Original Paper

Phonetica 2008;65:131–147
DOI: 10.1159/000144077

Received: April 15, 2006
Accepted: May 2, 2008

Production and Perception of Temporal Patterns in Native and Non-Native Speech

Tessa Bent[a] Ann R. Bradlow[b] Bruce L. Smith[c]

[a]Department of Psychological and Brain Sciences, Indiana University, Bloomington, Ind., [b]Northwestern University, Evanston, Ill., and [c]University of Utah, Salt Lake City, Utah, USA

Abstract

Two experiments examined production and perception of English temporal patterns by native and non-native participants. Experiment 1 indicated that native and non-native (L1 = Chinese) talkers differed significantly in their production of one English duration pattern (i.e., vowel lengthening before voiced versus voiceless consonants) but not another (i.e., tense versus lax vowels). Experiment 2 tested native and non-native listener identification of words that differed in voicing of the final consonant by the native and non-native talkers whose productions were substantially different in experiment 1. Results indicated that differences in native and non-native intelligibility may be partially explained by temporal pattern differences in vowel duration although other cues such as presence of stop releases and burst duration may also contribute. Additionally, speech intelligibility depends on shared phonetic knowledge between talkers and listeners rather than only on accuracy relative to idealized production norms.

Copyright © 2008 S. Karger AG, Basel

Introduction

The present study investigated production and perception of segment-level temporal patterns by native and non-native speakers of English. The overarching goal of this research was to identify systematic acoustic-phonetic features of foreign-accented speech and to relate these features to foreign-accented speech intelligibility. Specifically, in order to identify production features that differed between native and non-native talkers, the productions of two duration contrasts were compared between the talker groups. Word intelligibility was then related to the production of between-category differences for one of these duration contrasts. We hypothesize that foreign-accented speech intelligibility is determined, at least in part, by the degree to which talkers' and listeners' phonetic systems are similar to each other, rather than merely the degree to

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2008 S. Karger AG, Basel
0031–8388/08/0653–0131
$24.50/0
Accessible online at:
www.karger.com/pho

Tessa Bent
Department of Speech and Hearing Sciences
Indiana University, 200 S. Jordan Ave.
Bloomington, IN 47405 (USA)
Tel. +1 812 855 4202, Fax +1 812 855 5531
E-Mail tbent@indiana.edu

which the foreign-accented speech approximates native talker norms [see also Bent and Bradlow, 2003; Imai et al., 2005]. Key findings from previous research in support of this hypothesis come from studies showing relatively high speech recognition accuracy for foreign-accented speech when presented to non-native listeners with both matching and mismatching L1s versus native listeners [Bent and Bradlow, 2003; Imai et al., 2005] and native listener adaptation to foreign-accented speech [Bradlow and Bent, 2008; Clark and Garrett, 2004]. Both of these general findings suggest that the intelligibility of foreign-accented speech samples depends on the talker-listener relationship, and is partially independent of the degree to which the talker's speech differs from abstractly defined native talker norms. While previous studies have identified this perceptual pattern, they have not investigated how native and non-native listeners' attention to different acoustic-phonetic cues may influence talker intelligibility and give rise to the somewhat surprisingly high intelligibility of foreign-accented speech for non-native listeners. Accordingly, the current study is an attempt to explore some of the specific acoustic-phonetic features that may characterize foreign-accented English and that may be the basis for its 'enhanced' intelligibility for non-native listeners relative to native listeners.

Furthermore, this research addresses the related issue of how non-native speech should be compared to native speech when attempting to identify ways in which non-native speech is different from native speech. Foreign accented speech is at least partially defined as a deviation from native speaker norms. However, the large variability among native speakers makes the issue of how to define 'norm' more complex. Rather than viewing foreign accented productions as deviant from native talker averages, here we suggest that the range of variation among native talkers should be considered. This consideration of native talker *ranges* then allows us to investigate the link between differences in production and differences in perception *within* groups of native and non-native talkers, as well as across native and non-native talkers.

