Vol. 12.2 2014

Research in Language

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

RiL Vol. 12.2 2014

Research in Language

edited by
EWA WANIEK-KLIMCZAK

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

ŁÓDŹ 2014

EDITORS

Ewa Waniek-Klimczak – University of Łódź, Faculty of Philology
Department of English Language and Applied Linguistics
90-236 Łódź, Pomorska 171/173

PUBLISHING REVIEWERS

Hanna Pułaczewska (University of Regensburg), *Agnieszka Piskorska* (Warsaw University)

All papers published in RiL have been blind-reviewed by at least two independent readers

EDITING AND TYPESETTING

Przemysław Ostalski

COVER DESIGN

Barbara Grzejszczak

Printed directly from camera-ready materials provided to the Łódź University Press

Published by Łódź University Press
First Edition. W.06759.14.0.C

Printing sheets 6,25

ISSN 1731-7533

Łódź University Press
90-131 Łódź, Lindleya 8
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
phone (42) 665 58 63, fax (42) 665 58 62

CONTENTS

 DOI: 10.2478/rela-2014-0004

FORMULAIC SEQUENCES AS FLUENCY DEVICES IN THE ORAL PRODUCTION OF NATIVE SPEAKERS OF POLISH

EWA GUZ
John Paul II Catholic University of Lublin
ewasik@o2.pl

Abstract
In this paper we attempt to determine the nature and strength of the relationship between the use of formulaic sequences and productive fluency of native speakers of Polish. In particular, we seek to validate the claim that speech characterized by a higher incidence of formulaic sequences is produced more rapidly and with fewer hesitation phenomena. The analysis is based on monologic speeches delivered by 45 speakers of L1 Polish. The data include both the recordings and their transcriptions annotated for a number of objective fluency measures. In the first part of the study the total of formulaic sequences is established for each sample. This is followed by determining a set of temporal measures of the speakers' output (speech rate, articulation rate, mean length of runs, mean length of pauses, phonation time ratio). The study provides some preliminary evidence of the fluency-enhancing role of formulaic language. Our results show that the use of formulaic sequences is positively and significantly correlated with speech rate, mean length of runs and phonation time ratio. This suggests that a higher concentration of formulaic material in output is associated with faster speed of speech, longer stretches of speech between pauses and an increased amount of time filled with speech.

Keywords: formulaic sequences, speed fluency, breakdown fluency, temporal speech measures

1. Introduction

Research into L1 speech suggests that formulaic language contributes significantly to fluent language production (Wray and Perkins 2000; Wray 2002; Wray 2008; Pawley 2009). Pawley and Syder (1983, 2000) report that native-like oral fluency depends largely on the speaker's ability to rely on automatized repertoires of prefabricated chunks which reduce the amount of processing and encoding involved in speech production and afford the speaker the time to attend to other aspects of the speaking process. Selecting a formulaic rather than word-by-word formulation of the utterance results in the preservation of the speaker's cognitive resources and has a direct bearing on the temporal characteristics of the utterances, which are produced rapidly and smoothly without hesitations and pauses (Raupach 1980; Gatbonton and Segalowitz 1988; Towell, Hawkins and Bazergui 1996). The central goal of the present study is to establish whether the use of formulaic sequences in native speech is linked to the speaker's overall productive fluency. In particular, we attempt to determine the correlation between the number of formulaic sequences identified in the oral output of native speakers of Polish and a range of preselected speed and breakdown fluency measures. Our discussion begins with a brief definition of the two key concepts investigated in this study: fluency and formulaicity. This is followed by an overview of research into the relationship between fluency and the use of formulaic language in L1 and L2 speech. Finally, the research methodology adopted for the purposes of the study as well its results are described.

2. Fluency as a production/performance phenomenon

Speech fluency is a complex phenomenon which has been the subject of extensive research over the years. Earlier accounts of fluency saw it as a multifaceted construct which embraces a variety of aspects of language use ranging from sociolinguistic appropriateness through linguistic creativity to expressing ideas at length in coherent semantically dense sentences with few pauses and fillers (Fillmore 1979: 93). Later accounts define fluency as a psycholinguistic and speech-related phenomenon (Lennon 2000; Gatbonton and Segalowitz 2005) which has practical applications as a performance descriptor and indicator of progress (Chambers 1997). Understood in this way, fluency is synonymous with speech fluidity defined as the ability to efficiently "translate thought and communicative intentions into speech under the temporal constraints of online speech processing (Lennon 2000: 26)." Other accounts look at fluency from a more interactive perspective and postulate its additional 'perceived' or 'perceptive' dimension (Guillot 1999; Freed 2000; Kormos and Denes 2004; Segalowitz 2010; Götz 2013). As the main focus of this paper is speech production rather than reception, the definition of fluency adopted in this study corresponds to that of Segalowitz's (2010: 48) utterance fluency which he equates with the actual, quantifiable features of oral utterances gauged by a range of temporal measures reflecting the speed of speaking, time filled with speech vs silence, occurrence of pausing as well as hesitation and repair phenomena (for an extensive review of temporal measures used in fluency research see Kormos 2006). In fact, the need for adopting a range of clear-cut variables representing distinct aspects of speech is stressed by Skehan (2003), who recommends that any analysis of fluency should involve a separate treatment of such speech features as silence, speed of speaking, reformulation and automatization. In an attempt to provide a comprehensive and valid fluency description that avoids collapsing its distinctive temporal aspects into one dimension, Tavakoli and Skehan (2005) propose a division of fluency into speed, breakdown and repair fluency. In this framework, fluency is defined in purely temporal terms as a capacity to organize speech in real time and is captured by speed and breakdown fluency (Skehan 2003). The former denotes the number of words uttered per time unit, whereas the latter reflects time filled with speech and is expressed in measurements based on the number, length and location of pauses. A third, additional measure – repair fluency – corresponds to the number of false starts, misformulations, self-corrections and repetitions.

Taking into account the rationale underlying this study, which states that the use of formulaic language lessens the cognitive processing effort involved in online speech production, we have decided to adopt the approach to fluency proposed by Tavakoli and Skehan (2005) as one that captures accurately and reflects directly the temporal nature of speech production[1]. Basing on the calculation of the time spent speaking and pausing respectively, speed and breakdown fluency measures are indicative of the amount of effort needed by the speaker to produce utterances under real time constraints.

3. Definition of formulaic sequences

Recent years have seen a surge of scholarly interest in formulaic language in general and its involvement in the production of fluent L1 and L2 speech in particular (Wray 2002, 2008; Schmitt 2004; Meunier and Granger 2008ab; Corrigan et al. 2009ab; Wood 2001, 2006, 2007, 2010ab; Weinert 2010).

Although there seems to be a general consensus as to the reality of formulaic sequences, they are very difficult to define precisely. Despite the growing body of research in this area, there does not appear to be a single, all-inclusive definition of a formulaic sequence that would embrace the phenomenon in its entirety. Most researchers involved in investigating formulaic sequences emphasize that they constitute a notoriously heterogenous category which accommodates a variety of multiword strings marked by a varied degree of semantic, syntactic, phonological and pragmatic integrity (see for example Wray and Perkins 2000). Wray (2002: 9) provides a definition which has become widely recognized in formulaicity research and is gradually entering mainstream use. She defines a formulaic sequence as:

> (a) sequence, continuous or discontinuous of words or other elements, which is, or appears to be, prefabricated: that is stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.

It is worth noting that Wray's definition clearly prioritizes the holistic storage and retrieval of formulaic language as its prototypical feature. In fact, the psycholinguistic unity of formulaic sequences has been a recurrent theme in the majority of studies. More specifically, one observation that has been frequently made across studies with reference to the single-word-like status of formulaic sequences is that they contribute to the speaker's overall fluency by reducing the cognitive pressures involved in the construction of fluent speech. This line of thinking is based on two key assumptions concerning formulaic sequences on the one hand and the nature of speech processing on the other. These will be discussed in section 4.

[1] Repair fluency is excluded form the scope of the analysis as reformulation phenomena were scarce and appeared only in some of the examined samples.

4. Formulaic sequences as fluency-devices

The central concern of the present study is to investigate the relationship between productive fluency and the use of formulaic language in L1with the view to determining whether the use of formulaic sequences frees up cognitive resources for attending to other aspects of speech, which in turn may lead to an overall more fluent performance in terms of speed of speech and pausing.

The first premise which underlies our reasoning is that spontaneous speech is largely made up of prefabricated word sequences which constitute the preferred lexical choices of native speakers of a given language within a particular speech community (see Peters 1983; Erman and Warren 2000; Schmitt 2004; Wray 2002, 2008). This aspect of natural language use, referred to as ‘native-like selection’ (Pawley and Syder 1983), expresses the tendency of native language users to rely on a relatively closed set of ready-made prefabricated multiword strings rather than draw on freely from a pool of possible combinations of single words bound by the rules of grammar alone. Such ‘productive speech formulas’ constitute the major building blocks of speech and are memorized and holistically retrieved from long-term memory (Pawley 2009). Numerous scholars investigating the patterning of naturally occurring language (Altenberg 1990; Sinclair 1991; Renouf and Sinclair 1991; Aijmer 1996; Eeg-Olofsson and Altenberg 1996; Cowie 1998; Moon 1998; Biber et al. 1999; Hunston and Francis 2000) emphasize the lexically-driven character of spoken language and argue in favour of a view of language which accommodates the ability of native speakers to store, retrieve and recognize formulaic sequences. In such models, speech production is seen as based on two modes of processing; analytic grammar-based generation (and generalisation) as opposed to holistic, item-based handling of formulaic language.

The second dimension of speech which lies at the core of the present discussion is that L1 speech production is a complex mental and highly automatized process that occurs in the presence of heavy time and memory capacity constraints. Kormos (2006: 38) describes human speech as an effortless skill whose mastery is underpinned by automaticity. In psycholinguistic accounts of speech, automaticity is defined as a set of “utterance production skills which refer to the speed and ease of handling utterances (Gabonton and Segalowitz 1988: 473).” Apart from automaticity, production of rapid speech depends on the process of proceduralization whereby declarative (linguistic) knowledge is converted into procedural one (Levelt 1989; Anderson 1983). In this perspective, the cognitive advantage of formulaic sequences comes from their automatized/proceduralized nature, which allows speakers to exceed the limitations imposed by the size of their memory. Handled as single words, the automatized/proceduralized formulaic strings are retrieved directly from long-term memory (without conscious effort, attention or control) bypassing the storage limit of short-term memory and undergo processing in the working memory in larger units (Towell, Hawkins and Bazergui 1996; Lennon 2000; Segalowitz 2003). This, in turn, allows speakers to produce speech in a more fluent manner i.e. with fewer pauses and longer fluent runs between pauses.

Taken together, the findings presented so far suggest that the use of formulaic sequences aids the production of native speech in that it saves the cognitive resources and time, which can be used to attend to other aspects of speech such as e.g. discourse planning or manipulating information (Nattinger and DeCarrico 1992; Wray and Perkins 2000).

Although the validity of the relationship between fluency and formulaicity has been widely recognized, very few studies have investigated the exact nature of this association and little empirical evidence has been provided so far to validate its existence. Raupach (1984) is one of the first studies to emphasize the link between formulaicity and fluency. His study presents evidence of over-reliance on certain formulas (modifiers and rhetorical organizers) in German learners of L2 French. Raupach reports that his participants used formulaic sequences as a production strategy, or as 'islands of reliability.' Similarly, Dechert's (1984) investigation of German students of English (in a task which required them to retell a narrative in L2) reveals that the most fluent learners relied on a number of multiword sequences around which they pieced together a spoken narrative.

More recent accounts have focused on the significance of formulaic language in L2 fluency development. One such example is a series of longitudinal case studies conducted by David Wood (2004, 2006, 2009, 2010b) offering a fine grained analysis of the facilitating influence of automatized lexical phrases on the development of L2 temporal fluency. Wood (2004, 2006, 2010b) investigates fluency gains in individual ESL learners over an extended period of time and reports evidence of formulaic sequences functioning as fluency-enhancing devices. Wood (2009) presents the results of a longitudinal investigation of effects of focused instruction of formulaic sequences and fluency on the oral performance of L2 speakers. His results indicate a considerable improvement in temporal fluency after six weeks of focused instruction and a relationship between fluency and use of formulaic sequences in the learner speech samples. In a study based on a similar setup, De Jong, Halderman and Ross (2009) look at the effect of formulaic language instruction on speed and pause fluency in EFL learners at various levels of proficiency. On the one hand, their results show that an increase in the use of formulaic sequences leads to longer fluent runs. On the other hand, however, a higher number of formulaic strings in learner output is associated with an increase in the length of pauses. On the basis of these inconclusive results, the authors state that the use of formulaic sequences by L2 learners is not effortless and that the trained formulaic sequences were probably not stored as chunks and their retrieval was not automatized.

As regards investigating the fluency-enhancing role of formulaic language in L1 speech, to the knowledge of the author of the present investigation, there have been no studies to date that provide solid empirical evidence confirming this relationship between temporal fluency and the use of formulaic language. This study is an attempt at filling this gap.

5. The study

Basing on the assumption that formulaic sequences play some role in the speaker's productive fluency, our primary concern here is to investigate the relationship between the two aspects of native speech. In the analysis that follows, we set out to address the following research questions:

1. Is there a relationship between the use of formulaic sequences and temporal fluency in the speech of Polish native speakers?
2. Is there a correlation between the use of formulaic sequences and speed fluency as measured by articulation rate?
3. Is there a correlation between the use of formulaic sequences and breakdown fluency as measured by the mean length of runs, mean length of pauses and phonation time ratio?
4. Is there a correlation between the use of formulaic sequences and speech rate?

5.1 Participants

Participants included 45 undergraduate students (29 female, 16 male) at the University of Warsaw enrolled at the second year of a teacher training college and working towards their BA in teaching English as a Foreign Language. All of them were of Polish origin and spoke Polish as their first language. Prior to the study, participants were given extensive information about the character and purpose of the research. Written permissions were obtained to indicate explicit consent to participate in the study.

5.2 Procedure

The analysis is based on a 9,979 word data set consisting of L1 monologic speeches elicited through a personalising task (see Appendix 1 for a complete list of topics used as prompts). The data include both the recordings and their transcriptions annotated for pauses. In the first part of the study a set of temporal fluency measures of the speakers' output are determined. This is followed by establishing the total of formulaic sequences for each sample. Finally, to gauge the nature of the relationship between formulaicity and speech fluency Pearson product-moment coefficients are established for these two variables.

5.2.1 Measuring utterance fluency

As regards the pause annotation procedure, we followed a well-established research tradition (see Goldman-Eisler 1968; Towell, Hawkins and Bazergui 1996; Freed et al. 2004; Segalowitz and Freed 2004; Wood 2006; de Jong et al. 2012), where a pause is defined as a silence or a non-verbal filler of 0.25 seconds or longer. A visual representation was produced for each sample using PRAAT software (Boersma and Weenink 2005) to indicate pause boundaries and identify their duration. Pauses were then indicated in text using square brackets. Syllables were counted consulting the original recording when necessary. For reasons already explained, in this study we adopt the fluency measurement paradigm put forward by Skehan (2003, 2009) and Tavakoli and Skehan (2005) followed in many recent fluency studies (Housen, Kuiken and Vedder 2012). This means that we distinguish between two main dimensions of fluency: speed and breakdown fluency. Speed fluency is measured on the basis of articulation rate

(AR) which reflects the number of syllables divided by speaking time. Following on from the observations of Goldman-Eisler (1968), De Jong et al (2012: 124) argue that articulation rate constitutes a reliable fluency measure unaffected by situational and individual variability. However, speech rate of samples, which is expressed in the number of syllables per time unit and combines aspects of both speed and breakdown fluency was also established as it is regarded as one of the most reliable predictors of fluency (Kormos 2006: 162). Breakdown fluency refers to the frequency and length of pausing as expressed in measures such as: mean length of runs (MLR), mean length of pauses (MLP) and phonation time ratio (PhTR) denoting the total length of speech divided by the total time spent speaking) (Kormos 2006: 163). MLR is a measure of pause frequency in relation to words produced and is established by dividing the total of syllables by the total of runs. To arrive at a complete set of temporal measurements, a range of temporal input values was initially established for each sample including: overall speech time, pausing time, total of pauses (filled and silent) and total of runs between pauses. This was followed by computing the following set of fluency measures: AR, SR, MLR, MLP and PhTR.

5.2.2 Identification of formulaic sequences in the data

The second part of the procedure involved the identification of formulaic sequences in the collected data. The major problem inherent in extracting formulaic sequences from text in a consistent way is that their only truly essential characteristic (that could serve as the ultimate proof of the formulaic status) i.e. their holistic handling, is very difficult, if not impossible, to determine. In the current state of research there is no direct, unambiguous method of ascertaining that a string is stored and handled as a whole (Wray 2008: 113). Taking into account that formulaic status is not always signalled in a straightforward, objectively measurable way, it is particularly difficult to distinguish formulaic strings from grammatically regular and semantically transparent novel sequences. One way to overcome this procedural impasse is to adopt an eclectic approach (Read and Nation 2004: 33). In the absence of a single unquestionable indicator of formulaicity, the researchers have no alternative but to make use of the available criteria derived on the basis of many years of observation and research. Applying mixed criteria that originate from various areas of research and more than one mode of analysis has the advantage of increasing the likelihood of fishing out the majority of formulaic sequences from the dataset.

Keeping the above observations in mind, for the purpose of this study we have employed an identification procedure developed by Erman and Warren (2000), which relies on restricted exchangeability (also referred to as non-substitutability) of the constituent elements of a sequence as the main determinant of formulaic status (see also Wiktorsson 2001, 2003; Forsberg 2006; Forsberg and Fant 2010; Knutsson 2006). Restricted exchangeability expresses the degree of fixedness of a given sequence i.e. the fact that it displays some degree of resistance to change. Applying a restricted exchangeability test involves substituting at least one of the elements of a candidate string with a synonym and analyzing the resulting sequence in terms of changes in meaning, function or idiomaticity. For example, replacing the word ‘a doll’ in the expression ‘Thanks, you’re a doll!’ (‘You are very nice. Thank you.’), which is used to express gratitude for somebody’s help or kindness, with ‘a puppet’ or ‘a girl’s toy’ results in a radical change of meaning and the loss of the pragmatic function and idiomaticity. Borderline candidate strings (ones for which restricted exchangeability fails to produce consistent results) are subjected to a further verification procedure which involves testing them in terms of the twelve diagnostic formulaicity criteria outlined in Wray and Namba (2003) (see Appendix 2 for a complete list of their criteria). These features can be described as corresponding to the four major characteristics of formulaic language identified in available research i.e. form, meaning, function and provenance (origin) (Wray 2002).

To sum up, our identification procedure for extracting formulaic sequences consists of three stages. Firstly, samples are examined in search of sequences which appear to be formulaic and candidate formulaic strings are selected. The candidate strings are then subjected to restricted exchangeability test. If they are marked positive for restricted exchangeability, they are listed in the inventory of formulaic sequences. If the restricted exchangeability fails or cannot be applied, the candidate strings need to undergo further verification, which involves validating them in terms of Wray and Namba’s criteria. Once the formulaic strings are successfully identified, their total in each sample is calculated.

5.3 Results and discussion

The central issue addressed in this study concerns the validity of the claim that the use of formulaic language in native speech is linked to a speaker’s productive fluency. As already indicated, the apparent advantage of using automatized and/or proceduralized prefabricated, formulaic strings comes from the fact that their holistic retrieval decreases the processing load involved in online speech production. This grants the speaker with additional time and cognitive resources, which can be delegated to handle other aspects of speech (such as discourse planning or social interaction). As a result, speech is produced more smoothly and rapidly. Following this line of reasoning, higher incidence of formulaic sequences in speech should be linked with an increase in the speed of speaking and the length of runs and a decrease in the length and frequency of pausing.

To address the research questions formulated for this study, Pearson product-moment correlation coefficients were computed as a way of investigating the nature and strength of the relationship between the number of formulaic expressions identified in the speakers’ output and a range of selected speed and breakdown fluency variables. Table 1 summarises the results.

n=45	temporal measure	strength of relationship	significance level
speed fluency	AR	r =.266	p = n.s

n=45	temporal measure	strength of relationship	significance level
breakdown fluency	MLR	r. = .467	p ≤ .05
	MLP	r = -. 291	p = n.s.
	PhTR	r =.402	p ≤ .05
speed and breakdown fluency	SR	r =.457	p ≤ .05

Table 1: Formulaic sequences vs productive fluency

The first research question put forward for this study concerned the relationship between speed fluency (as measured by AR) and the use of formulaic sequences in native Polish speech. As illustrated in Table 1, articulation rate was reported not to correlate significantly with the number of formulaic strings found in the sample (r=.266, p=0.38, with the significance level set at .05). As regards speech rate, which combines speaking speed (AR) with overall pause time, a correlation for the data revealed that speech rate and the number of formulaic sequences were significantly related (r=.457). In other words, as predicted in the earlier parts of this paper, increases in the number of formulaic sequences were linked with increases in speech rate scores suggesting that participants whose output abounded in formulaic language were able to speak faster. This provides some preliminary evidence for the fluency-enhancing function of formulaic language with relation to the speed of speech production.

Our second research question was aimed at investigating the relationship between the use of formulaic sequences and the speaker's breakdown fluency (as measured by MLR, MLP and PhTR). Mean length of fluent runs (MLR) was found to correlate significantly with the number of formulaic strings (r=.467). This correlation indicates that a higher incidence of formulaic strings in speech was linked to an increase in the length of pause-free stretches. For mean length of pause, the correlation returned non-significant values (r=-.291, p=0.71). While the correlation was not significant relative to the standard alpha level of .05, the p-value was less than .10 suggesting that this weak negative correlation approaches significance. The existence of a negative correlation between the use of formulaic language and MLP would suggest that a higher number of prefabs in output is associated with a decrease in the mean length of pause. Combined with the correlation values for MLR, these findings could suggest that speakers relying on a more extensive use of formulaic strings produce discourse marked by fewer and shorter pauses.

However, as already stated, the MLP results need to be approached with caution as this correlation returned marginally significant values. The final measure employed as a gauge of breakdown fluency is that of phonation time ratio which, as De Jong et al. (2012: 124) point out, is a single measure which subsumes all the measures of silent pausing - MLR, MLP and the number of pauses. Overall, our results indicate a moderate positive correlation (r=.402) between PhTR and the number of uttered formulaic sequences. To put it differently, increases in the number of formulas used are linked to a higher proportion of time filled with speech.

Taken together, the correlation data obtained in the course of our analysis provide some useful insights into the issue of fluency-enhancing function of formulaic language. As regards both of the investigated dimensions of fluency – speed and breakdown fluency – we found that the use of formulaic strings in the oral native production of Polish native speakers is positively and significantly correlated with mean length of runs, speech rate and phonation time ratio. As we have seen, the data are consistent across the three fluency variables, all of which returned significant and positive values. It is worth pointing out that these three measures (MLR, PhTR and SR) capture the major temporal aspects of speech such as time spent speaking (as opposed to pausing) and the frequency and duration of pauses. Also, it is SR and MLR which are considered the most consistent and reliable temporal indicators of fluency (Kormos and Denes 2004; Wood 2004; Kormos 2006; Préfontaine 2010). On the basis of the speed and breakdown fluency data, it can be concluded that speech characterised by a higher incidence of prefabricated language consists of longer uninterrupted stretches of discourse and is produced with a higher number of words/syllables uttered per time unit. In short, speech which was rich in formulaic language is produced more rapidly and with fewer hesitation phenomena.

However, it has to be noted that although our results provide some preliminary evidence validating the link between formulaicity and fluency in L1 oral performance, at the same time we have not managed to gather a convincing body of data that would substantiate the claims concerning the relatively high strength of this relationship. In fact, the established correlation values did not exceed r=.5, which indicates a relatively weak relationship between the investigated variables. Additionally, two out of the five investigated fluency measures returned non-significant values. Although the number of participants was higher than in most previous research, the limitation of this study is its small sample size. However, the major issue facing any analysis of pause distribution in a larger body of naturally occurring discourse (native or non-native) is the lack of reliable pause annotation in the available spoken corpora (Dahlmann, Adolphs and Rodden 2007). The current annotation conventions fail to accommodate a number of essential aspects of pausing such as: the distinction between silent and filled pauses and pause duration (which are relevant in fluency research). Finally, some objections might be raised in connection with the identification procedure used to extract formulaic sequences from text, which involves some degree of subjectivity.

In sum, it can be stated that our research has lent some support to the claims concerning the fluency-enhancing function of formulaic language. From a psycholinguistic perspective, formulaic sequences provide adult language users with a kind of 'a shortcutting device' (Peters 1983: 2) by reducing the processing load involved in the production of speech. The considerable computational power involved in constructing each utterance afresh leads native speakers to "economize by stitching together language chunks (…) so that planning for the form and content of future utterances proceeds more smoothly" (Skehan 1998: 4). Relying on formulaic language allows language users to save planning time and processing effort and direct their attention to other aspects of speech. The major contribution of this study is that it relates the use of formulaic language to a set of objectively measured temporal fluency measures and so, provides a clear, numerical illustration of the speaker's processing gains.

In fact, ample evidence for the processing advantage of formulaic language comes from research on highly conventionalized speech of so called 'smooth talkers' (Schmitt and Carter 2004: 5) whose performance involves the need to maintain fluency in the face of considerable overload of working memory and additional external pressures. Kuiper (2004: 43-45) provides an extensive overview of relevant work in this area, which includes analyses of the performance of auctioneers, sport commentators of fast-paced sports such as ice-hockey and horse-racing and slow-paced sports such as cricket, hockey, weather forecasters and check-out assistants. The professionals from within all of these occupations are required to memorize large stocks of highly detailed information and engage in a kind of non-stop on-line processing to maintain a flow of fluent speech (be it a commentary, a forecast or a sales talk) in the face of additional pressures such as unexpected events, temporal constraints, the response of the audience etc. Kuiper's (2004) analysis of the speech of such live performers reveals that they avoid using novel formulations and tend to rely on formulaic speech which allows them to convey the maximum of information in the minimum of time, at the same time enabling them to maintain pause-free, non-hesitant speech.

6. Conclusion

The study reported in this paper investigated the relationship between L1 formulaicity and fluency. In the research we analysed speech samples collected from 45 native speakers of Polish in terms of the formulaic language used and temporal fluency. Pearson product-moment correlations were established to investigate the nature and strength of the link between the participants productive fluency and their use of formulaic sequences. The use of formulaic language was found to be positively and significantly correlated with speech rate, mean length of runs and phonation time ratio. This suggests that a higher concentration of formulaic material in participants' output is associated with faster speed of speech, longer stretches of speech between pauses and an increase in the amount of time filled with speech. No statistically significant correlations were found between participants' use of formulaic language and the articulation rate and mean length of pauses. To conclude, in our view this investigation provides some preliminary evidence for the facilitating role of formulaic sequences with relation to the speed of speech production and the occurrence of hesitation phenomena. However, it is recommended that the study is replicated with a larger body of data to provide further insight into the issue.

References

Aijmer, K. 1996. Conversational Routines in English. Longman: London and New York.

Altenberg, B. 1990. Speech as linear composition. In Caie, G. Haastrup, K., Jakobsen, A.L., Nielsen, J.E., Sevaldsen, J., Specht, H. and A. Zettersten, (eds.) Proceedings from the Fourth Nordic Conference for English Studies, Vol. 1. Department of English, University of Copenhagen: 133–143.

Anderson. J. R. 1983. The Architecture of Cognition. Cambridge, Massachusetts: Harvard University Press.

Biber, D., S. Johansson, S. G. Leech, S. Conrad, and E. Finegan. 1999. Longman Grammar of Spoken and Written English. London: Longman.

