Effects of listener factors and stimulus properties on the intelligibility, comprehensibility and accentedness of L2 speech

Abstract

The intelligibility, comprehensibility and foreign accentedness of native and Polish-accented English sentences were evaluated by six Polish, six Spanish and six English speakers. The non-native data were also analyzed for segmental and word stress errors. Results indicated that the three measures were partially independent of one another, supporting earlier findings that accented speech can be intelligible and comprehensible. An interlanguage speech intelligibility detriment was observed for Spanish listeners, but no clear evidence of an interlanguage speech intelligibility benefit was found, as non-native listeners never outperformed native listeners. The number of segmental errors, rather than lexical stress errors, was found to correlate with comprehensibility and accentedness ratings of non-native speech, but not with intelligibility scores. In general, the results point to a greater effect of stimulus properties than of listeners' L1s in the perception of non-native speech.

Keywords: intelligibility, comprehensibility, foreign accent, listener factors, segmental errors, word-stress errors.

1. Introduction

Foreign language or second language (L2) speech is typically accented, particularly when L2 learning begins after childhood. However, it is generally accepted that comfortably intelligible pronunciation, rather than native-like pronunciation, is the realistic and desirable goal of L2 speech learning (e.g., Abercrombie, 1949; Derwing & Munro, 2005; Morley, 1991; Munro, 2013; Munro & Derwing, 1999). This view is supported by empirical studies showing that foreign accent and intelligibility are partially independent measures (e.g., Munro & Derwing, 1995). Still, intelligibility and accentedness may vary depending on a number of factors including the L1 of the interlocutor (Bent & Bradlow, 2003; Stibbard & Lee, 2006); the level of L2 proficiency of the listeners and/or the talkers (Hayes-Harb, Smith, Bent, & Bradlow, 2008); lexical neighbourhood density (Imai, Walley, & Flege, 2005); familiarity with a given topic, speaker or accent (Gass & Varonis, 1984; Munro & Derwing, 1994); and the type and number of pronunciation errors (Caspers & Horloza, 2012; Kang, 2010; Munro & Derwing, 1999). This study evaluated further the influence of some of these factors, namely listener factors and stimulus properties, on the intelligibility, comprehensibility and foreign accentedness of Polish speakers' L2 English production by native and non-native speakers.

1.1. Intelligibility, comprehensibility and accentedness

Foreign accentedness can be defined as the listener's perception of how closely the pronunciation of an L2 speaker mirrors the pronunciation of a native speaker of a given language, whereas intelligibility refers to the degree to which a speaker's production is actually understood by a listener, and comprehensibility is concerned with the listeners' impressionistic observation of how easy it is to understand the L2 speech utterance (Kennedy & Trofimovich, 2008; Major, 2013; Munro, 2008). A number of studies have shown that speaking with a foreign accent does not necessarily affect the intelligibility of an utterance (e.g., Derwing & Munro, 1997; Munro & Derwing, 1999; Smiljanić & Bradlow, 2011). For instance, in a study that evaluated the accentedness, comprehensibility and intelligibility of semi-spontaneous L2 speech samples as perceived by native English speakers, Munro and Derwing (1999) observed stronger correlations between measures of intelligibility and comprehensibility than between measures of intelligibility and accentedness. Similarly, Smiljanić and Bradlow (2011) observed that judgments of accentedness remained constant while the degree of intelligibility changed under different noise conditions and was influenced by speaking styles.

1.2. Perception of L2 speech by native and non-native speakers

Interactions between non-native speakers are as frequent as communication between native and non-native speakers, particularly for widespread languages like English (e.g., Jenkins, 2002; Pickering, 2006). This fact has led researchers to explore these measures from the perspective of both native and non-native speakers of the target language. Bent and Bradlow

(2003) evaluated the intelligibility of one native and four non-native speakers of English by native and non-native listeners. Intelligibility was measured as the number of correctly transcribed keywords from short-sentence stimuli presented in noise.

The results showed that for non-native listeners, a high-proficiency non-native talker (and in one case a low proficiency talker) from the same native language background were as intelligible as a native talker. This was interpreted as an interlanguage speech intelligibility benefit for talkers (ISIB-T). Interestingly, this benefit was also observed with non-native listeners who did not share the L1 with the talkers (mismatched ISIB). Thus, a group of Chinese listeners, a group of Korean listeners and a mixed group of listeners from different L1 backgrounds found the short sentences produced by Korean and Chinese speakers of L2 English equally or more intelligible than those produced by native English talkers.

The idea of an ISIB was supported by a few other studies (Hayes-Harb et al., 2008; van Wijngaarden, 2001; van Wijngaarden, Steeneken, & Houtgast, 2002, among others). Hayes-Harb et al. (2008) found that Mandarin-speaking listeners performed better with Mandarin accented English than native English listeners did when both the non-native

listeners and talkers were low-proficiency, supporting an interlanguage speech intelligibility benefit for listeners (ISIB-L). In this case, intelligibility was measured as

correct identification of word-final consonant voicing in a word identification task. Using a similar measure, Xie and Fowler (2013) also reported an ISIB-L: Mandarin-speaking listeners outperformed English-speaking listeners in their identification of English final stops produced by a Chinese talker. An ISIB-T was also observed, but only for a group of Mandarin-speaking listeners in Beijing, not for a group of Mandarin-speaking listeners living in the US.

Other studies, however, have found no support for an ISIB. In a study by Munro, Derwing and Morton (2006), speakers of English, Japanese, Cantonese and Mandarin performed an orthographic transcription task to evaluate the intelligibility of English sentences uttered by speakers of Japanese, Cantonese, Spanish and Polish. Support for an ISIB was inconsistent: the Japanese listeners understood the Japanese talkers' production better than the native English listeners did (ISIB-L), and they also understood Japanese talkers better than they understood other L2 talkers (ISIB-T). However, these results were not replicated for the Cantonese listeners with respect to the Cantonese talkers' production, and there was no consistent evidence of a mismatched ISIB either. Similarly, Major, Fitzmaurice, Bunta and Balasubramanian (2002) had listeners of different L1 backgrounds evaluate the listening comprehension of lectures given by non-native speakers and found an ISIB-T with the Spanish listeners but not with the Chinese and Japanese listeners.

Finally, in a study investigating the intelligibility of English words produced by five speakers (one native, four non-native), Stibbard and Lee (2006) did not find any instance of ISIB. However, they observed what they termed as a (mismatched) interlanguage speech intelligibility detriment: Korean listeners' performance was significantly worse with a low proficiency Saudi-Arabian talker than with a low-proficiency Korean talker, while the opposite was true for the Saudi-Arabian listeners.

Munro et al. (2006) also evaluated the comprehensibility and accentedness of the L2 speech samples. As with intelligibility, they reported cases of ISIB-L and ISIB-T, but these were not consistent as they involved only a subset of the possible talker and listener combinations.

They also found no relationship between familiarity with a given type of foreign accent and the different measures studied. These outcomes led Munro et al. to conclude that perception and intelligibility of L2 speech is more likely determined by acoustic-phonetic properties of non-native speech (stimulus properties) than by listeners' factors (see also Munro, 2008).

The disparity of these results probably stems from methodological differences among studies regarding the actual type and size of speech samples evaluated (whole utterances vs. words or individual sounds, nondegraded speech vs. speech in noise), the task (transcriptions of whole utterances, word or consonant identification, and reaction times),

the number of talkers and listeners per condition, the type of languages involved, and the measure used for proficiency. Further, most studies have examined these issues with respect to intelligibility of L2 speech, with few studies addressing foreign accentedness and

comprehensibility. In addition, few studies have examined the effect of L2 speech errors, or stimulus properties, on the intelligibility and perception of L2 speech by both native and non-native listeners.

Thus, the goal of this study was to explore these issues further by evaluating the possible links between accentedness, comprehensibility and intelligibility of L2 speech from the perspective of native listeners as well as non-native listeners from the same L1 background as the L2 speakers (i.e., matched-L1 listeners) and non-native listeners from a different L1 background from the L2 speakers (i.e., mismatched-L1 listeners). In addition, this paper extends previous work by exploring the relationship between two types of pronunciation errors, segmental and word stress errors, and the perception of L2 speech by native listeners, matched-L1 non-native listeners, and mismatched-L1 non-native listeners. The role of L2 speech errors is discussed next.

