second language (L2) acquisition. Particularly when it comes to the L2 sound system, learners typically fall short of the native-speaker target. The result is varying degrees of non-nativelike accent. We can question whether this state of affairs is of concern, in the sense that (generally unobtainable) nativelike pronunciation may be an inappropriate goal for the L2 learner. Comprehensibility and communicative competence is probably a more realistic and appropriate goal for L2 learners. Likewise, the reasons for an L2 speaker diverging from an L1 speaker in acquisition of the sound system can be debated. Perhaps it is due to biological factors associated with maturation (e.g., due to a Critical Period – Lenneberg, 1967), to psychosocial factors such as identity (e.g., Block, 2007) and motivation (Gardner & Lambert, 1972), or else to transfer from the L1 (e.g., Major, 2008). This last factor, involving the shortcut of transferring knowledge from the first to the second language, is largely responsible for a listener's ability to detect an interlocutor's L1, based on their foreign accent. In general terms, the study reported on in this chapter addresses the issue of

¹ We would like thank Léa Cardoso (collaborator) and two research assistants, Lorena Dantas and Jennica Grimshaw, for their assistance in recruiting the participants and conducting the interviews. We would also like to acknowledge the financial support of the *Social Sciences and Humanities Research Council of Canada* (SSHRC).

what exactly is being transferred and of what exactly needs to be learned in the L2 sound system. The specific focus is on the acquisition of stops in certain syllabic contexts by Brazilian Portuguese learners of English, namely in medial coda (e.g., *do[k]tor*) and in word-final position (e.g., *magi[k]*). The over-riding concern is to investigate how word-final consonants are syllabified: as codas (the prevalent view) or as onsets of empty nuclei (the competing view).

The L1 of an L2 learner is often revealed by characteristic patterns of substitution of L1 consonants and vowels for L2 phonemes (e.g., of /θ ð/ by either /t d/ or /s z/ – see Brannen, 2011; Trofimovich & John, 2011) or else of variable deletion and hypercorrect epenthesis of the novel phoneme (e.g., of /h/ by Francophone ESL learners – see John & Cardoso, 2008; Mah, 2011). However, learners' difficulties are not always due to a mismatch between the L1 and L2 phoneme inventories. At times, problems arise due to differences in L1 and L2 syllable structure. For example, Brazilian Portuguese (BP) learners of English have considerable difficulty acquiring the voiceless stops /p k/ (among other consonants) in medial coda and word-final position in words such as *ca[p]tain* or *bisho[p]* and *do[k]tor* or *magi[k]*. The problem is not with the phonemes per se, since /p k/ appear in BP, but rather with their context of occurrence: in BP, /p k/ are limited to initial and medial onset position (see [p]itu 'type of shrimp', [k]a[p]a 'cover', and va[k]a 'cow').

The set of phonemes appearing in medial coda and word-final position in BP is limited to /s l r/, which are subject to variable processes of lenition (including deletion), and to the placeless nasal /N/, which surfaces as nasalization of the preceding vowel. Oral stops do not occur in these positions in BP, so when BP learners encounter English medial coda and word-final /p k/, they employ a strategy of vowel epenthesis such that the stops are realized as onsets: hence *ca[p]tain* → *ca[pi]tain*, *bisho[p]* → *bisho[pi]*, *do[k]tor* → *do[ki]tor*, and *magi[k]* → *magi[ki]*. Since the same process applies to native Portuguese words (*captar* → *ca[pi]tar*; *pacto* → *pa[ki]to*) and to loanwords to BP (*laptop* → *la[pi]to[pi]*; *chic* → *chi[ki]*) (Cantoni & Cristófaró Silva, 2008; Cristófaró Silva & Almeida, 2008; Nevins, 2008), the interlanguage phenomenon is best considered a case of L1 transfer. As BP ESL learners become more proficient, they eventually suppress the process of i-epenthesis, variably at first, then categorically at an advanced level of proficiency.

Our general research question is: What exactly are they learning as they acquire the ability to produce /p k/ in words like *chapter*, *bishop*, *doctor*, and *magic*? In medial position (*chapter* and *doctor*), it is fairly uncontroversial to claim that they are learning to syllabify /p k/ as codas

and thus to expand the set of coda consonants to include these voiceless stops. When it comes to word-final position (*bishop* and *magic*), however, the learning challenge is not so clear, since there are two views on how word-final (henceforth simply ‘final’) consonants are syllabified: according to the orthodox view (Selkirk, 1982; Blevins, 1995), final consonants are all codas, but according to the view put forward in Government Phonology (Kaye, 1990; Harris & Gussmann, 1998), final consonants are onsets of empty nuclei. A more nuanced view has also been proposed under which, depending on the language, final consonants are either codas or onsets of empty nuclei (Piggott, 1999).²

We adopt the view that final consonants can be either codas or onsets, with the former syllabification applying in BP and the latter in English. Arguments in support of this position are presented in the next section, as part of the literature review. Importantly, if we are right in our assessment of the differing syllabifications of final consonants in English and BP, we would expect that the acquisition of English /p k/ in medial coda and final position by BP learners represents a dual challenge: on the one hand, learners need to expand the set of coda consonants to include /p k/; on the other, they need to acquire a novel prosodic representation of final consonants (including /p k/) as onsets of empty nuclei. Hypothetically, then, if a dual challenge is involved, learners should acquire medial coda and final /p k/ separately rather than in unison. More strongly, we might expect that medial /p k/ will be acquired before final /p k/, under the intuition that expansion of the set of coda consonants represents a lesser challenge than acquisition of a novel prosodic representation.³ Crucially, our hypothesis will be falsified if learners show simultaneous rather than differential acquisition of medial coda and final /p k/.

² We have chosen to overlook the view that final consonants may be extrasyllabic appendices, partly because, in many such analyses (e.g., Ito, 1988), the final consonant is extrasyllabic only initially, being incorporated into a coda at some later stage in the derivation. See further arguments against the appendix in John (2014).

³ Precisely why this should be is hard to say (which is why we refer to an intuition), and we are grateful to an anonymous reviewer for challenging us on this point. If the presence of specific coda consonants and of empty nuclei in a representation are both determined by markedness constraints, it remains to be seen why one constraint should be more difficult to re-rank than another. The frequency distribution of these forms in the L2 input could always play a role, but we are not under the impression (albeit without having actually calculated frequency) that medial codas are particularly more or less frequent than final /p k/ in English.

The next section provides a review of previous research on BP acquisition of medial coda and final consonants, along with a more extensive presentation of the position that final consonants are codas in BP but onsets of empty nuclei in English. Section 3 presents information about the participants, about the procedures followed for data collection, and about the analysis used to test our hypothesis of differential rather than simultaneous acquisition of medial coda and word-final /p k/. The results are presented in the section 4, followed by a discussion and conclusion. Beyond the immediate focus on the syllabification of final consonants, we concern ourselves in these latter sections with the question of variation in interlanguage output, which is a particularly thorny issue for any theoretical framework.

6.2 Background

For the purposes of our study, the most important point to make regarding previous research into i-epenthesis in BP interlanguage is that no studies to date have examined epenthesis following medial codas, only following word-final consonants. This oversight is easy to understand if medial codas and final consonants have the same prosodic representation (i.e., both are codas in the orthodox view), but this is not the only view on the syllabification of final consonants: a competing analysis treats final consonants as onsets of empty nuclei. If this latter view is correct, we would expect medial coda and final consonants such as /p k/ to be acquired separately. This is what our study set out to investigate.

An early study (Major, 1986) reports that the quality of the vowel epenthesized varies from a high front vowel to schwa, but since this variation is not reported on in other studies, it may be a product of the variety of BP spoken by the participants (all but 5 of the 53 participants were from the state of São Paulo). Other research has found higher rates of epenthesis after final voiced rather than voiceless stops: $bdg > ptk$ (where “>” indicates “higher than”) (Baptista & Silva Filho, 1997). This finding is not surprising given that voiced obstruents are universally more marked in final position, as witnessed by the common process of final obstruent devoicing (see Brockhaus, 1995, on German; Gussmann, 1992, on Polish; Mascaro, 1987, on Catalan; and Yavaş, 1994, for an overview of devoicing in interlanguage). In addition, final coronal stops tend to be acquired before non-coronal stops: $td > pbkg$ (Cardoso, 2007). This pattern also reflects a universal tendency, namely for coronal place to be unmarked (Paradis & Prunet, 1991). The finding is of further interest since /t d/ are precisely the stops that do not occur in medial codas in English

(Harris, 1994; Harris & Gussmann, 1998). Interestingly, final stops in polysyllabic words such as *attack* are acquired before those in monosyllabic words such as *pack* (Cardoso, 2007). The greater epenthesis in monosyllabic words suggests the influence of a Word Minimality constraint (McCarthy & Prince, 1993) requiring forms to be minimally disyllabic (or, from the perspective that *pack* ends in an empty nucleus, to have two filled nuclei). Since the effects of this constraint are not found in BP nor in English, the phenomenon is best considered a case of the Emergence of the Unmarked in interlanguage (McCarthy & Prince, 1994; Broselow, Chen & Wang, 1998).

Finally, in the only study not restricted to absolute final consonants, Huf and Alves (2010) found higher rates of epenthesis after strictly final stops than after penultimate stops in final clusters: *tack* → *tac[ki]* > *tact* → *ta[ki]t*. The authors interpret these findings as demonstrating that initial stops in final coda clusters are acquired before truly final codas. From a Government Phonology perspective, however, the final consonant cluster in a word such as *tact* constitutes in fact a coda-onset sequence. Under this view, the findings suggest we may be on the right track in anticipating for our own study that medial codas are acquired before final consonants.

As proposed earlier, the reason that more researchers have not emulated Huf and Alves (2010) in investigating the acquisition of consonants in positions other than absolute final probably stems from the assumption that medial coda and final consonants have the same syllabification and are presumably acquired together. This assumption reflects the orthodox view on the syllabification of final consonants, according to which these universally occupy a coda (e.g., Selkirk, 1982; Blevins, 1995) – see Fig. 6-1a, which illustrates this view with respect to the final /k/ in the word *magic*. A competing view expressed in Government Phonology, however, holds that final consonants always syllabify as onsets followed by an empty nucleus (e.g., Kaye, 1990; Harris & Gussmann, 1998), as illustrated in Fig. 6-1b. A further view (Piggott, 1999) contends that languages can vary as to whether they syllabify their final consonants as codas (6-1a) or as onsets followed by empty nuclei (6-1b).

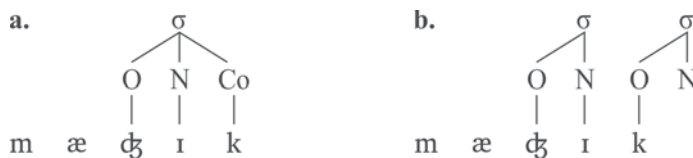


Fig. 6-1. Two views on the syllabic affiliation of final consonants

A number of arguments have been put forward in the Government Phonology literature to support the view that final consonants in English (and other languages) are onsets of empty nuclei (Harris & Gussmann, 1998). A general argument in favour of distinguishing between medial codas and final consonants is that the presence of one in a given language is no guarantee of the presence of the other: not all languages that instantiate medial codas also permit final consonants and vice versa. More specifically with respect to English, there is a mismatch between the set of consonants permitted in medial coda and in word-final position. Medial codas in English are restricted to sonorants and to select voiceless obstruents /p k f s m n ŋ l r/, whereas in final position virtually anything goes: with the exception of /h/ (as well as, arguably, the semi-vowels /j w/), any consonant in the English phoneme inventory may appear, namely /p b t d k g f v θ ð s z ʃ ʒ tʃ dʒ m n ŋ l r/.⁴ One possible reason a larger set of consonants is permitted in final rather than medial coda position is that final consonants are onsets rather than codas: the onset position licenses a greater range of contrasts than does the comparatively weak coda position. In addition, in the absence of a morpheme boundary, word-final consonant clusters in English are generally limited to pairs of consonants that occur word-internally in coda-onset sequences; for example, while *rt#*, *mp#* or *kt#* occur, the reverse sequences **tr#*, **pm#* or **tk#* do not. One explanation for this phonotactic restriction is, of course, that final clusters *are* coda-onset sequences in English, not coda clusters as an orthodox view would suggest.

A further argument in favour of viewing final consonants as onsets in English concerns the distribution of vowels in the preceding nucleus. Medial codas are generally preceded only by short vowels (with exceptions being limited to coronal codas as in *pastry*, *shoulder* or *counsel*). This restriction derives presumably from a binarity constraint which generally limits branching rhymes to maximally two timing slots (either a branching nucleus or a simplex vowel plus coda). Final consonants, however, can be preceded by any vowel in the inventory,

⁴ Some may object to our restrictive view of what appears in medial coda position, pointing to the voiced stops /b d g/ and the voiceless coronal /t/ in apparent exceptions such as *obnoxious*, *kidney*, *Ogden*, and *atlas*, but these have the profile of onsets of empty nuclei rather than codas. In most cases, the sequence of consonants involved in the putative exceptions either contravenes the falling or level sonority cline that typifies coda-onset sequences or else a morpheme boundary appears between the consonants, in which case the first consonant is morpheme-final rather than truly medial. In cases such as /bn/ in *obnoxious*, of course, both objections apply.

whether a short or long vowel or a diphthong. This ability of the final consonant to circumvent the binarity constraint provides further support for the view that final consonants are onsets rather than codas.

In BP, however, the arguments in favour of syllabifying final consonants as onsets do not apply. First, the pattern of a mismatch between medial coda and final position does not hold, since the same set of canonical coda consonants /s l r N/ appears in both locations. This identical distribution of consonants can be neatly accounted for if final consonants in BP are codas rather than onsets. Second, BP does not generally allow final clusters. Only singleton consonants occur word-finally, a pattern which is again consistent with a coda parse for final consonants in this language. Finally, the same set of short vowels /i e ε a u o ɔ/ appears before final and medial coda consonants in BP, although exceptionally diphthongs may also appear before /s/ or /N/ in both contexts. The inability of diphthongs to appear before final /l r/ is illustrated by disyllabic [xa.'uw] *Raul* (compare with ['paw.lu] *Paulo*) and trisyllabic [aw.da.'ix] *Aldair* (compare with ['kaj.ru] *Cairo*). For examples of diphthongs that are permitted exceptionally before /s/ or /N/, however, see [majs] *mais* 'more' and [pāw] *pão* 'bread'.

In sum, according to the various arguments presented above, final consonants are syllabified as onsets of empty nuclei in English, but as codas in BP. Under this analysis, as mentioned earlier, we anticipate that acquisition of medial coda and final /p k/ by BP learners of English represents a dual challenge. On the one hand, learners need to expand the set of consonants that appear in coda position to include the voiceless stops /p k/. On the other, they need to acquire the novel representation of final consonants as onsets of empty nuclei. Consequently, learners should exhibit differential acquisition of medial coda and final /p k/, with acquisition being demonstrated by suppression of the L1 transfer process of i-epenthesis. Finally, under the assumption that acquisition of novel syllabification is more challenging than expansion of the set of coda consonants, learners should acquire medial /p k/ before final /p k/. Our hypothesis of a dual challenge will, however, be falsified if learners show simultaneous acquisition of /p k/ in the two locations.

6.3 Method

This study employs a variationist approach to data collection and analysis (e.g., Labov, 2001; see also Cardoso, 2007, 2011, and John & Cardoso, 2008, for further examples of this approach as applied to L2 acquisition). In order to test our hypothesis, we collected data via three oral tasks from

two groups of 18 BP learners of English ($n=36$), one group in Montreal (Canada – an ESL context) and another group in Belém (Brazil – an EFL context).⁵ The data were analyzed using Goldvarb X (Sankoff, Tagliamonte & Smith, 2005), standard statistical software used in sociolinguistic/variationist studies to establish the relative contribution of a range of factors to the application of a variable process.

6.3.1 Participants

Of the 18 participants in Montreal, 8 showed no i-epenthesis in the tasks and were thus eliminated from the study. Although more extensive data collection might have uncovered occasional instances of i-epenthesis among these learners, for all intents and purposes these participants have passed the interlanguage stage where i-epenthesis occurs. Importantly, this indicates that acquisition of medial coda and word-final /p k/ is possible for BP ESL learners. Only the data of the remaining 10 participants showing some degree of i-epenthesis were retained for analysis. Among these participants, there were 8 females and 2 males, 9 ranging in age from 18 to 35 alongside one 64-year-old (mean: 29.1). In terms of self-declared proficiency level, 3 identified themselves as beginners, 6 as intermediate, and 1 as advanced (the 64-year-old outlier). The majority had arrived in Montreal in the 1–4 years preceding data collection, with one arriving 10 years before and another 17. The percentage of declared daily use of English ranged from 20–90 %.

Of the 18 participants in Belém, 2 showed no i-epenthesis. Advancement beyond the interlanguage stage of i-epenthesis is thus also attested in an EFL context. None had spent time in an English-speaking

⁵ Classifying Montreal as an English second language context is not straightforward but nonetheless justified in our view. Montreal is the largest city in Quebec, a province of Canada where French is the sole official language and where Francophones form a considerable majority in most regions. The situation is complicated in Montreal by the presence of a sizeable English-speaking and immigrant population. Montreal consequently is anomalous in Quebec in functioning to some extent as bilingual French-English, although the degree of French or English language use can vary considerably depending on the person, context and even the area of the city. At either extreme, it is in fact possible to function entirely in English or French. This point is underscored by the fact that, to the question about knowledge of official languages in the 2011 Canada census, 7.4% and 37% of the population of Montreal reported knowledge of only English or French respectively. Importantly from our perspective, there is thus the potential for BP speakers to have contact on a daily basis with English, as their reported daily percentage of use of English demonstrates.

country, so any acquisition of medial coda and final /p k/ occurred entirely in the EFL context. Unfortunately, the recordings for 2 participants were inaudible due to temporary technical problems with the microphone. Of the remaining 14 participants showing some degree of i-epenthesis, there were 6 females and 8 males, 11 ranging in age from 12 to 17 alongside one 28-year-old and two 46-year-olds (mean: 20.07). In terms of proficiency level, 8 identified themselves as beginners, 5 as intermediate and 1 as advanced. 12 participants declared that they had studied English for from 1 to 5 years, with one participant declaring only 2 months of study (and one 46-year-old participant failed to declare her length of study).

From the details provided, it is apparent that our two groups of ESL/EFL participants are not comparable, particularly in age and degree of contact, so our aim is not to carry out a comparative study of acquisition in the two contexts. Rather, in a spirit of thoroughness, the aim is to see whether the same patterns of acquisition are found in heterogeneous groups of learners. Similar results involving different groups of learners performing the same tasks will reinforce the impression that the findings truly are representative of the learning pattern.

6.3.2 Data collection

The participants were recorded performing three oral tasks involving: 1) real-word elicitation; 2) non-word repetition; and 3) non-word reading aloud. The real-word elicitation task was preceded by a training session in which the participants listened to a recording of a native speaker of English identify a series of 25 pictures shown on a series of PowerPoint slides (e.g., *Doctor. This man is a doctor.*). The purpose of the training session was to remind participants of certain highly frequent words in English with /p/ or /k/ in medial coda or final position (e.g., *doctor, captain, music, asleep*) and to set up an association between the image and the word. Six of the words used were distractors (e.g., *computer, window, chicken*). For the actual elicitation task, the participants saw the same set of images in a different order and had to produce the target word in a carrier sentence appearing on the screen. To further aid participants in recovering the target word, some of the letters were provided, though not those corresponding to /p k/ (e.g., *This man is a d__t__* to accompany the picture of a doctor). For the non-word repetition task, participants heard a non-word twice (e.g., *toctel, toctel*) and then inserted it at the end of a carrier phrase (e.g., *I can't find my _____*). For the final task, participants read aloud the same set of non-words as they appeared on the screen on successive slides (e.g., *TOctel*). Participants were asked to stress the

upper-case syllables in the non-words, and equal numbers of medial and final /p k/ occurred with preceding stressed and unstressed vowels. Both the real and non-words were presented in three randomized sets. So as to circumvent the potential confounding influence of the Word Minimality constraint, none of the target real and non-words were monosyllabic (i.e., in the sense of having a single filled nucleus). Apropos, the motivation for deciding to use non-words came from the relative difficulty of finding high-frequency non-monosyllabic words that end in /p k/.

The participants were recorded while performing three oral tasks by one of the researchers in a lab at an Anglophone university in Montreal, Canada, and by a research assistant at a public school in Belém, the capital city of the state of Pará in the north of Brazil. For the recording of participants' output, we used an Audio-Technica AT831b lavalier microphone attached to a Zoom H4 Digital Recorder. For the tasks involving oral stimuli (the task 1 training session and task 2), participants listened to the recordings over a Microsoft LX-400 headset connected to a computer.

6.3.3 Data analysis

The data were coded by one of the researchers (a native speaker of English) and checked by the other (a native speaker of Brazilian Portuguese) for a number of potential independent variables and the single dependent variable (presence vs absence of i-epenthesis). The crucial independent variable for the purposes of our hypothesis was whether /p k/ appeared in medial coda or word-final position. Further independent variables were included primarily for exploratory purposes (as is customary in variationist studies), since no hypotheses were associated with them. These were: proficiency level (beginner/intermediate/advanced); task (elicitation/repetition/reading aloud); word status (real/non-word); stress status of the preceding vowel (stressed/unstressed); and participants (1, 2, 3, and so on). If participants' self-evaluation of their proficiency level was accurate, we would expect a correlation between lower rates of i-epenthesis and higher proficiency, but otherwise no particular correlations were anticipated.

Once coded, the data were analyzed using Goldvarb X (Sankoff et al., 2005), which performs a regression analysis in order to establish the contribution of the various factors to the application of i-epenthesis. The program assigns to each factor a weight from 0 to 1 and determines whether factors contribute significantly, whether positively or negatively, to the process under study. Factor groups that are not significant or that

appear to interact with other factor groups can be eliminated in successive runs in order to refine the analysis.

6.4 Results

The data from Montreal and Belém were analyzed separately. The Montreal data were analyzed in three runs with the proficiency level and word status factor groups progressively eliminated from the analysis. The remaining factor groups were selected by Goldvarb X in the third run for both the step-up and step-down regression analyses, thus indicating that these groups contribute significantly to the variable phenomenon under investigation ($p < .05$). We present here the results from the third run, as this provided the most insightful analysis (for a more thorough discussion of Goldvarb X in the analysis of L2 data, see Cardoso, 2007).

<i>Factor groups</i>	<i>Factor weights/ %</i>			
Position	final: 0.899/ 12.4	medial: 0.103/ 0.3		
Task	elicitation: 0.803/ 9.6	repetition: 0.074/ 0.4	reading: 0.807/ 9.7	
Stress status	stressed: 0.364/ 4.1	unstressed: 0.684/ 9.2		
Participants	1: 0.213/ 1.5	2: 0.213/ 1.5	3: 0.232/ 1.5	4: 0.363/ 3.0
	5: 0.370/ 3.0	6: 0.485/ 4.4	7: 0.562/ 5.9	8: 0.686/ 8.8
	9: 0.804/ 11.9	10: 0.942/ 23.1		

Table 6-1. Montreal data – Factor weights and % assigned by Goldvarb X.

Overall, rates of epenthesis in the Montreal data were quite low (input probability: 0.004). All of the participants were more likely to produce medial coda and final /p k/ without [i] rather than with this epenthetic vowel. As illustrated in Table 6-1 above, the final analysis of the Montreal data (664 tokens) revealed rates of i-epenthesis that were significantly higher following /p k/ in final (0.899) rather than medial coda (0.103) position. Put differently, production of the target /p k/ without an epenthetic vowel was significantly more difficult in word-final than in medial coda position. Unexpectedly, the rates of i-epenthesis in the repetition task (0.074) were significantly lower than in the elicitation and reading-aloud tasks (0.803 and 0.807 respectively). Indeed, instances of i-

epenthesis in the repetition task were vanishingly rare (one token only), thus highlighting the importance of triangulation of data collection. Interestingly, i-epenthesis was also higher when the preceding nucleus was unstressed (0.684) (*'magic*) rather than stressed (0.364) (*at'tack*). Finally, four of the participants showed significantly greater i-epenthesis than the others, indicating varying degrees of proficiency in /p k/ production across the participants. As mentioned earlier, the other (eliminated) factor groups (word status and proficiency level) did not significantly influence rates of i-epenthesis.

The Belém data were likewise analyzed in successive runs during which the factor groups word status, proficiency level and stress status were progressively eliminated – unlike in the Montreal data, stress did not contribute to the likelihood of i-epenthesis. Two participants who had only single instances of i-epenthesis in their data (singletons) were also eliminated, since the statistical analysis requires the presence of systematic variation. The remaining factor groups (position within the word, task and participants) were all significant ($p < .05$; as indicated by the selection of these factor groups in Goldvarb's step-up and step-down regressive analyses).

The rates of epenthesis in the Belém data for the 12 participants remaining in the final run were low (input probability: 0.038), though higher than in the Montreal data (0.004). As shown in Table 6–2 below, the final analysis of the Belém data (798 tokens) revealed rates of i-epenthesis that were significantly higher following /p k/ in final (0.786) rather than medial coda (0.220) position. Once again, production of the target /p k/ without an epenthetic vowel was significantly more difficult in word-final than in medial coda position. Also, the rates of i-epenthesis in the repetition task (0.257) were significantly lower than in the elicitation and reading-aloud tasks (0.742 and 0.548). Finally, four of the participants showed significantly greater i-epenthesis than the others, indicating varying degrees of proficiency in /p k/ production across the participants. The other factor groups (word status, proficiency level and stress status) did not significantly influence rates of i-epenthesis.

To sum up, the Belém participants had overall higher rates of i-epenthesis than the Montreal participants: a total of 83 instances in 798 tokens or 10.4% (with 72 in final and 11 in medial position or 18.3% and 2.7% respectively) vs. 42 instances in 664 tokens or 6.3% (with 41 in final and 1 in medial position or 12.4% and 0.3% respectively). More importantly, these results show that regardless of learning context, learners acquire medial /p k/ before final /p k/, thus suggesting that medial and

word-final consonants may constitute different prosodic representations in the interlanguage of these learners, as will be discussed next.

<i>Factor groups</i>	<i>Factor weights/ %</i>			
Position	final: 0.786/ 18.3	medial: 0.220/ 2.7		
Task	elicitation: 0.742/ 17.8	repetition: 0.257/ 4.3	reading: 0.548/ 10.5	
Participants	1: 0.230/ 3	2: 0.230/ 3	3: 0.330/ 4.5	4: 0.391/ 5.9
	5: 0.391/ 5.9	6: 0.458/ 7.5	7: 0.487/ 6.8	8: 0.528/ 9
	9: 0.537/ 9.1	10: 0.606/ 11.9	11: 0.676/ 14.7	12: 0.943/ 43.3

Table 6-2. Belém data – Factor weights and % assigned by Goldvarb X.

6.5 Discussion

In this study investigating the acquisition of medial coda and final stops by BP ESL learners in Montreal and Belém, we found medial coda /p k/ to be acquired before final /p k/, with acquisition being determined by absence of the transfer process of i-epenthesis. These findings provide invaluable insight into an issue in phonological theory concerning the syllabification of final consonants: the prevalent view is to syllabify final consonants as codas; whereas an opposing view contends that final consonants are universally onsets of empty nuclei. Our analysis of final consonants in BP and English suggested that the former syllabification applies in BP and the latter in English, in line with what is proposed in Piggott (1999). That is, final consonants have the same syllabic affiliation as medial codas in BP, but a different syllabic affiliation in English. If this analysis is accurate, BP ESL/EFL learners are confronted with a considerable learning challenge: not only must they expand the set of medial codas to include such consonants as the voiceless stops /p k/, but they must also learn to syllabify final consonants as onsets of empty nuclei. Given that empty nuclei are not instantiated in BP (i.e., there are no configurations which require positing the presence of an empty nucleus), the latter part of the challenge thus requires development of a novel prosodic representation, presumably no easy feat.

Based on this hypothesis that acquisition of medial coda and final /p k/ constitutes a dual challenge for BP ESL/EFL learners, we predicted that

learners should exhibit differential (or sequential) acquisition of the stops in the two locations. Our hypothesis would have been falsified if we had observed simultaneous acquisition, but this is not at all what we found: our participants had significantly higher rates of i-epenthesis after final rather than medial /p k/. In a nutshell, we were unable to falsify our hypothesis. Thus, to answer the question posed in the title to our chapter (“Are word-final consonants codas?”), while they may very well be codas in BP, our findings are not consistent with such an analysis for English. The implications for phonological theory are that the orthodox view of final consonants being universally syllabified as codas is untenable.

Note, however, that it would be erroneous to claim that our findings actually confirm a distinction in the syllabification of final consonants in the two languages. Our results are merely consistent with such an analysis. It would also be possible for differential acquisition to occur if final consonants are onsets in both languages. In this case, acquisition would involve the dual challenge of expansion of the sets of both medial coda and final consonants. What is much harder to countenance is that differential acquisition should occur with final consonants being codas in both languages. In this case, the dual challenge would be reduced to expansion of the sets of medial and final coda consonants. A different syllabic parse for medial coda and final consonants thus seems unavoidable in English. In brief, under a circumscribed view of scientific practice (Popper, 1959), we have been unable to falsify our hypothesis, but resist making the claim of having positively confirmed it.

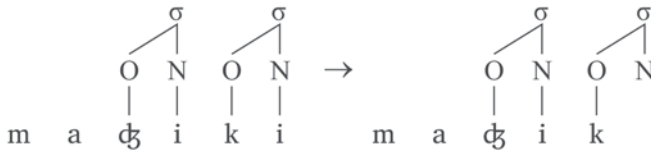
The preceding observations constitute the core implications of our findings, but there are other, more peripheral, issues which should also be addressed. In particular, two aspects of the findings require further consideration. First, the occurrence of significantly lower i-epenthesis in the non-word repetition task was unexpected. There is no obvious explanation for this task effect, but it serves as a valuable reminder to researchers of the vital importance of triangulating data collection in order to get an accurate portrait of the phenomenon under study. Second, it was not unexpected that i-epenthesis after medial coda and final /p k/ should be variable, since variability is a common characteristic of interlanguage. Nonetheless, variability requires some explanation. Although variation is a common feature of L1s, particularly in the domain of post-lexical processes (Kiparsky, 1982, 1985), it is even more frequent in L2 systems. Why should this be the case? At the risk of diverging from our principal concern with the syllabification of final consonants, we formulate here a tentative proposal.

To our knowledge, it is rare for variation to be attributed to the realm of phonetic implementation (i.e., to inconsistent articulation of a target form). Instead, variation is usually attributed to non-categorical phonological processes operating on lexical representations such that the surface form generated by the grammar varies. For example, variation can be generated via optional or variable rules (Labov, 1969; Cedergren & Sankoff, 1974) or via multiple grammars (Kroch, 1989). Variation can also be captured in Optimality Theory (OT – Prince & Smolensky, 1993) via partially ordered constraints (Kiparsky, 1993; Reynolds, 1994; Anttila, 1997), via stochastic OT (Boersma, 1997; Boersma & Hayes, 2001), or via lexically indexed constraints (Coetzee, 2009). While variable i-epenthesis could be due to some form of non-categorical process that applies to underlying /p k/ in medial coda and final position, we think an entirely different approach to variation is needed.

The first step towards understanding variation in i-epenthesis involves considering the notion of i-epenthesis itself, since the term is really a misnomer. The term, ubiquitous in the literature on BP, implies a process of synchronic insertion of a vowel such that, while the vowel is present in the surface form, it is missing from the lexical entry. This scenario, however, is inaccurate, since so-called epenthetic [i] is in fact part of L1 lexical entries in BP (see arguments in Cantoni & Cristófaró Silva, 2008, e.g., the fact that epenthetic [i] can bear stress, and in Cristófaró Silva & Almeida, 2008). Hence, surface forms such as [kapiti] *capto* ‘(I) capture, attract’ and [varigi] *Varig* (Brazilian airline) are underlying /kapitu/ and /varigi/, not /kaptu/ and /varig/. The adaptation of illicit codas and final consonants in BP occurs in the lexicon itself, not in the generation of surface forms. By extension, the same surely holds for epenthetic [i] in BP English: underlyingly, *magic* and *doctor* are /maɟʒiki/ and /dɔkitor/. Put differently, there is no synchronic process of i-epenthesis that BP speakers need to suppress in order to generate accurate output in English. Instead, what is required is revision of underlying forms such that *magic* and *doctor* are /maɟʒik/ and /dɔkitor/, not /maɟʒiki/ and /dɔkitor/.

We assume that syllable structure is recorded in the lexicon (Kaye & Lowenstamm, 1984), so as BP speakers progress in English, becoming capable of parsing medial /p k/ in coda position and final /p k/ in onset of an empty nucleus, they rewrite underlying forms such as *magic* and *doctor* as illustrated in Fig. 6-2a and b. The initial forms in Fig. 6-2a and b generate inaccurate output, with so-called i-epenthesis, whereas the revised forms generate target-like output.

a. /maɕʒiki/ → /maɕʒik/



b. /dɔkitɔr/ → /dɔktɔr/



Fig. 6-2. Revision of lexical entries in BP interlanguage.

We suggest that variation in i-epenthesis in BP-based interlanguage is better characterized as involving competing underlying representations (an approach which Anttila, 2002, following Müller, 1999, refers to as ‘pseudo-optionality’). The idea is that, when speakers rewrite lexical entries as shown in Figure 6–2, they do not actually overwrite or otherwise eradicate the original underlying representation; the replaced underlying representations continue to be present alongside the novel forms. Hence, at the moment of selection, the speaker can potentially access either form as a base for surface output. Initially, the ingrained habit is to access the form with an added vowel, but gradually, as speakers improve, they get more proficient at accessing the accurate form. At first, selection of the new underlying representation will be quite arduous, hence relatively infrequent and associated more with formal or careful speech, which favours attending to desired output.

To recap, our proposal is that competition between underlying representations is what is behind the variability of i-epenthesis in BP English. This is not to say, however, that competing underlying representations are responsible for all forms of L1 or L2 variation. Variation can doubtless have more than one source, including the computational system. Nonetheless, variation should not always be attributed to the computational realm; in some cases, the source of variation is in the lexicon itself.

The notion of competing underlying representations is virtually absent from the variationist literature. All the same, the analysis appears

promising for i-epenthesis in BP English, and it can probably account for other instances of variation too. For example, the speech of one of the authors of this chapter alternates between yod-ful and yod-less pronunciation of coronal oral/nasal stop + high back vowel sequences such as in the following words: *tube* [tjub]/[tub], *due* [dju]/[du], and *new* [nju]/[nu]. To our knowledge, this variation is not typical: the yod-ful forms are standard in British English, and the yod-less forms characterize standard North American English. We would contend that this variation in output is not due to a process of variable deletion or epenthesis (hence, not computational), but to variable accessing of yod-ful and yod-less underlying representations. The presence of the two underlying representations in this particular speaker's lexicon is due to extensive exposure to both British and North American English from an early age. At the moment of speaking, he thus has to choose between two potential underlying forms, subject to various contextual factors.

In sum, the competing representations approach may prove useful in accounting for variable phenomena beyond just i-epenthesis in BP English. Indeed, although we have yet to work out all the details, at first blush, competing representations could potentially account for phenomena such as gradual diffusion (in an L2, see Gatbonton, 1978) and for certain lexical frequency effects (Bybee, 2001). It is too early, however, to fully assess the range of explanatory coverage. The question is thus consigned to the category of future research.

6.6 Conclusion

Much research on the acquisition of an L2 sound system focuses on individual consonants and vowels, but this is not the only area in which languages can differ phonologically. At times, differences are found at the level of syllable structure. For example, BP ESL/EFL learners have difficulty with the stops /p k/ in medial coda and word-final position, not because these segments are absent from the L1 phoneme inventory, but because they are not found in these particular prosodic contexts in the L1. In order to acquire medial coda /p k/, BP learners need to expand the set of possible coda consonants. For the acquisition of final /p k/, however, the nature of the learning challenge depends on whether final consonants are syllabified as codas or as onsets of empty nuclei. If final /p k/ are codas, they should be acquired in tandem with their medial counterparts. On the other hand, if final /p k/ are onsets (as we hypothesized), they should be acquired separately from medial coda /p k/. This study thus set out to investigate whether English medial coda and final consonants have

identical syllabic affiliations. Our hypothesis would be falsified if BP ESL/EFL learners exhibited simultaneous acquisition of /p k/ in the two locations. This was not what we found: i-epenthesis rates were significantly lower following medial rather than final /p k/, findings which are consistent with our hypothesis.

Vowel epenthesis is nonetheless variable in both contexts, a state of affairs which is hard to capture. Rather than attribute variation to the computational system (e.g., to variable rules or to crucially unranked constraints), we propose that variable output could stem from competition between two phonological representations associated with single lexical entries: a representation where the supposedly epenthetic vowel is included (e.g., do/.ki/tor, magi/.ki/) and one where it is absent (e.g., do/k./tor, magi/.kØ/ – where Ø indicates an empty nucleus). A first step in acquisition would thus involve developing the novel phonological representations. A second step would involve learning to consistently access the novel form at the moment of speaking.

On a final note, our research has implications for second language teaching. First, teachers need to be aware of the developmental sequence in the acquisition of medial coda and final consonants by BP (and potentially other) learners. Since there are different schools of thought, we cannot make definite recommendations as to whether teachers should sequence their teaching to focus first on the easier context (medial codas) and then on the difficult context (final consonants) or whether they should focus all their effort on the difficult context under the assumption that easily acquired items will take care of themselves. This is an empirical issue that would require testing in further research (but see Cardoso & Collins, 2015, for the implementation of this hypothesis in the teaching of foreign codas and onset clusters). Nonetheless, confirmation of the presence of a dual challenge for BP learners in the acquisition of medial coda and final consonants should sensitize teachers and ultimately help them make informed choices.

References

- Anttila, A. (1997). Deriving variation from grammar: a study of Finnish genitives. In: Hinskens, F., Van Hout, R. & Wetzels, L. (Eds.), *Variation, Change and Phonological Theory*, pp. 35–68. Amsterdam: John Benjamins.
- (2002). Variation and phonological theory. In: Chambers, J., Trudgill, P. & Schilling-Estes, N. (Eds.), *Handbook of Language Variation and Change*, pp. 206–243. Oxford: Blackwell.

- Baptista, B. O. & Silva Filho, J. (1997). The influence of markedness and syllable contact on the production of English consonants by EFL learners. In: Leather, J. & James, A. (Eds.), *New Sounds 1997: Proceedings of the Third International Symposium on the Acquisition of Second Language Speech*, pp. 26–34. Klagenfurt: University of Klagenfurt.
- Blevins, J. (1995). The syllable in phonological theory. In: Goldsmith, J. A. (Ed.), *The Handbook of Phonological Theory*, pp. 206–244. Oxford: Blackwell.
- Block, D. (2007). *Second Language Identities*. London/New York: Continuum.
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Boersma, P. (1997). How we learn variation, optionality, and probability. In *L Proceedings of the Institute of Phonetic Sciences*, 21, pp. 43–58. Amsterdam: University of Amsterdam.
- Boersma, P. & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32/1, pp. 45–86.
- Brannen, K. J. (2011). *The Perception and Production of Interdental Fricatives in Second Language Acquisition*. Unpublished doctoral dissertation, McGill University.
- Brockhaus, W. (1995). *Final Devoicing in the Phonology of German*. Tübingen: Niemeyer.
- Broselow, E., Chen, S.-I. & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, 20, pp. 261–280.
- Cantoni, M. & Cristófaró Silva, T. (2008). Verbal stress assignment in Brazilian Portuguese and the prosodic interpretation of segmental sequences. In *Proceedings of the Speech Prosody Conference*, pp. 587–590.
- Cardoso, W. (2007). The variable development of English word-final stops by Brazilian Portuguese speakers: a stochastic optimality theoretic account. *Language Variation and Change*, 19, pp. 219–248.
- . (2011). The development of coda perception in second language phonology: A variationist perspective. *Second Language Research*, 27/4, pp. 433–465.
- Cardoso, W. & Collins, L. (2015). Developmental sequences in second language phonology: effects of exposure. Paper presented at the American Association for Applied Linguistics [ms. in preparation]. Toronto, Canada.

- Cedergren, H.J. & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, 50/2, pp. 333–355.
- Cristófaró Silva, T. & Almeida, L. (2008). On the nature of epenthetic vowels. In: Bisol, L. & Brescancini, C. R. (Eds.), *Contemporary Phonology in Brazil*, pp. 193–212. Cambridge: Cambridge Scholars Publishing.
- Coetzee, A. (2009). Phonological variation and lexical frequency. In: Schardl, A., Walkow, M. & Abdurrahman, M. (Eds.), *NELS 38: Proceedings of the North East Linguistic Society (Volume 1)*, pp. 189–202. Amherst: Graduate Linguistic Student Association (GLSA) of the University of Massachusetts.
- Gardner, R. C. & Lambert, W. E. (1972). *Attitudes and Motivation in Second Language Learning*. Rowley, MA: Newbury House.
- Gatbonton, E. (1978). Patterned phonetic variability in second language research: A gradual diffusion model. *Canadian Modern Language Review*, 34, pp. 335–347.
- . (1992). Resyllabification and delinking: The case of Polish voicing. *Linguistic Inquiry*, 23, pp. 29–56.
- Harris, J. (1994). *English Sound Structure*. Oxford: Blackwell.
- Harris, J. & Gussmann, E. (1998). Final codas: Why the west was wrong. In Cyran, E. (Ed.), *Structure and Interpretation in Phonology: Studies in Phonology*, pp. 139–162. Lublin: Folia.
- Huf, J. C. C. & Alves, U. K. (2010). A produção de /p/ et /k/ em codas simples e complexas do inglês (L2) por aprendizes gaúchos. *Verba Volant*, 1/1, pp. 1–27.
- Ito, J. (1988) *Syllable Theory in Prosodic Phonology*. New York: Garland Publishing.
- John, P. (2014). *The Phonology of Empty Categories*. Unpublished doctoral dissertation, Université du Québec à Montréal.
- John, P. & Cardoso, W. (2008). Francophone ESL learners and h-epenthesis. *Concordia Papers in Applied Linguistics (COPAL)*, 1, pp. 76–97.
- Kaye, J. (1990). ‘Coda’ licensing. *Phonology*, 7, pp. 301–330.
- Kaye, J. & Lowenstamm, J. (1984). De la syllabicité. In: Dell, F., Hirst, D. & Vergnaud, J.-R. (Eds.), *Forme sonore du langage*, pp. 123–159. Paris: Hermann.
- Kiparsky, P. (1982). From Cyclic to Lexical Phonology. In: van der Hulst, H. & Smith, N. (Eds.), *The Structure of Phonological Representations, vol. I*, pp. 131–175. Dordrecht: Foris.
- . (1985). Some consequences of Lexical Phonology. *Phonology Yearbook*, 2, pp. 83–138.

- . (1993). Variable rules. *Paper presented at the Rutgers Optimality Workshop*. New Brunswick, NJ.
- Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1, pp. 199–244.
- Labov, W. (1969). The logic of nonstandard English. *Georgetown Monographs on Language and Linguistics*, 22, pp. 1–31.
- . (2001). *Principles of Linguistic Change: Social Factors*. Malden, MA: Blackwell.
- Lenneberg, E.H. (1967). *Biological Foundations of Language*. New York: John Wiley and Sons.
- Major, R. (1986). Paragoge and degree of foreign accent in Brazilian English. *Second Language Research*, 2, pp. 53–71.
- . (2008). Transfer in second language phonology: A review. In: Edwards, J. G. H. & Zampini, M. L. (Eds.), *Phonology and Second Language Acquisition*, pp. 63–94. Amsterdam: John Benjamins.
- Mascaró, J. (1987). Underlying voicing recoverability of final devoiced obstruents in Catalan. *Journal of Phonetics*, 15, pp. 183–186.
- Mah, J. (2011). *Segmental Representations in Interlanguage Grammars: The Case of Francophones and English /h/*. Unpublished doctoral dissertation, McGill University.
- McCarthy, J. J. & Prince, A. (1993). *Prosodic Morphology I: Constraint Interaction and Satisfaction*. Unpublished manuscript, University of Massachusetts, Amherst, MA and Rutgers University, New Brunswick, NJ.
- McCarthy, J. J. & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. In: *Papers from the Annual Meeting of the North East Linguistic Society*, 24, pp. 333–379.
- Müller, G. (1999). Optionality in optimality-theoretic syntax. *GLOT International*, 4/51, pp. 3–8.
- Nevins, A. (2008). Review of *A Lateral Theory of Phonology: What is CVCV and Why Should it be?* by Tobias Scheer. *Lingua*, 118/3, pp. 425–434.
- Paradis, C. & Prunet, J-F. (Eds) (1991). *Phonetics and Phonology: The Special Status of Coronals*. San Diego: Academic Press.
- Piggott, G. (1999). At the right edge of words. *The Linguistic Review*, 16/2, pp. 143–185.
- Prince, A. & Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms., Rutgers University, New Brunswick, and University of Colorado, Boulder.
- Popper, K. (1959/1992). *The Logic of Scientific Discovery*. London: Routledge.

- Reynolds, W. (1994). *Variation and Phonological Theory*. PhD dissertation, University of Pennsylvania.
- Sankoff, D., Tagliamonte, S. & Smith, E. (2005). GoldVarb X: A variable rule application for Macintosh and Windows. Retrieved on October 15, 2015, from: <http://individual.utoronto.ca/tagliamonte/goldvarb.html>.
- Selkirk, E. (1982). The syllable. In: van der Hulst, H. & Smith, N. (Eds.), *The Structure of Phonological Representations (Part 2)*, pp. 337–383. Dordrecht: Foris.
- Trofimovich, P. & John, P. (2011). When *three equals tree*: Examining the nature of phonological entries in L2 lexicons of Quebec speakers of English. In Trofimovich, P. & McDonough, K. (Eds.), *Applying Priming Methods to L2 Learning, Teaching and Research*, pp. 105–129. Amsterdam: John Benjamins.
- Yavaş, M. (1994). Final stop devoicing in interlanguage. In: Yavaş, M. (Ed.), *First and Second Language Phonology*, pp. 267–282. San Diego: Singular.

CHAPTER SEVEN

PASSING FOR A NATIVE SPEAKER: PRODUCTION AND PERCEPTION

KSENIA GNEVSHEVA¹

7.1 Introduction

Linguistics has seen a heated discussion of whether ultimate attainment in second language speakers is possible (e.g., Abrahamsson & Hyltenstam, 2009; Bongaerts, 2000; Ioup et al., 1994). Coppieters (1987:565) argued that a native speaker ‘is someone who is accepted as such by the community referred to as that of [native] speakers’, in other words, who passes for a native speaker. A non-native speaker (NNS) of a language is said to ‘pass’ for a native speaker (NS) when he/she is believed to be a NS by native-speaking listeners. Second language (L2) speakers’ self-reports are often analyzed in an attempt to assess L2 speakers’ ability to pass for a native speaker and the conditions that may be conducive to a successful case of passing (e.g., Abrahamsson & Hyltenstam, 2009; Marx, 2002; Piller, 2002); however, using speaker self-reports usually means relying only on intermittent and unsystematic observations.

Obtaining nativeness judgments from raters in experimental conditions can add an extra element of control and allows us to explore the listeners’ perspectives (e.g., Abrahamsson & Hyltenstam, 2009; Bongaerts, 2000; Gnevsheva, 2015c; Ioup et al., 1994). In such perception experiments native-speaking listeners are usually asked to decide which of the clips were produced by a first language (L1) speaker of that language or to rate the clips on an accentedness scale. However, the focus on perception

¹ I am thankful to the 4th International Conference on English Pronunciation for a stimulating discussion and Kevin Watson, Jen Hay, and two anonymous reviewers for providing invaluable comments on earlier drafts of this chapter. All the remaining deficiencies are my sole responsibility. This research was partially supported by a University of Canterbury small research grant.

backgrounds the speaker and introduces a number of listener-dependent variables, such as familiarity with foreign accents (e.g., Kraut & Wulff, 2013).

