

Applications of the Buckeye GTA Corpus for L2 teaching and research

Jocelyn B. Hardman¹, Elizabeth McCullough²

¹ Department of Languages, University of Dayton, USA

² Department of Linguistics, Ohio State University, USA

hardmajb@notes.udayton.edu, eam@ling.ohio-state.edu

Abstract

The Buckeye GTA Corpus contains 9,664 L1 and L2 sentence productions by 89 talkers (27 American English, 19 Hindi, 23 Mandarin, & 20 Korean). A total of 5,696 sentences were read in English, with each talker contributing 64 sentences. Hindi, Mandarin, and Korean talkers also read 64 sentences each in their native languages, contributing a total of 3,968 sentences. Potential uses of the corpus are illustrated by research projects on classroom communication and acoustic phonetic patterns. These projects demonstrate how investigations in different disciplines can make use of the same corpus and provide converging data on second language phonological acquisition. **Index Terms:** corpus studies, pronunciation, intelligibility, phonetics, SLA, L2 speech production and perception

1. Introduction

Scholars of World Englishes and English as a Lingua Franca have argued for more investigation of intelligibility from the perspective of second language (L2) English interlocutors [1, 2, 3]. The increasing numbers of international students in the US requires ongoing assessment of standards for L2 English intelligibility. International Teaching Assistants have been researched in some depth, but very few studies have examined the oral English needs of international graduate students in their roles as students and researchers. These students are expected to participate orally in class, present at conferences, and conduct research in collaboration with international and American colleagues.

In fact, many US graduate schools now enroll a majority of international students in math, the sciences, and engineering. However, few studies have investigated the intelligibility of L2 English speech to these international students. The first author's study addresses this research gap by examining the intelligibility of Chinese graduate students as determined by their international peers. The second author's study examines the extent to which particular Hindi and Mandarin listeners in the first study exhibited native (L1) English patterns of contrast in their spoken English productions.

2. The Buckeye GTA Corpus

The Buckeye GTA Corpus contains 9,664 L1 and L2 sentences by 89 talkers (27 American, 19 Hindi, 23 Mandarin, & 20 Korean). A total of 5,696 BKB-R sentences were read in English, with each talker contributing 64 sentences. Hindi, Mandarin, and Korean talkers also read 64 sentences each in their native languages, contributing a total of 3,968 sentences. The breakdown of L1 and gender of the talkers recorded is provided in Table 1.

Table 1: Talkers in the Buckeye GTA Corpus

L1	Male	Female
American English	13	14
Hindi	12	7
Mandarin	11	12
Korean	2	18

2.1. Location and talkers

The corpus was collected on the main campus of a large public Midwestern university, where the international student enrollment is among the top 15 in the country and reflects the three foreign nationalities (Indian, Chinese, and Korean) most common on US campuses today [4]. The four L1s (Hindi, Mandarin, Korean, and American English) were selected because they constitute the top four L1s of students at US universities. All international talkers were graduate students who were certified as TAs (SPEAK score of 230 or higher out of 300, or an "unconditional pass" on the university's Mock Teaching Test) or had scored 26 or higher on the iBT TOEFL speaking section.

2.2. Materials and recording methods

The sentences included in the corpus are Lists 7 through 10 of the Bamford-Kowal-Bench Standard Sentence Test, revised for American English (BKB-R) [5]. BKB-R sentences are syntactically simple, and each contains 3-4 highly familiar key words, with a total of 50 key words per list.

Prior to recording, each participant filled out a language background questionnaire, reporting on their major, age at testing, city/region of origin, age of arrival in the US, age of first instruction, total length of English instruction, length of residence in the US, L1 and L2 usage, and familiarity with foreign-accented and Mandarin-accented English.

After completing separate perception and word familiarity rating tasks, each participant was seated in a sound-attenuated booth and provided with 64 printed English sentences. Each of the sentences was read aloud into a Shure SM10-A head-mounted microphone connected to a Symetrix SX302 dual microphone pre-amplifier. The sampling rate used for recording was 22050 kHz. L2 English participants were also recorded reading translations of each of these 64 sentences in their native language (Hindi, Mandarin, or Korean).

As more researchers begin to use the corpus for various projects, additional recordings and annotations will be made. The two studies below provide word familiarity ratings and intelligibility scores, and phonetic transcriptions and certain acoustic measurements, respectively.

3. Illustrative Project for Teaching

3.1. Research questions

Using Mandarin talkers from the Buckeye GTA Corpus, the first author sought to determine to what extent intelligibility was affected by the talkers' segmental pronunciation accuracy and L1, and the listeners' word familiarity and L1 [6].