1.3. L2 speech errors

A few studies have examined the relationship between specific pronunciation errors and the perception of L2 speech. Anderson-Hsieh, Johnson and Koehler (1992) found that errors in the production of L2 segments, syllable structure and prosody correlated with judgements of comprehensibility and accentedness of L2 speech, with prosody exhibiting the strongest

correlation. Similarly, accurate production of suprasegmental features like stress timing was found to correlate with accentedness ratings in a study by Trofimovich and Baker (2006).

Further, in a study that evaluated the effect of several suprasegmental measures on the degree of accentedness and comprehensibility of non-native speech, Kang (2010) found that pitch range and proportion of stressed words in an utterance were the best predictors of accent judgments. Finally, Munro and Derwing (1999) observed that intonation accuracy was most frequently found to correlate with accentedness and comprehensibility ratings, followed by segmental errors in the case of accentedness. Pronunciation errors were less clearly associated with intelligibility scores. The results of these correlational studies suggest that suprasegmental errors may play a more prominent role than segmental errors in L2 speech perception, and that pronunciation errors may be more closely related to perceptual impressions of L2 speech than to its intelligibility. Still, some more controlled studies involving manipulated stimuli (accurately and inaccurately stressed samples) have shown that lexical stress errors can in fact affect both comprehensibility (Hahn, 2004) and intelligibility (Field, 2005) of L2 English. In Field's study stress-related changes in vowel quality also played a role, which points to the joint effect of more than one type of mispronunciation.

Assessing the relative contribution of stress and segmental errors to the perception of foreign speech was the objective of two studies involving L2 Dutch (Caspers, 2010; Caspers & Horloza, 2012). Their main finding was that words containing both stress and

segmental errors produced by French and Chinese speakers of Dutch were less intelligible and less comprehensible to Dutch listeners than words that had either only stress or segmental

errors produced by the same L2 speakers. It is possible that it is the number of errors rather than the type of errors that has an impact on the intelligibility of L2 speech. Regarding accentedness, error type was unrelated to degree of accentedness for the French talkers, while segmental errors were more detrimental than word stress errors in the case of the Chinese talkers.

In order to investigate further the relationship between pronunciation errors and perceptual measures of L2 speech, the current study was designed to examine the influence of word-stress errors and segmental errors on the perception of Polish-accented English by native speakers of Polish, English and Spanish. Polish has fixed stress mostly falling on the penultimate syllable (Gussmann, 2007). By contrast, English and Spanish have variable stress, even if lexical stress patterns are partly determined by the syntactic category of the word or, in the case of Spanish, by whether the word ends in a consonant or a vowel (e.g., Spencer, 1996; Hualde, 2005). These characteristics allowed us to explore if Polish speakers would transfer Polish stress patterns into English (Marczak, 2008), and what consequences that may have for English listeners, Polish (matched-L1) listeners and Spanish (mismatched-L1) listeners. Moreover, the current study evaluated the intelligibility, comprehensibility and accentedness of L2 speech samples consisting of full sentences rather than isolated words, thus allowing for a more global assessment of the effect of stimulus properties, as well as listener factors, on overall perception of L2 speech.

Given the issues discussed above, this study addressed the following questions:

1. What is the interrelation of foreign-accentedness, comprehensibility and intelligibility of L2 English speech as assessed by native English listeners, matched-L1 non-native listeners and mismatched-L1 non-native listeners?

2. What is the effect of the listener's L1 on the intelligibility, comprehensibility and accentedness of non-native speech? Is there an interlanguage speech intelligibility benefit for matched-L1 and mismatched-L1 non-native listeners?

3. What is the relationship between two types of stimulus properties, segmental errors and word stress errors, and the intelligibility, comprehensibility and accentedness of non-native speech as perceived by native English listeners, matched-L1 non-native listener and mismatched-L1 non-native listeners?

2. Method

2.1. Talkers

The talkers in this study were eleven native Polish speakers of English, five male and six female, and two native speakers of Southern British English, one male and one female. The Polish speakers' age ranged from 22 to 29 (mean: 24.6). They had learned English as a foreign language in Poland for an average of 11.5 years (range 6-18 years), and their mean age of first exposure to English was 9 (range 5-15). Their level of self reported proficiency in English ranged from 1 to 4 as indicated on a five-point scale in which five was native-

like and one meant a low level of proficiency (mean: 3.4; standard deviation: 0.9). The amount of daily use of English in different contexts ranged from never or rarely to sometime-

and often. Except for one Polish resident in Spain, the remaining Polish speakers were temporarily residing in Spain, the majority being graduate students on a European student exchange program. The native English talkers were speakers of Standard Southern British English, one female and one male, residing in Spain. Only two native English talkers were recorded as it was assumed that there would be very little variability in listeners' responses to native productions.

2.2. Speech samples

The present study followed Munro et al. (2006) in evaluating L2 speech by means of whole utterances in order to achieve a more global and naturalistic measure. In order to explore the effect of stress errors, a set of sentences containing words with potentially difficult stress patterns was created. Difficult words included verb-noun contrasts (e.g., the noun 'conduct' vs. the verb 'to conduct'), contrasts between compound nouns and phrasal verbs (e.g., 'a turn-out' vs. 'to turn out') and words illustrating different stress patterns in multisyllabic words (e.g., 'noticeable,' 'development,' 'independence'). The participants were first presented with the list of content words present in the speech stimuli and were invited to ask for clarification of any unfamiliar words.

The elicitation consisted of a reading task in which the participants read aloud two blocks of twenty sentences displayed one by one on a computer screen. The speech samples were

recorded individually in a soundproof booth at the speech laboratory at Universitat Autonoma de Barcelona. Recordings were made with Audacity audio editing software and a Rode NT1-A Studio Condenser microphone at a 44.1 kHz sampling rate. A recording session took about 20 to 30 minutes. After the recording, participants filled in a short questionnaire about their personal and linguistic background.

Three different utterances were selected from each talker, giving a total of 39 distinct test sentences to be used in the listening tasks. The number of sentences per talker was limited to three to avoid potential fatigue effects affecting the listeners. The three sentences from each talker were selected ensuring that the number and type of words was comparable across all talkers' utterances. The average length of the sentences was 12 words, ranging from 8 to 17, including function words.

2.3. Listeners

The subjects for the listening task were six native speakers of English (one female, five male), six Polish non-native speakers of English (three female, three male) and six Spanish non-native speakers of English (two female, four male). The mean age of the participants in all three groups was 26. Table 1 presents the characteristics of each listener group.

The Polish listeners were university students spending a term in Spain as part of a European exchange program. Five Spanish-speaking subjects were from Spain and one from Venezuela. Four were undergraduate or graduate students, and two were currently employed. Finally, the group of English speakers included five UK citizens and one American. Four of them had been living and working in Spain for over a year, one was an exchange student and the sixth one visited Spain periodically. All participants reported normal hearing and none had been enrolled in an English pronunciation course, had a degree in English or were English language teachers.

Table 1. Personal and language background information for the three groups of listeners.

	English L. n = 6	Polish L. n = 6	Spanish L. n = 6
Age (group average)	30	23	24
Self reported proficiency[a]	-	4	4
Familiarity with Polish accented speech[a]	3.0	3.67	2.17
Age of first exposure to English (group average)	-	9	9
Length of EFL learning (group average)	-	13	13
Proficiency measure[b] (group average)	-	83	78

a. Based on a 5-point Likert scale (5 = High – native or native like –, 1 = Low).

b. Mean scores on the Oxford Placement Test results (grammar, out of 100).

