The Detection of Foreign Accent in Backwards Speech

Murray J. Munro*, Tracey M. Derwing† and Clifford S. Burgess*
*Simon Fraser University, Vancouver
†University of Alberta, Edmonton
E-mail: mjmunro@sfu.ca

ABSTRACT

In this paper we report the results of two foreign accent detection experiments. In both cases, native-English-speaking listeners heard reversed English speech samples from native speakers of English and Mandarin, and attempted to identify the samples that were Mandarin-accented. In general the listeners performed at above-chance levels. Moreover, successful performance on the task did not appear to depend on speaking rate differences that are known to occur between first and second language users. The results suggest that when detecting and evaluating foreign-accentedness, listeners can make use of properties of speech that are unrelated to semantic, grammatical, segmental, and prosodic goodness, but that remain available even in reversed speech.

1. INTRODUCTION

Even phonetically unsophisticated speakers of a language are quite adept at determining when their language is spoken by a non-native speaker. Work by Flege [1], for instance, has shown that foreign-accent detection can occur in very short speech samples. Listeners also seem to have little difficulty giving reliable ratings of the strength of a foreign accent. A number of studies have shown that straightforward counts of phonological errors (i.e., segmental insertions, deletions, and substitutions) in non-native utterances correlate positively with global accent ratings. In a study of oral reading passages Anderson-Hsieh, Johnson, and Koehler [2] observed a Pearson $\underline{r}$ of –.55 between the number of segmental errors and pronunciation goodness ratings. The correlation was even higher (–.74) when syllable structure errors were taken into account. In Munro and Derwing's [3] study of short excerpts of extemporaneous speech most untrained listeners' accent ratings correlated significantly with error counts. By using an index of accentedness determined by the frequencies of various types of errors in oral passages from Mexican-American ESL speakers, Brennan and Brennan [4] accounted for a very large part of the variance in accentedness ratings from trained listeners. Prosodic factors, such as rhythm and intonation have also been shown to influence accentedness judgments [2, 3].

However, factors other than segmental and prosodic goodness may also contribute to the detection and evaluation of accentedness. Our goal in this study was to explore the potency of such factors as cues to accentedness by assessing the accent detection ability of listeners presented with reversed speech.

Although connected backwards speech is readily recognizable as human speech, it contains little or no segmental or prosodic information that could presumably be used to identify a non-native speaker. When Black had phonetically-trained listeners transcribe pairs of reversed monosyllables, he found that some consonants and vowels were correctly transcribed [5]. However, connected backwards speech is not generally meaningful, and listeners are unlikely to have a concept of a segmentally "correct" backwards utterance. Therefore, typical phonemic substitutions, insertions, and deletions that betray a foreign accent would probably not be recognized as pronunciation errors. It is also unlikely that foreign-accented intonation contours or rhythmic patterns would be recognized. Nevertheless, Black [5] also noted that reversed speech has recognizable "timbre, rate, and naturalness and variability in pitch" (p. 173). It is conceivable that pitch, voice quality, and rate in backwards speech samples could be used by listeners in foreign accent detection. Modal pitch has been shown to vary across languages [6], and because it may be transferred from the first language to the second it may signal a foreign accent. In fact, it may have played a role in accent detection in a study by Munro [7]. Voice quality is also known to be a factor in accent. Its availability in backwards speech was discussed by Sherman [8], who noted improved validity in assessments of harshness and nasality in disordered speech when backwards rather than normal speech was presented to the evaluators. Munro and Derwing [9] have also observed small but significant correlations between speaking rates and listeners' foreign accent ratings. In general, even highly proficient second language users tend to speak at a slower rate than do native speakers. A speakers' rate may therefore be factored into an assessment of global accentedness. This study was conceived to determine whether listeners are able to detect a foreign accent on the basis of these kinds of potential cues.

2. EXPERIMENT 1

2.1. STIMULI

Twenty stimulus items were prepared, each item consisting of the first 36 words (six to ten seconds) of an intermediate-level ESL text that had been read aloud onto

ISBN 1-876346-48-5 © 2003 UAB

cassette tape in a sound-treated room. The speakers were ten native Mandarin speakers and ten native English speakers. The Mandarin speakers were all fluent users of English who had no difficulties reading the text. Nevertheless, they all exhibited a noticeable Mandarin accent in their English productions. The native English group spoke Canadian English.

All stimuli were recorded with the same audio equipment, digitized at 22 kHz (16-bit resolution), and normalized for peak amplitude. None of the recordings had obvious dysfluencies, such as false starts or hesitations. A waveform editing program was used to reverse each sample.

2.2 LISTENERS

The listeners were 16 native English speakers (eight female, eight male) recruited from undergraduate linguistics courses at Simon Fraser University. They ranged in age from 19 to 28 years, with a mean of 23.4 years. Six were fluent in a language other than English, but none were fluent in Mandarin or had Mandarin-speaking parents. All reported normal hearing.

