Production and Perception of Temporal Contrasts in Foreign-Accented English

Bruce Smith+, Ann R. Bradlow*, and Tessa Bent*

+Department of Communication Sciences and Disorders
University of Utah
390 S. 1530 E., Rm.1201 BEH S
Salt Lake City, UT 84112
bruce.smith@hsc.utah.edu

*Department of Linguistics
Northwestern University
2016 Sheridan Road
Evanston, IL 60208
abradlow@northwestern.edu
t-bent@northwestern.edu

ABSTRACT

Several temporal features of English were examined to determine the extent of their occurrence in the speech of talkers of Chinese-accented English who had relatively limited experience with spoken English. Data from sentences produced by ten non-native and ten native talkers indicated that the two groups did not differ significantly in the extent to which they realized the inherent duration difference between tense and lax vowels or for syllable-final vowel lengthening. Although the group of native English speakers showed substantially greater vowel lengthening before voiced versus before voiceless consonants than the Chinese-accented group, certain individual, non-native speakers did show patterns reasonably similar to those of native speakers. In order to investigate the perceptual consequences of these observations, a follow-up study was designed with a focus on the vowel duration difference before voiced versus voiceless consonants. This investigation tested the abilities of native and non-native listeners to identify words produced by native English-speaking subjects and by Chinese-accented talkers who exhibited varying amounts of vowel lengthening before voiced versus voiceless consonants, in order to assess the impact of this temporal parameter on intelligibility.

1 INTRODUCTION

It is well known (e.g., Flege [2]; Major [5]) that individuals who learn a second language (L2) as adults typically have an accent, which is partly a result of various phonemic and/or allophonic differences that “transfer” from their native language (L1). In addition, L2 phoneme patterns that don’t exist in L1 can also contribute to a speaker’s L2 accent. There are also other possible factors that may factor into non-native accents that have received considerably less attention than phoneme substitution or deletion patterns and allophonic differences. For instance, temporal-acoustic patterns of speech that are considered to be salient aspects of English have been investigated to a lesser degree regarding their possible contributions to non-native speakers’ accents (however, see, for instance, Flege [3]; Flege, Munro, and Skelton [4]). Such temporal properties can be inherent within a given segment; they may be influenced by an adjacent segment; or they may be related to a larger, prosodic unit.

One general question addressed in the present study was, thus, what speech production differences exist between native and non-native speakers’ patterns in these various temporal domains. Given that there is considerable variability among individual, native speakers with regard to many temporal parameters (e.g., Smith [6], [7]), however, there is a need for caution in assuming what non-native speakers might need to learn. In addition, although relationships between non-native speakers’ production and perception abilities have been explored in various studies of segmental substitution, relatively less is known about such relationships with regard to temporal properties of speech. It is not known, for instance, just how salient such temporal parameters might be in terms of non-native speakers being perceived as having an accent or the extent to which they might affect intelligibility.

2 METHODS

To investigate these various issues and questions related to temporal properties associated with speech production of non-native speakers, the speech of ten native and ten
non-native speakers was examined. The native talkers of English (6 F, 4 M) averaged about 19 years of age. The L1 of the ten non-native speakers of English (3 F, 7 M) was Chinese; these subjects had an average age of about 23 years. Their mean performance on two standardized tests of English proficiency (Test of English as a Foreign Language = 649; Test of Spoken English = 45) indicated that their English language abilities were sufficient to be admitted into a graduate program at a university in the United States; however, they all spoke English with a noticeable foreign accent. They had been in the United States for an average of only about 1 month, but they had studied English formally, on average, for approximately 10 years.

Each subject randomly produced a minimum of five repetitions of the target word: “cab/cap, cup/cup, faze/face, peas/peace, peg/pek, pig/pick” and “tech” in both target positions of the carrier phrase, “I like to say _____ more than _____” (with the restriction that the two target words were never the same in a given sentence). The temporal features of English that were examined to determine the extent of their occurrence in the speech of talkers of Chinese-accented English compared to native English talkers were: (1) “lengthening” of tense relative to lax vowels (i.e., a secondary cue to phoneme identity), (2) vowel lengthening before voiced versus voiceless obstruents (i.e., a segmentally-influenced effect), and (3) vowel and consonant lengthening in sentence-final vs. non-final position (i.e., a sentence-level effect).

In addition, to examine various questions pertaining to intelligibility of speakers with a foreign accent, two groups of listeners participated in a two-alternative, forced choice word identification task (chance = 50%) involving voiced/voiceless contrasts in final consonants, e.g., hear "cap," and identify it as "cap" or "cab". One group of subjects consisted of 20 listeners who were native speakers of English; the other group consisted of 35 listeners who were non-native speakers of English (27 were speakers of Mandarin or another Chinese dialect, and 8 were speakers of various other Western or Non-western languages).

Each listener heard five randomized repetitions of the following word pairs: “cab/cap, cub/cup, faze/face, peas/peace, peg/peck, and pig/pick” produced by four different speakers, presented in a blocked design. Two of the speakers were native and two were non-native subjects selected from the 20 subjects who participated in the production task; they were chosen on the basis of characteristics they exhibited in the production of vowels before voiced versus voiceless stops. Specifically, one of the native speakers exhibited the greatest amount of relative vowel lengthening before voiced stops, whereas the other one showed an “average” amount of lengthening. One of the non-native speakers was chosen because he showed the greatest amount of vowel lengthening prior to voiced stops of any of the ten non-native speakers, which was approximately the same amount of lengthening as the “average” native speaker. The other non-native speaker was selected because he had the least amount of lengthening of any of the non-native (and native) subjects.