2.4. Listening tasks

The listening tasks were an orthographic transcription task (OT task), an accentedness rating task and a comprehensibility rating task. The OT task and the foreign accent rating were performed first, followed by the comprehensibility judgement task. This order was adopted following the findings of previous studies showing that accentedness ratings tended to be harsher when the stimuli were heard a second time (Flege & Fletcher, 1992; Munro & Derwing, 1994, Munro et al., 2006).[1]

In the OT task the participants listened to each speech sample once and transcribed what they had heard in standard orthography on the handouts provided. Upon completion of the orthographic transcription of an utterance, participants judged its degree of foreign-accent by clicking on a 9-point Likert scale (1 = "No foreign accent", 9 = "Strong foreign accent"). A 9-point scale has been found appropriate for measuring foreign accentedness (Southwood & Flege, 1999). After completing the OT task and the accent rating, listeners heard each sentence once again in a new randomized order and rated the comprehensibility of each utterance on a 9-point Likert scale where 1 meant "easy to understand" and 9 meant "impossible to understand," following previous studies (Derwing & Munro, 1997; Munro & Derwing, 1999; Munro et al., 2006). The 39 individual sentences were randomly presented using Praat software (Boersma & Weenink, 2014).

Prior to the actual tasks, listeners performed a practice session consisting of five speech samples produced by an additional Polish speaker of English whose production was not used in the experiment tasks. The whole experiment lasted approximately 45 minutes, including a 5-minute practice session, 25 minutes for the OT and accentedness judgement tasks, a 5-minute break and 10 minutes for the comprehensibility judgment task.

At the end of each session participants filled out a background questionnaire.

3. Results

3.1. Data analysis

3.1.1. Listening tasks

The 702 orthographic transcriptions (39 sentences by 18 listeners) resulting from the OT task were analyzed to assess the intelligibility of the intended sentences. Transcription errors such as different lexical items (e.g., *contact* for *conduct*) or the omission of a content word were considered mistranscribed words. Minor errors such as trivial substitutions and use of contractions (e.g., *this is* for *it is*, *we'll* for *we will*), or the use of abbreviated forms (*photos* for *photographs*) were not computed as they did not affect the meaning of the sentence. The total number of transcription errors made by the English, Polish and Spanish listeners was 206, 538 and 793, respectively. An intelligibility score was calculated by dividing the number of words transcribed correctly by the total number of words in the sentence. Each of the 39 samples obtained 18 individual measures of intelligibility (one per

listener). The average score per listener group for each sample was also calculated, as well as the average for each talker across all three sentences. The results of the comprehensibility and accentedness rating tasks were calculated in a similar fashion, based on the ratings provided by the listeners on 9-point Likert scales for each sample and each talker.

3.1.2. Error analysis

The authors of this study examined the non-native recordings separately and listed the number of mispronunciations resulting from segmental and lexical stress errors. Lexical stress errors consisted of the wrong placement of the main stress in content words (e.g., 'astro*no*my' for 'a*stro*nomy or 'to *ob*ject' for 'to ob*ject'). Regarding segmental errors, a decision was made to include only errors that were perceptually salient such as phoneme substitutions (e.g., /t/ for /d/ in 'foo<u>d</u>', /ɛ/ for /æ/ in 'h<u>a</u>ppiness'), segment deletions (e.g., 'hun' for 'hunt') and insertions (/reɪnɪŋk/ for /reɪnɪŋ/, 'raining'). Phonetic errors that did not deviate dramatically from the native pronunciation were disregarded, e.g., differences in aspiration, or vowel quality variations that did not impede vowel identification. In this sense we followed previous studies that focused on phonemic errors (e.g., Caspers & Horloza, 2012). There was a high level of agreement between the two researchers regarding stress errors but there were some discrepancies in the segmental analysis (25% of the errors).

Cases of discrepancy were reevaluated by a closer visual (inspecting spectrograms) and auditory examination until an agreement on all errors was reached. A total of 37 word-stress errors were identified distributed among 27 of the 33 sentences, ranging from 0 to 3 per utterance. The number of segmental errors was 106, ranging from 1 to 8.

3.2. Inter-rater reliability and intergroup correlations

Table 2 presents the reliability coefficients (Cronbach's alpha) for each listener group based on the listeners' responses to the 39 native and non-native samples in each task. A high inter-rater agreement across all the listeners in each group was observed for all three measures, ranging from .730 to .967. The highest values for reliability were observed for accentedness ratings. All values were higher than .700, which is considered the lower limit for evidence of reliability (Cortina, 1994), indicating that the listeners were consistent in their assessments of the native and non-native samples.

Table 2. Reliability coefficient (Cronbach's alpha) by listener group (n = 6 per group) for each measure assessed

	Intelligibility	Comprehensibility	Accentedness
English listeners	.831	.885	.967
Polish listeners	.791	.730	.918
Spanish listeners	.876	.870	.917

In order to assess the level of agreement among listener groups, Pearson correlation analyses were conducted comparing the results obtained from English and Polish, English and Spanish, and Spanish and Polish listener groups on each of the three measures.

Most correlations reached a significance level of at least .01, *r*-values ranging from .533 to .777, while the correlation involving intelligibility scores for English and Polish listeners reached the .05 level ($r = .429$, $p = .013$). This indicates a generally high level of agreement among listeners in their intelligibility scores and perceptual judgments regardless of their L1.

3.3. General results

Across all 18 listeners, the average intelligibility score, comprehensibility rating and accentedness ratings for the English talkers were 89%, 1.45 and 1.31. The corresponding results for the Polish talkers were 81%, 3.58 and 6.18. Hence, the productions of the two English talkers were more accurately transcribed, easier to understand and less accented than the productions of the Polish talkers. The English listeners were the most successful at transcribing the native and non-native productions as intended, with average scores of 99% for the native English sentences and 91% for the sentences produced by the Polish talkers. The corresponding results for the Polish and Spanish listeners were 84% and 82% for the native English samples and 80.5% and 71% for the Polish accented samples, respectively.

The results for comprehensibility and accentedness were more consistent across listener groups. Table A1 in Appendix 1 provides the mean and standard deviation, median and range of scores obtained for the native English talkers and the Polish talkers for each of the three listener groups.

Figures 1, 2 and 3 show the distribution of intelligibility scores, comprehensibility ratings and accent ratings, respectively, assigned to each of the 33 L2 utterances by all listeners in each group, that is, the distribution of 198 scores per group (33 sentences x 6 listeners). As shown in Figure 1, the distribution of intelligibility scores was highly skewed, with the majority of the scores falling in the 80-100% correct range. Comprehensibility scores were more evenly distributed along the 9 points on the scale, although the distribution was also right-skewed, with more than 50% of the samples obtaining scores between 1 and 3 (where 1 meant 'easy to understand'). Finally the distribution of accentedness ratings is skewed towards the left, with most samples in the 6-8 range of the scale (where 9 meant 'strong foreign accent').