The focus of this investigation was on segmental temporal patterns that can be influenced by two different types of 'conditioning' factors, including those that are: (a) inherent to specific segments or (b) influenced by an adjacent segment. American English (the target language) presents clear cases of these types of temporal conditioning factors. While there are many other temporal patterns that could be investigated, these two patterns were selected as a starting point in this line of research. As a case of an intrinsic duration contrast, we explored duration differences between tense versus lax vowels; overall, for pairs of vowels with similar height and backness specifications, tense vowels tend to be longer than lax vowels [e.g. Crystal and House, 1988; Peterson and Lehiste, 1960; Stevens, 1998]. As a case of a local, adjacent-segment conditioning factor, we focused on the tendency to lengthen vowels before word-final voiced versus voiceless obstruents [e.g. Chen, 1970; Crowther and Mann, 1992; Denes, 1955; Peterson and Lehiste, 1960].

There are substantial differences in the extent to which these temporal features are realized cross-linguistically, which presents an interesting opportunity for investigating foreign-accented speech. While tense vowels are typically longer than lax vowels in English and other languages that have analogous vowel contrasts [e.g. German as shown by Strange and Bohn, 1998], various languages and even different dialects within a language can exhibit this tense-lax vowel duration difference to varying degrees. For example, within the British Isles, speakers of Scottish English tend to produce less of a duration distinction and more of a spectral distinction between tense

and lax vowels, as compared to Southern English speakers [Escudero, 2001]. Stevens [1998] also notes that there are various possible acoustic correlates of the tense-lax contrast that languages use to different degrees. Although increased vowel duration before voiced versus voiceless word-final consonants is also seen in a number of languages, the magnitude of this contrast differs across languages, and some languages do not appear to exhibit it at all [e.g. Chen, 1970; Flege and Port, 1981; Laeufer, 1992; Mack, 1982].

Temporal patterns that are not present or are not extensive in a talker's native language could influence his/her ability to produce those features in a second language in a native-like fashion. Although there have been many studies concerning how learning a second language may be affected in terms of voice onset time differences across languages [e.g., Flege, 1991; Flege and Eefting, 1986; Schmidt and Flege, 1995], other temporal parameters have been less extensively studied with respect to second language learning. Nevertheless, previous research has found that non-native talkers of English, particularly inexperienced ones, tend to produce less extensive vowel duration contrasts before voiced versus voiceless consonants [Flege, 1993; Flege and Hillenbrand, 1986; Flege, Munro and Skelton, 1992; Mack, 1982].

Although such studies suggest that non-native talkers' temporal patterns often differ from those of native talkers, considerable inter-talker differences in temporal features produced by native talkers make assessing the potential impact of deviation from native talker averages questionable. That is, although most descriptions of temporal patterning in English report group averages, as early as 1976 Klatt noted, 'There is considerable inter-speaker and intra-speaker variability in durational studies' [Klatt, 1976, p. 1208]. More recently, Smith [2000, 2002] has demonstrated that although a group of native speakers, on average, will tend to demonstrate various temporal patterns that are viewed as 'characteristic' of English, some individual subjects do not manifest certain temporal parameters at all, or manifest them to only a very limited extent. For instance, the lengthening of vowel durations in words in phrase-final position compared to non-final position across 10 subjects ranged from no evidence of this particular pattern to phrase-final vowels that were over 1.5 times as long as non-phrase-final vowels [Smith, 2000]. Thus, given that individual native speakers vary in the extent to which they realize a number of temporal parameters, a certain amount of caution is needed in assuming what non-native speakers may need to learn in order to approach native proficiency in terms of such patterns.

The central goal of the present study was to examine between-category differences in the production of two temporal-based contrasts across and within both native and non-native talkers of English and then to relate these individual- and group-level differences in production to variability in intelligibility for both native and non-native listeners. Previous work has established a relationship between non-native talker segmental production and intelligibility [e.g. Bent et al., 2007; Derwing and Munro, 1997; Munro and Derwing, 1995; Rogers, 1997]. Moreover, in an important attempt to relate changes in temporal characteristics of speech production to variation in overall intelligibility, Tajima et al. [1997] digitally modified the duration of acoustic segments of non-native speech to match corresponding samples by native talkers. They modified the speech by first segmenting the utterances into vowels, liquids, nasals, fricatives, and stops. After the utterances were divided into these categories, the durations of the non-native utterances were lengthened or shortened to match the segment durations for native speech using a dynamic time-warping algorithm, which temporally altered the

speech while maintaining its spectral characteristics. These manipulations significantly increased non-native speech intelligibility for native English listeners, suggesting that segmental temporal patterns over the course of a phrase-length utterance in non-native speech are important cues for effective communication. The present study further pursues the connection between specifics of non-native production and intelligibility by focusing on between-category differences in production of various duration contrasts in relation to the consequences of these duration differences for word intelligibility. The primary research questions were:

(1) What is the extent of inter-talker between-category differences for native and non-native talkers in their production of segmental temporal patterns related to segment-inherent duration differences and local phonetic contextual differences (experiment 1)?