Boersma, P. and D. Weenink. 2005. PRAAT. Retrieved June, 2013 from http://www.praat.org.

Chambers, F. 1997. What do we mean by fluency? System 25(4): 535–544.

Corrigan, R., E. A. Moravcsik, H. Ouali, and K. M. Wheatley. 2009a. Formulaic Language: Volume 1: Distribution and Historical change. Amsterdam: John Benjamins.

Corrigan, R., E. A. Moravcsik, H. Ouali, and K. M. Wheatley. 2009b. Formulaic Language: Volume 2: Acquisition, Loss, Psychological Reality, and Functional Explanations. Amsterdam: John Benjamins.

Cowie, A. P. 1998. Phraseology: Theory, Analysis and Applications. Oxford: Clarendon Press.

Dahlmann, I., Adolphs, S. and T. Rodden. 2007. Multi-word expressions fluency and pause annotation in spoken corpora. Paper presented 40th BAAL Annual Meeting on Technology, Ideology and Practice in Applied Linguistics, Edinburgh, UK, September 6–8, 2007.

De Jong, N., Halderman, L.K. and M. Ross. 2009. The effect of formulaic sequences training on fluency development in an ESL classroom. Paper presented at the American Association for Applied Linguistics conference 2009, Denver, CO, March 2009.

De Jong, N. H., M. P. Steinel, A Florijn, R. Schoonen, and J. H. Hulstijn. 2012. The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken and I. Vedder (eds.) Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA. Amsterdam: John Benjamins:121–142.

Dechert, H. W. 1984. Second language production: six hypothesis. In H. W. Dechert, D. Möhle and M. Rapuach (eds.) Second Language Productions. 211–230. Tübingen, Germany: Gunter Narr Verlag.

Eeg-Olofsson, M. and B. Altenberg. 1996. Recurrent word combinations in the London Lund Corpus: coverage and use of word-class tagging.” In C. E. Percy, C. E. Meyer, and I. Lancashire (eds.) Synchronic Corpus Linguistics: Papers from the Sixteenth International Conference on English Language Research on Computerized Corpora (ICAME 16). Amsterdam: Rodopi B.V.: 97–108.

Erman, B. and B. Warren. 2000. The idiom principle and the open choice principle. Text, 20 (1): 29–62.

Fillmore, C. J. 1979. On fluency. In D. Kempler and W. S. Y. Wang (eds.) Individual differences in language ability and language behavior, New York: Academic Press: 85–102.

Forsberg, F. 2006. Le Langage Préfabriqué en Francais Parlé L2. PhD Thesis, Stockholm University, Stockholm.

Forsberg, F., and L. Fant. 2010. Idiomatically speaking – effects of task variation on formulaic language in high proficient users of L2 French and Spanish. In D. Wood (ed.), Perspectives on Formulaic Language in Acquisition and Communication, New York: Continuum: 47–70.

Freed, B. 2000. Is fluency in the eyes (and ears) of the beholder? In H. Riggenbach (ed.) Perspectives on Fluency. University of Michigan Press: Ann Arbor: 243–265.

Freed, B. F., N. Segalowitz, and D. P. Dewey. 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(02): 275–301.

Gatbonton, E. and Segalowitz, N. 1988. Creative automatization: principles for promoting fluency within a communicative framework. TESOL Quarterly, 22 (3): 473–492.

Gatbonton, E. and Segalowitz, N. 2005. Rethinking communicative language teaching: a focus on access to fluency. Canadian Modern Language Review, 61: 325–353.

Goldman-Eisler, F. 1968. Psycholinguistics: Experiments in Spontaneous Speech. London: Academic Press.

Götz, S. 2013. Fluency in Native and Nonnative English Speech. Amsterdam: John Benjamins.

Guillot, M.-N. 1999. Fluency and its Teaching. Clevedon, UK: Multilingual Matters.

Housen, A., Kuiken, F. and I. Vedder. 2012. Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA. Amsterdam/Philadelphia: John Benjamins Publishing.

Hunston, S. and Francis, G. 2000. Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English. Amsterdam: John Benjamins.

Knutsson, R. 2006. Formulaic Language in L1 and L2. Licentiate Dissertation. Lund University, Lund.

Kormos, J. 2006. Speech Production and Second Language Acquisition. Mahwah, N.J.: Lawrence Erlbaum.

Kormos, J. and Dénes, M. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. System, 32: 146–164.

Kuiper, K. 2004. Formulaic performance in conventionalized varieties of speech. In N. Schmitt (ed.) Formulaic Sequences: Acquisition, Processing and Use. Amsterdam: Benjamins: 37–54.

Lennon, P. 2000. The lexical element in spoken second language fluency. In H. Riggenbach (ed.). Perspectives on Fluency. Ann Arbor: The University of Michigan Press: 25-42.

Levelt, W. 1989. Speaking from Intention to Articulation. Cambridge, MA: The MIT Press.

Meunier, F. and S. Granger. 2008a. Phraseology in Foreign Language Learning and Teaching. Amsterdam: John Benjamins.

Meunier, F. and S. Granger. 2008b. Phraseology: an Interdisciplinary Perspective. Amsterdam: John Benjamins.
Moon, R. 1998. Fixed Expressions and Idioms in English. A Corpus Based Approach. Oxford Studies in Lexicography and Lexicology. Oxford: Oxford University Press.
Nattinger, J. and J. DeCarrico. 1992. Lexical Phrases and Language Teaching. Oxford: Oxford University Press.
Pawley, A. 2009. Grammarians' languages versus humanists' languages and the place of speech act formulas in models of linguistic competence. In R. Corrigan, E. A. Moravcsik, H. Ouali, and K. M. Wheatley (eds) Formulaic Language: Volume 1: Distribution and Historical Change. Amsterdam: John Benjamins: 3–26.
Pawley, A. and F. H. Syder. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards, and R. W. Schmidt (eds.) Language and Communication. London: Longman: 191–225.
Pawley, A. and F. H. Syder. 2000. The one-clause-at-a-time hypothesis. In H. Riggenbach (ed.), Perspectives on fluency. Ann Arbor: University of Michigan Press: 163–199.
Peters, A. M. 1983. The Units of Language Acquisition. Cambridge: Cambridge University Press.
Préfontaine, Y. 2010. Differences in Perceived Fluency and Utterance Fluency across Speech Elicitation Tasks: A Pilot Study. Papers from Lancaster Univesrity Postgraduate Conference in Linguistics and language Teaching, 5:134–154.
Raupach, M. 1980. Temporal variables in first and second language speech production. In H.W. Dechert and M. Raupach (eds.) Temporal Variables in Speech. The Hague: Mouton: 263–270.
Read, J. and Nation, P. 2004. Measurement of formulaic sequences. In N. Schmitt (ed.) Formulaic Sequences: Acquisition, Processing and Use. Amsterdam: John Benjamins Publishing, 23-35.
Renouf, A. and J. Sinclair. 1991. Collocational frameworks in English. In K. Aijmer and B. Altenberg (eds.) English Corpus Linguistics: Studies in the Honour of Jan Svartvik. London, Longman: 128–143.
Schmitt, N. 2004. Formulaic Sequences. Amsterdam: John Benjamins.
Schmitt, N. and R. Carter. 2004. Formulaic sequences in action: an introduction. In N. Schmitt (ed.) Formulaic Sequences: Acquisition, Processing and Use. Amsterdam: John Benjamins: 1–22.
Segalowitz, N. 2003. Automacity and second languages. In C. Doughty and M. Long (eds.) The Handbook of Second Language Acquisition. Oxford, UK: Blackwell: 382–408.
Segalowitz, N. 2010. The Cognitive Bases of Second Language Fluency. New York: Routledge.
Segalowitz, N. and B. F. Freed. 2004. Context, contact and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. Studies in Second Language Acquisition 26(2): 173–199.
Sinclair J. 1991. Corpus, Concordance, Collocation. Oxford: Oxford University Press.
Skehan, P. 1998. A Cognitive Approach to Language Learning. Oxford: Oxford University Press.
Skehan, P. 2003. Task based instruction. Language Teaching, 36 (1): 1–14.

Skehan, P. 2009. Modelling second language performance: integrating complexity, accuracy, fluency, and lexis. Applied Linguistics 30(4): 510–532.

Tavakoli, P. and P. Skehan. 2005. Strategic planning, task structure, and performance testing. In R. Ellis (ed.) Planning and Task Performance in a Second Language. Amsterdam: John Benjamin: 239–273

Towell, R., Hawkins, R. and N. Bazergui. 1996. The development of fluency in advanced learners of French. Applied Linguistics 17: 84–119.

Weinert, R. 2010. Formulaicity and usage-based language: linguistic, psycholinguistic and acquisitional manifestations. In D. Wood (ed.) Perspectives on Formulaic Language. Acquisition and Communication. London/New York: Continuum: 1–22.

Wiktorsson, M. 2001. Register Differences between Prefabs in Native and EFL English. The Department of English in Lund: Working Papers in Linguistics 1: 85–94. Retrieved January 2009 from http://www.sol.lu.se/engelska/dokument/wp/vol01/Maria.pdf .

Wiktorsson, M. 2003. Learning Idiomaticity: A Corpus-based Study of Idiomatic Expressions in Learners' Written Production. Lund Studies in English. Vol. 105. Lund: Lund University.

Wood, D. 2001. In search of fluency: What is it and how can we teach it? Canadian Modern Language Review, 57: 573–589.

Wood, D. 2004. An empirical investigation into the facilitating role of automatized lexical phrases in second language fluency development. Journal of Language Learning, 2(1): 27–52.

Wood, D. 2006. Uses and functions of formulaic sequences in second language speech: an exploration of the foundations of fluency. The Canadian Modern Language Review 63: 13–33.

Wood, D. 2007. Mastering the English formula: fluency development of Japanese learners in a study abroad context. JALT Journal 29 (2): 209–230.

Wood, D. 2009. Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. Canadian Journal of Applied Linguistics, 12 (1), 39–57.

Wood, D. 2010a. Perspectives on Formulaic Language. Acquisition and Communication. London/New York: Continuum.

Wood, D. 2010b. Formulaic Language and Second Language Speech Fluency: Background, Evidence, and Classroom Applications. London/New York: Continuum.

Wray, A. 2002. Formulaic Language and the Lexicon. Cambridge: Cambridge University Press.

Wray, A. 2008. Formulaic Language: Pushing the Boundaries. Oxford: Oxford University Press.

Wray, A. and K. Namba. 2003. Formulaic language in a Japanese-English bilingual child: a practical approach to data analysis. Japan Journal for Multilingualism and Multiculturalism 9 (I): 24–51.

Wray, A. and M. R. Perkins 2000. The functions of formulaic language: an integrated model. Language and Communication 20(1): 1–8.

Appendix 1

1. Opowiedz o najbardziej cennej rzeczy jaką posiadasz lub przedmiocie, do którego jesteś szczególnie przywiązana/y. Adapted from Inside Out, Student's Book Upper intermediate p. 32, own translation.
2. Opowiedz o tym jak wyglądało twoje życie w wieku około 8 lat. Adapted from Inside Out, Student's Book Intermediate p. 49, own translation.
3. Opowiedz o sytuacji, kiedy robiła/eś coś po raz pierwszy. Adapted from Inside Out, Student's Book Advanced p. 87, own translation.
4. Opowiedz o momencie, w którym poczuła/eś przepływ adrenaliny. Adapted from Inside Out, Student's Book Intermediate p. 38, own translation.
5. Opowiedz o filmie, który zrobił na Tobie duże wrażenie lub kompletnie cię rozczarował. Adapted from Inside Out, Student's Book Upper intermediate p. 108, own translation.

Appendix 2

The Eleven Diagnostic Criteria for Identification of Formulaic Sequences. Adapted from Wray and Namba (2003).

1. A: By my judgement there is something grammatically unusual about this wordstring.
2. B: By my judgement, part or all of the wordstring lacks semantic transparency.
3. C: By my judgement, this wordstring is associated with a specific situation and/or register.
4. D: By my judgement, the wordstring as a whole performs a function in communication or discourse other than, or in addition to, conveying the meaning of the words themselves.
5. E: By my judgement, this precise formulation is the one most commonly used by this speaker/writer when conveying this idea.
6. F: By my judgement, the speaker/writer has accompanied this wordstring with an action, use of punctuation, or phonological pattern that gives it special status as a unit, and/or is repeating something s/he has just heard or read.
7. G: By my judgement, the speaker/writer, or someone else has marked this wordstring grammatically or lexically in a way that gives it special status as a unit.
8. H: By my judgement, based on direct evidence or my intuition, there is a greater than-chance-level probability that the speaker/writer will have encountered this precise formulation before, from other people.
9. I: By my judgement, although this wordstring is novel, it is a clear derivation, deliberate or otherwise, of something that can be demonstrated to be formulaic in its own right.
10. J: By my judgement, this wordstring is formulaic, but it has been unintentionally applied inappropriately.

11. K: By my judgement, this wordstring contains linguistic material that is too sophisticated, or not sophisticated enough, to match the speaker's general grammatical and lexical competence.

STOP RELEASE IN POLISH ENGLISH — IMPLICATIONS FOR PROSODIC CONSTITUENCY

GEOFFREY SCHWARTZ
Faculty of English, Adam Mickiewicz University in Poznań
geoff@ifa.amu.edu.pl

ANNA BALAS
Faculty of English, Adam Mickiewicz University in Poznań
abalas@wa.amu.edu.pl

ARKADIUSZ ROJCZYK
Institute of English, University of Silesia
arkadiusz.rojczyk@us.edu.pl

Abstract
Although there is little consensus on the relevance of non-contrastive allophonic processes in L2 speech acquisition, EFL pronunciation textbooks cover the suppression of stop release in coda position. The tendency for held stops in English is in stark opposition to a number of other languages, including Polish, in which plosive release is obligatory. This paper presents phonetic data on the acquisition of English unreleased stops by Polish learners. Results show that in addition to showing a tendency for the target language pattern of unreleased plosives, advanced learners may acquire more native-like VC formant transitions. From the functional perspective, languages with unreleased stops may be expected to have robust formant patterns on the final portion of the preceding vowel, which allow listeners to identify the final consonant when it lacks an audible release burst (see e.g. Wright 2004). From the perspective of syllabic positions, it may be said that 'coda' stops are obligatorily released in Polish, yet may be unreleased in English. Thus, the traditional term 'coda' is insufficient to describe the prosodic properties of post-vocalic stops in Polish and English. These differences may be captured in the Onset Prominence framework (Schwartz 2013). In languages with unreleased stops, the mechanism of submersion places post-vocalic stops at the bottom of the representational hierarchy where they may be subject to weakening. Submersion produces larger prosodic constituents and thus has phonological consequences beyond 'coda' behavior.

Keywords: coda stop release, L2 speech, phonetics-phonology

1. Introduction

In the area of second language (L2) phonological acquisition, the status of non-contrastive allophonic processes is not entirely clear. One influential theory, Best's Perceptual Assimilation Model (PAM; Best 1995, Best & Tyler 2007) is devoted to the perception of L2 *contrasts* rather than phonemes or allophones, yet success in acquisition is predicted on the basis of sub-phonemic phonetic detail. Flege's Speech Learning Model (SLM; Flege 1995) hypothesizes that L2 phonological acquisition is based on phonetic categories defined at the position-sensitive allophonic level, yet many SLM-inspired studies investigate the implementation of contrasts that are clearly phonological. Similar discrepancies may be found in the closely related area of loanword phonology (for a review, see e.g. Kang 2011), where researchers have made conflicting claims about the relative role of phonetic perception and universal phonological constraints. For example, Hindi and related languages, which contrast /t/ and retroflex /ʈ/, adapt English /t/ as /ʈ/ even though the retroflex stop is absent from the source language. This discrepancy has been attributed to the fact that retroflexion is a contrastive feature in Hindi; its phonological status in the L1 apparently overrides the phonetic properties of the L2 input (Arsenault 2009). In a contrasting example, speakers of French in Quebec and continental Europe show different strategies in their adaptations of English /θ/. In Quebec, speakers substitute /t/ while in Europe they use /s/. The difference is attributed to the fact that /s/ in Eurpean French is dental, while it is alveolar in Quebec French. Since the difference between dental and alveolar /s/ is not contrastive, the loanword phenomenon is claimed to be due to purely phonetic considerations. In sum, research into L2 phonology and loanword adaptation has found apparently conflicting evidence with regard to the role of non-contrastive phonetic features. In what follows, we shall offer a theoretical perspective that may help to reconcile this conflict.

While theoretically-oriented studies disagree about the status of allophonic detail in second language speech acquisition, didactically oriented works in applied English linguistics note that certain allophonic processes are crucial for achieving proficiency. Both Cruttenden (2001) and Cook (2000) devote significant attention to features such as the tapping of /t/ and syllabic productions of nasal consonants. The role of phonetic detail is especially evident in the case of voice contrasts in both word-initial and word-final position. Jenkins (2000) represents a somewhat liberal approach to English phonological acquisition espoused in research on English as a Lingua Franca (ELF). She nevertheless emphasizes the importance of voicing-related allophonic processes for the goal of international intelligibility. The effects of final voicing on the preceding vowel are argued to be crucial to enhance contrasts with a high functional load in the lexicon of English. The same is true of aspiration, which, when not acquired successfully can lead to misperception of the voicing specification of initial stops in learner English. At the same time, Jenkins disregards other allophonic processes, such as the realization of /l/ as 'clear' or 'dark', as being secondary and not crucial for intelligibility. Thus, once again it is difficult to gauge the relative importance of non-contrastive phonetic features in L2 acquisition.

One of the more interesting allophonic processes found in English is the suppression of stop release. Stops may, with varying degrees of frequency and stylistic consequences, remain unreleased in forms such as *kit̚, top̚, hot̚ dog̚,* and of course countless others. As these examples show, unreleased stops may appear before obstruents, as well in absolute final position. The tendency for held stops in English is in stark opposition to a number of other languages, including Polish, in which plosive release is obligatory (Dukiewicz & Sawicka 1995). However, since stop release is generally accepted to be a non-contrastive phonetic feature across languages, its nonphonological status has diminished the impetus for cross-linguistic studies. Consequently, unreleased stops have been largely neglected in the field of second language phonology.

This paper has two goals. First, phonetic data will be presented on the acquisition of English unreleased stops by Polish learners. Our results show that the suppression of stop release, despite its non-contrastive status, is clearly part of the acquisition process for L1 Polish speakers learning English. The second goal of this study is to explore a hypothesis concerning the representation of prosodic constituents, by which unreleased stops arise from purely phonological parameters. In other words, despite its non-contrastive status, stop release is a phonological phenomenon. The hypothesis is formulated within the Onset Prominence framework (Schwartz 2013). The essence of the proposal is that post-vocalic stops come in two varieties. In languages with unreleased stops they are joined with the preceding vowel into a single constituent by a process of *submersion*. In languages with obligatory release the submersion process is parametrically absent.

The submersion proposal makes additional predictions that go beyond the question of whether a stop is produced with an audible release. Since submerged stops are joined with the preceding vowel into a single constituent structure, VC sequences in languages with unreleased stops should be characterized by a greater degree of phonetic cohesiveness than in languages with obligatory release bursts. This cohesiveness may be measured in the VC formant transitions that are observable on vowels in pre-consonantal position. In our study, we measured these transitions in the speech of Polish learners of English. Our results show that in addition to showing a tendency for the target-language pattern of unreleased plosives, advanced learners acquire more native-like VC formant transitions. From the functional perspective, robust formant patterns may be seen as a perceptual license allowing for listener identification of unreleased stops (see e.g. Wright 2004). The acquisition of native-like VC sequences, with suppressed stop release and robust formant patterns, suggests that learners have acquired an aspect of English prosodic organization.

The rest of this paper will proceed as follows. Section 2 will provide background on the phonological and phonetic considerations that are relevant for the study of coda stop release. Section 3 will describe the experimental methods of our study and present the results. Section 4 will provide discussion of the phonological parameter of submersion and the framework in which it is derived.

2. Phonological and phonetic background

In the languages of the world, there are no known cases of a phonological system in which stop release is a contrastive property that may distinguish two phonemes. It is therefore generally assumed that the release of plosive consonants is not phonologically relevant, but rather represents a sub-phonemic phonetic detail. Although this is the mainstream view, some authors have argued that stop release is indeed incorporated into phonological grammars. Evidence for the phonological relevance of stop release may be found in the phonological properties of contour segments such as pre-nasalized stops and affricates; Steriade (1993) shows that only stops and affricates can be split into nasal and non-nasal portions and hypothesizes that stops must therefore be made of two separate root nodes. Others have noted the systematic nature of cross-linguistic generalizations regarding the release of stops in coda position as evidence of the phonological status of stop release. For example, in languages such as Korean, stop consonants in coda position are always produced without release. By contrast, in Polish plosive release is obligatory. Finally, the suppression of stop release may be optional, as we observe in English.

Cross-language interaction of coda stop release systems was observed by Kang (2003), who showed that the probability of epenthetic vowels following English coda stops in loanwords into Korean (*pad* →/pæ.tɨ/) is strongly correlated with the probability of stop release in the speech of L1 speakers. Kang noticed three such effects from a study of the TIMIT corpus of English of American English: (1) voiced stops, (2) stops following tense vowels, and (3) coronal stops were more likely to be released. When the source language stops are more likely to be released, the Korean adaptations showed a strong tendency to be characterized by an epenthetic vowel and an additional syllable. By contrast, the adaptations of English words with a tendency for unreleased stops matched with the native Korean pattern of suppressed stop release (*pack* →/pæk̚/). In sum, despite the non-contrastive status of stop release, there is reason to believe that the phenomenon is indeed phonological.

With regard to the area of English pronunciation pedagogy, textbooks devote some attention to the issue of plosive release. Cruttenden (2001) emphasizes the fact that stop release is suppressed in the first consonant of CC sequences, both at word boundaries and within words. Further, he notes that the failure to suppress stop release may contribute to a tendency for native English speakers to hear an additional [h] sound, or a CV sequence with a schwa in the case of voiced consonant sequences. In studies of cross-language intelligibility, such epenthesis processes have been found to have profound effects on the probability of correct identification of an utterance. This presumably stems from the fact that epenthetic segments affect the prosodic structure to be perceived. Thus, the acquisition of stop release suppression may be suggested to be an instrumental aspect of prosodic learning in L2 English. In the context of L1 Polish speakers learning English, previous studies by Bergier (2010) and Rojczyk et al. (2013) have investigated the production of unreleased stops.

The phonetic forces underlying stop release may be thought in terms of Lindblom's (1990) H&H theory of speech production. In Lindblom's model speech production is realized on a continuum of speech styles ranging from hyper-articulated (careful and emphatic) to hypo-articulated (casual) speech. Lindblom suggests that speakers adjust their utterances along this continuum in order to achieve the goal of 'sufficient discriminability' while economizing effort. In other words, speakers control their speech output according to whether they feel that their utterance will be understood. If a less effortful realization of a given word or phrase may be recovered on the basis of any number of linguistic or non-linguistic factors, speakers will minimize the effort in producing it. If listeners may identify the place of articulation of an unreleased stop, the impetus to spend the effort to produce stop release is lessened.

The recoverability of unreleased stops lies in the acoustic properties that may be observed on the later portion of the vowel in VC sequences (e.g. Wright 2004). The formation of stop closure produces distinct formant patterns on the preceding vowel as a function of the place of articulation, allowing listeners to identify where the constriction is made even in the absence of audible release. Consequently, cross-linguistic differences with regard to the obligatory or optional nature of stop release may be hypothesized to be related to cross-linguistic differences in the robustness of VC formant patterns. In languages that suppress stop release this acoustic robustness reflects a significant degree of phonetic cohesion between the V and the C. In languages like Polish with obligatory release, the acoustic transition is predicted to be less robust than in languages with optional or obligatory suppression of release, reflecting a smaller degree of articulatory coordination in the production of the VC sequence. Differences in phonetic coordination in turn suggest a difference in the prosodic affiliation of the C in VC sequences. In other words, an unreleased stop may be thought of as 'attached' to the preceding vowel, while obligatory stop release suggests that this is not the case. We will return to the implications of stop release in Section 4. Now, we shall turn to a description of the experimental study.

3. Acoustic study of stop release in the speech of Polish learners of English

On the basis of the phonetic and phonological considerations discussed earlier, we hypothesize that Polish learners at different levels of proficiency will differ with regard to the release of post-vocalic stop consonants, and these differences will also be reflected in the formant transitions on the preceding vowel. In what follows, we present an acoustic study investigating this hypothesis.

3.1 Experimental method

Fourteen Polish students of English took part in the experiment. Seven of the students were in the first year of English studies at the Instutute of English at the University of Silesia (Uniwersytet Śląski). The other seven were in advanced years of study (3^{rd} year and higher) at the Faculty of English at Adam Mickiewicz University (UAM) in Poznań. This division formed an independent variable (First Year/Advanced) for our analysis. The advanced students had completed rigorous training in English pronunciation over their first two years at university. The first year group had completed only 6 weeks of this training focusing almost entirely on the English vowel system.

The experiment elicited a series of 36 VC#C sequences in English in which the first consonant were stops. The data set was balanced, containing equal number of tokens with regard to C1 and C2 place of articulation (labial, coronal, dorsal), C1 voicing, and vowel quality (/iː/, /ɪ/, /æ/). Unrelated tokens used for another experiment were also included in the data elicitation task. A total of 504 tokens was analyzed (36 tokens * 14 speakers).

The tokens were presented to the participants on Power Point slides on a computer monitor housed inside a soundproof recording booth at the two universities. Both recording studios are equipped with high quality microphones and USB audio interfaces that allow for recording directly onto the hard drive of a computer. The recordings were made at sampling rate of 44 kHz, with 24 bit quantization.

Acoustic analysis was performed by hand with the help of the Praat program (Boersma & Weenink 2011) and focused on both the release of the post-vocalic stop consonant and the VC formant transitions. With regard to stop release, tokens were tagged as either Yes or No, on the basis of whether there was both a visual spike in the waveform as well as an audible burst. With regard to the vowels in the VC sequences, the following acoustic measurements were made.

- Overall vowel duration (including pre-vocalic /r/ and /l/) in milliseconds
- Duration of VC transition in milliseconds, from the end of the steady-portion of the vowel to the onset of stop closure.

From these measurements, one additional measure of VC formant transitions was calculated, %VC, defined as the transition duration divided by the vowel duration multiplied by 100. An illustration of the VC transition measurement is given in Figure 1, which is taken from a token of the phrase *drag down*. In the selected portion of the figure one may observe the formant excursion from the vowel target positions for the /æ/.

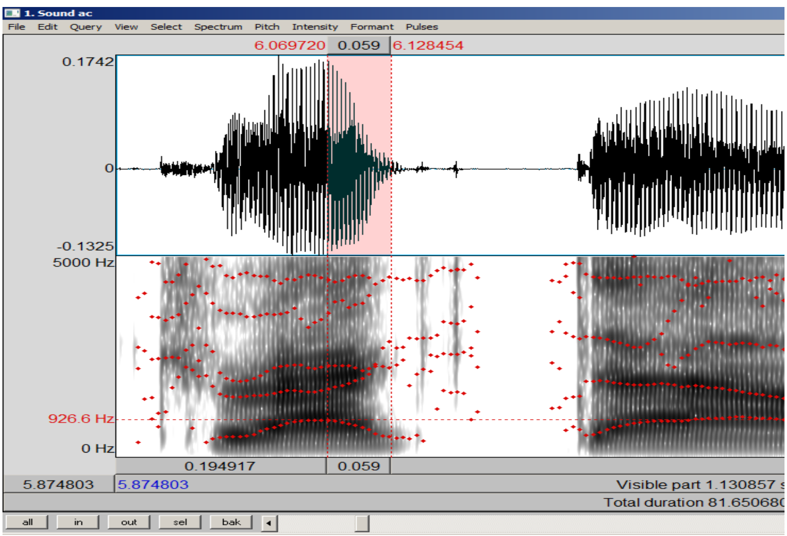

Figure 1: Example of VC transition measurement in *drag (down)*

3.2 Results

The first set of results is given in Figure 2, which shows the rate of stop release as a function of learner group. The first year group produced release bursts in 80% of the overall tokens, while the higher years' group produced unreleased stops in 51% of the tokens.

