PERCEPTION OF ALLOPHONIC CUES TO ENGLISH WORD BOUNDARIES BY POLISH LEARNERS: APPROXIMANT DEVOICING IN ENGLISH*

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

GEOFFREY SCHWARTZ
Adam Mickiewicz University in Poznań
geoff@wa.amu.edu.pl

ANNA BALAS
Adam Mickiewicz University in Poznań
abalas@wa.amu.edu.pl

Abstract
The study investigates the perception of devoicing of English /w, r, j, l/ after /p, t, k/ as a word-boundary cue by Polish listeners. Polish does not devoice sonorants following voiceless stops in word-initial positions. As a result, Polish learners are not made sensitive to sonorant devoicing as a segmentation cue. Higher-proficiency and lower-proficiency Polish learners of English participated in the task in which they recognised phrases such as buy train vs. bite rain or pie plot vs. pipe lot. The analysis of accuracy scores revealed that successful segmentation was only above chance level, indicating that sonorant voicing/devoicing cue was largely unattended to in identifying the boundary location. Moreover, higher proficiency did not lead to more successful segmentation. The analysis of reaction times showed an unclear pattern in which higher-proficiency listeners segmented the test phrases faster but not more accurately than lower-proficiency listeners. Finally, #CS sequences were recognised more accurately than C#S sequences, which was taken to suggest that the listeners may have had some limited knowledge that devoiced sonorants appear only in word-initial positions, but they treated voiced sonorants as equal candidates for word-final and word-initial positions.

Keywords: word boundary, devoicing, laryngeal coarticulation, allophonic cues

* The research reported in this paper 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'

1. Introduction

Unlike written language, continuous speech does not contain clearly-defined word boundaries (Lehiste 1972; Nakatani and Dukes 1977; Norris, McQueen, Cutler and Butterfield 1997). Rather, listeners are confronted with sequences of sounds which form meaningful units on a lexical level, but which are not separated by periods of silence to help the listener in their segmentation. Parsing strings of sounds into words relies on two sources of cues: phonetic-acoustic cues (Ainsworth 1986; Anderson and Port 1994; Boucher 1998; Christie 1977; Davidsen-Nielsen 1974; Lehiste 1960; Redford and Randall 2005), and higher-order cues such as word meaning, context and sentence structure (Cole, Jakimik and Cooper 1980; Kim, Stephens and Pitt 2012). However, the most effecting segmentation seems to be achieved when both sources are available (Mattys, White and Melhorn 2005; Norris et al. 1997). It has been suggested that the distribution of phonetic-acoustic cues in a given language is tracked by infants in the early stages of language acquisition to form a native-speech segmentation mechanism prior to the emergence of word meaning (Aslin, Saffran and Newport 1998; Brent and Cartwright 1996). Endress and Hauser (2010) suggest that languages carry universal prosodic cues that feed a universal mechanism for segmenting words across all languages. These cues are accessed by infants in the early acquisition of the ambient language, even before they are tuned to the language-specific properties. Although, in their study, adult English speakers effectively segmented words in unfamiliar languages such as French, Turkish and Hungarian when only prosodic cues to word boundaries were given, it is quite contrary to many reports that non-native segmentation is significantly compromised compared to the segmentation in the native language (e.g., Altenberg 2005; Cutler, Mehler, Norris and Segui 1992). It appears that, while there may be some language-independent word-boundary cues available, most segmentation relies on languages-specific cues. Such language-specific word-boundary information is attributed to differences in weighting and hierarchy of cues across languages (Cutler, Mehler, Norris and Segui 1986; Mehler, Dommergues, Frauenfelder and Segui 1981; Nakatani and Dukes 1977; Otake, Hatano, Cutler and Mehler 1993).

2. Phonetic-acoustic cues to word boundaries

Of many cues to word boundaries, temporal variability of segments seems to be the most extensively researched. Spanning a word boundary, both word-final (Beckman and Edwards 1990; Nakatani, O'Connor and Aston 1981; Oller 1973; Umeda 1975) and word-initial (Fougeron and Keating 1997 Oller 1973) sounds are lengthened. The results from studies investigating durational variability induced by the location of word boundaries are not, however, in complete

