LANGUAGE- AND TALKER-DEPENDENT VARIATION IN GLOBAL FEATURES OF NATIVE AND NON-NATIVE SPEECH

Ann R. Bradlow, Lauren Ackerman, L. Ann Burchfield, Lisa Hesterberg, Jenna Luque & Kelsey Mok

Department of Linguistics, Northwestern University, USA
abradlow@northwestern.edu; lmackerman@u.northwestern.edu; ann@u.northwestern.edu; lisahesterberg2013@u.northwestern.edu; kelseymok@u.northwestern.edu; SLPJenna@gmail.com

ABSTRACT

We motivate and present a corpus of scripted and spontaneous speech in both the native and the non-native language of talkers from various language backgrounds. Using corpus recordings from 11 native English and 11 late Mandarin-English bilinguals we compared speech timing across native English, native Mandarin, and Mandarin-accented English. Findings showed similarities across native Mandarin and native English in speaking rate and in reduction of the number of acoustic relative to orthographic syllables. The two languages differed in silence-to-speech ratio and in the number of words between pauses, possibly reflecting phrase-level structural differences between English and Mandarin. Non-native English had a significantly slower speaking rate and lower rate of syllable reduction than both native English and native Mandarin. But, non-native English was similar to native English in terms of silence-to-speech ratio and was similar to native Mandarin in terms of words per pause. Finally, some talker-specificity in terms of (non)optimal speech timing appeared to transfer from native to non-native speech within the Mandarin-English bilinguals. These findings provide an empirical base for testing how language-dependent, structural features combine with general features of non-native speech production and with talker-dependent features in determining foreign-language speech production.

Keywords: cross-language, second-language, speech timing, bilingualism, multi-lingual corpus

1. INTRODUCTION

The traditional approach to analyzing foreign-language speech emphasizes a comparison of the sound systems of the languages in question, seeking aspects of compatibility and conflict in terms of contrastive categories at sub-lexical, lexical, and prosodic levels. For instance, models of cross- and second-language speech perception and production e.g. [1, 3, 8] offer accounts for variation in the ease of acquisition of specific sound contrasts by individuals from specific native language backgrounds in terms of the relationship between the sound structures of the source and target languages.

In addition to source-vs.-target language structural (mis)matches, there are two other sources of variability in speech intelligibility between native and nonnative talkers. First, the cognitive-linguistic demands of non-native language production and perception may impact non-native speech regardless of the bilingual's native language and the target language. For example, speaking rate is typically slower in non-native than native speech (e.g. [4]). Second, even within monolingual, native talkers there is substantial variation in global acoustic-phonetic features (e.g. speaking rate, overall clarity) [2, 5]. Thus, there may be transfer of these talker-dependent features from native to non-native speech making inter-talker variation in native speech a significant predictor of non-native speech variation.

In our work, we attempt to examine all three sources of foreign-accented speech variability – language-dependent structure, general non-native speech production and perception features, and talker-dependent individual characteristics – simultaneously as each likely makes a crucial contribution to the cognitive representations and processes involved in speech communication in multi-lingual settings. This approach requires speech corpora that include both native and non-native speech recordings by talkers from various native language backgrounds. In this paper, we present such a corpus, and initial analyses of global-level speech timing across English and Mandarin (i.e. across native speakers of each language), across native English and Mandarin-accented English (i.e. across native and non-native speakers of English), and across native Mandarin and Mandarin-accented English (i.e. across the native and non-native languages within individual bilinguals).

2. THE ALLSSTAR CORPUS

2.1. Materials

Recording in the ALLSSTAR corpus (Archive of L1 and L2 Scripted and Spontaneous Transcripts and Recordings) include scripted materials, consisting of sentences and a paragraph-long passage. The corpus also includes spontaneous speech in response to two types of prompts: four simple picture story narratives and six open-ended questions (see Table 1 below).

Table 1: Materials in the ALLSSTAR Corpus.