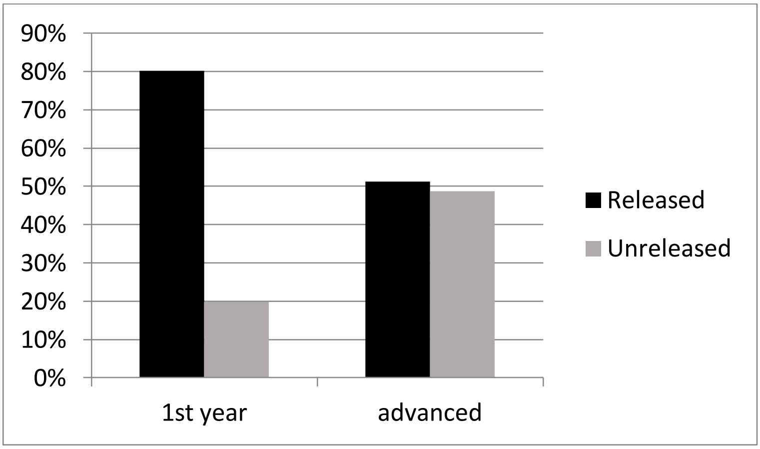

Figure 2: Rates of stop release (%) as function of learner group

To investigate how various independent variables affected the likelihood of stop release production, a binary logistic regression analysis was carried out, with release as the dependent variable. Learner group (advanced), homorganic place in C#C sequences, and %VC (higher), were all significant predictors of unreleased stops ($p<.001$ for each predictor). Higher VC transition duration also showed a tendency for unreleased stops, but this effect only approached significance ($p=.08$). The effects of vowel quality (tense vs. lax) and C1 voicing were not significant. The results of the regression analysis suggest that the suppression of stop release is indeed an important aspect of L2 phonological acquisition for the advanced group, and that there was indeed a link between the formant transitions on the preceding vowel and the probability of stop release.

Figure 3 shows the means of the VC transition parameters as a function of learner group. The Advanced group showed slightly higher measures for both parameters. A one way ANOVA was carried out on the mean values. In the case of %VC, the difference (29.6% vs. 28.3%) approached significance, $F(1, 502) = 2.96$; $p = .08$. For VC transition duration the difference (33.4 ms vs. 30.3 ms) was significant, $F(1, 502) = 7.87$; $p=.005$. These results suggest a modest effect of learner group on the acoustic robustness of VC transitions in C#C sequences.

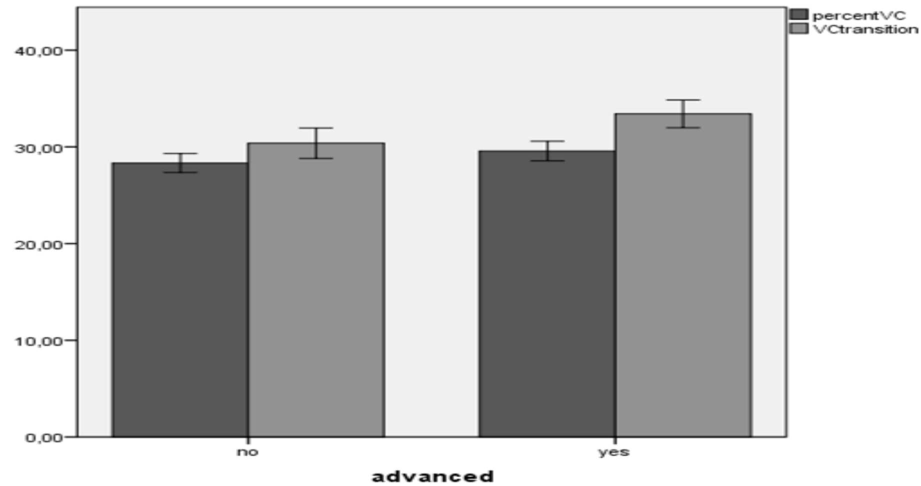

Figure 3: Mean VC transition measures as a function of learner group. Error bars show 95% confidence intervals

Figure 4 presents a summary of the results for the VC transitions parameters as a function of stop release across both groups. The unreleased stops had higher mean measures for both of the VC transition parameters (36.6 vs. 29.4 ms for VC duration; 32.3% vs. 27.2% for %VC). A one way ANOVA revealed that the differences in both parameters were significant ($F (1, 502) = 51.3$; $p < .001$ for %VC; $F (1, 502) = 42.6$; $p < .001$ for VC duration). These results suggest a robust link between the realization of the VC formant transitions and the probability of release.

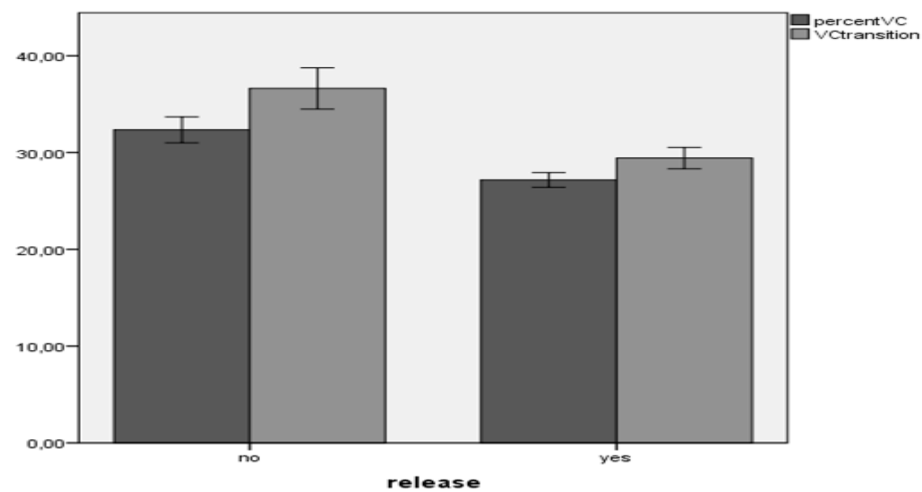

Figure 4: Mean VC transition measures as a function of stop release. Error bars show 95% confidence intervals.

In the case of the individual results, we may also observe a link between the robustness of VC formant transitions and the likelihood of unreleased stop production. This is shown in Figure 5, in which we see an inverse correlation between the mean %VC measures for each individual speaker and the number of stop release bursts (r=-.737, p =.003).

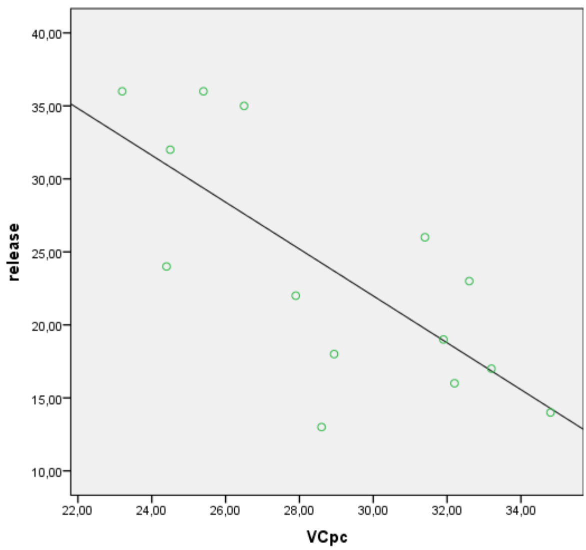

Figure 5: Individual results for stop release and %VC parameters

3.3 Discussion

The implications of our experimental results may be summed up as follows. In the speech of Polish learners of English, advanced learners show signs of acquiring target-like suppression of coda stop release in VC#C sequences. The production of stop releases is accompanied to some extent by changes in the quality of the preceding vowel as measured in the formant transitions in the approach to the stop closures. This link is seen in the significant differences in both vowel parameters as a function of stop release. That is, the unreleased tokens produced by the learners in both groups showed more robust VC transitions according to the measured parameters. The VC parameters as a function of learner group showed a more modest effect, but also indicated that acquisition of unreleased stops entails the acquisition of the formant patterns on the basis of which listeners may identify the stop consonant in coda position. The individual results also supported the tested hypothesis.

Taken together, these results suggest cross-language phonological interaction in the production of coda stop releases, providing general support for the notion that non-contrastive phonetic properties may play a significant role in L2 phonological acquisition. The results also suggest that the traditional phonological perspective, according to which a post-vocalic consonant in word-final position, regardless of language, occupies a ‘coda’ position, is in need of revision. In other words, it appears as though there are two types of ‘coda’ stops: those in languages like English and Korean with unreleased stops and greater phonetic cohesion with the preceding vowel, and the those in languages like Polish with obligatory release bursts and less robust VC transitions. Since this difference appears to be systematic, we feel it should be representable in an adequate phonological framework. In the next section, we shall offer a phonological description of this difference in the Onset Prominence framework (Schwartz 2013).

Despite the positive initial result found in our study, much more work remains to be done in two primary areas. First, additional acoustic descriptions should include the magnitude of formant excursion over the VC transition. This is a somewhat more complex methodological matter than simply measuring the duration of the transition. One possibility is to measure Euclidean distance in F1-F2 formant space, as is common practice in studies of diphthongs (e.g. Bogacka 2007) and diphthongization. We are planning a follow-up study in which this will be done while controlling for VC transition duration. The other area that needs to be pursued is the perceptual consequences of VC transitions in the two languages, as well as in the speech of L2 learners. For example, if the release burst is edited out of Polish-style tokens, will listeners still be able to identify the consonant? To what extent will this be based on English proficiency? These and other research questions will be posed in future work.

4. The representation of ‘codas’ in the Onset Prominence framework

In the Onset Prominence environment, segmental representations and syllabic structures are derived from the same representational hierarchy, which has its origins in the phonetic events associated with a stop-vowel sequence in initial position. An important aspect of the OP hierarchy is that syllabic structure is not built-up from vocalic nuclei. Rather, it is built down from onset consonants, the most ‘prominent’ of which are stops. Thus, the top layer of the hierarchy is labeled Closure, the defining property of stops (and also present in nasal consonants). This followed by Noise, which is associated with both stop release and fricatives. The next level down is Vocalic Onset (VO), representing both approximants and glides, as well as the initial position of vowels that house CV transitions facilitating the perception of obstruents. At the bottom of hierarchy are vowels.

The relation between individual segments and prosodic constituents is seen in Figure 6, which shows individual segmental representations in the English word *quick* on the left, and a ‘syllabified’ representation on the right. In the individual structures each segment type as housed at different levels of the hierarchy. The stops are at the top, below them is the glide, and below that the vowel. The fundamental mechanism for building syllabic structures from individual segments is referred to as absorption, by which lower-level vowel structures are merged with higher-level consonants to make a single tree. By this mechanism, the /kwɪ/ sequence in *quick* is joined into a single tree. By contrast, the final /k/ may not be absorbed into the preceding constituent. In Figure 6 it is placed underneath the vowel as a result of a different process referred to as *submersion* (Schwartz 2013).

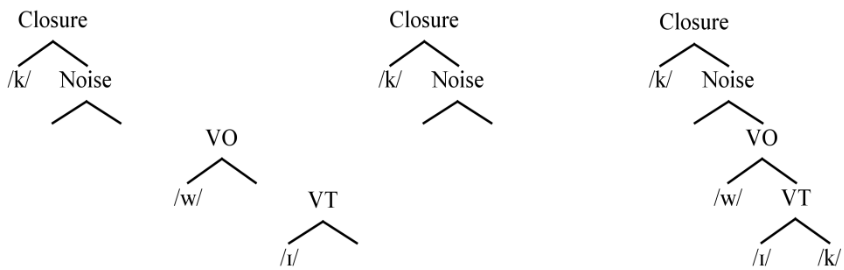

Figure 6: Segmental and constituent structures of English *quick*, with submerged final /k/

We suggest that the submersion process in *quick* is due to a parameter setting that enforces a more robust VC transition and allows for the possibility of unreleased stop. In Polish, submersion is hypothesized to be absent. The parametric difference is seen in Figure 7, which shows a representation of English *click* alongside the related Polish word *klik* (on the left). In *click*, the coda stop is at a lower level of the hierarchy and susceptible to lenition processes such as the suppression of stop release. The Polish coda /k/ remains at its underlying level, where release is obligatory.

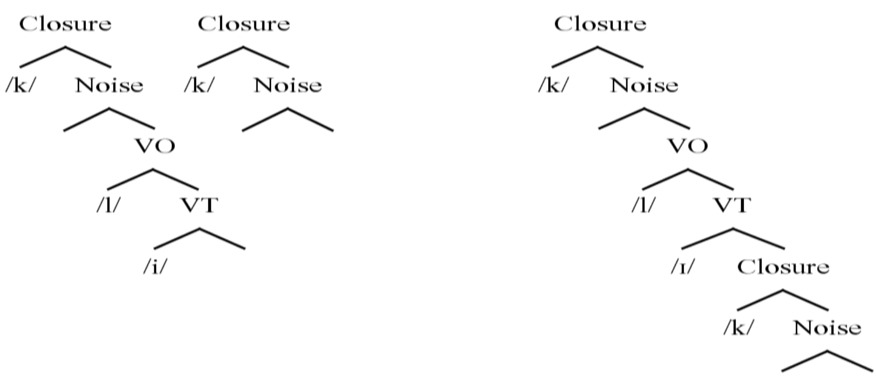

Figure 7: Polish *klik* (left) vs. English *click* (right) with submerged final /k/

While the representations in Figure 7 capture the cross-linguistic differences in coda stop behavior, the submersion parameter has prosodic implications that go beyond the mere question of stop release. Briefly stated, submersion is the same process that forms long vowels and vowel length contrasts, and creates VCV sequences in which the intervocalic consonant is subject to weakening processes. The vowel lengthening process is shown in Figure 8. Two adjacent vocalic structures, which may not undergo absorption since they are the same (VT) level of the OP hierarchy, may be joined by means of submersion producing a long vowel.

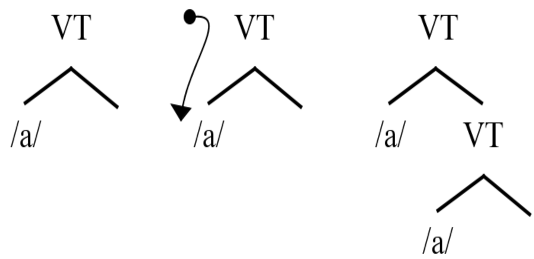

Figure 8: Submersion forming a long vowel

Figure 9 shows an additional case of submersion in the English word *pity*. In this case we see that not only single segmental structures, but also entire syllables may undergo submersion. Because of phonotactic constraints in English, the /t/ in pity is expected to syllabify as a coda, since lax vowels are always followed by a consonant in monosyllables. At the same time, universal onset maximization suggests that the /t/ should be an onset. This conflict gave rise to the controversial proposal of ambisyllabicity (Kahn 1976), by which the /t/ is shared by the two syllables in *pity*. The structure of *pity* in Figure 9 offers a natural expression of this ambiguity. The fact that the intervocalic /t/ is subject to weakening processes (flapping or glottaling) is a natural result of the consonant's lower position in the OP hierarchy.

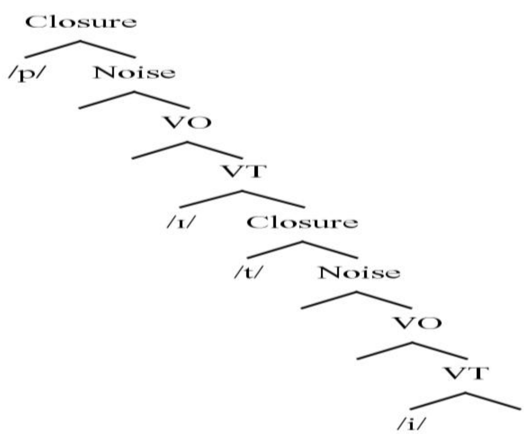

Figure 9: English *pity* with submerged final syllable

We have seen that submersion unites three seemingly unrelated features of English phonology: the presence unreleased stops, the presence of long vowels, and consonant weakening processes in VCV sequences. Polish by contrast lacks all of these phonological properties. Coda stops are always released (except in homorganic clusters), vowel length is absent, and there is little or no consonant weakening in VCV contexts (cf. Polish *PIT-y* 'tax returns'). In sum, the submersion parameter makes insightful predictions about a number of independent facts that may be observed in the phonologies of Polish and English.

5. Final remarks

This paper has presented phonetic data on stop release production and unreleased stops in the speech of Polish learners of English. We identified a link between the suppression of stop release and the acoustic robustness of VC transitions. Briefly stated, more robust transitions that allow listeners to identify stops in coda position, accompanied the acquisition of unreleased stops in coda position. The phonetic study has far reaching phonological implications that may be described in the Onset Prominence framework.

Acknowledgement

This research was supported by a grant from the Polish National Science Centre (Narodowe Centrum Nauki), project nr. UMO-2012/05/B/HS2/04036, "*Sandhi* in second language speech".

References

Arsenault, Paul. 2009. Coronal features and retroflexion in Dhivehi and other Indo-Aryan languages. *Toronto Working Papers in Linguistics* 30. 17–33.

Bergier, M. 2010. The occurrence of unreleased oral stops in English voice agreeing plosive clusters straddling word boundaries. Production experiments with Polish advanced learners of English. M. A. Thesis. University of Silesia.

Best, C. T. 1995. A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds), *Second language speech learning – the role of language experience in speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.

Boersma, P. & Weenink, D. 2011. Praat: doing phonetics by computer. [Computer program].Version 5.2.18.

Bogacka, A. 2007. Repopulating vowel space: English diphthong production by Polish learners of English. PhD dissertation. Adam Mickiwicz University in Poznań.

Cook, A. 2000. *American Accent Training*. Hauppauge, NY: Barron's.

Cruttenden, A. 2001. *Gimson's Pronunciation of English* (6th ed.). London: Arnold.

Dukiewicz, L. & I. Sawicka. 1995. *Gramatyka współczesnego języka polskiego – fonetyka i fonologia* [Grammar of modern Polish – phonetics and phonology]. Krakow: Wydawnictwo Instytutu Języka Polskiego PAN.

Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In W.Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Jenkins, J. 2000. The phonology of English as an international language. Oxford: Oxford University Press.

Kahn, D. 1976. *Syllable-based generalizations in English phonology.* PhD dissertation, Massachusetts Institute of Technology.

Kang, Y. 2003. Perceptual similarity in loanword adaptation: English post-vocalic word-final stops to Korean. *Phonology* 20 (2): 219-273.

Kang, Y. 2011. Loanword phonology. In van Oostendorp, Marc, Colin Ewen, Elizabeth Hume, and Keren Rice, eds., *Companion to Phonology*. Wiley-Blackwell, pp. 2258-2282.

Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (The Netherlands: Kluwer Academic), pp. 403-439.

Rojczyk, A., A. Porzuczek & M. Bergier. 2013. Immediate and distracted imitation in second language speech: unreleased plosives in English. *Research in Language* 11.1. 3-18. DOI: 10.2478/v10015-012-0007-7

Schwartz, G. 2013. A representational parameter for onsetless syllables. *Journal of Linguistics, 49 (3)*, 613-646. DOI: http://dx.doi.org/10.1017/S0022226712000436.

Steriade, D. 1993. Closure, release, and nasal contours. In M. K. Huffman and R. A. Krakow (Eds.). *Nasals, nasalization, and the velum* (pp. 401–470). San Diego: Academic Press.

Wright, R. 2004. Perceptual cue robustness and phonotactic constraints. In Hayes, B., R. Kirchner & D. Steriade (eds). *Phonetically Based Phonology*. Cambridge: Cambridge University Press, 34-57.

 DOI: 10.2478/rela-2014-0005

"I UNDERSTOOD YOU, BUT THERE WAS THIS PRONUNCIATION THING…": L2 PRONUNCIATION FEEDBACK IN ENGLISH/FRENCH TANDEM INTERACTIONS

CÉLINE HORGUES
University of Paris 3 – Sorbonne Nouvelle, France
celine.horgues@univ-paris3.fr

SYLWIA SCHEUER
University of Paris 3 – Sorbonne Nouvelle, France
sylwia_scheuer@yahoo.fr

Abstract
The role of corrective feedback (CF) in L2 development has been the topic of much discussion in SLA literature (see for example Sheen and Ellis 2011 for a recent overview). Researchers have focused their attention on CF provided either by language teachers or by fellow L2 learners, whereas relatively little is known about phonetic feedback offered in a non-institutional setting during peer-to-peer native/non-native interactions as is the case with tandem language learning. Tandem language exchanges represent a special learning environment, as each participant takes turns being the native and the non-native side of the dialogue. Thus, in contrast to the typical L2 learning setting, the hierarchical structure between the participants is fluid: the expert-novice power relationship evolves as the meeting progresses and the conversation switches from one language to the other.

In order to see how the distinguishing characteristics of tandem learning (such as solidarity and reciprocity) shape the process of L2 phonetic development in their own specific ways, we collected an English-French Tandem Corpus as part of the SITAF project (***S****pécificités des* ***I****nteractions verbales dans le cadre de* ***T****andems linguistiques* ***A****nglais-****F****rançais)*, launched at the University of Sorbonne Nouvelle-Paris 3 in October 2012. We gathered linguistic data – both video and audio recorded – from face-to-face conversational exchanges held by 21 pairs of undergraduate students, with each 'tandem' consisting of a native speaker of English and a native speaker of French. The dialogues and reading passages were recorded on two occasions separated by a 3-month interval.

The present paper offers a preliminary analysis of L2 pronunciation feedback on several renditions of the same text (*The North Wind and the Sun*), given to the French speakers by their English tandem partners. The passage was produced by each French participant three times: (1) during the 'monitored' reading, which was supervised by the English-speaking partner and which led to (2) the 'second reading' in the course of the first recording session, and then (3) the 'final reading' performed during the second recording session 3 months later.

Data analysis will allow us to address the following questions relating to the study of phonetic corrective feedback:

- What is corrected by the native-speaking partner (henceforth NS)? Segmental or prosodic errors? Phonemic or allophonic deviations?
- What is the corrective strategy adopted by the NS? Is it explicit correction, recast, or elicitation?
- What is the learner's uptake after receiving feedback?

We hope that our data brings a valuable and fairly unique contribution to SLA research, helping to establish which errors get corrected and how it may have implications for setting priorities in L2 pronunciation teaching.

Keywords: Tandem learning, L2 pronunciation, Corrective Feedback (CF), French learners of English

1. Introduction

This paper aims at presenting the preliminary results of work in progress on phonetic corrective feedback provided in the course of tandem interactions between native French and native English partners. This project was carried out in the framework of a junior researchers project entitled SITAF (***S**pécificités des **I**nteractions verbales dans le cadre de **T**andems linguistiques **A**nglais-**F**rançais),* which was launched at the University of Sorbonne Nouvelle-Paris 3 in October 2012. The SITAF research team is made up of ten members specialising in different but complementary research areas: phoneticians and phonologists, L1 and L2 acquisition specialists, didacticians and gesture specialists. Our Tandem Corpus aims to gather linguistic data, both verbal and non-verbal (with video recorded sequences), from conversational exchanges held by twenty-one French/English tandem pairs of undergraduate students at University Sorbonne-Nouvelle Paris 3. In the framework of a language tandem, the native speaker is the main provider of target-language input (*positive evidence*, see part 1) and occasionally is the provider of feedback on the partner's incorrect output (*negative evidence*). However he/she is not a professional teacher who has expert practice in teaching and correcting language features, which raises the following questions: Does the specific communicative setting of tandem interaction entail instances of corrective pronunciation feedback? If so, when, in what form and how often is phonetic feedback provided?

Corrective Feedback (CF) has been researched quite extensively when applied to learners receiving feedback from a language instructor in the formal environment of the classroom (Lyster and Ranta 1997, Mackey 1999, 2006, Gass 2003, Lyster *et al*. 2013). A lot of experimental research has looked into the strategies and effects of CF regarding grammatical development (*e.g.* Mackey 2006[1],), whereas the field of phonetic CF remains underrepresented (Lyster et al 2013[2]). What is more, very little is known about CF received in more informal learning contexts, as exemplified by the informal conversation with a native speaker of the target language in a language tandem

[1] In Mackey (2006), the target forms studied were questions, plurals and the past tense.

[2] Lyster et al. (2013: 22) *"Whether conducted in laboratories or classrooms, CF research has focused to a great extent on grammatical targets, reflecting the preoccupation with grammatical development in the study of SLA*". They only mention two of their studies (Sato & Lyster 2012), in relation to phonological CF (the acquisition of /ɹ/ and the development of fluency by Japanese learners of English).

exchange. This paper contributes to giving a new insight into the issue of pronunciation feedback, which, if evidenced at all during tandem interactions, might be different from that described for traditional language instruction.

The first part of the article will present the Tandem Corpus after having described how language tandem exchanges set a specific environment for L2 learning in general and L2 phonetic corrective feedback in particular. The key issues of which pronunciation features get corrected and how during the exchanges of the Tandem Corpus will be addressed in parts two and three, respectively. Finally, part four will raise the question of the effectiveness of pronunciation CF received during these same language exchanges.

2. Tandem learning: studying the implications for L2 pronunciation and collecting the Tandem Corpus

2.1 The acquisition of L2 pronunciation in tandem

O'Rourke (2005: 434) defines tandem learning as: "*an arrangement in which two native speakers of different languages communicate regularly with one another, each with the purpose of learning the other's language*". Tandem learning therefore represents an interesting and special form of language learning which often complements the more traditional instruction learners get through classroom teaching of the L2. The pedagogical benefits of tandem interactions for L2 learning have been pointed out in previous research.

First, contrary to the more traditional and hierarchical relation between a teacher and a learner the relation between tandem participants tends to be symmetric. Solidarity and role reversibility are at the basis of tandem learning. Indeed, the two participants will, in turn, construct two roles throughout the conversation exchange depending on which language is spoken: the role of the learner in the L2 and of the expert[3] in their L1. The native speaker is not expected to function as a teacher but rather, as an empathising peer taking part in maintaining a friendly, comfortable relationship which is not as face-threatening as, and more reassuring than, interacting with a teacher/assessor. This will be an essential factor in reducing the learners' inhibition about expressing themselves orally or to overcome their embarrassment at meeting pronunciation difficulties or having a foreign accent.

Tandem learning is also based on mutual assistance, learner commitment and learner motivation. Tandem participants generally sign up freely for such programs showing their genuine motivation to learn their partner's L1 but also to get to know "*their interlocutor as an individual*" (O'Rourke, 2005: 434).

[3] These roles are obviously idealised concepts rather than realities. The native speaker is considered by their tandem partner as a trustworthy representative of the target language and target language community and culture. This does not obviously mean that the native speakers effectively possess full mastery of their mother tongue or culture. One might doubt that such a skill is attainable anyway (see Kramsch 2003).

In addition to these positive socio-affective and psychological factors, during tandem interactions, learners are also exposed to very valuable L2 spoken input provided by the native speaker. Its quality lies in the fact that it is authentic and embodied input. Sufficient and quality exposure to L2 oral input is a well-known requirement of pronunciation learning. Through synchronous oral interaction, tandem partners get exposed to native input both in the form of what is called "positive evidence" (Long 1996, Gass 2003[4]) *i.e.* information on what is acceptable in the target language, but also "negative evidence" of the L2, *i.e.* information on what is unacceptable expressed through feedback on erronous learners' output. The latter type of evidence will be referred to in this article with the term "corrective feedback"[5] or CF. CF can be direct (explicit) or indirect (implicit).

