AN ACOUSTIC STUDY OF REAL AND IMAGINED FOREIGNER-DIRECTED SPEECH

Rebecca Scarborough[†], *Jason Brenier*[‡], *Yuan Zhao*[†], *Lauren Hall-Lew*[†], *Olga Dmitrieva*[†]

[†]Stanford University, [‡]University of Colorado
{rscar, yuanzhao, dialect, dmitro}@stanford.edu, jbrenier@colorado.edu

ABSTRACT

The acoustic properties of foreigner-directed speech are surprisingly understudied, and many existing studies evoke imagined interlocutors to elicit foreigner-directed speech. This study provides an acoustic comparison of foreigner-directed and native-directed speech in real and imaginary conditions. Ten native U.S. English speakers described the path between landmarks on a map to two confederate listeners (one native English speaker and one native Mandarin speaker) and to two imagined listeners (described as a native U.S. English speaker and a non-native speaker). Vowel duration, rate of speech, and vowel space size were examined across native/foreigner and real/imagined conditions. Stressed vowels were longer, rate of speech was slower, and vowel space distances were expanded in the foreigner-directed and imaginary conditions than in the native-directed and real ones. Speakers made acoustic-phonetic adjustments in foreigner-directed speech that are consistent with those seen in listener-directed clear speech, and these additional adjustments were made for both native and foreign listeners when the listener was imagined rather than real.

Keywords: foreigner-directed speech, clear speech, vowel hyperarticulation, rate of speech, vowel duration, imaginary listener

1. INTRODUCTION

Research has shown that speakers accommodate different communicative needs of their listeners. For instance, people talk more loudly and slowly in noisy environments than in quiet ones [8]. Similarly, speech directed toward hearing-impaired listeners is slower and has less phonological reduction (e.g., fewer reduced vowels and fewer unreleased word-final stops) than normal conversational speech [10]. Foreigner-directed speech (henceforth, FDS) is much cited as a similarly accommodative speech style. But though its syntactic and lexical properties have been reported (e.g., [6]), its acoustic properties are surprisingly understudied (cf [11], [12]). Furthermore, FDS, like other listener-directed speech styles, has often been elicited experimentally by asking a participant to speak to an imagined interlocutor, e.g., "speak clearly as if in a noisy room or to a hearing-impaired listener" [10] or "read as if speaking to a listener with a hearing loss or from a different language background" [4]. Thus, different types of special listeners and listening situations are not differentiated, and the accommodations they elicit are non-specific. And even to the extent that the acoustic properties of FDS are described, they are described with respect to a *hypothetical* foreign listener.

In light of this background, the current study seeks to describe the acoustic properties of speech directed to a non-native speaker. First, we ask whether there are listener-specific adaptations that characterize FDS, differentiating it from speech directed to a native speaker, and whether these properties are comparable to those found in other types of listener-directed speech. Second, we ask whether the speech elicited in an authentic foreigner-directed speech task is the same as speech elicited in hypothetical situations. By addressing these questions, we hope to better situate FDS in the broader context of clear speech.

2. METHODS

2.1. Participants

Ten participants (7 male, 3 female) took part in this study. All were native speakers of U.S. English. The participants were undergraduate students at Stanford University who received course credit for their participation. Each participant was also asked, post-test, for their amount of exposure to non-native English speakers.

Two confederates (both female) also took part in the experimental sessions. One confederate was

a native speaker of U.S. English; the other was a native speaker of Mandarin. The non-native confederate has been in the U.S. for less than three years and speaks noticeably Mandarin-accented English. Both confederates are authors on this paper.

2.2. Materials

One pair of maps (*q3ec6g* and *q3ec6f*) from the HCRC Map Task Corpus [1] was modified slightly to suit our task. In order to focus the participants' attention on the linguistic competence of their interlocutor, i.e., native vs. non-native speaker, we simplified the maps so that all landmarks were present on both the participants' maps and the confederate listeners' maps. Talk therefore focused only on the direction of the path, rather than negotiation of both the path and which landmarks were held in common. There were four different maps, i.e., maps with different paths through the same set of landmarks.