To help us understand the relationship between objective measures of production and subjective measures of perception in passing for a native speaker, the data from the same speakers are analyzed from both angles. Variation in perceived accentedness has been found to be predicted by variation in native-likeness in production of a wide range of segmental and supra-segmental features: voice onset time (VOT), vowel quality, F0, vowel duration, rate of speech, word stress, intonation, etc. (Anderson-Hsieh, Johnson, & Koehler, 1992; de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; McCullough, 2013; Munro, 1993; Munro & Derwing, 1995).

However, Munro and Derwing (2015) note that acoustic measures do not always correlate with perception ratings, and sometimes a second language speaker may be believed to be a native speaker by listeners despite some non-target-like production when assessed objectively, a phenomenon Abrahamsson and Hyltenstam (2008) called 'non-perceivable non-nativeness'. Such cases of non-perceivable non-nativeness suggest that the second language speaker's production of all segmental and suprasegmental features need not be nativelike for passing to occur. In fact, the findings of a perceptual dialectology study by Leach, Watson and Gnevshva (2016) suggest that listeners' ability to correctly identify L1 varieties may depend on the presence or absence of (salient) features in the stimulus. It may be possible that the presence or absence of (salient) features in the stimulus also affects a second language speaker's passing for a native speaker.

Studies of ultimate attainment often explore speakers' native-likeness in both production and perception but rarely do they correlate variation in native-likeness in these two domains. Abrahamsson and Hyltenstam (2009), for example, considered second language speakers' native-likeness from three different perspectives: self-reporting passing for a native speaker, passing for a native speaker in a perception experiment, and native-like performance in pronunciation, grammar, and other domains. They concluded that very few second language speakers performed in the native speaker range on production measures, despite self-reporting experiences and being regarded as a native speaker by many listeners in the perception experiment. The focus of the paper was on the effect of age of acquisition on ultimate attainment, and the measures of speakers' production were not directly compared to the results of perceived native-likeness. So exactly what linguistic features listeners responded to in the

signal when they made their judgments of second language speakers remains unclear, as does the degree to which native-likeness in production of different elements contributed to the overall perceived native-likeness.

Schmid, Gilbers and Nota (2014) collected L2 speakers' VOTs, vowel formant measurements for /ɜ/ and /æ/, and their ratings on a foreign accentedness scale. Only one L2 speaker out of 20 fell within the range of native speakers in perceived accentedness, but this speaker received a perfect score, meaning that he was perceived to have no foreign accent by any of the judges, and, therefore, passed for a native speaker. The authors scrutinized the speakers as a group and individually, and concluded that the two L2 speakers who scored within the NS range on production measures were not the ones judged to be most native-like in perception. On the other hand, the L2 speaker who passed for a NS produced some VOTs and /æ/ which were outside the native range. This may suggest that a certain degree of non-nativeness or non-target production of some (possibly less salient) elements may still be below the non-nativeness threshold in perception.

One way to explore the salience of individual features and their effect on passing or perceived accentedness is through acoustic manipulations of second language speech. Magen (1998), for example, found that in their ratings of foreign accent in non-native English speakers of L1 Spanish, listeners were sensitive to such features as vowel quality, consonant manner, and stress, but not to voicing.

An alternative is to collect listener comments and focus on what non-target-like elements they notice in L2 speech. Previous studies have found that listeners often comment on phonological features in general (Derwing & Munro, 1997; Moyer, 2004), but individual segments or suprasegmentals are rarely mentioned. Qualitative studies sometimes discuss individual segments that listeners comment on and which are, therefore, noticeable or salient. For example, in McKenzie (2015), listeners explicitly commented on many non-target-like pronunciations of consonants: /v/, /l/, /r/, /t/, /d/, /ð/, etc., but non-native-likeness in vowels was only mentioned in general.

Hayes-Harb and Hacking (2015: 54) also noted that, while the listeners commented a lot on consonants and even mentioned specific segments ('th' sounds like "d", sometimes like "t"), their comments on vowels were more general ('sounded different', 'foreign', and 'not English'). To address individual vowels, raters in Hayes-Harb and Hacking (2015: 55) often used imitation by providing examples from the NNESS' speech (e.g., 'call sounded more like "c[o]ll"'). Such imitation of speaker features may be reflective of the non-linguist listeners' lack of terms for description but, at the same time, a certain degree of awareness of noticeable differences

(Preston, 1996), which suggests that listener imitation of vowels can be used for analyzing salient features.

The goal of this chapter is to address feature salience in passing for a native speaker. It explores this through a detailed discussion of two L2 speakers' monophthongal vowel productions, native listeners' judgments of the speakers' origin in a perception experiment, and qualitative comments on their speech in comparison to those of a native speaker, specifically examples of native-like or non-native-like elements provided by the listeners in their comments.

7.2 Method

7.2.1 Speakers

24 native and non-native speakers of New Zealand English (NZE) were recruited for the study. This chapter focuses on a 23-year-old male near-native speaker of English, called Kahui, who was a German L1 speaker, and a 21-year-old female near-native speaker of English, called Emily, who was a Korean L1 speaker. The comments that these two speakers received from the listeners are compared to those of one male native speaker of NZE M. Both were students at the same university in New Zealand at the time of the study.

Kahui began his formal study of English at the age of about nine, in his 'English as a foreign language' classes at school. Later, as an adult, he visited England for one month with the purpose of preparing for a standardized test of English proficiency before he moved to New Zealand 18 months prior to this study. He reported using English almost 100% of the time with the exception of weekly Skype sessions with his family in Germany.

Emily started learning English with a tutor in her home country at the age of 10. She spent 1 month in Australia at the age of 11 before moving to New Zealand permanently at the age of 12 in order to continue her education there. She reported speaking both languages an approximately similar amount at the time of the study: Korean with her family and friends in Korea and some Korean friends in New Zealand as well as English with her flat-mates, at the university, and with some Korean friends.

7.2.2 Production analysis

The speakers were audio-recorded speaking English with a head-mounted Opus 55.18 MKII beyerdynamic microphone and an H4n Zoom audio-recorder in four different settings for the broader purpose of studying style-shifting in non-native English speakers (see Gnevsheva, 2015a). They recorded their (1) social interaction with friends and (2) short service encounters (e.g., paying for gas and buying a coffee at a gas station) and were also interviewed by the author (3) about their studies on the university campus and (4) about their family and childhood in their home country at home. The total duration of recording per speaker equalled approximately one hour. The collection of recordings was followed by an interview eliciting biographical and attitudinal information.

The recordings were orthographically transcribed and uploaded into the corpus Accents of Non-Native English (ANNE; Gnevsheva, 2015b). ANNE is an instantiation of LaBB-CAT, which allows you to store time-aligned audio files and transcriptions and search for elements of interest (Fromont & Hay, 2012). Naturally, passing for a native speaker would require native-likeness in several linguistic domains such as segmental features, intonation, grammar, and vocabulary, but here I concentrate only on the production of monophthongal vowels. This is because one of the main distinguishing characteristics of New Zealand English is the realization of some of its vowels, particularly the short front vowels TRAP, DRESS, and KIT and fronted NURSE and GOOSE (see below; I use Wells' (1982) lexical set to represent the intended vowels). For the purpose of analyzing the speakers' native-likeness in their production of monophthongs, vowel formant measurements for all lexically stressed vowels were automatically extracted from ANNE with Praat (Boersma & Weenink, 2009). The corresponding vowel spaces for the speakers were plotted based on the mean vowel formants in R (R Core Team, 2012).

7.2.3 Perception stimuli and procedure

Short clips of a minimum of 25 words were extracted from the recordings in the four different settings for the perception experiment. The mean length for all clips was 26.6 words (13 seconds). Where possible, clips were extracted after the initial 5 minutes of recording, where the speaker might have been adjusting to the recording situation. The recordings were normalized to remove variation in volume. An attempt was made to select clips without errors and disfluencies in order to avoid their effect on the judges' ratings of accentedness; however, it was impossible to find such a

clip for all speakers. The clips did not contain proper names or other extra-linguistic information that might draw attention to the speakers' foreignness. The presentation of speakers and settings was randomized. Three clips per setting were extracted for each speaker in order to lessen any potential effect of particular individual recordings (see Leach, Watson, and Gnevshva, 2016), resulting in a pool of 288 clips (24 speakers * 4 settings * 3 clips), but each individual listener only heard one randomly chosen clip for every speaker by setting combination (96 clips).

Table 7-1 represents the number of words with lexically stressed vowels of the classes DRESS, GOOSE, KIT, NURSE, and TRAP in the twelve clips extracted for the speakers of interest Kahui and Emily. One can see that the exact number of tokens varied for the two speakers, and on average Kahui produced a higher number of these five vowels, possibly due to slightly longer clips. However, the relative distribution of the vowels is comparable. For example, TRAP is the most common vowel of the five, followed by KIT and GOOSE, for both Kahui and Emily. This justifies future comparisons of vowel distributions in listener comments for the two speakers (see below).

<i>Speaker</i>	<i>DRESS</i>	<i>GOOSE</i>	<i>KIT</i>	<i>NURSE</i>	<i>TRAP</i>
Emily	11	26	28	11	32
Kahui	24	33	34	8	45

Table 7-1. Number of tokens for some lexically stressed vowels in Kahui and Emily's clips.

The perception experiment was conducted in a quiet linguistics laboratory on computers with E-prime 2.0 software (Psychology Software Tools, 2012). In the perception experiment, 30 native NZE-speaking listeners were randomly presented with the stimuli one at a time over head-phones and were asked to (1) rate the speaker on a 1–7 accentedness scale, (2) then guess the origin of the speaker, and (3) finally, indicate what it was in the clip that made them answer the way they did. The responses were entered on the keyboard: number keys were used for the 1st question and open-ended response was typed in answer boxes for the 2nd and 3rd questions. The task took about 1 hour, and the participants received a coffee voucher for their time.

The geographical assignments given by the listeners to the speakers in the 2nd question were systematized and then grouped into three major native-likeness categories: native speaker of NZE, native speaker of another variety of English, and a non-native English speaker. NS of NZE included any reference to NZ or its parts (e.g., 'kiwi', 'Auckland'). NS of

another variety subsumed references to Australia, Canada, Ireland, the UK, and the USA. Broad responses without a clear reference, such as ‘western country’, were added to ‘no response’. All remaining responses were added to the NNES category (e.g., ‘India’, ‘China’).

In the content analysis of the listeners’ qualitative responses to the 3rd question, their comments were categorized in the following classes: accent (holistic comments), example (the way he/she said ‘X’; see below), specific segments, suprasegmentals, vocabulary, grammar, content (extra-linguistic comments), and zero response.

7.3 Results

In this section I focus on two near-native L2 speakers of English: Kahui and Emily. I explore (1) whether their vowel production is similar to that of an ideal native speaker of NZE, (2) whether listeners judge them to be a native speaker of (New Zealand) English, and (3) whether and what vowel production listeners comment on without prompting when judging them to be native speakers of NZE, in comparison to that of a native speaker.

7.3.1 Kahui

NZE has evolved as a distinct variety of English with some distinguishing characteristics: TRAP (/æ/) is half-open, DRESS (/e/) is half-close, KIT (/ɪ/) is centralized, NURSE (/ɜ/) is front, START (/ɑ/), GOOSE (/u/) and STRUT (/ʌ/) are central (see Figure 7-1; Wells, 1982). From visual analysis of Fig. 7-1 below, it is clear that Kahui’s vowel space is very similar to the prototypical NZE vowel space (Wells, 1982). Many vowels are quite NZE-like: for example, TRAP is somewhat raised, DRESS is very high and fronted, KIT is centralized, and GOOSE and NURSE are front; however, the overlap is not perfect with Kahui’s LOT and STRUT vowels being higher compared to the NZE ideal. Despite the visually quite nativelike vowel-space, a speaker may not pass for a native speaker for other segmental (consonants), suprasegmental (intonation), grammatical (syntactic deviations), and other reasons.

As expressed by him in the post-recordings interview, Kahui believed that he could pass for a native speaker, and he had anecdotal evidence of such passing:

Most often they [people] realize that I’m not native when I tell them or when they ask me, “Where do you come from?” They expect an answer like Napier or Christchurch... (Kahui.AA)

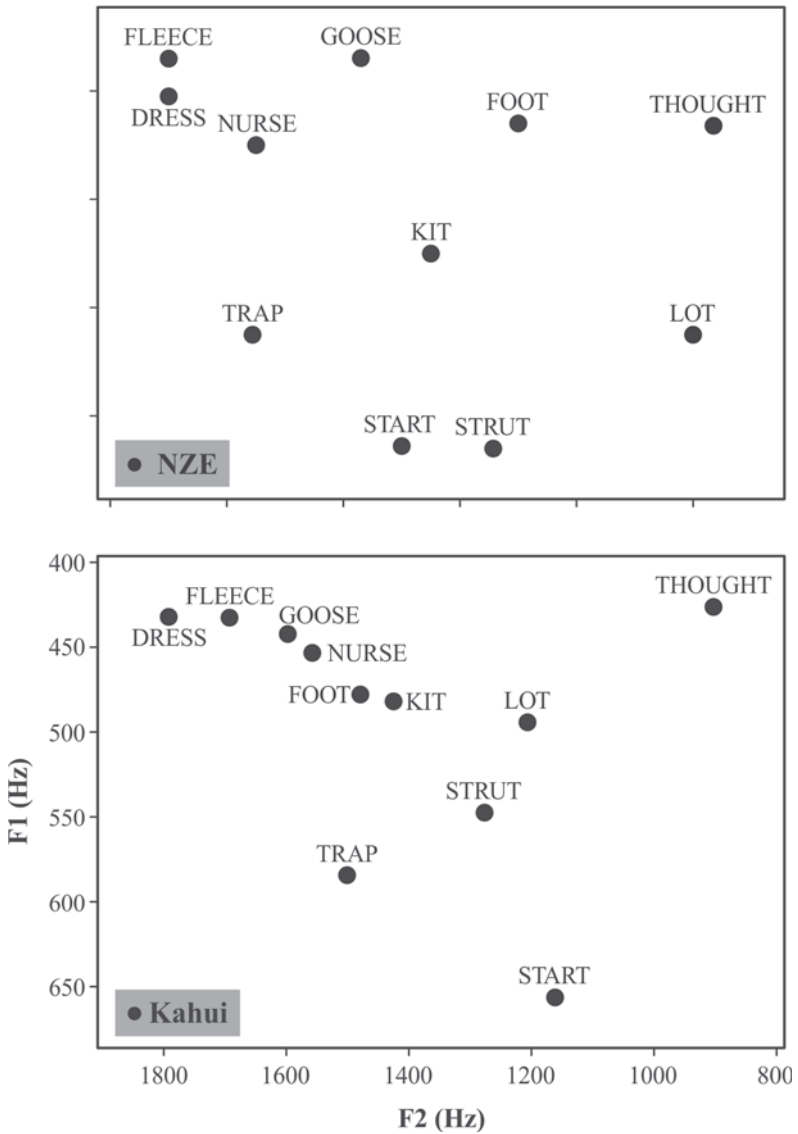


Fig. 7-1. Vowel spaces of NZE (top panel) and L1 German speaker Kahui (bottom panel).

His self-report of frequent passing is confirmed by his passing for a native speaker of different varieties of English in the perception experiment. In native listeners' perception, Kahui passed for a native speaker of NZE in the majority of cases (71.2%), a few times for a native speaker of other varieties of English (23.8%), and for just 2.5% of the listeners he did not pass for a native speaker, while 2.5% gave no response.

Content analysis of the listeners' comments revealed the following distribution of responses to Kahui's production: accent (46.6% of all comments), example (21.8%), segments (5.3%), suprasegmentals (15.0%), vocabulary (3.8%), grammar (0%), content (0%), and zero response (10%). Almost half of the descriptions were holistic comments which did not specify the features that they used to identify the origin of the speaker (e.g., 'kiwi accent'). Sometimes, the listeners indicated that they compared the clips to some sort of an ideal representation of the accent in their mind: '[s]ame accent as me', '...sounded like a friend of mine from Sydney', 'sounds like NZ TV presenter'. However, some listeners identified segmental features as a trigger (e.g., 'maybe australian [sic] with the vowel sounds'). Sometimes the listeners were more detailed and provided lexical examples/imitations (e.g., 'frish not fresh', which probably illustrates the raised quality of the DRESS vowel, typical of New Zealand English).

When the spelling of the example word illustrates what native or non-native feature was imitated ('frish not fresh'), the researcher can be fairly sure what feature was a salient marker of origin. But in many examples in this category the listeners simply used the word in its conventional spelling (e.g., 'the way he said ten') and without the listeners' orally imitating it, as in Hayes-Harb & Hacking (2015), it is difficult to tell which feature exactly elicited the listeners' reaction (e.g., stressed or unstressed vowels, consonants, suprasegmentals, or a combination of the above). However, for a speaker to be judged a native speaker, all the features in the example have to be native-like enough or not too non-native to be noticed. Because New Zealand English is quite distinct from other varieties in terms of its vowels (especially the short front vowels), I assume that listeners have to rely on the vowels in distinguishing New Zealand English from, for example, Australian English; therefore, I chose to analyze stressed vowels in the examples given when the speakers were judged a native speaker of New Zealand English.

All examples used by listeners when identifying Kahui as a native speaker of New Zealand English were categorized by lexically stressed vowel (see Table 7-2 for monophthongs). The listeners used five lexically stressed examples of the DRESS vowel, four of KIT, three each of GOOSE and TRAP, two of NURSE, and one each of FLEECE, START,

and STRUT. Most of these vowels are quite distinctive in NZE, and the listeners may be using them as a shibboleth. Interestingly, DRESS, and KIT, which are involved in a chain shift in NZE, and GOOSE, which is also one of the distinguishing characteristics of NZE, were commented on the largest number of times suggesting that the listeners focused on salient features representative of the variety.

<i>DRESS</i>	<i>FLEECE</i>	<i>GOOSE</i>	<i>KIT</i>
eleven	unbelievable	school	chilli
fresh		you	think
ten		youtube	

<i>NURSE</i>	<i>START</i>	<i>STRUT</i>	<i>TRAP</i>
working	mark	mum	thank

Table 7-2. Listeners' lexical examples when identifying Kahui as a NS of NZE.

Because of the small number of NNEs guesses, there were no examples provided by listeners to support their identification, but almost a quarter of listeners thought Kahui to be a NS of another English variety, and some of them used illustrations in their comments (see Table 7-3). The listeners used three lexical examples containing the stressed FLEECE vowel and one each of DRESS, GOOSE, STRUT, and THOUGHT. One can see that there is some difference between the vowels involved when Kahui was judged to be a NS of NZE and when he was judged to be a NS of another English variety. The short front vowels DRESS and KIT, as well as GOOSE, which are distinctive of NZE, are prevalent in Table 7-2 but only emerge twice in Table 7-3. Additionally, in Table 7-3, there are more illustrations of the FLEECE vowel which was only mentioned once in Table 7-2.

<i>DRESS</i>	<i>FLEECE</i>	<i>GOOSE</i>	<i>STRUT</i>	<i>THOUGHT</i>
ten	believe x 2 unbelievable	computes	suddenly	always

Table 7-3. Listeners' lexical examples when identifying Kahui as a NS of a different English variety.

For comparison, a native speaker of NZE, called M, also received many illustrations which contained characteristic NZE vowels: four of the DRESS vowel, two each of FLEECE, GOOSE, STRUT, and THOUGHT, and one each of KIT, LOT, NURSE, and START (Table 7-4), which

suggests that the listeners were noticing the native-like production of salient vowels.

<i>DRESS</i>	<i>FLEECE</i>	<i>GOOSE</i>	<i>KIT</i>	<i>LOT</i>
benefit	cheese	movies	will	what
guess	freeze	view		
then				
vet				

<i>NURSE</i>	<i>START</i>	<i>STRUT</i>	<i>THOUGHT</i>
perfectly	partner	lovely	awesome
		lunch	pause

Table 7-4. Listeners' lexical examples when identifying M as a NS of NZE.

To explore how the number of the characteristic vowels in a given clip influences its passing for a NS of NZE, along the lines of Leach, Watson and Gnevshva (2016), I calculated the number of words (Nwords), the number of lexically stressed vowels DRESS, KIT, TRAP, NURSE, and GOOSE, and the percentage of cases that it successfully passed for a NS of NZE (pass) for each of the twelve clips. I conducted Principle Components Analysis (PCA) on these data. The loading plot in Fig. 7-2 represents the relationships between the variables in the space of the first two components. We can see that the percentage of passing and the number of lexically stressed KIT vowels in a clip have similar heavy loadings. This suggests that they're positively correlated. On the other hand, the number of lexically stressed TRAP vowels seems to be negatively correlated with the percentage of passing. This suggests that the more KIT vowels were present in a given clip, the more it was likely to pass for a NS of NZE, and the more TRAP vowels were present, the more it was unlikely to pass. This negative correlation for TRAP suggests that this vowel may not have been native-like in Kahui's production, but one example containing the TRAP vowel was found in the listeners' comments when identifying him as a NS of NZE. Because there was only one example, it is difficult to draw conclusions based on these data.

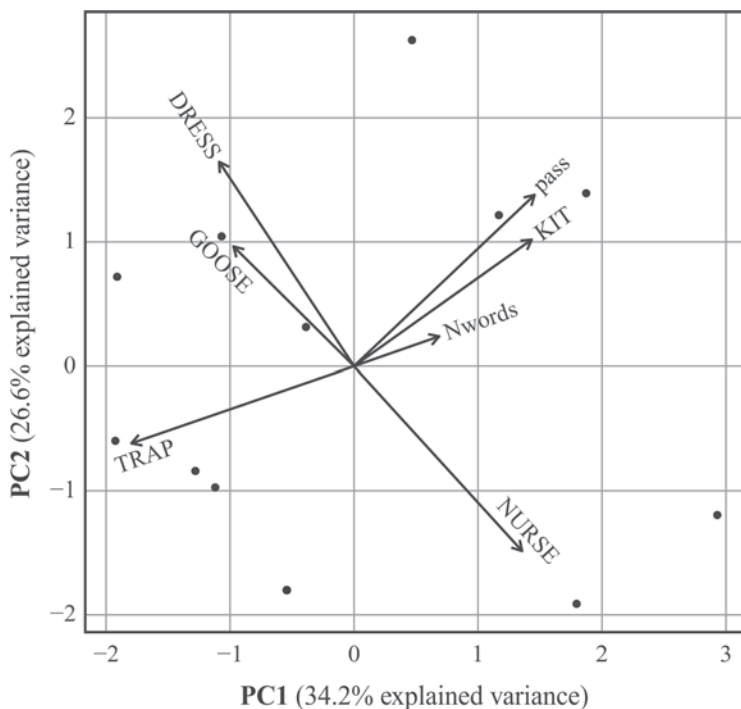


Fig. 7-2. The loading plot of PCA for Kahui.

7.3.2 Emily

The monophthongs produced by Emily were also quite native-like (Fig. 7-3): for instance, TRAP and DRESS are raised, GOOSE is fronted; however, there are some differences as Emily's KIT is high and front and very close to DRESS, NURSE is mid central, and LOT and STRUT are higher compared to the NZE ideal.

Emily also self-reported regular passing for a native speaker:

Oh, they usually ask me whether I'm from New Zealand or Australia yeah they know that I'm not from I wasn't born in here... but they ask me whether I was born in here cos through my accent... New Zea... they think I'm from New Zealand because of my accent... but I'm not sure cos... by my appearance so... (Emily.AA)

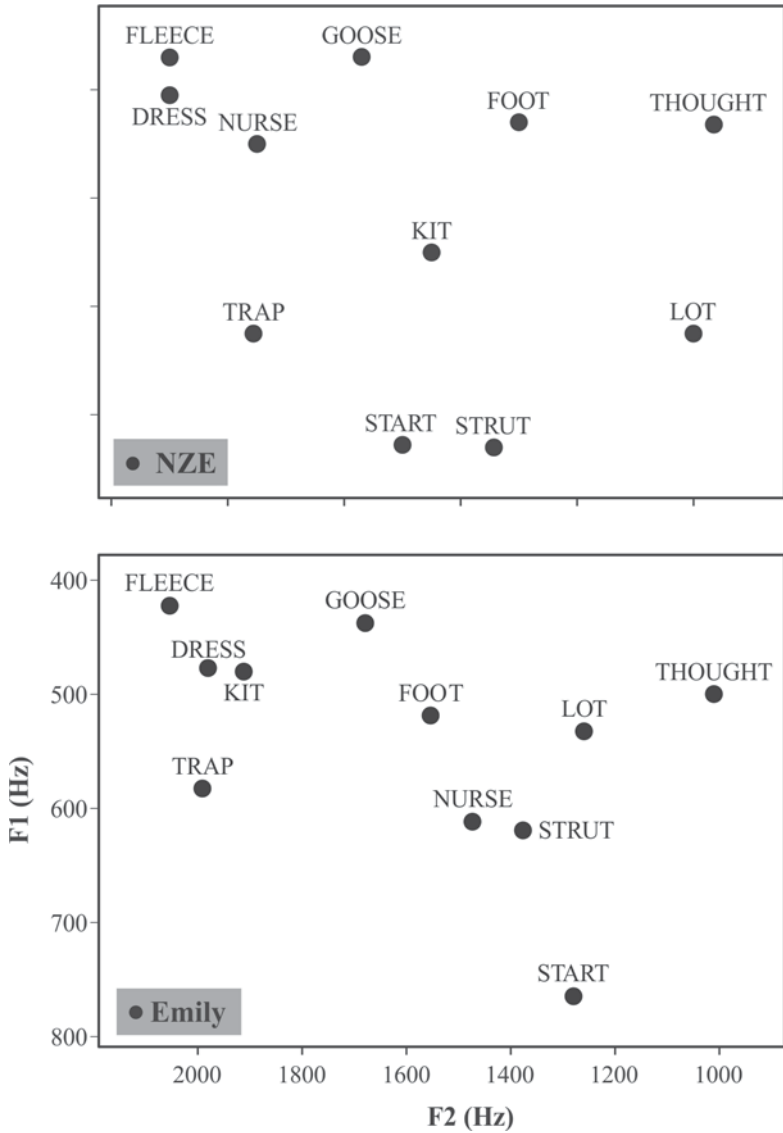


Fig. 7-3. Vowel spaces of NZE (top panel) and L1 Korean speaker Emily (bottom panel).

In native listeners' perception, Emily passed for a native speaker of NZE (48.7%) much less frequently than Kahui despite her quite NZE-like vowel space, which probably illustrates listeners' reliance on a wide range of cues including segmentals and suprasegmentals. She was believed to be a NNES by a large number of listeners (43.6%), and 4.3% identified her as a NS of another variety. Interestingly, several of the listeners who thought Emily to be from New Zealand and gave her a low accentedness score (more native-like) clarified that they believed her to be Māori ('sounds like a Maori [sic] girl'). In their justifications, the listeners commented on her pronunciation of consonants (e.g., 'fank you') illustrating how foreign influences can be heard as dialectal (Markham, 1997 as cited in Major, 2001) in the case of the voiceless dental fricative /θ/ which is absent in Korean.

Listeners' comments on Emily's production fell into the following categories: accent (52.6% of listeners), example (19.6%), segments (5.3%), suprasegmentals (12.0%), vocabulary (6.0%), grammar (0.8%), content (0%), and zero response (3.8%). When judging Emily to be a native speaker of NZE, listeners offered several lexical examples to support their identification. If we classify these examples by stressed monophthongs, they fall into five categories (Table 7-5). The listeners used four lexically stressed examples of the GOOSE and the TRAP vowels, two of the DRESS vowel, and one each of START and STRUT. Once again, we see the vowels that are representative of NZE. It is also a subset of Kahui's lexical examples (Table 7-2). Intriguingly, two of the categories that are missing for Emily in comparison to Kahui are KIT and NURSE, the two vowels which are less NZE-like in her vowel space (Fig. 7-3). I take this as evidence of the listeners' noticing that these two salient vowels were not target-like enough to trigger a native speaker of NZE guess.

<i>DRESS</i>	<i>GOOSE</i>	<i>START</i>	<i>STRUT</i>	<i>TRAP</i>
yeah	huge	market	just	dad
yep	you x 3			grandmother thank x 2

Table 7-5. Listeners' lexical examples when identifying Emily as a NS of NZE.

Emily received few NS of a different variety identifications, and no illustrations were used then. However, she was believed to be a NNES almost as often as a NS of NZE, and several lexical examples were used for justifications (Table 7-6). Some differences between the examples in Tables 7-5 and 7-6 can be observed: there are fewer examples of

distinguishing NZE vowels ('thank you' and 'tend') and examples of vowels which were not found in Table 7-5 (e.g., 'positions' and 'product').

<i>DRESS</i>	<i>FLEECE</i>	<i>GOOSE</i>	<i>KIT</i>	<i>LOT</i>	<i>START</i>	<i>TRAP</i>
tend	people	you	positions	product	market	thank

Table 7-6. Listeners' lexical examples when identifying Emily as a NNES.

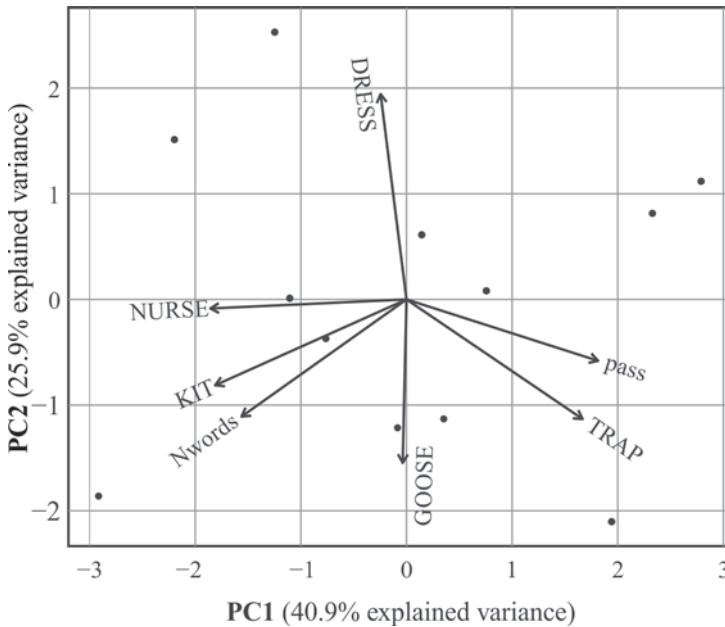


Fig. 7-4. The loading plot of PCA for Emily.

To explore how the number of the characteristic vowels in a given clip influences its passing for a NS of NZE for Emily, I calculated the number of words, the number of lexically stressed vowels DRESS, KIT, TRAP, NURSE, and GOOSE, and the percentage of cases that it successfully passed for a NS of NZE for each of the twelve clips in the same way that I did for Kahui. In Fig. 7-4 we can see that Emily's percentage of passing is positively correlated with the number of lexically stressed TRAP vowels in the clip. This suggests that the more of the TRAP vowels were present in the clips, the more likely she was to pass. Emily's native-like production of the vowel seen in Fig. 7-3 and the listeners' frequent

comments on this vowel when identifying her as a NS of NZE (Table 7-5) suggest that the listeners were relying on the TRAP vowel as a marker of NZE-likeness and a larger number of it present in a given clip supported their NS of NZE categorization.

7.4 Discussion

In this study the production of monophthongal vowels and passing for a native speaker of New Zealand English was scrutinized for two non-native speakers. The listeners' comments on certain features, which in their understanding influenced their speaker origin judgments, were contrasted with the speakers' actual production. Kahui's overall production of monophthongs was quite native-like, with NZE-like production of many of the characteristic vowels, despite no perfect overlap between his vowel space and the prototypical NZE vowel space. Kahui self-reported experience of passing for a NS, and his claim was supported by the perception experiment in which he was believed to be a native speaker of NZE almost 70% of the time. When the listeners had to justify their decision, the words mentioned as examples contained vowels which are different in other varieties of English and are salient markers of New Zealand identity (Hay & Drager, 2010); these were also produced by Kahui in a NZE-like fashion, according to his vowel space. This illustrates that there is some value in obtaining listeners' qualitative comments about their decision making in perceptual tests, even though work in perceptual dialectology and folk linguistics is traditionally cautious about trying to map such comments directly to speakers' linguistic production (see e.g., the discussion in Montgomery, 2007). The relatively frequent use of examples by listeners in the experiment (around 20% for both speakers) points to the importance of such comments and warrants further exploration.

Emily's near-native vowel space did not result in frequent passing comparable to that of Kahui. This may be reflective of her non-native-likeness in other linguistic and extra-linguistic domains. Additionally, unlike Kahui's, her production of the KIT and the NURSE vowels, distinguishing features of NZE, was less NZE-like, which may have made her non-native-likeness more noticeable through more deviations from the ideal NZE speaker in the listeners' expectation.

The speakers' frequent passing for a native speaker of New Zealand English despite some non-native features in their vowel spaces and potential deviations in other linguistic elements suggests that linguistic production need not be perfectly native-like for passing to occur. This

corroborates Schmid et al.'s (2014) finding that the L2 speaker that received the most native-like accentedness rating did not always fall within the native speaker range in production.

The difference between Kahui and Emily's production of vowels (Emily's less native-like production of KIT and NURSE) may be correlated with the difference in listeners' examples of words produced in the native-like manner (no examples of KIT- and NURSE-containing words for Emily when she was thought to be a NS of NZE and an example of KIT when she was believed to be a NNE). This suggests that certain linguistic features may be more salient to listeners or be more important markers of identity and/or origin, but also native-likeness in these salient features may be more important for passing.

This study highlighted the importance of native-like production of salient segments for high native-likeness judgments. Kahui's native-like production of characteristic NZE vowels and the listeners' noticing of it raises a further salience question which can potentially be addressed in future experiments: will native-likeness in characteristic vowels coupled with non-nativelikeness in other ones result in the same nativelikeness ratings as native-likeness in all vowels? Qualitative studies can help to shed light on the processes underlying listeners' judgments of speakers' nativeness and origin.

References

- Abrahamsson, N. & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30/4, pp. 481–509.
- Abrahamsson, N. & Hyltenstam, K. (2009). Age of onset and native-likeness in a second language: listener perception versus linguistic scrutiny. *Language Learning*, 59/2, pp. 249–306.
- Anderson-Hsieh, J., Johnson, R. & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42/4, pp. 529–555.
- Boersma, P. & Weenink, D. (2009). Praat: Doing Phonetics by Computer (Version 5.1.22). Retrieved on October, 12, 2012, from <http://www.praat.org/>
- Bongaerts, T., Mennen, S. & Van Der Slik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as second language. *Studia Linguistica*, 54/2, pp. 298–308.

- Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, 63/3, pp. 544–573.
- Derwing, T.M. & Munro, M.J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, 19/1, pp. 1–16.
- Fromont, R. & Hay, J. (2012). *LaBB-CAT: An Annotation Store*. Paper presented at Australasian Language Technology Association Workshop.
- Gnevshcheva, K. (2015a). Acoustic analysis in Accent of Non-Native English (ANNE) corpus. *International Journal of Learner Corpus Research*, 1/2, pp. 256–267.
- . (2015b). Style-shifting and intra-speaker variation in the vowel production of non-native speakers of New Zealand English. *Journal of Second Language Pronunciation*, 1/2, pp. 135–156.
- . (2015c). Within-speaker variation in passing for a native speaker. *International Journal of Bilingualism*. Advance online publication DOI: 10.1177/1367006915616197
- Hay, J. & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48/4, pp. 865–892.
- Hayes-Harb, R., & Hacking, J. (2015). Beyond rating data: What do listeners believe underlying their accentedness judgments? *Journal of Second Language Pronunciation*, 1(1), 43–64.
- Ioup, G., Boustagui, E., El Tigi, M. & Moselle, M. (1994). Reexamining the critical period hypothesis. A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, 16, pp. 73–98.
- de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R. & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34/01, pp. 5–34.
- Kraut, R. & Wulff, S. (2013). Foreign-accented speech perception ratings: a multifactorial case study. *Journal of Multilingual and Multicultural Development*, 34/3, pp. 249–263.
- Leach, H., Watson, K. & Gnevshcheva, K. (2016). Perceptual dialectology in northern England: Exploring the effects of proximity and salience on accent identification. *Journal of Sociolinguistics*, 20, pp. 192–211.
- Magen, H.S. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, 26/4, pp. 381–400.
- Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah: Lawrence Erlbaum Associates.
- Marx, N. (2002). Never quite a “Native Speaker”: Accent and identity in the L2 – and the L1. *The Canadian Modern Language Review*, 59/2, pp. 264–281.

- McCullough, E. A. (2013). Perceived foreign accent in three varieties of non-native English. *OSUWPL*, 60, pp. 51–66.
- McKenzie, R. M. (2015). The sociolinguistics of variety identification and categorisation: Free classification of varieties of spoken English amongst non-linguist listeners. *Language Awareness*, 24/2, pp. 150–168.
- Montgomery, Ch. (2007). *Northern English dialects: a perceptual approach*. Unpublished PhD thesis, University of Sheffield.
- Moyer, A. (2004). *Age, Accent, and Experience in Second Language Acquisition: An Integrated Approach to Critical Period Inquiry*. Clevedon: Multilingual Matters Limited.
- Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: acoustic measurements and accentedness ratings. *Language and Speech*, 36/1, pp. 39–66.
- . (2008). Foreign accent and speech intelligibility. In: Edwards, J. G. H. & Zampini, M. L. (Eds.), *Phonology and Second Language Acquisition*, pp. 193–218. Philadelphia: John Benjamins Publishing.
- Munro, M. J. & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45/1, pp. 73–97.
- Munro, M. J. & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation*, 1/1, pp. 11–42.
- Munro, M.J., Derwing, T.M. & Burgess, C.S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52, pp. 626–637.
- Munro, M. J., Derwing, T. M. & Holtby, A. K. (2011). *Evaluating individual variability in foreign accent comprehension*. Paper presented at The 3rd annual pronunciation in L2 learning & teaching conference, Iowa State University.
- Piller, I. (2002). Passing for a native speaker: identity and success in second language learning. *Journal of Sociolinguistics*, 6/2, pp. 179–206.
- Preston, D.R. (1996). Whaddayaknow? The modes of folk linguistic awareness. *Language Awareness*, 5/1, pp. 40–74.
- Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from <http://www.pstnet.com>.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved on October, 12, 2014, from <http://www.R-project.org/>

- Schmid, M. S., Gilbers, S. & Nota, A. (2014). Ultimate attainment in late second language acquisition: Phonetic and grammatical challenges in advanced Dutch–English bilingualism. *Second Language Research*, 30/2, pp. 129–157.
- Wells, J. C. (1982). *Accents of English*. Vol. 3. Cambridge, Cambridge University Press.

PART III:

MAPPING THE DIFFERENCES IN PROSODY

CHAPTER EIGHT

PHONETIC CONNECTEDNESS IN NON-NATIVE SPEECH

ŠÁRKA ŠIMÁČKOVÁ
AND VÁCLAV JONÁŠ PODLIPSKÝ¹

8.1 Introduction

Some learners of English as a second or even foreign language achieve a high degree of both general and specifically oral proficiency. Nonetheless, even such highly proficient learners are often rated as less fluent than native speakers (e.g. Riazantseva, 2001), and research into what impedes fluency in non-native speech is abundant (e.g. studies cited below in 8.1.1). Fluency is a phenomenon of great complexity (Segalowitz, 2010) and our perception of someone's speech as fluent is a product of a variety of factors. Consequently, researchers have approached the study of fluency from different perspectives, including linguistic, psycholinguistic and sociolinguistic perspectives, as discussed in the overview in Mora (2006). In this chapter we adopt a narrow view of fluency (for broad and narrow fluency see Lennon, 1990) and focus on the fluidity of producing phonetic segments of speech. Isaac and Trofimovich's (2011) succinct statement provides us with a working definition: in their words, fluency "denotes listeners' assessments of how smoothly and rapidly an utterance is spoken" (p. 114). However, rather than with fluency as experienced by the listener, which Segalowitz (2010) labels perceptual fluency, here we are concerned with aspects of utterance fluency measurable in speech recordings. We want to investigate phonetic properties of utterances that are more or less smooth as well as fast.

¹ The work of the first author was supported by the internal research grant FPVC2014/29 from the Faculty of Arts of Palacký University Olomouc.

8.1.1 Global and local fluency

Even within the narrow approach to the fluency of utterances, multiple contributing factors can be identified and researchers have employed various quantifiable global measures to gauge fluency. These measures are either temporal, such as speech rate and articulatory rate, duration of pauses, and duration of pause-free turns (Freed, Segalowitz & Dewey, 2004; Kormos & Dénes, 2004; Mora, 2006; Ginther, Dimova & Yang, 2010), or they are occurrence measures assessing hesitation phenomena, such as the frequency of filled and silent pauses, and repair phenomena, e.g. frequency of repetitions and repairs (e.g. Lennon, 1990; Riggenbach, 1991; van Gelderen, 1994; Freed, 1995, 2000; Kormos & Dénes, 2004). The temporal and occurrence measures reflect phenomena pervasive throughout spoken utterances: one speaks more or less slowly or fast, with or without pausing or hesitating. In this sense the speed of delivery and the amount of hesitation have a global influence on the fluency of an utterance, affecting perceivers at any moment of their listening to second-language (L2) speech. The global factors have been intensively researched and a link especially between temporal factors and overall fluency has been repeatedly observed (Lennon, 1990; Riggenbach, 1991; van Gelderen, 1994; Freed, 1995). It is methodologically convenient to operationalize utterance fluency as a relatively easily quantifiable measure, e.g. the number of syllables per second or the number of pauses in a clause, which can be correlated with the degree of perceived fluency. In addition, the global measures have an intuitive validity for listeners. We may easily become aware of a speaker's slow speech rate, long pauses, or too many "ehms". In studies of perceptual fluency, raters often have similar comments during their debriefing (e.g. Kormos & Dénes, 2004).

Occasionally, there have been calls for considering speech fluency in terms of qualitative as well as quantitative descriptors (Hieke, 1984, 1985; Anderson-Hsieh, Riney, & Koehler, 1994). In the former category, Hieke includes phonological processes which lead to reductions, simplifications, and linking in connected speech. Interestingly, raters' comments elicited during global fluency experiments indicate that their subjective evaluation of fluency is also affected by non-temporal factors, with learners' pronunciation of the non-native language being the most important among them (Freed, 1995: 136; Rossiter, 2009). However, in a recent study by Pinget, Bosker, Quené & de Jong (2014), perceived fluency and perceived accent were found to be only weakly related. Such contradictory reports suggest that the connection between perceived fluency and foreign-accentedness of L2 speech deserves to be investigated more closely. It is likely that not all deviations from native pronunciation contribute to the

perception of dysfluency, hence the lack of a robust link between fluency ratings and accent ratings. Here we propose that the foreign accent features which matter to fluency are the phonetic realizations of segments at word boundaries. These word-to-word transitions in connected speech can be viewed as local contributors to fluency which have an effect on the perception of smoothness of speech. An experienced non-native speaker of English may produce speech that is continuous, i.e. fluent globally, but not smooth, i.e. dysfluent locally.

8.1.2 Pronunciation of sounds at boundaries of words in continuous speech

In spoken English discourse, segments at boundaries of words which are tied together within a prosodic unit often undergo modifications which simplify the production of segment sequences, making the word boundary less audible and creating the impression of a smooth flow of speech. These modifications, which have been described as external sandhi processes (Cruttenden, 2014) or effects of gestural overlap (Browman & Goldstein, 1992), include assimilations, coalescences, elisions, liaison, and others. Their application is conditioned by the prosodic organization of speech (Nespor and Vogel, 1986) in such a way that a boundary process is more likely to take place when the adjacent words do not straddle a strong prosodic boundary. In this way, boundary processes fuse together words that form prosodic chunks. In non-native English speech such boundary phenomena are often sporadic (e.g. Zampini, 1997; Nguyen and Ingram, 2004). Their absence may result in reduced intelligibility due to distorted perception of prosodic structure, it may give rise to perceived hyper-articulation or, as we would like to propose, to perception of dysfluency.

For English-as-a-foreign-language (EFL) learners these sandhi processes are not automatic, and usually they must be learned as explicit knowledge (Kormos, 2011) that only gradually becomes routinized. Their application may also be subject to non-native constraints. One such L2 constraint was proposed by Cebrian (2000). His ‘interlanguage word-integrity effect’ describes learners’ tendency to “[treat] every word as a separate unit and [prevent] the articulatory synchronization of sounds belonging to different words” (p. 19). Cebrian suggests that in learners’ interlanguage, post-lexical processes only apply within a phonological word. Zsiga’s (2003) study of stop-stop sequences found partial support for this effect although a recent study of voicing assimilation in L2 (Simon, 2010) does not support it. Additionally, as suggested above, L2 learners are known to speak more slowly than native speakers even at

advanced proficiency levels (e.g. Lennon, 1990), which may result in inhibiting already automatized sandhi processes, as well as in producing a smaller gestural overlap compared to native speech.

In this chapter the degree of word-to-word fluency in English learners' utterances is discussed in the light of two production studies examining two kinds of boundary events. First, we consider the onset of vowel-initial words and the likelihood of them being linked to the preceding context as opposed to being separated by glottalization. Next, we examine the offset of words ending in a stop consonant and the likelihood of producing the final stop without a release in the context of a following stop.

8.1.3 Eliciting continuous speech from learners of English

Since it only makes sense to evaluate local fluency on continuous utterances, we concentrate on English learners who are able to deliver stretches of speech without halting and at a speed that is relatively high, if not necessarily as high as in native speech. The overall proficiency level achieved by the non-native participants in our two studies was C1 or higher according to the CEFR rating scale (Verhelst, Van Avermaet, Takala, Figueras & North, 2009) as evidenced by their results in a Certificate in Advanced English test. In terms of their learning experience the learners formed a fairly homogeneous group. They were young Czech adults in their early or mid-twenties who had learned English mainly via formal classroom instruction, starting around the age of nine. None of them reported a regular contact with native English speakers or a stay in an English-speaking country before reaching high school age (ca around 15 years). At the time of the data collection they were in the third year of a bachelor degree programme in English. At the university they came in daily contact with native speakers and they had received instruction in English phonetics. Also, as future English language professionals they were highly motivated to improve their pronunciation in general and fluency in particular. Based on our experience with this learner population we expected our participants to have reached a relatively high degree of automaticity in their L2 production.

As we said, smoothness of speech that we also call local fluency can only be evaluated on continuous utterances. To elicit maximally continuous English speech from our learners we used the same type of a simple reading task in both studies. Although we are aware that a reading task may reinforce learners' awareness of word boundaries, we reasoned that reading familiar short statements containing simple vocabulary would

help to eliminate pausing and other dysfluencies connected with speech planning.

8.2 Word-initial vowels in connected speech

Sound patterns in connected speech are always included as a topic in advanced textbooks (O'Connor, 1980; Pennington, 1996; Roach, 2009; Cruttenden, 2014), online sources on English pronunciation (e.g. BBC Learning English: Connected speech, 2009), and in accent reduction manuals (e.g. James & Smith, 2012). Clearly, there is a shared understanding among practitioners in English phonetics that connected speech phenomena are important for “good pronunciation”, as well as agreement that these phenomena represent a distinct problem for English L2 learners. In this section we expanded and reanalyzed data from an earlier study (Šimáčková, Podlipský & Kolářová, 2014) in which we asked to what extent Czech EFL learners link vowel-initial words to the preceding context as opposed to marking the beginning of such words with glottalization, as they usually do in their L1.