agreement as to the direction of lengthening. For example, Beckman and Edwards (1990) found shorter /pɒp/ and longer /ə/ in *poppe pose* compared to longer /pɒp/ and shorter /ə/ in *pop oppose*, which clearly points to word-final lengthening before the word boundary and no word-initial lengthening. On the other hand, Turk and Shattuck-Hufnagel (2000) reported significant word-initial lengthening and no word-final lengthening in *tune acquire* vs. *tuna choir*. Such differences may be attributed to the difficulty with disentangling boundary lengthening from polysyllabic shortening. For example, Turk and White (1999) found evidence for word-final lengthening in *shakedown stairs* vs. *shake downstairs*, in which *shake* was longer immediately before the word boundary, however, as they admit, such lengthening may as well be consistent with polysyllabic shortening, since *shake* in *shakedown* is both separated from the word boundary and simultaneously followed by another syllable. Despite the fact that the direction of word-boundary durational variability is not unequivocally established and the fact that speakers may indeed differ in how they signal word boundaries (Quené 1992; Shatzman and McQueen 2006), listeners seem to be quite efficient in using sound duration to segment words. Gow and Gordon (1995) showed that the word *lips* serves as an effective prime only if it is separated by a word boundary as in *two lips*, but not if it is a part of one word as in *tulips*. They concluded that most likely the participants may have attended to the duration of /l/, which was longer following the word boundary in *two lips* than word-internally in *tulips*. Moreover, different durations of /l/ resulted from different stress patterns in *two lips* and *tulips*. Smith and Hawkins (2012) observed significant individual variation of duration in signaling word boundaries in phrases such as *so she diced them* vs. *so she'd iced them*. A subsequent perception experiment in noise revealed that listeners coped with such variation if they were familiarized with a particular voice. It suggests that the perceptual mechanism is quite flexible in accommodating to individual strategies used by speakers in signaling word boundaries.

Apart from differences in duration, segments at word boundaries are also characterized by occurrence of allophonic realizations. More precisely, sounds directly adjacent to the word boundary have stronger articulation and greater spectral distinctiveness. For example, lip movements for /m/ are slower when preceded by a word boundary (Byrd and Saltzman 1988), and /l/ has a greater magnitude of velarization when followed by the word boundary (Umeda and Coker 1975). Segments separated by word boundaries are more immune to coarticulation. Krakow (1993) found that vowel nasalization induced by a following nasal consonant was significantly reduced when the two sounds were separated by a word boundary. Cole, Kim, Choi and Hasegawa-Johnson (2007) reported longer VOTs for word-initial than word-medial stops and this cue has been found to be attended to by native speakers of English in segmenting words (Altenberg 2005; Nakatani and Dukes 1977). All these studies point to the fact that, despite the frequent lack of observable points in the acoustic signal that

would indicate the division of sound sequences into words, the onset and offset of words is signaled by fine-grained spectral and temporal properties of sounds that listeners are sensitive to in speech segmentation. Another type of word-boundary strengthening is glottalization of word-initial vowels. Glottalization, unlike complete linking of sounds across the word boundary, appears to be a strong word-integrity marker. The actual frequency of glottalization may differ between languages. For example, the occurrence of glottalization in English is less widespread than in Polish (Schwartz 2013; Schwartz, Balas and Rojczyk 2014; Umeda 1978). Other factors that influence the incidence of glottalization may range from segmental, lexical, prosodic to sociolinguistic (Garellek 2014, Pompino-Marschall and Żygis 2010; Szczepek-Reed 2014). Acoustically glottalization is a strong candidate for an effective cue in speech segmentation, however the results from perception experiments do not clearly indicate that glottalization facilitates word segmentation as strongly as may be expected. In one of the studies, Bissiri, Lecumberri, Cooke and Volin (2011) compared word monitoring in English by Czech listeners, whose native language uses relatively frequent glottalization, with English and Spanish listeners, whose languages are characterized by across-boundary linking. Although the general pattern showed that glottalized words were recognized faster, the recognition of non-glottalized items was only slower by 11%, 9%, and 6% by British, Czech and Spanish listeners respectively. Moreover, the results revealed that these were English and not Czech listeners that benefited from glottalization more as indicated by shorter RTs, which runs counter to the expectation that experience with a feature in L1 leads to global greater sensitivity to this feature. In another word-monitoring task, Schwartz, Rojczyk and Balas (2015) had Polish listeners recognize target words in glottalized and non-glottalized sequences in English. The assumption was that, because Polish listeners are accustomed to word-initial glottalization as a word-boundary marker in their L1, they would respond faster to glottalized than non-glottalized items. The results revealed that, although glottalization did accelerate the word recognition, its effect was not as strong as expected, with RTs of 398 ms for glottalized and 411 ms for non-glottalized items. The authors interpreted this relative insensitivity to glottalization demonstrated by Polish listeners in terms of 'desensitization' (Bohn 1995) in L2 perception to the cue which is present in L1