Figure 1. Distribution of intelligibility scores obtained by the 33 L2 utterances.
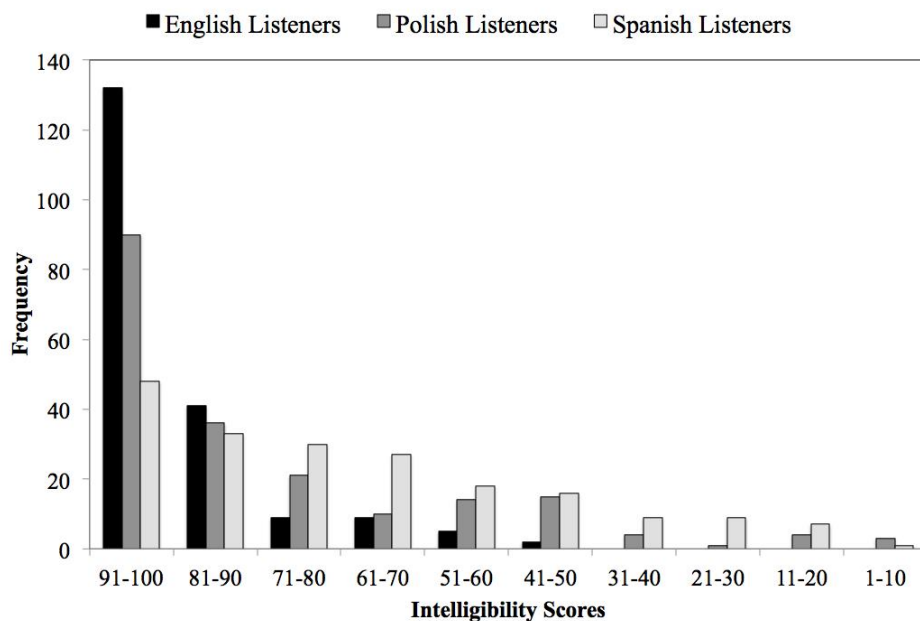
Figure 2. Distribution of comprehensibility ratings obtained by the 33 L2 utterances.
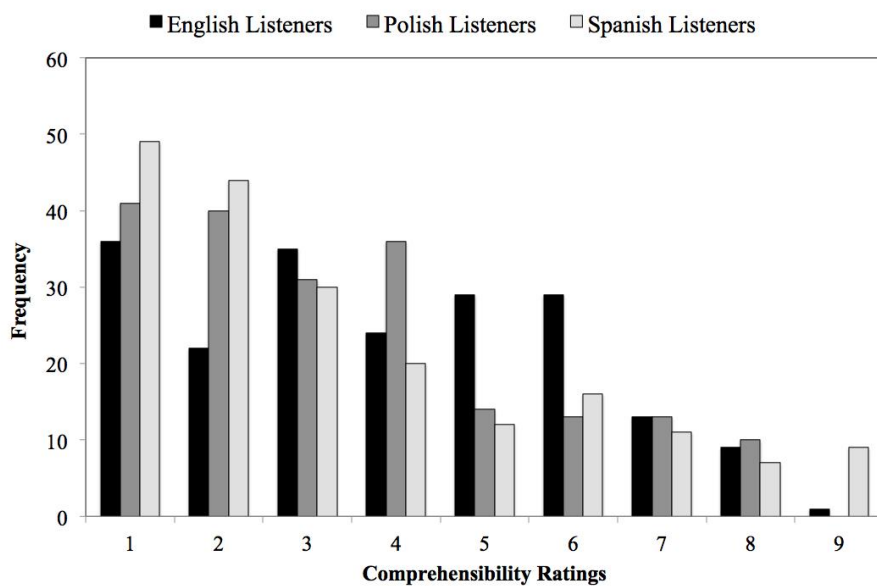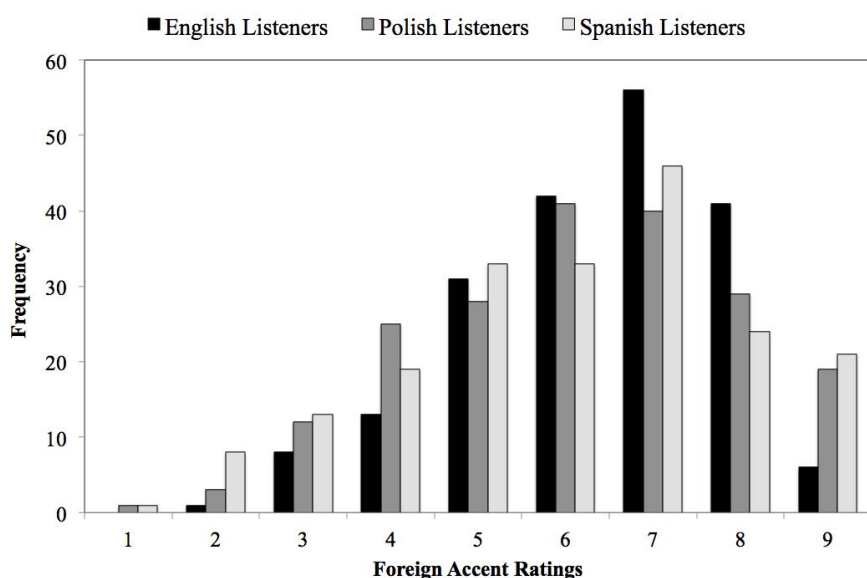

Figure 3. Distribution of foreign accent ratings obtained by the 33 L2 utterances.

Considering the average intelligibility score per sample per group, we find that for the English listeners eight of 33 L2 English utterances reached a score within the range of those obtained for native utterances (98-100%), and in fact the majority of the Polish accented samples obtained an accuracy score of 90% or higher (25 out of 33). The Polish listeners found 21 L2 utterances to be at least as intelligible as the native English productions (79-91%), and five of these were more intelligible than the native English samples (92-100%). In the case of the Spanish listeners, 26 L2 utterances fell within the range of scores obtained for the native samples (63-96%), and two more reached 100% intelligibility. These results show that some non-native productions were as intelligible as native productions both to native and non-native listeners.

The results of the rating tasks show that all three groups of listeners found the English talkers highly comprehensible and unaccented; for both measures and all listener groups the median rating assigned to the English talkers was the lowest in the scale, i.e., 1 out of 9. Similarly, all listeners found the L2 productions more difficult to understand and more accented than the native English productions. The English listeners found the Polish L2 productions less comprehensible (average of 3.9 out of 9) and more accented (6.36 out of 9) than did the two other groups of listeners (the average comprehension rating for both Polish and Spanish listeners was 3.42, the average accentedness rating was 6.12 and 6.05, respectively). For the English listeners, only one of the L2 utterances was comparable in comprehensibility to the native English utterances. This number reached 14 and 11 for the Polish and the Spanish listeners, respectively. By contrast, for all listener groups none of the sentences uttered by the Polish speakers of English obtained an accentedness rating that fell within the native speaker range. The relationship between the three measures is explored next.

3.4. Cross-task comparisons

Pearson correlation analyses were conducted on the average scores for each of the three measures across all listeners in each listener group for each non-native sample. The results are given in Table 3. In all instances but two the correlations were significant. The strongest correlations involved comprehensibility ratings and intelligibility scores. Notice that the negative correlation coefficients are due to the fact that for intelligibility the higher the

value, the greater the intelligibility, while in the rating tasks, the lower the value the more comprehensible and less accented the sample was. The strength of the correlation between accentedness and comprehensibility ratings was moderate to high.

The weakest correlations involved the comparison between accentedness ratings and intelligibility scores. Regarding group differences, the English listeners obtained the highest correlations for all three comparisons, while the Polish listeners yielded the weakest correlations. Unlike the results for Spanish and English listeners, there was no significant correlation observed for Polish listeners regarding intelligibility and accentedness, and only marginal significance for the correlation between accentedness and comprehensibility. It is possible that being more familiar with Polish-accented English (see Table 1), the Polish listeners' capacity to understand Polish-accented sentences was the least influenced by the degree of accentedness.

Table 3. Cross-task comparisons. Pearson correlation coefficients for each group and number of individual listeners in each group for whom correlations were significant.

Listener group	Intelligibility and Comprehensibility		Accentedness and Comprehensibility		Intelligibility and Accentedness	
English Ls (6)	-.825, $p < .001$	5 Ls	.804, $p < .001$	4 Ls	-.686, $p < .001$	3 Ls
Polish Ls (6)	-.667, $p < .001$	4 Ls	.344, $p = .05$	3 Ls	-.099 $p = .589$	1 Ls
Spanish Ls (6)	-.798, $p < .001$	6 Ls	.577, $p < .001$	3 Ls	-.380, $p < .05$	1 Ls

In order to evaluate if the individual listeners within each language group were consistent in their results, the cross-task comparisons were also assessed for each individual.

As shown in Table 3, 15 listeners yielded significant correlations between intelligibility scores and comprehensibility ratings, 10 between comprehensibility and foreign accent, and only five between intelligibility scores and accentedness ratings.

The individual results thus corroborate the general finding that intelligibility and comprehensibility are more strongly related while the relationship between accentedness and intelligibility is the weakest (Munro & Derwing, 1999).

3.5. Effects of listener L1 and talker L1

Figures 4, 5 and 6 illustrate the average intelligibility, comprehensibility and foreign accent scores assigned by each listener group to the two English native speakers and the 11 Polish speakers of English.