In English several ways of linking the pronunciation of consecutive words are commonly described. When one word ends in a consonant and the following word starts in a vowel, the word-final consonant may be resyllabified into the following onsetless syllable. The resyllabified consonant may also be an *r*. In non-rhotic varieties this “etymological” final *r* is not pronounced if it cannot be linked to the vowel at the beginning of the next word, hence the label “linking *r*”. When two vowels occur on either side of the word boundary, transient glides are likely to link the words. A final high front or a high back vowel may be linked to the initial vowel in the following word by the glides [j] or [w] respectively. A word-final non-high vowel may, in a non-rhotic accent, be linked to the vowel at the beginning of the next word by an epenthetic intrusive *r*, though the occurrence of such “non-etymological” *r* is variable and sociolinguistically conditioned (Mompeán-Gonzalez & Mompeán-Guillamón, 2009). Table 8-1 illustrates the types of linking with examples.

type	example	
Resyllabification	<i>seen us</i>	[si.nəs]
Linking <i>r</i>	<i>scare us</i>	[skeə.ɹəs]
Intrusive <i>r</i>	<i>saw us</i>	[sɔ.ɹəs]
Transient [j]	<i>see us</i>	[si ^j əs]
Transient [w]	<i>sue us</i>	[su ^w əs]

Table 8-1. Linking word-initial vowels in English.

While linking is common, native English speakers do not always connect a vowel-initial word to the preceding context. Instead, they may mark the beginning of the word with glottalization. Linking and initial glottalization can be construed as two competing strategies for dealing with an empty onset of a word-initial syllable. Although linking is preferred in English overall, glottalization is by no means uncommon. For example, Dilley, Shattuck-Hufnagel and Ostendorf (1996) report glottalization rates between 13% and 44% for five American English newsreaders, and the two British newsreaders in Bissiri and Volín (2010) produced glottalization in 28% and 34% vowel-initial contexts. Even if a host of factors conditioning initial glottalization have been explored, including speaking tempo, lexical frequency, or segmental context (Umeda, 1978; Dilley et al., 1996; Mompeán & Gómez, 2011), the occurrence of glottalization at the expense of linking is most consistently predicted by prosody. A word-initial vowel is more likely to be preglottalized if it is stressed, if it is in a pitch-accented word, and especially if it occurs at the beginning of an intonational phrase (Garellek, 2012).

The fact that in English vowel-initial words in connected speech are linked whenever prosodically possible presents Czech learners of English with a rather difficult problem because of the strong L1 bias towards initial glottalization. In Czech, glottalization before a word-initial vowel (or, sometimes morpheme-initial vowel) is a low-level phonological process repairing onsetless syllables. Czech speakers produce glottalization at boundaries of words, such as in *ona vidí obraz* ['ʔona vɨji: 'ʔobras] “she can see a picture”, or morphemes, e.g. in *vyobrazit* ['vɨʔobrazit] “to depict”. However, observations of casual connected Czech speech show that glottalization is an optional process. Resyllabification of the final consonant from the preceding word into an empty initial onset does occur (Volín, 2010; 2012) and a sequence such as *vím o tom* “I know about it” may come out as ['vi:.mo.tom] instead of ['vi:m.ʔo.tom]. Compared to English though, such linking is much less frequent and stylistically marked: it is likely to be regarded as a sign of sloppy pronunciation (Palková, 1994). One exception is regional differences. In Moravian Czech (in the East of the country), resyllabification of a final consonant is fairly common and goes unnoticed (Šimáčková, Podlipský and Chládková, 2012).

Despite this variation in the amount of initial glottalization in Czech, there is still a clear difference between the learners' L1 and L2. Thus, their task to learn English linking is quite complex. Firstly, learning to link involves suppressing the transfer of automatic glottal stop insertion. Instead, in the context of a preceding consonant the learner needs to allow

resyllabification to emerge. In the context of hiatus (*see us, sue us*), the learner also has to notice the transient glides in the English input, learn to form the transient glides and finally automatize gliding from the word-final high vowel to following initial vowel.

In our study, rates of linking and glottalization in the contexts for consonant resyllabification, linking *r*, and transient glides [j] and [w] were compared in the pronunciation of 10 native English speakers and 20 advanced Czech EFL learners selected from the population described in section 8.1.3. No predictions were made about the native speaker data, except that the rates of linking and glottalization were expected to conform to those reported in literature. Native speaker data served as a yardstick to which the L2 learners' production of linking could be compared. We made the following two predictions about the rates of linking versus glottalization in the learners' speech: (1) Since the learners' native dialect is Moravian Czech we predicted the highest rate of linking in the resyllabification context (resyllabification occurs in Moravian Czech whereas transient glides do not). (2) Considering that Czech-accented English is at least variably rhotic, i.e. final post-vocalic *r*'s are often pronounced and so can be resyllabified, we also predicted a higher rate of linking in the final *r* context compared to the transient glide contexts.

8.2.1 Methodology

In order to elicit continuous articulation without pauses and restarts we asked the participants to read simple short statements (4 to 7 words, averaging 6.6 syllables), each pronounceable as a single intonational phrase. Prior to the recording, the participants had enough time to familiarize themselves with the statements which were then presented one at a time on a computer screen at a pace comfortable for each speaker. The data were collected by two student assistants who were close to the participants in age and easily created an informal atmosphere. The data collector instructed each participant to read the displayed statement once for practice, then read it out loud as naturally as possible, and finally say it facing him or her. The order of sentences was individually randomized. Though the third repetition was expected to be the most natural one, in the end we judged the second versions as the most continuous ones and used them for analysis.

Altogether the stimulus set comprised 60 sentences covering five linking-glottalization contexts, shown in Table 8-2. Twenty sentences contained a vowel-initial target word placed after a word ending in a high vowel. In half of these cases the final high vowel was front, producing a

context for transient [j] insertion, in the other half it was back, producing a transient [w] insertion context. In another 20 sentences the word preceding the vowel-initial target ended in the letter ‘r’ creating an opportunity for the pronunciation of a linking *r*. The last 20 sentences contained word-final consonants followed by vowel-initial targets. Half of the consonants were voiced obstruents and half were sonorants, i.e. /l/ and nasals.

Context		Linking type	N of stimuli	Example
high V	i#V	transient [j]	10	I have <i>three empty</i> cans.
	u#V	transient [w]	10	I read <i>two articles</i> today.
final <i>r</i>	r#V	linking <i>r</i>	20	<i>Four agents</i> are missing.
final C	son#V	resyllabification	10	He <i>ran across</i> the street.
	obs#V	resyllabification	10	I <i>love open</i> spaces.

Table 8-2. An overview of the linking–glottalization contexts and example stimuli.

Voiceless obstruents were excluded to avoid irregularities connected with initialization of voicing. Also, final voiceless stops may themselves be preglottalized, which may lead to glottalization of the following vowel.

8.2.2 Analysis and results

We expanded Šimáčková, Podlipský and Kolářová’s study (2014) to include data from 10 native speakers, five British and five American. They were university students aged 20 to 30 with little or no knowledge of Czech. Thus, the current analysis includes between-group comparisons. On the other hand, we reduced the number of the non-native participants by excluding data from one speaker whose result was a clear outlier. His mean rate of linking was higher than that of any of the native speakers and 3.3 standard deviations above the mean of the EFL learner group.

Target sequences were analyzed both auditorily and by viewing spectrograms and waveforms in Praat (Boersma and Weenink, 2011). Monitoring several acoustic correlates of perceived glottalization described in previous studies (e.g. Dilley et al., 1996; Skarnitzl, 2004), each token was classified either as a full glottal stop (a period of silence followed by a burst and an onset of voicing), as non-canonical glottalization (including aperiodicity, creak, and diplophonia), or a case of linking (no glottalization detected auditorily or visually). For each participant and each context, mean proportions of linking, non-canonical glottalization, and full glottal stops were computed.

An overall linking score, computed as the mean proportion of linking across all contexts, was submitted to a one-way ANOVA, with Speaker group (Czech EFL learners, native speakers of American English, and of British English) as the independent variable. The effect of Speaker group was significant ($F[2, 27] = 10.5, p < .001$). A post-hoc Tukey HSD test showed that this overall proportion of linking was significantly lower for the EFL learners (.38) than for the American (.57) or British (.67) native speakers who did not differ from each other. As can be seen in Fig. 8-1, the group means conceal a great deal of individual variation, especially among the learners. Native speakers also varied, albeit to a lesser degree (standard deviations in learners', the American, and the British data was .16, .05, and .11 respectively). Interestingly, the speakers' means, rising from left to right in Fig. 8-1, show that all our British speakers produced more linking than the Americans.

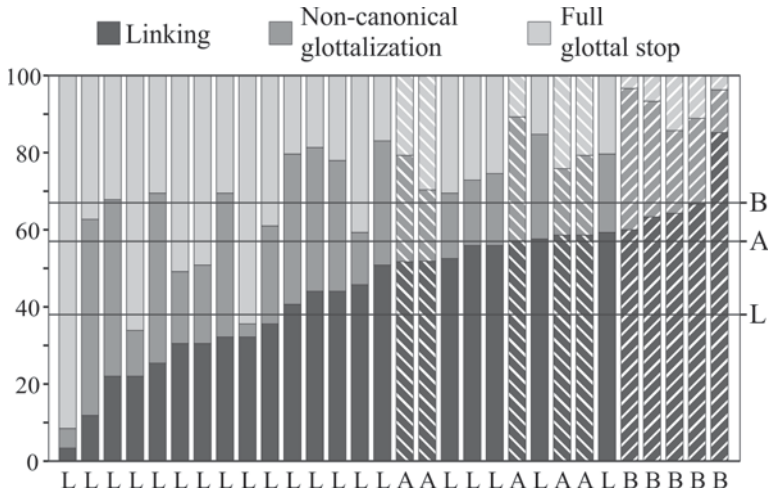


Fig. 8-1. Percentages of linking, non-canonical glottalization and full glottal stops produced by individual Czech EFL learners (L) and native speakers of American (A) and British (B) English. Group means are represented by horizontal lines.

Five subsequent one-way ANOVAs examined the linking scores in the individual segmental contexts. As can be seen in Fig. 8-2, the Czech learners most successfully approached the native rates of linking in the word-final obstruent context (obs#V), in which their proportion of linking was the highest, and the ANOVA found no significant effect of speaker group ($p = .38$). In contrast, in all the other contexts Speaker group had a

significant effect ($F(2,27) > 6.4, p < .01$). Tukey HSD tests confirmed that in all the other contexts the learners produced a significantly lower proportion of linking than the British native speakers ($p < .05$). Compared to the Americans, the learners produced a significantly ($p < .05$) lower proportion of transient [j]'s in the i#V context, and their proportion of linking in the final *r* context (r#V) was also lower, the difference from the Americans approaching significance ($p = .07$). In the final sonorant (son#V) and final high back vowel (u#V) contexts, the difference between the learners and American native speakers was not significant.

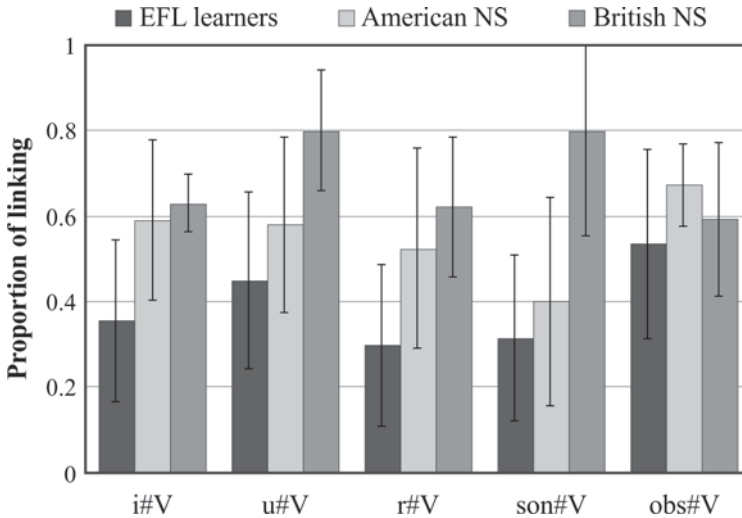


Fig. 8-2. The proportion of linking in each segmental context for each speaker group. Means and standard deviations.

8.2.3 Discussion

Our prediction that resyllabification of the final consonant would be more common than the occurrence of transient glides in the hiatus context in Moravian Czech learners' English was confirmed for obstruents but not for sonorants. Our prediction about linking *r* was not confirmed. Final *r* patterned with the other sonorants and was not linked to the following word more often than the preceding high vowels.

Our EFL learners' mean rate of *non-linking* (canonical and non-canonical glottalizations combined) was 62%. This is lower than the rates reported for the four Czech university students reading English news

bulletins in Bissiri and Volín (2010), probably due to the different difficulty of the elicitation materials or differences in the speakers' overall foreign-accentedness, but it is still higher than the rate of non-linking for our native English speakers: the learners clearly linked less and glottalized more. To evaluate to what degree this is due to transfer from L1, we can make a comparison with Volín's (2012) analysis of native Czech speech. He reported variation related to gender and speaking style. Professional male and female newsreaders produced 88% and 97% of word-initial vowels as glottalized. In spontaneous dialogues the rate of glottalization was significantly lower: 41% for male and 65% for female university students. The last-mentioned value seems most relevant to our discussion. Though it is true that our participants produced read speech, the simplicity of our reading material, the generous familiarization period, the informal nature of the recording session, and the repetition of the stimuli suggest a comparison with Volín's (2012) informal dialogue speech rather than with the reading of the news. Moreover, recall that 16 out of 20 our participants, also university students, were women. Thus, when we compare glottalization rates of our EFL learners' and Volín's Czech female speakers, we see that Czechs deal with vowel-initial words in continuous speech in the same way in their L2 as they do in their L1.

Although the group rates of linking suggest that Czech EFL learners have not acquired this aspect of connected speech, individual speakers' means show that there are learners who produce linking at English native speakers' rates. Individual means of 8 out of the 20 EFL learners fit within one standard deviation from the American group mean and two learners fit within one standard deviation from the British group mean.

The difference between the British and American speakers, though not statistically significant, merits comment. Our results suggest that British speakers link more than Americans. Perhaps it is not a coincidence that empirical studies of prevocalic glottalization, i.e. *non-linking*, have been carried out almost exclusively with Americans. This potential difference between British and American English is also likely to have contributed to the variation across the EFL learners, whose type of native speaker input we did not control for.

8.3 Word-final stop consonants in connected speech

While in the previous section we looked at how the onset of a word can be fused with the preceding context, now we focus on word offsets. In particular, we examine pronunciation of word-final stops which are adjacent to another stop in the next word, e.g. in *dig deep*. In native

English such stop-stop sequences are typically realized without an audible release of the first stop (e. g. Ladefoged, 1993). The absence of an audible release, or in short ‘unreleasing’, means that the combined closure of the two consecutive stops extends over the word boundary. This binds neighbouring words together perceptually. Unlike linking, this feature of connected speech may not always be included in English pronunciation teaching materials. Still, the presence of unexpected releases in non-native speech may weaken the listener’s impression of a smooth flow, as is illustrated by Ladefoged’s example *It’s a bigga day* caricaturing Italian-accented English (1993:52).

Unreleasing is usually treated in phonetic literature as a result of gestural overlap (e.g. Hardcastle and Roach, 1977) and has been shown to interact with the relative place of articulation of the two stops. If the two stops have the same place of articulation, the first one is unreleased; if the second stop in the sequence is labial, the release of the first stop is less likely (Henderson and Repp, 1982). More generally, if the second stop is more anterior, an audible release of the first stop is less likely because it is effectively hidden by the constriction of the second stop (Zsiga, 2003). However, in a study of spontaneous American speech this pattern was found only word-internally (Davidson, 2011) and some studies report the opposite effect (Byrd, 1996).

In Czech-accented English the stop-stop sequences are often realized with an audible release of the first stop as illustrated in Figure 8-3. This may be taken as interference from the L1; unreleasing is not typically discussed in textbooks of Czech phonology (e.g. Palková, 1994). However, in a relatively recent acoustic study of spontaneous Czech (Šimek, 2010), approximately a third of first stops were in fact unreleased. In English productions by advanced Czech learners, we observe quite a lot of between- and within-speaker variation. It is this variation that we aim to explore in this section. The potential contributing factors we consider are the interference of L1 timing of articulatory gestures, the relative place of articulation of the two stops, the interlanguage word-integrity effect, and articulatory rate. Six new advanced EFL learners drawn from the same participant pool as the learners in the glottalization study volunteered to be recorded. In addition, we collected baseline monolingual data from four English native speakers (2 females and 2 males) and six Czech native speakers (all female), all of whom reported low proficiency in foreign languages. In the following sections the three speaker groups are coded as L (learners), E (English near-monolinguals), and C (Czech near-monolinguals).

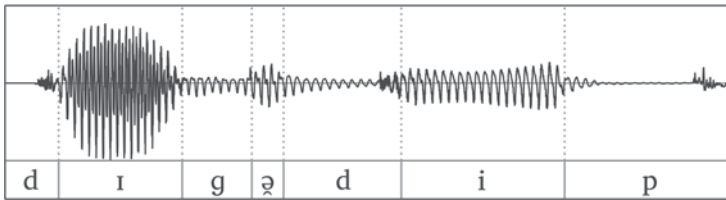


Fig. 8-3. The phrase *dig deep* pronounced by an advanced Czech EFL learner.

8.3.1 Methodology

To elicit productions of the target stop-stop sequences, we used 90 English and 90 Czech statements, manipulating 2 factors, displayed in Table 8-3. The first factor, ‘Boundary’, was whether the target sequence spanned a word boundary between two lexical words, as in *met Pam*, or only a syllable boundary within a compound, as in *output*. The second factor was the place of articulation of the second stop relative to the first one (‘Place of S2’), which was either homorganic (18 word-boundary sequences) or heterorganic (4 × 18 sequences). The heterorganic group was split between combinations with S2 ‘fronter’ than S1, as in *met Pam*, and combinations with ‘backer’ S2, as in *keep calm*. As in the previous study, the statements were short (7.1 syllables on average) and contained relatively common vocabulary. The data collection procedure was identical to the linking study except that only one repetition of each statement was elicited. After a familiarization period the participants read the statements in random order off a PC screen addressing a student assistant. We analysed the second reading of each sentence.

Boundary	Place of S2 (relative to S1)		
	heterorganic fronter	backer	homorganic
word boundary	<i>met Pam</i> (18)	<i>keep calm</i> (18)	<i>black colour</i> (18)
syllable boundary	<i>output</i> (18)	<i>fruitcake</i> (18)	–

Table 8-3. An overview of stop-stop sequence types according to the relative place of articulation of the two stops and the boundary.

8.3.2 Analysis and results

Target sequences were analyzed both auditorily and visually using Praat (Boersma and Weenink, 2011). Each sequence was coded for the type of S1 release. Four types of releases were coded: no release, an inaudible

release that was visible in the spectrogram, a weakly audible release, and a strong release. The first two and the last two release types were combined for the purposes of the statistical analysis as ‘unreleased’ and ‘released’. In addition, to estimate the amount of gestural overlap in each sequence, we measured the interval between the beginning of S1 closure and the end of S2 closure (‘closure duration’) and the duration of the entire sentence, and we calculated the closure duration normalized for articulatory rate (closure duration \times number of syllables in the sentence / sentence duration). For each participant, we then calculated the percentage of unreleased first stops (% unreleased), mean rate-normalized closure duration across their sentences, and the mean articulatory rate in syllables per second.

A repeated-measures (RM) ANOVA on % unreleased with the between-subject factor Group (L, E, C) and the within-subject factors Place of S2 (fronter, backer) and Boundary (word, syllable) found a significant main effect of Group ($F[2, 13] = 9.83, p = .00251$). As shown in Fig. 8-4a (on the left), the English speakers unreleased the first stop about 3 quarters of the time while the Czech monolinguals only unreleased it about a quarter of the time. The learners were in-between and, as shown by a post-hoc Tukey HSD test, differed significantly from the English natives ($p < .05$) but not from the Czechs. The factor Boundary was not significant and no interaction between Boundary and Group was found. Like in the glottalization study, we observed individual variation between learners (see Fig. 8-5). This time, we do not see a steady climb but a split between learners who do release final stops and those who do not.

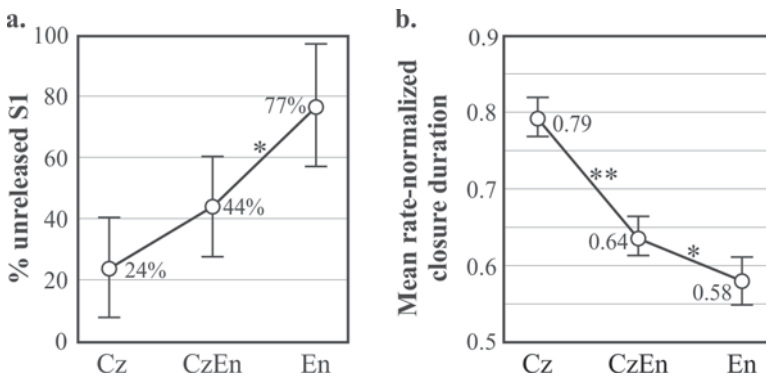


Fig. 8-4. **a.** Percentage of unreleased first stops by speaker group. **b.** mean rate-normalized S1+S2 closure duration by speaker group. (Cz = Czech, CzEn = Czech English, En = English)

Another RM ANOVA with the same factors found a significant effect of Group on the mean rate-normalized closure duration of S1+S2 ($F[2, 13] = 76.86, p < .001$). Fig. 8-4b shows that the durations of S1+S2 closures were the shortest for the English, the longest for the Czech monolinguals, and in-between for the learners who, as shown by a post-hoc Tukey test, differed significantly from both the Czechs ($p < .001$) and the English speakers ($p < .05$).

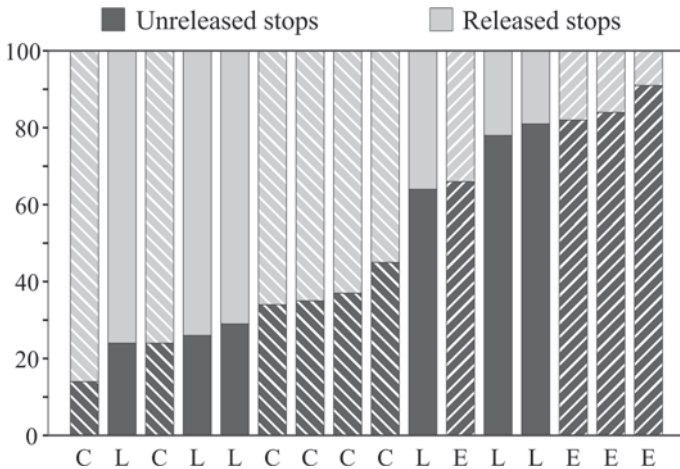


Fig. 8-5. Individual percentages of unreleased and released first stops in Czech native speech (C), English native speech (E), and English speech produced by Czech learners (L).

A RM ANOVA on % unreleased only in word-boundary sequences with Place of S2 (homorganic, frontier, backer) as the within-subject and Group as the between-subject factor showed that Place of S2 had a significant main effect ($F[2, 26] = 29.53, p < .001$) and it did not interact with Group. A post-hoc Tukey test showed that the % unreleased in homorganic sequences (78%) was higher ($p < .005$) than in sequences with a frontier S2 (60%), which in turn was higher ($p < .005$) than in sequences with a backer S2 (37%).

Finally, we examined articulatory rate. A one-way ANOVA found an effect of Group on speakers' mean articulatory rate ($F[2, 13] = 5.89, p = .015$). A post-hoc Tukey test revealed that the learners were slower than the Czechs ($p < .05$), which was the only significant difference. The correlation between the 6 learners' mean articulatory rates and their % unreleased was significant ($r \sim .95, p < .01$) only after excluding one

outlier (i.e. 16.66% of the data!). We also used *t*-tests to assess articulatory rate within each Czech EFL learner. For only a third of the learners was there a significant difference in articulatory rate between sentences they produced with unreleased first stops and those produced with a release (even with $\alpha = .10$).

8.3.3 Discussion

In terms of the percentage of their unreleased first stops, our learners were intermediate between the monolingual speakers of either language. Interestingly, the percentage of unreleasing seemed to be inversely related to the combined duration of both stop closures, indicating that a greater gestural overlap coincides with unreleasing.

What explains the difference between the learners and the native English speakers in unreleasing? First, the place of articulation of the second stop affected unreleasing for all three speaker groups in the same way, possibly reflecting a linguistic universal, and the non-native *vs.* native difference thus cannot be ascribed to the influence of this factor. Second, it was not the case that our learners would release word-final first stops significantly more often than first stops within a compound, and so the interlanguage word integrity effect can provide no explanation for the difference between the learners and the native English speakers either. Finally, our learners did not have a significantly slower articulatory rate than the native English speakers, nor was there within the learners a clear relationship between articulatory rate and unreleasing. Therefore, having eliminated all other factors, we ascribe the lower frequency of unreleasing in the Czech learners' English compared with native English to transfer from L1 Czech.

8.4 Conclusion

This chapter dealt with the topic of word-boundary processes in speech of advanced learners of English as a foreign language. We view these processes as creating local fluency, i.e. smooth word-to-word transitions. Taken together, the two production studies reported in this chapter show that in our learners' continuous speech, word boundaries often remain phonetically distinct and words are not fused together at their edges as frequently as they are in native English.

While we attributed both excessive vowel glottalization and unreleasing to transfer from L1, we suspect there are other aspects of non-native speech which may interfere with achieving local fluency. In our second

study we examined the idea that in non-native speech articulatory synchronization of sounds belonging to different words is restricted (the word integrity effect). Our study did not lend support to this idea: articulatory synchronization in stop-stop sequences (expressed as the percentage of unreleased first stops) was not greater within words than across the word boundary. Another potential inhibitor of local fluency in non-native speech is the rate of articulation. Could it be that the difference in the frequency of smooth word-to-word transitions in the speech of native speakers and advanced EFL learners is due to a difference in the speed with which articulatory movements are executed? In another study with Czech EFL learners not reported here we observed a connection between articulatory rate and the amount of linking (Šimáčková, Kolářová and Podlipský, 2014): when the learners were induced to speak faster they produced fewer intonation breaks and thus fewer glottalizations at the boundaries of intonational phrases, and they also linked more within intonational phrases, increasing significantly the rate of resyllabification. However, in our current study of stop-stop sequences spanning word boundaries, unreleasing could not be straightforwardly linked to articulatory rate.

In both studies we saw that learners' productions of the targeted boundary phenomena showed a great deal of inter-speaker variation. It remains to be demonstrated whether the learners who are more 'locally fluent', i.e. have more linking, unreleased stops, or other word-boundary phenomena in their speech, are actually *perceived* as more fluent. Although we assume they are, methodologically, corroborating or refuting this assumption appears to be much more complex than testing the link between perceived fluency ratings and 'global fluency' measures (reflecting tempo and hesitations). The effect of boundary processes on perceived connectedness of speech is bound to be cumulative. In other words, only a persistent absence (or a consistent application) of more of these processes is likely to evoke perceived disconnectedness (or smoothness) in the flow of non-native speech. Even if we manage to prepare stimuli differing in the incidence of various local-fluency phenomena and also not differing in their global fluency, and we do find a difference in their perceived smoothness, we will still not have conclusive evidence that any one of the occurring boundary processes influenced perception. At the same time, it is very likely that not all boundary processes are equally perceptually salient and have an effect of equal strength. Testing the effect of a single boundary process on perceived fluency, besides posing even greater challenges for the preparation of

stimuli, involves the risk that the effect will be weak or even absent just because the synergy with other boundary processes has been removed.

References

- Anderson-Hsieh, J., Riney, T. & Koehler, K. (1994). Connected speech modifications in the English of Japanese ESL learners. *Issues and Developments in English and Applied Linguistics*, 7, pp. 31–52.
- Bissiri, M. P. & Volín, J. (2010). Prosodic structure as a predictor of glottal stops before wordinitial vowels in Czech English. In: Vích, R. (Ed.), *Proceedings of the 20th Czech-German Workshop on Speech Processing*, pp. 23–28. Prague: Academy of Science.
- Boersma, P. & D. Weenink. (2011). Praat: Doing Phonetics by Computer (Version 5.0.40). Retrieved on November 11, 2011 from <http://www.praat.org/>.
- Browman, C. P. & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49/3–4, pp. 155–180.
- Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24/2, pp. 209–244.
- Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition*, 22/1, pp. 1–26.
- Cruttenden, A. (2014). *Gimson's Pronunciation of English* (8th ed.). London: Routledge.
- BBC Learning English: Connected speech. (2009). Retrieved on 29 October, 2015, from <http://www.bbc.co.uk/worldservice/learningenglish/grammar/pron/features/connected.shtml>
- Davidson, L. (2011). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, 53/8, pp. 1042–1058.
- Dilley, L., Shattuck-Hufnagel, S. & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24/4, pp. 423–444.
- Freed, B.F. (1995). *Second Language Acquisition in a Study Abroad Context*. Amsterdam: John Benjamins Publishing Company.
- . (2000). Is fluency, like beauty, in the eyes of the beholder? In: H. Riggensbach, (Ed.), *Perspectives on Fluency*, pp. 243–265. Ann Arbor, Michigan: University of Michigan Press.
- Freed, B. F., Segalowitz, N. & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom,

- study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, 26/2, pp. 275–301.
- Garellek, M. (2012). Glottal stops before word-initial vowels in American English: distribution and acoustic characteristics. *UCLA Working Papers in Phonetics*, 110, pp. 1–23.
- Ginther, A., Dimova, S. & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, 27/3, pp. 379–399.
- Hardcastle, W. J. & Roach, P. (1977). An instrumental investigation of coarticulation in stop consonant sequences. *Work in Progress Phonetics Laboratory, University of Reading*, 1, pp. 27–44.
- Henderson, J. B. & Repp, B. H. (1982). Is a stop consonant released when followed by another stop consonant? *Phonetica*, 39/2–3, pp. 71–82.
- Hieke, A. E. (1984). Linking as a marker of fluent speech. *Language and Speech*, 27/4, pp. 343–354.
- . (1985). A componential approach to oral fluency evaluation. *The Modern Language Journal*, 69/2, pp. 135–142.
- Isaacs, T. & Trofimovich, P. (2011). Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech. *Applied Psycholinguistics*, 32/1, pp. 113–140.
- James, L. & Smith, O. (2012). *Get Rid of Your Accent: The English Pronunciation and Speech Training Manual*. London: Business & Technical Communication Services.
- Kormos, J. (2011). Speech production and the cognition hypothesis. In: Robinson, P. (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*, pp. 39–60. Amsterdam: John Benjamins.
- Kormos, J. & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32/2, pp. 145–164.
- Ladefoged, P. (1993). *A Course in Phonetics* (3rd ed.). Orlando: Harcourt Brace Jovanovich.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40/3, pp. 387–417.
- Mompeán-Gonzalez, J. & Mompeán-Guillamón, P. (2009). /r/-liaison in English: An empirical study. *Cognitive Linguistics*, 20/4, pp. 733–776.
- Mompeán, J. & Gómez, F. A. (2011). Hiatus resolution strategies in non-rhotic English: the case of /r/-liaison. In: *Proceedings of 17th ICPHS*, pp. 1414–1417.

- Mora, J. C. (2006). Age Effects on Oral Fluency Development. In: Muñoz, C. (Ed.), *Age and the Rate of Foreign Language Learning* (Vol. 19), pp. 65–88. Clevedon; Buffalo; Toronto; Sydney: Multilingual Matters.
- Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.
- Nguyen, T. & Ingram, J. (2004). A corpus-based analysis of transfer effects and connected speech processes in Vietnamese English. In: *Proceedings of 10th Australian International Conference on Speech Science & Technology*, pp. 516–521. Sydney.
- O'Connor, J. D. (1980). *Better English Pronunciation*. Cambridge: Cambridge University Press.
- Palková, Z. (1994). *Fonetika a fonologie češtiny*. [Phonetics and phonology of Czech]. Prague: Karolinum.
- Pennington, M. C. (1996). *Phonology in English Language Teaching: An International Approach*. London; New York: Longman.
- Pinget, A.-F., Bosker, H. R., Quené, H. & de Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, 31, pp. 349–365.
- Riazantseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition*, 23/4, pp. 497–526.
- Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes*, 14, pp. 423–441.
- Roach, P. (2009). *English Phonetics and Phonology: A Practical Course* (4th ed.). Cambridge: Cambridge University Press.
- Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, 65/3, pp. 395–412.
- Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. New York: Routledge.
- Šimáčková, Š., Kolářová, K. & Podlipský, V. J. (2014). Tempo and connectedness of Czech-accented English speech. *Proceedings of the International Symposium on the Acquisition of Second Language Speech, Concordia Working Papers in Applied Linguistics*, 5/1, pp. 667–677.
- Šimáčková, Š., Podlipský, V. J. & Kolářová, K. (2014). Linking versus glottalization: (Dis)connectedness of Czech-accented English. *Proceedings of the International Symposium on the Acquisition of Second Language Speech, Concordia Working Papers in Applied Linguistics*, 5/1, pp. 678–692.

- Šimáčková, Š., Podlipský, V. J. & Chládková, K. (2012). Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association*, 42/2, pp. 225–232.
- Šimek, J. (2010). *Plosives in Czech: temporal characteristics and variability in realization*. MA thesis, Charles University in Prague.
- Simon, E. (2010). Phonological transfer of voicing and devoicing rules: Evidence from L1 Dutch and L2 English conversational speech. *Language Sciences*, 32/1, pp. 63–86.
- Skarnitzl, R. (2004). Acoustic categories of nonmodal phonation in the context of the Czech conjunction “a”. In: Palková, Z. & Veroňková, J. (Eds.), *Acta Universitatis Carolinae, Philologica 1/2004, Phonetica Pragensia X*, pp. 57–68. Prague: Karolinum.
- Umeda, N. (1978). Occurrence of glottal stops in fluent speech. *The Journal of the Acoustical Society of America*, 64, pp. 88–94.
- van Gelderen, A. (1994). Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures: Oral performances of students aged 11–12 years. *Language Testing*, 11, pp. 291–319.
- Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N. & North, B. (2009). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Volín, J. (2010). Fonetika a fonologie. [Phonetics and phonology]. In: Cvrček, V. (Ed.), *Mluvnice současné češtiny [Contemporary Grammar of Czech]*, 1, pp. 35–64. Prague: Karolinum.
- . (2012). Jak se v Čechách “rázuje”. [Glottal-stopping in Bohemia]. *Naše řeč*, 95/1, pp. 51–54.
- Zampini, M. L. (1997). L2 Spanish spirantization, prosodic domains and interlanguage rules. *Language Acquisition and Language Disorders*, 16, pp. 209–234.
- Zsiga, E. C. (2003). Articulatory timing in a second language. *Studies in Second Language Acquisition*, 25/3, pp. 399–432.

CHAPTER NINE

WEAK STRUCTURAL WORDS IN BRITISH AND CZECH ENGLISH

JAN VOLÍN AND TEREZIE JOHANÍKOVÁ¹

9.1 The importance of weak-form words

Human speech comprises chains of units that can be most of the time assessed as mutually contrastive in prominence. A unit can be perceived as stronger or weaker (i.e., more salient or less salient) than the one that precedes or follows. For certain types of analyses, the binary division to strong and weak units can be considered sufficient, but there are languages that would not be described satisfactorily with just two levels of prominence. If we want to be more precise and reliable about English lexical prominence patterns and we subscribe to the four-level analysis, we agree to operate with the following inventory of categories: primary stress, secondary stress, full unstress and weak unstress (e.g., Umeda, 1975: 442; Wells, 1990 or Fear et al., 1995, albeit with various names for the categories). In connected speech, i.e., in spoken utterances, the syllables carrying lexical primary stress may also manifest various degrees of relative prominence (the strongest of which should be the nuclear accent), but that is not the concern of our present investigation. On the contrary, we will be focusing on the pronunciation of certain monosyllabic structural words in spoken texts that are typically reduced or weak, i.e., the words that exhibit the lowest degree of prominence in the metrical structure of an English utterance (e.g., Cruttenden, 2014; Roach, 2009; Underhill, 1994).

The duration of such monosyllabic words is relatively short and their nucleus usually consists of a schwa or a close-mid vowel. They normally occur without pitch accents and with low sound levels as if living in the

¹ This chapter was supported by the grant provided by the Czech Science Foundation, GAČR, under no. 14-08084S GAČR.

shadow of the surrounding auto-semantic words (Umeda, 1975; Volín & Weingartová, 2014). These items will be termed *weak-form words* (WFW) in the current chapter. (Although in certain phonological accounts they may also be somehow misleadingly referred to as *structural* words or *grammatical* words, our understanding is that *weak-form words* constitute only a small subset of a much larger set of structural or grammatical words.)

Pronunciation manuals sometimes differ in the counts of such words. There are scholars who list about thirty of them, while others might offer more than fifty. Importantly, though, they agree on the core of the set. The weak forms of articles *the* and *a/an* are presented even in elementary textbooks, with the recommendation to use strong forms only if the following word begins with a vowel. Although the issue is a bit more complicated than that, beginner learners of English are guided in the right direction. As to conjunctions, *and*, *but*, *than* and *as* are not doubted and neither are prepositions *at*, *for*, *from*, *of* and *to*. The subsets of pronouns and auxiliary plus modal verbs are larger and some of their members raise a question of the status with regard to standard pronunciation. For instance, *you* and *your* are undoubted as to their WFW status, and neither are *she*, *her*, *he*, *his*, *him*, while *I* and *my* are questionable since their behaviour in more careful speaking styles is less consistent.

Speakers whose mother tongue is not stress-timed English may find it difficult to acquire the habit of weakening. If that happens, their monosyllabic structural words stay quite prominent (perceptually strong) in the chain of words and may attract unnecessary attention. What seems to be a more serious consequence is the distortion of the natural flow of speech or, in other words, alteration of the typical rhythmic patterns. Recent research suggests that if words are not temporally grouped into certain predictable configurations, the mental processing of speech becomes more demanding. In addition, cumbersome rhythmic patterns can be stigmatizing, i.e., they trigger negative stereotypical reactions in the minds of some listeners (Buxton, 1983; Ghitza & Greenberg, 2009; Volín, Poesová & Skarnitzl, 2014).

The Czech language, one of the West Slavic languages of the Indo-European family, does not have reduced vowels and weakened or elided consonants in its standard pronunciation. This is probably one of the reasons why Czech English as a specific foreign accent of the internationally used language often lacks reductions (Volín & Skarnitzl, 2005; Volín & Skarnitzl, 2010; Weingartová, Poesová & Volín, 2014). Since the monosyllabic structural words are of high frequency in spoken texts, the effect of their unconventional pronunciation can be quite

profound. Therefore, it might be useful to investigate the scale of the differences in the produced degree of reduction of such words by native vs. non-native speakers. Furthermore, it might be of interest to see how coherent the *weak-form words* are as a class. In other words, do Czech speakers of English treat individual members of the weak-form word set in the same manner? Such knowledge could be used in a more realistic modelling of Czech-accented English, which, in turn, is necessary for experiments focused on various effects of the sound of Czech English on the listener.

In our current study we decided to compare readings of identical texts by speakers of Czech English and by their native English counterparts. Read texts are sometimes criticised for not having ecological validity, but our material comprises news bulletins, and we argue that these are written to be read aloud and they do have their communicative objectives – they should inform listeners about various events. Clearly, the speech of news readers is supposed to be intelligible to wide audiences if the radio broadcast is meant seriously. Thus, we cannot dismiss read-aloud news as ecologically invalid. Their limitations have to be admitted, however. The potential findings in read-aloud texts should not be too hastily generalized across other speech styles.

What findings can be expected? The null hypothesis (H_0) would assume no differences in the phonological and phonetic forms of the words under investigation. As already noted above, the first alternative that comes to the mind of an observant person (hypothesis H_{A1}) speaks for greater reduction of weak-form words in the speech of native speakers of English. Given the outcomes of some recent studies (Weingartová, Poesová & Volín, 2014; Volín & Skarnitzl, 2010), it is highly unlikely that Czech speakers reduce WFW more than the natives (hypothesis H_{A2}). Many partial hypotheses could be stipulated about the behaviour of individual words or personal habits of individual speakers. We will try to address those when analysing the data.

9.2 Capturing segmental and temporal properties of WFW

Recordings of twelve speakers (6 Czech + 6 British) were used in the current study. With one potential exception, they were young adults with either progressing or completed university education. The exception was a professional BBC World Service news reader, whose education and age we were unable to establish. Apart from this news reader (whose speech sample was acquired directly from the BBC radio broadcast), the speakers were recorded in a sound treated studio of the Institute of Phonetics in

Prague. They were first given a text of the news bulletin read by the professional reader and spent some time familiarizing themselves with the contents. Then they were asked to read out the text in front of a studio microphone at a comfortable pace and without attempts to imitate any mannerisms they might have observed in professional newsreaders of the time. They were also reassured that any dysfluencies could be re-recorded so there is no need to be nervous about the performance. In reality, though, minor deviations from the text were ignored to reduce the number of corrections required from the speakers. Only substantial stutterings, hesitations, or mispronunciations were remedied.

The news bulletin comprised seven paragraphs of text with three additional sentences signalling the name of the broadcaster, the news reader and the location of the radio station. Altogether, there were 462 words in 26 sentences, ideally (i.e., canonically) consisting of 795 syllables. Each reading took about 4 minutes and the speakers produced between 40 and 50 breath-groups.

Word	n-text	n-sample	n-phon	structure
a	11	142	1	V
and	7	84	3	VCC
from	7	73	4	CCVC
has	6	72	3	CVC
have	6	70	3	CVC
in	13	159	2	VC
of	11	131	2	VC
on	5	58	2	VC
the	41	483	2	CV
to	14	167	2	CV

Table 9-1. Weak-form words selected for analyses. They are listed alphabetically with their counts in the text (*n-text*), in the whole recorded material (*n-sample*), and the numbers of phonemes (*n-phon*) in their canonical phonological form (*structure*).

We decided to analyse only those WFWs that occurred at least five times in the text. Table 9-1 shows that there were ten such items. Comparison of their counts in the text and in all the recordings indicates, however, that not all the speakers produced the words according to the scripted version of the news bulletin. There were some omissions and, occasionally, even insertions of WFWs. Altogether, our sample comprised 1439 words to be inspected.

The speech recordings were processed in the Praat analysis software (Boersma & Weenink, 2014). First, the boundaries of individual phones were manually labelled by the authors of this chapter, and then the information concerning durations of words and their constituting segments was retrieved. At the same time, the facts concerning the segmental make-up of the words were noted.

9.3 Do foreign speakers of English reduce weak-form words?

A cursory look at our data immediately suggested that the H_{AI} hypothesis might be correct. The WFWs produced by the native speakers of English were on average by 45.7 ms shorter than those uttered by the Czech speakers (Fig. 9-1 on the left, $t(1437) = 15.42$; $p < 0.001$). To make sure that this highly significant result was not an artefact of different articulation rates (native command of a language usually allows for faster speech, i.e., shorter durations of words), we calculated the mean ratio of the native vs. Czech articulation rates. The native speakers as a group turned out to be 1.143 times faster. This number was used as a constant to normalize the raw data, and, unless stated otherwise, all the following reported results are based on the normalized values.

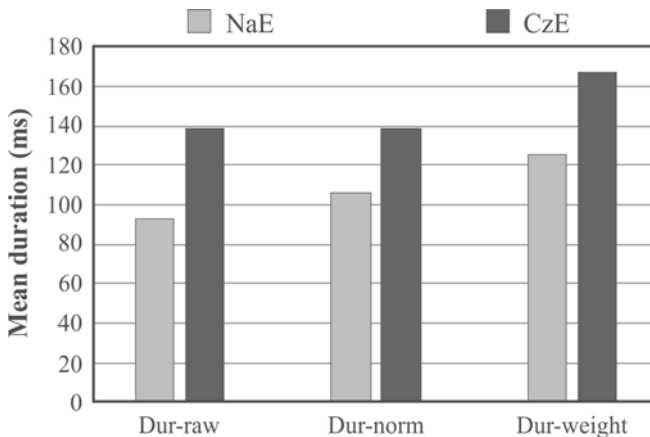


Fig. 9-1. Mean durations of all words. Native English is represented by light grey colour, Czech English is dark grey. *Dur-raw* are raw durations, *Dur-norm* are durations normalized for articulation rate, and the mean *Dur-weight* takes into account numbers of individual words.

After the normalization for unequal articulation rates, the duration of WFWs was still by 32.3 ms different (Fig. 9-1 in the middle, $t(1437) = 10.44$; $p < 0.001$). Yet even this result is somehow biased. First, our words were represented unevenly. For instance, there were more than 8 times as many definite articles *the* in the text than prepositions *on*. Second, the t-test on such a large number of items could be inflated. Therefore, we decided to represent each word only by its mean duration. The grand mean would then not be influenced by more frequent words (Fig. 9-1 on the right) and the degrees of freedom will drop to 9, which will make the calculation of statistical significance more rigorous. Even this stringent procedure produced a highly significant result in a t-test for correlated measures: $t(9) = 6.81$; $p < 0.001$. We can, therefore, conclude that Czech speakers of English produce weak-form words with longer relative durations. The difference is about 40 milliseconds even if we normalize articulation rates. Such difference is about 33.5% of the duration of the native English words, i.e., the Czech English WFWs are by about one third longer than the same words produced by native speakers of English.

This general result has to be further examined, however. It needs to be established whether the individual words behave analogically and each of them represents the group of selected weak-form words in a similar way.

Fig. 9-2 suggests that this is not the case: some differences are smaller, some are greater. Individual words do not display parallel behaviour. For instance, it can be observed that for the words *and*, *of* and *on*, that are second, seventh and eighth from the left the CzE durations (black bars) are more or less identical, while the NaE durations (grey bars) mutually differ and their pattern does not resemble that of the CzE. The mean duration of *on* is longer than *of* in the NaE sample while in the Czech sample it is slightly shorter.

To find out which of the differences are statistically significant, we calculated two-way ANOVA. It confirmed the interaction (meaning that the words do not display parallel behaviour): $F(9, 1419) = 11.54$; $p < 0.0001$. Post-hoc Tukey HSD test showed that the differences for the words *a* and *to* were only marginally significant ($p = 0.07$ and 0.1 respectively), for the article *the* it was significant ($p = 0.013$), and for the rest of the set they were highly significant ($p < 0.001$).

To understand this outcome better, we decided to analyse the words' segmental composition. The first set of analyses focused on the temporal features of vowels. Fig. 9-3 is analogous to Fig. 9-1 in design. It displays mean durations of vowels as they were retrieved (on the left), normalized for articulation rates differences (middle), and weighted for the unequal numbers of words in the text. The emerging pattern is clearly different

from what could be seen in Fig. 9-1. All three methods of measurements lead to the conclusion that vowels (or more precisely their durations) are not the major cause of longer weak-form words in Czech English.