3. The current study

In this study we contribute to the discussion on word segmentation in non-native speech by investigating how Polish speakers of English perceive English sonorant devoicing after /p, t, k/ as a word-boundary cue in sequences such as *buy train* vs. *bite rain*. Such allophonic devoicing is an example of laryngeal coarticulation (Ohala 1993) in English and may be interpreted as a by-product of

long-lag VOTs for English voiceless stops. In English sequences /p, t, k/ followed by /r, l, w, j/ the sonorants absorb the time interval between the stop release and the onset of voicing, which results in their widespread or even complete devoicing. Such allophonic devoicing is a logical word-boundary cue in those sequences for native speakers of English, because it only operates when a plosive and a sonorant constitute an onset cluster (Docherty 1992; Lehiste 1960). When they are separated by a word boundary, the laryngeal coarticulation of the plosive and the sonorant is blocked and the sonorant is fully voiced. Illustrating this regularity, in *buy train* /r/ is substantially devoiced, because it is preceded by voiceless /t/ in an onset cluster, while /r/ in *bite rain* is voiced because voiceless /t/ is separated from /r/ by a word boundary. Additionally, if there is glottal reinforcement of /t/, the glottis is closed, which blocks coarticulation. Contrary to English, Polish does not devoice sonorants following voiceless stops in onset clusters (e.g., *trawa* 'grass'), the only environment authorizing sonorant devoicing is in coda clusters (e.g., *wiatr* 'wind') (Gussman 1992). Figure 1 shows the spectrogram of the Polish word *plan* 'plan' and the English word *plan*. The sonorant /l/ in the onset /pl/ is fully voiced in Polish, but it is largely devoiced in English.

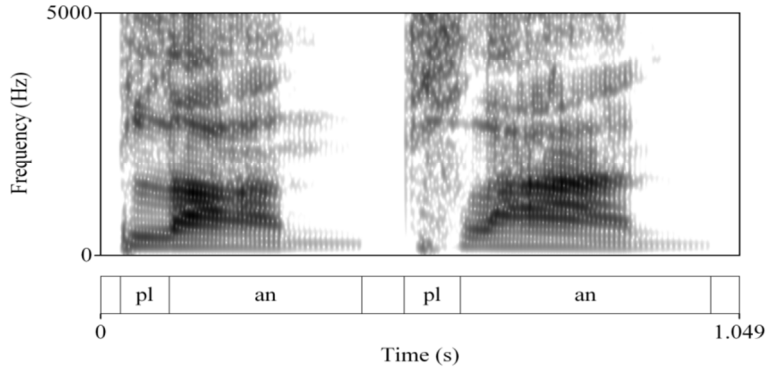

Figure 1. The sonorant /l/ in the onset sequence /pl/ in Polish *plan* 'plan' (left) and English *plan* (right)

The lack of laryngeal coarticulation between a voiceless plosive and the following sonorant is one of the features of Polish-accented English. Although we do not know any studies showing how the absence of laryngeal coarticulation interferes with intelligibility, it is a noticeable feature in the pronunciation of less proficient learners of English. In teaching English pronunciation to Polish learners, it is linked with training the production of long-lag VOT values for voiceless stops (Porzuczek, Rojczyk and Arabski 2013). As a result, Polish learners are faced with a two-stage learning task in order to use sonorant devoicing as a word-boundary cue. First, they must learn this laryngeal coarticulation and then they must use this knowledge in word segmentation.

Whether this feature can be learnt to serve as a boundary cue may be ascertained by comparing segmentation performance measured by accuracy scores and RTs by learners at different proficiency levels. More proficient learners are predicted to be more accurate and faster in their segmentation than less proficient learners.

Previous research on non-native word segmentation from voicing cues in English concentrated on voiceless-stop aspiration as a boundary cue. Altenberg (2005) tested Spanish listeners' segmentation of sequences such as *keeps parking* vs. *keep sparking*, where the word-boundary cue is the aspiration of word-initial /p/ in *parking* and the absence of aspiration in *sparking*. The results indicated that Spanish learners performed significantly worse than native speakers of English in exploiting aspiration as a boundary marker. The mean percentage of correct segmentation was 58.5% for Spanish and 96.7% for native English listeners. Ito and Strange (2009) tested the segmentation of similar sequences in English by Japanese listeners. Similar to Spanish, Japanese exploits only short-lag values for voiceless stops. The performance of Japanese listeners was better than that of Spanish listeners with the mean correct segmentation of 73.1%. There was also a positive correlation between language experience and performance, which was taken to suggest that sensitivity to aspiration cues is a learning process extended in time. Similar limited sensitivity to aspiration as a word-boundary cue was found by Shoemaker (2014) with French learners of English. As in the studies with Spanish and Japanese listeners, the French listeners were more sensitive to word-boundary glottalization than aspiration in their segmentation decisions.

3.1. Participants

A total of sixty-nine listeners participated in the experiment. There were 37 females and 32 males with the mean age of 21.7 years. They were students of English at the Institute of English, University of Silesia. The participants were recruited from two groups of different proficiency. The first lower-proficiency group included 43 first-year students, 20 females and 23 males, with 19.9 years of mean age. Their proficiency level in English ranged from B1 to B2 in the Common European Framework of Reference for Languages (CERFL). The second higher-proficiency group included 26 fifth-year students, 17 females and 9 males, with the mean age of 23.5 years. Their proficiency level in English ranged from C1 to C2 in CERFL. None of the participants reported any speech or hearing disorders. All participants had normal or corrected-to-normal vision. None of the participants had any observable manual disorders.