Figure 4. Mean intelligibility scores obtained by the English talkers and the Polish talkers
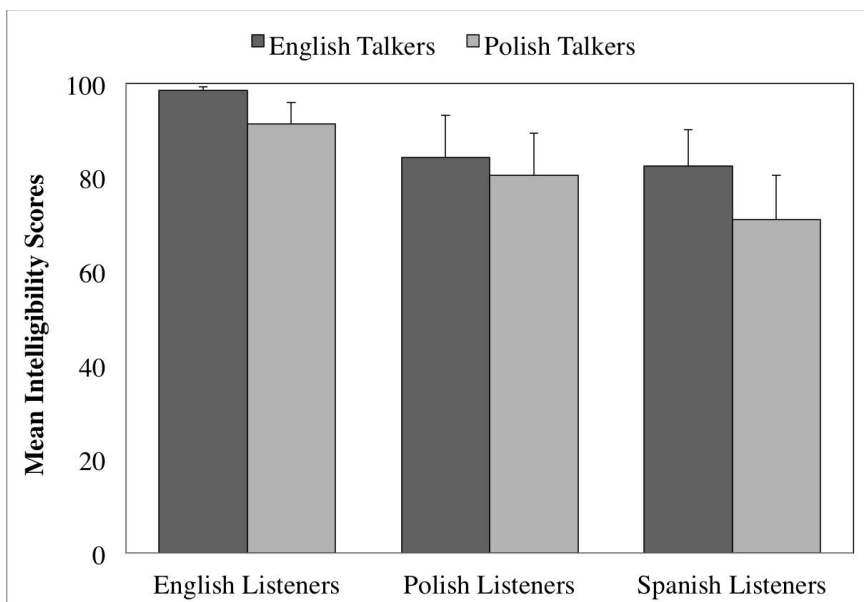

Figure 5. Mean comprehensibility ratings obtained by the English talkers and the Polish talkers
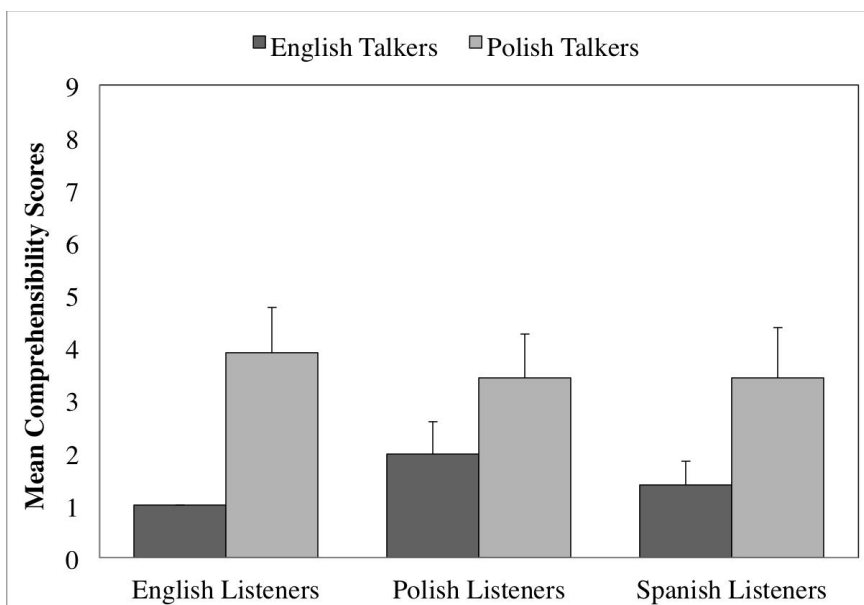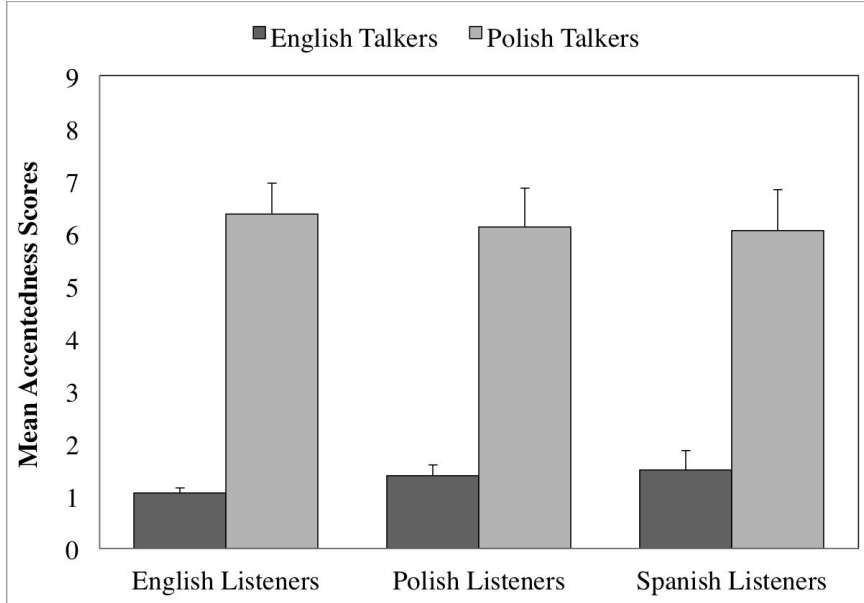
Figure 6. Mean accentedness ratings obtained by the English talkers and the Polish talkers

A 2 x 3 mixed design ANOVA was conducted on each dependent measure (intelligibility scores, comprehensibility ratings and accent ratings) exploring the effect of talker (English talkers, Polish talkers) as repeated measure and listener group (English, Polish and Spanish) as between groups factor. The intelligibility percent correct scores were log transformed to correct for non-normal distributions and render the proportional data amenable to parametric statistics. The results for intelligibility showed a significant main effect of talker L1, $F(1, 15)$ = 46.55, $p < .001$, $\eta^2 = .65$, and listener L1, $F(2, 15) = 7.5$, $p < .01$, $\eta^2 = .50$, and a talker L1 by listener L1 interaction of a modest effect, $F(2, 15) = 5.07$, $p < .05$, $\eta^2 = .14$.

With respect to comprehensibility ratings and foreign accent ratings, the main effect of talker L1 was significant in both cases: $F(1, 15) = 44.87$, $p < .001$, $\eta^2 = .71$, and $F(1, 15) = 589.67$, $p < .001$, $\eta^2 = .99$, respectively. However, the main effect of listener L1 and the two-way interaction did not reach significance in either case. Thus, the statistical analysis confirmed that all groups found the English talkers more comprehensible and less accented than the Polish talkers, regardless of the listeners' L1.

Regarding the main effect of listener L1 and the significant interaction obtained for the intelligibility scores, the results of Tukey HSD post-hoc tests showed that the English listeners significantly outperformed the Spanish listeners ($p < .01$) but their numerical advantage over the Polish listeners reached only marginal significance ($p = .052$). There was no difference between the Polish and the Spanish listeners ($p = .46$). Separate one-way ANOVAs were conducted on the data for English talkers and Polish talkers. The outcome replicated the general results. The effect of listener L1 was significant in both cases (English talkers: $F(2, 17) = 7.29$, $p < .01$, $\eta^2 = .50$, and Polish talkers: $F(2, 17) = 7.41$, $p < .01$, $\eta^2 = .49$). The corresponding post-hoc tests showed that the difference between English and Spanish listeners was significant in both cases ($p < .01$), while the English and Polish listeners differed significantly with respect to the English talkers' samples ($p < .05$) but not to the L2 utterances despite the English listeners' numerically higher scores (91% vs. 80.5%). No significant difference emerged between Spanish and Polish listeners.

The effect of talker L1 was explored further for each listener group by means of paired-sample t-tests comparing the average scores for English talkers and Polish talkers. The results showed that the English and Spanish listeners found the English talkers more intelligible than the Polish talkers: $t(5) = 5.43$, $p < .01$ and $t(5) = 4.93$, $p < .01$, respectively. However, in the case of the Polish listeners, the numerical difference (84% vs. 80.5%) did not reach significance, $t(5) = 1.56$, $p = .23$. Hence, only the Polish listeners found the Polish talkers as intelligible as the native English talkers.