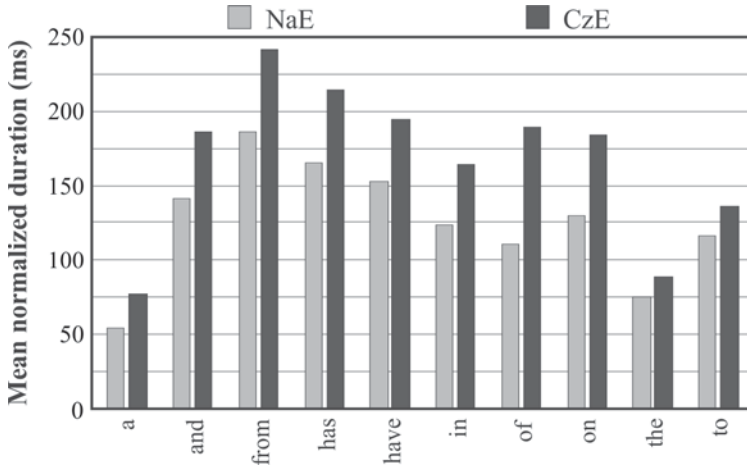


Fig. 9-2. Mean durations of individual words. Native English (NaE) is represented by light grey colour, Czech English (CzE) is dark grey. Durations normalized by articulation rate ratio.

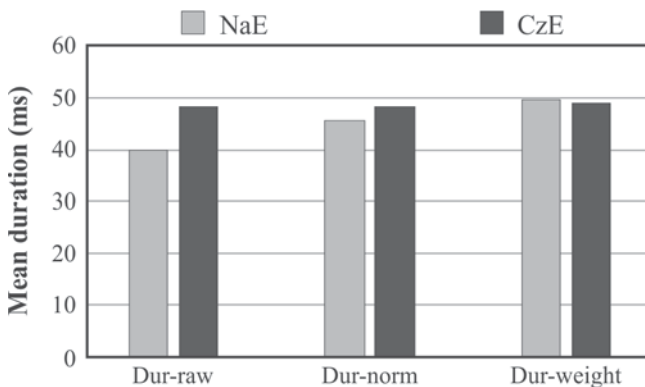


Fig. 9-3. Mean durations of vowels in WFVs. Native English is represented by light grey colour, Czech English is dark grey. *Dur-raw* are raw durations, *Dur-norm* are durations normalized for articulation rate, and the mean *Dur-weight* weighs the numbers of occurrences.

A vowel is a compulsory element in a monosyllabic English word, while consonants may be absent, present only in the coda or onset, or form clusters. Moreover, some consonants in WFWs can be elided. Therefore, the words in our sample had to be analysed separately in this respect.

The indefinite article does not possess a consonantal phoneme, but it is possible to strengthen it with an initial glottal stop, which, by its physiological nature, is a consonantal manoeuvre. Table 9-2 shows that NaE speakers seldom choose this option, while CzE speakers glottalize quite often. The Chi-square test returned an indication of a highly significant result: $\chi^2(1) = 38.46; p < 0.001$.

'a'	Glott.	Non-glott.
NaE	7	63
CzE	43	29

Table 9-2. Numbers of occurrences of initial glottalization of the word *a* by native (NaE) and Czech (CzE) speakers of English.

If we stipulate a hypothesis about the duration of the pure, modally phonated vowel in this word, the t-test finds the difference between NaE and CzE insignificant. However, the inclusion of creaky parts and other manifestations of glottal stop leads to a highly significant difference: $t(140) = 4.64; p < 0.001$. The native speakers make their *a* by 22 milliseconds shorter (that is 40% of its duration) than Czech speakers. (Calculated on normalized values, otherwise the difference is even greater.)

The conjunction *and* exhibited certain similarities with the preceding word. As displayed in Table 9-3, the Czech speakers pre-glottalized the initial vowel more often than NaE speakers. Although the effect is smaller than in the case of the word *a* (probably due to the frequent position of the conjunction at the beginning of a new prosodic unit), it is still highly significant: $\chi^2(1) = 12.32; p < 0.001$. Analogously, the CzE speakers produced the full coda (without /d/ elision) more often, yet this effect did not reach statistical significance ($p = 0.16$).

'and'	Glott.	Non-glott.	Coda/nd/	/d/ Elision
NaE	20	22	11	31
CzE	36	6	17	25

Table 9-3. Structural features of the word *and* as produced by native (NaE) and Czech (CzE) speakers of English. *Glott.* = word initial glottalization, *Coda /nd/* = no segmental elision, */d/ Elision* = the last segment deleted.

What was significant, though, was the difference in the duration of the consonantal coda and actually even the duration of /n/ itself: $t(82) = 3.91$; $p < 0.001$ and $t(82) = 2.93$; $p < 0.01$ respectively. Like in the indefinite article *a*, there is no significant difference in the duration of the vowel in *and*, but the CzE speakers still pronounce the word longer, partly due to the more frequent insertion of the pre-vocalic glottal stop, partly due to longer consonants and the lack of final /d/ elision.

The preposition *from* with its CCVC form was the longest weak-form word in our set. As presented above, its mean duration NaE speech was 183.5 ms, while in CzE it was 241.4 ms (Fig. 9-2). This difference is statistically highly significant: $t(71) = 4.66$; $p < 0.001$. Both the consonantal onset and vocalic nucleus contributed to this result: $t(71) = 4.32$; $p < 0.001$ and $t(71) = 3.13$; $p < 0.01$ respectively. Contrary to that, the nasal /m/ in the coda was virtually identical in both investigated accents of English. With regard to the onset, it was not only the duration of the initial fricative (76.3 ms against 94.0 ms) but also the fact that native speakers very often elided the post-initial /r/. Table 9-4 provides the numbers of occurrences of /r/ elisions (or ‘r-dropping’ in popular terminology).

‘from’	Present	Elided
NaE	6	31
CzE	30	6

Table 9-4. Elisions of post-initial /r/ in the word *from* by native (NaE) and Czech (CzE) speakers of English.

This almost symmetrical opposition in the behaviour of post-initial /r/ is highly significant: $\chi^2(1) = 32.88$; $p < 0.001$.

Auxiliary verbs *have* and *has* are both of CVC structure, however, both are prone to losing their initial laryngeal fricative and possibly even the vowel in the process called contraction (*there has been* → *there’s been*). It should be noted that no contractions were graphically suggested in our texts, yet some of the speakers chose to instantiate them nonetheless. Table 9-5 displays the numbers of contractions and elisions of /h/ for both forms of the verb. Due to the small numbers of occurrences no statistical significance was calculated.

One of the *have* tokens in the text was actually not a genuine auxiliary verb. It occurred in the phrase *have no confidence*, which in traditional grammar lends it a status of full verb (with the meaning *to possess*). One could argue, however, that in certain predictable contexts a frequent verb might be weakened and *have no confidence* could be such a predictable

collocation. Therefore, we did not exclude the word straight away and we wanted to know whether its ambiguous status really influences its phonetic form. The answer was positive: the mean duration of the true auxiliary *have* was 165.1 ms while the autosemantic *have* took 216.2 ms on average. Both NaE and CzE speakers contributed to this difference, although the NaE with greater weight. Twelve autosemantic instances of *have* (one by each speaker) were excluded from the subsequent analyses.

	'has'		'have'	
	Contraction	h-elision	Contraction	h-elision
NaE	4	2	1	11
CzE	1	3	0	0

Table 9-5. Reductions of the words *has* and *have* by native (NaE) and Czech (CzE) speakers of English. *Contraction* means elision of both the onset and the vocalic nucleus, *h-elision* concerns only the syllabic onset.

A two-way ANOVA was calculated to examine the durational difference between *has* and *have* in NaE and CzE accents. The main effect of accent was highly significant: $F(1, 126) = 38.34$; $p < 0.0001$, while the main effect of the word form was significant at the level of $\alpha = 0.01$: $F(1, 126) = 9.3$; $p < 0.01$. The word form *has* was longer by more than 20 milliseconds (despite its shorter spelling – the silent ‘e’ in *have* is an anachronism from today’s point of view). Fig. 9-4 also captures the absence of interaction between the two factors.

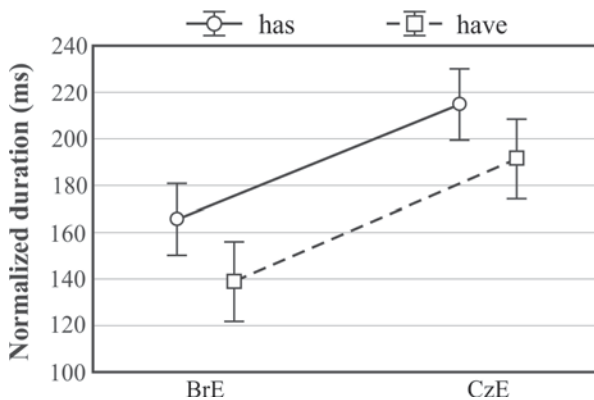


Fig. 9-4. Mean durations in milliseconds of the word forms *has* and *have* as produced by native (NaE) and Czech (CzE) speakers of English.

As to individual segments, the vowel contributed to the difference between NaE and CzE by mere 2 milliseconds – statistically an entirely insignificant outcome. Similarly, the onset /h/, if actually pronounced, is not significantly different in the two accents. Yet, it is more often elided in NaE. Finally, the duration of the coda is by 26 ms longer in CzE (both word forms contributed to this result equally). This result is highly statistically significant: $F(1, 126) = 46.03$; $p < 0.0001$. In summary, the durational difference between Czech and native productions of *has* and *have* are caused mainly by more frequent elisions of the consonant in the onset and shorter consonants in the coda in NaE.

The prepositions *in*, *of* and *on* are of the same word class and they all consist of a vowel followed by a consonant. Therefore, we will report about them jointly. The Czech speakers of English tended to provide these words with an onset in the form of glottalization (glottal stops or creaks) – see Table 9-6. The differences in the number of pre-glottalized initial vowels were all statistically highly significant (Fisher Exact Test had to be used on the preposition *on* due to a lower number of observations).

	'in'		'of'		'on'	
	Glott.	Non-glott	Glott.	Non-glott	Glott.	Non-glott
NaE	26	56	10	55	7	23
CzE	70	7	57	9	26	2

Table 9-6. Initial glottalization of the words *in*, *of*, *on* by native (NaE) and Czech (CzE) speakers of English.

All three prepositions were by tens of milliseconds longer in CzE than in NaE even after normalization for different articulation rates. Curiously, in the preposition *of* there were only three elisions of the coda in NaE and there was no elision in CzE. This lack of /v/ elisions was quite probably caused by the formal character of the read text and special significance of the recording session.

The differences in the durations of the prepositions were not caused by the vowels. On the contrary, if the vowels were measured as modally phonated segments, they were longer in NaE. This outcome is caused by the fact that pre-vocalic glottalization somehow leaks into the vowels. The Czech speakers, who regularly glottalized, seemingly produced shorter vowels since their creaky parts were measured as belonging to glottalization, i.e., onsets. But all three prepositions had shorter codas NaE: *in* by 14, *of* by 37 and *on* by 22 ms on average (for all of these $p < 0.001$). In summary, the highly significant differences in durations of the

three prepositions were caused by glottal onsets (arguably consonantal by nature) and longer codas.

The definite article *the* was the most frequent item in our sample, but at the same time it was the second shortest word (after the indefinite article *a* – see Fig. 9-2 above, which is in line with the findings of Umeda, 1975: 443). The difference in duration between NaE and CzE was only 13 ms, but because of the large number of tokens and relatively low variance it was ascertained as highly significant: $t(481) = 5.19$; $p < 0.001$. Unlike in the previous words, this difference is attributable almost entirely to the vowel: the native English vowels were by 11 ms shorter ($t(481) = 6.76$; $p < 0.001$), while consonants differed from the CzE ones by mere 2 ms ($t(481) = 1.42$; non-significant).

Preposition *to* belongs to a different word class than *the*, but its phonological structure CV is identical, and, more importantly, the phonetic manifestations of its status in our set of weak-form words were quite similar to the definite article. The CzE cases were on average by 19 ms longer, which is a significant result: $t(165) = 3.11$; $p < 0.01$. The vowel contributed with 15 ms ($t(165) = 4.72$; $p < 0.001$), while the consonantal onset with 4 ms ($t(165) = 0.66$; non-significant).

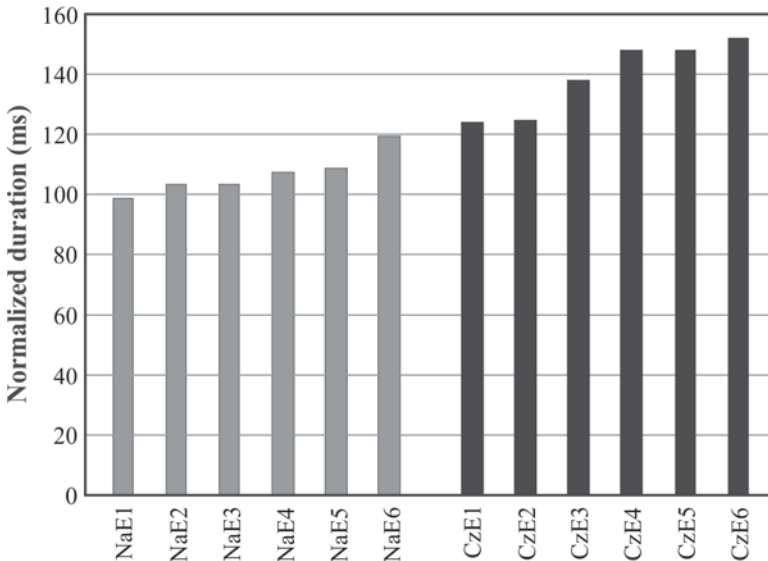


Fig. 9-5. Mean durations of weak-form words by individual speakers. The values were normalized for articulation rate differences.

As a final point it should be noted that individual speakers influenced the results in their own specific way. Fig. 9-5 reveals that there is no sharp division between native and Czech speakers of English. (Even though in the raw data there is a more conspicuous step between NaE6 and CzE1.) Individual speaking habits lead some Czech speakers of English to converge with native patterns. Contrary to that, the speaker NaE6 produces WFWs notably longer than the rest of the NaE sample. The generalization of our results then should be, as any broader view of this sort, cautious, and individuals must not be confused with arithmetic means of samples.

9.4 Discussion

In a language that a user acquires with comfortable mastery, many words become quite common, hence predictable, and articulatory effort does not need to be wasted on them. Complementarily to that, acquisition of perceptual skills makes listeners accustomed to a clearer pronunciation of less predictable lexical items and to blurry acoustic shapes of the easily anticipated units. One of the effects of reasonable reductions in articulatory effort is a perceived smooth flow of speech in stretches of satisfactory size. Ideally, the “flows” of both production and perception processes are intertwined and, as a consequence, the communication activities are effortless.

A learner of a foreign language, on the other hand, is not at ease with either producing or predicting words – he or she may struggle with each of them, and rather than thinking of a flow we might imagine search for individual bricks to be put one next to the other. This metaphor could indicate why we found significant differences in durations in all ten weak-form words in our sample. Czech speakers take the WFWs as individual regular bricks to be put next to some other bricks rather than as inconspicuous filling that holds the other bricks together.

An interesting and not entirely anticipated outcome of our analyses touched upon vowels. Although vowels are said to be the main carriers of prosodic information, our results invite certain attention in this respect. In seven out of ten words there were no significant differences between the two accents in the duration of vowels. At the same time, in eight out of ten words highly significant differences in the presence or duration of consonants were found. No simple description of this can be provided, however. For instance, nasal codas in prepositions *in* and *on* were significantly different between NaE and CzE, whereas nasal [m] in *from* was not.

Another purely theoretical question mark hangs over the pre-vocalic glottalization. Some word-initial vowels were preceded by glottal gestures that are constrictory by nature, hence they were classified as consonantal onsets. This happened undisputedly more often in CzE in all five words beginning with a vowel (*a*, *and*, *in*, *of*, *on*). Yet, glottalization also took the form of creaky phonation now and again. If that were classified as a secondary feature of the vowel, we would have to moderate our statement concerning the influence of vowels versus consonants on the durational difference. Be that as it may, though, the indefinite article *a*, for instance, is still by 40% longer in CzE than in NaE due to the presence of glottalization, regardless its phonetic status.

CzE weak-form words are on average by one-third longer than their NaE counterparts. The contribution of individual words is not equal, though. The greatest difference was found for the preposition *of*: it was by 78 ms longer in CzE than in NaE (i.e., by 71%). Contrary to that, the definite article *the* was only 14 ms longer in CzE. This difference was, however, quite consistent, so it would still be useful to know whether there is a noteworthy perceptual impact of this. Hopefully, future perceptual testing will be inspired and informed by our findings.

One final reservation has to be mentioned. The native speakers of English as a group spoke by about 14% faster than the Czech speakers. Therefore, all the raw durational values were normalized. We realize that normalizing all durations by one constant can be awkward since slower articulation rates do not mean linear lengthening of all segments. Nevertheless, this simple method was used to avoid false discovery caused by slower foreign accented speech. Modelling its temporal structure in detail is beyond the capacity of current phonetics.

References

- Boersma, P. & Weenink, D. (2014). *Praat: Doing Phonetics by Computer* (Version 5.4). Retrieved on October 14, 2014, from <http://www.praat.org>.
- Buxton, H. (1983). Temporal predictability in the perception of English speech. In: Cutler, A. & Ladd, D. R. (Eds.), *Prosody: Models and Measurements*, pp. 111–121. Berlin: Springer-Verlag.
- Cruttenden, A. (2014). *Gimson's Pronunciation of English* (8th ed.). London: Routledge.
- Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of Acoustical Society of America*, 97(3), pp. 1893–1904.

- Ghitza, O. & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66(1–2), pp. 113–126.
- Roach, P. (2009). *English Phonetics and Phonology: A Practical Course* (4th ed.). Cambridge: Cambridge University Press.
- Umeda, N. (1975) Vowel duration in American English. *Journal of the Acoustical Society of America*, 58/2, pp. 434–445.
- Underhill, A. (1994). *Sound Foundations*. London: Heinemann.
- Volín, J., Poesová, K. & Skarnitzl, R. (2014). The impact of rhythmic distortions of speech on personality assessment. *Research in Language*, 12/3, pp. 209–216.
- Volín, J. & Skarnitzl, R. (2005). Vybrané dynamické a melodické vlastnosti spojky ‘a’ v češtině a v angličtině [Selected dynamic and melodic characteristics of the conjunction ‘and’ in Czech and English]. In: Pořízka, P. & Polách, V. P. (Eds.), *Jazyky v kontaktu / jazyky v konfliktu a evropský jazykový prostor*. pp. 347–352. Olomouc: UPOL.
- Volín, J. & Skarnitzl, R. (2010). The strength of foreign accent in Czech English under adverse listening conditions. *Speech Communication*, 52, pp. 1010–1021.
- Volín, J. & Weingartová, L. (2014). Acoustic correlates of word stress as a cue to accent strength. *Research in Language*, 12/2, pp. 175–183.
- Weingartová, L., Poesová, K. & Volín, J. (2014). Prominence contrasts in Czech English as a predictor of learner’s proficiency. In: *Proceedings of Speech Prosody 2014*. Dublin: ProSIG & Trinity College.
- Wells, J. C. (1990). *Longman Pronunciation Dictionary*. Harlow: Longman.

CHAPTER TEN

ACOUSTIC CORRELATES OF TEMPORAL STRUCTURE IN NORTH-VIETNAMESE ENGLISH

ONDŘEJ SLÓWIK AND JAN VOLÍN¹

10.1 General information on Vietnamese

The Vietnamese language is spoken by approximately 92 million people in Vietnam. Over 1 million speakers of Vietnamese live in the USA. Other substantial Vietnamese communities are located in Australia (mainly people who fled from South Vietnam after 1975), France (descendants of those who came during the colonial era or after), and the former Eastern Bloc (inhabitants of North Vietnam coming during and after the Vietnamese-American war).

The Vietnamese language was originally spoken only by the Việt ethnic group settled in the Red River delta. Nowadays, the Việt people make up 88% of the overall Vietnamese population, and the remaining 12% comprise 53 ethnic minorities, who principally cherish their own culture and language despite the official status of Vietnamese.

It is a largely monosyllabic tonal language with six lexical tones in the standard dialect. All syllables tend to retain their tones at all times although the individual tonal realizations can differ from the canonical contours due to the effect of tonal coarticulation or the syllable's grammatical function. Vietnamese morphology is relatively simple: there is no declension or conjugation; temporality is expressed mainly by adverbials of time. Word classes overlap to a great extent and word order is crucial for successful comprehension of sentences.

In terms of orthography, Vietnamese uses *Chữ Quốc Ngữ* (National script) based on the Latin script with added diacritical signs introduced by

¹ The second author was supported by the grant provided by the Czech Science Foundation, GAČR, under no. 14-08084S.

the Portuguese colonizers to capture tonal differences and vowel quantity and quality.

It is estimated that about 60% of the Vietnamese vocabulary can be traced to Chinese (Trần Trí Dõi, 2011) but the languages are not genealogically related. Very often there are two words for a concept, one of which is considered purely Vietnamese and the other Sino-Vietnamese. Sino-Vietnamese words tend to be used in higher registers (literature, rituals, scientific terminology), whereas purely Vietnamese words are reserved for everyday conversation. Because of the colonial past, French borrowings are common especially in the field of technology and science (*ô tô* for *car*, *ni tơ* for *nitrogen*).

In our research, we examine the Hanoian dialect as described by Kirby (2011) because it has been considered the standard dialect since the reunification of Vietnam in 1976.

10.2 Stress in Vietnamese

Although a cursory look suggests that Vietnamese employs only monosyllabic words, it should not be thought of as a monosyllabic language. It possesses substantial numbers of di- and trisyllabic lexemes where the last element of the lexeme is usually stressed (Nguyễn & Ingram, 2007), e.g., *sinh viên* – student; *bưu điện* – post office; *người bán hàng* – seller. If a lexeme comprises two elements of equal importance, both elements are stressed equally (Nguyễn Văn Phúc, 2006), e.g., *bố mẹ* – parents (father mother); *vợ chồng* – married couple (wife husband). *Vợ* and *chồng* can stand alone, whereas in the case of *sinh viên* (student) the two syllables cannot stand separately and can only be combined with other syllables, hence the stress on the second constituent.

Thompson (1965) distinguished three degrees of sentential stress: a) *heavy* – for the morphemes which carry the heaviest load of new information, b) *weak* – for the morphemes with the lowest information conveying load (already mentioned or predictable), and c) *medium* – all remaining morphemes.

Cao Xuân Hạo (1998) points out that sentential stress is used to mark individual syntactic groupings, as it would often be difficult to establish which elements in the sentence belong together due to the lack of overt word-class markers or inflection. In the sentence: *Lan//đi mua cá// mí lì khế// về nấu canh* (*Lan went to buy fish as well as star fruit, then she returned home to make soup*), it is apparent that the stress tends to be placed phrase-finally. Sentence stress in Vietnamese is not linked to certain grammatical units (pronouns, prepositions, classifiers) but rather to

items that occur phrase finally (vocatives, temporal elements, intensifiers). In short phrases, both constituents can be stressed, e.g., *người cao* – tall people (or *people are tall*); *chó chạy* – dogs run.

Shiering, Bickel and Hildebrandt (2010) suggest that “a genuine stress domain [of Vietnamese] is preferably disyllabic and maximally trisyllabic. Within this domain, stress is realised on the final syllable in the default case. Crucially, this domain is computed irrespective of the morphosyntactic status of its constituent syllables, i.e. stress phonology does not distinguish between a word-level and a phrasal-level of prosodic structure. Metrically, polysyllabic words are thus indistinguishable from other combinations of syllables. Since the most complex structures which are referenced by the rules for iambic rhythm are phrasal, stress may most adequately be attributed to the prosodic domain of the Phonological Phrase.” This claim is also supported by the results of Nguyễn Thu Anh’s experiment (2014) investigating acoustic correlates of rhythmic structure of Vietnamese narrative speech.

Contrary to Chinese, Vietnamese does not contain atonal syllables. The apparent loss of tone in some Chinese syllables originates from the fact that they are unstressed. Unstressed syllables in Vietnamese tend to be shorter and the tone contour is far from canonical, but the tone is still present, even if less explicit.

10.3 Vietnamese English

The Vietnamese are generally pragmatic and flexible so teaching English as a foreign language is in high demand. Vietnam is open to foreign investments and command of English from B1 up can secure the user a significantly higher social and economic status. Despite high a level of proficiency in English ‘passive’ skills (i.e., listening and reading), many Vietnamese speakers encounter great difficulty in making themselves understood when speaking to both native speakers of English and non-native speakers from other linguistic backgrounds. These difficulties present themselves on both segmental and suprasegmental levels.

As vowel quantity is not a prominent factor in discrimination of Vietnamese vowels, it is difficult for the Vietnamese speakers to differentiate between long and short English vowels especially if combined with the effects of fortis and lenis codas.

There are no consonant clusters in Vietnamese, which leads to their simplification in English, specifically in syllabic codas because they carry a light semantic load in Vietnamese. Such simplification “... is clearly a problem when speaking English which places a heavy semantic load on

the coda in verb forms and other suffixes. Consonant clusters are generally simplified in Vietnamese-accented English to a degree that is not compatible with intelligibility” (Cunningham, 2009).

All final Vietnamese plosives [p, t, k] are glottally-reinforced [ʔp̚, ʔt̚, ʔk̚] (Singer, 2012), which causes the release to be very weak or even completely absent. Moreover, Vietnamese syllables ending with plosives can only carry the rising tone *sắc* or the glottalized falling tone *ngặng*. These two features lead to a very weak release of final plosives in English (or their possible complete omission) and to the fact that the preceding vowel shows either a rising pitch or strong glottalization. The abrupt change of pitch (especially in case of the rising tone *sắc*) renders the vowel perceptually shorter but at the same time more prominent, which can be confused for word-stress.

All vowel-initial words in Vietnamese begin with a glottal stop [ʔ], regardless of whether or not they are following a pause (Singer, 2012). This feature is present in casual everyday conversation as well as in careful news announcements. Vietnamese speakers therefore usually place glottal stops in front of word-initial English vowels and their speech sounds rather disconnected.

Singer (2012) further claims: “The six tones make Vietnamese sound very musical, if staccato. This too may carry over into speaking English. Vietnamese English tends to be very broken up, with short intonation phrases, an abundance of stresses, recognizably Vietnamese tones, and a great many glottal stops. Another difference between Vietnamese and English is that Vietnamese is a syllable-timed language, whereas English is a stress-timed language. (...) It is very hard for speakers of syllable-timed languages to get the hang of stress-timing, so this feature will persist, strengthening the staccato rhythm of Vietnamese English.”

10.4 Measurements of temporal organization in Vietnamese English

The aim of this study is to calculate the durational correlates of the rhythm class for English spoken by North Vietnamese speakers (inhabitants of Hanoi) according to Ramus, Nespore and Mehler (1999), Low, Grabe and Nolan (2000) or Grabe and Low (2002), and Dellwo and Wagner (2003) or Wagner and Dellwo (2004). The outcome will be compared with values extracted for RP British English and the Vietnamese language spoken in Hanoi.

Global temporal metrics of RP Standard have been calculated in numerous studies, and the results have always shown a clear tendency

towards stress-timing. To our best knowledge, calculations for Vietnamese English and Vietnamese itself have not been carried out yet. Vietnamese scholars claim Vietnamese to be a syllable-timed language but they do not support their claim with empirical evidence (Cao Xuân Hạo, 1998; Nguyễn Văn Phúc, 2005). Their claim might be advocated with reference to a relatively simple structure of Vietnamese syllables. However, the prosodic manifestation of metrical structure (the appearance of lengthened and shortened syllabic nuclei) might alter the surface rhythmic impression. We wanted to know whether the global temporal metrics can capture the differences between Vietnamese (Vtm) and Vietnamese English (ViE) and relate them in a meaningful way to the Southern British Standard (BrE).

The benchmark measurements for the Southern British Standard were extracted from eight recordings of BBC World News bulletins. Four men and four women, all professional news readers, read bulletins of seven paragraphs comprising about 500 words. The recordings were made in years 2001 and 2002, and are part of the sound database of the Prague Institute of Phonetics. Eight young adult speakers of Vietnamese English (4 male + 4 female) were asked to familiarize themselves with the texts of the new bulletins and subsequently read them out in front of a recording device. They were all university educated inhabitants of Hanoi, the capital city, with the proficiency level of B2. Recordings of the Vietnamese language (Hanoi dialect) were also made to gain insight into the possible influence of the temporal characteristics of Vietnamese on Vietnamese English. All the recordings were segmented and manually labelled in Praat (Boersma & Weenink, 2014). Altogether over 40,000 phone boundaries were processed.

The best-known seven metrics were used out of the available choice: normalized Pairwise Variability Indices for vocalic and consonantal intervals (nPVI-V and nPVI-C respectively), percentage of vowel durations in the utterance duration (%V), standard deviations of consonantal and vocalic interval duration (ΔC and ΔV) and their normalized versions, i.e., coefficients of variation (Varco-C and Varco-V). The first two measures were suggested by Low, Grabe and Nolan (2000), the next three by Ramus, Nespor and Mehler (1999) and the normalizing improvement was suggested by Dellwo and Wagner (2003). However, for the purpose of comparison of data with other studies it is necessary to note that the pairwise variability indices were calculated with a modified formula (after Gibbon & Gut, 2001):

$$PVI = 100 \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{d_k + d_{k+1}} \right| / (m - 1)$$

What is different from the original formula of Low, Grabe and Nolan's (2000) is the absence of division of the sum of two neighbouring intervals ($d_k + d_{k+1}$). This does not change the patterns found in results but makes the calculation simpler and the range of resulting numbers easier to grasp mentally.

Another less common procedure that we used was to calculate each metric three times. Once for the speech material as it was, then again but with the last word in each intonation phrase excluded, and finally with the exclusion of the phrase-initial stress-groups, highly infrequent words and foreign names. The first degree of data cleansing (exclusion of phrase-final words) was done in order to eliminate the influence of phrase-final lengthening on durations of segments. The phrase-final lengthening increases (sometimes dramatically) the variability of durations but does not really contribute to the rhythm type of a language. The second degree of data pruning was to eliminate foreign or highly unusual phonotactic patterns and possible hesitations or dysfluencies, together with a potential brief increase in articulation rate phrase-initially. In the following figures and paragraph these three conditions will be labelled *Raw*, *Prune-1* and *Prune-2*, respectively. The individual metrics will be presented one by one in order to make the outcomes of the measurements more transparent. Vocalic metrics will be presented before the consonantal ones.

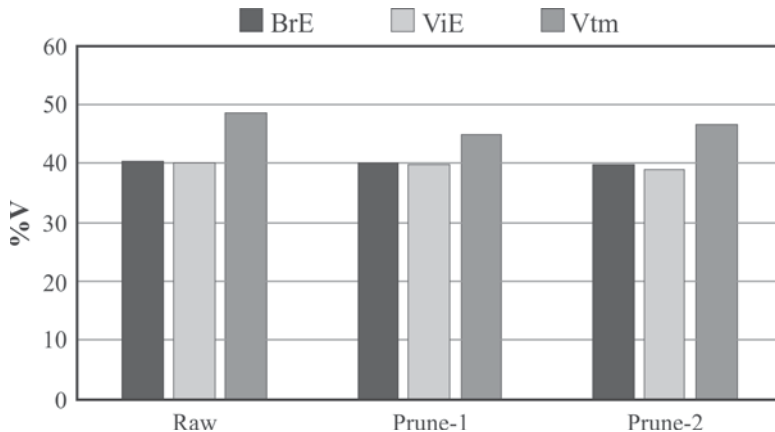


Fig. 10-1. Mean values of the percentage of vowel durations (%V) in the utterance under the conditions Raw, Prune-1 and Prune-2 (see text). Black columns represent native English (BrE), light grey columns Vietnamese English (ViE), and dark grey columns Vietnamese (Vtm).

Fig. 10-1 presents the resulting values for the metric traditionally labelled as %V, i.e., the percentage of vowel durations in the utterance duration. Given that only vowels and consonants are distinguished, it provides information about the ratio of these two phone classes in connected speech. Clearly, Vietnamese is more vocalic than English. Although the rhythm of Vietnamese English is perceptually different from Native English, the metric %V does not reflect the impression. One-way ANOVA with Tukey HSD post-hoc test confirmed a significant difference of Vietnamese against the other two types in all three conditions. For Prune-2: $F(2, 21) = 41.8; p < 0.001$.

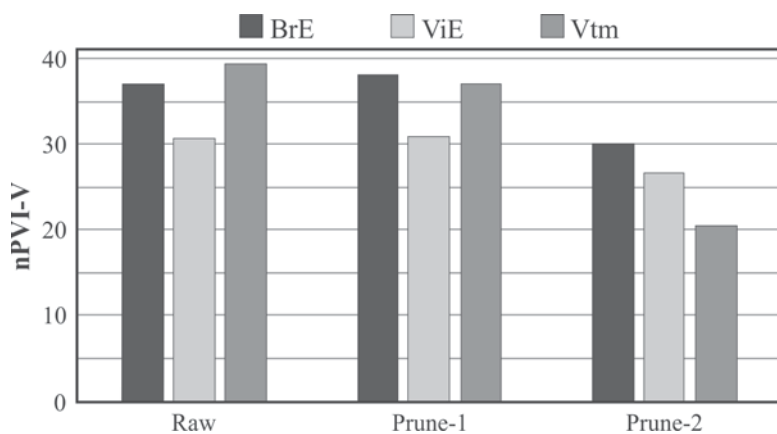


Fig. 10-2. Mean values of the normalized Pairwise Variability Index of vocalic intervals (nPVI-V) under the conditions Raw, Prune-1 and Prune-2 (see text), for native English (BrE), Vietnamese English (ViE), and Vietnamese (Vtm).

The situation is more dynamic in Fig. 10-2 which captures measurements of the normalized Pairwise Variability Index (nPVI). This metric has no units and expresses an overall degree of disparity between neighbouring elements – here the durations of vocalic intervals. The resulting values vary between about 20 and 40. The raw data suggest that vocalic variation is the highest in Vietnamese, lower in English and lowest in Vietnamese English. The result is similar after the exclusion of phrase-final words. However, if the material is cleansed of all phrase-initial, phrase-final and other temporally dubious parts (i.e., Prune-2), English is portrayed as most vocally variable, Vietnamese least and Vietnamese English in between. One-way ANOVA for Prune-2: $F(2, 21) = 53.8; p < 0.001$ and post-hoc Tukey HSD found all three types different from each other with high

significance. The results for n-PVI-V indicate that local changes in tempo can affect the global metrics quite substantially – the high values for Vietnamese under Raw and Prune-1 conditions are attributable to phrase-final lengthening, phrase initial acceleration and, possibly, the pronunciation of foreign names and unusual lexical items.

A similar caveat occurs if we compare Figures 10-3 and 10-4. The metric ΔV (Fig. 10-3) is the standard deviation from the mean duration of vocalic intervals and Varco-V is the same measure normalized by the mean. As a consequence, ΔV is in milliseconds while Varco-V is a ratio without units. More importantly, though, both measures suggest different rhythmic patterns in the comparison of English, Vietnamese and Vietnamese English. The metric Varco-V should be more indicative since it is normalized for different articulation rates. Native speech is generally known to be faster than the use of a non-native language. Therefore, the information that the duration of English vowels is more variable than that of Vietnamese and Vietnamese English is more useful than the picture provided by Fig. 10-3 where Vietnamese English appears as most variable, but this is an artefact of the slowest speech tempo there. One-way ANOVA for Varco-V under Prune-2: $F(2, 21) = 30.2$; $p < 0.001$. Post-hoc Tukey HSD found all three types different from each other, but BrE from ViE with only marginal significance ($p = 0.073$), while the other differences with significance on the level of $\alpha = 0.001$.

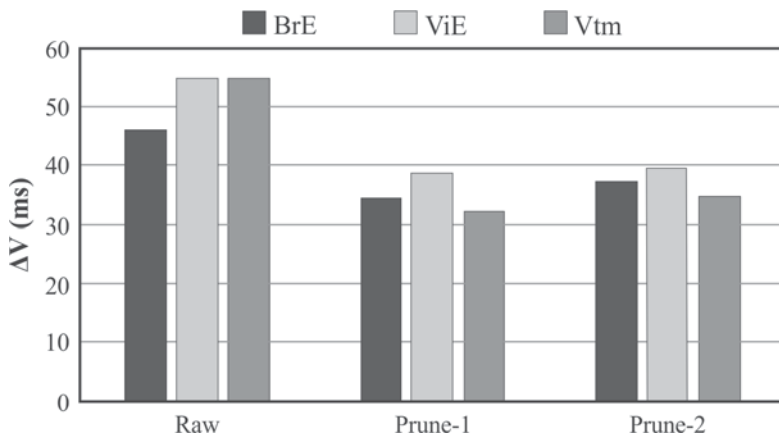


Fig. 10-3. Mean values of the standard deviation of durations of vocalic intervals (ΔV) under the conditions Raw, Prune-1 and Prune-2, for native English (BrE), Vietnamese English (ViE), and Vietnamese (Vtm).

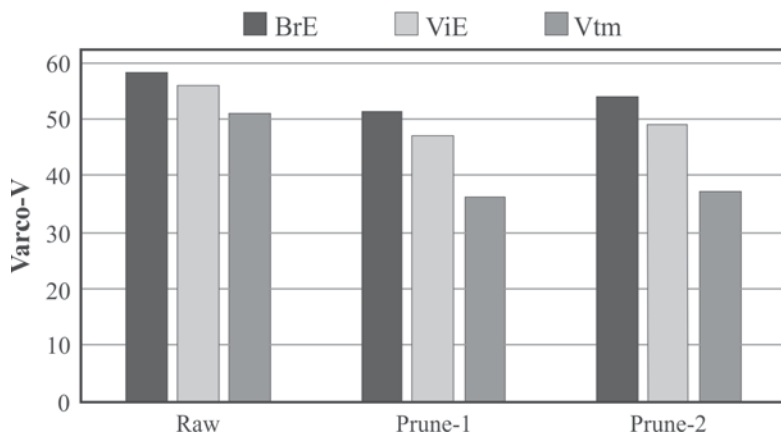


Fig. 10-4. Mean values of the variation coefficient of vocalic interval durations (Varco-V) under the conditions Raw, Prune-1 and Prune-2, for native English (BrE), Vietnamese English (ViE), and Vietnamese (Vtm).

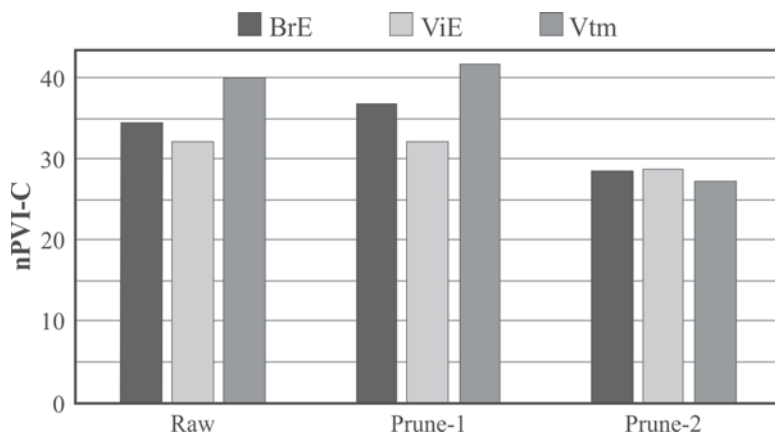


Fig. 10-5. Mean values of the normalized Pairwise Variability Index of consonantal intervals (n-PVI-C) under the conditions Raw, Prune-1 and Prune-2, for native English (BrE), Vietnamese English (ViE), and Vietnamese (Vtm).

Consonantal n-PVI is displayed in Fig. 10-5. In raw data and after first pruning (Prune-1) Vietnamese consonants seem to vary most in their durations. However, after the elimination of items with the local changes in tempo it turns out that all three types of material exhibit about the same pairwise consonantal variation ($F(2, 21) = 0.95$; $p > 0.39$). Otherwise, the

range of values is similar to that of the vocalic n-PVI (see Fig. 10-2 above). The apparent high variation of Vietnamese consonants is clearly attributable to the phrase-final and phrase-initial segments.

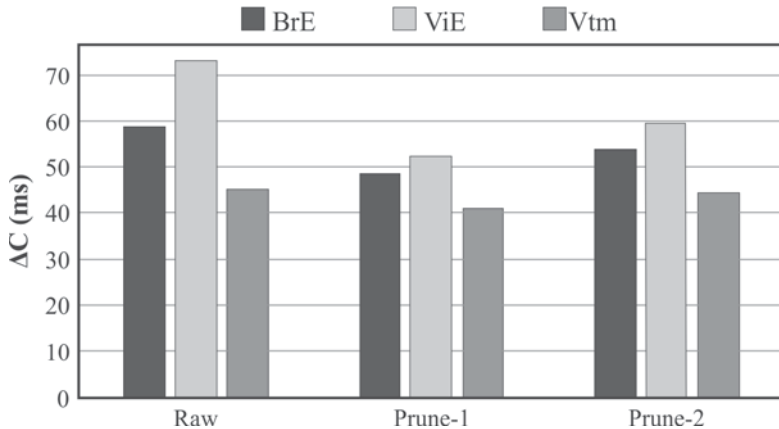


Fig. 10-6. Mean values of the standard deviation of durations of consonantal intervals (ΔC) under the conditions Raw, Prune-1 and Prune-2, for native English (BrE), Vietnamese English (ViE), and Vietnamese (Vtm).

The comparison of Figures 10-6 and 10-7 leads to the conclusion analogical to the one found in Figures 10-3 and 10-4. Standard deviations (here ΔC) are markedly affected by the overall tempo. That explains why the relatively slow Vietnamese English reached the highest values. If normalized for tempo (Varco-C in Fig. 10-7), this non-native accent falls between the mother tongue and the target language of the speakers. This is, however, not the case in the raw data where the influence of phrase-final lengthening of consonants by Vietnamese speakers of English is still obvious.

In terms of statistical significance, Vietnamese is different from the other two types under all three conditions with high significance, while ViE is different from BrE only marginally under the Prune-1 condition ($p < 0.1$). It could be argued that this marginal significance is only due to the very low number of degrees of freedom. One-way ANOVA for Prune-2: $F(2, 21) = 10.3$; $p < 0.001$.

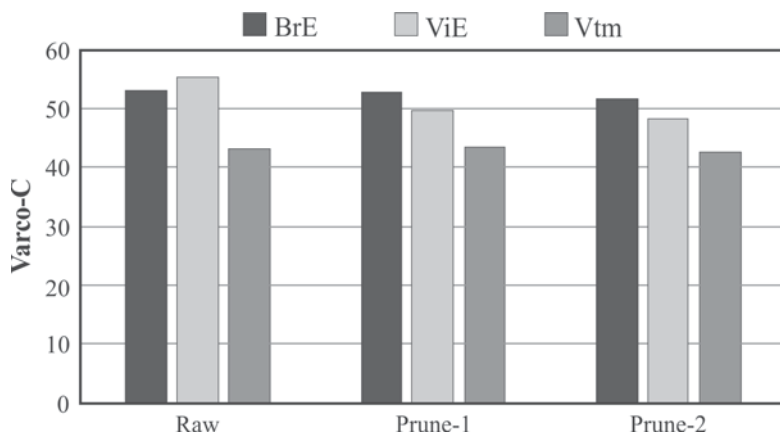


Fig. 10-7. Mean values of the variation coefficient of consonantal interval durations (Varco-C) under the conditions Raw, Prune-1 and Prune-2, for native English (BrE), Vietnamese English (ViE), and Vietnamese (Vtm).

10.5 Interpretation of the results

Despite the general belief that pairwise variability indices and standards deviations of durations are “rhythm metrics”, we suggest that these and similar measures only capture a certain component of temporal structure, to which speech rhythm is somehow linked. Although the occasional correlation of these measures with human intuitions about rhythm classes is encouraging, researchers are far from offering a credible model of speech rhythm perception.

The fact that we measured each metric three times in progressively pruned material provided evidence about the influence of changes in tempo at the edges of prosodic phrases (see also Volín & Skarnitzl, 2007). If musicians play with changes in tempo, i.e., slowing down and speeding up, for instance, in order to make a musical piece less machine-like or more dramatic, they do not change the rhythmic values of individual notes. Similarly, acceleration at the beginning of a prosodic phrase and deceleration at its end does not change the rhythm, since rhythm builds on relative values. Global metrics, even if mathematically normalized, do not reflect perceptual normalization.

In about 40,000 consonants and vowels that were carefully measured, there were two persistent problems that must be mentioned for the correct utilization of our results. First, it was the status of the high front glide in Vietnamese. Traditionally scripted as ‘j’, the glide in post-vocalic position

mostly contributed to the impression of a closing diphthong. Since the English closing diphthongs were perceptually equivalent, we treated the glide as part of a vocalic interval – we would consider that inconsistent to interpret the same sound as a diphthong in English and a vowel followed by consonant in Vietnamese. Second, the glottalized tones *ngã* and *nặng* did not have a uniform effect on the rhythm in all their renderings.

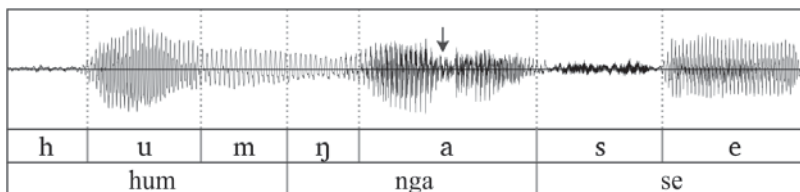


Fig. 10-8. An example of the glottalized falling tone over the syllable *ngã*. Glottal event marked with a black arrow above the soundwave.

In some cases, the glottalization was moderate and did not disrupt the impression of one rhythmic pulse. In other cases (slower tempo and glottal stop), the syllable was clearly divided into two rhythmic pulses (Fig. 10-8). Thus, the syllables with the glottalized tones *ngã* and *nặng* had to be treated *ad hoc*.

Despite these two controversies we believe our results are reliable and can be used for cross-language or cross-accent comparisons and for further modifications of approaches to the so-called “rhythm metrics”.

With regard to vowels, %V seems to be most tightly linked to the phonotactics of the language and, therefore, does not indicate any difference between native and Vietnamese-accented English. On the other hand, Varco-V and n-PVI-V of pruned data (Prune-2 condition) point in one common direction: native English exhibits the greatest variation in vocalic durations, native Vietnamese the smallest, and Vietnamese English falls in between.

The same pattern is provided by the consonantal metric Varco-C under the conditions Prune-1 and Prune-2. Other metrics seem to be affected by global or local variation in articulation rates, while n-PVI-C (Fig. 10-5) in pruned data does not indicate any difference between the three types of speech material. Although there are no consonant clusters in Vietnamese, the durations of individual consonants in our recordings are far from regular.

In conclusion, it can be stated that for the given type of material, i.e., the read monologue, global temporal metrics indicate stronger syllable-timing in Vietnamese and stronger stress-timing in English, while

Vietnamese English displays the properties of a typical interlanguage. This pattern is, however, not apparent in the raw material or through non-normalized standard deviations.