3.2. Methods

The sentences from the first 6 male talkers recorded (3 American English L1, 3 Mandarin L1) were used as the audio stimuli in a word-recognition-in-noise experiment. The sentences were equalized in amplitude, mixed with white noise at a +5-dB signal-to-noise ratio, and 60 were presented in a counterbalanced design to listeners at a comfortable listening level (65 dB). A total of 72 listeners (18 American English L1, 18 Hindi L1, 18 Mandarin L1, & 18 Korean L1) orthographically transcribed each sentence. These listeners also contributed audio recordings of 64 sentences each in their L1 and L2 to the Buckeye GTA Corpus. The talkers' intelligibility scores were calculated based on accurate transcriptions of the key words in each sentence.

3.3. Results

A barplot of key word percent correct intelligibility rates by talker L1 and listener L1 is presented in Figure 1.

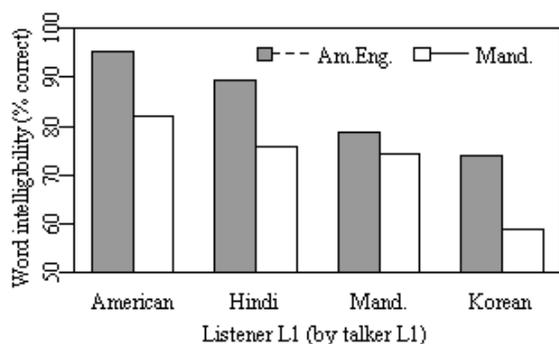


Figure 1: Intelligibility by talker L1 and listener L1

All listeners found the American English talkers to be more intelligible overall than the Mandarin talkers. The Mandarin talkers were most easily understood by the American listeners (82%) and most difficult to understand for Korean listeners (58%). Listeners with American English and Hindi L1s appeared to be more successful in understanding American English (95% and 89%, respectively) than were those with Mandarin and Korean L1s (78% and 73%, respectively). American listeners were most successful when hearing American English and performed nearly at ceiling (95%).

An interaction between talker L1 and listener L1 can clearly be seen for the Mandarin listeners when listening to Mandarin talkers. Since listeners with Mandarin and Korean L1s performed nearly equivalently when listening to American English (78% and 73%, respectively), they were apparently at equivalent levels of oral proficiency in English. However, when listening to Mandarin-accented English, Mandarin listeners (74%) performed much more successfully than did Korean listeners (58%). The difficulties of L2 talker-L2

listener communication are shown most clearly by these Korean listeners, who understood only slightly more than half the Mandarin-accented words.

Statistical analyses using a series of mixed-effects logistic regression models revealed that talker L1, listener L1, and listener word familiarity were significant predictors of intelligibility. Talker segmental accuracy at the "TA-certified" proficiency level was not a significant predictor of intelligibility.

3.4. Discussion

The finding that an L1 match between talker and listener was a benefit to intelligibility is important for English Language Teaching, since most teachers are either native speakers of English or share the L1 of their students. The first author recommends including more diversity of accents in the audio and video materials for international students preparing to study or conduct research in English L1 countries.

For effective communication in the academic sphere, international and American students alike would benefit from discipline-specific vocabulary building and linguistic training to improve their ability to accommodate the accent diversity that has become a reality of higher education today.

4. Illustrative Project for Research

4.1. Research questions

The second study is a pilot for a larger project on acoustic timing patterns in speech samples from the Buckeye GTA Corpus [7]. This study examines phonetic patterns in productions of voiced versus voiceless stops, and the extent to which L2 talkers exhibit native-like patterns of contrast. Stops were chosen as the segments of interest because they display a number of easily measurable and well-studied phonetic properties.

Some foreign accent research focuses on segmental differences between a talker's L1 and L2, and the ensuing effects on the phonetics of production in the L2. That is, a foreign accent can arise when a talker's L2 contains phonemes that his L1 does not. Because the talker is not accustomed to producing these sounds, his initial productions often differ phonetically from those of native talkers of the L2. Subsequent developments in the production of this new phoneme are of great research interest.

However, it is maintained that "we cannot account for foreign accent strictly in terms of segmental phonemic [...] differences between languages, for even non-segmental differences in temporal implementation carry over from one language to another" [8]. Although cross-linguistic differences in phoneme inventories are certainly one source of foreign accent, additional explanations lie at levels below the segment. For instance, while French and English both have /t/ phonemes, the VOT values of /t/ are shorter in French than in English. An L1 English talker who uses English /t/ VOT values when producing French /t/ will not sound like an L1 French talker [9].