In addition to this verbal evidence about the target language, through face-to-face tandem, participants also have access to useful non-verbal cues. Indeed, body gestures and facial movements enable partners to better interpret their interlocutor's message, to identify or express instances of communication breakdowns, to make on-line interactional adjustments, etc. As far as pronunciation is concerned, the fact that the participants are literally positioned face-to-face allows for gestural or facial elicitation of what the target pronunciation should sound and look like (simultaneously). Indeed, participants have direct visual access to some of their native interlocutor's articulatory gestures (lips, jaws, possibly the tongue), which can be very valuable visual support for French learners to grasp the two renditions of the interdental fricatives <th> in English, for instance (see examples in parts 2).

Despite all the pedagogical benefits for the acquisition of L2 pronunciation listed above, language tandem also has some drawbacks which researchers have underlined (see Brammerts and Calvert 2002). The main limitation is that, in face-to-face oral tandem, the spoken input is ephemeral and is therefore highly demanding on the learners' attention and memory skills. The positive socio-affective factors mentioned above might in some instances also act negatively on L2 pronunciation learning. Many tandem participants will naturally focus on content and smooth communication, task completion rather than form accuracy. Some native speakers will tend to develop a tolerance to errors or erroneous pronunciation in the speech of their tandem partner (Brammerts and Calvert 2002, 2003) and will minimise expression and comprehension problems, simply because it might feel socially awkward to point to errors in your peer's output.

By and large, however, we believe the benefits of tandem learning largely outweigh these limitations.

[4] Gass (2003: 225) defines positive evidence as comprising "*the set of well-formed sentences to which learners are exposed*" and negative evidence as "*the type if information that is provided to learners concerning the incorrectness of an utterance*".

[5] See Lyster and Ranta (1997) and El Tatawy (2002) for a discussion on the terminology and its interchangeability (corrective evidence, negative evidence, negative feedback, repair, focus-on-form).

2.2 The English/French Tandem Corpus

Twenty-one pairs of participants were recruited through an online questionnaire available on our university website. It aimed at providing information about the participants' language profile (self-assessed proficiency level, language background) and their general interests to ensure tailored pairing up of the participants. All participants volunteered freely for the tandem program and later organised their meetings autonomously. They met between 2 and 23 times with a mean frequency of 12 meetings over a three-month period (February-May 2013). None of the speakers was bilingual in the other language[6].

The 21 native French speakers (labelled *F01>F21*) were all undergraduate students[7] in English studies for the most part. Their proficiency level varied from upper-intermediate to advanced. The 21 native English speakers (labelled *A01>A21*) were exchange students at our university and represented a range of dialectal variety (American, British, Irish, Australian). Their proficiency level in the target language was certainly more varied than that that of our French speakers because it was highly dependent on the length of their stay in France at the start of the experiment and on the age when they started learning L2 French (which was less homogeneous than that of the French learners of English).

We recorded the tandem pairs at two points in time: the first session in January/February 2013 was organised about a week after the participants' first face-to-face encounter at the introductory meeting, and the second recording session was scheduled three months later in April/May 2013.

The technical set-up used for the two recording sessions was the university recording studio where tandem partners were seated face to face. We used 3 cameras (one in the direction of each participant, and one capturing the general interactional frame), and 2 microphones positioned 10 cm above the speakers' heads. The individual footage of the two speakers was then edited to appear both in the same video frame.

More details about the speakers' profiles, recruitment method and experimental design can be found in Horgues and Scheuer (2014, forthcoming).

To prompt interaction between tandem participants, the speakers were first recorded performing two semi-spontaneous speech tasks: two games eliciting argumentation and story telling. The uses of French and English were clearly separated with an instructed switch after 30 mins. The games were then followed by a reading task whose objective was to collect controlled, hence directly comparable speech data for all speakers. The reading passage, *The North Wind and the Sun* (see appendix 1), and its French version, were selected. The choice of this particular reading passage was motivated by the reference to the numerous phonetic studies having previously used it to explore phonetic variation in English (native and non-native varieties), and also because its French version is used for similar purposes in French phonetics as well. However, we are well-aware that this passage was initially designed to elicit phonemic variation and that in that respect, it is not the best suited for the analysis of some suprasegmental features like

[6] Although some speakers were bilingual in another language (*e.g.* Guadeloupean French, Algerian Arabic, Costa Rican Spanish).

[7] From the first to the third year of the degree.

intonational focus, the prosodic marking of information structure, etc. This will have to be taken into account when considering the speech features studied (part 2).

Only one section of the corpus speech data will be used at this stage in our analysis: the various renditions of the reading passage in L2 English. Upon monitoring the two recording sessions, we informally observed that the reading task (where the learners are more naturally more focused on form than communication) entailed a higher frequency of CF than spontaneous game-like activities. Therefore the results presented in this study are limited to this particular speaking style and would have to be compared with more spontaneous speech in further analyses.

Below are the details of how the reading task in L2 English was performed during the two recording sessions (see fig. 1)

- in the first session, the French speakers performed what is called the first or "*monitored*" reading which encouraged interaction and feedback from the native interlocutor, and for which the instructions explicitly said:
  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*
- this reading was then immediately followed by a second and (hopefully improved) reading, "*second reading*" during which the tandem partner was no longer supposed to intervene:
- *and then a second time on your own (no interruption)*
- in the second session, 3 months later, the same speakers were simply asked to read the same passage (with no specific instruction as to the monitoring). This corresponds to the "*final reading*".[8]

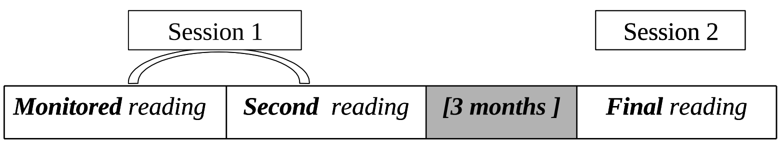

Figure 1: The L2 reading task in Tandem Corpus (The North Wind and the Sun)

Let's now turn our attention to the focus of CF in this reading task, *i.e.* the types of errors the native English partners tended to correct.

3. What was corrected?

The basic statistics on the corrective feedback provided by the English participants about their French partners' performance on the reading task are presented below (see also fig. 2):

[8] After the learners performed their final readings in the L2, all speakers also read the text in their mother tongue. This L1 control data was not used for the specific research question studied in this article, but it will be analysed in further studies.

- we observed 108 instances (tokens) of CF across the three readings in the two recording sessions;
- 103 of them (*i.e.* 95.4%) regarded segmental errors;
- of those:
    - 58.3% related to vocalic errors (*e.g.* 'wind' /ˈwɪnd/ incorrectly rendered as */ˈwaɪnd/, 'sun' pronounced *[ˈsɜn])
    - 25.2% concerned consonantal errors (*e.g.* 'obliged' commonly pronounced with a medial [ʒ] or [g]);
    - 16.5% involved 'mixed category' errors, *i.e.* when the NS's intervention targeted both a vocalic and a consonantal realisation at the same time (*e.g.* 'cl<u>os</u>ely' rendered as *[ˈklɒzli]);
- the remaining 5 instances of CF (*i.e.* 4.6%) regarded suprasegmental matters, which in our case were limited to lexical stress (*e.g.* 'considered' pronounced 'CONsidered').

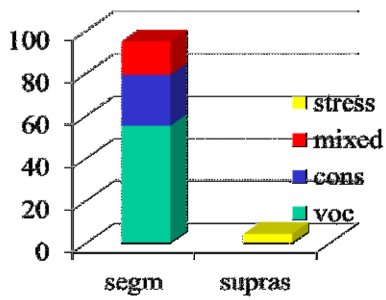

Figure 2: Instances of corrective feedback

A couple of acknowledgements are in order here. Firstly, even the few examples of CF given above reflect considerable variation among the tokens in terms of the nature of the error being corrected, which might call for a number of separate treatments in the future. For example, certain erroneous pronunciations arose (presumably) from incorrect phonemic representations of particular lexical items, whereas some others pointed to potentially more important, global problems, *i.e.* imperfect realisations of particular English phones across the board. 'Wind' mispronounced */ˈwaɪnd/ illustrates the former, and /ʌ/ rendered as [ɜ] (as in *[ˈsɜn] for /ˈsʌn/) the latter category. Secondly, the fact that nearly 60% of all instances of CF pertained to vowels, and just over 25% to consonants, does not in itself prove that vocalic errors are more serious than consonantal ones: it could simply mean that vowels are targeted more often because they are mispronounced more often than consonants. Either way, our results tend to highlight the pre-eminent position of vowels in the course of L2 pronunciation teaching and learning.

Another key question to be addressed in the context of error typology regards the reasons why certain phonetic deviations (vocalic and consonantal alike) attract native speakers' attention whereas some others are ignored, as far as CF goes. Although the motivations underlying the individual decisions to correct or not to correct are ultimately complex and subjective, it stands to reason that they are grounded in certain overarching considerations such as intelligibility or impression of a strong foreign accent.

While it can be argued that any mispronunciation has the capacity for miscommunication, mistakes that result in actual or potential lexical confusion – or simply compromise intelligibility – will naturally rank higher in this hierarchy. Therefore, errors that involved (near) minimal pairs were naturally expected to be CF magnets in our study. This can be illustrated with the following examples: [ˈzen] for ‘then’, ‘cloak’ pronounced like ‘clock’, or the almost archetypal ‘hungry’ for ‘angry’ substitution (which indeed led to genuine communication breakdown during *A09-F09* conversational exchange), as well as the rendition of ‘blew’ as [ˈbliː] by speaker *F04*, which was simply not understood at all by the native listener (see part 4). This tendency – *i.e.* to correct errors that result in one lexical item morphing into another – does not seem to merit an in-depth analysis here: after all, the primary purpose of language is communication, and if communication is in danger of being inadvertently hindered, the native speaker is justified in feeling compelled to intervene.

In view of the above considerations, what is perhaps more interesting is precisely instances of corrective feedback when intelligibility was *not* at stake. Some native speakers indeed showed a stronger, or at least longer, reaction to pronunciation errors that did not – by their own admission – impede communication than to those that did have the potential for lexical confusion. To paraphrase the comment which features in the title of our paper (made by speaker *A02*), there were certain “pronunciation things” that the native English participants were not ready to ignore even though they were able to understand everything that was being said. The erroneous rendition of the <th> could serve as a prime example in this context. This virtually proverbial and remarkably widespread error in L2 English speech has received a lot of attention from SLA researchers and EFL practitioners alike, for a whole array of reasons. To name but a few, Brennan and Brennan’s (1981) classic study of foreign accent in the English of Mexican immigrants to the US showed no relationship between the frequency of this phonetic deviation and accentedness ratings. Similar results were reported in Scheuer (2002) for Polish learners of English. In another influential volume, Jenkins (2000) states in no uncertain terms that substitutions of other consonants for the English dentals are inconsequential to international intelligibility, which means that these consonants should not be prioritised in teaching English as a Lingua Franca. These findings would seem to point to the conclusion that ‘th’ mispronunciations, frequent as they may be in L2 speech, do not necessarily deserve the high standing that is sometimes accorded to them in EFL pronunciation instruction.

Yet, in spite of its apparent status as being inconsequential to communication and relatively indifferent to the strength of perceived foreign accent, this type of error was singled out for correction – and sometimes even mini-speeches – by several of our native English participants. For example, speaker *A13* commented on his French partner’s renditions of <th> as [s] in the following way: “‘North’, with a ‘th’ at the end. That’s probably a tricky one, but, really, get the /θ/: ‘north’ /…/ Again, I completely understood you, but /…/”. In much the same vein, although as if speaking on behalf of native English speakers in general, participant *A15* reassured his partner as follows: “The only suggestion that I could make for you was the /θ/ sound /…/ I mean I… we could completely unders I’m sure… I could completely understand you, and everyone else could, but… erm… instead of [zi] it’s /ˈðiː/”.

It is perhaps worth noting that the authors of the "I understood you but" comments appeared at a loss to justify their preoccupation with this pronunciation problem. Naturally, no such justification was demanded of them. As usual, a variety of motives may have driven their decision to focus on the 'non-th's. Errors involving interdentals are relatively easy to spot and point out, as their place of articulation lends itself to simple description and demonstration. Consequently, the correctors may have felt fairly confident about offering advice on how to rectify this particular kind of mistake. On the other hand, this insistence on the accurate pronunciation of the dental fricative could be interpreted as the native speaker's attempt to correct an annoying – rather than communicatively confusing – error. This ties in with Markham's (1997: 101) observation that *"[c]uriously, the more negatively judged errors are ones which do not cause lexical confusion /.../ – they are simply non-native pronunciations –,* [sic] *whereas the more acceptable errors can cause lexical confusion."* Further support for the notion that dental fricatives may belong in the 'annoying' category comes from the questionnaires that our participants were asked to fill in on completion of the recording sessions. These provide invaluable insights into young native speakers' beliefs about and perceptions of a French accent in English, as well as into those aspects thereof which they find irritating and/or detrimental to intelligibility. 8 out of the 21 subjects explicitly mentioned 'th's in this context, and half of them (4) went as far as branding this type of mistake as annoying without necessarily hindering comprehension.

4. How was it corrected?

In spite of the fact that tandem exchanges represent a unique learning environment, where the power structure is symmetric and fluid, within each given task one participant was clearly the novice and the other one the expert, who was therefore – implicitly or explicitly – expected to provide assistance and guidance without souring the friendly atmosphere. In a recent state-of-the-art article, Sheen and Ellis (2011: 606) conclude that *"[l]earners almost invariably express a wish to be corrected"*, although CF is a highly complex issue where no overall ideal strategy might necessarily be identified.

Corrective feedback provided by our native English participants indeed took different forms, as a function of the gravity of the error and – presumably – the personal preferences of the corrector, or the rapport between the tandem partners. In the present analysis we will be distinguishing just three categories of CF: explicit comments or explanations, recasts, and clarification requests (see *e.g.* El Tatawy 2002). Needless to say, certain instances of feedback represented complex cases where more than one corrective strategy was used at a time, for example when a recast was immediately followed by an explicit comment. In such cases we aimed to identify the dominant strategy and we labelled the CF token accordingly. The overall statistics, graphically presented in fig. 3, look as follows:

- Recasts accounted for 61.1% (60 out of 108) of all instances of CF across the three readings in the two recording sessions. Lyster and Ranta (1997: 46) define this – generally implicit – corrective strategy as one involving *"the teacher's reformulation of all or part of a student's utterance, minus the error"*. A classic example from our corpus is provided by the following exchange between speaker

A17 and speaker *F17*, who previously pronounced the word 'obliged' with an [i] vowel. The NS supplies the correct pronunciation of the word, without explicitly stating that the NNS's rendition was erroneous. The strategy seems to work, at least for the time being:
A17: Obliged.
F17: Obliged.
A17: Obliged. Yeah.
F17: OK.

- Explicit comments represented 25.9%, *i.e.* 28 tokens. These included short statements such as "wind, not [ˈwaɪnd]" (speaker *A11*), but also slightly longer descriptive sequences such as the ones regarding <th> (speakers *A13* and *A15*) mentioned in section 2.
- The remaining 13% (14 cases) were requests for clarification or repetition. To quote a somewhat extreme example, speaker *A16* reacted to her partner's rendition of one of the sentences (which was erroneous in more ways than one) by exclaiming *"Whoa, read that again!"*

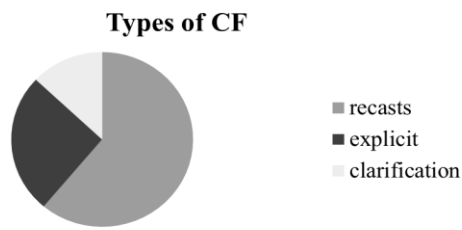

Figure 3: Instances of CF: corrective strategies

The fact that recasts proved to be the form favoured by our participants comes as no surprise. 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 dominant position. However, this indirectness comes at a price: corrections in the form of recasts are not always easily made sense of by the recipient, especially if more than one item is being corrected at a time. This is evident first and foremost in cases of automatic phonetic processes, such as the intervocalic intrusive (sandhi) [h], which is a common feature of French-accented L2 English speech. The following extract from the *A18*-*F18* interaction provides a good illustration of this problem (the sections in bold indicate the erroneous pronunciations picked up on by the native speaker):

F18: Then the North Wind blew **h**as hard as he could but the more he blew the more clo**[z]**ely did the traveller fold his cloak around him.
A18: /…/ BLEW AS HARD
F18: [silence]
A18: Blew as hard as he could.
F18: OK, blew **h**as hard as he could.
A18: <u>AS</u> hard.
F18: [laughing] As hard as he could.

A18: And closely.

F18: Closely. Je recommence: The North Wind blew **h**as hard as he could but the more he blew the more clo**s**ely did the traveller fold his cloak around him.

Evidently, it took as many as three recasts on the part the *A18* speaker (which, however, was by no means a record number) to finally bring about an h-less rendition of *as* in her NNS partner, which, alas, proved to be a rather short-term improvement.

Having explored the scope and modalities of CF during the tandem interactions, we are now turning our attention to the key issue of its effectiveness.

5. The effect of pronunciation CF during tandem exchanges

In this paper, we are very cautious when using the term "learner uptake" which often appears in the literature on CF. Indeed, learner uptake[9] typically refers to the learner's immediate response after receiving feedback from a professional provider of CF (*i.e.* typically a language teacher), and generally in a formal setting (*i.e.* the classroom), which is clearly not the case in the tandem setting.

At this stage in the analysis, we have looked at the French learners' uptake following feedback their English partner provided during the *monitored reading* (session 1). The idea is to see whether, and to what extent, erroneous pronunciation pointed out by the monitoring partner is subsequently modified positively (*i.e.* repaired) by the learner. French learners' uptake, or response to feedback, was studied at three points in time.

a) immediately upon receiving feedback (session 1)

b) during the second reading directly afterwards (still session 1)

c) during the final reading (session 2).

While carrying out the auditory analysis of the modifications the learners applied to items they had received feedback for, we realised that we needed to account for two types of learner's repair:

- total repairs: when an error pointed out by the tandem partner was fully repaired at one point in time. For example, *F04* had initially mispronounced *wrapped* as *[ˈrept], but then correctly modified it to [ˈræpt] immediately implementing her partner's (*A04*) CF.
- partial repairs: to refer both to cases where i) one item occurs several times in the reading passage (eg. *wind*, *cloak)* and is not repaired systematically, or where ii) the erronous pronunciation pointed out by the English-speaking partner is not totally attended to or repaired. An example of this is the term "closely" for which the French learner (*F12*) seemed to be unable[10] to attend to the correction of both

[9] For Lyster and Ranta (1997: 49): "*Uptake, in our model refers to a student's utterance that immediately follows the teacher's feedback and that constitutes a reaction in some way to the teacher's intention to draw attention to some aspect of the student's initial utterance (this overall intention is clear although the teacher's specific linguistic focus may not be*)*"*. Mackey (2006: 407) reports a slighlty different definition by Ellis *et al.* (2001): "*in which the learner utterance was optional and could occur not only after feedback, but also after any interlocutor utterance that provided information about a linguistic feature*".

[10] Immediate uptake, first session.

the unvoiced fricative /s/ (which she incorrectly realised as voiced[11]) *and* the diphthong at the same time. Her various attempts at repetition show that she can only repair one of these features at a time. This appears to be an illustration of the cognitive overload the learners are faced with. It also raises the question of their (in)ability to grasp the focus and the scope of their correcting partner's feedback.

The percentages[12] of repair at these 3 points in time (horizontal axis: *immediately, 2nd reading, final reading*) are presented in the 3 bars of the graph below (fig. 4). Total repairs appear in solid black, and partial in checked grey.

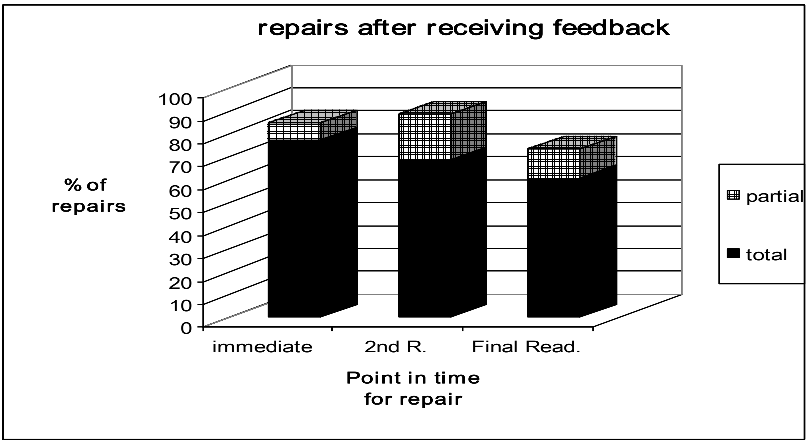

Figure 4: French learners' repairs after receiving pronunciation CF from their English-speaking tandem partners

	moment of repair		
type of repair	*immediate*	*2nd Read.*	*final Read.*
partial	7.6	20.2	12.6
total	77.2	68.4	60.7
both types	84.8	88.6	73.3

Table 1: Percentages of repair at 3 points in time (21 French speakers of L2 English)

A few general trends emerge and contribute to giving an insight into the uptake perspective (fig. 4 and table 1 above):

[11] Certainly under the phonotactic influence of L1 French where, in an intervocalic environment, <s> is systematically voiced /z/, and where the unvoiced form /s/ has to be indicated by a variation in spelling <ss>.

[12] Calculated in relation to the number of errors pointed out by the English-speaking partner during the *monitored reading* (*i.e.* potential repairs): overall a total 79 instances of CF provided by the 21 native-English speakers.

a) Globally, over two thirds of the errors pointed out by the tandem partner during the *monitored reading* then get repaired by the French learners during subsequent readings (either in same session or in the second session three months later). From an SLA perspective, this is a fairly encouraging result.

b) There is an overall degradation of the rate of all types of repairs over time between the two sessions (84.8% vs 73.3%)[13] but not within one recording session (84.8% vs 88.6%). This quantitative degradation of repair rate over the three-month period does not come as a surprise given the detrimental influence of decreasing memory trace over time.

c) More interestingly, there appears to be a qualitative degradation when looking at learners' delayed responses. Indeed, the number of *partial* repairs increased in the 2nd reading of the first session (20.2%) when compared to their rate immediately upon receiving feedback (7.6%). Accordingly, the number of *total* repairs dropped significantly between the two readings[14]. Correct instantaneous modification does not always lead to effective, complete repair even five minutes later. This observation underlines the loss of accuracy and the lack of permanency in the learners' uptake/ response to CF, which are certainly linked to the cognitive limitations of memory span and overload restraining the learners' performance.

These observations are in line with the results of previous research showing that *"the effects of corrective feedback are almost always gradual and cumulative rather than instantaneous and categorical*" (Doughty and Williams, 1998: 40, in El Tatawy, 2002: 15).

Let us emphasize that it is difficult to interpret the repairs occurring in the *final reading* (2nd session) as cases of actual *uptake* resulting from CF received in the first session three months earlier. Indeed, it was technically impossible to control the input the learners were exposed to in between the two sessions. Not only did the French learners attend English language classes but they also had access to other sources of English input to various degrees (TV films, personal interactions with their tandem partners, with other anglophones in Paris or during travels). For this reason, the positive (or negative) modifications occurring in the final reading cannot be interpreted as direct effects of CF received in the first session. Many factors other than CF might have come into play so the cause-effect relation cannot be established as it could be for the two readings of the first session.

Beyond these very general trends, it is interesting to start looking at individual learner differences when analysing repair in relation to received CF.

a) A few cases illustrate what could be called *consistent repair*, *i.e.* when an error is immediately repaired and then is systematically attended to in subsequent renditions over the two sessions. This was the case for *"wind"* initially mispronounced as *[ˈwaɪnd] but then corrected systematically by four French learners (*F01, F05, F15, F21*).

b) Other cases exemplify *non-permanent repair*: when an error was corrected in the first session (both immediately and in the second reading), but where the learner reverts back to the initial error 3 months later (2nd session). This time, *F04*

[13] The trend, however, fails to reach statistical significance (paired *t* test, $p>.05$).

[14] Paired *t* test, significant at $p<.05$.

illustrates a situation which, from a teacher's perspective, might seem rather frustrating. Indeed, this learner initially mispronounced the verb "blew" as *[ˈbliː], instantaneously received CF from her English-speaking partner[15], and fixed it to [ˈbluː] both immediately and then again in the second rendition of the reading passage. Her body language seemed to reveal that she seemed surprised at her own mistake, and seemed to fully take in her partner's CF about the correct pronunciation. However, in the final reading she fell back to her initial mispronunciation [ˈbliː], which her partner did not fail to pick up on again in her after-task remarks.

c) And conversely, cases of *late repair*: when the repair is delayed until the final reading (2nd session), and is either absent or incomplete (partial) in the first session. This situation is rather rare compared to the first two types of repair. *F02* provided an example with her cluster simplification of *–ed* at the end of *obliged* [əˈblaɪʤd > *əˈblaɪʒd]. This simplification led her partner (*A02*) to provide CF during the monitored reading[16]. *F02* seems to have disregarded this CF in the first session but only implemented it in the final reading. As previously mentioned, since other input influences cannot be factored out, it is not possible to consider this modification towards the target form to be the sole and direct consequence of CF received in the first session.

Future research will aim at proposing more fine-grained analyses of error treatment sequences both from the correcting partner's perspective (solicited vs spontaneous CF, simple/complex recasts and other combined strategies) and the learner's perspective (subcategories of "needs repair" uptake following Lyster and Ranta 1997), and will include some attention to the contribution of non-verbal features on both parts.

6. Conclusion

By definition, language tandem partners do not provide as expert, accurate and systematic CF as professional language teachers are supposed to. Quite importantly, they are not expected to, either. However, the positive socio-affective and psychological assets attached to the tandem setting might compensate this failing by providing a non-threatening, comfortable and collaborative environment for learners. Reduced-stress levels are facilitative of L2 pronunciation learning. Although sometimes imperfect and insufficient, the CF provided by the language partners plays a part in raising the learners' awareness about the difference between their output and target form. Significantly, attention to form and noticing the gap have been described as essential steps on the way towards L2 development (Long 1996, Lyster and Ranta 1997, Schmidt 1990, Mackey 2006[17]). To improve the pedagogical benefits of language tandem programmes

[15] Whose comprehension was genuinely impaired by the learner's erroneous pronunciation, as shown by her clarification request inviting *F04* to rephrase: *"the more he....?"*.

[16] In the form of a recast "obliged" with hyperaticulation of the ending.

[17] Mackey (2006: 408): "*Attention and awareness in particular have been identified as two cognitive processes that mediate input and L2 development through interaction (e.g. Gass and Varonis 1994; Robinson 1995, 2001, 2003; Long 1996; Gass 1997, Mackey et al. 2000; Philip 2003). Long (1996), for example, claims that selective attention (along with the learner's*

developed in various educational institutions, we would therefore advocate providing the tandem partners with some awareness-raising training on the strategies necessary to provide and receive CF effectively in the course of autonomous tandem interactions.

Acknowledgments

The SITAF project was financed thanks to a research grant obtained from the Conseil Scientifique de l'Université Sorbonne-Nouvelle Paris 3 (Projet Innovant, Jeunes Chercheurs, 2012-2014). Part of the orthographic transcription will be financially sponsored by two research laboratories (LPP, Univ. Paris3 and UMR 7110, Univ. Paris7).

We are very grateful to the SITAF team members for their invaluable help when designing the protocol, recruiting participants, monitoring the recording sessions and transcribing the data.