References

- Boersma, P. & Weenink, D. (2014). Praat: Doing Phonetics by Computer (Version 5.4). Retrieved on October 14, 2014, from <http://www.praat.org>.
- Cao Xuân Hạo (1998). Trọng âm và các quan hệ ngữ pháp trong tiếng Việt (Role of Stress in Vietnamese Grammar) in *Tiếng Việt mấy vấn đề ngữ âm ngữ pháp ngữ nghĩa (Vietnamese – Selected Problems in Phonetics, Grammar and Semantics)*. Hanoi: NXB Giáo Dục. 137–157.
- Cunningham, U. (2009). Phonetic correlates of unintelligibility in Vietnamese-accented English. *Proceedings, Fonetik 2009*. University of Stockholm.
- Dellwo, V. & Wagner, P. (2003). Relations between language rhythm and speech rate. In: *Proceedings of 15th ICPHS*, 471–474. Barcelona: UAB & IPA.
- Gibbon, D. & Gut, U. (2001). Measuring speech rhythm. In: *Proceedings of Eurospeech*, 91–94. Aalborg: ISCA.
- Grabe, E. & Low E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In: Gussenhoven, C. & Warner, N. (Eds.), *Papers in Laboratory Phonology 7*, 515–546. Berlin: Mouton de Gruyter.
- Kirby, J. P. (2011). Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association*, 41(3), 381–392.
- Low, E. L., Grabe, E. & Nolan, F. (2000). Quantitative characterisations of speech rhythm: Syllable timing in Singapore English. *Language and Speech*, 43, 377–401.
- Nguyễn Văn Phúc (2005). *Ngữ âm tiếng Việt thực hành (Practical Phonetics of Vietnamese)*. Hanoi: NXB Đại Học Quốc Gia Hà Nội.
- Nguyễn, Anh-Thư T. & Ingram, J. C. L. (2007). Stress and tone Sandhi in Vietnamese Reduplications. *Mon-Khmer Studies*, 37, 15–40.
- Ramus, F., Nespors, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292.
- Schiering, R., Bickel, B. & Hildebrandt, K. (2010). The prosodic word is not universal, but emergent. *Journal of Linguistics*, 46, 657–709.
- Singer, E. (2012). *Vietnamese English*. University of York. Canada.

- Thompson, L. C. (1965). *A Vietnamese Grammar*. Seattle: University of Washington Press.
- Trần Trí Dõi (2011). *Giáo trình lịch sử tiếng Việt* (Textbook to the History of the Vietnamese Language). Hà Nội: NXB giáo dục Việt Nam.
- Volín, J. & Skarnitzl, R. (2007). Temporal downtrends in Czech read speech. In: Proceedings of 8th Interspeech, 442–445. Antwerpen: ISCA.
- Wagner, P. & Dellwo, V. (2004). Introducing YARD and re-introducing isochrony to rhythm research. In: *Proceedings of Speech Prosody 2004*. Nara: ProSIG.

CHAPTER ELEVEN

PATTERNS OF ARTICULATION RATE IN ENGLISH / FRENCH TANDEM INTERACTIONS

BARBARA KÜHNERT
AND TANJA KOCJANČIČ ANTOLÍK¹

11.1 Introduction

The aim of this chapter is to explore patterns of articulation rate in spontaneous interactions of English and French tandem pairs. Articulation rate is a prosodic feature of the temporal organization of speech and represents the pace at which segments are produced. It is commonly defined as a measure of speaking rate in which all pauses are excluded from the speakers' phrases, i.e. speaker-specific ways of conveying information, such as pausing, hesitations or fillers, are disregarded.

Early studies hypothesized that articulation rate was relatively constant and invariant within and across speakers (e.g. Goldman-Eisler, 1961). Subsequent research, though, showed that articulation rate may vary with the mode of delivery (speaking *vs.* reading rate; Crystal & House, 1990), across different languages (Grosjean & Deschamps, 1972, for a comparison between English and French), across different regional variants of the same language (Jacewicz, Fox & Wei, 2010), and may be gender- and/or age-specific (Whiteside, 1996; Avanzi, Dubosson & Schwab, 2012).

11.1.1 Articulation rate in L2 learners

In research of second or foreign language (henceforth L2) acquisition, articulation rate, as one parameter of speech tempo, is often considered to

¹ This work was partially funded by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

be associated with notions of L2 proficiency and intelligibility (e.g. Munro & Derwing, 2001; Trouvain & Möbius, 2014). Essentially, the increasing mastery of an L2 “results in a faster overall articulation rate, a process that might be described as ‘becoming more fluent’” (Gut, 2012: 90).

For example, L2 learners have often been described as having a slower articulation rate in their second than in their native language, such as Korean, Italian or German learners of English (Guion, Flege, Liu & Yeni-Komshian, 2000; Gut, 2009), or German learners of French (Raupach, 1980). As for the difference in articulation rate between native and non-native speakers, the results in the literature are somewhat more mixed. On the one hand, it has been observed that L1 speakers articulate faster in a given language than L2 speakers. By way of illustration, German learners of English have been reported to be significantly slower than native speakers of British English (Gut, 2009), Spanish learners of French to be slower than native French speakers from Switzerland (Barquero Armesto, 2012). On the other hand, there is evidence that the articulation rate of native and non-native speakers might, in fact, be similar. For instance, Swiss-German learners of French did not show a slower articulation rate than native Swiss-French speakers, regardless whether the learners were from Neuchâtel (Schwab, Dubosson & Avanzi, 2012) or Zurich and Bern (Avanzi, Dubosson & Schwab, 2012).

11.1.2 Accommodation of articulation rate

Another notion of interest in the context of L2 acquisition and articulation rate is the notion of accommodation – also called adaptation, alignment or entrainment. In the broadest sense of the term, phonetic accommodation refers to the modifications made by the speakers in response to the situation and the listener (Coupland & Giles, 1988). Initial studies advocated that speech accommodation is the consequence of either automatic or social processes. However, more current research suggests that a combination of both processes may provide a more comprehensive account (for an overview, see Babel, 2009).

The speech addressed to non-native interlocutors, also called ‘foreigner-directed speech’ or ‘foreigner talk’, is often characterized as being “delivered at a slower pace” (Ellis, 1997:45) or produced with a “slow speech rate” and “long pauses” (Gass & Selinker, 2013:341). To date, however, few studies have looked at the acoustic properties of foreigner-directed speech and reports vary as to whether speakers apply indeed slower articulation rates. Although Biersack, Kempe & Knapton (2005) found that participants lowered their speech rate when addressing

foreign interlocutors, they seemed to do so primarily by lengthening the duration of the pauses in their utterances, whereas no significant speech rate correlates of foreigner talk occurred in a study by Uther, Knoll & Burnham (2007). Scarborough and colleagues (2007), who examined foreigner-directed speech to an imagined and a real interlocutor, reported that speakers adjusted conversational tempo in both cases but that they talked even more slowly in conditions involving imaginary interlocutors.

While studies on foreigner talk look at accommodation from a global perspective, studies on ‘phonetic convergence’ assess accommodation from a dynamic point of view, i.e. try to evaluate to what extent speakers become more similar (or dissimilar) during the course of an interaction (for an overview, see Pardo, 2013). With regard to temporal measures and articulation rate, research once more yielded relatively inconsistent and variable outcomes. For instance, in interactions between native speakers, Street (1982) confirmed rate convergence between talkers participating in an interview, in contrast to a more recent study in which talkers involved in a map task showed no cross-correlation of articulation rate (Pardo, Cajori Jay & Krauss, 2010). Likewise, in mixed interactions between L1 and L2 speakers, instances of both convergence and divergence have been reported. Based on similarity judgements by an independent group of listeners, Kim, Horton and Bradlow (2011) concluded that talker-listener adjustments vary as a function of interlocutor language distance. Speakers who are closer in linguistic distance are more likely to converge than speakers who are further apart. Specifically, the authors found that greater convergence occurred in conversations between talkers of two different dialects of American English than in conversations between native speakers of American English and Korean L2 learners of English. Rao (2013), however, only partially confirmed this hypothesis in a subsequent acoustic analysis. Interactions between native speakers of American English resulted in convergence whereas both patterns, convergence and divergence, were observed in inter-dialectal interactions between American and Indian English talkers and in mixed language interactions between American English talkers and Spanish learners of English.

11.1.3 Goals of the current study

Tandem interactions between two individual native speakers of different L1 backgrounds are deemed useful to shed more light on some of the inconsistent findings reported in the literature. Tandem learning is based on the principle of reciprocity whereby pairs of native speakers aim to learn each other’s language and both learners contribute equally in terms

of the respective L1/L2 input (Penman, 2002). Thus, unlike in previous studies which are based on comparisons of different speaker groups, one using their native language, another their L2, the analysis of tandem conversations permits an evaluation of articulation rate of the *same* speakers, talking about the *same* topics in the *same* setting in two different languages. Specifically, we examine the following questions: (i) how do native speakers of American English and Standard French adjust their L1 articulation rate in conversations with L2 learners; (ii) how does the articulation rate in their L1 compare to the articulation rate in their L2 speech, and, (iii) can we observe any convergence of articulation rate between the native and non-native speakers in the course of a tandem conversation?

11.2 Method

Using part of the SITAF tandem corpus (Horgues & Scheuer, 2014), we analysed the spontaneous speech of five speaker pairs, with one member of the pair being a native French speaker, the other member a native speaker of American English (henceforth GA).

11.2.1 Participants and data collection

The French speakers were all female first or second-year undergraduate students of English at the University of Paris 3. The American speakers, four female and one male, were exchange students at the same university from different parts of the United States. This implies that we did not control for possible dialectal differences in articulation rate (henceforth AR) for a given language. The ages of the French speakers ranged from 17 to 21 years with an average of 9.2 years of learning English; the ages of the American speakers ranged from 19 to 20 years with an average of 6.4 years of learning French. All participants were naive to the purpose of the study.

The conversations were recorded digitally in a soundproof room at the University of Paris 3. Each speaker participated in three conversational interactions: an L1-L1 conversation with another speaker of the same native language, henceforth called *control* conversation; and two L1-L2 conversations with the tandem partner (English and French), in other words in the tandem interactions each speaker was once using his/her L1 and once using his/her L2. Each conversation, in turn, was composed of two communicative activities which ensured that the two participants were involved in a reciprocal dialogue.

11.2.2 Data analysis

The durations of the control conversations ranged from 5:10 to 12:31 mins (mean 7:06 mins), the tandem conversation ranged from 4:48 to 10:20 mins (mean 7:10 mins). The resulting corpus contained approximately 4 hours of speech. All conversations were first orthographically transcribed, including the annotation of speech turns, using the Transcriber toolkit (Barras, Geoffrois, Wu & Liberman, 2001).

The transcriptions of the English conversations were then semi-automatically aligned with the Munich automatic segmentation system MAuS (Schiel, Draxler & Harrington, 2011) and alignments were manually verified and corrected by inspecting waveforms and spectrograms using PRAAT (Boersma & Weenink, 2014). The French conversations were manually annotated by an experienced transcriber with the PRAAT software.

Excluding samples produced with background noise, laughter, overlapping speech or obvious disfluencies, such as false starts or hesitations, we subsequently extracted all articulation phases demarcated by two pauses. Following the study on articulation rate by Jacewicz, Fox and Wei (2010) we only considered inter-pausal stretches between 3 and 20 syllables. Pause durations were identified as silent intervals longer than 100 ms, except in cases of adjacent stop clusters in which the combined closure duration sometimes exceeded the 100 ms threshold. Stretches of less than 3 syllables were discarded as they mainly contained short backchannels, such as *yes, right, you know* or *oui, bon, c'est vrai*; stretches of more than 20 syllables were discarded as they were few in numbers and unequally distributed between speakers. About 4% of the overall number of stretches produced during the conversations were eliminated in this way, and the final data set included 1475 inter-pausal intervals from all subjects.

After measuring the onset and offset of each phrase, articulation rate was measured in syllables per second, which was calculated by dividing the duration of each phrase by the number of syllables as produced, i.e. the actual phonetically realised syllables. In a study examining the optimal linguistic unit to reflect temporal variance in speech, Trouvain et al. (2010) demonstrated that the 'realised syllable' is a good candidate to monitor AR and is one of the most frequently applied measures (for an overview, see Fletcher, 2013).

Fig. 11-1 shows the distribution of the raw data of AR during the tandem conversations which will be further analysed below. It suffices here to note that the majority of the stretches contained 12 syllables or less and that the productions of the L2 learners were not restricted to stretches

of only a few syllables. As can be seen, both French and GA speakers realised a variety of stretches with low and high syllable count in their L1 as well as their L2.

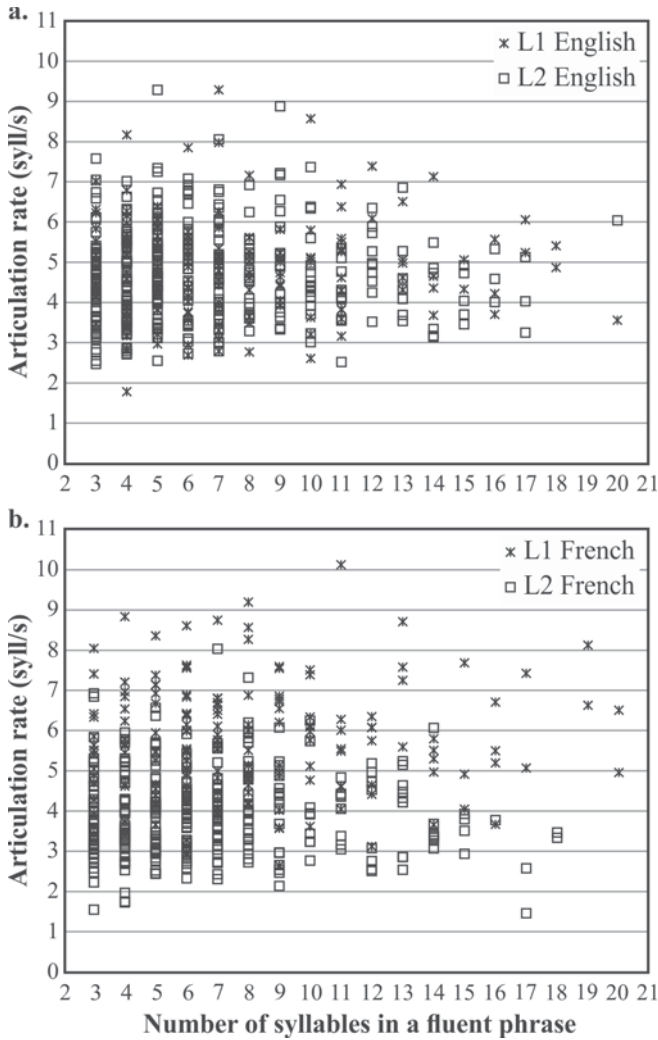


Fig. 11-1. Raw values of articulation rate (syll/s) per number of syllables in a phrase. **a.** English tandem sessions; **b.** French tandem sessions.

Statistical significance of the results was tested by applying a linear mixed-effects model (lme4 package by Douglas, Maechler, Bolker & Walker, 2015) in the R statistical computing software (R Core Team, 2012). Differences in AR were tested separately for three data subsets: (i) control conversation, (ii) conversations only in L1, and (iii) L1-L2 conversations with the tandem partner. AR was entered as a dependent variable and speakers as random effects in all models, while fixed effects varied across the three data subsets and were entered as follows: (i) L1 language (English or French), (ii) L1 language (English or French), session (control or tandem session in L1), interaction of factors L1 language and session, and (iii) language (English or French), language status (language spoken as L1 or as L2) and interaction of factors language and language status. In order to obtain p-values, a likelihood ratio test was used to compare the full model with a specified fixed factor to the model without it.

11.3 Results

11.3.1 Global patterns of articulation rate

A summary of the mean AR for each language and type of conversation is given in Fig. 11-2.

Looking at the control conversations first, the production of the French speakers showed a higher average AR than the productions of the GA speakers (6.78 syll/s vs. 4.88 syll/s). The linear mixed effects model revealed a significant effect of *Language* ($\chi^2(1) = 12.855, p < 0.001$). While a language-specific comparison of AR is not of central interest in the present study, the results are consistent with previous studies evaluating AR in different languages and in which AR values reached regularly higher values in French than in English (e.g. Dellwo & Wagner, 2003; Pellegrino, Coupé & Marsico, 2011). However, this outcome in itself is not surprising as it is well-known that AR can be influenced by a number of language-specific factors, such as syllable complexity, consonant phonotactics or reduction processes (for an overview, see Schiering, 2007). As a consequence, French with its simpler syllable structure and less complex consonant clusters is produced faster than English.

What is more interesting is the observation that only the French speakers showed a prominent overall adjustment in their L1 AR when addressing L2 learners. In comparison to the conversations with the native interlocutors, all French speakers slowed down significantly in their

measured AR from an overall mean of 6.78 syll/s to 5.75 syll/s. No such accommodation in L1 AR to the L2 speech of the French learners could be observed on the part of the GA speakers during the English conversations. Quite coincidentally, the GA talkers' average AR during the control sessions and during the English tandem sessions was identical (4.88 syll/s). The statistical analysis confirmed that the factor *Language* is a significant predictor of AR in the L1 condition ($\chi^2(1) = 9.439, p < 0.01$). In addition, there is a significant impact of *Session* ($\chi^2(1) = 25.201, p < 0.001$) and, most importantly, a significant *Interaction* between *Language* and *Session* ($\chi^2(1) = 23.718, p < 0.001$).

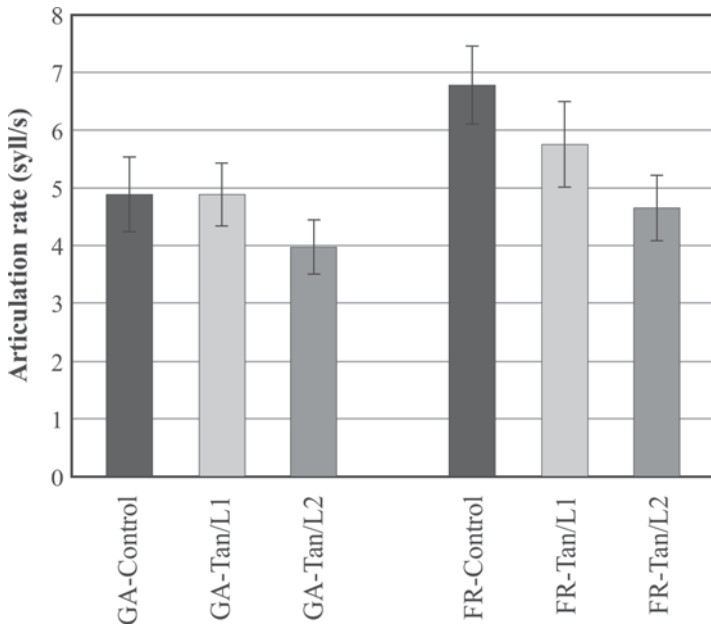


Fig. 11-2. Means and sd. of AR (syll/s) for, from left to right, GA speakers in control conversations (GA-Control), in tandem conversations speaking English (GA-Tan/L1) and in tandem conversations speaking French (GA-Tan/L2), followed by French speakers in control conversations (FR-Control), in tandem conversations speaking French (FR-Tan/L1) and in tandem conversations speaking English (FR-Tan/L2).

Finally, we tested whether AR differed when the speakers used their L1 and when they used their L2 during the tandem sessions. Both speaker groups were faster in their native than their second language. Nevertheless,

as can be seen in Fig. 11-2 when comparing FR/L2 with GA/L1, the French talkers were not significantly slower than the native GA partners when speaking English (4.65 syll/s versus 4.88 syll/s). The GA participants, on the other hand, had significantly lower AR values than the corresponding French talkers when speaking French (3.97 syll/s versus 5.75 syll/s) – even though, as outlined above, the latter already considerably slowed down their average AR during the French tandem sessions in order to adjust to the L2 learners. The statistical analysis revealed no significant main effect of *Language* ($\chi^2(1) = 0.373, p = 0.54$), but a significant effect of *Language Status* (being spoken as L1 or L2) ($\chi^2(1) = 169.09, p < 0.001$) as well as a significant *Interaction* between *Language* and *Language Status* ($\chi^2(1) = 6.388, p < 0.05$).

11.3.2 Dynamic patterns of articulation rate

In order to test for possible accommodation of AR between speakers during a single tandem session and to evaluate whether speakers increased in similarity as the conversations progressed, we followed a design proposed by Levitan and Hirschberg (2011). As mentioned above, each tandem session consisted of two conversational activities. We therefore divided each session into the two activities (corresponding roughly to the first and second half of the conversation), calculated the average AR for both parts and then compared the differences in AR between the first and second part for individual speaker pairs. We assumed that convergence in AR occurred if the similarity between speaker averages was closer in the second part than their similarity in the first part of the tandem session.

Table 11-1 gives a summary of the results. Overall, there is a general tendency of the speakers, albeit to different degrees, to become more similar to their partners, i.e. differences between the partners' AR were smaller in the second part of the conversations. Note that negative values in the table indicate that the speaker with an initially slower AR showed a higher AR than his interlocutor in the second part of the interactions. In the French tandem conversations this trend was almost exclusively due to a reduction of AR on the part of the native French speakers. In other words, the overall reduced speech tempo of the French participants for the benefit of the GA interlocutors reported above seems to be the reflection of a gradual slowdown of AR in the course of the conversations. In the English tandem sessions, the patterns of AR convergence were more heterogeneous. Both native GA speakers and French L2 speakers accommodated via a reciprocal in- and decrease of AR.

Speaker Pair	Conversation					
	<i>English Tandem</i>			<i>French Tandem</i>		
	Part 1	Part 2	Diff.	Part 1	Part 2	Diff.
FR1	4.65	4.97		5.62	5.36	
GA1	5.20	4.16	0.55 > - 0.23	3.92	3.82	1.70 > 1.54
FR2	4.07	4.39		5.28	5.04	
GA2	5.06	5.10	0.99 > 0.71	3.84	3.78	1.44 > 1.26
FR3	4.53	4.73		5.30	4.25	
GA3	6.19	5.50	1.66 > 0.77	4.78	4.70	0.52 < - 0.35
FR4	4.85	4.81		7.47	6.64	
GA4	4.54	4.57	0.31 > 0.24	3.82	4.02	3.65 > 2.62
FR5	4.88	4.85		5.20	4.80	
GA5	4.35	4.72	0.53 > 0.13	3.58	3.48	1.62 > 1.32

Table 11-1. Mean AR values and AR differences (syll/s) of paired tandem partners. In the column ‘Diff’, the first number of each pair indicates the difference in the first part, the second number the difference in the second part of the tandem conversations in English (left) and French (right). Negative values implicate that the initially slower speaker became the faster speaker in the course of the conversation.

Individual tandem partners, therefore, displayed quite different accommodation patterns of AR depending on the session and language use. Looking for example at the values of speaker pair GA5/FR5 in Table 11-1, it can be noticed, subject FR5 shows the aforementioned reduction, subject GA5 a relative constant AR during the French tandem conversation. During the English tandem conversation, by contrast, speaker FR5 displays little or no change in AR, while speaker GA5’s AR increases.

11.4 Discussion

As in previous studies, the results of our analysis of the control conditions confirmed a cross-linguistic difference in AR between French speakers and speakers of American English, with the former being produced substantially faster than the latter (Grosjean & Deschamps, 1972; Dellwo & Wagner, 2003). This difference can be partially explained by language-specific properties. It has been widely demonstrated that AR is influenced by a number of factors, such as syllable complexity, consonant phonotactics or reduction processes in unstressed syllable position (Dauer, 1983; Schiering, 2007). As a consequence, the articulation of syllables in French, with its simpler syllable structure and less complex consonant

clusters, requires less time and is produced faster. For these reasons, studies on speech rhythm often use a normalisation procedure to adjust the articulation between different languages (e.g., Dellwo & Wagner, 2013). Such corrections, however, are usually applied to L1 speech, and our primary interest does not lie in a comparison of AR between the two languages but rather in a comparison of relative adjustments of AR within one language.

In absolute terms, the AR values in the present study seem to be marginally slower than those reported in some other studies. For French 6.78 syll/s rather than 7.3 syll/s or 7.18 syll/s (Dellwo & Wagner, 2003; Pellegrino, Coupé & Marsico, 2011); for General American 4.88 syll/s rather than 5.12 syll/s (Jacewicz, Fox, O'Neill & Salmons, 2009). These differences could be related to differences in the conversational task in the studies. Spontaneous speech in previous research frequently consists of short unconstrained talks of the speaker answering to a set of questions (e.g. Jacewicz, Fox, O'Neill & Salmons, 2009; Kim, Ackerman, Burchfield et al., 2013). This is quite different from the interactive dialogue structure in tandem conversations. The give-and-take nature of the interactions is likely to result in a higher amount of variation which, in turn, may be reflected in a somewhat slower overall articulation rate.

Our findings are furthermore consistent with reports that speakers are generally slower in their L2 than in their native language (Cucchiari, Strik & Boves, 2000; Gut, 2009). However, our results do not confirm that L2 articulation rates are on average slower than those of L1 speakers. While GA speakers were slower in their L2 French productions as compared to the French native speakers, there was no difference at the group level between French L2 learners and the GA partners in the English tandem sessions. Strikingly, three out of the five French L2 learners produced higher AR values than the L1 speakers. These results suggest that L2 articulation rates are not only related to L1 rates on an individual level – a factor which we did not analyse in the present study – but also that L2 rate variations are dependent on the language background. The intrinsically high and slow rates of French and American English, although motivated by the language-specific properties outlined above, interweave with individual rate habits. This allows French learners of English to closely approach the native speakers' level whereas this is not the case in the opposite situation for GA learners of French. A similar language-specific effect has been observed in a recent study by Kim and colleagues (2013) on rate variations in bilingual speakers of different-language backgrounds. While the authors found some evidence for language-specific rate variation across different languages with, for

example, Turkish and Mandarin being faster than Korean or Spanish, and AR was generally faster in the dominant than in the second language (English), there was also a significant influence of the rate of the L1 on the AR patterns in English across the different language groups. Thus, like in the present study, the results indicated language-specific and status-specific (first- vs. second-language) effects on the rate.

In addition, our data showed some instances of speaker accommodation. On a global level, French speakers significantly slowed down their L1 productions when talking to L2 learners, while no such rate adjustments to the L2 interlocutors were observed in the English tandem sessions. As mentioned earlier, only a few studies have directly focused on the acoustic properties of ‘foreigner talk’ and we are not aware of any studies that directly compared ‘foreigner-directed speech’ of different language backgrounds. We can therefore only speculate at present whether different assessments of the interlocutors’ L2 level by the French and GA speakers play a role, or whether these one-sided adjustments are, once more, guided by the underlying faster AR of French, favouring a reduction of the conversational tempo.

With regards to phonetic accommodation on an intra-session level, the majority of tandem pairs were more similar in their AR during the second part of the conversations. Interestingly, those rate adjustments did not reveal an L1 leader / L2 follower dynamic as could have been assumed. For instance, Trouvain and Moebius (2014) found that L2 learners increased their L2 articulation rate in read sentences after direct exposure to a model sentence produced by an L1 speaker and interpreted these results as an effect of convergence. In the present spontaneous interactions, it was rather the native speakers who adjusted their rate in the French tandem interactions, while in the English sessions the convergent tendencies were mutual, regardless of the language status. In other words, the contribution to convergence of an individual speaker in a given speaker pair differed according to the language used. A more fine-grained analysis of phonetic convergence of AR during the tandem conversations, including local adjustments in AR at turn exchanges (cf. Levitan & Hirschberg, 2011), will be necessary in the future.

One factor that has not been considered in our analysis is the length of the stretch or ‘phrase-length’. In several studies, articulation rate has been shown to vary with the number of syllables in a phrase. For instance, Quené (2008) reported that in Dutch phrase length is a significant predictor of speech tempo. Longer phrases with more syllables are expected to be produced at a faster rate in comparison to shorter phrases due to anticipatory shortening, i.e. speakers shorten their syllables if they

anticipate more syllables within a phrase (see also Crystal & House, 1990). However, the tendency towards higher articulation rate in longer utterances could not be confirmed in a study on German spontaneous speech (Trouvain, Koreman, Erriquez & Braun, 2001) nor in a study on spontaneous speech in two distinct varieties of American English (Jacewicz, Fox & Wei, 2010). As has been pointed out in Fig. 11-1, the speakers in our study produced a variety of short and long stretches in both their L1 and their L2, although with a slight tendency for native speakers to have a higher number of longer phrases than L2 learners. Nevertheless, as Fig. 11-1 also reveals, there seems to be no strict correlation between AR and phrase-length; rather, given the spontaneous quality of the interactions, the productions display a high amount of variance (cf. Trouvain, Koreman, Erriquez & Braun, 2001). If anything, the French learners' AR in their L2 matched the performance of the native speakers, *despite* the fact that their productions contained more stretches with fewer syllables. Thus, although we did not consider phrase length in our analysis, this should not have affected the interpretation of our results significantly.

It should also be noted that our data do not permit to draw any conclusions with regard to perceived accentedness or comprehensibility of L2 learners. Faster articulation rates do not automatically translate into an increase of perceived L2 proficiency. In fact, as Munro and Derwing (2001) have shown, listeners tend to assign lower comprehensibility judgments to L2 speech that is produced with either very slow *or* very fast AR. In a future study, therefore, we intend to obtain perceptual assessments of the tandem speakers' L2 performances.

In summary, the current results provide several distinct findings with regard to the articulation rate in spontaneous interactions between French and GA speakers. Consistent with previous research, the L1 articulation rate was distinctly faster in French than American English and all subjects spoke more slowly in their L2. On average, L2 performances were only slower for GA learners of French but not for French learners of English. In addition, only French speakers tended to clearly accommodate their conversational tempo when talking to foreign interlocutors, indicating that the French speakers' production patterns might reflect the inherent overall faster AR of their L1. Finally, we found some evidence of convergence of AR during the course of the individual conversations, but the patterns of the speaker pairs varied according to whether they used their L1 or their L2 in the different tandem sessions.

References

- Avanzi, M., Dubosson, P. & Schwab, S. (2012). Effects of dialectal origin on articulation rate in French. In: *Proceedings of Interspeech 2012*, pp. 651–654.
- Babel, M. E. (2009). *Phonetic and Social Selectivity in Speech Accommodation*. Doctoral Dissertation. Berkeley: University of California.
- Barquero Armesto, M. (2012). A comparative study on accentual structure between Spanish learners of French interlanguage and French native speakers. In: *Proceedings of the 6th International Conference on Speech Prosody*, pp. 603–606.
- Barras, C., Geoffrois, E., Wu, Z. & Liberman, M. (2001). Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication*, 33(1–2), pp. 5–22.
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4, *Journal of Statistical Software*, 67, pp. 1–48.
- Biersack, S., Kempe, V. & Knapton, L. (2005). FineTuning speech registers: a comparison of the prosodic features of child-directed and foreigner-directed speech. In: *Proceedings of the 9th European Conference on Speech Communication and Technology*, pp. 2401–2404.
- Boersma, P. & Weenink, D. (2014). *Praat: Doing Phonetics by Computer* (Version 5.4.08). Retrieved from <http://www.praat.org>.
- Coupland, N. & Giles, H. (1988). Introduction: The communicative contexts of accommodation. *Language and Communication*, 8, pp. 175–182.
- Crystal, T. H. & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, 88, pp. 101–112.
- Cucchiari, C., Strik, H. & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, 107(2), pp. 989–999.
- Dellwo, V. & Wagner, P. S. (2003). Relations between language rhythm and speech rate. In: *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 471–474.
- Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, pp. 51–62.

- Ellis, R. (1997). *Second Language Acquisition*. Oxford: Oxford University Press.
- Fletcher, J. (2013). The prosody of speech: timing and rhythm. In: Hardcastle, W. J., Laver, J. & Gibbon, F. (Eds.), *The Handbook of Phonetic Sciences*, pp. 523–602. Chichester: Wiley-Blackwell Publishing.
- Gas, S. & Selinker, L. (2013). *Second Language Acquisition: An Introductory Course* (2nd ed.). London: Routledge.
- Goldman-Eisler, F. (1961). The significance of changes in the rate of articulation. *Language and Speech*, 41, pp. 171–188.
- Grosjean, F. & Deschamps, A. (1972). Analyse des variables temporelles du français spontané. *Phonetica*, 26, pp. 129–156.
- Guion, S. G., Flege, J. E., Liu, S. & Yeni-Komshian, G. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, 21, pp. 205–228.
- Gut, U. (2012). Rhythm in L2 speech. In: Gibbon, D. Hirst, D. & Campbell, N. (Eds.), *Rhythm, Melody and Harmony in Speech: Studies in Honour of Wiktor Jassem, Special edition of Speech and Language Technology*, pp. 83–94.
- . (2009). *Non-native Prosody. A Corpus-based Analysis of the Phonetic and Phonological Properties of L2 English and L2 German*. Frankfurt: Peter Lang.
- Horgues, C. & Scheuer, S. (2014). Why some things are better done in tandem? In: *Proceedings of the Third International Conference on English Pronunciation: Issues and Practices*, pp. 41–44.
- Jacewicz, E., Fox, R. A. & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, 128, pp. 839–850.
- Jacewicz, E., Fox, R., O’Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, 21(2), pp. 233–256.
- Kim, M., Ackerman, L., Burchfield, L., Dawdy-Hesterberg, L., Luque, J., Mok, K. & Bradlow, A. (2013). Rate variation as a talker-specific/language-general property in bilingual speakers. *Journal of the Acoustical Society of America*, 133(5), p. 3574.
- Kim, M., Horton, W. & Bradlow, A. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Journal of Laboratory Phonology*, 2, pp. 125–156.
- Levitan, R. & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech 2011*, pp. 3081–3084.

- Munro, M. & Derwing, T. (2001). Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23, pp. 451–468.
- Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, 4: 559; doi: 10.3389/fpsyg.2013.00559.
- Pardo, J. S., Cajori Jay, I. & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception & Psychophysics*, 72(8), pp. 2254–2264.
- Pellegrino, F., Coupé, C. & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 87(3), pp. 539–558.
- Penman, C. (2002). Evaluating tandem interactions. In: *The Guide to Good Practice for learning and teaching in Languages, Linguistics and Area Studies*. LTSN Subject Centre for Languages, Linguistics and Area Studies, University of Southampton.
- Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America*, 123(2), pp. 1104–1113.
- R Core Team (2012). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rao, G. (2013). *Measuring phonetic convergence: segmental and suprasegmental speech adaptations during native and non-native talker interactions*. UT Electronic Theses and Dissertations. Retrieved from <http://hdl.handle.net/2152/23105>.
- Raupach, M. (1980). Temporal variables in first and second language speech production. In: Dechert, H. & Raupach, M. (Eds.), *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*, pp. 263–270, The Hague: Mouton.
- Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L., Dmitrieva, O. (2007). An acoustic study of real and imagined foreigner-directed speech. In: *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 2165–2168.
- Trouvain, J. & Möbius, B. (2014). Sources of variation of articulation rate in native and non-native speech: comparisons of French and German. In: *Proceedings of Speech Prosody 2014*, pp. 275–279.
- Trouvain, J., Koreman, J., Erriquez, A. & Braun, B. (2001). Articulation rate measures and their relations to phone classification of spontaneous and read German speech. In: *Proceedings of the ISCA Workshop on Adaptation Methods for Speech Recognition*, pp. 155–158.
- Schiering, R. (2007). The phonological basis of linguistic rhythm. Cross-linguistic data and diachronic interpretation. *Sprachtypologie und Universalienforschung*, 60, pp. 337–359.

- Schwab, S., Dubosson, P. & Avanzi, M. (2012). Etude de l'influence de la variété dialectale sur la vitesse d'articulation en français. *Actes des XIX^e Journées d'Etudes de la Parole*, pp. 521–528.
- Uther, M., Knoll, M.A., Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, 49, pp. 2–7.
- Whiteside, S. (1996). Temporal-based acoustic-phonetic pattern in read speech: Some evidence for speaker sex differences. *Journal of the International Phonetic Association*, 26, pp. 23–40.

PART IV:

**METHODOLOGY OF TESTING AND
TEACHING**

CHAPTER TWELVE

EVALUATING THE ESSENTIALS: THE PLACE OF PROSODY IN ORAL PRODUCTION

DAN FROST AND JEAN O'DONNELL

12.1 Introduction

As in many countries, university students in France study a foreign language no matter what they major in. In France, this is compulsory for all students, most of whom study English, and conditions vary a great deal from one context to another (Taillefer, 2002). Studies such as EPTiES¹ (Henderson et al., 2012) have shown that pronunciation teaching is often neglected by both teachers and by teacher trainers. In France, as we shall see, there is a great need to work on certain aspects of pronunciation.

The French Ministry of Education and Research has specified the levels which it expects learners to achieve in their foreign languages at various stages of their education, and for LSP/LAP² learners, this means B2 according to the CEFRL³ on arrival at university (Goullier, 2005: 38). In fact, most students arrive after eight–ten years of secondary schooling with a level that is closer to A2 (Taillefer, 2007; Macré, 2015). The Innovalangues project⁴ is an attempt to address this issue: its primary mission is to help to bring the levels of LSP/LAP learners to a certified B2 level as defined by the CEFRL over the three years of their university career and the main thrust of the project is the creation of a “digital ecosystem” based around the platform Claroline Connect (Masperi & Quintin, 2014). The Innovalangues project is a six-year project which started in 2012 and is coordinated by Monica Masperi at Université Stendhal

¹ English Pronunciation Teaching in Europe Survey.

² Languages for Specific Purposes / Languages for Academic Purposes.

³ Common European Framework for Reference in Languages (COE, 2001).

⁴ <http://innovalangues.fr/>

(Grenoble 3). It is composed of a team of about 60 teachers, researchers, resource developers, IT developers and administrative staff, some of whom are employed full-time and some part-time to work on the project paid for with substantial funding from the *ANR (Agence Nationale pour la Recherche)* – the research funding body of the French Ministry of Education and Research. Initially, the target learners are students in Grenoble, but the project will also reach learners nationally and internationally and already has several national & international partners. The project contains several teams, and the subject of this chapter is part of the work of one of those teams: *THEMPPO*⁵ (Frost & Picavet, 2014). For reasons we shall see in the next part, *THEMPPO* was created to work specifically on the prosody of English on two fronts: firstly, the development of teaching resources (tools, media and activities) for use within the digital eco-system as part of a blended learning programme and secondly, a series of teacher-training seminars. The approach adopted by the team is essentially an articulatory approach (Honikman, 1964), especially towards the beginning of each course or learning path. We believe that as pronunciation is a complex set of physical gestures and as prosody is so iconic (Pennington, 1996: 137) and physical (Messum, 2009), work must begin with the body and the articulators. After “raising awareness to the articulatory settings” of English, constant reference is made to the body and articulators and various kinesthetic techniques are often used to reinforce the acquisition of stress patterns, etc. The pedagogical approach and the tools and activities which we have developed are presented elsewhere (Frost & Picavet, 2014; Frost & Guy, forthcoming), and are not the main focus of the present work. This chapter concentrates solely on the conception and calibration of one of those tools, i.e., the prosody-based descriptors for assessing oral production in English.

The descriptors are based on the CEFRL scales, which are, for various laudable political, pedagogical and linguistic reasons, plurilinguistic and based on communicative competence. We are aware, however, that our starting point is a tool which is neither language-specific nor form-based. We have developed an assessment tool which is similar in *form* to the CEFRL descriptors but is concerned only with “phonological control”, i.e., the subjects’ accent in English. In undertaking the development of this tool, we are attempting to address two main research questions: Firstly, is it feasible to peg such a tool to the CEFRL levels? And secondly, is such a tool useful, for teaching, for assessment and for research? In the next section, we will examine issues surrounding our choice to focus on

⁵ *THEMatique Prosodie et Production Orale*.

prosody, including the differences between English and French prosody, language transfer, pronunciation in language instruction and relevant assessment issues. In the following section, we will describe the conception and calibration of the descriptors and finally, we will present the results of a preliminary pilot study and discuss some of the implications.

12.2 Theoretical background

12.2.1 English and French prosody

French and English, despite sharing much of their vocabulary, are extremely different both phonetically and phonologically, particularly in the domain of prosody (see Frost, 2011 for a more complete contrastive analysis). What interests us particularly as teachers of English is how prosody is used to mark stress, both at the word level and at the level of the tone unit (TU). It is of course unrealistic to separate intonation and prominence in natural speech, but in order to identify some of the problems faced by French learners of English, let us look first at issues concerned more with intonation, then prominence, and finally segmental questions. The prosodic and segmental features which are treated in this section explain to a large extent our choices in developing the prosody-based descriptors which we will present later in the chapter.

Intonation is very closely related to prominence, especially in English, and the range between the high and low points for an average native speaker (NS) of English is greater than for an average French speaker (see Campione & Véronis, 1998 for a comparison of F₀ range across five European languages). Across larger prosodic units, in particular tone units, a French learner of English typically has flatter intonation patterns than an English NS, with a step-up or a step-down at the end of a TU, as we will see later.

Regarding stress, English and French are very different. French does not have lexical stress and it exhibits relative prominence to a lesser degree than in English (Rossi, 1979). To compound difficulties for French NSs, this “relative prominence” is fixed, i.e., it is usually present only on the last syllable of prosodic units (Dahan & Bernard, 1996). These units may be shorter “stress groups” (Di Cristo, 1998) or longer “accentual phrases” (Jun & Fougeron, 1995). This final prominence, marked by a lengthening and often a fall in F₀, may be explained by articulatory features, such as there simply being less acoustic energy at the end of a breath group. Wenk and Wioland (1982: 204) therefore describe French as

“trailer-timed” and English as “leader-timed”. This leads to an alternation between strong (prominent) syllables and weak (reduced) syllables, which is what characterises the rhythm of English. The concept of isochrony (Pike, 1945) has been questioned by many authors (see Bertinetto, 1989 for an overview), and is clearly too simplistic a model for describing the reality of natural speech, although many teachers and researchers use models inspired by Pike and Bolinger’s work even today (See Dickerson, 2015) and in our work, we have found that treating English as “stress-timed” and French as “syllable-timed” over shorter TUs can be useful for raising the awareness of learners and teachers to different metrical structures. There are many possible explanations for this, not least of which is the relative importance of the acoustic cues F0, amplitude, duration and formant structure in English. F0, perceived as pitch, is essential to producing and perceiving prominence in English, and it is this which led Bolinger (1958) to call the marking of stress in English “pitch prominence”. Prominence is key to segmenting the speech signal, and research points to the relative importance of F0 compared to other cues in English compared to French (Frost, 2011; Coughlin & Tremblay, 2012). French, however, marks the group-final syllable mentioned above with an increased duration – this is not to say that amplitude and F0 are not factors, but syllable-lengthening is the most salient feature (Benguerel, 1973; Di Cristo, 1998; Lacheret-Dujour & Beaugendre, 1999: 41; Jun & Fougeron, 2000; Astesano, 2001).

Finally, there are also many segmental differences between French and English. French has simple vowels, some of which are nasalised. French vowels are all tense and there are far fewer than in English – only 10–13 oral vowels and 3–4 nasal vowels (Fougeron & Smith, 1993) whereas English has a complex system comprising of lax short vowels, long tensing vowels, diphthongs and triphthongs. As for consonants, English has consonants which do not exist in French (/h/, /θ/ and /ð/) and there are many differences in the place and manner of plosives, fricatives and glides. We are, however, less concerned with these segmental differences apart from those which have a close relationship with the prosodic features of English, as we shall explain in the next part. Segmental features are important to this work for three reasons. Firstly, at the syllable level, unstressed syllables are often reduced in English. As Jenkins (2000: 147) points out, the weak/strong syllable alternation is a characteristic feature of all varieties of English. We believe this is at the heart of the production and comprehension problems that many learners, especially French NSs, have with English. Jenkins goes on to say that weak forms are “unteachable”, an assertion which is not backed up by research and which

we refute strongly. Secondly, we consider full, unreduced vowels to be important as they ‘carry’ the stress. It is therefore important to focus to some degree on the “correct” pronunciation of full vowels so as to differentiate them from reduced vowels. Thirdly, regarding consonants, we are essentially concerned only with phonotactic phenomena, particularly at word boundaries, as they are problematic for oral production and comprehension for many French NSs learning English.

The differences between French and English pronunciation for pedagogical purposes (i.e., the priorities which research has led us to establish) may be summed up in Table 12-1:

		English	French
Prosody	Rhythm & stress	<ul style="list-style-type: none"> • Lexical stress • F0 an important cue • + Stress-timed • Strongly marked nuclear stress • Final lengthening if nuclear stress is final 	<ul style="list-style-type: none"> • No lexical stress • F0 a less important cue • + Syllable-timed • Weakly marked nuclear stress • Evident final lengthening
	Intonation	<ul style="list-style-type: none"> • Large range • Smooth and varied contours throughout TUs 	<ul style="list-style-type: none"> • Narrow range • Step up / step down changes more frequent
Segmental	Syllables	<ul style="list-style-type: none"> • Very frequent reductions (mainly /ə/) 	<ul style="list-style-type: none"> • Reductions are rare
	Vowels	<ul style="list-style-type: none"> • + Lax • Complex (short and long simple vowels, diphthongs, triphthongs) 	<ul style="list-style-type: none"> • + Tense • Simple vowels only
	Consonants	<ul style="list-style-type: none"> • Deletion, assimilation, etc. very frequent 	<ul style="list-style-type: none"> • Deletion, assimilation, etc. less frequent

Table 12-1. Some differences between English and French pronunciation.

12.2.2 Language transfer: Perception, pronunciation & prosody

The differences between the phonetics and phonology of French and English which we outlined in the previous part are relevant to language teaching because of the phenomena of language transfer, or interference (Weinrich, 1953). Although it is not within the scope of this chapter to provide a full review of language transfer and L2 phonology (for a critical review, see, e.g., Major, 2008), we would like to highlight a few points which are relevant to the current study. Firstly, some of the very first work on transfer was done in an attempt to understand issues related to pronunciation and perception. The Prague Circle was already aware of “phonological deafness” (Polivanov, 1931; Trubetzkoy, 1939) well before the instruments of modern acoustic phonetics were able to measure such perceptual phenomena. Secondly, the degree to which negative transfer may be problematical depends on the languages concerned. The differences between English and French prosody are substantial, as we saw in the previous section. Lado’s “contrastive analysis hypothesis” (1957) suggests that the greater the difference between a language feature in a learner’s L1 and the target language, the harder it will be to learn and Eckmann’s “markedness differential hypothesis” (1977) supports this idea. Thirdly, following Selinker’s definition of interlanguage (1972) and Corder’s work on learner errors (1981), many teachers chose to accept the effects of language transfer on pronunciation, because fluency was the emphasis in the communicative approach which dominated English teaching for this period. The integration of interlanguage into Krashen’s “Monitor Model” (1981) was an example of this. Fourthly, many researchers have preferred to focus on segmental features. For example, Flege’s “Speech Learning Model” (Flege, 1995; Flege, Schirru & MacKay, 2003) and MacWhinney’s “Unified Model” (MacWhinney, 2008) devote considerable importance to the impact of L1 phonology on the production and perception of other languages, but as far as our purposes are concerned, do not place enough emphasis on prosody. Finally, however, recent work on transfer has focused more on prosody, as teachers and researchers are increasingly concerned with intelligibility and teaching pronunciation for comprehension. Dolbec and Santi refer to a “linguistic filter” (1995: 46) and Dupoux and Peperkamp and their colleagues in Paris identified stress in English as being problematical for French and Spanish NS learners coining the term “stress deafness” (Dupoux & Peperkamp, 1999; Dupoux, Peperkamp & Sebastien-Galles, 2001; Peperkamp & Dupoux, 2002). Furthermore, research also shows that

prosody is one of the most stubborn features of a learner's L1, even with proficient language users (Bailey, Plunket & Scarpa, 1999; Flege, Schirru & MacKay, 2003; Gabriel & Kireva, 2014). As we have seen in the previous section, the differences between English and French prosody are both numerous and considerable, and in the next section, we shall examine the implications of L1 transfer for our pronunciation instruction and our pedagogical choices.