Type	Source	Languages
Sentences	Hearing in Noise Test (HINT) [16], n=120.	Available in English, Mandarin, French, Japanese, Korean, Portuguese, Spanish, Turkish.
	From *The Little Prince* [15], n=30.	Available in all languages except Gishu, Gujarati.
	From *The Universal Declaration of Human Rights* [17], n=20.	Available in all languages except Gishu, Gujarati.
Paragraph	*The North Wind and the Sun* reading passage [6].	Available in all languages except Gishu, Gujarati.
Spontaneous speech	Picture story narratives [9, 10, 11, 12]	All languages.
	Open response questions, ~5 minutes	All languages.

2.2. Talkers and recording procedure

All native and nonnative talkers were recruited from the student population at Northwestern University. All were paid for their participation or received course credit. All reported normal speech and hearing at the time of testing, and were 18-34 years of age. Participants were recorded in a sound-treated booth. They spoke into a Shure SM81 Condenser Microphone and their speech was recorded using a Marantz PMD 670 flash recorder. All talkers completed a language background questionnaire and performed an English sentence-in-noise recognition test (the HINT test, [16]) before beginning the recording of the sentences, paragraph and spontaneous speech in English. The nonnative talkers returned the following day for a second recording session during which they recorded the sentences, paragraph and spontaneous speech recordings in their native language (where available). All scripted materials were presented in the standard orthography of the language. Each session took approximately 1-1.5 hours. In addition, all non-native talkers took a commercial test of English proficiency (Pearson's VERSANT test, [14]) at a separate time. To date, we have complete sets of recordings from 50 talkers from 18 language backgrounds (Table 2) in addition to 20 native English talkers. Additional recordings by more talkers in the sparsely represented languages and by talkers of additional languages are ongoing.

Table 2: ALLSSTAR talkers recorded to date.

Language (number of talkers)	
Chinese (Mandarin) n=18)	Hebrew (n=1)
Chinese (Taiwanese) (n=2)	Hindi (n=2)
Chinese (Singapore) (n=1)	Japanese (n=2)
Farsi (n=2)	Korean (n=2)
French (n=1)	Portuguese (Brazilian) (n=4)
German (n=1)	Russian (n=2)
Greek (n=1)	Spanish (n=2)
Gujarati (n=1)	Turkish (n=6)
Gishu (n=1)	Vietnamese (n=1)
Native English (n=20)	

2.3. Corpus access and storage

As they become available, transcriptions of the speech recordings are force-aligned, and stored with the digital speech files in a web-based speech recording archive, OSCAAR: Online speech corpus archive & analysis resource ([13]). To date, most of the Mandarin, English and Mandarin-accented English materials have been transcribed and aligned.

3. SPEECH TIMING ANALYSIS

We explored features of speech timing at a global (rather than local phonetic) level in the ALLSSTAR paragraph recordings (NWS) by native English (n=11) and native Mandarin talkers (n=11). These analyses allowed us to compare global timing across native English and native Mandarin, native English and Mandarin-accented English, and native Mandarin and Mandarin-accented English within these Mandarin-English bilinguals.

3.1. Speech timing measures

For each paragraph recording, we obtained four speech timing measures. First, we calculated the number of syllables per second based on the number of orthographic syllables divided by the duration of the paragraph recording with major disfluencies (e.g. coughs) excluded. The orthographic syllable count for the English NWS recording was based on the expected, standard citation form reading of the paragraph which yielded a count of 141 syllables. For the Mandarin paragraph the number of orthographic syllables was

based on the number of Chinese characters in the text of the Mandarin NWS passage which yielded a count of 162 syllable-like units. Second, we calculated the proportion of the total paragraph reading duration (excluding major disfluencies) that was devoted to inter-word pauses of at least 100 milliseconds. Third, we compared the number of orthographic syllables to the number of acoustic syllables in each talker's recordings. The number of acoustic syllables was obtained using an automatic syllable counting algorithm implemented as a Praat script [7]. This algorithm counts peaks in intensity that are preceded and followed by dips in intensity. Substantial phonetic reduction processes would be reflected by a decrease in the number of acoustic syllables compared to the number of orthographic syllables, calculated as (acoustic minus orthographic) / orthographic. Finally, we counted the number of words per silent pause as a measure of "chunking" in the paragraph reading. Note that while the Mandarin paragraph contained more words than the English paragraph (133 vs. 113), the average number of syllables per word and segments per syllable were similar (1.25 and 1. 22 syllables/word, 2.7 and 2.6 segments/syllable for English and Mandarin, respectively).

