

# Perception of Foreign Accent in Spontaneous L2 English Speech

Jiahong Yuan<sup>1\*</sup>, Yue Jiang<sup>2\*</sup>, Ziang Song<sup>1</sup>

<sup>1</sup>University of Pennsylvania, Philadelphia, U.S.A.

<sup>2</sup>Xi'an Jiaotong University, Xi'an, China

\*The authors equally contributed to this work.

jiahong@ling.upenn.edu, jiang8@mail.xjtu.edu.cn, ziang@seas.upenn.edu

## Abstract

This study compared English and Mandarin Chinese listeners' assessments of foreign accent in spontaneous English spoken by speakers of eight L1 languages. Mandarin Chinese listeners perceived lower degree of foreign accent than native listeners, and were less sensitive to Mandarin and Cantonese accents than to the other accents, especially French, Spanish, and Russian. Acoustic analysis suggests that English and Mandarin Chinese listeners relied on different cues in the perception of Foreign Accent in L2 English.

**Index Terms:** Foreign Accent, Perception, Prosody, L1 effect

## 1. Introduction

The impact of the native language (L1) sound system on the perception of non-native speech sounds has been a central problem in developing theories of cross-language speech perception and second language (L2) speech learning, e.g., the Perceptual Assimilation Model [1] and the Speech Learning Model [2]. Both of the models regard phonetic similarity between L1 and L2 as key in predicting L2 learners' success or difficulty in perceptually distinguishing segmental contrasts in L2 [3]. One of the most extensively studied phenomena along this line of research, for example, is Japanese listeners' difficulty in perceiving the English /l-/r/ distinction [4].

Much less is known, however, about the perception of accented L2 speech by L2 learners. A few studies have found that the speech of non-native speakers is more intelligible to non-native listeners than to native listeners [5, 6]. It was also found that for non-native English listeners, intelligibility of a high-proficiency non-native English speaker from a different native language background was greater than or equal to the intelligibility of a native English speaker, the so called "mismatched interlanguage speech intelligibility benefit" [6]. The study [6] suggested that the observed mismatched interlanguage benefit was unlikely to be the result of structural similarities between the native languages of the speakers and listeners, but rather due to certain tendencies in foreign accented English regardless of native language background.

In this study, we intend to compare native and non-native judges' assessment of spontaneous L2 accented English speech of speakers from different native language backgrounds, and to find answers to the following two questions: 1) Can non-native judges reliably assess degree of foreign accent, as perceived by native judges? 2) How is non-native judges' assessments affected by L2 speakers' native language backgrounds?

We also attempt to investigate what cues are employed by native and non-native judges to assess degree of foreign accent. It has been found that prosodic deviations contribute more than or at least equally to segmental deviations in perceived foreign accent by native listeners [7, 8]. [9] reported that the intelligibility of Chinese-accented English phrases was

significantly improved after they were manipulated so that the segments in the phrases were given native segment durations, while their spectral and source characteristics were retained. [10] demonstrated that although non-proficient English learners with Japanese L1 were not less sensitive than the native American English speakers in discriminating the durations of stressed vowels in English, their judgments of what is natural for a stressed vowel spanned a broader range of vowel durations than for the native speakers. [11] showed that perceptual judgments of foreign accent in non-native German and English correlated primarily with speaking rate; pitch range and pitch movement played little role. [12] manipulated L2 Norwegian utterances of speakers from different L1 language backgrounds by changing the pitch and duration patterns to native Norwegian values. The study showed that the effects of the manipulations on degree of foreign accent were different for different L1 groups. The French and German L1 groups gained most from the pitch manipulation; the English, Tamil, and Chinese groups gained most from the duration manipulation; and the Russian and Persian groups were equally affected by the two manipulations.

In the following sections we first introduce the data set in Section 2, then we present a perception experiment and acoustic analysis in Sections 3 and 4, respectively. Finally, we end with discussion and conclusions in Section 5.

## 2. Data

The data were selected from the CSLU Foreign Accented English Corpus [13]. The corpus consists of spontaneous speech of English by native speakers of 23 L1 languages, in which the speakers were asked to speak about themselves in English for up to 20 seconds. Three native American English speakers (hereafter as native judges) judged the utterances' accents on a 4-point scale (see Figure 1 for the definition): negligible/no accent (denoted as '1'), mild accent ('2'), strong accent ('3'), and very strong accent ('4').