3.2. Stimulus materials

The stimulus materials were six pairs of sequences differing in the location of word boundary.

C#S	#CS
bite rain	*buy train*
float weed	*flow tweed*
wake lock	*way clock*
sake west	*say quest*
pipe lot	*pie plot*
rope ride	*row pride*

The targets words were all familiar to the learners, as shown by a follow-up short questionnaire. A female native speaker of American English recorded the test stimuli in a carrier phrase 'Say__again'. The speaker was instructed to speak naturally, however not too fast. She was also instructed to avoid emphasising the word boundary by inserting pauses or glottalization. The recording took place in a sound-proof booth in the Acoustic-Phonetic Laboratory, Institute of English, University of Silesia. The signal was captured at 44100 Hz (24 bit quantization) through a headset dynamic Sennheiser HMD 26 microphone fed by a USBPre2 (Sound Devices) amplifier. Together with familiarization items and fillers, the speaker recorded 26 pairs of sequences.

The recordings were analysed acoustically from waveform and spectrogram in Praat (Boersma 2001). No traces of glottalization or periods of silence marking word boundaries were detected. Figures 2 and 3 show the phrases *row pride* and *rope ride*. The sonorant /r/ following /p/ in *pride* is voiceless almost in its full portion, while /r/ in *ride* after a word boundary is voiced.

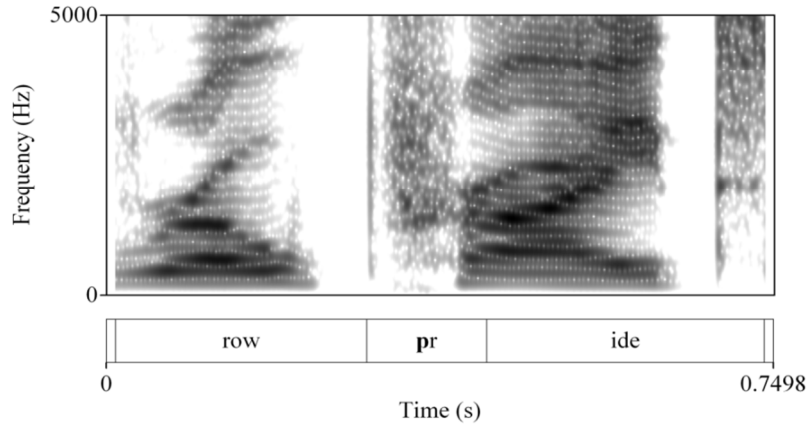

Figure 2. Devoiced /r/ in *row pride*

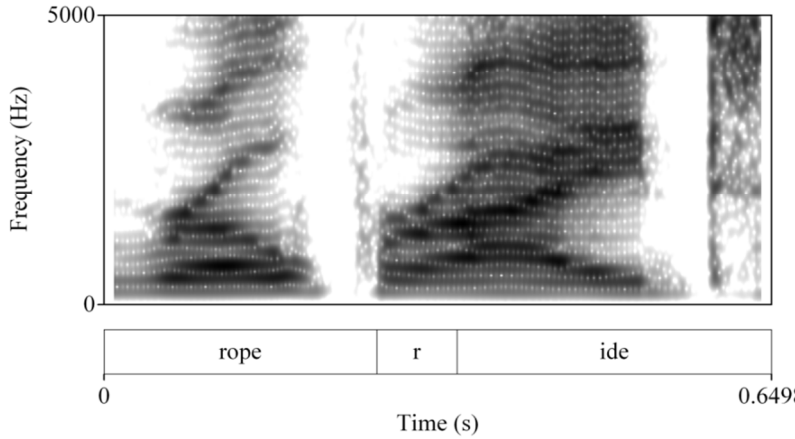

Figure 3. Voiced /r/ in *rope ride*

Moreover, since English has a tendency for glottal reinforcement of the word-final stops, especially voiceless alveolar /t/, the stimuli were inspected for any manifestation of glottal reinforcement. No instances of glottal reinforcement were found. As noted by a reviewer, another cue that may have distinguished the test phrases was vowel/diphthong duration. In the C#S sequences vowel/diphthong duration is expected to decrease as a result of pre-fortis clipping. The measurements of vocalic portions in the test pairs revealed that indeed vowels and diphthongs in the C#S sequences were shorter (M=140.3; SD=20.4) than in the #CS sequences (M=158.3: SD=24.2) [t(6)=4.44, p<.01].

Two native speakers of American English and one native speaker of British English verified the validity of the recorded test phrases. Four pairs were judged as unambiguous as to the location of word boundary, whereas two pairs were reported to be relatively ambiguous. The acoustic analysis of those two pairs revealed that the speaker had produced relatively long release bursts in *bite rain* and *float weed*, which may have been perceived as partial devoicing of the following sonorant. Accordingly, the release bursts in those phrases were reduced by removing their portion from the acoustic signal. Such manipulations were found to reduce the ambiguity reported earlier. Finally, all phrases were normalized for intensity at 70 dB for the experiment.