As phonetic categories are always variable to some degree, even in L1 talkers, why might such differences in the productions of an L2 talker matter? One possibility is that an L2 talker who is not exhibiting native-like implementation of segments is, as a result, not making native-like contrasts between similar segments, which could have serious implications for his intelligibility in the L2. For instance, if an L1 French talker exhibited VOT values for French /t/ in

productions of English /t/, his productions of English /t/ may be confused with his productions of English /d/ by English listeners.

4.2. Methods

For this pilot study, measurements were made on the English productions of 9 male talkers in the Buckeye GTA Corpus (3 American English L1, 3 Hindi L1, and 3 Mandarin L1). These Hindi and Mandarin talkers also contributed intelligibility data in the study described above. The 64 sentences recorded by each talker contained a total of 195 stop targets. For each target, measurements were made manually by examining the waveform and spectrogram.

4.3. Results

As word onsets are very important to speech recognition [10], the preliminary data for VOT in word-initial position is reported here. Histograms for all 9 talkers are presented in Figure 2. Based on previous studies, it was expected that native English talkers would generally contrast short lag voicing with long lag voicing. Not surprisingly, this was exactly the pattern that was found. The native English talkers displayed voiced and voiceless distributions that overlap slightly in the short lag VOT range, and produced only a few instances of lead voicing.

The distributions of the productions of L1 Hindi talkers looked rather different due to the prevalence of lead voicing in voiced tokens. Thus, the token distributions prevented this pattern from being fully native-like. It is likely that the large

VOT range in these talkers' productions was a result of their experience with prevoiced stops in Hindi.

Overall, the patterns of the L1 Mandarin talkers were much more native-like, as they generally realized the English contrast as one of short lag VOT versus long lag VOT. Talker M2 exhibited some prevoiced productions, but fewer than any of the L1 Hindi talkers.

4.4. Discussion

The data examined in this study showed that in some cases, L2 English talkers display stop timing patterns that differ from those produced by L1 English talkers. Overall, including additional data not presented here, L1 Hindi talkers tended to look least native-like with respect to VOT, while L1 Mandarin talkers deviated most from native talkers in the patterns of preceding vowel duration and closure duration. In many cases, non-native deviations from native patterns could be explained by patterns in a talker's L1.

However, non-native talkers with the same language did not always pattern together. For instance, although talkers M1 and M3 showed a relatively native-like pattern for word-initial VOT, talker M2 deviated slightly from the pattern by having more prevoiced productions. Production patterns are certainly influenced by a talker's L1, but to some extent they also vary on the level of the individual.

This pilot project included data from only 9 of the 89 talkers in the Buckeye GTA Corpus. The study is currently being extended to the remaining 80 talkers.

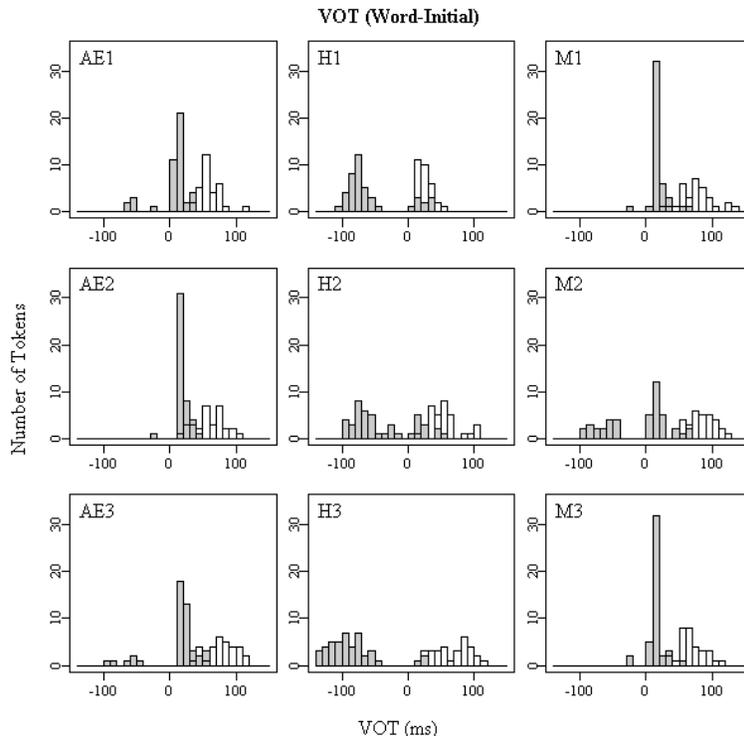


Figure 2: Word-initial VOT production patterns

5. Applications of the Buckeye GTA Corpus

5.1. Applications to teaching

The Buckeye GTA Corpus has many potential uses for both teaching and research. In the language classroom, the recordings of International Teaching Assistants may serve as targets for pronunciation teaching, especially for students who share the same L1. As models of the L2 English likely to be encountered in a US university classroom, the recordings may also be used to enhance listening comprehension for both international and American students. By using the recordings for focused linguistic training, students who have not been exposed to a variety of accents can gain more experience with the accents that they are most likely to encounter in an American university.