12.2.3 Pronunciation and prosody instruction

The importance given to teaching pronunciation has waxed and waned over the years, with little emphasis on oral language skills at all in formal education before the direct methods of the early 20th century. Behaviourist ideas and the development of analogue language laboratories saw an expansion of repetitive drilling and a focus on accuracy in pronunciation. The “Army method” which drew heavily on behaviourist ideas developed by Skinner led to the audio active comparative drills used in schools and universities throughout the world for decades. More recently, pronunciation was neglected throughout the nineties and early part of this century during the hegemony of the communicative approach, with authors such as Judy Gilbert referring to pronunciation as the orphan of L2 teaching (Gilbert, 2010), but things are changing now. Certainly the Web as we know it today with much user-generated media-rich content, downloading and streaming of music, films and TV series involves more active engagement with audio and video than ever before. Many authors and teachers have been pushing for the integration of pronunciation into English language teaching (ELT) (Kjellin, 1999b; Henderson, 2008; Gilbert, 2008, 2010; Derwing, 2010; Munro & Derwing, 2015). Moreover, the existence of conferences such as EPIP (English Pronunciation Issues and Practices) and the inauguration of the *Journal of Second Language Pronunciation* in 2015 are evidence of this renewed interest. There has, however, been a major shift in the goals of pronunciation instruction over the last decade: the more pragmatic targets of intelligibility and comprehension are now central to the work of many teachers and researchers (Zielinski, 2006; Munro & Derwing, 2011, 2015; Harding, 2012). This is one of the key factors in our choice to focus on prosody in our pedagogical approach. All too often, teachers in France focus on features such as /θ/ and /ð/ which do not impair intelligibility or impinge on comprehension to the same degree – after all, Irish speakers of English realise inter-dental fricatives as alveolar plosives (Hickey, 2004) and -th- in parts of London and elsewhere is often pronounced as /f/ or /v/

(Kerswill, 2006) without causing communication problems. Research shows that prosody is important for the perception of a foreign accent (Jilke, 2000) but more importantly, it is one of the major factors contributing to the intelligibility and comprehension of L2 speakers (Hahn, 2004; Munro & Derwing, 2011; Piske, 2012; Acton, Baker, Burri & Teaman, 2013).

There are as many different ways of learning and teaching pronunciation as there are learning situations, and various studies have demonstrated the effectiveness of teaching pronunciation (see for example Scarcella & Oxford, 1994 and Saito, 2012 for overviews). As described in the introduction to this chapter, the context of this work is courses of 20 to 60 hours per year with undergraduate students or adults who are not language specialists. On such short courses with non-specialists, our work and the work of others (Derwing & Munro, 1998, 2005; Kjellin, 1999b; Celik, 2001; Murphy, 2004) has shown that focusing on prosody is an efficient use of limited time. In the French context, other authors have chosen to work on prosody (Cooke, 1993; Herry, Nishinuma & Ghio, 2003; Stenton, 2011; Horner, 2013, 2014) for the same reasons as we describe. The realisation of prosodic features, particularly stress, involves increased acoustic energy and physical effort. We therefore consider it extremely important to work on the body, breathing and the articulators. This view is shared by other teachers and researchers interested in prosody (Kjellin, 1999a, 1999b; Borrell & Salsignac, 2002; Messum, 2009; Souleine, 2014). As we mentioned in the introduction, our approach therefore begins with raising awareness and training of the body and the voice before progressing to freer production tasks. In the context of the Innovalangues project, this work is carried out not only in classrooms, but also in an online environment (mentioned in the introduction) using video recordings and playback tools and a variety of tasks. If we are to put prosody at the centre of our teaching, then we need a tool to define objectives, assess progress and allow learners to assess their own progress. As we shall see in the next part, however, there is a paucity of such tools in language teaching.

12.2.4 Pronunciation and prosody instruction

In the French university system when it comes to assessing what is referred to as either spoken language, speaking, oral production, speaking proficiency, fluency, speaking skills, oral expression, communicative language competence or speech, students are generally graded on a scale ranging from 0 (lowest grade) to 20 (highest grade) with 10 representing a

pass. Grades on oral exams are usually based on in-house scales devised either by individual teachers or small teaching teams all of whom have a great deal of freedom regarding teaching and assessment practices. Lack of a common reference frequently leads to incoherence of grades obtained by students not only within a given course but also from one course to another. The CEFRL (see Table 12-2) is rarely actually used as an in-house tool to assess students' oral competence, indeed only one of the 32 scales within the CEFRL is devoted to pronunciation and is entitled "phonological control" (Council of Europe, 2001: 117). Despite this fact, Henderson et al. (2012) found that it was the most frequently quoted set of scales by teachers when asked what, if any, tools they used to assess pronunciation.

PHONOLOGICAL CONTROL	
C2	As C1
C1	Can vary intonation and place sentence stress correctly in order to express meaning.
B2	Has acquired a clear, natural, pronunciation and intonation.
B1	Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.
A2	Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time.
A1	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group.

Table 12-2. CEFRL "phonological control" descriptors (COE, 2001: 117).

These "phonological control" descriptors can be criticised on several fronts. For example the use of the terms "*clear and natural*" at level B2 tends to reinforce the perception of the model of the native speaker. This runs contrary to what is stated in the global scales with respect to C2, the highest level attainable in the CEFRL global scales, defined as "the degree of precision, appropriateness and ease with the language which typifies the speech of those who have been highly successful language learners" (Council of Europe, 2001: 36). Harding (2013) points out that the use of vague terminology can be interpreted differently by different assessors, for example "clear enough to be understood", "can be understood with some effort", "foreign accent is sometimes evident", etc.

The issue of the acquisition and therefore the assessing of language competence existing on a continuum is a much-debated topic (Fulcher,

2004). Horner (2014) questions whether or not the acquisition of phonological control can be considered as a hierarchy. When for example should intonation be taught? At what stage should word stress be introduced? Unfortunately, the lack of research in general, and more specifically regarding French learners of English, means we cannot provide definite answers to such questions. Thus scales with ascending levels are not an interpretation of SLA findings as there are as many routes to learn a language as there are learners (Luoma, 2004). This means that creating scales is a real challenge and scales will reflect the designer's beliefs, experience and understanding of the learning process. An example of a re-working of the phonological control scales is proposed by Horner (2014), who tentatively suggests a new grid for pronunciation which is more holistic than analytic. He insists on the importance of intelligibility and is strongly inspired by the CEFRL, thus he accepts to a certain degree the construct of a hierarchy.

Despite the criticism levelled at the CEFRL "phonological control" descriptors, they can be considered as a starting point in attempting to assess pronunciation. As our primary concern is assessing prosody and due to the fact there are very few methods available to do so – the PEPS-C test (Peppé & McCann, 2003) is a rare example, but designed for native speakers with autism and Asperger's – the CEFRL thus paved the way for the development of the prosody descriptors presented in this chapter.

12.3 Designing the descriptors

The prosody descriptors take account of pronunciation factors not for the mere sake of correct pronunciation according to some native model or other, but with the aim of improving intelligibility and comprehension. They were developed mainly for assessment purposes but also as a tool for raising learner and teacher awareness and for structuring learning objectives, and they came into being because no other suitable tool existed. During the ELLO project (Frost & O'Donnell, 2013), we used the CEFRL scales of descriptors for speaking to assess hundreds of subjects over a three-year period. The subjects also assessed their own performances using the same scales. We found it quite constraining that the scales didn't address the particular prosody-related language-specific issues which are, as we have seen, not only responsible for intelligibility problems, but also for problems understanding spoken English. Indeed the subjects often mentioned pronunciation in the surveys and interviews we conducted. In devising the first set of descriptors, we were therefore thinking primarily of assessment, both by teachers and by learners

themselves, but also of the pedagogical value of such a tool. The primary stated goal of the CEFRL is to improve practices – assessment is just a means to an end and the scales themselves encourage the various stakeholders “to reflect on their current practice” and, amongst other things, provide a common basis for “the elaboration of language syllabuses”, (COE, 2001: 1). As we have also seen, prosody is one of the key factors in intelligibility and comprehension problems, especially in English and especially for French learners. Although we refer to the descriptors as “prosody descriptors”, segmental features are included. It is not therefore a question of segmental features *or* prosody: segmental features are therefore dealt with in a way that makes them *secondary* to prosody. This is why there is a column for reduced syllables, one for full vowels and one for connected speech phenomena, but not a column simply for the correct realisation of consonants such as /θ/ and /ð/.

The development of the current prosody descriptors is based on the conviction that prosody can be described analytically rather than holistically. In our approach, prosody and its closely linked segmental features are broken down into components, namely rhythm and stress, intonation, syllables, vowels and consonants. In keeping with the CEFRL and on account of our experience as teachers and researchers we acknowledge a certain hierarchy, although not universal or totally rigid, in the progression of the acquisition of the above prosodic features. Whereas the CEFRL descriptors are not language-specific, the tool which we have devised is specific to English and has been calibrated for use by French learners. The implications of this choice will be discussed later in this chapter. The prosody descriptors are calibrated to the CEFRL oral production descriptors, but it is of course the case that a learner may have relatively poor phonological control, but good grammatical and lexical control or vice versa. Consequently, the “level” which the prosody-based descriptors permit a user to attribute to a given performance for a given subject may or may not correspond exactly to overall oral performance as determined by the CEFRL descriptors for speaking. In this section, we will describe the choices we made and the procedures we adopted for the design and calibration of the descriptors. The prosody descriptors and corresponding assessment sheets were initially designed intuitively based on twenty years of teaching experience, observations and assessment. The descriptors then underwent a period of revision, calibration and piloting that has lasted two years. In line with the CEFRL guidelines, the grid and assessment sheet were designed to be as positive, clear, brief and independent as possible. We also set out to ensure that they would be “flexible”, “dynamic”, “user-friendly”, and “non-dogmatic” (COE, 2001:

pp. 7–8). “Can do” statements were used to ensure the overall perspective of the grid would be perceived positively by students and teachers alike.

The starting point for the calibration process was a set of “full” prosody descriptors containing eight columns (Appendix 3) – four with quantitative criteria (QTT) and four with qualitative criteria (QLT). The original eight columns were as follows:

1. Word Stress (QTT)
2. Pause & Tone Units (QLT)
3. Focus (Nuclear & Contrastive Stress) (QLT)
4. Rhythm (number of beats) (QTT)
5. Reductions (QTT)
6. Stressed & Unstressed Vowel (QLT)
7. Connected Speech (QTT)
8. Intonation (QLT)

The full set of descriptors with eight columns was then calibrated in three stages. Firstly, there was a reading task, secondly two spontaneous speech tasks and thirdly, a preliminary pilot study. As we shall see below, to make the tool easier to use, the final version of the descriptors contains only five columns (see Appendices 1 and 2).

12.3.1 Reading task

It was decided to begin with a reading text because in this way we could control for all the features which interest us; although the prosody and the phonotactic phenomena of spontaneous speech would be more interesting, it would be hard to compare like with like. A text – “The Mallory text” – containing all the prosody features in the descriptors was selected, then read and recorded by two native North American speakers and two native British speakers. Using an orthographic transcription, the prosody features were annotated by three native speakers to obtain a harmonised annotation.

Some of the features were obligatory, such as correctly placing word stress, but some of the features were optional: the number of phonotactic phenomena realized at word boundary level, for example, depends on speaker style, speed of delivery, etc.

The calibration using the reading task was carried out using the full set of prosody descriptors, i.e., all eight columns mentioned above. The annotated text broke down into twenty-eight individual potential tone units (TUs). For each of these TUs, the number of potential occurrences of the

four QTT features was calculated (yielding between 0 and 117 possible “scores” per feature/column). For the QLT features a possible score (0, 1 or 2) was decided upon (yielding between 0 and 50 possible “scores” per feature/column). (See Appendix 3 for the calibration sheet for the reading task).

Three representative subjects’ recordings (A2, B1 and B2+) were then selected. The overall oral production level of these speakers had previously been obtained by double expert assessment and self-assessment using the CEFRL overall speaking descriptors. These three speakers then recorded the Mallory text. Two raters spent three days listening to and “scoring” each of the three recordings TU by TU. The total number of scores for each of the eight prosody features per subject was obtained independently by each of the raters, then discussed (with further listening when necessary) in order to reach a consensus for each subject’s “scores”. Based on the results of the discussions the descriptors underwent several adjustments. “Can do” statements were reworded, moved from one level to another, others were added or omitted and a simplified “basic” version of the descriptors was adopted, which evolved constantly over the next few weeks until the current version, version 15. (See Appendix 1 for the full set of descriptors, and Appendix 2 for the accompanying assessment sheet).

12.3.2 Spontaneous speech task

The reading task was a useful way of controlling for most of the features which interested us, but a prosody assessment needs to take account of spontaneous speech. As the reading text contained no interaction, the intonation was very repetitive and the number of phonotactic phenomena was quite limited. We therefore decided on two spontaneous tasks: one monologue and one interaction. Two sets of video recordings, (i.e., monologues and interactions) from the *ELLO* project were used in the second stage of the calibration process, both of which were obtained following the WebCEF project protocol (Bijmens, 2009). The monologues involved subjects describing a 30-second television advertisement (two minutes) and the interactions involved the students chatting about their mobile phones (five minutes). The two raters (watched and) listened to twenty subjects’ monologues and interactions independently (forty recordings in all) several times without knowing the subjects’ CEFRL level. They used the “basic” descriptors to give each a “prosody level”. This was followed by a discussion and harmonisation phase. A log of issues, solutions and comments was kept. For each of the twenty

monologues and interactions the “prosody level” obtained was compared to the overall CEFRL speaking level previously obtained in the *ELLO* project. This led to further vertical realignment and verification of the descriptors as well as to the current set of “basic” descriptors along with its corresponding assessment sheet.

12.3.3 Pilot study

The third stage of the calibration, still ongoing, is the piloting of the descriptors. The first part of this process, the initial pilot study, involved contacting teachers, trainee teachers and students, fourteen in all, to test out the current version of the descriptors and corresponding assessment sheet on a video recording (the same task as used in the WebCEF and *ELLO* projects). The participants were also asked to reply to a 6-point, 22-item Likert scale questionnaire on the ease of use and usefulness of the descriptors. Based on the results of the pilot study, work is underway to develop a handbook and to road-test the descriptors in a variety of learning situations. These tests will result in further modifications to the descriptors and to the handbook and will lead to a tool which will be usable by teachers, trainers and students alike.

12.4 Results and discussion

12.4.1 Stage 1 calibration – Reading

Applying the descriptors to the reading of the “Mallory text” resulted in the subjects obtaining high scores across the board for “Word stress” and “Rhythm (number of beats)”. These two features were initially considered intuitively to be discriminating features at the different levels on the grid. Fig. 12-1 shows that all three subjects produced approximately 90% of all potential occurrences of words stress irrespective of their level. Similarly, all three subjects produced between 70% and 80% of all possible beats. This indicated that subjects were sensitive to the phenomena of word stress and rhythm but marked both to varying degrees in their speech. Consequently, in order to differentiate between the three subjects’ levels, the concept of the “quality” (use of the acoustic cues to stress, i.e., amplitude, duration, F0 curve and formant structure) or how well the words were stressed thus replaced the “quantity” (i.e., stressed or not) to differentiate the subjects’ scores. The original eight columns were merged into five columns (see Appendix 1):

1. Rhythm and stress
2. Reduced syllables
3. Stressed and unreduced vowels
4. Connected speech
5. Intonation

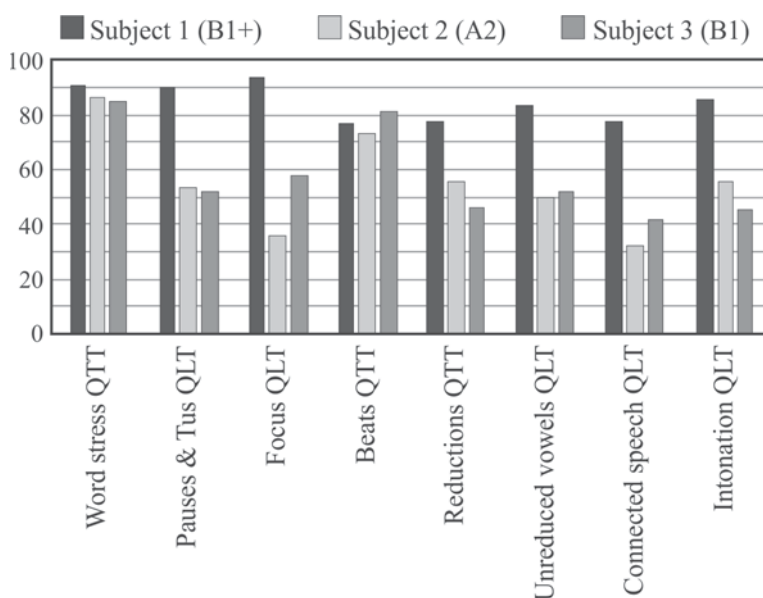


Fig. 12-1. 8 criteria for measuring prosody profiles (reading scores/100).

12.4.2 Stage 2 calibration – Spontaneous speech

This stage involved the spontaneous speech tasks (a monologue and an interaction) and actually comprised two steps: the first with all twenty subjects, and then some “fine tuning” based on the results of three selected subjects. Firstly, the average levels for the twenty subjects were considerably higher for both monologues and interactions when assessed with the prosody descriptors than when assessed using the CEFRL scales. Fig. 12-2 shows that the average level (twenty subjects) for the monologue and interaction using the CEFRL was A2+ whilst with the prosody descriptors the average level was B1+. It was therefore clear that several of our descriptors were too “generous”. This led to further adjustments in wording and vertical realignment, i.e., certain “can do” statements being

moved down one or two levels. For example, in the pre-calibrated version “can usually place word stress correctly” was in B2, whereas in the post-calibrated version, this descriptor is at A2 level.

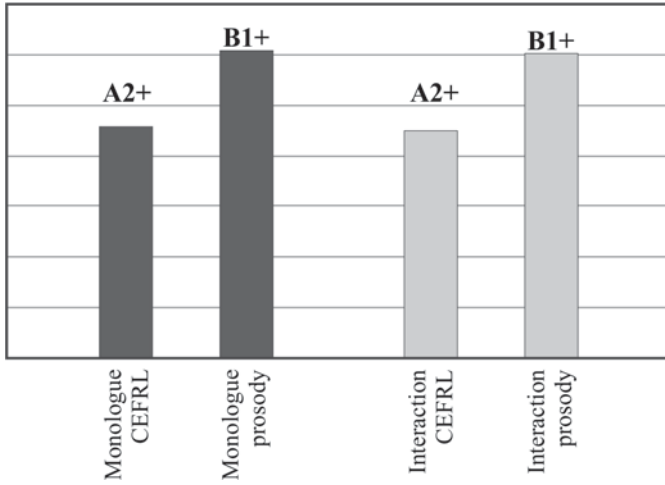


Fig. 12-2. Pre-calibration CEFRL and prosody levels for monologues and interactions (averages for all 20 subjects).

After this realignment and rewording, we looked at the individual subjects’ levels and identified three subjects from across the range of levels who had a marked discrepancy between their level using the CEFRL descriptors and their level using the prosody descriptors (subject 6 = B2/C1, subject 13 = A2/B2, subject 18 = A2/B2+ respectively). This enabled us to “fine-tune” the descriptors by making minor adjustments, especially in the “Rhythm and stress”, “Stressed and unreduced vowels” and “Phonotactics” columns.

If we compare Fig. 12-3 and Fig. 12-4 below we can see that the levels obtained after calibration were closer to the CEFRL levels. In other words, the level of a given subject’s oral performance regarding prosody using the present tool is closer to the subject’s overall oral production level obtained using the CEFRL descriptors. These final modifications resulted in the current version (V15) of the prosody descriptors and the next stage was to let other users use them in a pilot study.

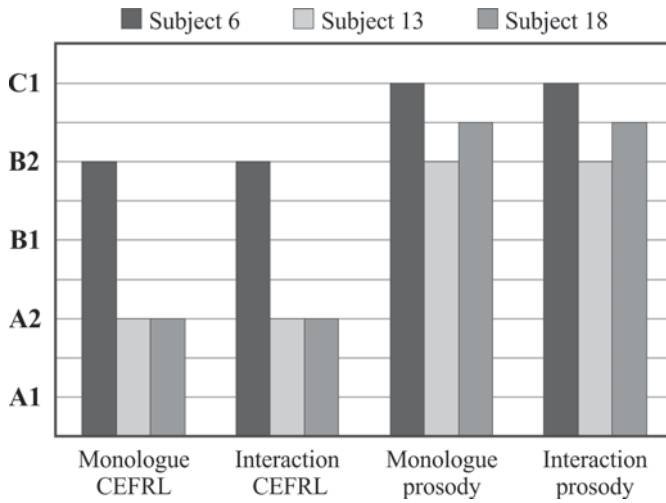


Fig. 12-3. CEFR and prosody levels for monologues and interactions before stage 2 calibration.

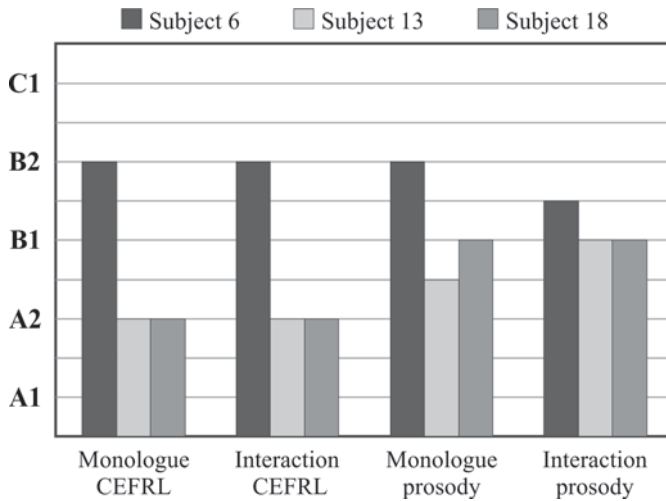


Fig. 12-4. CEFR and prosody levels for monologues and interactions after stage 2 calibration.

12.4.3 Initial pilot study

Firstly, all fourteen respondents were generally satisfied with the tool, and in the main, negative comments referred to context and training-related issues. The prosody descriptors were perceived as being both user-friendly and a good awareness-raising tool. More specifically, of the fourteen users questioned, ten of them agreed to varying degrees with the statement that the prosody descriptors were easy to use. Only four subjects found them difficult or very difficult to use. Suggestions were made that practice in using the descriptors was necessary as they referred to some concepts with which they were unfamiliar. As to the usefulness of the descriptors as an awareness-raising tool, thirteen respondents were in agreement to some extent and remarked that the descriptors enabled them to grasp the nuances between the otherwise unfamiliar elements of prosody. There was a consensus that stress was the easiest element to assess whilst connected speech was usually considered the most difficult. Finally, about half of those questioned stated that some training was required to use the tool or that a reminder of the basic technical terms and definitions related to phonology was necessary.

12.5 Conclusion

The research questions which we set ourselves were whether it was feasible to develop such a tool and to peg it to the CEFRL levels and whether such a tool would be useful for assessment, pedagogical and research purposes. We have shown that it is possible to develop a set of descriptors based on prosody and to peg them to the CEFRL levels, however the calibration process is never going to be entirely satisfactory, as it will inevitably involve subjectivity on the part of the assessor and because each individual learner has his or her own profile with different strengths and weaknesses. As the CEFRL document states: "A scale, like a test, has validity in relation to contexts in which it has been shown to work. Validation – which involves some quantitative analysis – is an ongoing and, theoretically never-ending, process." (COE, 2001: 22). The content and calibration of these descriptors will continue based on the results of ongoing pilots in different contexts. The second research question mentions three areas where the descriptors may prove useful: assessment, setting learning objectives and research. As an assessment tool, the descriptors will be used by a number of teachers and learners both within the Innovalanguages project and in other learning situations to assess learners' needs and direct them towards sequences of remedial activities

accordingly. These prosody-based descriptors are therefore, in the same way as the CEFRL descriptors, a tool for diagnostic, summative and formative assessment (*ibid* 186). The remedial activities are still being developed and will be piloted in the next year, but the descriptors have already proved valuable as they form a key part of the syllabus for developing learning objectives for oral production activities for the Innovalangues platform. Finally, on the question of research, Munro & Derwing (2015: 13) call for further defining of the construct: “Although much pronunciation-related research investigates accent, comprehensibility, and intelligibility, there has been far from perfect unanimity on how these constructs should be defined and operationalized.” We believe that this tool contributes to defining some of the fundamental elements which are essential for intelligibility and comprehension in the context where this work is taking place.

References

- Acton, W., Baker, A., Burri, M. & Teaman, B. (2013). Preliminaries to haptic-integrated pronunciation instruction. In: Levis, J. & LeVelle, K. (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference, August. 2012*, pp. 234–244. Ames, IA: Iowa State University.
- Alazard, C., Astésano, C. & Billières, M. (2010). The Implicit Prosody Hypothesis applied to Foreign Language Learning: From oral abilities to reading skills. In: *Proceedings of the 5th Speech Prosody 2010*. Retrieved from <http://www.speechprosody2010.illinois.edu/papers/100648.pdf>.
- Astésano, C. (2001). *Rythme et accentuation en français. Invariance et variabilité stylistique*. Paris: Editions L’Harmattan, Collection Langue et Parole.
- Bailey, T., Plunket, K. & Scarpa, E. (1999). A cross-linguistic study in learning prosodic rhythms: Rules, constraints and similarities. *Language and Speech*, 42(1), pp. 1–38.
- Benguerel, A-P. (1973). Corrélats physiologiques de l’accent en français. *Phonetica*, 27, pp. 21–35.
- Bertinetto, P.M. (1989). Reflections on the dichotomy “stress” vs. “syllable-timing”. *Revue de phonétique appliquée*, 91–93, pp. 99–130.
- Bijnens, H. (2009). (Ed.) WebCEF. *Collaborative evaluation of oral language skills through the web*. Leuven: AVNet, K.U.Leuven.
- Bolinger, D. (1958). A theory of pitch accent in English. *Word*, 14, pp. 109–149.

- Borrell, A. & Salsignac, J. (2002). Importance de la prosodie en didactique des langues (application au FLE). In: Renard, R. (Ed.), *Apprentissage d'une langue étrangère/seconde. Vol. 2*, pp. 163–182. De Boeck Supérieur: Pédagogies en développement.
- Campione, E. & Véronis, J. (1998). A statistical study of pitch target points in five languages. In: *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP'98)*, pp. 1391–1394.
- Celik, M. (2001). Teaching English Intonation to EFL/ESL students. *The Internet TESL Journal*, 7/12. Retrieved on September 15, 2015, from <http://iteslj.org/Techniques/Celik-Intonation.html>
- Coughlin, C. & Tremblay, A. (2012). The role of prosodic information in L2 speech segmentation. In: *Proceedings of the 35th Boston University Conference on Language Development (Supplement)*. Somerville, MA: Cascadilla Press. Retrieved from <http://www.bu.edu/buclid/files/2012/03/CoughlinTremblayBUCLDProceedingsFinal.pdf>
- Cooke, R. (1993). Reducing word stress errors: Time restricted help for ESP students. *ASp*, 2, pp. 164–175.
- Corder S.P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- Council of Europe. (2001). *A Common European Framework of Reference for learning, teaching and assessment*. Cambridge: Cambridge University Press.
- Cutler, A., Dahan, D. & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40(2), pp. 141–201.
- Dahan, D. & Bernard, J.-M. (1996). Interspeaker variability in emphatic accent production in French. *Language and Speech*, 39/4, pp. 341–374.
- Derwing, T. M. & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39, pp. 379–397.
- Derwing, T. M., Munro, M. J. & Wiebe, G. (1998). Evidence in favour of a broad framework for pronunciation. *Language Learning*, 48/3, pp. 393–410.
- Derwing, T. (2010). Utopian goals for pronunciation teaching. In: Levis, J. & LeVelle, K. (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, pp. 24–37. Ames, IA: Iowa State University.

- Di Cristo, A. (1998). Intonation in French. In: Hirst, D. & Di Cristo, A. (Eds.), *Intonation Systems: A Survey of Twenty Languages*, pp. 3–45. Cambridge: Cambridge University Press.
- Dickerson, W. (2015). A nail in the coffin of stress-timed rhythm. In: Levis, J., Mohammed, R., Qian, M. & Zhou, Z. (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference*, pp. 184–196. Ames, IA: Iowa State University.
- Dolbec, J. & Santi, S. (1995). Effet du filtre linguistique sur la perception de l'accent: étude exploratoire, *Travaux de l'Institut de Phonétique d'Aix*, 16, pp. 41–60.
- Dupoux, E. & Peperkamp, S. (1999). Fossil markers of language development: Phonological “deafness” in adult speech processing. In: Laks, B. & Durand, J. (Eds.), *Cognitive Phonology*, pp.168–190. Oxford: Oxford University Press.
- Dupoux, E., Peperkamp, S. & Sebastian-Galles, N. (2001). A robust method to study stress-deafness. *Journal of the Acoustical Society of America*, 110(3), pp. 1606–1618.
- Flege, J. E., Schirru, C. & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40(4), pp. 467–491.
- Fougeron, C. & Smith, C. L. (1993). Illustrations of the IPA: French. *Journal of the International Phonetic Association*, 23(2) pp. 73–76.
- Fraser, H. (2006). Helping teachers help students with pronunciation: A cognitive approach. *Prospect: An Australian Journal of TESOL*, 21/1, pp. 80–95.
- Frost, D. (2011). Stress and cues to relative prominence in English and French: A perceptual study. *Journal of the International Phonetic Association*, 41(1), pp. 67–84.
- Frost, D. & O'Donnell, J. (2013). Combatting the “can't do mentality”: Expert, peer & self-assessment in a French university context. In: Colpaert, J., Simons, M., Aerts, A. & Oberhofer, M. (Eds.), *Proceedings of the 2nd International Conference “Language Testing in Europe: Time for a New Framework?”*, pp. 104–109. Antwerp: University of Antwerp.
- Frost, D. & Picavet, F. (2014). Putting prosody first – some practical solutions to a perennial problem: The Innovalangues project. *Research in Language*, 12/2, pp. 1–11.
- Frost, D. & Guy, R. (Forthcoming). L'innovation est le ton qui fait la chanson dans le secteur LANSAD: musique et prosodie dans le projet Innovalangues. *Recherche et pratiques pédagogiques en langues de spécialité – Cahiers de l'APLIUT*.

- Fulcher, G. (2004). Are Europe's tests being built on an 'unsafe' framework? *Guardian weekly*, 18 March, 2004. Retrieved on September 15, 2015, from <http://www.theguardian.com/education/2004/mar/18/tefl2>
- Gabriel, C. & Kireva, E. (2014). Prosodic transfer in learner and contrast varieties. Speech rhythm and intonation of Buenos Aires Spanish and L2 Castillian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition*, 36(2), pp. 257–281.
- Gervain, J. & Werker, J. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, 4, pp. 1–6.
- Gilbert, J. (2008). *Teaching Pronunciation Using the Prosody Pyramid*. Cambridge: Cambridge University Press.
- . (2010). Pronunciation as orphan: What can be done? *As We Speak (TESOL Newsletter)*, 7(2) pp. 1–9.
- Goullier, F. (2005). *Les outils du Conseil de l'Europe en classe de langue. Cadre européen commun et Portfolios*. Paris: Didier.
- Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27, pp. 315–330.
- Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38/2, pp. 201–223.
- Harding, L. (2012). Nateness or intelligibility: Locating the construct in pronunciation scales (roundtable presentation). In: *Proceedings of the 4th Annual Pronunciation in Second Language Learning and Teaching conference*. Vancouver.
- . (2013). Investigating the construct underlying the phonological control scale. Presented at *Language Testing in Europe: Time for a new framework?* Antwerp: University of Antwerp.
- Henderson, A. (2008). Towards intelligibility: Designing short pronunciation courses for advanced field experts. *ASp*, 53/54, pp. 89–110.
- Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U. & Curnick, L. (2012). The English Teaching in Europe Survey: Selected results. *Research in Language*, 10(1), pp. 5–28.
- Herry, N., Nishinuma, Y., & Ghio A. (2003). Auto-apprentissage de la prosodie anglaise. In: Desmoulins, C., Marquet, P. & Nissen, E. (Eds.), *EIAH 2003 Environnements Informatiques pour l'Apprentissage Humain, Actes de la conférence*, pp. 113–116. Retrieved on September 15, 2015, from <http://hal.archives-ouvertes.fr/hal-00136760/en/>
- Hickey, R. (2004). *A Sound Atlas of Irish English*. Berlin/New York: Mouton de Gruyter.

- Hirst, D. & Ding, H. (2015). Using melody metrics to compare English speech read by native speakers and by L2 Chinese speakers from Shanghai. In: *Proceedings of Interspeech*.
- Honikman, B. (1964). Articulatory settings. In: Abercrombie, D., Fry, D. B., MacCarthy, P. A. D., Scott, N. C. & Trim, J. L. M. (Eds.), *In Honour of Daniel Jones*, pp. 73–84. London: Longman.
- Horner, D. (2013). Towards a new phonological grid. In: Galaczi, E. & Weir, C. (Eds.), *Exploring Language Frameworks: Proceedings of the ALTE Krakow Conference, Studies in Language Testing*, pp. 227–251. Cambridge: Cambridge University Press.
- . (2014). And what about *Testing* pronunciation? A critical look at the CEFR Pronunciation Grid and a proposal for improvement. In: van der Doel, R. & Rupp, L. (Eds.), *Pronunciation Matters*, pp. 109–124. Amsterdam: VU University Press.
- Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent*. Ph.D. dissertation, Institut für Maschinelle Sprachverarbeitung, University of Stuttgart.
- Jun, S.-A. & Fougeron, C. (1995). The accentual phrase and the prosodic structure of French. In: *Proceedings of the 13th International Congress of Phonetic Sciences, vol. 2*, pp. 722–725.
- Jun, S.-A. & Fougeron, C. (2000). A phonological model of French intonation. In: Botinis, A. (Ed.), *Intonation: Analysis, modeling and technology*, pp. 209–242. Dordrecht: Kluwer.
- Kerswill, P. (2006). RP, Standard English and the standard/non-standard relationship. In: Britain, D. (Ed.), *Language in the British Isles*, pp. 34–51. Cambridge: Cambridge University Press.
- Kjellin, O. (1999a). Accent addition: Prosody and perception facilitates second language learning. In: Fujimura, O., Joseph, B. D. & Palek, B. (Eds.), *Proceedings of LP'98 (Linguistics and Phonetics Conference), vol. 2*, pp. 373–398. Prague: The Karolinum Press.
- Kjellin, O. (1999b). Five cornerstones for second-language acquisition – the neurophysiological opportunist's way. Unpublished, pp. 1–9.
- Krashen S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.
- Lacheret-Dujour, A. & Beaugendre, F. (1999). *La prosodie du français*. Paris: CNRS.
- Lado, R. (1957). *Linguistics Across Cultures*. Ann Arbor: Michigan University Press.

- Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
- MacWhinney, B. (2008). A Unified Model. In: Robinson, P. & Ellis, N. (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, pp. 341–371. Abingdon: Routledge.
- Macré, N. (2015). Situation paradoxale ou *catch 22*: le cas d'un dispositif hybride mis en place pour des étudiants LANSAD de niveau A2/B1 en anglais en première année d'université, *Recherche et pratiques pédagogiques en langues de spécialité*, 34(1), pp. 106–126.
- Major, R. (2008). Transfer in second language phonology: A review. In: Edwards, J. G. & Zampini, M. (Eds.), *Phonology and Second Language Acquisition*, pp. 63–94. Amsterdam: John Benjamins.
- Maspero, M. & Quintin, J-J. (2014). L'innovation selon Innovalangues. *LEND - Lingua e Nuova Didattica* 1, pp. 6–15.
- Messum, P. (2009). Grounding stress in expiratory activity. *Speak Out*, 41, pp. 12–15.
- Munro, M. J. & Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching*, 44/3, pp. 316–327.
- Munro, M. J. & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century. A point of view. *Journal of Second Language Pronunciation*, 1, pp. 11–42.
- Murphy, J. (2004). Attending to word stress while learning new vocabulary. *English for Specific Purposes*, 23, pp. 67–83.
- Peperkamp, S. & Dupoux, E. (2002). A typological study of stress 'deafness'. In: Gussenhoven, C. & Warner, N. (Eds.), *Laboratory Phonology* 7, pp. 203–240. Berlin: Mouton de Gruyter.
- Peppé, S. & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics and Phonetics*, 17, pp. 345–354.
- Pike, K. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- Piske, T. (2012). Factors affecting the perception and production of L2 prosody: Research results and their implications for the teaching of foreign languages. In: Romero-Trillo, J.(Ed.), *Pragmatics, Prosody and English Language Teaching*, pp. 41–59. New York: Springer.
- Polivanov E. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4, pp. 79–96.
- Rossi, M. (1979). Le français, langue sans accent? In: Fónagy, I. & Léon, P. (Eds.), *L'accent en français contemporain*, pp. 93–106. Paris: Didier.

- Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, 46/4, pp. 842–854.
- Scarcella, R. & Oxford, R. (1994). Second language pronunciation: State of the art in instruction. *System*, 22(2), pp. 221–230.
- Selinker L. (1972). Interlanguage. *IRAL*, 10, pp. 209–231.
- Soulaine, S. (2014). Le défi d’une transposition didactique inédite: création d’un gestographe. *E-CRINI*, 6, pp. 1–21.
- Stenton, A.J. (2011). Managing the monolingual mindset. SWANS: An authoring system for raising awareness of L2 lexical stress patterns and for inhibiting mother-tongue interference. *ReCALL*, 18, pp. 75–85.
- Taillefer, G. (2002). L’anglais dans les formations spécialisées à l’Université: un cheveu sur la soupe? Peut-on rendre le plat plus appétissant? *Asp*, 37–38, pp. 155–172.
- . (2007). Le défi culturel de la mise en œuvre du Cadre européen commun de référence pour les langues: implications pour l’enseignement supérieur français. *Les Cahiers de l’APLIUT*, 16/2, pp. 33–49.
- Trubetzkoy, N. S. (1939). Grundzüge der Phonologie. *Travaux du cercle linguistique de Prague* 7.
- Weinreich U. (1953). *Languages in Contact, Findings and Problems*. New York: Linguistic Circle of New York.
- Wenk, B. J. & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10, pp. 193–216.
- Zielinski, B. (2006). The intelligibility cocktail: An interaction between speaker and listener ingredients. *Prospect: An Australian Journal of TESOL*, 21/1, pp. 22–45.

Appendix 1: Prosody descriptors (V15)

	RHYTHM & STRESS	INTONATION
Level	<p>Is the correct syllable stressed & marked correctly? (word stress & focus, i.e. nuclear and contrastive stress) The stressed syllable should be:</p> <ul style="list-style-type: none"> • higher (there is some pitch change) • louder • longer • pronounced more clearly / "correctly" 	<p>Intonation</p> <ul style="list-style-type: none"> • Are the intonation patterns varied and appropriate to the speaker's intentions? • Is the range clearly marked or is the intonation flat? <p>NB1: Monologues may not provide opportunities for different patterns. NB2: Some younger native speakers use almost exclusively HRT (High rising Terminal). This is also true of learners exposed to lots of recent English, including film & TV series.</p>
C2	Can place and mark word stress & nuclear stress at will using all 4 cues without disrupting flow. Errors are extremely rare.	Can produce native level & natural-sounding intonation patterns appropriately, including for attitudes, emotions, humour, etc.
C1	Can almost always correctly place word stress and nuclear stress. Can clearly mark stress with all 4 cues. Errors are rare.	Can produce nearly all natural-sounding intonation patterns appropriately, even for attitudes, emotions, humour, etc.
B2	Can correctly place both word stress and nuclear stress nearly all of the time using all 4 cues to varying degrees.	Can produce a wide variety of appropriate patterns, including some attitudes, emotions, humour, etc. Can often produce a good range between high and low tones.
B1	Can correctly place word stress most of the time. Can correctly place nuclear stress most of the time, especially in shorter sentences. Can use the 4 cues but not consistently.	Can produce a variety of appropriate patterns, including the more obvious attitudes, emotions, etc., Can sometimes produce higher tones.
A2	Can usually place word stress correctly. Nuclear stress is placed correctly some of the time. Occasionally uses all 4 cues together.	Can produce some appropriate patterns, especially on shorter sentences. Range between low and high is minimal.
A1	Can audibly place stress on isolated words or short sentences only. Limited control of the cues which mark stress so the stressed syllable is usually difficult for the listener to identify.	Can occasionally produce appropriate intonation patterns on short learnt phrases (e.g. greetings). Range between low and high is minimal.

SOUNDS	
Level	<p>Reduced syllables <u>Reduced syllables are usually:</u></p> <ul style="list-style-type: none"> • less high • less loud • shorter • pronounced less clearly / "correctly" <p>E.g. schwa /ə/, (doct<u>or</u>); /t/, short final /t/, etc. (happ<u>y</u>); syllabic /n/ & /l/</p>
C2	<p>Can reduce the full range of forms, including syllabic /n/ & /l/, etc. resulting in a natural-sounding alternation of strong & weak forms. Strong & weak alternation is evident most of the time.</p>
C1	Can reduce nearly the full range of forms. Strong & weak alternation is evident most of the time.
B2	Can partially or fully reduce most possible reductions. Strong & weak alternation is often evident.
B1	Can partially reduce half of all reduced syllables. Some full reductions, especially schwa. Strong & weak alternation is evident.
A2	Can produce a few partially reduced syllables, mainly schwas.
A1	Can rarely if ever reduce syllables; reduced syllables are usually pronounced the same as stressed syllables.
	<p>Stressed & unreduced vowels</p> <ul style="list-style-type: none"> • Are the stressed & unreduced vowels closer to English or closer to the speaker's native language?
	<p>Connected speech (phonotactics)</p> <ul style="list-style-type: none"> • Contractions ('ll, it's, gonna etc.) • Linking (e.g. an egg) • Linking with /j/ & /w/ (e.g. go-<u>u</u>-way) • Deletion of final /t/ & /d/, etc., (e.g. first question) • Assimilation (e.g. Breaking Bad > Breakingbad) • Geminates (e.g. keep <u>pl</u>aying) • Etc. <p>NB: Higher level speakers will pause between sentences and tone</p> <p>Can produce native speaker level connected speech phenomena making for a smooth and natural sounding flow. Can use more or fewer connected speech phenomena according to register, speed, etc.</p> <p>Can produce mainly smooth and natural-sounding connected speech. Nearly the full range of phenomena, etc., but the occasional missed opportunity, especially in longer sentences. Hesitations are common.</p> <p>Can produce a variety of phenomena, including assimilation & deletion well over half of the time, especially in shorter sentences without many hesitations. Can produce fluent stretches of less "staccato-sounding" speech with lots of contractions.</p> <p>Can produce about half of all possible linking phenomena, including deleted /t/ and /d/, assimilations, etc., but hesitations are fairly frequent in longer sentences.</p> <p>Can occasionally produce contractions ('ll, 'd, gonna, wanna, etc.)</p> <p>Can link between words less than half the time with word-final consonants and some /j/ & /w/ between words (e.g. go out).</p> <p>Fragment-hesitations</p> <p>Can produce basic and isolated contractions (e.g. I'm, it's, gonna, etc.). Can produce word internal /j/ & /w/ linking (going). Occasional linking between words when a word-final consonant is followed by an initial vowel (e.g. <u>in</u>-<u>ter</u>-est). Very frequent & long hesitations.</p>

SOUNDS	
Level	NB: Higher level speakers will pause between sentences and tone
	<p>Reduced syllables <i>Reduced syllables are usually:</i></p> <ul style="list-style-type: none"> • less high • less loud • shorter • pronounced less clearly / "correctly" <p>E.g. schwa /s/, (doctor); /t/, short final /l/, etc. (happy);</p>
	<p>Stressed & unreduced vowels • Are the stressed & unreduced vowels closer to English or closer to the speaker's native language?</p>
	<p>Connected speech (phonotactics)</p> <ul style="list-style-type: none"> • Contractions ('ll, it's, gonna etc.) • Linking (e.g. an egg) • Linking with /j/ & /w/ (e.g. go-/w/-away) • Deletion of final /t/ & /d/, etc., (e.g. first question) • Assimilation (e.g. Breaking Bad > Breakinbad) • Geminates (e.g. keep playing) • Etc.
C2	+
C1	+
B2	+
B1	+
A2	+
A1	+

Comments:

Appendix 3: Calibration sheet (V11)

TU no	Word stress QTT			Pauses and Tone Units (TUs) (0/1/2) - Y/N QLT			Focus: nuclear & contrastive stress (0/1/2) -			Rhythm (no of beats)/QTT				Reductions (schwa, /l/ & /n/) QTT			Stressed & unreduced vowels (0/1/2) QLT			Connected speech QTT			Intonation (0/1/2) QLT					
	real	notes	max	real	notes	max	real	notes	max	min	max	real	notes	max	real	notes	max	real	notes	max	real	notes	max	real	notes	max		
1	1		2			2					3	4					2			2						2		
2	1		2			2					3	3					4			4						2		
3	0		2			2					1	1					0			2						2		
4	2		2			2					4	5					4			2						2		
5	1		2			2					3	4					4			2						2		
6	0		2			2					1	1					0			2						2		
7	4		2			2					4	6					4			2						2		
8	6		2			2					7	9					8			2						2		
9	6		2			2					7	9					11			2						2		
10	2		2			2					3	3					3			2						2		
11	2		2			2					5	6					5			2						2		
12	0		2			2					1	2					1			2						2		

TU no	Word stress			Pauses and Tone Units			Focus			Rhythm				Reductions			Stressed & unreduced			Connected speech			Intonation					
	max	notes	real	max	notes	real	max	notes	real	min	max	notes	real	max	notes	real	max	notes	real	max	notes	real	max	notes	real	max	notes	
13	2			2			2			2	3			1			2			3			2			2		
14	2			2			8			8	9			7			2			6			2			2		
15	0			2			1			1	2			0			2			2			2			2		
16	6			2			8			8	10			8			2			12			2			2		
17	6			2			7			7	7			10			2			8			2			2		
18	0			2			2			2	4			1			2			1			2			2		
19	1			2			4			4	7			3			2			5			2			2		
20	0			2			1			1	1			1			2			1			2			2		
21	2			2			1			1	2			1			2			1			2			2		
22	0			2			2			2	2			1			2			2			2			2		
23	1			2			3			3	5			2			2			4			2			2		
24	2			2			4			4	5			3			2			5			2			2		
25	0			1			5			5	7			1			2			6			2			2		
Totals:	46			49			90			90	117			85			50			95			50			95		

TU no	Possible Tone Units (TUs). NB - there may be more or fewer depending on speed, stylistic factors, etc.
1	(LF)"The first question which you will ask,
2	(LR) and which I must try to answer,
3	(LF) Is this:
4	(LF) 'What is the use of climbing Mount Everest?'
5	(LR) And my answer must at once be
6	(LF) 'It is no use.
7	(LF) 'There is not the slightest prospect of any gain whatsoever.
8	(FR) Oh, we may learn a little about the behaviour of the human body at high altitudes,
9	(LR) and possibly medical men may turn our observation to some account for the purposes of aviation
10	(LF) But otherwise nothing will come of it.
11	(LR) We shall not bring back a single bit of gold or silver,
12	(LR) not a gem,
13	(LR) nor any coal or iron.
14	(LR) We shall not find a single foot of earth that can be planted with crops to raise food.
15	(RF) So it is no use
16	(LF) If you cannot understand that there is something in man which responds to the challenge of this mountain and goes out to meet it,
17	(LR) that the struggle is the struggle of life itself upward and forever upward,
18	(LF) then you won't see why we go.
19	(LF) What we get from this adventure is just sheer joy.
20	(LR) And joy,
21	(LR) after all
22	(LF) is the end of life.
23	(LR) We don't live to eat and make money.
24	(RF) We eat and make money to be able to live.
25	(RF) That is what life means and what life is for."