3.6. Role of lexical stress errors and segmental errors

In order to explore the relationship between pronunciation errors and the perception of L2 speech, a series of correlations were conducted on the intelligibility scores, comprehensibility and accentedness ratings and the number of lexical stress errors and segmental errors obtained for the 33 non-native utterances (see section 3.1.2 above). There was no significant correlation between either number of segmental errors or number of stress errors and intelligibility scores for any of the three listener groups. The correlation between segmental errors and comprehensibility ratings reached significance for the English listeners ($r = .357$, $p < .05$), and was almost significant for the Polish listeners ($r = .350$, $p = .057$). There was also a marginally significant correlation between word stress errors and the English listeners' comprehensibility ratings ($r = .328$, $p = .062$).

The correlations with foreign accent ratings reached significance for all three groups (English listeners: $r = .351$, $p < .05$; Polish listeners: $r = .508$, $p < .01$; Spanish listeners: $r = .539$, $p < .01$).

Thus, the number of segmental and stress-based errors was not related to intelligibility scores. By contrast, the number of segmental errors did correspond to an increase in comprehensibility ratings and particularly accentedness ratings, although the correlations were moderate. Correlations between utterance length and each of the three measures were not significant.

4. Discussion

This study investigated the contribution of listener factors and stimulus properties to the perception of L2 speech. Three groups of listeners (native English listeners, matched-L1 non-native listeners and mismatched-L1 non-native listeners) evaluated the intelligibility, comprehensibility and foreign accentedness of a set of English utterances produced by Polish and English speakers. A high degree of consistency was observed in all listeners' responses regardless of the listeners' L1. This outcome indicates that perception of L2 speech is not necessarily affected by the linguistic background of the listeners, in agreement with previous findings (Munro et al., 2006). Cross-task comparisons showed that the production of Polish speakers of English was perceived to be fairly comprehensible and mostly intelligible despite being moderately to heavily accented. Strong correlations were found between intelligibility scores and comprehensibility ratings, and moderate to strong correlations between comprehensibility and accentedness ratings. The least consistent correlations involved intelligibility scores and accentedness ratings, which were strong for English listeners, moderate for Spanish listeners and did not reach significance

for Polish listeners. It is possible that Polish listeners' greater familiarity with Polish-accented speech may have resulted in a higher tolerance for deviation from native English pronunciation.

For all native and non-native listener groups accentedness ratings were harsher than comprehensibility ratings, replicating earlier findings (Munro & Derwing, 1999). The general outcome thus provides further evidence that accented speech can be highly intelligible, and that intelligibility, comprehensibility and accentedness are partially independent dimensions of L2 speech (Derwing & Munro, 1997; Munro & Derwing, 1999; Smiljanić & Bradlow, 2011). These findings strengthen the view that achieving intelligibility, rather than native-likeness, should be the ultimate goal in foreign language teaching and learning.

The second research question addressed possible listener effects by evaluating the prevalence of an interlanguage speech intelligibility benefit (Bent & Bradlow, 2003). All three listener groups found L2 speech less intelligible than native speech, although for Polish listeners the numerical difference (80.5% vs. 84%) did not reach significance. Hence there was no evidence of a mismatched ISIB for talkers, but the results for the Polish listeners may point to a matched ISIB-T. Still, it is not an actual benefit in the sense that Polish listeners did not find Polish talkers more intelligible than native English speech (Stibbard & Lee, 2006).

With respect to listener group comparisons, native English listeners outperformed both Polish and Spanish listeners in the correct transcription of Polish-accented utterances (91% vs. 80.5% and 71%, respectively).

This difference reached statistical significance in the case of English and Spanish listeners, lending no support for a mismatched ISIB-L (Hayes-Harb el at., 2008). In fact, the Spanish listeners' overall poorer performance indicates an interlanguage speech intelligibility detriment for non-native listeners who differ in L1 background from the non-native talkers (Stibbard & Lee, 2006). The lack of significant difference between English and Polish listeners in their transcription of the L2 speech samples could lend support to a matched ISIB-L (Bent & Bradlow, 2003). However, the numerical superiority of the English listeners' results weakens the support, especially under an actual benefit interpretation (Stibbard & Lee, 2006).

The mismatched interlanguage speech intelligibility detriment may be linked to he Spanish listeners' relative lack of familiarity with Polish-accented English. Thus it is possible that Spanish listeners had the greatest difficulty with features that are characteristic of Polish-accented English but not of Spanish-accented English. For example, Polish speakers were found to make a number of word stress errors, which presumably stem from the fixed nature of the Polish stress system, as opposed to the English and Spanish free-stress systems.

An examination of the results of the orthographic transcription task revealed that the six Polish listeners mistranscribed or omitted a total of 34 wrongly stressed words, ranging from 1 to 13 errors per listener, with an average of 5.7 words. The English listeners failed to understand 32 wrongly stressed words, ranging from 3 to 7 per listener, with an average of 5.3

Finally the Spanish listeners made 58 errors, ranging from 5 to 12 per listener and averaging 9.7. The Spanish listeners, therefore, appeared to have greater difficulty than the Polish listeners in understanding inaccurately stressed words.[2] Presumably a mismatched ISIB would be more likely in cases where the non-native features of the accented production were common to the speech of both non-native talkers and listeners. However, Bent and Bradlow (2003) reported that a group of listeners from a variety of L1 backgrounds found Korean-accented and Chinese-accented English more intelligibile than native English speech. Bent and Bradlow thus argued that a mismatched ISIB could not simply be due to similarities between the interlanguages of different listener groups, but to general tendencies or strategies common to non-native speakers of the same TL. It is not apparent, however, how those general tendencies might emerge and be shared by speakers of different L1s. The explanation for the presence or absence of an ISIB thus deserves further research possibly involving controlled studies investigating the role of shared vs. L1-specific interlanguage features in the perception of accented speech by different groups of non-native listeners.[3]

Our results are not in line with previous findings showing an ISIB (Bent & Bradlow, 2003; Hayes-Harb et al., 2008; Xie & Fowler, 2013) and are more in agreement with Stibbard and Lee's (2006) and Munro et al.'s (2006) lack of consistent ISIB effects. Methodological differences across studies may provide another explanation for the discrepant results, as discussed earlier. For example, different measures of intelligibility were used (full sentence transcription Munro et al.'s and our study vs. final consonant identification in Hayes-Harb et al.'s and Xie and Fowler's).

Further, the number of talkers also varied across studies, ranging from two talkers per L1 in Bent and Bradlow (2003) and Stibbard and Lee (2006) (one high-proficiency and one low-proficiency) to 11-12 non-native talkers per group in Munro et al. (2006) and the current study. It seems reasonable to assume that a small sample of talkers may result in the listeners' greater dependence on the characteristics of particular individuals and thus render the results less generalizable (Stibbard & Lee, 2006).

A few studies have also shown that the potential effect of the listener's L1 may be modulated by L2 proficiency (Bent & Bradlow, 2003; Hayes-Harb et al., 2008; Stibbard & Lee, 2006; van Wijngaarden et al., 2002). However, their findings are not consistent with regards to whether the determining factor is the talker's or the listener's proficiency, or whether the effect is stronger for high-proficiency or low-proficiency L2 speakers.

Again methodological differences regarding the number of talkers and listeners make cross-study comparisons difficult. The current study did not aim to explore the effect of proficiency, as no distinct levels of proficiency were tested. However, as an exploratory inspection, we examined the correlation between the individual listeners' results in an English grammar test and their performance in the three tasks. The listeners' proficiency scores were not found to correlate with comprehensibility or accent ratings. However, perhaps unsurprisingly, the higher the listeners' proficiency, the better they transcribed both native and non-native utterances. This was so regardless of the L1, pointing to a greater role of listener proficiency than listener L1 in the ability to understand L2 utterances correctly. Nonetheless, comprehending the nature of the interrelation between listener and talker proficiency and the perception of L2 speech lies beyond the scope of this study as it requires a more specifically designed methodology.