2.3. Procedure

Participants were asked to describe the route indicated on each map. Each participant gave directions four times, once in each of four experimental conditions, varying by interlocutor. In one condition, participants described the route to an *imagined foreigner*. In the second, participants described the route to an *imagined native speaker of English*. In the third, participants described the route to a *real confederate foreigner* who was actually present in the recording booth, and in the last condition, the participants described the route to a *real confederate native speaker*. The order of the conditions was varied across participants.

In the confederate conditions, the two interlocutors were seated across from one another, separated by a low divider which hid their maps but did not interfere with face-to-face visual contact.

All sessions were recorded to DAT in a sound-attenuated booth using an SM10A Shure head-mounted microphone worn by the participant.

2.4. Data preparation

The audio data was transferred to hard drive and downsampled to 22 kHz. The audio signal was then roughly manually segmented into intonation phrases of 1 to 20 words (ranging from approximately 20 to 600 ms) and orthographically transcribed using the Transcriber software package [2]. Only the participants' utterances were transcribed; confederates' speech and overlapping speech were ignored.

Timestamps were obtained from the hand-segmented utterance transcripts and used to generate forced word and phone alignments with the Sonic continuous speech recognition package [9]. Acoustic models and a language model trained on the Switchboard corpus of spontaneous conversational speech [7] were used for automatic alignment. The resulting corpus contains approximately 64 minutes of speech, consisting of 3315 intonation phrases, 13,901 words, and 43,751 phones.

2.5. Measures

All acoustic measures for our study were extracted from the alignment files using Praat [3]. All target words (map landmarks) and stressed vowels were identified automatically with hand-generated dictionaries.

Duration and formant values were measured for each vowel in the corpus. Only data for stressed vowels in target words are examined here. F1 and F2 were measured automatically at the vowel midpoint using a Burg LPC-based algorithm in Praat. Formant values of all test vowels were then visually inspected and corrected manually when necessary. Average pairwise distance for all vowels, [i]-[u]-[a] triangle area, and point vowel distances (i-u, i-a, and a-u) were calculated from the formant data for each speaker in each condition as measures of vowel space dispersion [5].

Rate of speech (words per second and phones per second) as well as average phone and word durations were calculated for each intonation phrase.

3. RESULTS

All measures were submitted to analysis by repeated measures ANOVAs with factors of Listener Language (native vs. foreign) and Task Authenticity (real vs. imaginary).

3.1. Rate of speech

Results from the analysis of phones per second (PPS) showed a significant main effect of Listener Language ($F[1,9]=7.399$, $p=.024$), due to the fact that speakers produced significantly fewer phones per second in the foreigner-directed speech conditions. Figure 1 shows average PPS across

conditions. An analysis of words per second (WPS) showed a significant main effect of Task Authenticity (F[1,9]=6.483, p=.031), indicating that speakers produced fewer words per second in conditions involving imaginary interlocutors. Figure 2 shows WPS across conditions. No significant interactions were found for either PPS or WPS.

Figure 1: Average number of phones per second.

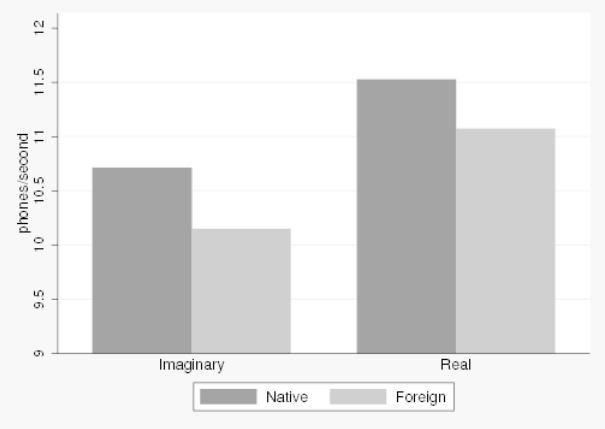
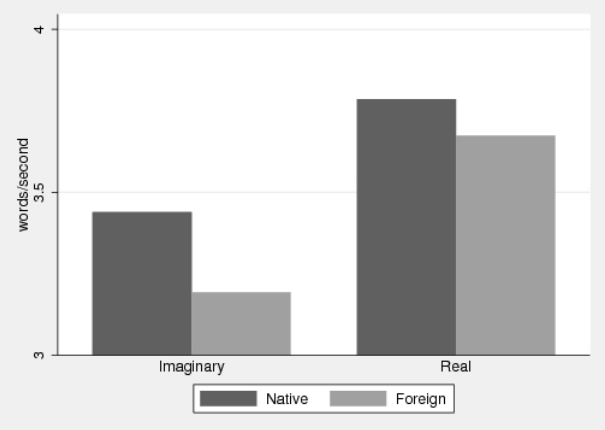