3.2. Results

Table 3 shows the four speech timing measures for the three sets of paragraph recordings (native English, native Mandarin, and Mandarin-accented English). For each measure, we compared (1) native English and native Mandarin, (2) native English and Mandarin-accented English, and (3) native Mandarin and Mandarin-accented English by means of two-tailed t-tests. For comparisons within Mandarin-English bilinguals, i.e. (3), we conducted paired t-tests; for other comparisons, i.e. (1) and (2), we conducted unpaired t-tests. The alpha level was p=0.004 (to correct for the total of 12 comparisons).

3.2.1. Native English vs. native Mandarin

In terms of orthographic syllables per second (including inter-word silent pauses, excluding major disfluencies), we found no significant difference between English and Mandarin (orthographic syllable rates of 4.44 and 4.36, for English and Mandarin respectively). Furthermore, English and Mandarin did not differ in terms of reduction from the orthographic syllable count to the acoustic syllable count (-16.3% and -17.3%, for English and Mandarin respectively). In contrast, the proportion of silence in the English paragraphs tended to be smaller than the proportion of silence in the Mandarin paragraphs (16.3% vs. 25.2% for English and Mandarin, respectively; t(10)=3.525, p=0.005). While the individual silences in English and Mandarin did not differ in average duration (approx. 0.525 msec. for both.), the average number of words between pauses tended to be larger in English than in Mandarin (11.6 and 7.8, respectively; t(10)=3.536, p=0.005), possibly reflecting a difference of language structure at the word and phrase levels.

Table 3: Means and standard deviations (in parentheses) for four speech timing measures in native English (n=11), native Mandarin (n=11), and Mandarin-accented English recordings.

	orthographic syll./sec.	% Silence	% syllable reduction (acoust-ortho)/ ortho)	Words per pause
Eng.	4.44 (0.32)	16.3 (3.8)	-16.25 (5.6)	11.6 (0.09)
Mand.	4.36 (0.81)	25.2 (8.1)	-17.23 (9.5)	7.8 (0.12)
Mand.-Eng.	3.07 (0.43)	19.1 (4.9)	+1.03 (9.9)	7.0 (0.21)

3.2.2. Native English vs. Mandarin-accented English

The number of orthographic syllables per second was significantly different in native and Mandarin-accented English (4.44 vs. 3.07, respectively; t(10)=13.93, p<.0001), replicating the well-known slower speaking rate of non-native versus native speech (e.g. [15]). There was also a significant difference between native and Mandarin-accented English in terms of the acoustic versus orthographic syllable count difference (change of -16.3% and +1.03%, for native and Mandarin-accented English respectively; t(10)=6.05, p<.0001). This suggests that non-native, Mandarin-accented English may exhibit substantially less syllable reduction than native English. Indeed the non-native English recordings occasionally displayed syllable addition. The proportion of silence did not differ across native and Mandarin-accented English. However the number of words between pauses was greater for native than Mandarin-accented English (11.6 and 7.0, respectively; t(10)=4.577, p=.001), suggesting that the non-native talkers tended to "chunk" the reading passage into smaller sections than the native talkers. This may be related to planning difficulties associated with foreign-language production. However, whether this is a feature of non-native speech in general (regardless of the particular native and non-native language) or related to these talkers' native language (i.e.

Mandarin for which we observed a similar number of words per pause), remains to be tested with the other language recordings in the corpus.

3.2.3. Native Mandarin vs. Mandarin-accented English

The Mandarin-English bilinguals had a significantly slower speaking rate in the non-native than native language (3.07 vs. 4.36, respectively; $t(10)=4.62$, $p=0.001$), and significantly greater syllable reduction in their native Mandarin than in their non-native English as expressed by % syllable change across the orthographic and acoustic syllable counts (-17.25% vs. +1.03% for native Mandarin and non-native English, respectively; $t(10)=3.99$, $p=0.003$). These talkers' Mandarin and English paragraphs did not differ in % silence and words per pause.