Eight L1 languages were included in this study: three tone languages - Cantonese, Mandarin, and Vietnamese; four stress languages - German, French, Spanish, and Russian; and one pitch accent language - Japanese. Based on the average of the three native judges' assessment scores, we randomly selected utterances from five accent levels: 2.0, 2.33, 2.67, 3.0, and 3.33. The three judges' scores on the same selected utterance were either the same or differed by only one degree. Three utterances were selected for each accent level and each L1 language, making a total of 120 utterances (3\*5\*8 = 120).

As the CSLU corpus does not contain transcripts, a native English speaker manually transcribed the 120 utterances at the word level. The phonetic transcription and segmentation were then obtained through forced alignment, as described in Section 4.

### 3. Perception experiment

#### 3.1. Procedure

A perception experiment was conducted using a web-based software developed by the third author of the paper. The subjects listened to the utterances in a randomized order, and assessed foreign accent for each utterance by choosing from the four levels shown on the screen. The definition of the levels of accent, which is the same as used in the CSLU corpus, was also displayed on the screen. Five additional utterances from the same CSLU corpus were placed at the beginning for the subjects to familiarize themselves with the test. The subjects may listen to an utterance for multiple times before moving forward to the next one. A screenshot of the user interface is shown in Figure 1.

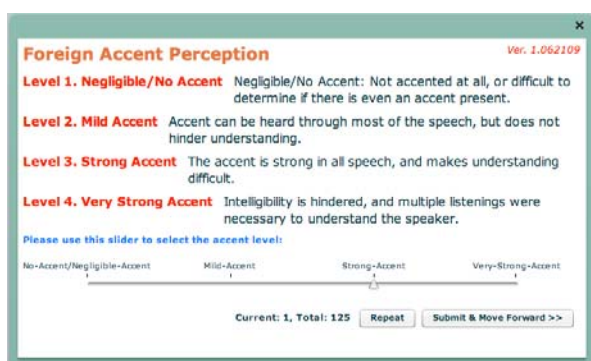


Figure 1: A screenshot of the perception experiment user interface. The definition of accent levels is displayed.

Eight graduate students in the English department at Xi'an Jiaotong University in China participated in the experiment. They are all native speakers of Mandarin Chinese and can speak fluent or near-fluent English (hereafter as Mandarin judges). All of them have studied English in school for more than 10 years, but none of them has lived in an English-speaking country or region.

#### 3.2. Results

Figure 2 is a comparison of the assessment scores from the perception experiment with those from the native judges contained in the CSLU corpus.

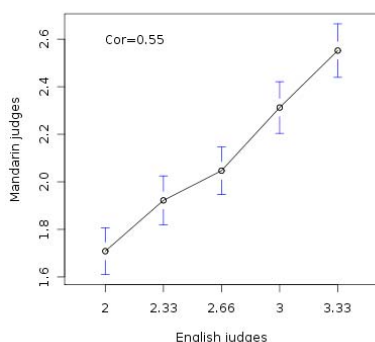


Figure 2: Comparison of the judgments made by the native and Mandarin Judges.

We can see from Figure 2 that Mandarin judges' average assessment scores are lower than native judges' at every accent level. The average scores from the Mandarin judges are

1.71, 1.92, 2.05, 2.31, and 2.55 respectively at the five levels of 2.0, 2.33, 2.67, 3.0, and 3.33 defined by native judges' assessments. Having lower assessment scores at the low accent levels suggests that Mandarin judges were less sensitive to foreign accent in L2 speech than native judges. From the definition shown in Figure 1, high levels of accent, 3 and 4, indicate that the intelligibility of the utterance is hindered. Having lower assessment scores at the high accent levels, as shown in Figure 2, suggests that the intelligibility of an L2 speaker with a strong accent is greater to Mandarin judges than to the native judges.

We can also see from Figure 2 that the Mandarin and native judges showed similarity when assessing degree of foreign accent. The mean assessment scores from the Mandarin judges' are significantly different from each other for every pair of the accent levels (defined by the native judgments) with a difference of more than 0.5 (t-tests,  $p < 0.01$ ), e.g., between the levels of 2 and 2.67 ( $p=0.003$ ). The correlation between the native and Mandarin per-utterance average scores is 0.55 ( $df = 118, t = 7.24, p < 0.001$ ).