Student: _____ **Assessor:** _____

Institution: _____ **Class:** _____ **Date:** _____

CHAPTER THIRTEEN

L2 PRONUNCIATION FEEDBACK IN ENGLISH-FRENCH TANDEM CONVERSATIONS

CÉLINE HORGUES AND SYLWIA SCHEUER¹

13.1 Introduction

13.1.1 Research background

So far, most SLA research on Corrective Feedback (CF) has focused on feedback provided by language teachers in the field of L2 morphosyntax and vocabulary (Lyster & Ranta, 1997; El Tatawy, 2002; Mackey, 2006 and Lyster et al., 2013). A few studies have reported that the language area most frequently targeted by teachers' CF in the classroom was grammar (e.g., Sheen, 2006²). Paradoxically, some other studies have pointed to pronunciation and vocabulary CF being more noticeable for learners than morphosyntactic CF, which was found to be less likely to lead to uptake (Mackey et al., 2000; Lyster, 1998 quoted in Sheen, 2006; Saito & Lyster, 2012). However, the study of the process of providing and receiving pronunciation feedback (CF focus, CF type, learner uptake) has largely been neglected, hence leaving the question of the effectiveness of various pronunciation CF strategies almost unexplored (see for example Saito & Lyster, 2011 and Saito & Lyster, 2012, on the role of recasts in the development of Japanese learners' L2 English).

¹ We are very grateful for all the help this research project received from the SITAF research team, the SITAF participants, our research team Sesylia/Prismes, as well as the funding provided the Ortolang-Ircom programme. This work was partially funded by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

² Sheen 2006 reports that 50% of the recasts observed in their study targeted grammar and that vocabulary recasts were slightly more frequent than pronunciation-focused recasts.

The chapter offers findings from the analysis of corrective feedback given to the native French speakers by their native English-speaking tandem partners as part of the SITAF corpus collected at the University of Paris 3. The corpus, described at length in Horgues & Scheuer (2015), consists of around 25 hours of video-recorded, face-to-face interactions held by 21 pairs of native-French speaking and native-English speaking tandem participants. The participants were all students aged between 17 and 22 and none were balanced English-French bilinguals. The 21 native French-speaking students (coded F01 to F21) were English language specialists for the most part, with a self-assessed level in L2 English of 7.2/10 for their mean proficiency and 6.8/10 for oral expression in particular. The 21 English-speaking students (coded A01 to A21) came from various Anglophone countries (USA, Canada, UK, Ireland) and self-assessed their level in L2 French as 6.9/10 for the mean proficiency and 6.6/10 for oral expression³.

The speakers were recorded on two occasions – in February (session 1) and May 2013 (session 2) – while performing three types of tasks. Two of them were communication activities, *Liar-Liar* (Game 1; storytelling) and *Like Minds* (Game 2; debating), while the last was a monitored reading task. In Game 1, the L2 learner had to tell a story containing three lies that the native-speaking partner had to identify by asking questions. In Game 2, both participants had to give their opinion on a potentially controversial subject before determining the degree of like-mindedness between them. The text used for the monitored reading task was *The North Wind and the Sun* (see Appendix 1). Although all the participants got to perform all three tasks in their respective L1 and L2 at least once during the recording sessions, our analysis will only be concerned with the English portion of the data. We have previously reported on the CF provided by the native speakers (NSs) during the reading task (Horgues & Scheuer, 2014). In this previous study, we found that almost all the pronunciation feedback provided by the NSs during the reading task targeted segmental matters (95%) – predominantly vowels (58%), ahead of consonants and mixed-category segmental errors – and that the preferred CF strategy used by NSs was recast (61%) before explicit comments (26%) and clarification requests or repetition (13%). The present study expands this line of research by offering an analysis of L2 pronunciation feedback given to the native French-speaking partners during the two conversation tasks (story-

³ Out of the 5 self-assessed evaluations (oral expression, oral comprehension, written expression, written comprehension, mean score), oral expression was the only score where the difference between the two speaker groups reached statistical significance.

telling and debating). Reference will also be made to the metadata collected through the questionnaires filled in by the participants on completing the tandem programme (a general questionnaire on their overall tandem experience and another questionnaire specifically focusing on pronunciation issues; see appendices 2 and 3), with a view to comparing their introspective judgements – as CF providers or receivers – with the corrective practices actually observed in the corpus.

13.1.2 Research questions

Drawing on the assets of our corpus (longitudinal face-to-face video recordings of tandem pairs performing semi-spontaneous conversation tasks), we would like to address the following research questions: (1) What gets corrected by the native speaker: segmental or prosodic errors? (2) What is the corrective strategy adopted by the NS: recast, explicit correction, or clarification request⁴? (3) Is the correction solicited by the learner in some way, or is it spontaneously provided by the NS participant? (4) What is the learner's uptake after receiving feedback? (5) How do body gestures supplement both the corrective audio input and the CF request and uptake? Based on informal observation of tandem interactions and on the results reported in the SLA literature regarding the teachers' preferred language areas for CF, one of our overarching research hypotheses is that pronunciation errors are weak magnets for corrective feedback in spontaneous tandem interactions, with a vast majority of CF instances focusing on syntax and vocabulary.

13.2 Method

We use the term 'corrective feedback' to refer to the negative evidence given by the native speaker to their tandem partner during the recorded interactions. Gass (2003: 225) defines negative evidence as "the type of information that is provided to learners concerning the incorrectness of an utterance". In the present analysis, we will be distinguishing three basic categories of CF: explicit comments (the NS provides metalinguistic information explicitly: "you can't say X..."), clarification requests ("what do you mean by X?") and recasts. Lyster & Ranta (1997: 46) define recast as a corrective strategy involving "the teacher's reformulation of all or part of a student's utterance, minus the error"; see also El Tatawy (2002). In our context, this can be illustrated with the following exchange: F06

⁴ These terms will be defined in the methodology section below.

(NNS) “*and I fall*”; A06 (NS) “*Oh, you fell!*”. If one corrective sequence entails different feedback moves (for example recast+explicit correction), we describe it as combined CF (what Sheen, 2006, calls *combination recasts* as part of the multi-move recast). Importantly, we do not make use of two other CF types present in Lyster and Ranta’s (1997) typology: *elicitation* and *repetition of the learner’s error*, which are absent from our peer-to-peer interactions. These CF strategies seem to be restricted to teachers’ corrective style and avoided by tandem participants possibly because they reinforce the asymmetry between the two conversation interactants.

We count corrective feedback as spontaneous if no appeal, be it verbal or non-verbal, is made to the native speaker by the learner. On the other hand, the non-native participant may solicit feedback explicitly or implicitly. In the current study, the former label is applied to cases of explicit *verbal* requests – e.g., “*is it /pra:g/ or /preig/?*” – whereas implicit requests are conceptualised as various types of non-verbal, vocal or visual appeal, such as hesitation marks, unfinished sentences, rising tones, questioning gazes or gestures, etc.

Finally, learner uptake is understood following Lyster & Ranta’s (1997: 49) definition as “a student’s utterance that immediately follows the teacher’s feedback and that constitutes a reaction in some way” to that feedback. We distinguish between (a) total uptake, characterised by (reasonable) conformity to the model provided by the NS expert (‘successful uptake’ or ‘repair’ for Sheen, 2006); (b) partial uptake, where only part of the correction has been implemented by the learner (‘partial repair’ target in Lyster & Ranta, 1997); (c) failed uptake, where the NNS attempts but fails to repeat the model form (e.g., by repeating the initial error, ‘off target’ in Lyster & Ranta, 1997) and (d) no uptake, in which case there is no observable reaction whatsoever to the CF and the NNS continues on the same topic / proceeds to a topic shift. We also use the label ‘no uptake’ to refer to cases where the learner simply acknowledges the NS’s contribution through minimal verbal back channelling (“*yes*”, “*okay*”, “*hm hm*”). Our use of *uptake* therefore relates to any attempt by the learner at modifying their output in response to the CF provided by the NS. When uptake is not totally successful, we have deemed it helpful to distinguish three situations: partially successful attempt, failed attempt or no attempt at modifying one’s output. It is noteworthy that Lyster and Ranta (1997) and Sheen (2006) do not make this distinction and use the umbrella label: *needs repair* instead. In contrast to Sheen (2006), we have also decided to exclude acknowledgment markers from uptake results as it is impossible to state whether they actually represent a learner’s response

to the CF provided by their interlocutor or if they simply pertain to the discursive need of keeping the conversation going after what was interpreted as a confirmation check. Since the above-quoted studies focused on a communicative setting markedly different from ours (institutional classroom instruction vs. spontaneous peer-to-peer conversation), we considered it necessary to fine tune the typology of uptake categories in order to better suit the specificities of the tandem context and, consequently, to allow for a more insightful analysis of the data.

13.3 Results

13.3.1 General CF findings

We analysed seven hours of video-recorded interactional speech (Game 1 and Game 2 in both recording sessions) and identified a total of 158 instances of corrective feedback. In accordance with our initial hypothesis, pronunciation did not constitute the primary target of native speakers' interventions, accounting for 27.8% (44) of all CF instances at best (i.e., when combinations of foci – such as grammar/syntax and phonetics – are taken into account), and just 19.6% if we consider pronunciation alone. The favourite area targeted by the experts in our study was vocabulary with 52.5% of all cases (e.g., A01 needing to explain the term *a psychic*), while pure grammatical (morphosyntactic) errors only accounted for just under 13% of all CF occurrences (“[F08]: *why we had to get off the sea, to get out...?* [A08]: *get out of the sea*”), with the remainder split between the various mixed-focus categories (“[F12]: *some are just in jail because they...they are thief so stuff like that,* [A12]: *hm hm they are thieves yeah.*”) (see Fig. 13-1).

The mixed-focus instances naturally pose methodological problems when it comes to establishing the motives behind a native speaker's corrective behaviour. When the learner's utterance is incorrect in more ways than one and is subsequently repaired by the expert by means of recast, the exact reasons for the CF intervention are to a large extent a matter of speculation. The case of a French participant using the expression *'interior in leather *['li:ðər]* illustrates this point. Her American partner recasts it as *'leather interior'*, insisting on the correct word order and praising her for eventually getting it right (and resuming the flow of the conversation). However, the French speaker persisted in her erroneous rendition of the vowel: *'leather *['li:ðər] interior'*. We still classify such cases as 'mixed focus' and include them in our subsequent

analysis of phonetic CF, even though there is no way of knowing whether the pronunciation issue – had it stood alone – would have merited the NS's reaction at all.

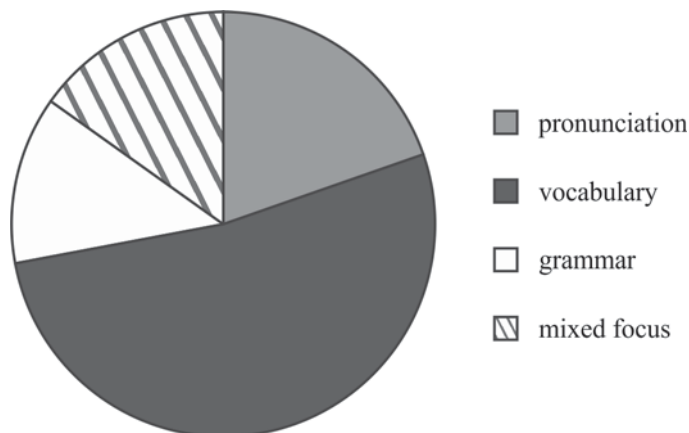


Fig. 13-1. Distribution of CF focus across the 158 corrective instances.

The predominance of vocabulary as the focus of the English NSs' corrective interventions is in line with their declared CF goals, as per the post-recording questionnaires filled in by all participants (Appendix 2, question 10). 17 out of the 21 Anglophones claimed to correct lexical mistakes 'always' or 'almost always'. For comparison, grammar got 13 hits in these frequency categories whereas pronunciation only got 9⁵. Interestingly, this distribution did not quite tally with the impressions of their French tandem partners, who thought to be corrected 'always' or 'almost always' more or less across the board, casting 15 votes for vocabulary and 12 each for grammar and pronunciation.

Game 1 (storytelling) attracted 2.6 times more corrective instances than Game 2 (debating): 114 and 44, respectively. The same proportion is found in the case of phonetic CF considered alone (discussed in section 13.3.2 below): 32 vs. 12 occurrences. This difference is hardly surprising,

⁵ In the general questionnaire in appendix 2, question 10, the answers of the 21 NSs regarding the frequency with which they corrected each language field were split as follows: vocabulary [always: 8, almost always: 9, sometimes: 2, often: 2], grammar [always: 6, almost always: 7, sometimes: 2, often: 4, almost never: 2] and pronunciation [always: 2, almost always: 7, sometimes: 7, often: 3, almost never: 2].

bearing in mind that the aim of the storytelling task (the *Liar, Liar* game) was to identify three lies incorporated by the NNS partner into their story, thus making the native-speaking participant focus more on their interlocutor's discourse than on their own. In the debating game, however, there was a better balance between both interlocutors' contributions to the discussion.

13.3.2 Phonetic CF

Segmental or prosodic matters?

Unlike the reading task (analysed in Horgues & Scheuer, 2014), where only a tiny minority – 4.6% – of CF instances regarded suprasegmental matters (predominantly word stress), the conversation data presents a less unbalanced picture. 'Pure' segmental errors appear to have acted as major triggers in 29 cases (65.9%), with the remaining 15 – three of which had a potential segmental overlay – divided between word stress (12, or 27.3%, e.g., *'prisoners* being incorrectly stressed as **pri'soners* or *Fer'rari* mispronounced **Ferra'ri*) and syllable count, i.e., the learner adding or 'losing' a syllable (3, or 6.8%, e.g. *cluedo* mispronounced with an extra medial syllable: **clu-e-do* or *tu'ition* being pronounced as two syllables only **tui-tion*). This drop – i.e., between the reading and the speaking data – in the share of segmental corrections in the overall CF jigsaw is highly significant ($p < .0005$).

The consequent growth in relative importance of lexical stress between reading and spontaneous speech deserves further consideration. To begin with, the two speaking tasks and corresponding speaking styles might also entail a difference in the segmental vs. suprasegmental complexity of the speech material for French learners of L2 English, which will need to be further explored. The nature and format of the reading text *The North Wind and the Sun* might also have an impact on the quantity and focus of the CF provided: it is mainly made up of monosyllabic words⁶ and was initially devised – and is still mostly used – to exemplify segmental (particularly vocalic) differences between different dialectal varieties of English.

Moreover, speech characteristics showing high potential for leading to unintelligibility may be more of a distraction in the case of spontaneous L2 discourse than read speech, which is arguably easier to process by virtue

⁶ Monosyllabic words represent 83 % of the text, which only contains five different words of three syllables or more: *disputing*, *traveller*, *succeeded*, *considered*, *immediately*.

of being grammatically correct and thus more predictable. This might explain why mistakes involving lexical stress were relatively powerful magnets for corrective feedback in the conversational exchanges, since, according to Roach (2009: 79), “incorrect stress placement is a major cause of intelligibility problems for foreign learners, and is therefore a subject that needs to be treated very seriously”. In a similar vein, Cruttenden (2001: 235) declares that in English, where there is no word accent pattern, “the differing accentual patterns of words are as important to their recognition as is the sequence of phonemes”, which would indeed make such deviations well worthy of the listener’s attention.⁷

Again in contrast to our 2014 reading study, vocalic errors no longer seem to be such glaringly obvious culprits: 44.8% of all 29 segmental corrections might be attributed to vowels (e.g., *meadow* pronounced with an [i:], 38% to consonants (e.g., *sixth* pronounced as if it was **thixth*), and the remaining 17.2% represent a mixed V+C category (e.g., *hotel* pronounced **[o'tel]*). The respective figures in our reading findings, as regards types of segmental CF, were 58.3%, 25.2% and 16.5%. Although the differences between the vocalic and consonantal CF contributions are statistically non-significant, they may still be indicative of certain broad tendencies. In this context, it is worthy of note that ‘vowels’ was the top pick among the Anglophone participants when they were asked to specify in retrospect which aspects of their partners’ pronunciation they believed they had corrected (18 ‘yes’ answers in the pronunciation questionnaire, question 6 – see Appendix 2), followed by consonants (11) and word stress (10), while intonation and rhythm only scored (5) and (2), respectively.

On the other hand, potentially conflicting forces might be at work in the vowel vs. consonant CF competition. Much as the native participants might have wished to prioritise their partners’ inaccurate vowel renditions in their corrective endeavours, they could have been expected to instinctively show more tolerance towards vocalic rather than consonantal errors⁸, since, as Ladefoged (2001: 28) points out, “accents of English

⁷ The importance of word stress placement is certainly not limited to its potential for mis- or lack of understanding, but is also due to extralinguistic considerations, as evidenced by the following comment found in one of the respectable on-line resources for learning and teaching English: “Even if the speaker can be understood, mistakes with word stress can make the listener feel irritated, or perhaps even amused, and could prevent good communication from taking place” (Teaching English, <https://www.teachingenglish.org.uk/article/word-stress>).

⁸ Saito & Lyster (2012: 396) refer to the same characteristics of the English vowel inventories resulting in a limited communicative saliency of vowels: “L2

differ more in their use of vowels than in their use of consonants”. This variability seems to be in line with the universal tendency whereby “[v]owel systems vary greatly in their complexity from language to language”, although in the case of English the situation may be argued to be compounded by the fact that the language “happens to be relatively rich in vowel contrasts, with the added complexity that the vowel system is by no means uniform across the English-speaking world” (Clark & Yallop, 1995: 29). Specifically in the context of learner pronunciation and Minimum General Intelligibility as a possible performance target, Cruttenden (2001: 309) identifies this relative instability of vowels as a factor diminishing their importance in L2 speech, noting that “the major forms of native-speaker English exhibit considerable homogeneity in their consonant systems, which offer no further possibilities of simplification if they are to retain some resemblance to a natural system. Most of the simplification will therefore be in the vowel system”. This also ties in with Jenkins’s proposal for a minimal set of features claimed to be essential to safeguarding international intelligibility, presented in her influential publications on English as a Lingua Franca (ELF) communication (e.g., 2000, 2007). Her findings seem to point to the crucial importance of accurate rendition of consonant – but not vowel – quality. A case in point is the following quote from a pronunciation instructor interviewed as part of Saito & Lyster’s study (2012: 395): “Whereas /r/ is something that is salient to me, I could not distinguish the /æ/ sound instantly. It is hard to perceive. English /r/-/l/ interferes with communication more. Native speakers would still understand vowel difference”.

Corrective strategy

Just like in our reading study, recast proved by far the predominant strategy employed in our peer-to-peer interactions. 37 of the 44 cases (84%) involved this method, although 8 of those featured recast combined with another strategy. Still, recast in its ‘pure’ form accounted for nearly 66% of all phonetic CF instances. The runner-up was clarification request, which was found in 12 cases (of which 5 were in combination with recast, as in “*Who? Prisoners?*” asked A03), whereas explicit correction was employed merely 3 times, always accompanied by recast as in the following exchange: “[F17]: *we could erm... swim in the river... and all, so, and sunbath, so it was really... cool*; [A17]: *sunbathe*; [F17]: *sunbathe*;

pronunciation research has shown that vowel inventories dramatically differ between regional dialects of English.”

[A17]: *bathe, yeah*; [F17]: *ok, cause you say bath*, [A17]: *right, err, bath, but then to bathe oneself, so sunbathe*".

The special place of recast in our analyses merits a few comments. In the SLA literature it has been reported as the predominantly preferred corrective strategy used by language teachers in various teaching backgrounds (e.g., Lyster et al., 2013). Its numerical dominance in our tandem interactions, too, was only to be expected: by its very nature, a recast is indirect and non-threatening, and therefore ideally suited for the type of peer-to-peer interaction where neither party particularly wishes to emphasize their superior position. Moreover, recasts are also minimally disruptive or obtrusive⁹ in this context where the tandem partners tend to focus on smooth and friendly communication and task completion, rather than on language accuracy¹⁰. This observation echoes the claims made by various researchers that, in a more conventional SLA setting, "recasts serve an ideal pedagogical function, arguably because they enable teachers to implicitly draw students' attention to the accurate use of language without interrupting the flow of classroom discourse" (Saito & Lyster, 2012: 387; also Long, 2007).

The lowly position of the other corrective strategies, notably that of explicit correction, may partly be due to more mundane causes. The participants might not have always been in a position to provide an explanation – especially in the case of sub-phonemic articulatory inaccuracies – by virtue of lacking the necessary background knowledge of their mother tongue, or even descriptive linguistic terms in general, which could inform their explicit pronunciation feedback. The problem would be aggravated in the case of vowels, which, unlike consonants, do not lend themselves to neat articulatory descriptions and may therefore be more difficult to pinpoint, thus leaving NSs at a loss to explain the nature of the error committed by their partners, or to suggest an efficient way of rectifying it.

However, opting for recast as the corrective strategy presupposes the ability – on the part of the corrector providing the recast – to still understand

⁹ Sheen (2006: 364): "It is arguably more appropriate to think of recasts as themselves constituting an implicit/explicit continuum, as reflected in Doughty and William's (1998) 'unobtrusiveness/obtrusiveness continuum'."

¹⁰ The metadata gathered through the questionnaires tends to confirm that language accuracy is indeed not the top priority here: 19 out of our 21 Anglophones chose 'not wanting to interrupt the flow of their ideas' as a reason they had *not* corrected their French interlocutors (question 7 in the questionnaire, Appendix 3).

their interlocutor's overall message¹¹ and thus to retrieve a possible interpretation of an erroneous utterance. Consequently, one could hypothesise that when recast – in spite of its obvious virtues – is *not* employed, this is quite possibly because the native listener did not comprehend the L2 discourse. Being able to establish which types of error tend to make the listener resort to a clarification request (hence implying incomprehension) would be precious from the point of view of L2 pedagogy. At this early stage of our corpus analysis we must limit ourselves to the general observation that word stress mistakes triggered a clarification request in nearly half of the cases (5 out of 12), which may indeed point to their high capacity for miscommunication.

In view of the above considerations, it is remarkable that Jenkins not only fails to include word stress on her list of core features, but even goes as far as branding it as unteachable and potentially reducing flexibility. Word stress – in contrast to tonic/nuclear stress – therefore figures among those non-core features that, based on her analysis of extensive miscommunication data, did not prove necessary for successful interactions among NNSs of English. On the other hand, it has to be borne in mind that the conversational exchanges discussed in the present contribution do not fall under the category of ELF communication, so one cannot exclude the possibility that native English speakers, such as the ones featuring in our study, show much lower tolerance towards this type of pronunciation error than their non-native counterparts.

Solicited or spontaneous?

Phonetic feedback was solicited roughly as often as it was not (23 vs. 21 cases, respectively). When some sort of appeal to the native speaker did occur, it was predominantly implicit, usually executed through prolonged gazing at their interlocutor, hesitating tempo and rising tone. Only 13.6% (6) of CF instances followed an explicit verbal request on the part of the learner, as in: “[F05] *if you are in jail for perpetuity...* [A05]: *what?* [F05]: *I don't know if you can say that, in French perpétuité?, you're gonna s-you're gonna die in prison* [A05]: *oh for ever, perpetuity*”.

¹¹ Sheen (2006: 365) refers to comprehensibility as a prerequisite for a teacher's recast: “There is also the conundrum of how a teacher can properly reformulate a learner's utterance unless he or she understands the learner's communicative intentions although teachers become ‘interlanguage experts’ in understanding the meaning of their students' utterances”.

Learner uptake

Of the four options we considered, ‘no uptake’ turned out to be the most frequent one: it followed 24 (54.5%) out of the 44 CF occurrences. Very often, the learner did not show any visible or audible reaction to the CF and went on with their discourse¹². If uptake did occur, however, it was predominantly ‘total’ (12, or 27.3%), whereas only 3 cases (6.8%) were labelled as ‘failed uptake’. This relative scarcity of uptake of any kind (45.4%) is perhaps less surprising than it might initially appear: after all, the majority of corrections were carried out by means of recast, which – by its very nature – is non-explicit and therefore often too discreet to be appreciated as such. Consequently, the corrective function of recasts is sometimes not perceived by the recipient, especially if more than one item is corrected at a time (e.g. the inflectional ending *and* the stressed vowel of a verb, as in a NS recasting ‘*he sit* *['si:t]’ as ‘*he sits*’). This result is in line with previous research pointing to the limited effectiveness of recasts for L2 learners who sometimes fail to grasp its didactic function in the classroom, too (El Tatawy, 2002 and Lyster et al., 2013). A good illustration of this problem is also provided by an extract from Game 1 where native English speaker A11 reacts to his partner’s pronunciation of the ‘t’ in *castle* with “*in a castle you said?*”. French speaker F11 fails to notice the corrective function of this intervention pointing to the silent ‘t’ and goes on to repeat her mistake (“*yes, a casTle*”), presumably interpreting the NS’s utterance as a genuine question or a confirmation check – a valid assumption in the context of an activity whose aim was to establish which details of your partner’s story were not true. The question of the saliency or the ambiguity of recasts for learners has recently been explored by other researchers (Sheen, 2006; Saito & Lyster, 2011 and 2012) and further variables will need to be taken into consideration to account for the saliency of recasts such as: i) the intonation/mode of recasts (interrogative *vs.* declarative recasts having been shown to have a clearer corrective function for learners – Sheen, 2006), ii) the length and format of the recast – reduced recast/isolated word/embedded recasts, one change/multiple changes, and iii) the type of modification concerned (substitution *vs.* deletion/addition).

¹² E.g., “[F01]: *We played a game, it was called the cluedo I think; [A01] (frowning, looking sideways to figure out which word her tandem partner intended to use): oh cluedo! I have heard of it, I have never played; [F01]: yeah, yeah and we did that.*”.

Multimodality

All three stages of corrective feedback (i.e., CF request, provision and uptake) were found to be highly multimodal in our SITAF corpus (Debras, Horgues & Scheuer, 2015). NSs tend to provide phonetic feedback combining visual cues to support their verbal content and learners also frequently rely on non-verbal strategies when attempting to take up phonetic CF (face movements, hand gestures and visual alignment with the expert's articulatory movements). For example, when A11 corrected his partner's (F11) pronunciation of *geese* (which she mispronounced as the singular form *goose*):

A11: *yeah, you can say for... erm... if there's more than one goose, they're geese*¹³ (see Fig. 13-2)

F11: *geese!* (see Fig. 13-3)

A11: *ok changes to e-e- in the middle*¹⁴

F11: *ok yeah*¹⁵, *so, geese* (Fig. 13-4)

A11: *ok, so there were geese, ok*¹⁶

F11: *geese, hens...*

Her stretched lips and the visible articulatory tension she performs clearly reveal her intention of mirroring the NS's model pronunciation of long /i:/ in *geese* through visual alignment (see Du Bois, 2007, on multimodal alignment).

Interestingly, the role of multimodal resources to support feedback has now become the centre of attention in CF research, as for example the contribution of mutual eye-gaze to the effectiveness of recasts (McDonough et al., 2015). Along with suprasegmental aspects (declarative mode and emphatic stress on key items, as suggested in Sheen, 2006), visual cues (eye-gaze, face movements, hand gestures) participate in enhancing the saliency of CF for the NNS learner.

¹³ He accompanies this explanation with a metalinguistic or pragmatic gesture (framing the word item) in the upward direction as if to offer his proposition.

¹⁴ He accompanies this metalinguistic explanation with a pointing gesture referring to the spelling form.

¹⁵ F11 replicates a similar type of pointing gesture symbolizing spelling.

¹⁶ A11 supports his final feedback with an interpersonal (or interactive) gesture in the direction of his tandem partner. For the typology of gestures analyzed in the SITAF corpus see Debras, Horgues & Scheuer, 2015.



Fig. 13-2. *Yeah, if there is more than one goose they're GEESE.*



Fig. 13-3. *F11: GEESE!*



Fig. 13-4. *F11: oh yeah, so, GEESE.*

13.4 Discussion and conclusions

We hope that our present contribution brings a new insight into the study of CF by showing how peer-to-peer tandem interactions induce a favourable environment for L2 learners not only to be exposed to valuable phonetic feedback provided by their native-speaking counterparts (positive evidence), but also to be proactive in soliciting this feedback (negative evidence), which is essential for their L2 phonetic development. Most SLA studies seem to present the L2 learner as a mere *recipient* of pronunciation CF provided by the teacher but our study shows how L2 learners can also play an active role in the CF process.

In addition, the negotiation for meaning occurring between the tandem partners is framed in the context of a fairly authentic and communicatively meaningful environment rather than during decontextualized instruction, which again is a facilitating factor for the development of the L2 sound system (Sheen, 2006¹⁷; Saito & Lyster, 2011 and 2012).

The quantity, type, form and impact of the phonetic feedback are partly dependent on factors such as the speaking task and corresponding speaking style (read speech *vs.* spontaneous speech), the instructions given to the participants and the interlocutors' learning and collaborative profiles. A more thorough investigation of which phonetic errors tend to be corrected, how and why, would be necessary. It would be interesting to study to what extent the CF strategy chosen by the NS is determined by the CF focus entailed by the learner's errors, and/or whether the quality of learner uptake is linked to the type of CF strategy used by the NS or the CF focus concerned. However, several methodological caveats will have to be considered: it is sometimes difficult to identify the corrective focus intended by a NS's correction¹⁸, and to have access to their reasons for deciding to intervene or not (except through *a posteriori* introspective recall¹⁹) but also to assess the short-term and long-term efficacy of

¹⁷ Sheen (2006: 364): "Long (1996) contends that recasts which arise naturally from negotiating for meaning create opportunities for acquisition".

¹⁸ See also Sheen's (2006: 365) remark about the difficulty of identifying the focus of teachers' recasts: "a speaker's intention cannot easily be determined unless there is an obvious verbal communication breakdown in the interaction".

¹⁹ The reasons the 21 native-English speakers mentioned for deciding *not* to correct their partner's errors were: Their mistakes are too small (20), They don't want to interrupt the flow of their ideas (19), They can understand them despite their mistakes (15), They don't want to make their partner feel uncomfortable (9), They don't think it's polite to correct or to interrupt their partner (9), It would be impossible to correct all their mistakes (5), They can't hear their mistakes (3),

pronunciation CF in L2 acquisition. In a further perspective, we would like to take a closer look at the details of the timing of pronunciation CF, i.e., how the CF episode develops in time: in fact, the CF provision by the NS and CF uptake by the NNS are not always immediate and it might be enlightening to study their delayed effects, too. Indeed, a NS might decide not to intervene right after the NNS's error and wait until the end of their turn or a part of the conversation to give feedback. Likewise, a NNS might not immediately repair their error (and show no immediately observable uptake) but might still notice the mismatch between their initial output and the negative evidence of the NS's CF, which, following Schmidt's (1990) Noticing Hypothesis²⁰, would still contribute to their L2 phonological development.

References

- Asari, Y. (2013). The relationship between the timing of recasts, repair and learners' proficiency levels: analysis of NS-NNS recast episodes. *Pan-Pacific Association of Applied Linguistics*, 16(2), pp. 21–34.
- Clark, J. & Yallop, C. (1995). *An Introduction to Phonetics and Phonology*. Oxford and Cambridge, MA: Blackwell.
- Cruttenden, A. (2001). *Gimson's Pronunciation of English*. London: Arnold.
- Debras, C., Horgues, C. & Scheuer, S. (2015). The multimodality of corrective feedback in tandem interactions. In: Molina-Plaza, S. (Ed.), *Multimodal Communication in the 21st Century: Professional and Academic Challenges. Procedia – Social and Behavioural Sciences*, vol. 212, pp. 16–22.
- Du Bois, J. (2007). The stance triangle. In: Englebretson, R. (Ed.), *Stancetaking in Discourse*, pp. 139–182. Amsterdam and Philadelphia: John Benjamins.

They don't like it when other people correct their mistakes when they're speaking (2), They don't think correcting them would be helpful (1), Their partner doesn't take your corrections on board (0). [Appendix 3, Pronunciation questionnaire, Question 7].

²⁰ “Long (1996) contends that the juxtaposition recasts create between learners' erroneous output and target forms aids language acquisition. He argues that the negative evidence provided by recasts aids the process of ‘cognitive comparison’ and is more effective than positive evidence in the form of ‘models’”. (Sheen, 2006: 364).

- El Tatawy, M. (2002). Corrective feedback in second language acquisition. *Working Papers in TESOL & Applied Linguistics*, pp. 1–19. Teachers' College, Columbia University.
- Gass, S. M. (2003). Input and interaction. In: Doughty, C. J. & Long, M. H. (Eds.), *The Handbook of Second Language Acquisition*, pp. 224–255. Oxford: Blackwell.
- Horgues, C. & Scheuer, S. (2014). “I understood you, but there was this pronunciation thing...”: L2 pronunciation feedback in English/French tandem interactions. *Research in Language*, 12(2), pp. 145–161.
- Horgues, C. & Scheuer, S. (2015). Why some things are better done in tandem? In: Mompeán, J. A. & Fouz-González, J. (Eds.), *Investigating English Pronunciation: Current Trends and Directions*, pp. 47–82. Basingstoke and New York: Palgrave Macmillan.
- Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- . (2007). *English as a Lingua Franca: Attitude and Identity*. Oxford: Oxford University Press.
- Ladefoged, P. (2001). *A Course in Phonetics* (4th ed.). Boston: Heinle & Heinle Publishers Inc.
- Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum.
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19(1), pp. 37–66.
- Lyster, R., Saito, K. & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46, pp. 1–40.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, 27(3), pp. 405–430.
- McDonough, K., Crowther, D., Kiestra, P. & Trofimovich, P. (2015). Exploring the potential relationship between eye-gaze and English L2 speakers' responses to recasts. *Second Language Research*, 31(4), pp. 563–575.
- Roach, P. (2009). *English Phonetics and Phonology: A Practical Course* (4th ed.). Cambridge: Cambridge University Press.
- Saito, K. & Lyster, R. (2011). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɪ/ by Japanese learners of English. *Language Learning*, 62, pp. 593–633.
- Saito, K. & Lyster, R. (2012). Investigating the pedagogical potential of recasts for L2 vowel acquisition. *TESOL Quarterly*, 46(2), pp. 387–398.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), pp. 129–58.

Sheen Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research*, 10(4), pp. 361–392.

Teaching English. Retrieved on October 15, 2015, from <https://www.teachingenglish.org.uk/article/word-stress>

Appendix 1: The reading passage

(for the francophone participant. Read this instruction aloud.)

Please read the following text twice:

- *once with your tandem partner helping you especially if he/does not understand what you are saying or if your reading is unclear*
- *and then a second time on your own (no interruption).*

The North Wind and the Sun

The North Wind and the Sun were disputing which of them was stronger, when a traveller came along wrapped in a warm cloak*.

They agreed that the one who first succeeded in making the traveller take his cloak off should be considered stronger than the other.

Then the North Wind blew as hard as he could, but the more he blew, the more closely did the traveller fold his cloak around him; and at last the North Wind gave up the attempt.

Then the Sun shone out warmly, and immediately the traveller took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.

(* a cloak is a type of coat)

(Pour le participant anglophone. Lisez cette consigne à haute voix)

Lisez le texte ci-dessous deux fois :

- *une première fois avec l'aide de votre binôme qui vous aidera s'il/elle ne comprend pas ce que vous dites ou si la lecture n'est pas claire.*
- *et une deuxième fois tout seul (sans interruption).*

La bise* et le soleil

La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort, quand ils ont vu un voyageur qui s'avavançait, enveloppé dans son manteau.

Ils sont tombés d'accord, que celui qui arriverait le premier à faire ôter* son manteau au voyageur, serait regardé comme le plus fort.

Alors la bise s'est mise à souffler de toute sa force, mais plus elle soufflait, plus le voyageur serrait son manteau autour de lui; et à la fin, la bise a renoncé à le lui faire ôter.

Alors le soleil a commencé à briller et au bout d'un moment, le voyageur, réchauffé, a ôté son manteau.

Ainsi la bise a dû reconnaître que le soleil était le plus fort des deux.

(* ici la bise : un vent très froid * ôter : retirer/enlever)

Appendix 2: The general questionnaire

1. Number of tandem meetings with your partner since the January 31, 2013 meeting:

Did you find the number of meetings to be (circle your answer):

insufficient / sufficient / too frequent?

2. How often did you meet your tandem partner? (circle your answer):

*twice per week/ once per week/ once every other week/ once every 20 days/
once per month*

Did you find the frequency of meetings to be (circle your response):

insufficient / sufficient / too frequent?

3. On average, how long did your tandem meetings last?

30 mins / 1hour / 1h30 / 2 hours / more than 2 hours

4. What percentage of time was spent speaking French in your tandem conversations?

[0 , 10 , 20 , 30 , 40 , 50 , 60 , 70 , 80 , 90 , 100] % of the time

5. What percentage of time was spent speaking English in your tandem conversations?

[0 , 10 , 20 , 30 , 40 , 50 , 60 , 70 , 80 , 90 , 100] % of the time

6. Over the semester, have you had the opportunity to converse with other French speakers:

*daily/ several times a week/ once a week / a few times over the semester/
no*

Explain:

7. During your tandem conversations:

	always	almost always	often	sometimes	almost never	never
You begin the meeting in the same language (which language:)						
You start speaking in one language at the beginning of the conversation and then switched to the other language						
You speak in one of the two languages most of the time						
You speak one language at one meeting and the other language at the next meeting						
You switch from one language to the other throughout the conversation (for example when there was a comprehension problem)						
You don't mix languages, except to ask specific vocabulary or grammar questions						
Other (explain):						

8. During a conversation in a foreign language, if you have doubts about how to express something or if you have problems expressing yourself (grammar, vocabulary, pronunciation):

	always	almost always	often	sometimes	almost never	never
You stop and explain your problem in the foreign language						
You stop and explain your problem in your native language						
You continue to speak and wait for your partner to react						

9. When you make a mistake in French:

	always	almost always	often	sometimes	almost never	never
Your tandem partner corrects you						
Your tandem partner corrects your vocabulary						
Your tandem partner corrects your grammar						
Your tandem partner corrects your pronunciation						

10. When your tandem partner makes a mistake in English:

	always	almost always	often	sometimes	almost never	never
You correct him/her						
You correct his/her vocabulary						
You correct his/her grammar						
You correct his/her pronunciation						

11. When your partner tells you something in English:

	always	almost always	often	sometimes	almost never	never
You listen and try not to interrupt						
You listen and ask questions so as to help the conversation going on						
You interrupt your partner when you cannot understand what is said						
You interrupt your partner when he/she makes a mistake						

12. Did you notice any differences in your partner's conversational habits that would be linked to his/her culture? YES / NO If yes, explain:

13. Have ever helped your partner in his/her academic work? YES/NO

14. Has your partner ever helped you in your academic work? YES/NO

15. Working in tandem with your partner allowed you to improve your knowledge and skills in which of the following domains?

(Circle your answer)

	0 = no improvement			5 = much improvement		
culture	0	1	2	3	4	5
phonetics/pronunciation	0	1	2	3	4	5
grammar	0	1	2	3	4	5
vocabulary	0	1	2	3	4	5
general ease of expression	0	1	2	3	4	5

16. Evaluate your confidence in speaking French before beginning your tandem meetings and after:

	0 = less confident					10 = more confident					
before:	0	1	2	3	4	5	6	7	8	9	10

after:	0	1	2	3	4	5	6	7	8	9	10
--------	---	---	---	---	---	---	---	---	---	---	----

17. On a scale of 0 to 10, how was your experience with working in tandem with your partner?

	0 = the most negative					10 = the most positive					
	0	1	2	3	4	5	6	7	8	9	10

18. What is your overall impression of your tandem conversations? (obstacles encountered, benefits gained)

Appendix 3: The pronunciation questionnaire

Tandem and pronunciation (English-speaking participants)

When speaking with your tandem partner in English:

1. Do you adapt the way you speak to them? YES / NO. If so, what changes do you make?

	YES	NO
I speak more slowly		
I articulate more clearly		
I speak louder		
My intonation is clearer		
My vocabulary is more simple		
I use more straightforward sentence structure		
I use shorter sentences		

2. Now choose one of the terms below to describe your tandem partner's accent [in English]:

very strong / quite strong / moderate / quite slight / slight / very slight / no accent at all

3. What do you think about the French accent in English as a rule:

4. I correct my partner's pronunciation (circle one of the following):

systematically, whenever I hear a mistake / almost always / often / sometimes / only when they ask me to / only when they ask me a specific question about a particular word / hardly ever / never

5. I prefer to correct my partner's pronunciation (circle one of the following):

on the spur of the moment / at the end of their sentence / when they've finished saying what they have to say / at the end of our tandem session

6. What exactly do you correct when it comes to your partner's pronunciation?

	YES	NO	Please give examples:
Their intonation			
The rhythm of their English			
Their word stress			
The speed at which they speak			
Their consonants			
Their vowels			
Their general fluency			

Other:

7. When you **don't** correct your partner's pronunciation, it's because:

	YES	NO
Their mistakes are too small		
It would be impossible to correct all their mistakes		
You don't want to make your partner feel uncomfortable		
You don't want to interrupt the flow of their ideas		
You don't think it's polite to correct or to interrupt your partner		
You don't like it when other people correct your mistakes when you're speaking		
You don't think correcting them would be helpful		
You can understand them despite their mistakes		
Your partner doesn't take your corrections on board		
You can't hear their mistakes		

8. Your partner's pronunciation in English prevents you from understanding them in English% of the time:

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Please provide specific examples from your tandem experience:

9. The advice you have given your partner has helped them to improve their pronunciation

I haven't noticed any changes / yes, most of aspects of his-her pronunciation / some aspects of his-her pronunciation / no, not really / no, definitely not

Say what has improved:

List any remaining difficulties:

10. Which pronunciation problems annoy you the most [in English by French speakers] even if they don't hinder actual comprehension?

When you're speaking with your partner [in French]:

1. Would you say that your accent is:

very strong / quite strong / moderate / quite slight / slight / very slight / no accent at all

2. What do you think about the English accent in French in general?

3. What aspects of French pronunciation do you think you still need to improve?

4. Your partner corrects your pronunciation (circle one of the following):

systematically when they hear a mistake / almost always / often / sometimes / only when I ask them to / only when I ask them a specific question about a word / hardly ever / never

5. Your partner prefers to correct your pronunciation (circle one of the following):

on the spur of the moment / at the end of my sentence / when I've finished saying what I have to say / during the tandem assessment session

6. What does your partner correct about your [French] pronunciation?

	YES	NO	Please give examples:
Your intonation			
The rhythm of your French			
Your word stress			
The speed at which you speak			
Your consonants			
Your vowels			
Your general fluency			

Other:

7. How often does your partner correct your pronunciation?

	YES	NO
Too often because it interrupts the flow of conversation		
Too often because it makes you feel uncomfortable		
Very often but you find it helpful		
From time to time and you're okay with that		
Rarely and you think it's a shame		
Hardly ever, which is fine because you don't like being interrupted when you're speaking		

8. Your pronunciation [in French] prevents you from making yourself understood by your partner% of the time:

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Please provide specific examples:

9. Speaking with your partner and receiving advice from them have helped you to improve your pronunciation (circle)

I haven't noticed any changes / yes, most aspects of my pronunciation/ some aspects of my pronunciation/ no, not really / no, definitely not

Say what has improved:

List any remaining difficulties:

CHAPTER FOURTEEN

PHONETIC INSPIRATIONS IN AUTHENTIC MATERIALS: STIMULATING STUDENTS' PHONETIC AWARENESS

MARTA NOWACKA

14.1 Introduction

14.1.1 The importance of authentic materials in phonetic training

In this chapter, we discuss the use of authentic materials in a pronunciation syllabus to boost learners' metaphonetic competence (Sobkowiak, 2003) and also to motivate them. The aim is to present some humorous listening, visual and printed resources in the form of excerpts from films, commercials, interviews, billboards, advertising slogans, cartoons and popular press articles that can add sparkle to the phonetic discussion, raise students' phonetic awareness and competence and thus support and enhance their pronunciation learning.

Celce-Murcia et al. (1996) and Szpyra-Kozłowska (2014) recommend using a variety of authentic materials and Internet resources to develop learners' pronunciation skills and autonomy and also to make pronunciation teaching more attractive. In addition, a significant number of authors (Brown, 1990; Shockey, 2003; Field, 2008; Hancock, 2012; Cauldwell, 2013) point to the great importance of listening skills and explicit formal instruction of casual speech processes for better understanding of colloquial conversational English, which can be achieved through the application of authentic resources.

The benefits of using authentic materials for listening, oral and pronunciation practice are undeniable (Underwood, 1989; Thorn, 2012; López Casoli & Paderni, 2013). Underwood's (1989: 100) description of authentic materials encompasses such elements as natural rhythm,

intonation and pronunciation, some overlap between speakers including interruptions, normal rate of delivery, relatively unstructured language, incomplete sentences, false starts, hesitations, background noises and voices, natural starts and stops, and less densely packed information than is the case in written language.

Among the advantages of listening to a variety of authentic texts we can list familiarization with multiple voices and preferably accents, including social and regional ones, and also becoming acquainted with speech articulated with a variety of voice qualities, e.g., whispering or conveying emotions, which resembles the real life context.

Shockey (2003) and Cauldwell (2013) note that for successful communication it is vital to understand casual speech phonology. As phonology is a facility that allows us to use the sound system of our native language, we must also think of it in terms of perception. Thus, training students of English phonetics in phonology of listening by means of authentic materials seems well-justified. Shockey (2003: 123) advocates the inclusion of perception of conversational speech in English courses for those planning to live in English-speaking countries and she suggests explicit teaching of types of phonological reduction and where they are likely to be encountered. Brown (1990) advises that a foreign learner will have to learn to listen like a native speaker to extract the message from a predominantly acoustic signal, taking moments of greater and lesser prominence into account, to make guesses about the content and reinterpret them if necessary.

Foreign learners, like first-language learners, should have intensive experience with a variety of different styles of speech, which could allow them to subconsciously deduce the relationships between and amongst them. Pisoni and Lively (1995: 454) stress the importance of high-variability training, in which the new phonetic contrast is spoken by a variety of speakers in several different phonetic environments, to promote the development of robust perceptual categories.

From the above-mentioned arguments for the use of authentic materials and their potential benefits for a learner's performance a positive picture emerges. Any disadvantages of their implementation in the phonetic curriculum seem to be unlikely and it is to be hoped that they would be outweighed by the linguistic and non-linguistic advantages, such as their positive effect on learners' listening comprehension skills (Brown, 1990; Shockey, 2003; Cauldwell, 2013) or on the use of appropriate intonation (López Casoli & Paderni, 2013). Although authentic materials do not usually belong to weekly pronunciation training, in which the phonetic difficulty of words and chunks is gradually increased and tailored to the

needs of the group, it is believed that such sources could serve as a break from a routine class dominated by ‘listen-and-repeat’, ‘transcribe’ or ‘read-from-transcription’ activities.

14.2 Method

For the purpose of this analysis we have adopted Sobkowiak’s (2003) taxonomy of phonetic trivia. We group them according to the phonetic features included in the traditional descriptive grammar course, such as segments, prosody, spelling-to-sound correspondence and various accent-oriented issues.

An attempt has been made to fulfil McGrath’s (2002) criteria of appropriate selection and evaluation of authentic materials for use in the pronunciation classroom, i.e. relevance to textbook and students’ needs, topic interest, cultural appropriateness, logistical considerations, cognitive and linguistic demands, quality and exploitability. The choice of materials presented here is appropriate to the freshman’s phonetic needs and a result of the author’s fascination with how and why English sounds the way it does, which is typical for phoneticians.