3.3. Procedure

The participants were tested individually in the Acoustic-Phonetic Laboratory, Institute of English, University of Silesia. The experiment was run by E-Prime and accuracy scores and RTs were collected using the Serial Response Box (Psychology Software Tools). The audio stimuli were fed by Philips SBC HP840

headphones at a comfortable listening level. The participants were seated in front of the 17-inch monitor and were told to put their left-hand and right-hand index fingers on the flashed left and right buttons on the response box. They were instructed that first they would see a pair of two phrases, e.g., *bite rain* on the left of the screen and *buy train* on the right of the screen and next they would hear one of the two phrases spoken through headphones. Their task was to press the left button or the right button to indicate which phrase they had just heard. They were told that their RTs were collected and were encouraged to press the button as soon as they had made the decision. Each cycle of stimulus presentation had the following structure. First, the get-ready screen was displayed for 2500 ms with 'get ready' in the centre. Next, the pair of two phrases was displayed for 3000 ms with e.g, *buy train* aligned to the left of the screen and *bite rain* aligned to the right of the screen. This time was estimated to be sufficient for the participants to read the choice phrases and remember their location. After 3000 ms the audio stimulus was played through headphones and the listeners indicated by pressing the button which phrase they had heard. The time limit for making a decision was 4000 ms. After that, another trial began. The experiment started with a welcome screen and the participants proceeded to the familiarization with 6 phrases not included in the analysis. After the familiarization phase, another screen informed the participants that the experiment proper would begin. No feedback was provided informing the listeners if their choice was correct or not. All presentations in the experimental part were randomized for each individual listener. In order to balance hand preference in measuring RTs, two phrases of the same pair were presented on one side of the screen. For example, if *buy train* was left-aligned and required pressing a button with a left hand in the trial with the recording *buy train*, so *bite rain* would also be left-aligned in the trial with the recording *bite rain*. Similarly, if *wake lock* was right-aligned with the target *wake lock*, so *way clock* would also be right-aligned with the target *way clock*. It guaranteed that the RTs were only the result of differences in word segmentation and not of differences in hand preference. Each listener was presented with a 6 training phrases, 12 test phrases and 84 filler phrases. The filler phrases were analogical sequences with two word-boundary locations, but containing different cues (e.g., *Lou spills* vs. *loose pills* or *I learn* vs. *I'll earn*). Each session lasted approximately 15 minutes.

4. Analysis and results

The correct segmentation was calculated as the rate of correctly recognized test phrases to the total number of presented test phrases. RTs were calculated only for correct segmentation from the onset of the audio stimulus to the registered response. The RTs were trimmed for values lower than 500 ms and higher than

2500 ms to remove spurious responses and outliers. According to these criteria a total of 38 trials were discarded. Table 1 presents accuracy scores and RTs for each phrase broken by proficiency (lower proficiency for first-year students and upper proficiency for fifth-year students).

Table 1. Mean accuracy scores and RTs for the test items in each group

	Accuracy in %		RTs in ms (SD)	
	lower proficiency	higher proficiency	lower proficiency	higher proficiency
pipe lot	44.2	73.1	980 (480)	1149 (400)
pie plot	60.5	57.7	1292 (439)	1143 (500)
wake lock	69.8	46.2	1212 (375)	1176 (500)
way clock	76.7	84.6	1330 (478)	1053 (416)
rope ride	51.2	38.5	1207 (501)	1139 (417)
row pride	62.8	50.0	1195 (357)	1220 (406)
float weed	67.4	69.2	1198 (384)	1168 (343)
flow tweed	60.5	76.9	1239 (376)	1270 (433)
bite rain	58.1	61.5	1143 (446)	1085 (312)
buy train	79.1	76.9	1271 (478)	1037 (356)
sake west	48.8	46.2	1183 (378)	1179 (481)
say quest	79.1	73.1	1328 (439)	1183 (503)
TOTAL MEAN	63.2	62.8	1228 (430)	1149 (417)

The overall accuracy score for both groups was 63%. There was no difference in accuracy between the lower-proficiency (63.2%) and higher-proficiency (62.8%) group, which indicates that more proficient learners were not more sensitive to sonorant devoicing as a word-boundary cue. The comparison of the boundary location (C#S e.g., *bite rain* vs. #CS e.g., *buy train*) revealed that the #CS sequences were recognized more accurately (69.8%) than C#S sequences (56.3%) [$\chi^2(1)=16.23$, $p<.01$]. The analysis broken by proficiency reflected the same pattern of more accurate segmentation of #CS sequences for each group. Lower-proficiency learners segmented the #CS sequences more accurately (69.8%) than C#S sequences (56.6%) [$\chi^2(1)=9.63$, $p<.01$]. Similarly, higher-proficiency students performed better with the #CS sequences (69.9%) than with C#S sequences (55.8%) [$\chi^2=6.64$, $p<.05$].