Many thanks to the video and sound engineers at the University for their technical support (among others Charly Comino and Hervé Profit).

References

Brammerts, H. and M. Calvert. 2002. Apprendre en communiquant. In B. Helming (ed.) L'Apprentissage Autonome des Langues en Tandem. Paris: Didier: 31-40.

Brammerts, H. and M. Calvert. 2003. Learning by communicating in tandem. In T. Lewis and L. Walker (eds) Autonomous Language Learning in Tandem. Sheffield: Academy Electronic Publications: 45-60.

Brennan, E. M. and J. S. Brennan. 1981. Measurements of accent and attitude toward Mexican-American speech. Journal of Psycholinguistic Research 10: 487-501.

Doughty, C. and J. Williams. 1998. Focus on Form in Classroom Second Language Acquisition. New York: Cambridge University Press.

Ellis, R. 1997. Second Language Acquisition. Oxford: Oxford University Press.

Ellis, R., H. Basturkmen and S. Loewen. 2001. Learner uptake in communicative ESL lessons. Language Learning 51: 281-318.

El Tatawy, M. 2002. Corrective feedback in second language acquisition. Working Papers in TESOL & Applied Linguistics. Teachers' College, Columbia University: 1-19.

Gass, S. M. 2003. Input and interaction. In C. J. Doughty and M. H. Long (eds) The Handbook of Second Language Acquisition. Oxford: Blackwell 224-255.Helming, B. 2002. L'Apprentissage Autonome des Langues en Tandem. Paris: Didier.

Horgues C. and S. Scheuer (forthcoming). 2014. Why some things are better done in tandem?

---

developing L2 processing capacity) mediates the L2 acquisition process. Negotiated interaction is claimed to be particularly useful in this regard, as the interactional feedback can help direct the learner' own interlanguage form (i.e. 'noticing the gap', Schmidt and Frota (1986), while at the same time providing the learners with opportunities to produce modified output. (Swain 1995, 1998, 2005). [...] Schmidt (1995, 2001) and Robinson (1995, 2001, 2003) argue that learners must conciously notice input in order for it to become intake. This claim is generally referred to as the Noticing Hypothesis (Schmidt 1990, 1993, 1995)."

Jenkins, J. 2000. The Phonology of English as an International Language. Oxford: Oxford University Press.

Kramsch, C. 2003. The privilege of the non-native speaker. In C. Blyth (ed.) The Sociolinguistics of Foreign-Language Classrooms: Contributions of the Native, the Near-Native, and the Non-Native Speaker. Issues in Language Program Direction, A Series if Annual Volumes. Boston: Thomson, Heinle: 251-262.

Long, M. 1996. The role of the linguistic environment in second language acquisition. In W. Ritchie and T. Bhatia. Handbook of Second Language Acquisition. San Diego: Academic Press: 413–468.

Lyster, R. and L. Ranta. 1997. Corrective feedback and learner uptake. Studies in Second Language Acquisition 19(1): 37-66.

Lyster, R., K. Saito and M. Sato. 2013. Oral corrective feedback in second language classrooms. Language Teaching 46: 1-40. DOI: 10.1017/S0261444812000365

Mackey, A. 1999. Input, interaction, and second language development. Studies in Second Language Acquisition 21: 557-587.

Mackey, A. 2006. Feedback, noticing and instructed second language learning. Applied Linguistics 27/3: 405-430. DOI: 10.1093/applin/ami051

Markham, D. 1997. Phonetic Imitation, Accent, and the Learner. Lund: Lund University Press.

O'Rourke, B. 2005. Form-focused interaction in online tandem learning. CALICO Journal, vol. 22 n 3: 433-466.

Sato M. and R. Lyster. 2012. Peer interaction and corrective feedback for accuracy and fluency development. Studies in Second Language Acquisition 54: 591-686. DOI: 10.1017/S0272263112000356

Scheuer, S. 2002. What makes foreign accent sound foreign? In J. Leather and A. James (eds) New Sounds 2000. Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech. Klagenfurt: University of Klagenfurt: 306-314.

Schmidt, R. W. 1990. The role of consciousness in second language learning. Applied Linguistics 11: 128-158.

Sheen, Y. and R. Ellis. 2011. Corrective feedback in language teaching. In E. Hinkel (ed.) The Handbook of Research in Second Language Teaching and Learning Vol. 2. New York: Routledge: 593-610.

The Tandem Server (Bochum): http://www.slf.ruhr-uni-bochum.de/.

Appendix

INSTRUCTIONS FOR THE READING PASSAGE

Note that the layout of the reading passage given to the participants was different from the one presented here. The text appeared horizontally and ensured no sentence was visually interrupted by line breaks (impact on prosodic phrasing).

....……………

(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 dîtes 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'avançait, 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)

UNSTRESSED VOWELS IN GERMAN LEARNER ENGLISH: AN INSTRUMENTAL STUDY

LUKAS SÖNNING
University of Bamberg
lukas.soenning@uni-bamberg.de

Abstract
This study investigates the production of vowels in unstressed syllables by advanced German learners of English in comparison with native speakers of Standard Southern British English. Two acoustic properties were measured: duration and formant structure. The results indicate that duration of unstressed vowels is similar in the two groups, though there is some variation depending on the phonetic context. In terms of formant structure, learners produce slightly higher F_1 and considerably lower F_2, the difference in F_2 being statistically significant for each learner. Formant values varied as a function of context and orthographic representation of the vowel.

Keywords: vowel reduction, acoustic analysis, SLA, L1 German, L2 English

1. Introduction

In English, vowels occurring in unstressed syllables are reduced – they are articulated with a more central position of the tongue, a narrower jaw-opening and a loss of lip rounding (Delattre 1982). Acoustically, this is reflected in their duration and formant structure. One characteristic feature of German Learner English (GLE) is a lack of vowel reduction in unstressed syllables (e.g. Parkes 2001). This paper gives an exploratory account of the production of vowels in unstressed syllables by advanced German Learners of English in comparison with native speakers of Standard Southern British English. Two acoustic properties are investigated: duration and formant structure.

The paper will start out with an overview of instrumental approaches to speech rhythm, a prosodic construct that is closely related to the properties of unstressed vowels and has received considerable attention by researchers in the past two decades. Though located at the same end of the rhythmic continuum, English and German differ with respect to the properties of unstressed vowels. A brief account of relevant contrasts will be given. With transfer from L1 being an important factor in L2 phonological acquisition (e.g. Major 2008), these contrasts are assumed to play a role in the interlanguage of German leaners. While research on unstressed vowels in GLE remains scarce, studies including other L1-L2 combinations have shed light on possible universal tendencies in learner speech. Based on the contrastive account as well as the findings of previous studies, two hypotheses are formulated and tested in this study.

2. Background

2.1 Speech rhythm

The occurrence of unstressed vowels in connected speech is closely related to the rhythmical properties of a language or accent. Speech rhythm is generally defined as “a perceived regularity of prominent units in speech” (Crystal 2008: 417). Acoustic correlates of prominence are duration, vowel quality, pitch, and intensity. A long-standing claim is that languages can be grouped into two rhythmic classes: syllable-timed and stress-timed (Pike 1945, Abercrombie 1967). This rhythm-class hypothesis still awaits empirical verification (e.g. Roach 1982), and the binary opposition has given way to the view that languages vary along a continuum. The notion of isochrony has been questioned by Dauer (1983), who claims that rhythmic differences between languages reflect a number of phonological properties, such as syllable structure, length as a distinctive feature in vowels, and the (non-)existence of vowel reduction.

Recent attempts to quantify the rhythmic properties of languages have relied on durational measurements to capture the rhythmic qualities of speech. These rhythm metrics can be grouped into global and local measurements. Global measurements include the calculation of (i) the proportion of vocalic intervals (%V) in an utterance and (ii) the standard deviation of vocalic and consonantal interval durations, either as a raw (ΔV and ΔC, Ramus et al. 1999) or rate-normalized measure (VarcoV and VarcoC, Dellwo and Wagner 2003). Low %V values indicate a high degree of vowel reduction and/or high complexity of syllable structure (i.e. stress-timing properties). Syllable structure complexity and vowel reduction are positively correlated with (Varco) ΔC and (Varco) ΔV respectively.

While global measurements pay no attention to the linear arrangement of events, local measures rely on the ratio of successive interval durations (i.e. successive syllables or vocalic/consonantal intervals). Metrics that have been proposed differ in terms of formulae and, crucially, in the choice of intervals for comparison. Low and Grabe’s (1995) Pairwise Variability Index (PVI) and Gibbon and Gut’s (2001) Rhythm Ratio include all successive units, while Gut’s (2003) Syllable Ratio only covers stressed/ unstressed syllable pairs. Generally, local rhythm metrics calculate a single value per speaker or speaker group by averaging durational ratios of successive intervals. A higher degree of temporal vowel reduction is reflected in a higher ratio between successive syllables or vowels.

2.2 Unstressed vowels: English vs. German

German and English are both considered stress-timed languages (Kohler 1995, Giegerich 1992); they share a number of typical properties. Apart from having a complex syllable structure, both languages (i) distinguish stressed and unstressed syllables in terms of quality and quantity; (ii) have the short central vowel [ə] and (iii) show schwa deletion and syllabic consonants as extreme forms of reduction. These features allow the compression of syllable nuclei to (theoretically) achieve isochrony between stressed syllables.

However, the distribution of schwa vowels in German is more restricted (Kaltenbacher 1998). In simple lexemes, they are only found in stem-final syllables (*Hase*, [ˈhaːzə]) or inflectional affixes (*ge-dacht*, [gəˈdaxt]; *denk-e*, [ˈdɛŋkə]). In English, there are no morphological restrictions. In contrast to English, German has a second, more open schwa vowel [ɐ], which occurs in contrast with [ə] (*bitte* [ˈbɪtə], *bitter* [ˈbɪtɐ]). The distribution of schwa also differs in complex lexemes. In both languages, morphophonological processes apply to derived words such as *photography* and *Fotografie*, which differ from their base in terms of primary and/or secondary stress placement. In German, vowel reduction in such cases can be observed as a shortening of long vowels (*Foto* [ˈfo:to] – *Fotograf* [fotoˈgra:f] – *Fotografie* [fotograˈfi:]). A change of vowel quality from tense to lax is less frequent, and vowels are never reduced to schwa. Through productive morphophonological processes in English, on the other hand, unstressed vowels are shortened and centralized to(wards) schwa (*photo* [ˈfəʊtəʊ] – *photograph* [ˈfəʊtəgrɑ:f] – *photography* [fəˈtɒgrəfi]).

In general, the quality of unstressed vowels in polysyllabic words shows a higher degree of reduction in English. In an acoustic analysis of derivational word pairs in four languages, Delattre (1981) compared the first two formants of stressed vowels (base form) with the respective unstressed ones (derivative). In terms of distance in the F_1 x F_2 vowel space, the degree of vowel reduction in German was much smaller than in English.

In connected speech, closed-class function words (e.g. determiners, pronouns, conjunctions, auxiliaries) can undergo reduction in both languages (*und* [ʊnt]→[ənt]→[ən]→[n]; *and* [ænd]→[ənd]→[ən]→[n]). In German, however, these reduction processes are stylistically marked; they only occur in informal speaking styles (Kohler 1995, Wesener 1999). In clear speech, syllable nuclei in monosyllabic function words are not reduced to [ə]. In English, the weak form of function words (which involves [ə] in many cases) is the unmarked variant, even in formal speech. Thus, while both languages show reduction in function words, a centralization of vowel quality is much more common in English, which is primarily due to stylistic differences.

2.3 Unstressed vowels in L2 acquisition

Past research on the presence/absence of reduction phenomena in learner speech has shown this to be an area of difficulty in L2 acquisition. Studies differ in scope and methodology. In an instrumental study of the pronunciation of four derivational word pairs, Flege and Bohn (1989) found that Spanish learners of English lack vowel reduction in unstressed syllables in terms of quality and quantity. Lee et al. (2006) investigated acoustic properties of unstressed vowels in late Japanese and late Korean bilinguals. Compared with native speakers, less vowel reduction and an influence of orthography was found in the productions of the bilinguals, whose unstressed vowels were scattered wider in the F_1 x F_2 space. Gut (2006) analyzed the acoustic properties of several affixes in reading passages and retellings produced by English, Chinese and

Italian learners of German. In post-stress syllables of the type C+<en>, F_1 and F_2 values differed significantly from the native speakers (English learners: F_1 only).

A number of studies have looked at the pronunciation of function words in connected speech. Ghazali and Bouchhioua (2003) auditorily analyzed 15 sentences read by Tunisian learners of English. The strong form of function words (e.g. *for*, *to*, *that*) was produced in 92.5% of the cases. Comparing Japanese learners with native speakers, Aoyama and Guion (2007) analyzed the duration of 3 function words embedded in sentences; these were significantly longer in the speech of learners. Similar findings are reported by Porzuczek (2010) for Polish learners of English. Compared to native speakers, they produced longer vowels in the function word *to* in a reading passage.

Acoustic studies on unstressed vowels in GLE have applied rhythm metrics, thus producing a more general assessment of its rhythmical properties. Comparing German learners with native speakers of British English, Gut (2009) analyzed durational differences of successive stressed/unstressed syllable pairs in a reading passage. The native speakers' Syllable Ratio (2.50:1) was larger than that of the learners (2.23:1), indicating that, while German learners transfer quantity reduction, they do not reach the level of native speakers. Ordin et al. (2011) investigated the timing patterns of GLE at various proficiency levels. The VarcoV and vocalic PVI measurements showed that the variation of vowel durations increases with language proficiency. This might indicate that more advanced learners show a higher durational reduction of unstressed vowels. Due to the lack of a native speaker control group, however, their results cannot be compared with SSBE speech.

Transfer from L1 plays an important role in L2 phonology (e.g. Major 2008). Rhythmic interference in unstressed syllables was observed by Gut (2003), who identified transfer of L1 properties in the speech of Polish, Chinese and Italian learners of German. In a review of past research on L2 stress patterns, Broselow and Kang (2013) conclude that prosodic similarity between L1 and L2 facilitates acquisition; errors tend to reflect L1 influence. A general feature of non-native speech seems to be a tendency to overarticulate compared to native speakers (Barry 2007).

3. Aims and method

3.1 Aims of the study

Research on unstressed vowels in German Learner English has mostly relied on auditory descriptions (e.g. Pascoe 1996, Parkes 2001, Dretzke 2006). This study aims to identify the acoustic properties of unstressed vowels in advanced German Learner English. Based on the prosodic similarities and differences of unstressed vowels described above, German learners are expected to show negative transfer in the reduction of vowel quality. Positive transfer of the reduction of vowel duration is expected. Based on the findings reported by Gut (2009), however, durational reduction in GLE is expected to be smaller than in SSBE. Thus, this study sets out to test two hypotheses: Compared to native speakers, learners produce unstressed vowels with (i) a longer duration and (ii) different F_1 and F_2 values.

3.2 Method and data

Recordings of 4 advanced learners (female, aged 19-29, university students) and 3 native speakers of Standard Southern British English (female, aged 19-30) were analyzed. There were two reading tasks. Task 1 consisted of 66 short phrases, which served to elicit canonical vowels. Structured in a similar way (*I said ..., not ...*), they contained two monosyllabic words. These either formed a minimal pair differing in the vowel or they rhymed, the second slot eliciting (nonsense) words with the syllable frames [hVd] and [hVt] (Peterson and Barney 1952). Task 1 produced 7-12 tokens of each monophthong. These measurements served as reference values for the reduction of vowel quality. Task 2 consisted of 28 sentences, which were designed to elicit (a) stressed vowels (5-7 tokens of each monophthong per speaker) and (b) unstressed vowels in three contexts: in function words (e.g. *to*, *from*, *some*, *have*; 23 tokens per speaker); in polysyllablic lexemes in pre-stress (e.g. *<u>con</u>sider*, *<u>a</u>gain*, *in<u>for</u>mation*; 20 tokens per speaker) and post-stress position (e.g. *num<u>ber</u>*, *pro<u>ba</u>bly*, *atten<u>tion</u>*; 17 tokens per speaker).

To arrive at comparable speech rates in task 2, native speakers were asked to imagine they were reading to a non-native speaker; learners were asked to read in a way they felt comfortable. The speech rate (syllables/second) was calculated for each speaker using the 5% Winsorized mean to avoid the influence of outliers (Wilcox 2012). Pre-pausal syllables were excluded from analysis. The groups differed in speech rate: learners read at a slower rate (M = 5.15, SD = 0.26) than native speakers (M = 5.53, SD = 0.64).

The acoustic analysis was carried out in Praat. Vowel duration was determined using the onset and offset of F_2. Deleted vowels were included in the analysis (duration = 0). Due to the differences in speech rate, the vowel durations were standardized to z-scores for each speaker. This standardization was based on the stressed and unstressed vowels measured in task 2. Z-scores express the duration of vocalic segments in terms of standard deviations away from the mean. To avoid the influence of outliers, the standardization was based on the 5% Winsorized mean and the 5% Winsorized standard deviation. The vowel target was determined visually at the point of maximal displacement (Di Paolo et al. 2011). F_1 and F_2 were transformed to Bark (Traunmüller 1997) using the package `vowels` in R (Kendall and Thomas 2013).

4. Results

This section starts with a descriptive summary of the duration and formant measurements, followed by the results of the hypothesis tests. The boxplots in Figure 1 compare the distribution of unstressed vowel durations in the two groups across the different contexts. Looking at all contexts, there appears to be no difference between the groups. The variation between the three contexts is similar in both groups: vowels in post-stress position were on average longer, those in pre-stress position shorter. Differences between the groups emerge in the three contexts. While the patterns in the weak form words are very similar, there is a clear difference in pre-stress position, with learners producing shorter vowels. In post-stress position, native speakers produce shorter segments.

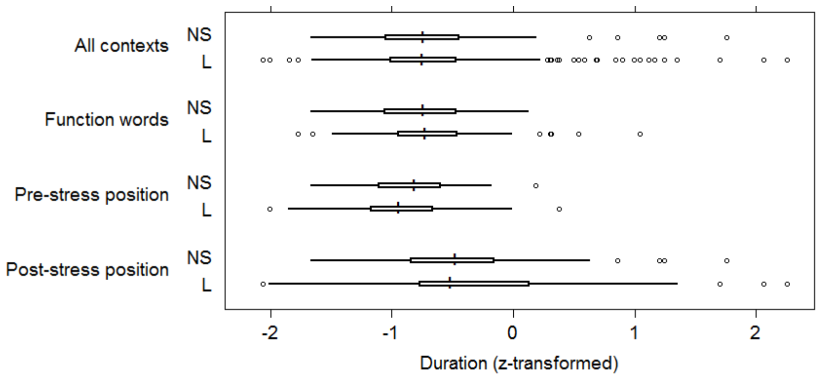

Figure 1: Duration of unstressed vowels by group and context (NS = native speakers; L = learners)

Figures 2a-c show the Bark-transformed formant measurements. The mean values of 8 canonical monophthongs elicited in task 1 are represented as IPA symbols. They serve as reference points of vowel quality. The boxplots at the margins give more detailed information about the distribution of the measurements in the two groups.

Figure 2a plots F_1 and F_2 of all unstressed vowels by group. The location of the centroids reveals that, on average, the learners produced higher F_1 and lower F_2 values. The difference in F_2 is more pronounced, which is clearly illustrated by the boxplots. The ellipses around the centroids, which were drawn with the function `dataEllipse` in the package `car` in R (Fox 2013), contain 50% of the tokens. The area of the two ellipses shows that the variation in the learner group was higher. A comparison of the 3 different contexts reveals that this tendency is less pronounced in post-stress syllables, where the groups produce more similar formant values. The F_1 and F_2 values measured in the weak form words are shown in Figure 2b. The distribution is similar to Figure 2a, but the differences are slightly larger. Higher F_1 and lower F_2 values seem to reflect the influence of the strong form of the function words (e.g. *has*, *have*, *than*, *some*; *was*, *of*, *from*, *for*). The shape of the data ellipses and the F_2 boxplots show that the variation in F_2 was larger in the learner group. To detect a possible influence of orthography (cf. Lee et al. 2006) the set of vowels in pre-stress position was split by their orthographic representation – <o> vs. <a>. While pre-stress <a>-vowels (9 tokens per speaker) differ only slightly, the difference in pre-stress <o>-vowels (8 tokens per speaker) is remarkable. Pre-stress <o> vowels are shown in Figure 2c. The ellipses do not overlap. This is due to the striking difference in F_2.

In the inferential analysis, Wilcoxon tests were used to determine whether the individual learners differed significantly from the group of native speakers regarding duration, F_1 and F_2 of all unstressed vowels. Table 1 lists the results. There was no significant difference in vowel duration. Only Learner 3 had a significantly higher F_1. The differences in F_2, which were described above, were statistically significant for all learners.

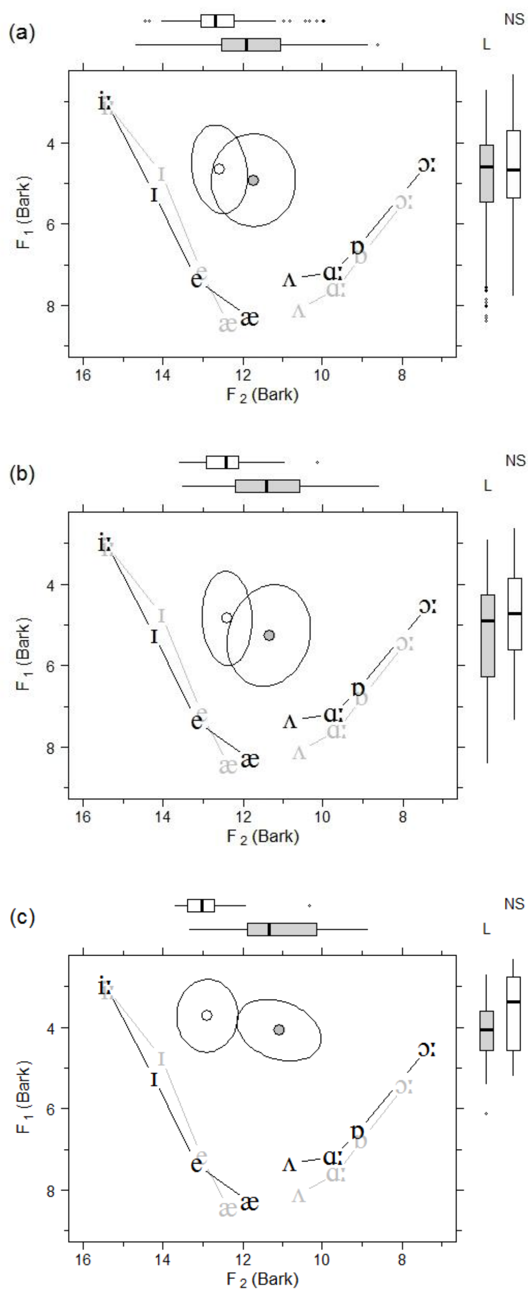

Figure 2: F_1-by-F_2 plot of unstressed vowels (a) in all contexts; (b) in function words; (c) in pre-stress position with orthography <o>. 50% data ellipses are drawn around the centroids (white: NS, grey: L); boxplots show the distribution of F_1 and F_2 by group; IPA symbols represent the centroids of canonical vowels elicited in task 1 (grey: L; black: NS)

	Duration		F_1 (Bark)		F_2 (Bark)	
	Mdn	*p*	*Mdn*	*p*	*Mdn*	*p*
L1	-.77	n.s.	4.2	n.s.	12.1	*
L2	-.69	n.s.	4.6	n.s.	11.8	***
L3	-.73	n.s.	5.8	***	11.9	***
L4	-.77	n.s.	4.3	n.s.	11.9	***
NS	**-.75**		**4.7**		**12.7**	

Table 1: Results of Wilcoxon tests: comparison of each learner with the native speaker group (significance levels: n.s. $p > .05$; * $p < .05$; ** $p < .01$; *** $p < .001$)

5. Discussion

This study investigated two acoustic properties of unstressed vowels in advanced GLE. No significant differences between learners and native speakers were found in the duration of unstressed vowels. Contrary to hypothesis (i), the descriptive analysis showed that the durational measurements were very similar. As predicted by hypothesis (ii), differences were found in vowel quality: learners produced slightly higher F_1 and considerably lower F_2 values, the difference in F_2 being statistically significant for each learner. An exploration of the formant values in the different contexts showed that in post-stress position differences between the groups were less pronounced, while in pre-stress position there was a clear influence of orthography. These findings indicate that in polysyllabic lexemes, the lack of deprominencing of unstressed vowels is higher in pre-stress position than in post-stress position. Differences in vowel quality were also found in weak forms words.

These findings could be explained in terms of transfer from the native language. L1 reduction processes are transferred to L2, leading to positive transfer of vowel duration and negative transfer of vowel quality in unstressed syllables. However, this interpretation overlooks two findings of previous research, which were stated above (cf. section 2.3.): (1) there appears to be a universal tendency of learners to overarticulate; (2) a lack of durational reduction in GLE was reported by Gut (2009). A possible explanation is that the advanced learners in this study have developed beyond the stage where the duration of unstressed vowels is influenced by universal factors. This would be in line with Major's (2001) Ontogeny and Phylogeny Model of second language acquisition. Major claims that the influence of universals first increases and then decreases in the chronological course of second language acquisition. It is possible that this universal feature surfaced at an earlier stage in the interlanguage of the learners investigated in this study. Possibly, further development resulted in a substitution of these universal structures with target language structures. It has to be kept in mind, however, that the findings of this study are not directly comparable to Gut (2009), since vowel reduction was operationalized in different ways.

The use of a reading task for data elicitation is a methodological weakness of this study. First, it is questionable whether this speaking style is representative of learner speech – especially for an investigation of reduction phenomena. Second, the clear influence of orthography revealed in the descriptive analysis might be a characteristic feature of reading style, but not necessarily learner speech.

Duration and vowel quality are acoustic correlates of prominence, and thus contribute to the rhythmic properties of speech. The findings of this study seem to suggest that the differences in speech rhythm between advanced German learners of English and native speakers of British English might concern vowel quality rather than vowel duration. It must be kept in mind, however, that the measurement of selected vocalic intervals in utterances does not allow generalizations about the rhythmic properties of speech. Future research will attempt to bridge the gap between this "local" approach, i.e. the comparison of selected intervals across speakers and groups, and a more "global" approach, i.e. an assessment of rhythmic properties that includes all vocalic intervals in the speech signal.

References

Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Aoyama, K. and S. Guion. 2007. Prosody in second language acquisition: Acoustic analyses of duration and F0 range. In M. Munro and O. Bohn (eds) *Language experience in second language speech learning*. Amsterdam: Benjamins. 281-297.

Barry, W. 2007. Rhythm as an L2 problem: How prosodic is it? In J. Trouvain and U. Gut (eds) *Non-native prosody: phonetic description and teaching practice*. Berlin: Mouton de Gruyter. 97- 120.

Broselow, E. and Y. Kang. 2013. Phonology and speech. In J. Herschensohnand M. Young-Scholten (eds) *The Cambridge handbook of SLA*. Cambridge: CUP. 529-553.

Crystal, D. 2008. *A dictionary of linguistics and phonetics*. Malden: Blackwell.

Dauer, R. 1983. Stress-timing and syllable-timing reanalysed. *Journal of Phonetics* 11: 51-62.

Delattre, P. 1981. An acoustic and articulatory study of vowel reduction in four languages. In P. Delattre (ed) *Studies in comparative phonetics*. Heidelberg: Julius Groos. 63-93.

Dellwo, V. and P. Wagner. 2003. Relations between language rhythm and speech rate. *Proceedings of the 15th ICPhS*, Barcelona, Spain. 471-474.