Figure 3 plots the Mandarin judges' assessment scores for different L1 languages. The native judges' average scores for the L1 languages are the same, due to the way the utterances were selected. We can see from the figure that for Mandarin judges the perceived degree of accent is lower for the French, Spanish, and Russian accented utterances. The scores of the Japanese, German, and Vietnamese accented utterances are in the middle. From t-tests, the three groups, Mandarin and Cantonese (group 1), Japanese, German and Vietnamese (group 2), and French, Spanish and Russian (group 3), are pairwise different from each other ( $p \leq 0.01$ ). This result suggests that L1 language backgrounds affect non-native judges' assessment of degree of foreign accent. The Mandarin judges showed three levels of preference to different L1 languages, suggesting that structural similarities or differences between the L1 languages of the speaker and the listener play an important role in the listener's perception of foreign accent of the speaker.

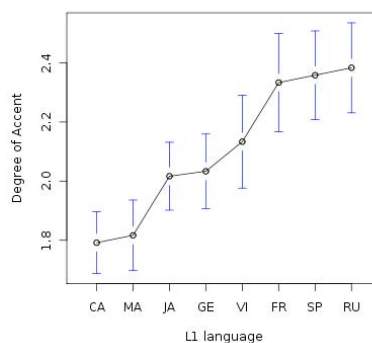


Figure 3: Judgments made by Mandarin judges on different L1 accents.

Figure 4 shows the mean scores for the three groups of L1 languages at different accent levels. As can be seen from the figure, the perceived degree of accent for group 3 (French, Spanish, and Russian) is higher than that for the other two groups at every accent level. This result suggests that Mandarin listeners generally disfavored or were more sensitive to French, Spanish, and Russian accents in L2 English.

We can also see from Figure 4 that the assessment scores of the Mandarin and Cantonese accented utterances at the accent levels of 3.0 and 3.3 were much lower, i.e., only 1.58

and 1.69 respectively. From the assessment instruction shown in Figure 1, we may infer that the utterances at these accent levels were difficult to understand for the native English judges. There was, however, no difficulty for the Mandarin judges to understand these utterances, as their assessment scores were below 2.0. This result is consistent with the previous studies [5, 6], i.e., that the speech of non-native speakers is more intelligible to non-native listeners than to native listeners.

Comparing the scores of group 1 and group2 languages at different accent levels in Figure 4, we can see that the difference is much larger at the higher levels than at the lower levels. The complicated effect of L1 language background on listeners' perception of foreign accent is interesting, but not surprising, because languages differ on many dimensions, e.g., segmental, prosodic, etc., and each dimension may play a different role in L2 speech production and perception.

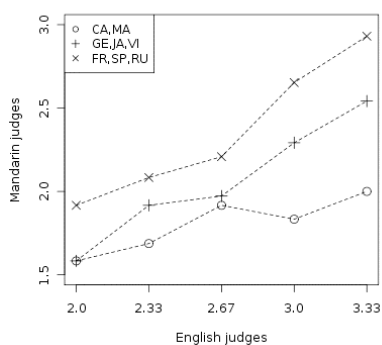


Figure 4: The mean scores for the three groups of L1 languages at different accent levels.

#### 4. Acoustic Analysis

The purpose of the acoustic analysis is to explore what cues are important for assessing degree of foreign accent in L2 speech, and whether native and non-native judges employ the cues differently.

The utterances were phonetically aligned and segmented using the Penn Phonetics Lab Forced Aligner (<http://www.ling.upenn.edu/phonetics/p2fa>). The average log likelihood score of an utterance associated with the forced alignment procedure was used to measure the phonetic deviation of the segments in the utterance from L1 English. The success of using likelihood scores from forced alignment to measure acoustic differences has been demonstrated in [14].