Regarding the relationship between listener L1 and comprehensibility and accentedness of L2 speech, all three groups of listeners found Polish-accented English less comprehensible and more accented than native English speech. Similarly, the three groups provided comparable comprehensibility and accentedness ratings for Polish-accented English. Munro et al. (2006) also reported no link between listener L1 and accentedness ratings, and an inconsistent effect of listener L1 on comprehensibility. Given the limited effect of listener L1, Munro et al. concluded that properties in the speech itself may play a stronger role than the listener's language background in the perception of L2 speech.

Evaluating the effect of stimulus properties was in fact the third goal of this study. The contribution of two different types of pronunciation errors, segmental and word stress errors, to the perception of L2 speech was assessed. Lexical stress was investigated due to the typological difference between English, characterized by variable stress, and Polish, which has fixed stress, which led us to hypothesize that Polish speakers would make stress placement errors. As discussed earlier, the Polish speakers of English were indeed found to make a number of word stress errors (e.g., 'deve*lop*ment' for 'de*vel*opment', 'philo*so*phers' for 'phi*lo*sophers, 'noti*cea*ble' for '*no*ticeable'), as well as a variety of segmental errors. Correlation analyses indicated that pronunciation errors may be closely related to accentedness ratings, followed by comprehensibility ratings, but they appeared to be unrelated to intelligibility scores. This result replicates the pattern observed in Munro and Derwing (1999), who evaluated phonetic, phonemic and intonation errors, and further emphasizes that deviations from native pronunciation may be more closely related to listeners' impressionistic assessments of non-native production than to the intelligibility of non-native speech.

It is possible that the inclusion of only two types of pronunciation errors may not be enough to assess the role of stimulus properties in native and non-native perception of L2 speech. This limitation is discussed further below.

Regarding the role of different types of errors, our results indicate that comprehensibility and accentedness ratings are more closely related to the number of segmental errors than to the number of stress errors. This is in line with the outcome reported by Casper (2010) for Chinese-accented Dutch words. However, it could be argued that the measure we adopted to compare the contribution of the two error types was not appropriate for two reasons. First, the number of potential segmental errors was larger than the number of potential word stress errors simply because there were many more segments than words per utterance. Secondly, given that stress placement is closely related to vowel quality in English, many segmental errors could be the direct result of word stress errors. For instance, Field (2005) found that erroneous stress placement together with mispronunciations involving vowel quality affected intelligibility of L2 speech by native and non-native listeners. Further, Caspers and Horloza (2012) reported that the intelligibility and comprehensibility of French-accented and Chinese-accented Dutch words were compromised the most when the L2 words contained both segmental errors and stress errors. In fact, we found that the number of segmental and word stress errors per utterance was significantly correlated ($r = .449$, $p < .01$).

Thus vowel mispronunciations could be the consequence of erroneous stress placement, as in the wrongly stressed words 'deve*lop*ment' and 'noti*cea*ble' pronounced with a full vowel in the wrongly stressed syllable and a reduced vowel in the wrongly unstressed syllable. Therefore, segmental errors, some of which are due to stress misallocation, may constitute one feature of L2 English speech that contributes to the perception of foreign accent. The fact that a greater number of errors corresponded to an increase in accented ratings for all three

36

groups of listeners, regardless of the listeners' L1, points to the importance of stimulus properties in L2 speech perception.

This study had a number of limitations, such as the limited number of speaker L1s and the fact that the level of foreign language proficiency was not fully addressed in the case of the listeners and was not investigated in the case of the talkers. As pointed out above, another important limitation is that the evaluation of the stimulus properties was restricted to two types of pronunciation errors, segmental and lexical stress errors, evaluated globally. This approach may have missed other factors such as the type of segmental error (vowels vs. consonants) and the location within the word (cf. Bent, Bradlow & Smith, 2007). In addition, an analysis limited to segmental and word-stress mispronunciations is clearly insufficient. Other aspects of pronunciation have been found to interact with the degree of foreign accent, such as intonation accuracy (Munro & Derwing, 1999), prosody and syllable structure (Anderson-Hsieh et al., 1992), or speech rate and pause duration (Trofimovich & Baker, 2006). Foreign language pronunciation practice thus should not be limited to consonant and vowel production.

For instance, Derwing, Munro and Wiebe (1998) found that learners benefitted more from a global training program designed for practising a variety of pronunciation features, including stress, rhythm, intonation, and speech rate, than from training focused solely on segmental contrasts. In line with Derwing et al.'s finding, our results also point to the importance of drawing English learners' attention to the effect of word stress (as well as sentence stress) on the pronunciation of individual sounds in English.

5. Conclusions

The study reported in this paper indicates that accentedness, intelligibility and comprehensibility are, to some extent, independent dimensions of L2 speech and should not be confused with one another. Further, it adds to a growing body of research showing that even heavily-accented L2 speech can be highly comprehensible and intelligible for native and non-native listeners. Regarding listener effects, no strong evidence was found to fully support an interlanguage speech intelligibility benefit for the listeners. In this study, the non-native speakers had no advantage over native speakers, as the Polish and Spanish listeners never outperformed the English native listeners in the orthographic transcription task. However, support for an interlanguage speech intelligibility detriment was observed since the Spanish listeners were the least accurate in providing orthographic transcriptions of the utterances produced by Polish speakers of English. It is possible that features of L2 speech that are specific to Polish-accented English may account for the Spanish listeners' greater difficulty, although this issue requires further investigation by means of carefully designed studies for that question. Listener L1 was not found to influence the other two dimensions of L2 speech examined, namely comprehensibility and accentedness.

By contrast, the properties of the stimuli appeared to be related to the perception of foreign accent, and to a lesser extent to the comprehensibility of L2 speech, regardless of the listeners' L1s. Segmental errors were more detrimental than word stress errors, although the two types of errors were in part related in the current case.

No relationship emerged between segmental or word-stress errors and intelligibility. It is possible that other aspects of pronunciation not examined in this paper are more closely related to intelligibility.

References

Abercrombie, D. (1949). Teaching pronunciation. *English Language Teaching Journal*, *3*(5), 113-122.

Anderson-Hsieh, J., Johnson, R. & Koehler, K. (1992). The relationship between native speaker judgements of non-native pronunciation and deviance in segmental, prosody, and syllable structure. *Language Learning*, *42*, 529-555.

Bent, T. & Bradlow, A.R. (2003). The interlanguage speech intelligibility benefit. *Journal of Acoustical Society of America*, *114*(3), 1600-1610.

Bent, T., A. Bradlow & Smith, B. (2007). Segmental errors in different word position and their effects on intelligibility of non-native speech. In O-S. Bohn & M.J. Munro (Eds.), *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege.* Amsterdam/Philadelphia: John Benjamins, pp. 331-347.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved 4 October 2014 from http://www.praat.org/

Caspers, J. (2010). The influence of erroneous stress position and segmental errors on intelligibility, comprehensibility and foreign accent in Dutch as a second language. In van Kampen, J., Nouwen, R. (eds.), *Linguistics in the Netherlands 2010*. John Benjamins: Amsterdam.17-29.

Caspers, J. & Horloza, K. (2012). Intelligibility of non-natively produced Dutch words: Interaction between segmental and suprasegmental errors. *Phonetica, 69*, 94-107.

Cortina, J.M. (1994). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104.

Derwing, T.M. & Munro, M.J. (1997). Accent, intelligibility and comprehensibility. Evidence from four L1s. *Studies in Second Language Acquisition, 20*, 1-16.

Derwing, T. M. & Munro, M.J. (2005). Language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*, 379-397.

Derwing, T.M., Munro, M.J., & Wiebe, G.E. (1998). Evidence in favour of a broad framework for pronunciation instruction. *Language Learning, 48*(3), 393-410.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*(3), 399-423.

Flege, J.E. & Fletcher, K.L. (1992). Talker and listener effects on degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125-3134.