Di Paolo, M., Yaeger-Dror, M. and A. Beckford Wassink. 2011. Analyzing vowels. In M. Di Paolo and M. Yaeger-Dror (eds) *Sociophonetics. A student's guide*. New York: Routledge. 87-106.

Dretzke, B. 2006. Ausspracheschulung im Fremd-sprachenunterricht. In U. Jung (ed) *Praktische Handreichung für Fremdsprachenlehrer*. Frankfurt: Peter Lang. 132-140.

Flege, J. and O. Bohn. 1989. An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition* 11: 35-62.

Fox, J. 2013. *Package 'car'*. R-package version 2.0-17. http://cran.r-project.org/web/packages/car/car.pdf.

Ghazali, S. and N. Bouchhioua. 2003. The learning of English prosodic structures by speakers of Tunisian Arabic: word stress and weak forms. *Proceedings of the 15th ICPhS*, Barcelona, Spain. 961-964.

Gibbon, D. and U. Gut. 2001. Measuring speech rhythm. *Proceedings of Eurospeech*, Aalborg, Denmark. 91-94.

Giegerich, H.J. 1992. *English Phonology*. Cambridge: Cambridge University Press.

Gut, U. 2003. Non-native speech rhythm in German. *Proceedings of the 15th ICPhS*, Barcelona, Spain. 2437-2440.

Gut, U. 2006. Unstressed vowels in non-native German, *Speech Prosody 2006. Proceedings of the 3rd International Conference on Speech Prosody*, eds. R. Hoffmann and H. Mixdorff. Dresden, Germany.

Gut, U. 2009. *Non-native speech. A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt: Peter Lang.

Kaltenbacher, E. 1998. Zum Sprachrhythmus des Deutschen und seinem Erwerb. In H. Wegener (ed) *Eine zweite Sprache lernen*. Tübingen: Narr. 21-38.

Kendall, T. and E.R. Thomas. 2013. *Package 'vowels'*. R-package version 1.2. http://ncslaap.lib.ncsu.edu/tools/norm/.

Kohler, J. 1995. *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt Verlag.

Lee, B., Guion, S.B. and T. Harada. 2006. Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition* 28(3): 487-513.

Low, E.-L. and E. Grabe. 1995. Prosodic patterns in Singapore English. *Proceedings of the 13th ICPhS*. 636-639.

Major, R. 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology.* Mahwah: Erlbaum.

Major, R. 2008. Transfer in second language phonology: A review. In J. Hansen Edwards and M. Zampini (eds) *Phonology and second language acquisition*. Amsterdam: Benjamins. 63-94.

Ordin, M., Polyanskaya, L. and C. Ulbrich. 2011. Acquisition of timing patterns in second language. *Proceedings of INTERSPEECH 2011*, 1129-1132.

Parkes, G. 2001. *The Mistakes Clinic for German-speaking Learners of English.* Southampton: Englang.

Pascoe, G. 1996. Pronunciation Analysis and Teaching. In W. Barry and A. Addison (eds) *Phonus 2*. Saarbrücken: Institut für Phonetik. 109-122.

Peterson, G.E. and H.E. Barney. 1952. Control methods used in a study of vowels. *Journal of the Acoustical Society of America* 24: 175-184.

Pike, K.L. 1945. *The intonation of American English*. Ann Arbor: Michigan University Press.

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*, ed.. Newcastle upon Tyne: Cambridge Scholars Publishing. 309-322.

Ramus, F., Nespor, M. and J. Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73: 265-292.

Roach, P. 1982. On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. In D. Crystal (ed) *Linguistic controversies. Essays in linguistic theory and practice*. London: Arnold. 73-79.

Traunmüller, H. 1997. *Auditory scales of frequency representation*. http://www.ling.su.se/staff/hartmut/bark.

Wesener, T. 1999. The phonetics of function words in German spontaneous speech. In K. Kohler (ed) *Phrase-level phonetics and phonology*. Kiel: Universität Kiel. 327-377.

Wilcox, R. 2012. *Modern statistics for the social and behavioral sciences.* Boca Raton: CRC Press.

ACOUSTIC CORRELATES OF WORD STRESS AS A CUE TO ACCENT STRENGTH

JAN VOLÍN
Metropolitan University Prague, Czech Republic
jan.volin@ff.cuni.cz

LENKA WEINGARTOVÁ
Institute of Phonetics, Prague, Czech Republic
lenka.weingartova@ff.cuni.cz

Abstract
Due to the clear interference of their mother tongue prosody, many Czech learners produce their English with a conspicuous foreign accent. The goal of the present study is to investigate the acoustic cues that differentiate stressed and unstressed syllabic nuclei and identify individual details concerning their contribution to the specific sound of Czech English. Speech production of sixteen female non-professional Czech and British speakers was analysed with the sounds segmented on a word and phone level and with both canonical and actual stress positions manually marked. Prior to analyses the strength of the foreign accent was assessed in a perception test. Subsequently, stressed and unstressed vowels were measured with respect to their duration, amplitude, fundamental frequency and spectral slope. Our results show that, in general, Czech speakers use much less acoustic marking of stress than the British subjects. The difference is most prominent in the domains of fundamental frequency and amplitude. The Czech speakers also deviate from the canonical placement of stress, shifting it frequently to the first syllable. On the other hand, they seem to approximate the needed durational difference quite successfully. These outcomes support the concept of language interference since they correspond with the existing linguistic knowledge about Czech and English word stress. The study adds specific details concerning the extent of this interference in four acoustic dimensions.

Keywords: Czech English, word stress, duration, F0, SPL, spectral slope

1. Introduction

A foreign accent is a multidimensional phenomenon: its manifestations can be explored with respect to vowels or consonants, intonation or rhythm, or, from the acoustic point of view, in the domains of frequency, intensity, timing, or spectral properties of speech units. Interestingly, the term *accent* itself refers not only to specific pronunciation patterns, but in many languages also to specific prominence lent to a syllable or a word (e.g., Polish *akcent*, English *accent*, French *accent*, German *Akzent*, or Czech *přízvuk*, which also bears both meanings). Although traditional dialectology focused mainly on segmental aspects of regional accents in its descriptions, the term itself is motivated prosodically as if to suggest that one of the most conspicuous features of 'pronunciation other than the reference standard' is prominence distribution and prominence manifestation. This aspect of foreign accentedness is also the focus of our present study.

Despite the generally accepted awareness of the differences between native and non-native treatment of stress patterning, their detailed descriptions, let alone applicable models are still largely missing for most languages and accents. On the other hand, there is a growing body of research reflecting many aspects of the problem. One of the strands of the research shows that the consequences of deviations from expected prominence patterns can be quite important. Crystal (1996: 9) reports cases in which unusual prominence distribution caused communicational problems. In less extreme cases it could be predicted that the foreign speaker is understood, but the extra processing demands may perhaps cause irritation on the part of the listener. It follows that, among many other things, regular stress patterns in speech may be linked to a positive acceptance of the speaker.

Several reliable acoustic correlates of word stress are recognized for English and other languages: F0, duration and intensity were experimented with already in the 1950s (e.g., Fry, 1955; Fry 1958), but also in the following decades (e.g., Klatt, 1976 or Beckmann, 1986). The parameters of spectral slope were acknowledged as important and added later (e.g., Sluijter and van Heuven, 1996). However, stressed and unstressed syllables in Czech do not differ systematically in any of these characteristics (Janota and Palková, 1974; Palková and Volín, 2003; Volín, 2008); although the results concerning spectral slope are still preliminary.

Discussing prominence naturally invites some attention to the non-prominence which provides the indispensable background to perceptually salient elements. Vowel reduction processes are partially responsible for the very specific sound of native English. Standard Czech, on the other hand, requires all vowels to materialize in their full, unreduced forms regardless of their position in the word or phrase. Czech English as an interlanguage is expected to produce a mixture of these two tendencies. Vowel strengthening and weakening in polysyllabic chains creates contrasts that our current study intends to map.

2. Data and method

Recordings of 16 female non-professional speakers aged 20–25 years were used, eight native Czech speakers and eight native English speakers of Southern British Standard. None of them reported any hearing disorder or speech impediment. They were asked to read out an English BBC news bulletin text of 4–4.5 minutes in duration. The speakers were instructed to familiarize themselves with the text beforehand and to read it as naturally as possible. The Czech subjects 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. The British subjects were recorded with a portable professional device Edirol HR-09, with the sampling frequency of 48 kHz and 16-bit resolution. These recordings were afterwards downsampled to 32 kHz to match the sampling frequency of the Czech items.

All the vowels in the recorded utterances were manually labelled in Praat (Boersma and Weenink, 2013) by experienced phoneticians. Apart from the identity and boundaries of the vowels, we carefully annotated the positions of both canonical (according to Wells, 2008) and actually observed word stresses. Not all vowels were selected for analysis: diphthongs and vowels confined to foreign proper names (such as "Arafat") or in words pronounced dysfluently were discarded. Eventually, a total number of 10044 vowels were analyzed, 32 % of them stressed and 68 % unstressed. The identities of the vowels were clustered into six types: /i/, /e/, /a/, /o/, /u/, /ə/ in order to allow for direct comparison of Czech and British vowels. The factual status of the vowel with regard to word stress will be referred to with capital S for stressed and capital U for unstressed vowels.

After the labelling process, the following acoustic measurements were extracted with the help of Praat scripts for each of the vowels:

- duration (in ms)
- fundamental frequency (F0, in semitones relative to 100 Hz)
- sound pressure level (SPL, in dB)
- spectral slope (measure α, in dB)

F0, SPL and spectral slope were measured in the middle third of each vowel to reduce transitional effects from neighbouring consonants or possible annotation inaccuracies. The spectral slope was expressed with the measure α, i.e., the difference between the spectral energy in the frequency bands 0–1000 Hz, and 1000–16000 Hz.

Statistical significance of the results was assessed by two-way analyses of variance (ANOVAs) for independent measures with the two-level factors of LANGUAGE (BrE, CzE) and STRESS (S, U), this being the actual realization of the stress by the given speaker. In order not to inflate the significance of the results, the items were grouped according to speaker, language, vowel type and real stress status. In this way, 180 degrees of freedom instead of 10038 for the ungrouped set were obtained.

The tangibility of the Czech speakers' foreign accent was verified by a perception test. 20 native Czech listeners with working knowledge of English heard two utterances from each of our sixteen speakers in a random order and were asked to judge their native-likeness on a five-point Likert scale, where at one edge number 1 signified "certainly a native English speaker" and at the other edge number 5 meant "certainly a Czech speaker of English". The utterances were selected to satisfy the following conditions: no dysfluencies, no less common proper names, and comparable length of around 5-6 seconds.

3. Results

Figure 1 shows the listeners' answers in the perception test. It is obvious that the group of Czech speakers (left) is clearly separated from the British speakers (right). Also, the fact that the mean scores of the Czech speakers are in all cases between 4 and 5 indicates that their accent is audible and distinguishable even for lay listeners. Even though speaker Cz4 exhibits the most native-like accent, her speech is not at all confusable with native production.

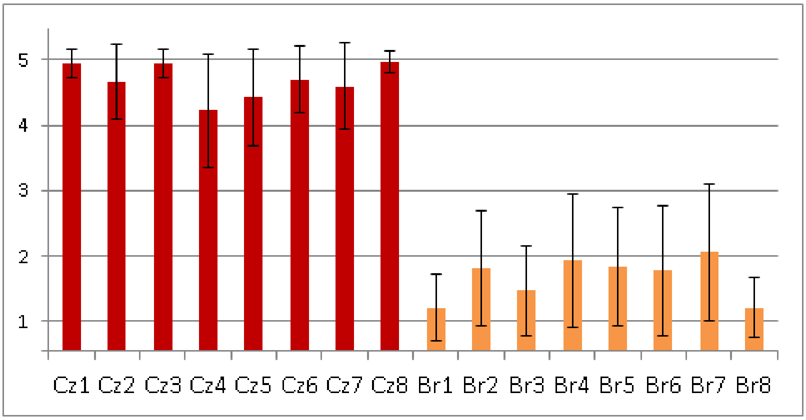

Figure 1: Results of the perception test showing means and standard deviations of the listeners' answers: 1 = "certainly a native speaker", 5 = "certainly a Czech speaker"

Since the information about canonical word stress was available, the alterations in stress placement in the production of both groups of speakers could be assessed (see Figure 2). As expected, the Czech speakers displayed far more alterations than the British speakers. From a total of over 3000 words, 331 words uttered by the Czech speakers and 92 words in the case of British speakers exhibited some form of alteration. Most of them can be attributed to shift of stress to the first syllable which corresponds to the stress pattern which is obligatory in Czech. Interestingly, British speakers also exhibited several shifts to the first syllable (18 words).

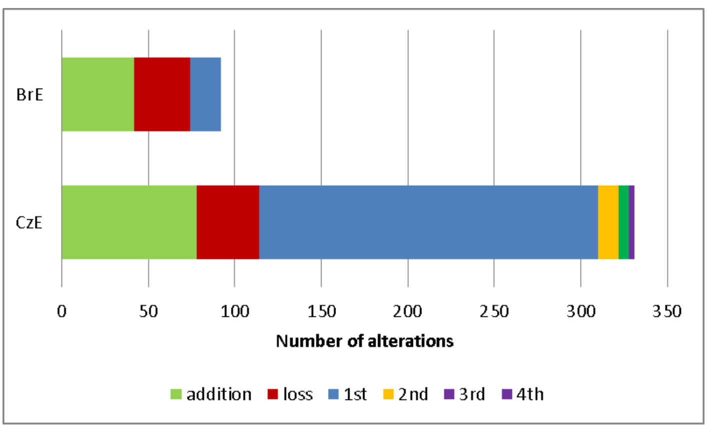

Figure 2: Number of alterations in the material. Addition: a canonically unstressed word received stress; loss: a canonically stressed word does not receive stress; 1st, 2nd, 3rd or 4th: stress shift to the respective syllable.

The results concerning durations of stressed vs. unstressed syllables are displayed in Figure 3. Both Czech and British speakers modify their vowel duration according to the stress status of the vowel; the stressed ones are significantly longer than the unstressed. The two-way ANOVA returns a highly significant effect of STRESS: $F(1, 180) = 80$, $p < 0.001$. However, the difference between Czech and British speakers is not significant.

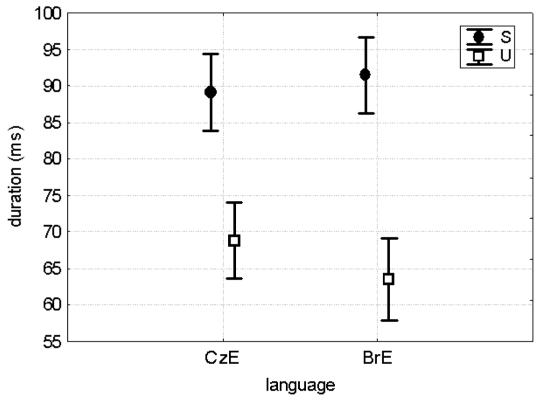

Figure 3: Means of stressed (S) and unstressed (U) vowel durations for Czech (CzE) and British (BrE) speakers. Whiskers indicate 95% confidence interval.

Turning to F0, the results look quite different (see Figure 4). The interaction LANGUAGE*STRESS is significant: $F(1, 180) = 4.42$, $p = 0.037$, which indicates that the Czech speakers treat F0 in stressed and unstressed vowels differently from the British speakers. The trend in Czech speakers appears reversed – unstressed display higher F0 than stressed – but this particular difference is not statistically significant.

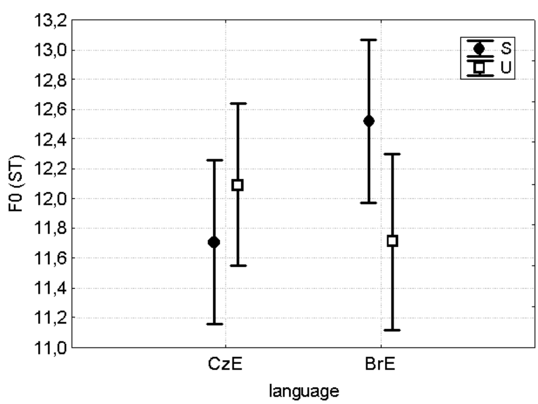

Figure 4: F0 means stressed (S) and unstressed (U) vowels for Czech (CzE) and British (BrE) speakers. Whiskers indicate 95% confidence interval.

Figure 5 shows the differences in spectral slope. It should be noted that caution is needed in this case since the α measure used here is sensitive to speaker and vowel identity (see Weingartová and Volín, 2014) and the vowels of all speakers are pooled together. Even so, both effects were significant, STRESS: $F(1, 180) = 4.01$, $p = 0.047$ and LANGUAGE: $F(1, 180) = 9.68$, $p = 0.002$, while the interaction was not. The significance of the STRESS difference is contributed to mainly by the unstressed British vowels which show the steepest spectral slope. SPL (shown in Figure 6) grants the most clear-cut results from all acoustic correlates of word stress and shows the largest difference between both speaker groups. The effects are significant, STRESS: $F(1, 180) = 62.76$, $p < 0.001$ and LANGUAGE: $F(1, 180) = 9.89$, $p = 0.002$, as well as the interaction: STRESS* LANGUAGE: $F(1, 180) = 16.79$, $p < 0.001$.

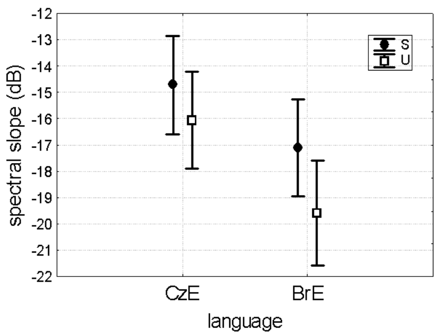

Figure 5: Spectral slope means stressed (S) and unstressed (U) vowels for Czech (CzE) and British (BrE) speakers. Whiskers indicate 95% confidence interval.

While the Czech speakers produce their stressed vowels with only slightly higher SPL (approximately 0.8 dB on average), the British speakers' difference is evidently much more conspicuous (around 2.5 dB on average). It is noteworthy that the unstressed vowels' level is more or less identical for both speaker groups.

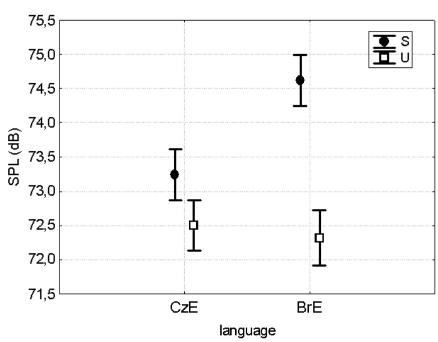

Figure 6: Sound pressure level (SPL) means stressed (S) and unstressed (U) vowels for Czech (CzE) and British (BrE) speakers. Whiskers indicate 95% conf. interval.

4. Discussion and conclusions

We have seen that Czech speakers realize English word stress differently from British speakers in several ways. First, they deviate more substantially from the canonical placement of stress, frequently shifting it to the first syllable of a word. This could be attributed to interference from their native language where the first syllable is stressed obligatorily. Interestingly enough, the British speakers also display some alterations in stress placement. Stress addition or loss on some words may be due to individual rhythmic patterning. When a stress shift on a word in our sample did occur, it was always in favour of the first syllable. As Cutler (2005) already noted, this could be caused by distributional asymmetry of the English word stress patterns.

Concerning the individual acoustic correlates of stress it can be said that Czech speakers in general use much less acoustic marking of stressed vowels than the British. The difference is most prominently seen in F0 (Figure 4) – where the Czech trend is in fact reversed (that is, lower F0 in stressed vowels). As Volín has demonstrated, the predominant F0 pattern in Czech stress-groups is a post-stress rise or L*+H (Volín, 2008). In news reading this is true not only for non-final, but also for final stress-groups, i.e., the post-stress rise occurs very often even in nuclear positions.

Similarly conspicuous difference between Czech and British speakers of English was found in their treatment of SPL (Figure 6), with Czech speakers showing little difference between stressed and unstressed syllables. In this case it could be argued that the perceptual importance of the English free stress is greater than that of the fixed Czech stress. Hence, such a physiologically expensive feature as SPL is habitually avoided by the Czech speakers.

On the other hand, the Czech speakers seem to achieve the required durational difference quite successfully (Figure 3). It can be inferred from the statistical analysis that they are almost native-like in their treatment of the temporal difference between stressed and unstressed vowels even though duration is not typically used as marking of Czech stressed vowels. However, the Czech language does exploit vowel duration to achieve the phonological quantity difference between long and short vocalic phonemes. It is perhaps possible to speculate that the speakers are somehow more sensitive to the temporal differences due to their phonological importance and, therefore, are able to transfer them to a different function in a foreign language. This speculation would, however, require corroboration from other languages.

When interpreting spectral slope differences, caution is needed, since all metrics (including α used here) are sensitive to vowel and speaker identity (Weingartová and Volín, 2014). If we compare the results with our earlier study on *schwa* (Volín *et al.*, 2013), the current results regarding spectral slope are in agreement but much clearer, due to the fact that the identity of the vowel was controlled for. Nevertheless, our findings show a significant effect of stress caused primarily by the unstressed vowels of British speakers, which exhibit the steepest spectral slope. This can be attributed to a lower vocal effort when pronouncing these vowels. Czech speakers display, on the one hand, a smaller difference between stressed and unstressed vowels and, on the other hand, a flatter spectral slope in general. We hypothesize that the experimental task of reading a relatively difficult text in a foreign language could elicit a tenser voice quality which is indeed associated with a less steep spectral tilt (e.g., Hammarberg *et al.*, 1980).

The differences between Czech and British realizations of word stress in English can be ordered according to their significance as follows: The most prominent marker of Czech-accented speech is SPL (in relation to perceived loudness). F0 (in relation to perceived pitch) and spectral slope (in relation to vocal effort) are also treated differently by the native and non-native speakers, but not to the same extent as SPL. Durational differences, on the other hand, show little divergence between both groups and seem to be easy to acquire for the Czech group of speakers: they behave native-like in this respect.

In conclusion, we can summarise the outcome of this study as follows: while the native speakers of British English behave in accordance with earlier literature on acoustic correlates of word stress (e.g. Fry, 1955, Beckman, 1986 or Sluijter and van Heuven, 1996), the Czech group produces significantly different stress manifestations. Since the detail of Czech word stress is not satisfactorily described as yet, our findings might help to illuminate the Czech prominence patterns that cause interference with the native ones. From a didactic point of view, this may be eventually useful for foreign language teaching and learning purposes.

Acknowledgements

The support of the Programme of Scientific Areas Development at Charles University in Prague (PRVOUK), subsection 10 – Linguistics: Social Group Variation is acknowledged. The second author was supported by the project “Acoustic correlates of word stress in Czech, English and Czech English” awarded by the Faculty of Arts, Charles University in Prague in the framework of Specific Academic Research Projects 2014.

References

Beckman, M. E. 1986. *Stress and Non-Stress Accent*. Dordrecht: Foris.

Boersma, P. and Weenink, D. 2013. *Praat: doing phonetics by computer* (version 5.3.41). Retrieved from http://www.praat.org/.

Crystal, D. 1996. The past, present and future of English rhythm. In M. Vaughan-Rees (ed) *Changes in Pronunciation*. Whitstable: IATEFL Pronsig.

Cutler, A. 2005. Lexical stress. In D. B. Pisoni and R. E. Remez (eds) *The handbook of speech perception.* Oxford: Blackwell: 264-289.

Fry, D. B. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustic Society of America* 27: 765-768.

Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech 1*: 126-152.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J. and Wedin, L. 1980. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngologica* 90: 441-451.

Janota, P. and Palková, Z. 1974. Auditory evaluation of stress under the influence of context. *AUC Philologica 2/1974, Phonetica Pragensia,* 4: 29-59.

Klatt, D. H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59: 1208-1221.

Palková, Z. and Volín, J. 2003. The role of F0 contours in determining foot boundaries in Czech. *Proceedings of the 15th ICPhS* 2: 1783-1786. Barcelona: UAB.

Sluijter, A. and van Heuven, V. 1996. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of the 4th International Conference on Spoken Language Processing*, 630-633. Philadelphia.

Volín, J. 2008. Z intonace čtených zpravodajství: výška první slabiky v taktu [Intonation in newsreading: pitch of the first syllable in a stress group]. *Čeština doma a ve světě* 3-4: 89-96.

Volín, J., Weingartová, L. and Skarnitzl, R. 2013. Spectral Characteristics of Schwa in Czech Accented English. *Research in Language* 11(1): 31-39. DOI: 10.2478/v10015-012-0008-6

Weingartová, L. and Volín, J. 2014. Short-term spectral slope measures and their sensitivity to speaker, vowel identity and prominence. *Akustické listy* 20(1): 5-12.

Wells, J. C. 2008. *Longman Pronunciation Dictionary (3rd ed.)*. Harlow: Pearson Education Limited.

LANGUAGE EXPERIENCE AND PHONETIC TRAINING AS FACTORS INFLUENCING TIMING ORGANISATION IN POLISH LEARNERS OF ENGLISH

ANNA GRALIŃSKA-BRAWATA
University of Łódź
agrabra@uni.lodz.pl

Abstract
The paper investigates the dynamics of speech rhythm in Polish learners of English and, specifically, how rhythm measurements revealing durational characteristics of vocalic and consonantal intervals through the measures (%V, ΔV, ΔC, VarcoV, VarcoC and nPVI) change along the process of second language acquisition as a result of language experience and phonetic training, and influence rhythmic characteristics of L2 English. The data used for the analysis come from 30 Polish first-year students of the University of Łódź recorded reading two texts (English and Polish) during two recording sessions separated by a 7-month period of language studies and compared to the data obtained from the recordings of native speakers of English. The experiment aims at verifying whether the participants achieve progress in the rhythm measure scores under the influence of language experience and phonetic training, as it has already been confirmed that general proficiency of non-native speakers of English is a key factor contributing to the successful production of rhythmic patterns in English (Waniek-Klimczak 2009, Roach 2002). The results have shown no substantial and consistent progress for the whole group and across all the measures. Statistical tests, however, have revealed significant changes in the subjects' performance with respect to the vocalic measures ΔV and VarcoV. This may reflect the effect of the type of phonetic training the students are offered, which is segment-based with particular emphasis on vowels.

Keywords: non-native speech rhythm, rhythm measures

1. Introduction

The rhythm of speech is generally understood in terms of the arrangement of alternating stressed and unstressed elements, which are considered temporal organisational units. These elements vary between languages that have been claimed to fall into stress-timed and syllable-timed classes (Pike 1945, Abercrombie 1967), the identity of which, together with the concept of isochrony, has long been the subject of heated discussions. Instead of clear-cut categories between rhythm classes, Dauer (1983, 1987) proposed a model according to which languages can be located on a continuum based on phonetic and phonological criteria with typically stress-timed languages at one end and typically syllable-timed languages at the other end. This model contributed to the development of acoustic correlates of rhythm, or rhythm measures (metrics) that seem to give some evidence for the existence of quantitative rhythmic differences between rhythm classes. The most commonly used include:

- %V – timing proportions of vocalic intervals in an utterance (Ramus et al.1999),
- ΔV and ΔC - standard deviation of vocalic and consonantal intervals (Ramus et al. 1999),
- nPVI (normalised Pairwise Variability Index) - the mean of the differences between successive vocalic intervals divided by the sum of these intervals (Grabe and Low 2002),
- VarcoV and VarcoC - standard deviation of vocalic and consonantal intervals divided by the mean vocalic interval duration (Dellwo and Wagner 2003, White and Mattys 2007).[1]

The measures are based on timing relations between vocalic and consonantal durations, and on the conviction that rhythm type depends on consonantal and vocalic variability.
Motivations for investigating the rhythm of speech are numerous, and commonly are grounded within the context of language teaching (see e.g. Adams 1979), as it is believed that linguistic rhythm contributes substantially to the intelligibility of foreign speech and, consequently, it is one of the most important acquisition targets for language learners. To confirm the importance of rhythm in English over other elements on the lexical level of the language, it is claimed that in the case of foreign speech, poorly pronounced words that conform to the right rhythm have more chances to be intelligible than correctly articulated words with wrong rhythmic structure; in other words, correct rhythmic patterns in L2 speech enhance intelligibility (Avery and Ehrlich 1992, Roach 2002).