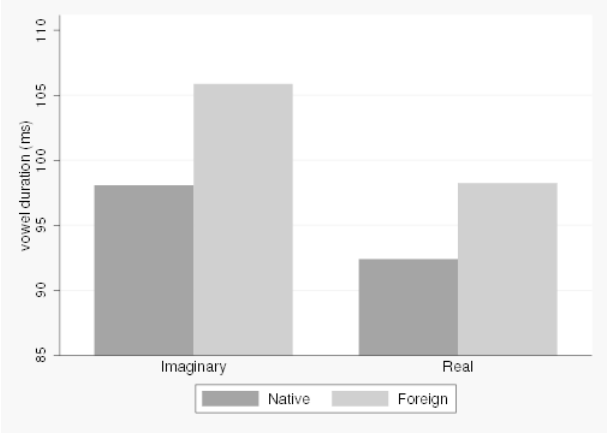
Figure 2: Average number of words per second.



3.2. Vowel duration

An analysis of mean vowel durations for stressed vowels in target words showed a significant effect of Listener Language (F[1,9]=5.834, p=.039). When talking to foreigners, speakers produced significantly longer vowels (M=102.1) than when speaking to native speakers (M=95.2). Figure 3 illustrates the pattern of average vowel duration across conditions.

Figure 3: Average vowel duration.



3.3. Vowel quality

Spectral hyperarticulation was also analyzed based on measures of average pairwise distance, vowel triangle area, and point vowel distances (i-u, i-a, and a-u). Average pairwise difference showed a significant effect of Task Authenticity (F[1,9]=4.582, p=.045) and a marginally significant effect of Listener Language (F[1,9]=3.511, p=.076), shown in Figure 4. These results indicate that the vowel space, illustrated in Figure 5, was more expanded for imagined and for foreign interlocutors. Neither triangle area nor point vowel distances showed significant effects for either Listener Language or Task Authenticity or their interaction.

The pattern for vowel area and point vowel distances, while not significant, were similar to the results for pairwise distance, suggesting hyperarticulation in both foreign and imaginary conditions.

Figure 4: Average pairwise distance.

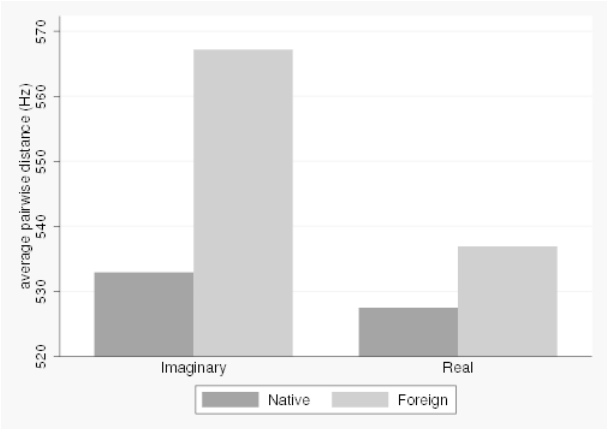
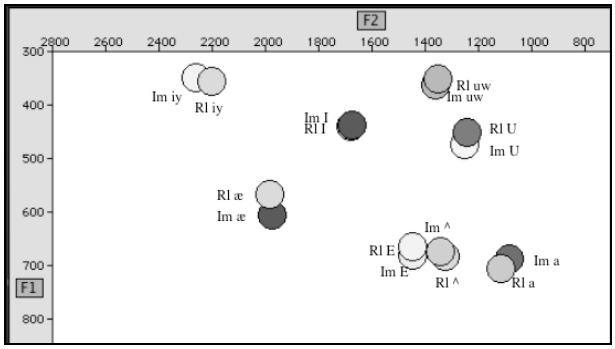


Figure 5: Acoustic vowel space.