$F_0$ s of the utterances were extracted and smoothed using Praat. Various prosodic features were then computed from  $F_0$ s and segment boundaries. The features are listed in Table 1. The feature  $F_{0\_var}$  was obtained by counting the times the  $F_0$  contour of an utterance crosses its mean  $F_0$ , divided by the total number of  $F_0$  points in the contour. In the literature  $pvi$  [15] and  $pV/stdV$  [16] were proposed to measure the rhythmic characteristics of languages.  $pV$  is the proportion of duration of the vocalic segments in an utterance, i.e., the sum of vocalic intervals divided by the total duration of the utterance;  $stdV$  is the standard deviation of the vocalic intervals in the utterance.  $pvi$  refers to Pairwise Variability Index, it is defined as below, where  $m$  is the total number of vowels in an utterance, and  $d_k$  is the duration of the  $k$ th vowel.

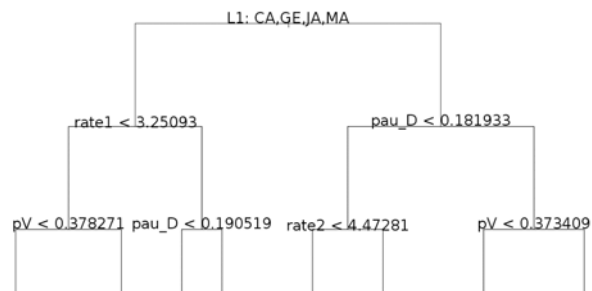
$$PVI = 100 \times \left[ \frac{\sum_{k=1}^{m-1} |d_k - d_{k+1}|}{(\sum_{k=1}^{m-1} (d_k + d_{k+1})/2)} / (m-1) \right]$$

Table 1. Acoustic features used for Regression trees.

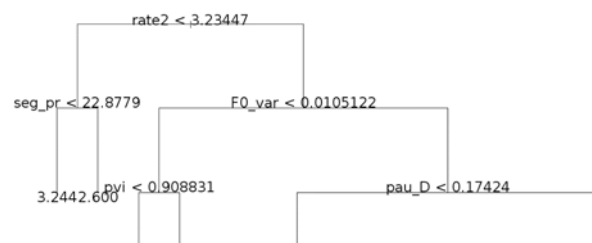
Feature	Description
L1	L1 language: CA, GE, FR, JA, MA, SP, RU, VI.
seg_pr	The average log likelihood of the utterance from forced alignment.
$F_{0\_var}$	The number of times the $F_0$ contour crosses the mean $F_0$ divided by the total number of $F_0$ s.
pvi	Normalized Pairwise Variability Index on vowels.
pau_N	The number of pauses.
pau_D	The total duration of pauses.
rate1	The number of syllables per minute; pauses were not excluded in the calculation.
rate2	The number of syllables per minute; pauses were excluded in the calculation.
pV	The proportion of vocalic intervals.
stdV	The standard deviation of vocalic intervals.
stress_dur	The average duration difference between stressed and unstressed vowels.
stress_ $F_0$	The average $F_0$ difference between stressed and unstressed vowels.

Based on the features in Table 1, regression trees were built to predict native and Mandarin judges' assessment scores, respectively. The first three levels of the trees are shown below.

##### 1. Regression tree for Mandarin judges' judgments:



##### 2. Regression tree for native English judges' judgments:



We can make the following conclusions from comparing these two trees. First, the native English and Mandarin Chinese judges relied on different cues when assessing degree of foreign accent in L2 English. Secondly, speaking rate and pauses were more important cues contributing to degree of perceived foreign accent than the other prosodic cues such as the duration and  $F_0$  difference between stressed and unstressed

syllables. Finally, the phonetic deviation of segments from L1 norms and the  $F_0$  variation were important cues in the English judges' perception of foreign accent in L2 English. The Mandarin Chinese judges were, however, either not able to detect these cues or not able to properly apply them when assessing foreign accent in L2 English.

## 5. Conclusions and Discussion

The present study demonstrates that in general, Mandarin Chinese listeners are able to assess degree of foreign accent in L2 English speech much like native English listeners. This result agrees with previous studies [17, 18].

The study also shows that Mandarin listeners were less sensitive to the mild accent in L2 English speech than native listeners, and they could better understand strong accented L2 English speech than native listeners. These results complement the 'interlanguage speech intelligibility benefit' finding [6], which states that for non-native listeners, intelligibility of a high-proficiency non-native speaker was equal to or greater than the intelligibility of a native speaker. Our results suggested that, from the perspective of comparing native and non-native listeners, a non-native speaker's speech was less accented and more intelligible to a non-native listener than to a native listener.