### 3 RESULTS AND DISCUSSION

#### 3.1 Production

The native English speakers and the Chinese-accented talkers differed in the extent to which they realized some, but not all duration contrasts. For example, Figure 1 shows the native English speakers, as a group, had a considerably greater contrast in vowel durations preceding voiced versus voiceless consonants (in both final and non-final position) than the non-native group. For final and non-final positions combined, both groups’ vowels preceding voiceless consonants averaged 140-145 msec in duration. Before voiced consonants, however, the native speakers’ vowels averaged 220 msec versus 167 for the non-native speakers, i.e., on average, a 55% versus 15% msec “lengthening effect,” respectively.

![Figure 1](image.png)

**Figure 1.** Average vowel durations preceding voiced and voiceless stops in sentence-final and non-final position for the native and non-native speaker groups.

Furthermore, as can be seen in Figure 2, there was quite limited overlap between the individual native and non-native speakers for this particular, segmentally-conditioned lengthening effect. The range of vowel lengthening was from about 25-80% for the native English speakers [mean = 1.58 lengthening (i.e., 58%) in non-final position, and 1.53 in final position, where 1.0 = no difference]. In contrast, the non-native speakers ranged from no lengthening of vowels before voiced obstruents to about 20% (mean = 1.12 in non-final and 1.17 in final). A Mann-Whitney U test indicated that the group differences were significant (p < .001) in both final and non-final position. It is quite possible that the reason for the difference between the two groups is that Mandarin has very few post-vocalic consonants (and no final stops);
therefore, this particular lengthening pattern of English is not something the Chinese-accented speakers had had any experience with in their native language.

In contrast to the substantial difference between the two groups for this particular temporal parameter, the group of native English speakers and the group of Chinese-accented talkers did not differ in the extent to which they realized the inherent duration difference between tense and lax vowels. In general, the amount of “lengthening” of tense vowels relative to lax vowels ranged from about 5-30% for both the native and non-native speakers. There were only two or three non-native speakers whose values fell outside that range, usually having greater tense/lax ratios than those observed for native speakers. The non-native speakers may have been able to produce this contrast adequately because it is often explicitly taught in English pronunciation courses and/or because it occurs allophonically in Chinese. The two groups of speakers also did not differ significantly in the extent to which they lengthened sentence-final vowels or consonants relative to non-final vowels and consonants, although there was somewhat of a tendency (p < .10) for the native English talkers to show somewhat greater sentence-final lengthening than the Chinese-accented talkers, especially for consonants.

One interesting finding in this experiment, which can be seen in Figure 3, was that the 20 native English-speaking listeners were significantly more accurate (ANOVA, p < .0001) in judging productions by the “good” non-native speaker (Non-nat., Max. V = 87%) than they were for the “average” native speaker (Native, Avg. V = 80%). This occurred despite the fact that both “Native, Avg. V” and “Non-native, Max. V” had very similar amounts of relative vowel lengthening before voiced consonants in their speech. This result may have been at least partially due to the non-native talkers often having had stronger final consonant release bursts than the native speakers (although this is an observation that has not yet been empirically verified). Another interesting observation in the present study was that the non-native listeners were more accurate (ANOVA, p < .0001) in judging the

---

**Figure 2.** Relative vowel lengthening before voiced versus voiceless consonants in sentence-final and non-final position for individual, native and non-native speakers. (The arrows indicate speakers selected for perception testing, who are described in more detail in the following section.)

**Figure 3.** Percent correct identification by native and non-native listeners on a one-interval, two-alternative forced choice word identification task involving final consonant voiced/voiceless production contrasts by native and non-native speakers.
Contrasts, suggest that non-native listeners don’t typically perform as having less of an accent than other non-native speakers. Native contribute various noted lengthening “transfer” deletion their well substantial to unaccented speakers. In their productions they reported the greater temporal-speech lengthening directly before voiced consonants (Native, Max V and Non-nat., Min. V). These observations seem comparable to findings reported by Tajima, Port and Dalby ([8]) that non-native speakers may become substantially more intelligible if they concentrate on improving temporal-acoustic patterns in their speech to be more comparable to those of native speakers.

4 CONCLUSIONS

In order to more fully understand the task that non-native speakers face in attempting to acquire relatively unaccented and intelligible English, it is important to investigate their performance with regard to various acoustic-phonetic parameters. Furthermore, it is essential to consider their performance within the context of the substantial variability that exists among native talkers, as well as with respect to the perceptual consequences of their productions. As with phoneme substitution and deletion patterns that are often observed when speakers are learning a second language, it was found in the present study that certain fine-grained temporal patterns may “transfer” from a non-native speaker’s L1 and/or may be quite easily taught to non-native speakers (e.g., tense vs. lax vowel contrast), whereas others may not (vowel lengthening preceding voiced obstruents). It was also noted that both native and non-native speakers tend to exhibit quite large ranges of performance in producing various temporal patterns. Given all the factors that can contribute to an accent, however, it is not known if non-native speakers whose productions fall within the range of native speakers’ productions would necessarily be judged as having less of an accent than other non-native speakers.

In terms of perception, results from the present study suggest that non-native listeners don’t typically perform as well as native listeners with regard to important temporal contrasts, but also that non-native listeners may be more accurate in recognizing temporal contrasts of non-native relative to native talkers. We are currently conducting a follow-up study investigating whether non-native listeners who exhibit a high degree of sensitivity to the vowel lengthening difference before voiced versus voiceless consonants in their perception of native-accented English also show native-like production for this temporal pattern of English.

5 REFERENCES