Finally, within this relatively small sample of Mandarin-English bilinguals, we explored whether some of the variation in non-native speech timing could be accounted for by variation in native speech timing. Across the four speech timing measures, three showed no within-talker correlation between the native Mandarin and non-native English paragraph readings (orthographic syllable rate, % silence, and words per pause). In contrast, we observed a weak negative correlation (correlation coeff.= -0.38) for % syllable reduction in native Mandarin and non-native English. This indicates that talkers with relatively high rates of syllable reduction in their native Mandarin tended to exhibit relatively low rates of syllable reduction and in many cases (6 out of 11 talkers) actually exhibited syllable addition in their non-native English. While this relationship remains to be explored further with a larger sample of bilinguals we can speculate that relatively casual native speech and (excessively) hyper-articulated non-native speech are both signs of efficient talker-to-listener accommodation: native speech to an assumed native listener can afford to be very hypo-articulated while non-native speech may benefit from hyper-articulation.

4. SUMMARY AND CONCLUSIONS

This exploration of speech timing at global level suggests that native Mandarin and native English may be similar in rates of speech and connected speech reduction, but differ in the pause-to-speech ratio and number of words per pause, possibly reflecting word and/or phrase structure differences across English and Mandarin. Mandarin-accented English had a significantly slower speaking rate and lower rate of syllable reduction than both native English and native Mandarin. But, Mandarin-accented English was similar to native English in silence-to-speech ratio, and was similar to native Mandarin in terms of words per pause. Finally, some talker-specificity in terms of (non)optimal speech timing may transfer from native to non-native speech. Our next step is to expand our analyses to other materials (especially spontaneous speech), to other languages, and to other global phonetic parameters. This will ultimately provide the basis for hypothesis-driven experiments that advance our understanding of the confluence of language- and talker-dependency in native and non-native speech production and perception.

5. REFERENCES

[1] Best, C., McRoberts, G., Goodell, E. 2001. American listeners' perception of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoust. Soc. Amer.* 109, 775-794.

[2] 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.

[3] Flege, J.E. 1995. Second language speech learning: theory, findings, and problems. In Strange, W. (ed.), *Speech Perception and Linguistics Experience. Issues in Cross-Language Research*. Timonium, MD: York Press, 229-273.

[4] Guion, S.G., Flege, J.E., Liu, S., Yeni-Komshian, G. 2000. Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics* 21, 205-228.

[5] Hazan, V., Markham, D. 2004. Acoustic-phonetic correlates of talker intelligibility for adults and children. *J. Acoust. Soc. Amer.* 116(5), 3108-3118.

[6] International Phonetic Association. 1999. *The Handbook of the International Phonetic Association. 1999.* Cambridge: Cambridge University Press.

[7] de Jong, N.H., Wempe, T. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41(2), 385-390.

[8] Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T. 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded. *Phil. Trans. Roy. Soc. B* 363, 979-1000.

[9] Mayer, M. 1973. *Bubble Bubble.* New York: Parents' Magazine Press.

[10] Mayer, M. 1974. Bear's new clothes. *Two Moral Tales*. New York: Four Winds Press.

[11] Mayer, M. 1974. Bird's new hat. *Two Moral Tales*. New York: Four Winds Press.

[12] Mayer, M. 1974. Just a pig at heart. *Two Moral Tales*. New York: Four Winds Press.

[13] OSCAAR: Online speech corpus archive & analysis resource. Northwestern University. *http://oscaar.ling.northwestern.edu/*

[14] Pearson, Knowledge Technologies group, VERSANT Test of Spoken English. *http://www.ordinate.com/*

[15] de Saint-Exupéry, A. 1943. *Le Petit Prince*. Harcourt Inc.

[16] Soli, S.D, Wong, L.L.N. 2008. Assessment of speech intelligibility in noise with the hearing in noise test. *Intl. J. Audiology* 47, 356-361.

[17] United Nations. 1948. The universal declaration of human rights. *http://www.un.org/en/documents/udhr/index.shtml*