Gass, S. M. & Varonis, E.M. (1984). The effect of familiarity on the comprehensibility of non-native speech. *Language Learning*, *34*, 65-89.

Gussmann, E. (2007). *The Phonology of Polish*. Oxford: Oxford University Press.

Hahn, L.D. (2004). Primary stress and intelligibility: Research to motivate teaching of suprasegmentals. *TESOL Quarterly*, *38*(2), 201-223.

Hayes-Harb, R., Smith, B., Bent, T., & Bradlow, A. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrast. *Journal of Phonetics*, *36*, 664-679.

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge: Cambridge University Press.

Imai, S., Walley, A.C., & Flege, J.E. (2005). Lexical frequency and neighbourhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of Acoustical Society of America*, *117*, 896-907.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics, 23*(1), 83-103.

Kang, O. (2010). Relative silence of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System, 38*, 301-315.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review, 64*, 459-490.

Major, R.C. (2013). Foreign accent. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics.* Wiley-Blackwell.

Major, R.C, Fitzmaurice, S.F., Bunta, F., & Balasubramanian, C. (2002). The effects of non-native accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly, 36*(2), 173–190.

Marczak, A. (2008). Production of English stress beats: The case of Polish speakers. *Poznan Studies in Contemporary Linguistics*, *44*(4), 545-568.

Morley, J. (1991). The pronunciation component of teaching English to speakers of other languages. *TESOL Quarterly, 25*, 481-520.

Munro, M.J. (2008). Foreign accent and speech intelligibility. In J.G. Hansen Edwards, & M.L. Zampini (Eds.), *Phonology and Second Language Acquisition.* John Benjamins: Amsterdam. 193-218.

Munro, M.J. (2013). Intelligibility. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics.* Wiley-Blackwell.

Munro, M.J. & Derwing, T.M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing, 3*, 253-266.

Munro, M.J. & Derwing, T.M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289-306.

Munro, M.J. & Derwing, T.M. (1999). Foreign Accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*, Supplement 1: 285-310.

Munro, M.J., Derwing T.M., & S.L. Morton. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*, 111-131.

Pickering, L. (2006). Current research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics, 26*, 219-233.

Smiljanić, R., & Bradlow, A.R. (2011). Bidirectional clear speech perception benefit for native and high proficient non-native talkers and listeners: Intelligibility and accentedness. *Journal of Acoustical Society of America*, *130*(6), 4020-4031.

Southwood, M., & Flege, J.E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics and Phonetics*, *13*, 335-349.

Spencer, A. (1996). *Phonology: theory and description*. Oxford: Blackwell.

Stibbard, R.M., & J.I. Lee. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *Journal of the Acoustical Society of America*, *120*, 433-442.

Trofimovich, P. & Baker, W. (2006). Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*, 1-30.

van Wijngaarden, S.J. (2001). The intelligibility of non-native Dutch speech. *Speech Communication*, *35*, 103-113.

van Wijngaarden, S.J., Steeneken, H.J.M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America*, *111*, 1906-1916.

Villafaña Rojas, V. del C. (2011). On the acquisition of English stress by Spanish native speakers. In Wrembel, M., Kul, M., Dziubalska-Kołaczyk, K. (eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010*, Vol. II. Frankfurt am Main: Peter Lang. 275-286.

Xie, X. & Fowler, C.A. (2013). Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English. *Journal of Phonetics*, *41*, 369-378.

---

[1] In another study by Munro and Derwing (1995), task order was not found to have an effect on comprehensibility and accentedness ratings.

[2] Notice however that stress errors are not the only difference between Spanish-accented and Polish-accented English, and Spanish speakers may also make stress placement errors in English (e.g., Villafaña Rojas, 2011).

[3] A closer inspection of the Polish-accented samples revealed only a relatively small subset of segmental errors that could be conclusively identified as Polish-specific, providing insufficient data to explore this issue further. A specifically-designed study contrasting clearly defined Polish-specific features of accented English with features common to different interlanguages would be necessary.

Appendix 1. Descriptive statistics for the three listening tasks.

Table A1
Results of the three listening tasks

	Intelligibility			Comprehensibility			Foreign Accentedness		
	Mean (SD)	Median	Range	Mean (SD)	Median	Range	Mean (SD)	Median	Range
English Talkers (n = 2, 6 utterances)									
English Listeners (n = 6)	99 (2)	100	90-100	1 (0)	1	1-1	1.06 (0.23)	1	1-2
Polish Listeners (n = 6)	84 (22)	92	10-100	1.97 (1.52)	1	1-7	1.39 (0.49)	1	1-2
Spanish Listeners (n = 6)	82 (19)	90	20-100	1.39 (1.1)	1	1-7	1.5 (0.88)	1	1-5
All Listeners (n = 18)	89 (18)	100	10-100	1.45 (1.15)	1	1-7	1.31 (0.62)	1	1-5
Polish Talkers (n = 11, 33 utterances)									
English Listeners (n = 6)	91 (11)	100	43-100	3.9 (2.11)	4	1-9	6.36 (1.46)	7	2-9
Polish Listeners (n = 6)	80.5 (22)	89	6-100	3.42 (2.06)	3	1-8	6.12 (1.79)	6	1-9
Spanish Listeners (n = 6)	71 (23)	75	10-100	3.42 (2.35)	3	1-9	6.05 (1.9)	6	1-9
All Listeners (n = 18)	81 (21)	89	6-100	3.58 (2.19)	3	1-9	6.18 (1.73)	6	1-9

Appendix 2. List of sentences used in the experiments listed by speaker (ES = English speaker, PS = Polish speaker).

ES1
1.Cambridge certificates open the door to remarkable job opportunities in Europe.
2. In order to set up your database you need to restart your computer.
3. The president's car is usually followed by a police escort.

ES2
4. I wanted to become a diplomat but in the end I studied psychology.
5. He told us about the main causes of the greenhouse effect and we found it very interesting
6. You can decrease your weight by eating right and exercising.

PS1
7. This is not democratic. It is the antithesis of democracy.
8. The fitness instructor started them on a new workout.
9. I object to violence because when it appears to do good, the good is only temporary.

PS2
10. Her parents do not approve of the wedding because she is catholic and he is a Muslim.
11. This new cake recipe did not turn out very well.
12. Meanwhile, philosophers debate whether it's right to clone an individual.

PS3
13. His work was instrumental in developing links with European organizations.
14. The drive-through is open even if the dining room is closed.
15. Please, check out the upcoming debate schedule as soon as possible.

PS4
16. Alcohol has a noticeable effect on the body.
17. I hope the sun breaks through the clouds soon, it's been raining all week!
18. Astronomy was an object of interest already in ancient times.

PS5
19. We hired a professional photographer to take some photographs during
the wedding ceremony.
20. The political effects of food cutbacks could be devastating.
21. She keeps all drugs and alcohol out of the house because her boyfriend is a recovering addict.

PS6
22. Unhappiness does not unite people, but separates them.
23. Rising consumer prices have forced shoppers to hunt for bargains and cut back on overall spending.
24. It is highly probable that they don't permit smoking in this restaurant.

PS7

25. Some people believe that the development of capitalism is good for mankind.

26. We will have to carry on the best we can even without his help.

27. The best way to find information about changeable social trends is to conduct a survey.

PS8

28. How do we reach an acceptable level of data security?

29. Only after I asked her to do so, she showed me her bus pass.

30. A remake of the movie "Planet of the Apes" was a huge success.

PS9

31. We can still find some furniture from the Colonial period in antique shops.

32. Yesterday, they went to the gym to work out.

33. In the desert, there is a big contrast between temperatures in the day and at night.

PS10

34. He is a very irresponsible and chaotic person!

35. While I was waiting for my taxi to arrive, I saw Mary's bus pass.

36. I will present his ideas to the company tomorrow in an oral presentation.

PS11

37. It is often said that adolescence is a time for developing independence.

38. You have to walk over the bridge in order to get to the opposite side of the city.

39. It was really nice of him to suggest that.