The RTs were analysed in a two-way mixed ANOVA with an independent between-subject variable of proficiency (low-proficiency/high-proficiency) and an independent repeated-measures variable of a sequence type (#CS/C#S). There was a main effect of proficiency on RTs [$F(1, 482)=3.9$, $p<.05$], caused by shorter reaction times in the higher-proficiency group (M=1149; SE=31.6) than in the lower-proficiency group (M=1228; SE=24.4). There was also a main effect of a sequence type [$F(1, 202)=6.6$, $p<.05$], indicating that the C#S sequences (M=1149; SE=28.2) were identified quicker than the corresponding #CS sequences (M=1259; SE=30.6). The analysis of interaction between proficiency and sequence type showed that the significantly longer RTs for the

#CS sequences were contributed to the lower-proficiency group. While the higher-proficiency groups did not differ in their RTs for the C#S sequences (*M*=1153; *SE*=47.1) compared to #CS sequences (*M*=1096; *SE*=46.2) [F(1, 74)=.69, p>.05], in the lower-proficiency group the RTs for the #CS sequences were significantly longer (*M*=1293; *SE*=39.3) than for the C#S sequences (*M*=1164; *SE*=38.3) [F(1, 124)=5.66, p<.05].

5. Discussion

The study had the following objectives: (1) investigate the accuracy of word segmentation from the cue of sonorant devoicing after voiceless stops; (2) analyse reaction times in correct segmentations; (3) compare the performance of lower- and higher-proficiency learners of English. The accuracy scores revealed that all tested items were recognized slightly above a chance level. Higher proficiency did not contribute to more accurate recognition. Both groups performed better with the #CS sequences (*buy train*) than with the C#S sequences (*bite rain*). The analysis of RTs showed that, although the higher-proficiency listeners were not more accurate, they were faster in their correct identifications. Moreover, despite the fact that C#S sequences were identified less accurately, they were identified faster. The between-group analysis revealed that overall faster RTs for the C#S identifications were caused by slower RTs for the #CS sequences in the lower-proficiency group. In other words, while the higher-proficiency listeners reacted with the same speed to both types of the sequences, lower-proficiency listeners provided significantly slower RTs for the #CS sequences, which resulted in overall longer RTs for this sequence. In the following discussion, we will try to provide interpretation of this complicated pattern.

The overall accuracy of 63% suggests that the listeners were not able to perceive sonorant devoicing as a word-boundary cue. This result for Polish learners is very similar to the one obtained for Spanish learners (58.5%) by Altenberg (2005) and slightly lower than for Japanese learners (Ito and Strange 2009). It adds to the previous results that non-native speakers are outperformed by native speakers, because non-native speakers transfer segmentation strategies from their L1 into L2 (Otake et al. 1993; Cutler et al. 1986; Mehler et al. 1981; Otake et al. 1993). The reason for poorer performance by our listeners may have two interpretations. They were not sensitive to sonorant devoicing as a segmentation cue in that they did not learn to associate devoicing as aligning the sonorant to the preceding stop in an onset cluster. Another equally valid interpretation is that they did not learn to associate voiced sonorants as indicating that a preceding stop is separated from the sonorant by a word boundary. The results indicating that the #CS sequences were segmented more effectively than the C#S sequences points to the latter interpretation. Poorer

performance with the C#S sequences may mean that the listeners treated voiced sonorants as equally likely to be preceded by a stop or a word boundary. Since sonorants in Polish are voiced irrespective of whether they are word-initial or follow a word-initial voiceless stop, they do not serve as word boundary cues. As a result Polish listeners treated English voiced sonorants as uninformative as to the word boundary location. In other words, while the listeners may have noticed, to a limited extent, that sonorant devoicing cues the #SC sequences, they did not notice that voiced sonorants consistently cue the S#C sequences. As noted by a reviewer, another possibility is that the listeners attended to the differences in vowel/diphthong duration caused by pre-fortis clipping in the C#S sequences. Although we cannot ignore this suggestion, we believe that the mean difference in the vocal portions of only 18 ms is not sufficient to account for better segmentation of the #CS sequences. Firstly, sonorant devoicing seems to be a much more robust perceptual cue compared to a temporal cue with the mean difference of only 18 ms and standard deviations of 20 ms and 24 ms. Secondly, a previous study showed that Polish learners, even of high proficiency, are relatively insensitive to vowel duration as a voicing cue in English (Rojczyk 2010).