1.1 Rhythmic status of English and Polish

The rhythmic status of English is undisputable. From the beginnings of speech rhythm studies, it has frequently been described as a prototypical stress-timed language (Abercrombie 1967: 97, Dauer 1983: 56, Laver 1994: 529) with a tendency for complex syllable structures, variable stress pattern and vowel reductions.

The rhythmic status of Polish is an interesting and unresolved issue as no consensus among phoneticians and phonologists has been reached with regard to its affiliation with stress- or syllable-timed class of languages (e.g. Hayes and Puppel 1985). Rubach and Booij (1985) claim that Polish, although not as typical as English or Russian, is a stress-timed language. Avery and Ehrlich, on the other hand, argue that "Polish is a syllable-timed language and lacks the reduction of vowels so important to English rhythm" (1992: 145). In more recent literature describing Polish rhythmic structure, most attention is devoted to the fact that it appears to be a mixed-type language. It is due to the fact that its features do not match those of typically stress- or syllable-timed

[1] There are other rhythm measures introduced to speech rhythm investigations, e.g. rPVI-C (Grabe and Low 2002), PVI-CV(Barry et al. 2003), med nPVI-V, med rPVI-C (Ferragne and Pellegrino 2004), nPVI-CV (Asu and Nolan 2005), YARD (Wagner and Dellwo 2004). These are not discussed in the paper as they are not employed in the project presented here.

languages (Dauer 1987, Nespor 1990, Ramus et al. 1999). Namely, Polish presents a large variety of syllable types and great syllable complexity, which places it among stress-timed languages. However, it exhibits no phonological vowel reduction at normal speech rates, a characteristic of syllable-timed languages.

1.2 Second language rhythm and rhythm metrics

The studies investigating second language acquisition of English rhythm with the application of rhythm measures in general point to the considerable impact of L1 phonological system (e.g. Low et al. 2000 for Singapore English; Jian 2004 for Taiwanese English; Mok and Dellwo 2008 for Cantonese and Mandarin English) and English proficiency level of L2 speakers on the rhythm metrics (e.g. Ordin et al. 2011 for German English; Shport 2009 for Japanese English).

Research on Polish or Polish English with the use of rhythm metrics is scarce. The experiments conducted by Grabe and Low (2002) and Wagner (2012) belong, to the best knowledge of the author of this paper, to the very few studies that employ calculations of rhythm metrics. The results of their work show that L1 timing patterns and its phonotactic structure are observable factors influencing rhythm metrics. This is true particularly in terms of the variability of vocalic interval duration as “segmental parameter of vowel duration (is an element) deciding about the temporal organisation at higher levels” (Waniek-Klimczak 2005: 9).

Generally, when it comes to the acquisition of English rhythm in the context of L2 speech, the main stages in the development of appropriate timing relations follow similar patterns in the case of both native and non-native speakers, pointing to proficiency as the factor responsible for producing more adequate rhythm (Waniek-Klimczak 2009).

2. Experiment

This paper investigates the effect of L2 training on the rhythmic organisation of English in Polish advanced learners of English, adopting an acoustic phonetics approach to describe rhythm in terms of temporal speech measures. The main assumption adopted in the present study is that increased language experience and phonetic training that Polish learners of English are offered during English studies at university level, affects timing organisation in Polish speakers of English. It is based on the belief that the potential difficulties related to English rhythm that Polish learners of English have, are mainly caused by insufficient fluency and general proficiency in their L2 rather than the stress vs. syllable timed characteristics of English vs. Polish. In view of the recent research, the process of L2 rhythm acquisition of Polish students proceeds in a similar way as in English children who pass through a syllable-timing phase before they are able to perform proper reductions and other prosodic features typical of stress-timing at around 4 years of age. This is grounded in the statement made by Waniek-Klimczak who claims that "a gradual shift towards stress-timing organization of English follows a natural developmental process. It is the stage in second language acquisition and not first language interference that is the main source of difficulty in the acquisition of stress-based English rhythm for native speakers of Polish" (2009: 377). The study also aims to verify the usefulness of selected rhythm metrics proposed in the literature for the study of L2 English.

The following research questions have been formulated:

- Does Polish L1 influence the production of English L2 on the rhythmic level of speech as performed by Polish advanced learners of English? If yes, is the influence reflected through rhythm metrics?
- Does language experience and phonetic training that Polish learners of English are offered during English studies at a university level, affect timing organisation in Polish speakers of English? If yes, what is the direction of change and which rhythm metrics can reflect progress?

2.1 Participants

The data used for the analysis come from 30 Polish first-year students of English at the University of Łódź aged between 19-24. The recording for the present study took place while the respondents attended numerous courses in English as part of the major programme of their studies, which ensures considerable exposure to the language. The phonetic training the students are offered provides intensive input directing their attention to various pronunciation issues, however it is not focused on the rhythm itself. The syllabus of the practical phonetics course for first-year students concentrates mostly on recognising, discriminating and producing the sounds of English on the segmental rather than suprasegmental level of British English standard pronunciation, with some very basic hints of such prosodic elements as prominence, rhythm and intonation.

The selection of participants was not controlled for subjective perception of native/non-native accent or objective fluency tests. All first-year students are expected to meet the requirements of intermediate or upper-intermediate level of language proficiency and they must pass the *Matura* exam at the extended level in English at the end of their secondary education.

Gender and the knowledge of other foreign languages were the variables that were not considered for the experiment, although subjects were asked to answer questions connected to their sex and the level of their proficiency in other foreign languages. Generally, females outnumbered males with the proportion of 25 to 5.

2.1.1 Reference group

The recordings of non-native speech were compared to 11 native English speakers recorded for the purposes of the present experiment. Ten of them form a group assumed to speak standard British English. Their age ranged between 24 and 60 at the time of the recording. Five of them are female and five are male. All the speakers have similar educational background, terminating their education at university level and specialising in instrumental performance.

2.2 Materials

The reading passage used in the experiment was one of Aesop's Fables, *The North Wind and the Sun,* frequently used in phonetic investigations (recommended by the IPA for indicating phonemic contrasts in English) or speech rhythm research (e.g. Grabe and Low 2002).

The material that has been selected for phonetic measurements consists of four phrases within which no pauses or breaks in the flow of speech are to be expected (see Appendix 1). The intention was to investigate phrases of different length and that is why Phrase 1 and Phrase 2 are longer (11 syllables each) and Phrase 3 and Phrase 4 are shorter (5 and 7 syllables).

2.3 Procedure

There were two recording sessions. The first session took place at the beginning of the winter semester (October 2009), prior to the practical phonetics course; the second was arranged at the end of summer semester (May 2010), close to the end of the course. Before the recording the participants were given time to read the text, and asked to read aloud highlighted words that the author predicted might cause some difficulty for Polish speakers, both in terms of word stress and vowel quality. These are as follows: *wind, disputing, cloak, succeeded, blew, attempt, obliged.* Having read and, if necessary, repeated aloud the problematic words, the subjects were asked to read the English text first in a natural manner with their normal tempo (defined in terms of Dellwo, 2010), then to read the same text faster, trying to make a noticeable difference between the two tempos. The whole procedure took place twice and was conducted in an identical way. All the speech samples used in the experiment (L2 and L1 English) were recorded with the use of MXL Studio 1 USB microphone directly to the laptop computer in the .wav format in a relatively quiet, but not soundproof room at the Institute of English Studies, University of Łódź. They were later segmented manually into vocalic and consonantal intervals with the use of Praat, version 5.0.29, mainly according to the criteria employed by Grabe and Low (2002). Both normal tempo and fast tempo variables are taken into consideration.

2.4 Assessment criteria

The experiment aims at verifying whether the participants achieve progress in rhythm measure scores under the influence of language experience and phonetic training. The calculated mean scores of each subject's performance in each of the two recording sessions were compared with the mean scores obtained by the group of native speakers. All the scores have been analysed separately for normal and fast speech. Progress is understood here as the tendency of the Polish subject scores in Recording 2 to approximate the native subject scores to a greater extent than in Recording 1.

Generally, the progress is marked with '+' and its lack is marked with '-'. There is no label for the individual performance that demonstrates the lack of change over time.

Because the obtained scores differ to a various degree across rhythm measures, different assessment criteria have been chosen for identifying progress in each rhythm measure and for each participant.

- For %V, which exhibits the smallest variability between scores, with the difference between mean values of minimum and maximum score being 13.6, progress has been defined in terms of the difference between Rec.1 and Rec.2 for both normal and fast tempo of speech is at least 0.5.
- For ΔV, VarcoV and VarcoC and nPVI, which presents greater variability with the difference between mean values of minimum and maximum score being 43.5, 36.4, 43.6 and 44.7, progress is marked when the difference between Rec.1 and Rec.2 for both normal and fast tempo of speech is at least 1.0.
- For ΔC, which exhibits the greatest variability between scores with the difference between mean values of minimum and maximum score being 60.8, progress has been assumed when the difference between Rec.1 and Rec.2 for both normal and fast tempo of speech is at least 2.0.

Additionally, when no change defined above as progress is observed but both Rec.1 and Rec.2 scores approximate or are identical with the native score, then they are rated as progress. This is motivated by the decision that progress is assessed by means of binary +/- system and choosing '-' could be unfair in a situation when pre- and post-training scores reflect the native normal and fast speech scores and no regress is observed. There are 11 such cases to be observed in the mean data for 4 phrases. If a rhythm metric score for Rec.1 is higher than the corresponding native one and then changes its value in Rec.2 to the one that is slightly below the native score, the change is treated as progress provided that Rec.2 score does not deviate from the native score by more than 1.5 for %V, 2.5 for ΔV, ΔC, VarcoV, VarcoC and nPVI and 8.0 for ΔC (due to its considerable variability) and reflects the tendency of native performance (higher values for Rec.1 and lower values for Rec.2, or conversely lower values for Rec.1 and higher values for Rec.2). If this tendency is retained in non-native speech, but the scores exceed the ones obtained by native speakers by more 2.5 (in italics), no progress is assumed. There are several instances of obtaining reversed scores for normal and fast speech in L2 English and L1 English mean scores, i.e. Rec.1 fast tempo (F) score approximates native F score and Rec.2 F score approximates native normal tempo (N) score. For consistency reasons, in such situations no progress is marked, except when both N and F non-native scores have values falling between the N and F native values.

2.5 Results and analysis

Table 1 shows the average results for 30 Polish learners of English. '+' or '-' has been assigned to each of the measures for each individual learner according to the criteria described in the section above. The final column indicates the sum of '+s' achieved across all measures and tempos by individual participants. The two bottom lines provide the sum of pluses and the percentage of the achieved progress for all individual measures across all the 30 speakers.

Phrases 1-4													
	%V		**ΔV**		**ΔC**		**VarcoV**		**VarcoC**		**nPVI**		**Sum**
	N	**F**	**N**	**F**	**N**	**F**	**N**	**F**	**N**	**F**	**N**	**F**	
1	-	-	+	-	+	+	+	-	+	-	-	+	6
2	+	-	-	-	-	-	+	-	-	-	-	+	3
3	+	-	+	-	-	+	+	-	-	+	+	-	6
4	-	+	-	-	-	+	+	-	+	-	+	-	5
5	-	+	-	-	-	-	+	+	+	-	+	-	5
6	-	+	+	+	+	+	+	-	+	+	-	-	8
7	-	-	+	+	-	+	+	-	-	+	+	-	6
8	+	+	-	+	-	+	+	+	+	-	-	+	8
9	-	-	+	+	-	-	-	-	-	+	+	-	4
10	-	-	-	-	-	+	-	+	-	+	-	+	4
11	-	+	+	-	+	+	-	+	+	+	-	+	8
12	-	+	-	-	-	+	+	+	+	-	+	+	7
13	-	+	-	-	-	+	+	-	-	-	+	-	4
14	-	+	+	-	-	+	+	-	-	+	+	-	6
15	-	-	-	-	-	-	-	-	+	+	-	-	2
16	-	+	+	+	-	+	-	+	-	-	-	+	6
17	-	-	-	+	-	+	+	-	+	+	-	-	5
18	-	-	-	-	+	+	+	+	-	-	-	+	5
19	+	-	+	-	-	+	-	-	+	-	-	-	4
20	+	-	-	-	-	-	-	-	-	+	-	-	2
21	+	+	+	-	+	+	+	+	+	-	+	+	10
22	-	+	+	+	+	+	+	+	+	-	+	+	10
23	-	-	-	+	-	+	+	-	-	+	-	-	4
24	+	-	+	-	+	-	+	-	-	+	-	-	5
25	+	+	+	-	-	-	+	+	-	-	-	-	5
26	+	-	-	-	-	-	-	+	+	+	-	+	5
27	+	-	-	-	-	+	-	-	-	-	-	-	2
28	-	-	-	-	-	-	+	+	+	-	+	+	5
29	+	+	+	+	+	+	-	-	+	+	+	+	10
30	-	+	-	-	+	+	-	-	-	-	+	+	5
sum	11	14	14	9	9	**21**	**19**	12	15	14	13	14	6 mean
%	37%	47%	47%	30%	30%	**70%**	**63%**	40%	**50%**	47%	43%	47%	

Table 1: Progress assessment of rhythm metrics for an individual participant for Phrases 1-4. '+' - progress. '-' - no progress.

Overall, all the participants can be said to progress in at least one rhythm measure getting at least two pluses (3 speakers), which are assigned to VarcoC (3 pluses), %V (2 pluses) and ΔC (1 plus). None of the subjects has progressed in all the measures and for two tempos getting 12 pluses. The greatest progress has been accomplished in the performance of learners 21, 22 and 29 who have been assigned 10 pluses each. The average number of pluses across subjects is 6. Interestingly, the tempo does not seem to be an influential factor in the achieved progress, as it can be observed that the higher number of pluses within each of the six measures is distributed equally for normal (3) and fast (3) tempo (%V-F, ΔV-N, ΔC-F, VarcoV-N, VarcoC-N, nPVI-F). In other words, no pattern of improved scores for normal or fast tempo is to be found across all the measures. It must be remembered that Table1 presents averaged data for four phrases and there are considerable differences between phrases in the achieved progress. Table 2 shows the minimum and maximum number of pluses with the numbers of respective subjects for each of the four phrases and the measure which has been assigned the greatest number of pluses. What can be noticed here is the apparent inconsistency of speakers in their progress across measures. For instance, speaker 15 who in average data in Table 1 gets only 2 pluses for VarcoC and 1 plus for %V in Phrase 2, gets the highest number of 15 pluses in Phrase 4. Similarly, speaker 27 gets the minimum of 2 pluses in the averaged progress results and the maximum of 10 pluses in Phrase 3, Speaker 21 who scores 10 pluses in averaged results, gets 10 pluses in Phrase 2, which seems to be the only directly observed consistency across phrases.

Phrase	Min. nr of +	Max. nr of +	Measure with the greatest nr of +
1	0 (speaker 25)	9 (speaker 10)	ΔC (N-18, F-19)
2	1 (speaker 15, 19)	10 (speaker18, 21)	nPVI (F-19)
3	3 (speaker 4, 9, 13)	10 (speaker 26, 27)	ΔV (N-20) VarcoC (F-20)
4	1 (speaker 3, 10)	8 (speaker 15)	%V (F-17)

Table 2: Minimum and maximum numbers of pluses for each of the phrases and the measure with the greatest number of pluses in each of the phrases.

The inconsistencies are noticeable also with regard to the rhythm measures within which the greatest progress has been attained. Table 1 shows that the maximum of 21 subjects progressed in ΔC scores (fast tempo) and 19 subjects in VarcoV (normal tempo), while the analysis of individual phrases reveals that the each measure undergoes some degree of change. In Phrase1 the greatest number of speakers have improved in terms of ΔC normal tempo (18), and fast tempo (19). In Phrase 2, it is nPVI (fast speech) with 19 speakers. In Phrase 3, it is ΔV (normal tempo) and VarcoC (fast tempo) with 20 speakers each. Finally, in Phrase4, %V (fast tempo) with 17 speakers is the leading measure. Taking into consideration the progress of mean scores for all the phrases within individual measures according to the criteria formulated above and depicted in Table 3, the improvement can be noticed for all the phrases, except for VarcoC (which surprisingly gets the maximum number of pluses in Phrase 3). Partial progress (only normal or fast tempo score has improved) can be observed in ΔC and nPVI. The between-phrase differences may substantially contribute to the differences in the assessed progress between phrases. They included length differences, with Phrases 1 and 2 being longer (11 syllables each), and Phrases 3 and 4 being shorter (5 and 7 syllables), and consequently the differences in the numbers and durations of vocalic and consonantal intervals. Particular phrases may also contain various articulatory challenges with which different learners deal in different ways, and this in turn may affect the timing relations within phrases and consequently the rhythm measures scores. The %V scores do not improve, but they are in close proximity, not exceeding the native score for more than 1.0, and thus classified as progress.

Metric	tempo	Rec.1	Rec.2	Progress	Native
		mean	mean	mean	mean
V%	N	39.8	40.2	+	39.6
	F	40.2	40.6	+	39.6
ΔV	N	50.3	52.6	+	52.2
	F	46.6	46.8	+	46.8
ΔC	N	89.3	86.0	+	87.2
	F	74.8	73.8	-	68.3
varcoV	N	47.5	50.9	+	54.9
	F	49.7	50.8	+	57.8
varcoC	N	57.1	57.4	-	60.2
	F	55.2	55.9	-	59.0
nPVI	N	51.1	53.4	+	58.8
	F	51.6	50.7	-	66.3

Table 3: Assessing progress in the performance of rhythm measures between Recording 1 and Recording 2 for Phrases 1-4 with the scores of native speakers of English. '+' - progress. '-' - no progress.

The improvement scores presented in Tables 1-3 seem to be randomly distributed across phrases and speakers. In order to verify whether there are significant statistical differences between the rhythm metric scores obtained by the group of Polish learners of English between two recordings, a statistical analysis of repeated measures design has been employed and presented in Table 4.

measure	%V		ΔV		ΔC		VarcoV		VarcoC		nPVI	
tempo	N	F	N	F	N	F	N	F	N	F	N	F
p	0.1262	0.2341	0.0266	0.4132	0.0709	0.2615	0.0002	0.1913	0.4211	0.2206	0.0608	0.2473

Table 4: The results of one-tailed paired t test verifying significant differences in L2 English between two recordings, N=30.

It shows that statistically significant differences have been achieved only in terms of ΔV normal tempo (p=0.0266) and VarcoV normal tempo (p=0.0002). ΔC in fast tempo and nPVI in normal tempo obtained scores close to significance level (p=0.0709 and p=0.0608 respectively).

3. Discussion

The general assumption of the experiment presented here is that with increased exposure to English and regular formal instruction concerning English pronunciation, L2 speech rhythm should improve and approximate native speech rhythm production. It has already been confirmed that general proficiency of non-native speakers of English is a key factor contributing to the successful production of rhythmic patterns in English (Waniek-Klimczak 2009, Roach 2002). However, very few studies have provided evidence to rhythmic differences being connected with language learners' proficiency level. Dellwo et al. (2009) have not observed any significant changes in rhythmic scores for Spanish learners of English calculated twice after pre- and post-training recording sessions. Similarly, Gut (2012) has found no effect of either the training course in pronunciation or the stay abroad on rhythm measures in her comparison of the rhythm of language learners after a 6-month training and before and after a 9-month stay abroad in the target language country. One of the few studies that indicates the difference between learners at different competence levels is the one conducted by Stockmal et al. (2005). They found that the metrics able to show significant difference between beginners and more advanced learners of Latvian with L1 Russian are ΔC and consonantal PVI. %V and nPVI have not been found to differentiate between the two groups of learners.

The results of the present study provide some weak support for Stockmal et al.'s findings in that the highest number of improved scores has been noted for ΔC as well. The surprising and contradictory result is that this is true for the measure's fast speech version, while the normal speech ΔC, together with ΔV for fast tempo, got the worst improvement scores. Statistical tests, however, have not verified the above results and revealed significant changes in the subjects' performance with respect to the vocalic measures ΔV and VarcoV in normal tempo. The direction of change within the rhythm measures has been identified to lead towards the native English speakers' scores as depicted in Table 5 below.

	Recording1	**Recording2**	**Native English**
ΔV	50.3	52.6	52.2
VarcoV	47.4	50.9	54.9

Table 5: The mean results of ΔV and VarcoV normal tempo obtained for Recording 1 and Recording 2, and compared to native English speakers' scores.

Although statistical differences have been identified for normal tempo, when analysed across all the measures, speech tempo does not appear to have a crucial impact on the learners' performance. In faster tempo L1 English can be expected to use more reduction, hence it becomes more difficult for its learners. Ideally, faster L2 English should be closer to normal tempo L1 English. However, no regularity of that kind is to be found within the data, as improved scores are similarly distributed for normal and fast tempo across the rhythm metrics.

Out of six measures, only two (ΔV and VarcoV) and only in normal tempo have revealed significant differences between scores. The fact that the progress *was* verified for two measures gives hope that, given some more time, the learners may manage to master their pronunciation to the degree that may be reflected by other rhythmic measures.

4. Conclusion

The analyses of rhythm metric scores for 30 subjects' pre- and post-training performance have revealed significant differences in terms of two vocalic measures (ΔV and VarcoV). No substantial and consistent progress has been observed for the whole group and across all the measures. Although there are instances of observable improvement of scores in relation to native speakers' scores in the case of individual subjects and measures, one cannot speak of global progress and an essential change of participants' linguistic behaviour as:

- the average number of pluses indicating progress across all the measures for the group is 6 which covers only half of the 12 potential improvement points, suggesting that the progress is partial;
- different measure is 'the winner' in different phrases and the means for all phrases do not necessarily reflect reliably the tendencies observed in each individual phrase;
- the subject who is an outlier and performs progress in terms of the majority of measures according to one phrase scores much worse according to the remaining phrases.

The reasons why no clear and consistent change in all the rhythm measures is observed between the recordings are not obvious. The situation may result from the fact that rhythm metrics of the type discussed above do not constitute reliable indicators for changes in L2 proficiency, or that the differences in rhythmic proficiency of the learners between the two recordings are not large enough to be measurable with the metrics applied in this study. Perhaps, the type of training (mainly segment-based) contributes to the situation, as focusing on the qualities of individual sounds may disturb the acquisition and production of prosodically satisfactory utterances. As far as the exposure to English is concerned, it remains true that the majority of lecturers, including pronunciation teachers, are Polish, and their speech rhythm patterns may also diverge from the ones represented by native speakers of English.

Despite the prevailing conviction that teaching stress-timing of English is ineffective and should be omitted in favour of segment-based training (Jenkins 2000), the fact that English as performed by Polish learners reveals changes in the rhythm measure scores suggests that speech rhythm is teachable and should constitute an important element of formal instruction in the pronunciation teaching syllabuses.

References

Abercrombie, D. 1967. Elements of general phonetics. Edinburgh: Edinburgh University Press.

Adams, C. 1979. English Speech Rhythm and the Foreign Learner. The Hague: Mouton.

Asu, E.L., Nolan, F. 2005. Estonian rhythm and the Pairwise Variability Index. In: Proceedings, FONETIK 2005, Department of Linguistics, Göteborg University.

Avery, P. and Ehrlich, S. 1992. Teaching American English Pronunciation. Oxford: Oxford University Press.

Barry, W.J., Andreeva, B, Russo, M., Dimitrova, S. and Kostadinova, T. 2003. Do Rhythm Measures Tell us Anything about Language Type? In: D. Recasens, M.J.Solé and J. Romero, editors, Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003), Barcelona: 2693-2696.

Boersma, P. And Weenink, D. 2008. Praat:Doing phonetics by computer (verion5.0.29). Retrieved 14th July 2008, from <http://www.praat.org/>.

Dauer, R. 1987. Phonetic and phonological components of language rhythm. Proceedings of the 11th International Congress of Phonetic Sciences, Tallin, Vol.5: 447-450.
Dauer, R. M. 1983. Stress-timing and syllable-timing reanalysed. Journal of Phonetics 11: 51-69.
Dellwo, V., F. Gutierrez Diez, and N. Gavalda. 2009. The development of measurable speech rhythm in. Spanish speakers of English In Actas de XI Simposio Internacional de Comunicacion Social, Santiago de Cuba: 594-597.
Dellwo, V. 2010. Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence. PhD-Dissertation, Universität Bonn (electronic publication: http://hss.ulb.uni-bonn.de:90 /2010/2003/2003.htm).
Dellwo, V., Wagner, P. 2000. Relations between language rhythm and speech rate. Proceedings of the International Congress of Phonetic Science. Barcelona: 471-474.
Ferragne, E. and Pellegrino, F. 2008. Le rythme dans les dialectes de l'anglais: une affaire d'intensité ? Actes de Journées d'Etude de la Parole, Avignon, 9-13 juin 2008.
Grabe, E. and Low, E.L. 2002. Durational variability in speech and the rhythm class hypothesis. In: C. Gussenhoven and N. Warner (eds) Papers in Laboratory Phonology 7. Cambridge: CUP.
Gut, U. 2012. Rhythm in L2 Speech. In Gibbon, Dafydd (ed.), Speech and Language Technology. Poznań: 83-94.
Hayes, B. and S. Puppel. 1985. On the rhythm rule in Polish. In: van der Hulst, H. and N. Smith. (eds) Advances in non-linear phonology. Dordrecht: Foris Publications: 59-81.
Jenkins, J. 2000. The Phonology of English as an International Language. Oxford: Oxford University Press.
Jian, Hua-Li. 2004. On the syllable timing in Taiwan English, In B. Bel /. I. Marlien (eds) Proceedings of Speech Prosody 2004, Nara (Japan) 23-26 March 2004, ISCA Special Interest Group on Speech Prosody (SproSIG), 247-250.
Laver, J. 1994. Principles of phonetics. Cambridge: Cambridge University Press.
Low, E. L., Grabe E. and Nolan F. 2000. Quantitative characterization of speech rhythm: Syllable-timing in Singapore English. Language and Speech 43: 377-401.
Mok, P., Dellwo, V. 2008. Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. Speech Prosody 2008, Campinas, 423-426.
Nespor, M. 1990. On the rhythm parameter in phonology, Roca Iggy M., (ed.), Logical Issues in Language Acquisition, Dordrecht: Foris: 157-175.
Ordin, M., Polyanskaya, L., Ulbrich, Ch. 2011. Acquisition of Timing Patterns in Second Language. INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association, Florence, Italy, August 27-31, 2011, Florence, Italy; 08/2011.
Pike, K.L. 1945. The intonation of American English. Ann Arbor: University of Michigan Press.
Ramus, F., Nespor, M., and Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. Cognition, 72, 1-28.

Roach, P. 2002. Studying rhythm and timing in English speech: Scientific curiosity, or a classroom necessity? In: E. Waniek-Klimczak and P. J. Melia (eds) Accents and Speech in Teaching English Phonetics and Phonology. Łódź Studies in Language Vol.5. Frankfurt am Main: Peter Lang: 199-206.

Rubach, J. and Booij, G.E. 1985. A grid theory of stress in Polish. Lingua 66, North Holland: 281-319.

Shport, I. 2009. The effect of oral proficiency on production of rhythm in spontaneous second language (L2) Japanese speech. The Journal of the Acoustical Society of America vol. 125: 2728-2728.