For Applied Linguistics research purposes, the Buckeye GTA Corpus can be used in further intelligibility studies. Results from these studies can help tailor pronunciation and listening comprehension teaching to meet the needs of students at US universities. Careful transcriptions and acoustic analyses of the recordings compared to the listeners' key word transcriptions can reveal English L1 and L2 speech perceptions of each listener L1 group. These perceptions and intelligibility scores will reveal the features of L2 English speech that negatively impact intelligibility for each listener L1.

5.2. Applications to research

Although it was developed primarily to address concerns about intelligibility in academia, the Buckeye GTA Corpus is also valuable for scholars who are not directly concerned with educational applications. For instance, the recordings, or words or sounds excised from them, may be used as audio stimuli in any of a number of perception tasks. In addition to having four different L1 groups, each L1 group contains at least 19 talkers, allowing for much talker variability.

The study of speech production, as well, may greatly benefit from the use of this corpus. A fundamental step of any acoustic phonetic research program requires collecting high-quality recordings of the type of speech of interest. While using an existing set of recordings gives the researcher less control of the material recorded, such an option saves a considerable amount of time and resources if an appropriate corpus can be found.

The Buckeye GTA Corpus consists of read sentence lists which are balanced syntactically and morphologically, but not phonetically. As such, it is clearly not appropriate for all phonetic investigations. However, all talkers read the same 64 sentences. Thus, although the phonological contexts are not controlled, this is alleviated somewhat by having an identical sample of speech from each talker. In addition, the availability of comparable recordings in the L1 of each L2 English talker allows for direct phonetic comparisons between an individual talker's L1 and L2 productions.

The value of taking advantage of existing corpora cannot be understated. While the Buckeye GTA Corpus does have its limitations, it has numerous potential applications in language research, particularly when recruiting and recording a large number of L2 English talkers is not logistically feasible.

6. Conclusion

The two studies described above used the Buckeye GTA Corpus for very different purposes, but both contribute analyses of the corpus which shed light on the processes of second language phonological acquisition. Since theories of L2 phonological acquisition are informed both by speech perception data and by speech production data, further analyses of this corpus can contribute to the field in two significant ways.

First, the perception data from intelligibility studies, like the first study reported here, can be combined with the production analyses, as in the second study, to explore how perception and production vary. Second, both L1 and L2 productions by the same talkers with three different L1s allow a view into each talker's L1-L2 phonological space at a specific L2 proficiency level. These two types of phonological comparisons provide converging data on L2 phonological acquisition at an underinvestigated L2 developmental level.

We hope that the Buckeye GTA Corpus will be valuable for other scholars in linguistics, applied linguistics, and second language acquisition as we collaborate in further analyses of the corpus toward better understanding of L2 phonological acquisition.

7. References

- [1] B. B. Kachru, "Standards, codification and sociolinguistic realism: The English language in the Outer Circle," in *English in the World: Teaching and learning the language and literatures*, R. Quirk and H.G. Widdowson, Eds. Cambridge: Cambridge University Press, 1985, pp. 11-30.
- [2] B. B. Kachru and C. L. Nelson, "World Englishes," in *Sociolinguistics and Language Teaching*, S. L. McKay and N. H. Hornberger, Eds. Cambridge: Cambridge University Press, 1996, pp. 71-102.
- [3] J. Jenkins, *English as a Lingua Franca: Attitude and identity*. Oxford: Oxford University Press, 2007.
- [4] Institute of International Education, "Open Doors Report 2009," *Institute of International Education*. [Online]. Available: <http://opendoors.iienetwork.org>.
- [5] J. Bench and J. Bamford, *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children*. London: Academic Press, 1979.
- [6] J. B. Hardman, "The intelligibility of Chinese-accented English to international and American students at a US university," Ph.D. dissertation, The Ohio State University, Columbus, OH, USA, 2010.
- [7] E. McCullough, "Timing patterns of stops in non-native English," unpublished manuscript, The Ohio State University, Columbus, OH, USA, 2009.
- [8] J. E. Flege and R. Port, "Cross-language phonetic interference: Arabic to English," *Language and Speech*, vol. 24, no. 2, pp. 125-146, June 1981.
- [9] J. E. Flege, "The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification," *Journal of Phonetics*, vol. 15, pp. 47-65, Jan. 1987.
- [10] W. Marslen-Wilson and P. Zwitserlood, "Accessing spoken words: The importance of word onsets," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 15, no. 3, pp. 576-585, Aug. 1989.