The interpretation of collected RTs is less clear. The higher-proficiency listeners were significantly faster in their correct segmentations than the lower-proficiency listeners despite the fact that they were not more accurate. In other words, the higher-proficiency listeners seemed to be more decided in their responses, even though they were not more correct. Such a result is difficult to interpret if it is not followed by correlated accuracy. On the one hand, it may suggest that higher-proficiency learners were more confident about their segmentation strategies despite the fact that those segmentation strategies turned out to generate incorrect decisions. On the other hand, it may suggest that higher-proficiency listeners were generally more experienced in processing spoken English and more confident in reacting to English stimuli. If this was the case, the tested cue had limited impact on their reaction times in the experiment. The analysis of RTs for the sequence type also demonstrates a pattern that escapes clear interpretations. The C#S sequences were recognised faster than the #CS sequences, but, at the same time, they were recognised less accurately. However, the by-group analysis revealed that this difference was mainly contributed to by the lower-proficiency group. It may additionally suggest that the lower-proficiency listeners were less confident about their responses than the higher-proficiency listeners. Nevertheless, as emphasised earlier, any definitive conclusions are precluded by the similarly low accuracy scores in both groups. The RTs obtained in the current study cannot be compared to the studies by Altenberg (2005) and Ito and Strange (2009), because the RTs were not collected in these studies.

The following conclusions summarize the objectives of the study:

1. sonorant voicing/devoicing in English is not perceived effectively as a boundary cue by Polish listeners
2. higher proficiency does not lead to more successful segmentation
3. sequences with devoiced word-initial sonorant are recognized more successfully than sequences with word-final voiced sonorant

References

Ainsworth, W. 1986. Pitch change as a cue to syllabification. Journal of Phonetics 14. 257-264.

Altenberg, E. P. 2005. The perception of word boundaries in second language. *Second Language Research* 21(4): 325-358.

Anderson, S. and R. Port. 1994. Evidence for syllable structure, stress and juncture from segmental durations. Journal of Phonetics 22. 283-315.

Aslin, R. N., Jenny R. Saffran, and Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. Psychological Science 9. 321-324.

Beckman, M. E. and J. Edwards. 1990. Lengthening and shortening and the nature of prosodic constituency. In J. Kingston and M.. E. Beckman (eds.), *Laboratory Phonology I,* 152-178. Cambridge: Cambridge University Press.

Bissiri, Maria. P., Maria L. Lecumberri, Martin Cooke and Jan Volín. 2011. The role of word-initial glottal stops in recognizing English words. Proceedings of Interspeech 2011. Florence, Italy.

Boersma, P. 2001. Praat, a system for doing phonetics by computer. Glot International 10. 341-345.

Bohn, O.-S. 1995. Cross-language speech perception in adults – first language transfer does not tell it all. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 275-300. Baltimore: York Press.

Boucher, Victor J. 1998. A parameter for syllabification for VstopV and relative timing invariance. Journal of Phonetics 16. 299-326.

Brent, M. R. and T. A. Cartwright. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. Cognition 61. 93-125.

Byrd, D. and E. Saltzman. 1998. Intergestural dynamics of multiple prosodic boundaries. Journal of Phonetics 26. 173-199.

Christie, W. M. 1977. Some multiple cues for juncture in English. General Linguistics 17. 212-222.

Cole, J., H. Kim, H. Choi, and M. Hasegawa-Johnson. 2007. Prosodic effects on acoustic cues to stop voicing and place of articulation. Evidence from Radio News speech. Journal of Phonetics 35. 180-209.

Cole, R. A., J. Jakimik, and W. E. Cooper. 1980. Segmenting speech into words. Journal of the Acoustical Society of America 64. 1323-1332.

Cutler, A., J. Mehler, D. Norris and J. Segui. 1986. The syllable's differing role in the segmentation of French and English. Journal of Memory and Language 25(4). 385-400.

Cutler, A., J. Mehler, D. Norris and J. Segui. 1992. The nonlingual nature of speech segmentation by bilinguals. Cognitive Psychology 24. 381-410.

Davidsen-Nielsen, N. 1974. Syllabification in English words with medial sp, st, sk. Journal of Phonetics 2. 15-45.

Docherty, G. J. 1992. *The timing of voicing in British English obstruents*. New York: Foris Publications.

Endress, A. D. and M. D. Hauser. 2010. Word segmentation with universal prosodic cues. Cognitive Psychology 61. 177-199.

Fougeron, C. and P. Keating. 1997. Articulatory strengthening at edges of prosodic domain. Journal of the Acoustical Society of America 101. 3728-3740.

Garellek, M. 2014. Voice quality strengthening and glottalization. Journal of Phonetics 45. 106-113.

Gow, D. W. and P. C. Gordon. 1995. Lexical and prelexical influences on word segmentation: Evidence from priming. Journal of Experimental Psychology: Human Perception and Performance 21. 344-359.

Gussman, E. 1992. Resyllabification and delinking. The case of Polish voicing. Linguistic Inquiry 23. 29-56.