Stockmal, V., D. Marcus, and D. Bond. 2005. Measures of Native and Non-Native Rhythm in a Quantity Language. Language and Speech 48: 55-63.

Wagner, P.S. 2012. Speech rhythm in native and non-native Polish. Proceedings of the International Conference of Experimental Linguistics ExLing 2012: 27-29 August 2012, Athens, Greece.

Wagner, P.S., Dellwo, V. 2004. Introducing YARD (Yet Another Rhythm Determination) And Re-Introducing Isochrony to Rhythm Research. Speech Prosody 2004, Nara, Japan.

Waniek-Klimczak, E. 2009. Sources of difficulty in the acquisition of English rhythm by Polish learners: A developmental perspective. In M. Wysocka (ed.) On Language Structure, Acquisition and Teaching: 359-169. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Waniek-Klimczak, E. 2005. Temporal Parameters in Second Language Speech: An Applied Linguistic Phonetics Approach, Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

White, L. and Mattys, S. 2007. Rhythmic typology and variation in first and second languages. In P. Prietro, J. Mascaró, and M. J. Solé (eds) Segmental and prosodic issues in romance phonology (Current issues in linguistic theory series). Amsterdam, Philadelphia: John Benjamins.

Appendix 1

Phrase1: *should be considered stronger than the other*

ʃ	ə	d	b	iː	k	ə	n	s	ɪ	d	ə	d	s	t	r	ɒ	ŋ	g	ə
C	V	C	C	V	C	V	C	C	V	C	V	C	C	C	C	V	C	C	V
ð	ə	n	ð	i	ʌ	ð	ə												
C	V	C	C	V	V	C	V												

Phrase2: *and at last the North Wind gave up the attempt*

ə	n	d	ə	t	l	ɑː	s	t	ð	ə	n	ɔː	θ	w	ɪ	n	d	g	eɪ	v
V	C	C	V	C	C	V	C	C	C	V	C	V	C	C	V	C	C	C	V	C
ʌ	p	ð	i	ə	t	e	m	p	t											
V	C	C	V	V	C	V	C	C	C											

Phrase3: *wrapped in a warm cloak*

r	æ	p	t	ɪ	n	ə	w	ɔː	m	k	l	əʊ	k
C	V	C	C	V	C	V	C	V	C	C	C	V	C

Phrase4: *then the Sun shone out warmly*

ð	e	n	ð	ə	s	ʌ	n	ʃ	ɒ	n	aʊ	t	w	ɔː	m	l	i
C	V	C	C	V	C	V	C	C	V	C	V	C	C	V	C	C	V

 DOI: 10.2478/rela-2014-0007

ASSIMILATION OF VOICING IN CZECH SPEAKERS OF ENGLISH: THE EFFECT OF THE DEGREE OF ACCENTEDNESS

RADEK SKARNITZL
Institute of Phonetics, Faculty of Arts, Charles University in Prague
skarnitzl@mup.cz

PAVEL ŠTURM
Institute of Phonetics, Faculty of Arts, Charles University in Prague
Pavel.Sturm@ff.cuni.cz

Abstract
Czech and English are languages which differ with respect to the implementation of voicing. Unlike in English, there is a considerable agreement between phonological (systemic) and phonetic (actual) voicing in Czech, and, more importantly, the two languages have different strategies for the assimilation of voicing across the word boundary. The present study investigates the voicing in word-final obstruents in Czech speakers of English with the specific aim of ascertaining whether the degree of the speakers' foreign accent correlates with the way they treat English obstruents in assimilatory contexts. L2 speakers, divided into three groups of varying accentedness, were examined employing *categorization* and a *voicing profile* method for establishing the presence/absence of voicing. The results suggest that speakers with a different degree of Czech accent do differ in their realization of voicing in the way predicted by a negative transfer of assimilatory habits from Czech.

Keywords: foreign accent, voicing, assimilation of voicing, Czech, English

1. Introduction

Voicing is a phenomenon which distinguishes members of speech sound pairs like /t/-/d/ or /s/-/z/. A cursory look at the IPA chart may suggest that voicing is a trivial binary contrast, accomplished by a simple "switch", the vocal folds. A number of studies showed that this is not the case: the laryngeal muscles are activated and suppressed in a complex manner (see, e.g., Hirose & Gay, 1972; Löfqvist & Yoshioka, 1980; Ridouane, Fuchs & Hoole, 2006).

On the linguistic level, it is necessary to distinguish, on the one hand, voicing from the perspective of the phonological system of the given language, the so-called **phonological voicing**, and, on the other hand, voicing defined by the actual presence of vocal fold vibration, or what we call **phonetic voicing**. We may then talk about the phonetic implementation of phonological voicing, and this is a characteristic in which individual languages may differ. Thus, while no one would dispute the existence of a two-way voicing contrast – in other words the existence of phonological pairs like /p/-/b/ or /s/-/z/ – in both Czech and English, the voicing contrast in these two languages is implemented differently from the phonetic point of view.

In Czech, we observe a considerable agreement between phonological and phonetic voicing (Skarnitzl, 2011). That means that phonologically voiceless obstruents like /p/ or /s/ are really pronounced as voiceless, without vocal fold vibration, while phonologically voiced /b/ or /z/ are typically produced with vibration of the vocal folds. Although devoicing of voiced obstruents does occur in Czech (Skarnitzl, 2011: Chapter 9), it may be regarded as online, momentary reduction of laryngeal gestures, rather than a systematic phenomenon. In addition, Czech neutralizes voicing word finally, so that words like *les* (*forest*, NOM, SG) and *lez* (*crawl*, IMPERATIVE, SG-2) will have identical surface phonetic representation, /les/.

In English, phonologically voiceless obstruents are pronounced as phonetically voiceless (and plosives are accompanied by aspiration in many positions), but phonologically voiced obstruents become partially or completely devoiced in many contexts (see, e.g., Ogden, 2009, p. 99ff.; Roach, 2009, p. 26ff.). For example, in the words *dog* or *buzz*, none of the consonants will be pronounced with full voicing but only with partial voicing; they will be transcribed [d̥ɒg̊] and [b̥ʌz̥]. As a descriptive category, voicing is therefore not very useful in some languages, and **tenseness** has come to be exploited to account for differences between phonologically voiced and voiceless obstruents (Ladefoged & Maddieson, 1996, p. 95; Butcher, 2004; Kohler, 1984). The former are called lenis or lax, suggesting a weaker pronunciation, while the latter are called fortis or tense, suggesting stronger articulation.

In the context of second language acquisition, the differences in the implementation of phonological voicing in Czech and English are likely to result in negative transfer from the learner's native language into the second language. We have to realize that the laryngeal settings associated with the distinction between fully and partially voiced obstruents are extremely fine. We believe that, due to this subtle difference, a fully voiced obstruent like [z] and a partially voiced or devoiced obstruent like [z̥] will be categorized as similar sounds in Flege's Speech Learning Model (Flege, 1987; 1995); the same would probably apply for a devoiced [z̥] and voiceless [s]. It is well known that our speech perception system is tuned into the phonological system of our native language (L1) and that we have learned to ignore the "ballast" of acoustic differences which are irrelevant for the distinguishing of L1 phonemic categories (see, e.g., Jusczyk, 1993; Kuhl & Iverson, 1995). It is precisely the relatively small acoustic differences between similar L1 and L2 sounds which are, according to Flege, rather difficult to acquire for foreign learners. Furthermore, pronunciation instruction in English as a foreign language only seldom targets voicing (Skarnitzl, 2002), and at least Czech students of English are rarely aware that voicing functions differently in the two languages.

On the one hand, therefore, we are talking about subtle distinctions on the laryngeal level and we might want to dismiss voicing as a marginal, irrelevant phenomenon in the acquisition of English as a second language. It is true that, to our knowledge, voicing has not been the target of much research from the perspective of second language acquisition. However, we believe that when these L1–L2 differences in the implementation of voicing are combined with divergent strategies for the **assimilation of voicing** across the word boundary, the transfer effect may become much more perceptible for native listeners.

In Czech, (mostly) regressive assimilation of voicing is very pervasive: *pes*[s] *plaval* (*the dog was swimming*) – *pes*[z] *běžel* (*the dog was running*). In contrast, voicing assimilation is very rare in English: according to Cruttenden, only phonologically voiced fricatives may be realized as voiceless by some speakers "if the two words form part of a close-knit group" (Cruttenden, 2008, p. 299f.). He mentions *with*[θ] *thanks*, *of*[f] *course* or *was*[s] *sent* as examples. As the examples indicate, voicing assimilation concerns only grammatical words and, most importantly, voicing may only be lost, never gained. The pronunciation of the phrase *Have a nice day* with [naɪz deɪ] rather than [naɪs d̥eɪ] leads to an impression of a foreign accent (*cf.* Roach, 2009, p. 212).

In a preliminary study, Skarnitzl & Poesová (2008) investigated the pronunciation of word-final obstruents by five Czech speakers of English who had been evaluated as having a strong Czech accent (see Skarnitzl, Volín & Drenková, 2005 for more detail). The results showed that, for instance, word-final neutralization of voicing was nearly ubiquitous in voiced–voiceless contexts across the word boundary (e.g., *calle<u>d P</u>eter*), or that in voiceless–voiced contexts (e.g., *ba<u>ck d</u>oor*) the voiceless obstruent assimilated to the following voiced obstruent in approximately one half of all items.

The present study investigates voicing in word-final obstruents in Czech speakers of English; however, in addition to the previous study, we are adding the degree of the speakers' foreign accent into play as a factor. In other words, we are interested in whether the degree of the speakers' foreign accent correlates with the way they treat English obstruents in assimilatory contexts. Specifically, we hypothesize that speakers with a strong Czech accent in English will neutralize and assimilate voicing word-finally, depending on the underlying voicing status of the neighbouring sounds, while speakers who have a near-native English will not assimilate voicing in voiceless–voiced contexts, nor neutralize voicing in voiced–voiceless contexts.

2. Method

We examined twelve Czech female speakers of English, aged 20–25, all students of English and American Studies at the Faculty of Arts in Prague. Four of the speakers had been previously (Skarnitzl et al., 2005) classified as having a near-native English accent (henceforth referred to as Group A), four as manifesting a recognizable but not strong Czech accent (Group B), and four as having a strong Czech accent (Group C). The speakers were asked to read, after sufficient time for preparation, a news bulletin from BBC World Service. The recordings were obtained at the sound-treated recording studio of the Institute of Phonetics in Prague, using a studio electret microphone IMG ECM 2000 and digitized at the sampling rate of 22,050 Hz.

The recordings were automatically segmented using the Penn Phonetics Lab Forced Aligner (P2FA; Yuan & Liberman, 2008), and the boundaries of the target speech sounds were then manually adjusted following the recommendations listed in Machač & Skarnitzl (2009). By target sounds we will mean two consecutive phones – the word-final obstruent and the initial sound of the following word; since the voicing of the former may not be independent of that of the latter in Czech speakers of English, we were interested in both of them.

In order to assess the voicing of the target sounds, fundamental frequency (F0) information – or more specifically, the presence or absence of F0 – was extracted at eleven equidistant points throughout the final obstruent using a Praat (Boersma & Weenink, 2013) script.

In subsequent analyses, we evaluated the degree of voicing using two methods. First, we characterized each word-final obstruent using the **categorization** proposed by Smith (1997) for [z] in American English: *voiceless* when voicing was present in less than 25 % of the sound's duration, *voiced* when voicing was present in over 90 % of the sound's duration, and *partially voiced* in the remaining cases. Second, we applied the **voicing profile** method (Shih & Möbius, 1998; Möbius, 2004) to assess the dynamic changes of the probability of voicing in normalized time. The probability of voicing at the given point of the obstruent's duration refers to the ratio of voiced realizations to all realizations; for instance, there is a 95% probability of voicing in the temporal midpoint of Czech [z], but only a 85% probability in the midpoint of [ʒ], and as low as a 50% probability of voicing in the middle of the fricative trill [r̝] (Skarnitzl, 2011, p. 215).

3. Results

The results suggest that speakers with a different degree of Czech accent do differ in their realization of voicing. Disregarding sequences of voiceless obstruents, we distinguished three assimilatory contexts at word boundaries: (1) the *voiced–voiceless* context (*vd-vl*), where we predicted the loss of voicing of the first obstruent in the least proficient speakers; (2) the *voiceless–voiced* context (*vl-vd*), in which we anticipated regressive assimilation of voicing in the least proficient speakers; and (3) the *voiced–voiced* context (*vd-vd*), where two voiced obstruents meet. Here, we predicted partial voicing of both obstruents in the proficient group, but either a full voicing of both obstruents or a loss of voicing of the first obstruent for the less proficient speakers.

As regards the **voiced–voiceless** context, the "Czech" neutralization (i.e., loss) of word-final voicing was strongest in Group C and weakest in Group A. There was a monotonic relationship between the number of realizations categorized as phonetically voiceless and the degree of Czech accent (Fig. 1a). The differences between the categories and groups were highly significant: χ^2 (4; n = 849) = 39.2, p < 0.001. The voicing profile, which adds a dynamic dimension to the results, revealed similar tendencies: the probability of voicing of the word-final obstruent decreased most rapidly in Group C (at 40% of the sound's duration it was already low), and slowest in Group A (see Fig. 1b).

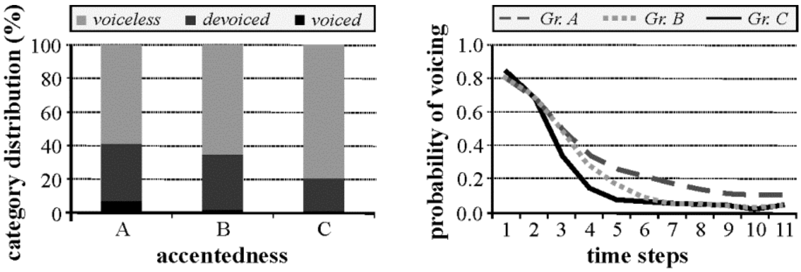

Figure 1: *Voiced–voiceless assimilatory context* for speakers with different degrees of accentedness (A = "native-like English", C = "strong Czech accent"): **a.** categorization according to Smith (1997); **b.** the voicing profile.

Unfortunately, the inverse **voiceless–voiced** context did not yield enough cases to allow drawing serious conclusions (*n* = 84). Group A behaved as expected, i.e. the proficient speakers kept the final obstruent voiceless in more cases than the other two groups, and relatively rarely assimilated voicing. However, Group B proved to be more "Czech-like" than Group C. This holds true both for the categorization according to Smith and for the voicing profile (Fig. 2a and 2b).

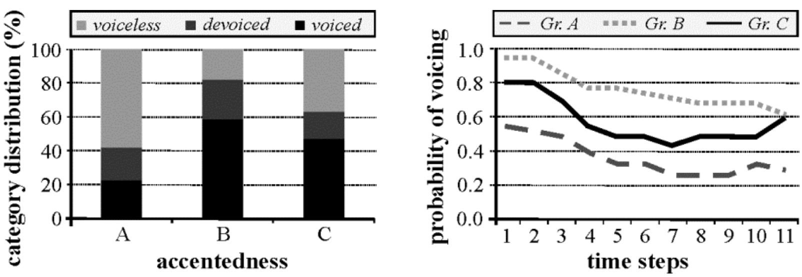

Figure 2: *Voiceless–voiced assimilatory context* for speakers with different degrees of accentedness (A = "native-like English", C = "strong Czech accent"): **a.** categorization according to Smith (1997); **b.** the voicing profile.

In sequences of **two voiced obstruents** across the word boundary, Group A manifested the greatest proportion of not fully voiced tokens, indicating the English-like process of word-final devoicing. The differences between the groups were significant: χ^2 (4; n = 284) = 9.54, $p < 0.05$. Although the probability of voicing was quite high for all groups, Group A has a drop in the profile towards the end, indicating a tendency for partial voicing (Fig. 3a and 3b).

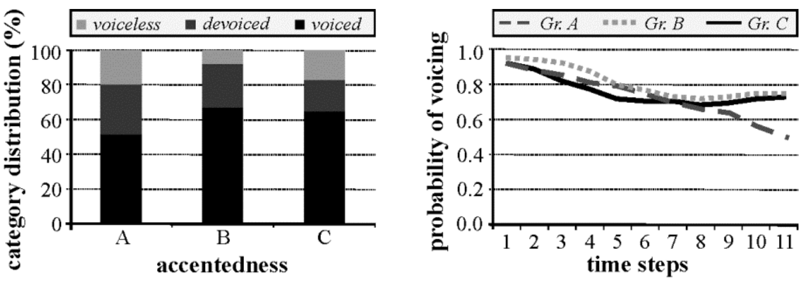

Figure 3: *Voiced–voiced assimilatory context* for speakers with different degrees of accentedness (A = "native-like English", C = "strong Czech accent"):
a. categorization according to Smith (1997); **b.** the voicing profile.

In addition to these overall results, we analyzed the data with respect to **individual speakers** (four in each group). The first context (*vd-vl*) showed a gradual decline in the voicing profile for all Group A speakers as opposed to a sharp drop for the speakers from Group C. The individual speakers thus behaved uniformly, with the exception of a negligible two-cluster pattern in Group A.

The inverse context (*vl-vd*) revealed significant between-speaker differences. As can be seen in Fig. 4a, two speakers from Group A had a low probability of voicing, especially towards the end of the sound, while one seemed to voice all instances. However, this speaker contributed only four instances of this assimilatory context, which might explain her outlying position. Group B was more uniform, with three speakers assimilating voicing in all items and only one speaker demonstrating a lesser tendency to assimilation (Fig. 4b), comparable to speakers in Group A. Group C, with the highest degree of Czech accent, ranged from 0% probability to 100% probability of voicing; however, the two speakers in these extreme positions contributed only 2 and 3 items, respectively. The *vl-vd* assimilatory context must therefore be considered unrepresentative and the results only tentative.

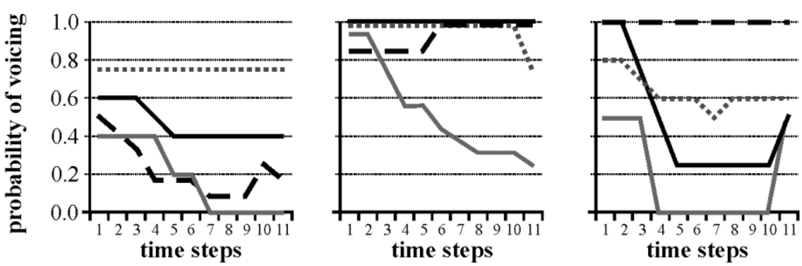

Figure 4: Voicing profiles (*voiceless–voiced assimilatory context*) for three groups of speakers (individual speakers denoted by different lines): **a.** native-like English (group A); **b.** mild Czech accent (group B) and **c.** strong Czech accent (group C).

The last context (*vd-vd*) did not reveal any significant differences between individual speakers within the groups. In Group A there was one speaker who diverged slightly, but not markedly, in the direction towards full voicing. Group B was completely uniform, and Group C also included one subject who slightly diverged from the rest of the group towards full voicing.

Lastly, we were interested in other linguistic effects, such as the **type of word** the final obstruent appeared in. Restricting our attention to the two contexts with a sufficient number of cases (*vd-vl* and *vd-vd*), we found that the presence of voicing was stronger in synsemantic (function) than in autosemantic (content) words (see Fig. 5a and 5b). For instance, devoicing of the first consonant in *vd-vd* sequences was stronger in words like *called Barry* than *his brother*. Similarly, *world cup* in the other context was devoiced more frequently than *could face*.

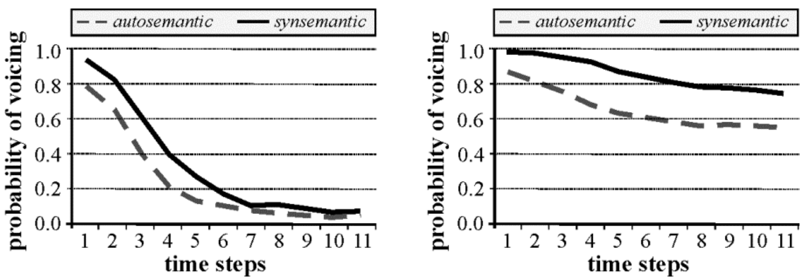

Figure 5: Voicing profiles for autosemantic vs. synsemantic words: **a.** voiced–voiceless assimilatory context and **b.** voiced–voiced assimilatory context.

The vd-vd context is of special interest: whereas speakers from Group A and B did not make any marked distinction between autosemantic and synsemantic words (Fig. 6a and 6b), speakers from Group C behaved quite differentially with respect to these classes (Fig. 6c). In their productions, the word-final obstruent in autosemantic words was devoiced in contrast to synsemantic words, in which it was fully voiced.

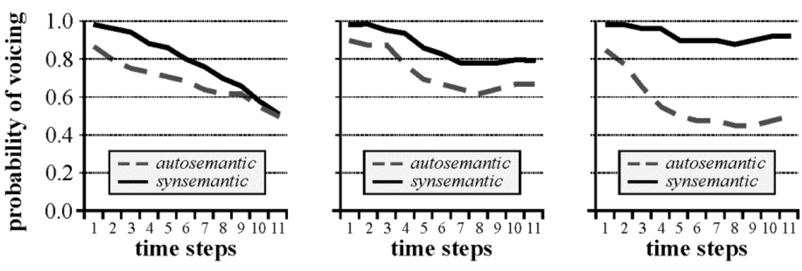

Figure 6: Voicing profiles for autosemantic vs. synsemantic words split into three groups of speakers (*voiced–voiced assimilatory context*): a. native-like English (group A); b. mild Czech accent (group B) and c. strong Czech accent (group C).

4. Discussion

The objective of this study was to examine the realization of voicing in Czech speakers of English with a differing degree of foreign accent. We focused on several types of sequences of two obstruents. First, we investigated the sequences of a voiced and a voiceless obstruent, such as *called Peter*. In native English, the word-final [d] tends to become partially devoiced, but does not turn into its voiceless (or, more precisely, fortis) counterpart, [kɔːld̥ piːtə]. In Czech, word-final voicing is neutralized, and we thus predicted a greater degree of devoicing in the speakers with a stronger Czech accent in their English and a lower degree of devoicing in the near-native-like group. Our results confirmed this hypothesis, with Group A producing the fewest voiceless items (i.e., with voicing present in less than 25 % of the sound's duration) and Group C the most (Figure 1). In English, some degree of devoicing of a final voiced obstruent is expected even if the following obstruent is phonologically voiced, such as *called Barry*. A similar tendency was observed in our most proficient speakers, who devoiced more than the other two groups (Figure 3). The results regarding the voiceless–voiced context (*back door*) are only tentative, given the low number of occurrences in our data, but they again show that speakers in Group A behave in accordance with expectations: most of their realizations are not fully voiced, which means that they did not produce the incorrect assimilation of voicing to the following voiced obstruent (Figure 2).

The data lend support to our division of speakers into three groups according to the degree of foreign accent, but we can still observe some idiosyncratic tendencies (Figure 4). That is to be expected, since the original evaluations are based on the listeners' overall impression in which the implementation of voicing presumably plays a marginal role.

Assimilation of voicing is typically not treated in much detail in the classic textbooks of English phonetics and phonology (see the Introduction). At the level of fine phonetic detail, however, the relationship between phonological and phonetic voicing in obstruent clusters across the word boundary may be more complex. From our cursory analyses, the phonetic voicing of the word-final obstruent does appear to partly depend, even in English, on the phonetic voicing of the following word-initial obstruent. In other words, the degree of voicing of the two sounds is not completely independent. It might therefore be useful, in our future research, to incorporate into the analyses the phonetic voicing of the following speech sound.

The more important line of our research will focus on the evaluation of wrong voicing assimilation by native listeners. While we do not expect this to be a phenomenon which should be stigmatizing, it may still play a role in the evaluation of a foreign-accented speaker (see, e.g., Gluszek & Dovidio, 2010).

Acknowledgement

The second author of this study was supported by the project GAČR P406/14-08084S, and by the Programme of Scientific Areas Development at Charles University in Prague (PRVOUK), subsection 10 – Linguistics: Social Group Variation.

References

Boersma, P. and Weenink, D. 2013. Praat: doing phonetics by computer (Version 5.3.57). Retrieved from http://www.praat.org

Butcher, A. 2004. “Fortis/lenis” revisited one more time: the aerodynamics of some oral stop contrasts in three continents. Clinical Linguistics & Phonetics 18: 547–557.

Cruttenden, A. 2008. Gimson’s pronunciation of English. London: Hodder Education.

Flege, J. E. 1987. The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. Journal of Phonetics 15: 47–65.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed) Speech perception and linguistic experience: issues in cross-language research. Baltimore: York Press: 233–277.

Gluszek, A. and 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: 214–237. DOI: 10.1177/1088868309359288

Hirose, H. and Gay, T. 1972. The activity of the intrinsic laryngeal muscles in voicing control: An electromyographic study. Phonetica 25: 140–164.

Jusczyk, P. W. 1993. From general to language-specific capacities: the WRAPSA model of how speech perception works. Journal of Phonetics 21: 3–28.

Kohler, K. 1984. Phonetic explanation in phonology: the feature fortis/lenis. Phonetica 41: 150–174.

Kuhl, P. K. and Iverson, P. 1995. Linguistic experience and the “Perceptual Magnet Effect.” In W. Strange (ed) Speech perception and linguistic experience: issues in cross-language research. Baltimore: York Press: 121–154.

Ladefoged, P. and Maddieson, I. 1996. The sounds of the world’s languages. Oxford: Blackwell Publishers.

Löfqvist, A. and Yoshioka, H. 1980. Laryngeal activity in Swedish obstruent clusters. Journal of the Acoustical Society of America 68: 792–801.

Machač, P. and Skarnitzl, R. 2009. Principles of Phonetic Segmentation. Praha: Epocha.

Möbius, B. 2004. Corpus-based investigations on the phonetics of consonant voicing. Folia Linguistica 38: 5–26. DOI: 10.1515/flin.2004.38.1-2.5

Ogden, R. 2009. An introduction to English phonetics. Edinburgh: Edinburgh University Press.

Ridouane, R., Fuchs, S. and Hoole, P. 2006. Laryngeal adjustments in the production of voiceless obstruent clusters in Berber. In J. Harrington and M. Tabain (eds) Speech production: models, phonetic processes, and techniques. New York: Psychology Press: 275–297.

Roach, P. 2009. English phonetics and phonology: A practical course. Cambridge: Cambridge University Press.

Shih, C. and Möbius, B. 1998. Contextual effects on voicing profiles of German and Mandarin consonants. Proceedings of ICSLP 1998: 3135–3138.

Skarnitzl, R. 2002. Pronunciation practice overview in current textbooks of English. ATE Newsletter 13: 45–51.

Skarnitzl, R. 2011. Znělostní kontrast nejen v češtině. Praha: Nakladatelství Epocha.

Skarnitzl, R. and Poesová, K. 2008. Typology of voicing changes in Czech English. In A. Grmelová, L. Dušková, M. Farrell and R. Pípalová (eds) Plurality and Diversity in English Studies – Proceedings from the Third Prague Conference on Linguistics and Literary Studies. Praha: Faculty of Education, Charles University in Prague: 8–17.

Skarnitzl, R., Volín, J. and Drenková, L. 2005. Tangibility of foreign accents in speech: The case of Czech English. In A. Grmelová, L. Dušková and M. Farrell (eds) Proceedings from the Second Prague Conference on Linguistics and Literary Studies. Praha: Faculty of Education, Charles University in Prague: 11–20.

Smith, C. L. 1997. The devoicing of /z/ in American English: effects of local and prosodic context. Journal of Phonetics 25: 471–500.

Yuan, J. and Liberman, M. 2008. Speaker identification on the SCOTUS corpus. Proceedings of Acoustics ’08.