With regard to the effect of the speaker's L1 language, Mandarin listeners showed three levels of preference. They are, from high to low, Cantonese and Mandarin (group 1), German, Japanese, and Vietnamese (group 2), and French, Spanish, and Russian (group 3). This result suggests that structural similarities or differences between the L1 languages of the speaker and the listener play an important role in the listener's perception of foreign accent of the speaker. We also found that how a non-native speaker's L1 language affects a non-native listener's perception of the speaker's foreign accent is complicated, and the effect seems to show different patterns at different accent levels.

Acoustic analysis suggests that speaking rate and pauses play a more important role in perception of foreign accent for both native and non-native listeners. The phonetic deviation of segments and the  $F_0$  variation were important cues in the native judges' perception of foreign accent in L2 English. The Mandarin Chinese judges were, however, either not able to detect these cues or not able to properly apply them when assessing foreign accent in L2 English.

## 6. Acknowledgements

The authors thank the subjects who participated in the perception experiment, and thank Rebecca Marsh for transcribing the foreign accented English utterances.

## 7. References

[1] Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange, pp. 171–204.

[2] Flege, J. (1995). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange, pp. 233–277.

[3] Strange, W. (2007). "Cross-language phonetic similarity of vowels: Theoretical and methodological issues," in *Language experience in second language speech learning: In honor of James Emil Flege*, edited by O.S. Bohn and M. J. Munro, pp. 35–55.

[4] Aoyama, K., Flege, J., Guion, S., Yamada, T. and Akahane-Yamada, R. (2004). "Perceived phonetic dissimilarity and L2

speech learning: The case of Japanese /r/ and English /l/ and /r/," *Journal of Phonetics*, 23, 233–250.

[5] Jongman, A. and Wade, T. (2007). "Acoustic variability and perceptual learning: The case of non-native accented speech," in *Language experience in second language speech learning: In honor of James Emil Flege*, edited by O.S. Bohn and M. J. Munro, pp. 135–150.

[6] Bent, T. and Bradlow, A. R. (2002). "The interlanguage speech intelligibility benefit," *Journal of the Acoustical Society of America*, 114 (3), 1600–1610.

[7] Anderson-Hsieh, J., Johnson, J. and Koehler, K. (1992). "The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure," *Language Learning*, 42, 529–555.

[8] Boula de Mareuil, P. and Vieru-Dimulescu, B. (2006). "The contribution of prosody to the perception of foreign accent," *Phonetica*, 63, 247–267.

[9] Tajima, K., Port, R. and Dalby, J. (1997). "Effects of temporal correction on intelligibility of foreign accented English," *Journal of Phonetics*, 25, 1–24.

[10] Mochizuki-Sudo, M. and Kiritani, S. (1991). "Production and perception of stress-related durational patterns in Japanese learners of English," *Journal of Phonetics*, 19, 231–248.

[11] Gut, U. (2007). "Foreign Accent," in *Speaker Classification I: Fundamentals, Features, and Methods*, edited by C. Muller, pp. 75–87.

[12] Holm, S. (2008). *Intonational and durational contributions to the perception of foreign-accented Norwegian: An experimental phonetic investigation*, Ph.D. thesis, Norwegian University of Science and Technology.

[13] Lander, T. (2007). CSLU: Foreign Accented English Release 1.2, Linguistic Data Consortium, Philadelphia.

[14] Yuan, J. and Liberman, M. (2009). "Investigating /l/ variation in English through forced alignment," *Interspeech, 2009*, pp.

[15] Grabe, E. and Low, E. L. (2002). "Acoustic correlates of rhythm class," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner, pp. 515–546.

[16] Ramus, F., and Nespors, M. and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," *Cognition*, 73, 265–292.

[17] MacKay, I., Flege, J. and Imai, S. (2006). "Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent," *Applied Psycholinguistics*, 27, 157–183.

[18] Flege, J. (1988). "Factors affecting degree of perceived foreign accent in English sentences," *Journal of the Acoustical Society of America*, 84, 70–79.