Ito, K. and W. Strange. 2009. Perception of allophonic cues to English word boundaries by Japanese second language learners of English. Journal of the Acoustical Society of America 125(4). 2348-2360.

Kim, D., J. D. Stephens and M. A. Pitt. 2012. How does context play *a part* in splitting words *apart*? Production and perception of word boundaries in casual speech. Journal of Memory and Language 66. 509-529.

Krakow, R. A. 1993. Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In Stephen R. Anderson, Marie. K. Huffman, Rena. A. Krakow and Patricia A. Keating. (eds.), *Phonetics and phonology: Nasals, nasalization, and the velum,* 87-116. San Diego: Academic Press.

Lehiste, I. 1960. An acoustic-phonetic study of internal open juncture. Phonetica 5. 1-54.

Lehiste, I. 1972. The timing of utterances and linguistic boundaries. Journal of the Acoustical Society of America 51. 2018-2024.

Mattys, S. L., L. White, and J. F. Melhorn. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. Journal of Experimental Psychology: General 134. 477-500.

Mehler, J., J. Y. Dommergues, U. Frauenfelder and J. Segui. 1981. The syllable's role in speech segmentation. Journal of Verbal Learning and Verbal Behavior 20(3). 298-305.

Nakatani, L. and K. D. Dukes. 1977. Locus of segmental cues for word juncture. Journal of the Acoustical Society of America 62. 714-719.

Nakatani, L., K. D. O'Connor and C. H. Aston. 1981. Prosodic aspects of American English speech rhythm. Phonetica 38. 84-106.

Norris, D., J. M. McQueen, A. Cutler and S. Butterfield. 1997. The possible word-constraint in the segmentation of continuous speech. Cognitive Psychology 34. 191-243.

Kimbrough, O. D. 1973. The effect of position in utterance on speech segment duration in English. Journal of the Acoustical Society of America 54. 1235-1247.

Quené, H. 1992. Durational cues for word segmentation in Dutch. Journal of Phonetics 20. 331-350.

Ohala, J. J. 1993. Coarticulation and phonology. Language and Speech 36. 155-170.

Otake, T., G. Hatano, A. Cutler and J. Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. Journal of Memory and Language 32(2). 258-278.

Pompino-Marschall, B. and M. Żygis. 2010. Glottal marking of vowel-initial words in German. ZAS Papers in Linguistics 52. 1-17.

Porzuczek, A., A. Rojczyk and J. Arabski. 2013. *Praktyczny kurs wymowy angielskiej dla Polaków*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Redford, M. A. and P. Randall. 2005. The role of juncture cues and phonological knowledge in English syllabification judgments. Journal of Phonetics 33. 27-46.

Rojczyk, A. 2010. Preceding vowel duration as a cue to the consonant voicing contrast: Perception experiments with Polish-English bilinguals. In E. Waniek-Klimczak (ed.), *Issues in accents of English: Variability and norm,* 341-360. Newcastle upon Tyne: Cambridge Scholars Publishing.

Schwartz, G. 2013. Vowel hiatus at Polish word boundaries – phonetic realization and phonological implications. Poznań Studies in Contemporary Linguistics 49(4). 557-585.

Schwartz, G., A. Balas and A. Rojczyk. 2014. External sandhi in L2 segmental phonetics – final (de)voicing in Polish English. Concordia Working Papers in Applied Linguistics 5. 637-649.

Schwartz, G., A. Rojczyk and A. Balas. 2015. Monitoring English *sandhi* linking – a study of Polish listeners' L2 perception. Research in Language 13(1). 61-76.

Shatzman, K. B., and J. M. McQueen. 2006. Segment duration as a cue to word boundaries in spoken-word recognition. Perception and Psychophysics 68(1). 1-16.

Shoemaker, E. 2014. The exploitation of subphonemic acoustic detail in L2 speech segmentation. Studies in Second Language Acquisition 36(4). 709-731.

Smith, R. H. and S. Hawkins. 2012. Production and perception of speaker-specific phonetic detail at word boundaries. Journal of Phonetics 40. 213-233.

Szczepek-Reed, B. 2014. Phonetic practices for action formation: Glottalization versus linking of TCU-initial vowels in German. Journal of Pragmatics 62. 13-29.

Turk, A. E., and S. Shattuck-Hufnagel. 2000. Word-boundary-related duraton patterns in English. Journal of Phonetics 28. 397-440.

Turk, A. E., and L. White. 1999. Structural influences on accentual lengthening in American English. Journal of Phonetics 27. 171-206.

Umeda, N. 1975. Vowel duration in American English. Journal of the Acoustical Society of America 58. 434-445.

Umeda, N. 1978. Occurance of glottal stops in fluent speech. Journal of the Acoustical Society of America 64. 88-94.

Umeda, N. and C. H. Coker. 1975. Subphonemic detail in American English. In G. Fant and M. A. A. Tatham (eds.), *Auditory analysis and perception,* 539-564. London: Academic Press.