4. DISCUSSION

This study compared acoustic properties of native U.S. English speakers' speech to native and non-native U.S. English listeners. It further compared their productions to real and imagined interlocutors.

Together, the two measures of rate of speech and the vowel duration results indicate that speakers adjusted conversational tempo according to the linguistic experience of their listeners. They talked more slowly and with longer stressed vowels to foreigners than to native speakers. In addition, the average pairwise distance between vowels was marginally larger for foreign interlocutors. (Interestingly, speech rate correlates of FDS were not found to be significant in [12], though spectral hyperarticulation was.) These data provide evidence that FDS is indeed an acoustically distinct speech style from standard native-directed speech. The adjustments are consistent with those seen in other listener-directed speech styles, such as speech in noise or speech to the hearing impaired, which also show decreased rate of speech, increased phone duration, and expanded vowel space in the accommodative condition [10]. In these cases, as in the current study, speakers produced a signal that was "clearer" and presumably easier to process when speaking to listeners who may have had extra processing difficulties (in the current case, due to limited language experience).

Speakers also produced longer, slower speech with more expanded vowels when speaking to imagined as opposed to real listeners. These data suggest that interpersonal interactions speed up the tempo of a speech event. Furthermore, they indicate that, although the patterns of accommodation are similar, FDS (and possibly other listener-directed styles as well) elicited in the absence of a real listener is not in fact acoustically identical to genuine FDS (or listener-directed clear speech). We take our findings to indicate that methodologies for future studies of clear or listener-directed speech should involve communicatively authentic elicitation tasks.

We predict that future analysis of this or similar data will reveal additional distinctive acoustic properties of listener-directed speech in FDS, e.g., higher mean f0 or greater f0 range and less phonological reduction (segment deletion or unreleased final stops).

5. REFERENCES

[1] Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G.M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S. and Weinert, R. 1991. The HCRC Map Task Corpus. *Language and Speech 34, 351-366.*

[2] Barras, C., Geoffrois, E., Wu, Z., Liberman, M. 2001. Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication* 33, 5-22.

[3] Boersma, P., Weenink, D. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 341-345.

[4] Bradlow, A. 2002. Confluent talker-and listener-related forces in clear speech production. *Laboratory Phonology* 7, 241–273.

[5] Bradlow, A.R., Torretta, G.M., Pisoni, D.B. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20, 255-272.

[6] Ferguson, C. 1975. Towards a characterization of English foreigner talk. *Anthropological Linguistics* 17, 1–17.

[7] Godfrey, J., Holliman, E., McDaniel, J. 1992. SWITCHBOARD: telephone speech corpus for research and development. *Proc. ICASSP*, 517-520.

[8] Lombard, E. 1911. Le signe de l'élévation de la voix. *Annales des maladies de l'oreille, du larynx, du nez et du pharynx* 37, 101-119.

[9] Pellom, B. 2001. Sonic: The University of Colorado Continuous Speech Recognizer. *Center for Spoken Language Research*, University of Colorado, Boulder, CO Rep. TR-CSLR-2001-01.

[10] Picheny, M.A., Durlach, N.I., Braida, L.D. 1986. Speaking Clearly for the Hard of Hearing II Acoustic Characteristics of Clear and Conversational Speech. *JSHR* 29, 434-436.

[11] Smith, C. 2006. Investigating speaker adaptation to listeners. Presentation at the *Linguistics Society of America*, Albuquerque, New Mexico.

[12] Uther, M., Knoll, M.A., Burnham, D. 2007. Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication* 49, 2-7.