2.3 PROCEDURE

During individual sessions, the listeners were instructed to make a binary judgment of each stimulus item, indicating whether the item was spoken with a Mandarin accent or a native English accent. They were first presented with a normal sample of a Mandarin speaker and a Canadian English speaker reading the English text so as to heighten their awareness of the differences between the two kinds of speech. They then carried out a practice task with four backwards speech samples. None of the sample stimuli were used in the actual experiment. The listeners indicated their responses by pressing one of two buttons marked "Mandarin" and "English" on the computer screen. After the practice session, each listener heard a different randomization of the 20 test stimuli. The test items were presented through headphones at a comfortable volume. Each participant took about five minutes to complete the self-paced task.

2.4 RESULTS

Each listener's score (in terms of hits and false alarms) was converted to an A' sensitivity score, which permits a fairly straightforward interpretation of the outcome of a binary judgment task [10]. An A' score of 1 indicates perfect sensitivity, while a score of .5 indicates performance at chance levels. The mean score for the 16 listeners was .788. Scores from the individual participants are shown in Figure 1.

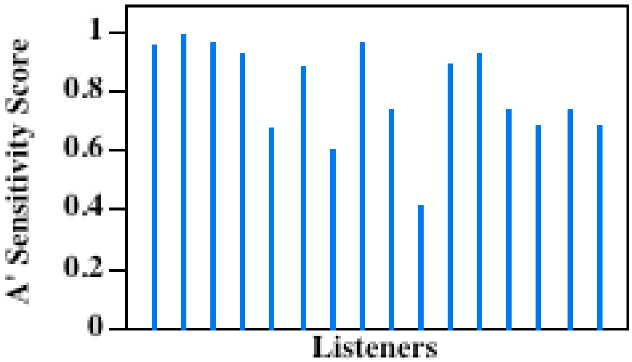

Figure 1: A' sensitivity scores for each of the 16 listeners in Experiment 1. A score of 1 indicates perfect sensitivity; a score of .5 indicates no sensitivity.

The data were submitted to a one-sample *t*-test to determine whether the listeners' A' scores differed significantly from a "chance" level of .5. The analysis revealed that they did [$t(15) = 7.109$, $p < .0001$]. (Note: In all statistical tests reported here, we adopted $p < .05$ as the criterion for significance.)

3. EXPERIMENT 2

The results of Experiment 1 indicate that the listeners were able to separate the stimuli reliably into native and foreign-accented categories. However, the basis for their judgments is unknown. Although it is possible that the listeners made use of cues to accentedness such as voice quality, it is also possible that their responses were the result of paying attention to superficial aspects of the stimuli that may have permitted categorization into two groups. For example, it is well-established that foreign-accented speech tends to be produced at a slower rate than native-produced speech [9]. If, as expected, the Mandarin-accented speech samples in Experiment 1 were noticeably slower than the native English samples, the listeners might have been able to score at above-chance levels on the task simply by paying attention to speaking rates. Experiment 2 was carried out to investigate this possibility.

3.1 STIMULI

The 20 stimulus items used in the previous experiment were modified so as to equate their durations. We first measured the durations of the original stimuli to the nearest 1 ms from the point of onset of the first word to the offset of the last. The mean duration for the Native English tokens was 6.880 s; for the Mandarin-accented tokens it was 8.735 s. Ranges for both groups are shown in Figure 2.

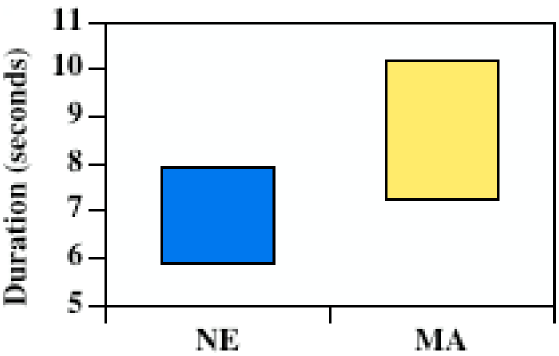

Figure 2: Duration ranges for the unmodified Native English (NE) and Mandarin-accented (MA) stimuli used in Experiment 1.

A comparison of the mean durations for the two groups revealed that the Mandarin-accented tokens were significantly longer [$t(18) = 4.549$, $p < .0002$] than the native English tokens. Using a waveform editing program with a "time-warping" function, we equalized the stimulus durations by adjusting all stimuli to the mean duration across the two groups (i.e., 7.8 s). This process introduced no evident changes in pitch to the stimuli. It entailed temporally compressing seven of the Mandarin-accented utterances and expanding eight of the native English ones. An aural inspection of the manipulated tokens indicated that the modifications had not introduced noticeable distortions to any of the stimuli that might make them stand out for the listeners.