Firstly, within segments our examination starts with notorious Polish substitutions of ‘th’ but it also touches upon th-fronting and exemplifies a notion of a phonotactic constraint in English. In the context of Polish learners, interdental fricatives are rendered as mainly dental plosives /t, d/ but also labiodental fricatives /f, v/ and even dental fricatives /s, z/, which is confirmed by a large number of researchers including Szpyra-Kozłowska et al. (2002), Sobkowiak (1996, 2000, 2012), Gonet and Pietroń (2006), Nowacka (2008) and Porzuczek et al. (2013). The presentation of foreign erroneous renditions of ‘th’ is achieved by exposing learners to two pieces of authentic materials, i.e. *Decathlon*’s (DecathlonPolska, 2014) commercial together with its advertising slogan *The Cathlons* and also to an interview with Donald Tusk (Tusk, 2014), President of the European Council. The examination of non-native ‘th’ substitutions leads us to one of the native renditions of /ð/, that is th-fronting exemplified in an interview with David Beckham, which was observed by Lindsay (2014). After contrasting a native English with Polish enunciation of interdental fricatives, we aim at showing an English accent in the Polish language to familiarize our learners with the notion of a foreign accent and phonotactic constraints. This has been done with the use of quotations from the popular newspaper *The Guardian* on ‘accentism’ (Buist, 2013; Muir, 2014) and a Polish advertisement for

Arctic mineral water in which Cindy Crawford pronounces a Polish sentence with an English accent (bbs7000, 2010).

The second area of interest is English prosody and here we have selected an issue of uptalk, the use of terminal rising tones for statements applied by some younger native-speakers in many parts of the English-speaking world. To familiarize our subjects with this intonation contour, which has attracted much unfavourable comment in recent years, we have chosen a few types of authentic materials such as: an excerpt of a conversation between an uptalker, Lucy, and Paul Slippery from the TV series *Fortysomething* (Majka327, 2008), a youtube clip in which an uptalk user provides her reasons for the use of upspeech (LiveScienceVideos, 2013), a cartoon in which the image of a valley girl is exaggerated (Webcomic/YAFGC, n.d.), various imitations of Valley Girls' uptalk in Ellen's Show in a game *Heads up!*, performed by Fergie (TheEllenShow, 2014d), Jennifer Love Hewitt (TheEllenShow, 2014b) and Emily Blunt (TheEllenShow, 2014c), a fragment of a popular press article on this issue (Goman, 1993) and a sketch by Catherine Tate presenting a Valley Girl's monologue (BBCWorldwide, 2007).

The third issue selected for analysis is spelling-to-sound correspondence in English, e.g. *Cockburn's* and *Chipotle*, based on clever pronunciation-oriented advertising campaigns. We then move to a word used in the IT sector, and point to an intervention by the creator of *gif* on how his invention should be pronounced correctly (Gif's inventor, 2013; The Webby Awards, 2013). We also focus on questions on pronunciation in a popular student quiz show (University Challenge, 2015), the rendition of fashion (i-D's A-Z, 2014) and brand names (Getting Your Brand Pronounced Correctly, 2013; Stuart & Kakoyiannis, 2014), as well as words which are most commonly mispronounced even by native speakers of English (Crossley, 2014).

Our analysis finishes with a focus on the image of British speakers and their accents. We examine the stereotypes concerning the British accent in Ellen DeGeneres's *Heads-up* with Fergie (TheEllenShow, 2014d), Jennifer Love-Hewitt (TheEllenShow, 2014b) and Joseph Gordon-Levitt (TheEllenShow, 2014a); negative associations of posh RP in the Jaguar F-type Coupe 2014 commercial *Good to be bad* (TV Top Car – welcome to the Top, 2014) and its continuation *The Art of Villainy* (Mapauto, 2014) and this accent's attractive appeal in *Love actually* comedy film (Defender Productor Mix, 2011; Love actually script, n.d.), *Dilbert's* cartoons (Adams, 2005) and *Time Out* Global Dating survey (Bourn, 2015) and the reactions to it (Gabbatt, 2015; Scott, 2015 and Singh, 2015).

The authentic materials which have been gathered for the purpose of this text can be divided into three groups: audio-visual, audio and visual respectively. Among the first category of audio-visual sources we find: video-clips of commercials encompassing *Decathlon's* (DecathlonPolska, 2014), *Arctic* (bbs7000, 2010), *Cockburn's Boat* (n.d.), *Cockburn's Rescue* (n.d.) and *12 incher* (Compartiendo Publicidad, 2012), the Jaguar F-type Coupe 2014 commercial *Good to be bad* (TV Top Car – welcome to the Top, 2014) and its continuation *The Art of Villainy* (Mapauto, 2014); interviews with David Beckham (Lindsay, 2014), Donald Tusk (Tusk, 2014), the founder of *Chipotle* (Chipotle Mexican Grill, 2011), Emily Blunt (TheEllenShow, 2014c), Fergie (TheEllenShow, 2014d), Jennifer Love-Hewitt (TheEllenShow, 2014b) and Joseph Gordon-Levitt (TheEllenShow, 2014a); TV series/films: *Fortysomething*, *Love actually* (Defender Productor Mix, 2011; Love actually script, n.d.) and other types of videos: Catherine Tate's sketch on *Valley Girl* (BBCWorldwide, 2007), acceptance speech by Steve Wilhite, the creator of *gif*, at the 7th Webby Awards (Gif's inventor, 2013; The Webby Awards, 2013), i-D's A-Z of fashion name pronunciation (i-D's A-Z, 2014), and *Business Insider* clip on brand names (Stuart & Kakoyiannis, 2014).

Visual materials such as excerpts from: cartoons including *Valley Girl elf* (Webcomic/YAFGC, n.d.) and Adams' (2005) *Dilbert*; poster advertisements, e.g. *Cockburn's*' 'Pronounce Responsibly' campaign (Compartiendo Publicidad, 2012; *Cockburn's Boat*, n.d.; *Cockburn's Passport*, n.d.; *Cockburn's Rescue*, n.d.; *Cockburn's Submarine*, n.d.; Those famous 70's ads, n.d.) and *Chipotle* (Chipotle Mexican Grill, 2011; Swift, 2014), posts (Scathe to Reddit and Creative Review magazine both in Gif's inventor, 2013), popular press (*Australian Times*, *Daily Mail*, *The Guardian*, *The Telegraph*, *Time Out* and *The New York Times*) on a variety of pronunciation-oriented topics such as accentism, uptalk, unpredictability of some English proper names, the correct pronunciation of an IT invention, University Challenge questions on pronunciation (University Challenge, 2015) and the attractive appeal of British English.

The last group encompasses two mostly audio materials: pronunciation of a proper name on *Forvo* (n.d.) pronunciation database as well as a youtube recording on the reasons for uptalk in a conversation.

It is believed that this collection of authentic sources should be sufficient to convince the reader that pronunciation issues are plentiful and their abundance rather than their scarcity may be inhibiting, yet it proves that pronunciation teaching can be diversified by real life texts and speech samples.

14.3 Results

14.3.1 From segments via phonotactic constraints to foreign accent

To start with segmental features, we would like to point to a negative transfer from Polish into English, exemplified by a consonantal substitution of interdental fricatives, a voiced one with a Polish dental /d/ or /z/ and the voiceless one with /f/, /s/ or /t/. The intention is to familiarize students with non-native substitutions of sounds resulting in varied degrees of foreign accent and also with native realizations such as th-fronting.

The first video which we take into account is Decathlon's (DecathlonPolska, 2014) advertising slogan *The Cathlons* and the catchphrase *The Cathlons: sport for all, all for sport*. The clever play on words, i.e. the brand name *Decathlon* and its close phonetic equivalent *The Cathlons* has been applied in the commercial so as to reinforce the word *Decathlon*, and to make consumers remember the name of the sportswear retail chains and the brand itself. In the advert the family of the Cathlons finds the best Christmas presents for their nearest family members in Decathlons. The advertisement makes use of consumers' knowledge of English, i.e. their familiarity with the formation of a collective family surname by the implementation of a definite article 'the' and adding a suffix -s to the noun constituting a surname, which also accounts for foreigners' frequent mispronunciation of a voiced dental fricative and its substitutions with /d/. Although this rendition of dental fricative is typical for foreign-accented speech, this could be a starting point for a discussion about native renditions of /ð/ such as th-fronting, which we support with fragments from two interviews with David Beckham in which he applies th-fronting in the function words *with* and *through* but not in content words like *things* or *think*.

For comparison, we also include Polish-accented pronunciation of 'th's in a fragment from an interview with the President of the European Council, Donald Tusk, (Tusk, 2014, 0:40–1:06) in which he replaces both dental fricatives with Polish equivalents. The replacements concern /ð/ pronounced as /z/ in *with* and as /d/ in *the*, *that*, *this* and *their*, and also /θ/, which is rendered as /f/ in *ethnic* and *three* but as /s/ in *everything*, *something* and /t/ in *thousand*. In general, whereas the substitution of dental fricatives with /f/ and /v/, although non-standard, is observed in English and referred to as th-fronting, the substitution of /θ/ and /ð/ with dental plosives, well-known in Irish English and Black Urban Vernacular

of the U.S., appears to be stigmatizing and contributes to a foreign accent in English. It should be borne in mind that these were not cases of δ -reduction, which is a process that exists in conversational English.

To make it easier for first-year students to imagine what the notion of a foreign accent represents we selected a video in which the Polish language is spoken by an English speaker with a noticeable degree of foreign sounding in one word. This material could in turn lead to a discussion about the differences in syllable structure between native and target languages. Such phonotactic constraints can be illustrated in the commercial for *Arctic* mineral water in which Cindy Crawford is saying a sentence “Twoj dzień, twoja woda” [Your day, your water.] (bbs7000, 2010, 0.25). Her rendition of the word /tfuj/ sounds foreign because of the initial consonant cluster ‘tf’, which functions in Polish but not in English. Her native phonology is clearly guiding her, as in her realization, she inserts a schwa between /t/ and /f/, then forms a good English onset and thus compensates for a forbidden one in English.

These materials could also lead to a conversation about the notion of foreign-accent and the impressions one makes on listeners when one speaks with varied degrees of foreign accent. Let me start with Buist’s (2013) quotation from *The Guardian* about foreigners’ speech: “[t]he tips of their tongues touching their teeth when they say ‘t’, or saying ‘dose’ instead of ‘those’ seems a flimsy basis for mistrust. So, do we think foreigners are dishonest or stupid?”

Although the above-mentioned provocative question can be left unanswered, it seems to be a teacher’s duty to make their students aware of this insidious impact of accent. The learners should realize that their non-native pronunciation of sounds can undermine the credibility of the whole utterance, as Lev-Ari and Keysar (2010:3) confirm: “[a]ccent might reduce the credibility of non-native job seekers, eyewitnesses, reporters or news anchors.”

In a popular daily newspaper Muir (2014) admits that accents matter in modern Britain. He points to ‘accentism’, i.e. being discriminated against because of how one speaks, claiming that “[n]o one should feel compelled to change the way they interact. If they do so because it works to their advantage, that’s different. Then it’s about choice, not obligation or coercion.” Muir defines the roles of the listener and the speaker in a way which, although true in the ideal world, does not always happen in real conversation: “[t]he responsibility of the listener is to be tolerant, celebratory perhaps, and to adapt when necessary. The requirement on the speaker is to make themselves understood, whatever lilt they choose.”

14.3.2 Prosody: uptalk

Firstly, by exposing the learners to an excerpt from *Fortysomething* (Majka327, 2008, 0:04–1:07), starring and directed by Hugh Laurie, we wish to obtain students' observations on uptalk, a rising intonational pattern, applied by a young woman, Lucy, in the following statements: "You must be Paul? Dan's dad? I'm Dan's girlfriend? Lucy? Laura's sister...?"

One of the issues that we present is the impression which uptalk exerts on the listener. This intonational pattern sounds as if the speaker was desperately craving the interlocutor's attention, although, as we learn from Amanda Ritchart, a youtube user (LiveScienceVideos, 2013), it is politeness which is the main role of uptalk in her speech: "... For an example, you know how, when you're at Starbucks and they're asking you for the name, I always go 'Amanda↗'. And it's not because I don't know my name but it's just the way I talk, so it's almost like a polite way of telling them my name, but this stuff is uptalk." To a non-uptalker, this rising pattern sounds as if the speaker were not sure of her identity, or felt she could not assert it although Wells (2006:37) notes that "the pragmatic context calls not for checking or querying, but for assertion of something the speaker certainly knows." Another example of uptalk is when a person is giving directions, which we included in our study, for example, "Ok, this's at the college↗, keep going↗, towards the grocer's store↗..." On the basis of this explanation we can conclude that politeness is one of the attitudinal functions of this tone which is in agreement with Bradford's (2008) observations that upspeak has affective and referential communicative functions, as it promotes a sense of solidarity and empathy between interlocutors, reduces speaker-hearer distance and it signals salient pieces of information, encouraging the hearer's involvement in the conversation.

Then there are of course humorous cartoons in which the notion they represent, in this case uptalk, also referred to as *Valley Girls' talk* is exaggerated (Webcomic/YAFGC, n.d.). In the cartoon, found on the tvtropes site, one can see two young, but rather different, women, A & B, sitting in a bar. A is looking at B with some kind of compassion but also disapproval and comes to the conclusion: "She's a valley elf!". B, engaged in her story, is gesturing and saying: "And I was like oh-MIGAWD!" The obvious thing is that the cartoon is not favourable towards uptalk speakers.

As another source of phonetic inspiration we selected the game "Heads up!" from *Ellen's* show. In this timed game, Ellen DeGeneres' guests are asked to speak with different accents, whose names are displayed on the screen. Ellen's job is to guess the accent just by listening to the guest's

imitation. Our suggestion is to listen to Valley Girl's interpretations by the following celebrities: Fergie (TheEllenShow, 2014d, 0:21–0:25) saying “Oh my gosh, like;” Jennifer Love Hewitt (TheEllenShow, 2014b, 0:59–1:03) with “Oh, my God;” and Emily Blunt (TheEllenShow, 2014c, 1:27–1:36), reciting a tongue twister “How much wood ...” with a lot of rising tones and eye rolling. Having done this task, it seems obvious that the image of the speaker which has been created is far from being positive and not even close to neutral. The students of phonetics should be aware that uptalk can be perceived as confusing or/and disrespectful when a distance in a discourse is required instead of a social bonding.

The discussion on uptalk could also be supported by an excerpt from an article from the popular press. Goman (1993) asks rhetorically whether you would be comfortable hearing a rising tone at the end of a pilot's sentences such as “[t]his is Captain McCormick? Your pilot? We'll be flying to Denver? Our cruising altitude will be, like, 30,000 feet?” As noted by Bradford (2008:252) an irritating impact of uptalk on especially older generation of listeners can be caused by “a misjudgement of the social situation on the part of the upspeaker and a misconstrual of the upspeaker motivations on the part of the hearer(s).”

As a culmination of the demonstration of uptalk we recall the classic sketch by Catherine Tate (BBCWorldwide, 2007), which is a Valley Girl's monologue, starting with the line: “ ... so I'm in this bar and he comes over and I'm thinking: ‘Oh, my God;’ and he's like: ‘Hi,’ and I'm like: ‘Hi,’ and he's like: ‘Are you on your own?;’ and I'm like: ‘yes’...” Once again this piece of authentic material focuses on the negative association of an uptalk intonation contour.

Taking the above examples into consideration, it has to be admitted that these presentations of uptalk might be derogatory, that is why the next step could be to find such materials in which, if at all, we can see a positive or neutral approach of listeners to this tone.

14.3.3 Spelling-to-sound correspondence and proper names

The next issue is spelling-to-sound correspondence and enunciation of proper names. Poole (2014) expresses the startling view that “arguments over pronunciation are mostly proxies for snobbery and class one-upmanship.” To support his claim with some examples, such as *expresso*, *nucular* or *bruschetta*, he further adds that “[t]he completely unpredictable pronunciations of many proper names in English, for example, act as a kind of secret code for the elect.”

It cannot be denied that in the UK great importance is attached to pronunciation by media, and we can find evidence of this in everyday articles in TV programmes, shows and advertisements. Pronunciation-oriented issues appear not only in advertising campaigns, but also in TV programs such as the academic quiz show *University Challenge* (2015) and they even find their way into the IT sector, e.g. during the Webby Awards, where the creator of the *gif* graphics format taught the audience how the name should be enunciated correctly, so as to dispel the ambiguity and also to make them laugh (Gif's inventor, 2013; The Webby Awards, 2013, 0:55–1:12).

To rectify the brand name mispronunciation, *Chipotle* and especially *Cockburn's* have excellently implemented a time-consuming but effective technique. They have guided the market to pronounce the brand name correctly through the use of press and online advertisements. *Cockburn's*, a wine brand, prepared port's advertising comeback in the UK with the 'Pronounce Responsibly' campaign in 2011, undeniably displaying great intrinsic humour. It consists of both light-hearted posters and online videos, playing on the common mispronunciation of the name *Cockburn's*. Each poster conceals the letters 'CK' in the word *Cockburn's* using one of three techniques: a strategically placed rip, a port glass or pixilation giving rise to a blurred image (Macleod, 2011). As regards the *Cockburn's* port online video campaign, it was developed by BETC London around the continued 'pronounce correctly' brand strategy. The video entitled '12 inches' (Compartiendo Publicidad, 2012) is a part of *Cockburn's*' re-launch ad campaign, which echoes the affectionately remembered classic *Cockburn's* TV advertising from the 1970s, 'Cockburn's Special Reserve'. As we learn from the *Cockburn's* website, in 1971 the British advertising agency Collett Dickenson Pearce created a series of TV ads, directed by Jon Ritchie, to launch their Port and these ads immediately met with enormous success. The must-see pronunciation-oriented *Cockburn's* series involves such episodes as: *Boat*, *Passport*, *Rescue* and *Submarine*.

In *Cockburn's Boat* (n.d.) advertisement, set on a Titanic-style lifeboat, we can see some upper-class survivors, in luxurious ballroom dresses, along with some working-class survivors and a few members of the crew. From their conversation we learn that they had just finished dinner before the ship sank. The captain says "I'm afraid I might have some rather bad news for some of you. Dave's only had time to bring the port and it's my after dinner tittle but what about the rest of you?" The first officer replies: "I should say so, sir." One passenger makes sure: "Cockburn's, is it?" and pronounces *Cockburn's* with a LOT vowel in the first syllable. Then the captain corrects the passenger's enunciation by correctly pronouncing the

name: “*Cockburn’s*.” At the same time the first officer reacts to this mispronunciation by laughing and repeating this erroneous pronunciation: “Huh, huh *Cockburn’s* huh, huh, good.” Then the captain says: “You mean, *Cockburn’s*. Yes, special reserve.” We can still hear the first officer’s laughter and his repetitive mispronunciation of the name: “*Cockburn’s* huh, huh, good.” The commercial finishes with the statement: “Cockburn’s. After dinner, a bottle of port is really all you need. Cockburn’s Special Reserve, a very fine bottle of port.” In the last seconds of this commercial we can hear one female voice with an upper-class accent saying: “Did anyone bring the petits fours?”

In *Cockburn’s Rescue* (n.d.), the crew of a British ship is rescued by a Russian ship, and at dinner time they sit at a table, have a conversation and as a thank-you-gift the British captain serves a bottle of port, saying that after dinner he enjoys a glass of it with his wife. The Russian captain happens to mispronounce the name of the port as in his spelling-guided judgement /k/ should be found in *Cockburn’s* and the British first officer makes him aware of the silent nature of the letters ‘ck’ in its name. The Russian captain pronounces *Cockburn’s* correctly now and immediately starts applying the phonetic rule to other words with the same final letters ‘ck’, naming objects in the room, such as *clock* and a *sock*. Once again, he is corrected by being told that in the above-mentioned words the sound /k/ should not be mute, so he makes an attempt to test this new hypothesis and he implements LOT and /k/ in the final syllable of *Moscow*. The commercial finishes with the line: “After dinner you can’t go wrong with Cockburn’s Special Reserve. A very fine bottle of port.”

In the recent half-minute-long *Cockburn’s 12 incher* (Compartiendo Publicidad, 2012) we can hear Grieg’s Concerto in A minor Op. 16 performed by a pianist, giving a recital. After a while the camera moves away from the performer and we can see a man in the audience who appears to be sitting far from the stage. We soon learn that this is an optical illusion because the pianist and its piano turn out to be toy-sized. What follows is the image of the man shaking his head with disbelief and asking the genie facing him: “Seriously, a 12 inch pianist?” after which we hear the well-known tagline “Cockburn’s. Pronounce responsibly.”

Our investigation into the unpredictable pronunciation of proper names moves to *Chipotle*¹ pronounced /tʃɪˈpɒtleɪ/, the name of an American company selling Mexican fast-food (Chipotle Mexican Grill, 2011; Swift, 2014). The spelling of this borrowing does not lend itself to clear

¹ The word *Chipotle* comes from the Nahuatl, i.e. Aztec word *chilpochtli*, which means ‘smoked chilli.’ It is used in Mexican cuisine.

rendering in English. Swift (2014) reports that Mother, the creative agency for Chipotle's first UK campaign, decided to teach consumers what the correct pronunciation of the word is. This was done in print and poster ads showing images of Chipotle's food along with its most common mispronunciations, such as crossed out versions of "~~CHI-POLE TAY, CHI-POTTLE, SHI-POT-LAY~~", and finally its correct, respelled, transliterated rendition "CHI-POAT-LAY". The advertisement also features the strapline "delicious however you say it". The recommended pronunciation can be heard in an interview with the founder of the brand, pronouncing it as /tʃi'pəʊtleɪ/ (Chipotle Mexican Grill, 2011, 0:32–0:42).

Pronunciation happens to be a hot issue even in the IT sector. During the 17th Webby Awards² in 2013 Steve Wilhite, the creator of the Gif (Graphics Interchange Format), a data compressing technique, announced that his invention should be pronounced with /dʒ/ rather than with /g/: "It's pronounced 'jif' not 'gif'" (The Webby Awards, 2013, 0:55–1:12). The Oxford English Dictionary and the American Heritage Dictionary both allow two optional pronunciations with a post-alveolar voiced affricate and a velar voiced plosive, based on poll preference results. This intervention of the coiner of the acronym in the defence of the correct rendition of the word sparked a flood of responses from web users, who played on the spelling and sound discrepancy, such as a post by Scathe to Reddit: "Unless we are going to pronounce graphics as jraphics he makes zero sense," or the reaction of *Creative Review* magazine: "Well theer you jo." (Gif's inventor, 2013).

Our next step is pronunciation questions asked in TV quiz shows. In the show *University Challenge* (2015) teams of students from the UK universities answer questions on all manner of subjects. The final of the 2015 series between competing teams from Magdalen College,³ Oxford and Gonville and Caius College,⁴ Cambridge, featured, according to *The Telegraph* the toughest twenty questions ever asked on the program. Among them three bonus questions relate, to some extent, to pronunciation as they concern heteronyms. The questions are as follows:

- a) Which heteronyms are the names of a port in Alabama and a type of kinetic art invented in the 1930s by Alexander Calder?

² Webby Awards, regarded as the Internet equivalent of the Oscars, honour the year's best websites, and online films and videos. They are known for their five-word limit on acceptance speeches.

³ /,mɑːdlɪn 'kɒlɪdʒ/

⁴ /,ɡɒnvl̩ ən kiːs 'kɒlɪdʒ/

- b) Which two heteronyms are words used to describe workers who have joined together for self-protection, and a chemical compound that has not dissociated electrically?
- c) Which Indian film director and American photographer and Dadaist had heteronymic surnames? (The Telegraph, April 13, 2015)⁵

Another challenging pronunciation topic a student can focus on is the hard-to-pronounce names in the global fashion industry, which may include words from a variety of languages, e.g. *Givenchy*, *Hermès* or *Loewe*. In the i-D's video (i-D's A-Z, 2014), which is an A-Z compilation of fashion's trickiest names, we are taken through the alphabet as pronounced by Lindsay Wixon and twenty-six other models dressed in the clothes of the designers that they are representing. In yet another video, by *Business Insider* (Stuart & Kakoyiannis, 2014), fifteen challenging brand names, including *Hermès*, *Fage* and *Stella Artois*, are pronounced and the viewers are also reminded to avoid their common mispronunciations.

For a non-native teacher of pronunciation, aware of the frequent lack of a straightforward correlation between spelling and pronunciation, memorization of quite large numbers of proper names and other deceptive words is simply an inseparable part of a life-long process of learning English as a foreign language. It seems that, as a natural consequence, students of English Departments, future interpreters and teachers, should be sensitized to this issue and should also be taught to build on their phonetic awareness in this respect. A course in English Phonetics, could, start with Poole's (2014) quotation about the annoying mispronunciation of the word *pronunciation*, as this erroneous enunciation is one of the likely ways that first year students may say the word: "Possibly the most amusingly disastrous is the mispronunciation of "pronunciation" as "pronounceiation", which hurls the sensitive listener into a hellish abyss of faulty self-reference."

Crossley (2014) reports the findings of a study on frequently mispronounced common words and place names, which point to the fact that more than 75% of British people mispronounce the following words: *Ely* (59%), *Keighley* (40%), *Sherbet* (40%), *et cetera* (34%), *St Pancras* (33%), *espresso* (26%), *bruschetta* (25%), *often* (24%), *prescription* (21%) and *Greenwich* (16%). It is important for teachers to be aware of the fact that it is not only foreigners who have insufficient knowledge of the correct pronunciation of some words. In the classroom these examples

⁵ Answers: a) Mobile /mə(ʊ)'bi:l, 'məʊbi:l/ and mobile /'məʊbaɪl/ (moh-beel, moh-bile); b) Unionised /'ju:niənaɪzd/ and unionized /,ʌn'aɪnənaɪzd/; c) Satyajit Ray /rai/ and Man Ray /reɪ/ (wry, ray).

could serve as words with tricky pronunciation, which students practise and memorize as unique phonetic entities.

We also find it beneficial for the students of phonetics to introduce an amateur pronunciation database like *Forvo* (n.d.), the largest online pronunciation reference, in which words are pronounced by native speakers. In 2013 *Forvo* was among the ‘50 Best Websites of the Year’ in the ranking run by *Time* magazine. In 2017, one could choose from around 3 million words from 344 languages or ask for help with the pronunciation of new words. The reason for pointing to *Forvo* (n.d.) is to familiarize students with this platform, where they can quickly confirm their phonetic guesses concerning, for example, the name of the actress, Siobhan Hewlett, the uptalker from the series *Fortysomething*. The spelling of this Irish forename does not make it clear how the word should be rendered. The likely erroneous options, based on the spelling could be /si'ɒbhən/⁶*, /sai'ɒbən/* or /'sɒnbən/*; however, *Forvo*'s (n.d.) speakers *PaulJWrite* from the UK, *katisings* from the USA and *Mollydub* from Ireland suggest altogether two versions: /ʃə'vɔ:n/ or /ʃɪ'vɔ:n/, both regarded as standard pronunciations (How to pronounce Siobhan, 2008).

14.3.4 British accent: its villainous and sexy appeal

The final area of interest is the image of the British speaker and their accent. We wish to examine the stereotypes concerning the British accent in Ellen's DeGeneres's *Heads-up*, negative associations of posh RP in the Jaguar F-type Coupe 2014 commercial *Good to be bad* (TV Top Car – welcome to the Top, 2014) and its continuation *The Art of Villainy* (Mapauto, 2014, 0:22–1.23) and also its attractive appeal in the comedy *Love actually* (Defender Productor Mix, 2011; *Love actually* script, n.d.), *Dilbert*'s cartoons (Adams, 2005) and the Time Out Global Dating survey (Bourn, 2015).

In Jaguar's *Good to be Bad* (TV Top Car – welcome to the Top, 2014) commercial three British actors: Tom Hiddleston, Mark Strong and Ben Kingsley exhibit the essence of luxury and show how good it is to be bad. Sir Ben Kingsley asks at the start: “Have you ever noticed how in Hollywood movies all the villains are played by Brits?” Mark Strong answers: “Maybe we just sound right,” and proceeds to take his Jaguar for a ride. Tom Hiddleston, who is flying over London in a helicopter, adds “We're more focused, more precise, always one step ahead”. And then we hear “with a certain style, an eye for detail and we are obsessed by power.

⁶ The asterix * stands for erroneous pronunciation and transcription.

Stiff upper lip is key.” We may ask what it means to sound right or more precise, or focus on the expression “stiff upper lip,” which could also lead to different lip shapes while producing vowels, etc.

In another commercial for the Jaguar F-type coupe, we can see Tom Hiddleston stating that Brits are believed to play the best villains and then rhetorically asking what makes a great villain (Mapauto, 2014). His monologue includes a long description of the essential prerequisites: “Firstly, you need to sound distinct. To speak with eloquence that lets everyone know who is in charge. A villain should have style. A suit should always be bespoke, razor-sharp like your wit. It is important that a villain has the means to stay one step ahead. World domination starts with an attention to detail.” (Mapauto, 2014, 0:22–1:23)

This could raise the issues of what it means to sound distinct or whether attention to detail matters in pronunciation. We could also wonder what is involved in speaking in such a way that “lets everyone know who is in charge,” which might lead us to the function of different tones such as a likely use of frequent fall-rises in a chairperson’s speech.

As regards the stereotypes connected with the British accent, we could once again make us of Ellen’s *Heads up!* game. In these interviews three American actors, Joseph Gordon-Levitt (TheEllenShow, 2014a, 0:47–0:51), Fergie (TheEllenShow, 2014a, 0:33–0:40) and Jennifer Love Hewitt (TheEllenShow, 2014b, 0:54–0:58) are asked to imitate the British accent. As one can note, they make reference to a typical British attribute, which is either ‘a pint’ of beer or ‘a cup’ and also ‘a spot of tea’.

Scott Adams’ (2005) *Dilbert* cartoons also cover the issue of the sex appeal of the British accent, even a fake one. In the cartoon of May 31, 2005 the Boss is talking to a cat, Catbert: “I thought I hired a genius. But he turned out to be an ordinary guy faking a British accent.” Then this man comes in and greets his female colleague: “ ’Ello, bird. ’Ow ’bout a spot o’tea? Whot do you say, gov’nor?” Alice is wondering: “Who’s making those sexy sounds?” The light-hearted message we get from it is that even an artificial British accent does wonders and is favoured by the opposite sex.

The same can be found in the *Dilbert* cartoon of June 01, 2005 in which we can witness a conversation in an office between two female colleagues, here called A for Alice and B for the other interlocutor, just before a male colleague with a British accent comes in, referred to as C: B: “I think I’m in love with the new guy because of his fake British accent;” A: “You’re married;” B: “I am? Wow! His British accent made me forget;” C: “I say, old beans, did anyone see my brolly on the lift?” A: “Swoon;” B: “I’m single.”

Another comment about the cuteness of the British accent, “American girls would dig me with my cute British accent”, comes from the comedy *Love actually* (Defender Productor Mix, 2011; *Love actually* script, n.d.), in which Colin’s accent acts as an aphrodisiac. In the key accent scene, set in a bar, we can see that the distinctiveness of British pronunciation is a good start for a conversation. Much to their disappointment, the girls soon find out that vocabulary does not differ so much between their own and Colin’s varieties of English.⁷

S: Oh my God. Are you from England? C: Yes. S: Oh, that is so cute. Hi, I’m Stacey. Jeannie? J: Yeah? S: This is Colin. C: Frissell. J: Cute name. Jeannie. S: He’s from England. C: Yep, Basildon. J: Oh. S: Oh. J: Wait till Carol Anne gets here. She’s crazy about English guys. CA: Hey, girls. J: Carol Anne, come, meet Colin. He’s from England. CA: Well, step aside ladies this one is on me. Hey, gorgeous. C: (Gentle growl). ... S: That is so funny. What do you call that? C: ‘Bottle’, Girls: ‘Bottle.’ CA: What about this? C: Er, ‘straw.’ Girls: ‘Straw.’ J: What about this? C: ‘Table.’ Girls: ‘Table’ Oh, it’s the same.

Students may also be motivated to excel in the British accent by the *Time Out* Global Dating survey (Bourn 2015), which revealed that the sexiest foreign accent in the world is British, chosen by 26.7 %. Inhabitants of Sydney, New York and Paris are most likely to be infatuated with a speaker of the British accent.

This immediately became a hot issue in the press (Gabbatt, 2015; Scott, 2015 and Singh, 2015). There was not, however, any distinction made between regional accents because the survey failed to specify the results and to point to any one specific British accent. Confusingly, we learn that ‘Scottish’ counted separately in the poll, coming in eighth. However, it is assumed that the British accent addressed in the survey was a Hugh-Grant-style standard English.

The questionnaire involved 11,000 respondents from 24 international cities all over the world, including London, Paris, New York, Chicago, Los Angeles, Hong Kong, Kuala Lumpur, Singapore, Sydney, Melbourne, Tokyo and Beijing. The following three varieties of English came next: American (8.7%), Irish (8.1%) and Australian (8%). The French accent as L1 was relegated to fifth place (7.7%) and lost its title as the language of love.

⁷ In the following passage the names are abbreviated to: C for Colin, CA for Caroline Anne, J for Jeannie and S for Stacey.

Scott (2015) makes the observation that although the British accent has been crowned as the most dateable, a glaring problem remains as there is no such thing as a uniform British accent. She further comments:

[p]erhaps it's the perfectly annunciated 'BBC English' made popular by Hugh Grant and John Cleese ... Then again, it could very likely be the northern English spoken by the Game of Thrones Stark Family ... Or is it the Geordie accent that has taken off thanks to Cheryl Cole and the orange cast of the Geordie Shore? ... Or maybe there are fetishes for Cockney, Welsh, Lancashire, Brummie, Bristolian, Liverpoolian or Cornish accents.

14.4 Conclusion

This chapter has aimed at applying elements of pop culture in pronunciation teaching on the basis of fragments of a variety of authentic materials. The selected excerpts of speech or written texts have touched upon a wide array of phonetic issues, which are usually discussed in a descriptive grammar course. In this analysis, we have grouped them into four parts, i.e. segments, prosody, spelling-to-sound correspondence and the British accent. The first category covers such features as 'th' native and non-native substitutions, th-fronting, phonotactic constraints and the notion of a foreign accent. Our focus on intonation has been deliberately limited to uptalk which, as observed by Wells (2006:38), should not be overdone by EFL learners as it is never essential and may annoy non-uptalkers who regard it as an inappropriate choice of tone.

The discrepancy between a letter and a sound is exemplified with unique pronunciation of proper names with reference to geographical places, brands and fashion designers' and inventors' surnames. Some British accent oriented issues such as stereotypes, negative associations of posh RP, and also its attractive appeal, complete our discussion. We also recommend the use of an online pronunciation reference database along with authentic materials of visual, audio and audio-visual nature.

It is hoped that we have managed to evidence that a phonetic course can be enriched by carefully selected elements of pop culture and authentic materials which encompass a number of phonetic issues. We have made an attempt to show that some examples of Internet trivia can stimulate a phonetic discussion and exemplify a pronunciation-oriented topic in a convincing, humorous and memorable way. We have also intended to refresh this well-known source of inspiration by exemplifying the richness of real texts and spoken samples in the realm of English phonetics.

It is believed that this discussion will encourage pronunciation teachers to implement authentic materials in their pronunciation teaching, not only in order to make the course more attractive for learners but also to make the students responsible for their own pronunciation and development of their receptive skills as well as knowledge, which in turn should lead to their greater sensitivity as listeners and gradually greater intelligibility as English users.

References

- Adams, S. (2005). Dilbert [Cartoons], Tuesday May 31, 2005 & Wednesday June 01, 2005, UFS Inc., Universal Uclick. Retrieved on October 26, 2015, from http://dilbert.com/search_results?page=16&sort=date_asc&year=2005
- BBCWorldwide. (2007, March 2). Valley girl – in this bar – The Catherine Tate Show – BBC comedy [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=-fGZtrBeDcQ>
- bbs7000. (2010, April 18). Cindy Crawford – Arctic [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=bJTqK1-BkDM>
- Bourn, Ch. (2015). The world of dating in 2015. *Time Out*, February 4, 2015. Retrieved on October 22, 2015, from <http://www.timeout.com/dating-2015/>
- Bradford, B. (2008). The rise of ‘upspeak’. In: Collins, B. & Mees, I. M. (Eds.), *Practical Phonetics and Phonology*, pp. 248–253. London: Routledge.
- Buist, E. (2013). Is Britain becoming a nation of accent snobs? *The Guardian*, 2 October 2013. Retrieved on October 22, 2015, from <http://www.theguardian.com/commentisfree/2013/oct/02/britain-accent-sound-foreign-snobs>
- Brown, G. (1990). *Listening to Spoken English*. Harlow: Pearson Education.
- Cauldwell, R. (2013). *Phonology for Listening: Teaching the Stream of Speech*. Birmingham: Speech in Action.
- Celce-Murcia, M. D., Brinton, M. & Goodwin, J. M. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Oxford: Oxford University Press.
- Chipotle Mexican Grill. (2011, August 24). Chipotle story – how is all started [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=wmH73Diqf5Q>

- Cockburn's Boat, Passport, Rescue, Submarine [Video files]. (n.d). Retrieved on October 22, 2015, from <http://www.cockburns.com/index.php?p=famous-70s-ads>
- Compartiendo Publicidad. (2012, February 13). Cockburn's Port: 12 incher (REINO UNIDO) [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=187PIyFY-So>
- Crossley, L. (2014). The words we can't get right: Why 80 per cent of British people struggle to pronounce common words including espresso and prescription (and women are the worst), *Daily Mail*, 10 March 2014. Retrieved on October 22, 2015, from <http://www.dailymail.co.uk/news/article-2577432/More-80-cent-British-people-speak-properly-Most-commonly-mispronounced-words-include-espresso-prescription-women-worst.html#ixzz3Y3Hf0n9c>
- DecathlonPolska. (2014, December 4). TheCathlons [Video file]. Retrieved on October 22, 2015, from https://www.youtube.com/watch?v=MS-HqY_Ogm4
- Defender Productor Mix. (2011, May 21). Love actually – USA Scene [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=UtGjDHE2oVk>
- Field, J. (2008). *Listening in the Language Classroom*. Cambridge: Cambridge University Press.
- Forvo: All the words in the world pronounced by native speakers [Online pronunciation dictionary]. (n.d.). Retrieved on October 31, 2017, from <http://www.forvo.com/>
- Gabbatt, A. (2015). Roger that: British confirmed as world's sexiest accent. But which one? *The Guardian*, 12 February 2015. Retrieved on October 22, 2015, from <http://www.theguardian.com/lifeandstyle/2015/feb/12/roger-that-worlds-sexiest-accent-confirmed-british-but-which-one>
- Getting Your Brand Name Pronounced Correctly. (2013). *Great Brands*. Retrieved on October 22, 2015, from <https://greatbrands.wordpress.com/2013/03/24/getting-your-brand-name-pronounced-correctly/>
- Gif's inventor says ignore dictionaries and say 'Jif'. (2013, May 22). *BBC Technology*. Retrieved on October 22, 2015, from <http://www.bbc.com/news/technology-22620473>
- Goman, J. (1993). On Language; Like, Uptalk?, *The New York Times*, 15 August 1993. Retrieved on October 22, 2015, from <http://www.nytimes.com/1993/08/15/magazine/on-language-like-uptalk.html>

- Gonet, W. & Pietroń, G. (2006). English interdental fricatives in the speech of Polish learners of English. In: Sobkowiak, W. & Waniek-Klimczak, E. (Eds.), *Zeszyty Naukowe Państwowej Wyższej Szkoły Zawodowej w Płocku, Dydaktyka fonetyki języka obcego*, Soczewka 27–27 kwietnia 2005, pp. 155–165. Płock: Wydawnicwo Państwowej Wyższej Szkoły Zawodowej.
- Hancock, M. (2012). Pronunciation for listening: Understanding authentic English. Retrieved on October 22, 2015, from <http://hancockmcdonald.com/sites/hancockmcdonald.com/files/filedownloads/IATEFL%20Poland%20talk%20article.pdf>
- How to pronounce Siobhan. (2008, May 11). Forvo: All the words in the world Pronounced [Audio file]. Retrieved on October 22, 2015, from <http://pl.forvo.com/word/siobhan/#en>
- i-D's A-Z of fashion name pronunciation [Video file]. (2014, March 24). *The Telegraph Fashion*. Retrieved on October 22, 2015, from <http://fashion.telegraph.co.uk/videos/TMG10718752/s-A-Z-of-fashion-name-pronunciation.html>
- Lev-Ari, S. & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46, pp.1093–1096.
- Lindsay, G. (2014, October 31). Fings to come [Blog]. Speech talk: thoughts on English, speech & language. Retrieved on October 22, 2015, from <http://englishspeechservices.com/blog/fings-to-come/>
- LiveScienceVideos. (2013, December 05). Examples of Valley Girl “UpTalk” [Video file]. Retrieved on October 22, 2015, from https://www.youtube.com/watch?v=n_WM-Af5svs
- López Casoli, M. & Paderni, R. G. (2013). Using authentic materials in the intonation classroom. In: Przedlacka, J., Maidment, J. & Ashby, M. (Eds.), *Proceedings of the Phonetics Teaching and Learning Conference*, pp. 63–65.
- Love actually script [Script]. (n.d). Retrieved on October 22, 2015, from <http://www.veryabc.cn/movie/uploads/script/loveactually.txt>
- Macleod, D. (2011). Cockburn's Port Pronounce Responsibly, 8 December 2011, Inspiration Room. Retrieved on October 22, 2015, from <http://theinspirationroom.com/daily/2011/cockburns-port-pronounce-responsibly/>
- Majka327. (2008, September 11). Fortysomething – Episode 2 – part 2/5 [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=ZSHCB1ZAGQ8&list=PL6FC40AD9CFDA8E80&index=7>

- Mapauto. (2014, April 4). British Villains – Tom Hiddleston en Jaguar F-Type Coupé [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=nOgDNmJExVc>
- McGrath, I. (2002). *Materials Evaluation and Design for Language Teaching*. Edinburgh: Edinburgh University Press.
- Muir, H. (2014). Do accents matter in modern Britain? *The Guardian*, 14 July 2014. Retrieved on October 22, 2015, from <http://www.theguardian.com/uk-news/2014/jul/14/do-accent-matter-modern-britain>
- Nowacka, M. (2008). Phonetic Attainment in Polish University and College Students of English: a Study in the Productive and Receptive Pronunciation Skills. Unpublished Ph.D. dissertation, Lublin: Uniwersytet Marii Curie-Skłodowskiej.
- Pisoni, D. B. & Lively, S. E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In: Strange, W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research*, pp. 433–462. Timonium, MD: York Press.
- Poole, S. (2014). Speaking proper: Does it matter if we can't pronounce mispronunciation? *The Guardian*, 11 March 2014. Retrieved on October 22, 2015, from <http://www.theguardian.com/science/2014/mar/11/speaking-pronounce-mispronunciation-english-elocution-expresso-nucular>
- Porzuczek, A., Rojczyk, A. & Arabski, J. (2013). *Praktyczny Kurs Wymowy Angielskiej dla Polaków*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Scott, K. (2015). British wins sexiest accent while Aussie accent comes in fourth. *Australian Times*, 16 February 2015. Retrieved on October 22, 2015, from <http://www.austliantimes.co.uk/british-wins-sexiest-accent-while-aussie-accent-comes-in-fourth/>
- Shockey, L. (2003). *Sound Patterns of Spoken English*. Oxford: Blackwell.
- Singh, A. (2015). British accent is more attractive than French, says survey. *The Telegraph*, 8 February 2015. Retrieved on October 22, 2015, from <http://www.telegraph.co.uk/news/newstopping/howaboutthat/11399001/British-accent-is-more-attractive-than-French-says-survey.html>
- Sobkowiak, W. (1996). *English Phonetics for Poles*. Poznań: Bene Nati.
- . (2000). *Fonetyka angielska w pigulce 2.0*. Konin: Wydawnictwo Wyższej Szkoły Zawodowej.

- (2003). Materiały ulotne jako źródło metakompetencji fonetycznej. In: Sobkowiak, W. & Waniek-Klimczak, E. (Eds.), *Zeszyt Naukowy Instytutu Neofilologii Państwowej Wyższej Szkoły Zawodowej w Płocku Vol. 5, Dydaktyka Fonetyki Języka Obcego*, pp. 151–166. Płock: Wydawnictwo Państwowej Wyższej Szkoły Zawodowej.
- (2012). This is Tom = /Zyzys'Tom/ Pronunciation in Beginners' EFL Textbooks Then and Now. *Research in Language*, 10/1, pp. 111–122.
- Stuart, M. & Kakoyiannis, A. (2014, September 10) The Right way to say 15 brand names you're mispronouncing all the time [Video file]. Business Insider. Retrieved on October 22, 2015, from <http://www.businessinsider.com/mispronounce-difficult-to-say-brands-2014-9>
- Swift, J. (2014). Chipotle helps the British with pronunciation in first UK campaign. *Campaign*, 19 March 2014. Retrieved on October 22, 2015, from <http://www.campaignlive.co.uk/news/1285798/>.
- Szpyra-Kozłowska, J., Frankiewicz, J. & Gonet, W. (2002). Aspekty fonetyki angielskiej nauczane w polskich szkołach średnich. [Aspects of English phonetics taught in Polish secondary schools]. In: Sobkowiak, W. & Waniek-Klimczak, E. (Eds.), *Zeszyty Naukowe Państwowej Wyższej Szkoły Zawodowej w Płocku, Dydaktyka fonetyki języka obcego I*, pp. 9–27. Płock: Wydawnictwo Państwowej Wyższej Szkoły Zawodowej.
- Szpyra-Kozłowska, J. (2014). *Pronunciation in EFL Instruction: A Research-Based Approach*. Bristol: Multilingual Matters.
- TheEllenShow. (2014a, January 9). Joseph Gordon-Levitt Does Accents with 'Heads Up!' [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=380gVwbnjJA>
- (2014b, April 1). Jennifer Love Hewitt Plays 'Heads Up!' with Ellen [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=rSiIO6EkHkE> (0:54–0:58)
- (2014c, May 14). Emily Blunt and Ellen Play 'Heads Up!' [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=yzS0KbYZQpU>
- (2014d, October 23). Ellen and Fergie Play 'Heads Up!' [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=xDpKOiciUBU>
- Those famous 70's ads: a Port with character and humour. (n.d.). The Home of Cockburn's Port Wine | It's a Family Thing. Retrieved on October 22, 2015, from <http://www.cockburns.com/index.php?p=famous-70s-ads>

- Thorn, Sh. (2012). Debunking authentic listening. *Modern English Teacher*, 21(2), pp. 65–69.
- Tusk, D. [Donald Tusk]. (2014, November 28). Donald Tusk: president-elect of the European Council – A biography in his own words [Video file]. Retrieved on October 22, 2015, from <https://www.youtube.com/watch?v=6BU6jB4fu3A>
- TV Top Car – welcome to the Top. (2014, October 28). Jaguar[:] It[']s good to be bad | TV Top Car [Video file]. Retrieved on October 22, 2015, from https://www.youtube.com/watch?v=B-_30ZRZAI0
- Underwood, M. (1989). *Teaching Listening*. New York: Longman Group.
- University Challenge: Are these the 20 toughest questions Jeremy Paxman has ever asked. (2015, April 13) *The Telegraph*. Retrieved on October 22, 2015, from <http://www.telegraph.co.uk/culture/tvandradio/10744941/University-Challenge-Are-these-the-20-toughest-questions-Jeremy-Paxman-has-ever-asked.html>
- Webcomic/YetAnotherFantasyGamerComic. (n.d.). Retrieved from http://static.tvtropes.org/pmwiki/pub/images/valleyelf_6391.jpg
- Wells, J. C. (2006). *English Intonation: An Introduction*. Cambridge: Cambridge University Press.