3.2 LISTENERS

A new cohort of 27 listeners was recruited. This group consisted of 20 women and 7 men between 18 and 50 years of age. All were native speakers of English. Only one of the participants reported speaking a second language (French) fluently. However, she had not actually lived in a French-speaking area. Eleven of the participants reported that they spoke regularly with non-native speakers of English and had extensive experience with foreign-accented speech. All participants reported normal hearing.

3.3 PROCEDURE

The procedure was identical to that used in Experiment 1, except that the participants completed the task in a quiet room at a different location.

3.4 RESULTS

As before, each listener's score was converted to an A' value, and the data were submitted to a one-sample *t*-test to determine whether the listeners' A' scores differed significantly from .5. This time, the mean A' score was .722, which was indeed significantly better than chance [$t(26) = 8.755$, $p < .0001$]. Although the mean A' score in Experiment 2 appeared to be slightly lower than that obtained in Experiment 1, a further test comparing the performance of the listeners across the two experiments revealed that the difference failed to reach significance [$t(41) = 1.452$, $p = .1542$]. A final comparison, this time between the 11 listeners who reported regular exposure to accented speech (mean A' score = .773) and the 16 who did not (mean A' score = 687), also failed to yield a significant difference [$t(25) = 1.732$, $p = .0956$].

4. DISCUSSION AND CONCLUSIONS

In Experiment 1, our listeners correctly labelled backwards speech samples of about six to ten seconds in duration as native-produced or Mandarin-accented. In Experiment 2, we replicated this finding with a new group of listeners who heard stimuli that had been modified so as to equate the durations of the speech samples. When presented with the temporally-manipulated speech, they not only performed at above-chance levels, but were no less successful at the task than were the listeners in Experiment 1. Thus, we found no support for the proposal that the listeners in the first experiment had simply been attending to superficial differences in speaking rate in making their judgments. We also found no indication that listeners who reported regular exposure to foreign-accented speech performed any better at this accent detection task than those who had no such exposure.

These experiments are a step in gaining a better understanding of the ways in which listeners determine whether a particular individual shares their native language or not. Our results indicate that despite the very limited list of potential cues to non-nativeness that exist in backwards speech, listeners seem able to find enough information in such speech to detect a foreign accent. Although we cannot yet pinpoint the exact cues that our listeners used, it is quite possible that they attended to long-term properties of the speech samples such as voice quality.

Several directions for future research are suggested by these findings. First, we plan to replicate the study with speech samples from other groups of speakers to determine whether some types of foreign accents are more readily detected than others in backwards speech. Also, because of the limitations of the temporal compression and expansion techniques used in Experiment 2, it would be useful to carry out an experiment using stimuli produced naturally at similar rates of speech. Finally, work is clearly called for to establish the nature of the differences between the native and foreign-accented speech samples, in terms of such characteristics as voice quality.

ACKNOWLEDGMENTS

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada. The authors thank Kama Jamieson for her assistance with data collection.

REFERENCES

[1] J. E. Flege, "The detection of French accent by American listeners," *Journal of the Acoustical Society of America*, vol. 76, 692-707, 1984.

[2] J. Anderson-Hsieh, R. Johnson, and K. Koehler, "The relationship between native speaker judgments of

ISBN 1-876346-48-5 © 2003 UAB

nonnative pronunciation and deviance in segmentals, prosody, and syllable structure," *Language Learning*, vol. 42, 529-555, 1992.

[3] M. J. Munro and T. M. Derwing, "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners," *Language Learning*, vol. 49, Supplement 1, 285-310, 1999.

[4] E. Brennan and J. Brennan, "Accent scaling and language attitudes: Reactions to Mexican American English speech," *Language and Speech*, vol. 24, 207-221, 1981.

[5] J. W. Black, "The 'phonemic' content of backward-reproduced speech," *Journal of Speech and Hearing Research*, vol. 16, 165-174, 1973.

[6] W. Majewski, H. Hollien, and J. Zalewski, "Speaking fundamental frequency of Polish adult males," *Phonetica*, vol. 25, 119-125, 1972.

[7] M. J. Munro, "Non-segmental factors in foreign accent: Ratings of filtered speech," *Studies in Second Language Acquisition*, vol. 17, 17-34, 1995.

[8] D. Sherman. "The merits of backward playing of connected speech in the scaling of voice quality disorders," *Journal of Speech and Hearing Disorders*, vol. 19, 312-321, 1954.

[9] M. J. Munro and T. M. Derwing, "Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate," *Studies in Second Language Acquisition*, vol. 23, 451-468, 2001.

[10] N. A. MacMillan and C. D. Creelman, *Detection Theory: A User's Guide,* Cambridge MA: Cambridge University Press, 1991.