(2) How do the between-category differences in the production of a temporal pattern within and across native and non-native talkers influence native and non-native listeners' abilities to accurately identify words in minimal pairs (experiment 2)?

Experiment 1: Production of English Temporal Contrasts by Native and Non-Native Talkers

Experiment 1 examined native and non-native talkers' productions of two duration contrasts in English including: (a) the duration of tense relative to lax vowels (a segment-intrinsic, secondary cue to phoneme identity) and (b) vowel duration before voiced versus voiceless obstruents (a context-conditioned effect).

Both of these contrasts do not exist in Mandarin. First, Mandarin has a smaller vowel inventory compared to English with only six vowels /i, e, u, o, ɑ, y/ and does not have a tense-lax vowel contrast as all the vowels in Mandarin are considered tense [Chen, 2006]. Second, because the coda position in Mandarin syllables is limited to either open syllables or nasal codas, there is not a voiced/voiceless distinction in final position.

Methods

Talkers

Ten native speakers of English (6 female and 4 male) with a mean age of 19 years from various geographic locations around the United States participated. These native speakers were all undergraduate students at Northwestern University. Ten non-native speakers of English whose first language was Chinese (3 females and 7 males) with a mean age of 23 years also participated. All Chinese talkers were native speakers of Mandarin in that all schooling including university had been conducted in Mandarin. Additionally, many of the participants spoke another dialect of Chinese at home. 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 Northwestern University; however, they all spoke English with a noticeable foreign accent. They had been in the United States for an average of only 1 month, but they had studied English formally in China for 10 years, on average.

Stimuli and Task

The 20 talkers were recorded individually in a sound-attenuated booth. The recordings were made on an Ariel Proport A/D soundboard with a Shure SM81 microphone. Each subject was recorded

Table 1. Word pairs for experiment 1

Word pairs targeting the tense versus lax vowel contrast			
(1)	deep	dip	/i/-/I/
(2)	ease	is	/i/-/I/
(3)	take	tech	/e/-/ɛ/
(4)	phase	fez	/e/-/ɛ/
Word pairs targeting vowel duration contrast before voiced versus voiceless obstruents			
(5)	peace	peas	/i/
(6)	pick	pig	/I/
(7)	peck	peg	/ɛ/
(8)	face	phase	/e/
(9)	cap	cab	/æ/
(10)	cup	cub	/ʌ/

as s/he randomly produced a minimum of five repetitions of the 20 target words: 'cab, cap, cop, cub, cup, deep, dip, ease, face, fez, is, peace, peas, peck, peg, phase, pick, pig, take, tech' in both 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).[1] Word pairs are listed according to the contrast they target in table 1.

Note that the word 'phase' was paired with 'fez' for the tense versus lax comparison and with 'face' for the voiced versus voiceless following consonant comparison. After each subject's session, recordings were converted to the WAV format with a 16-kHz sampling rate and transferred to a PC-based computer.

Measurements

The various segments in table 1 were analyzed to examine the duration patterns for tense versus lax vowels and vowels before voiced versus voiceless obstruents. All measurements were made by the third author. Each talker's productions were analyzed by examining acoustic waveforms (SoundEdit 16, v. 2.0.7) displayed on a Macintosh G4 computer. Segmentation of the consonants and vowels was based on commonly utilized acoustic characteristics associated with substantial changes in waveform shape and/or amplitude, consonant release bursts, and other relevant acoustic events [Smith et al., 1986]. Sentence-final stops were measured if they were released. Voice onset time was measured separately from vowel duration. Intra-judge reliability was evaluated by having the investigator who performed all the original acoustic analyses remeasure the data for one randomly selected, non-native subject after a period of approximately 3 months. Inter-judge reliability was also assessed by having a different investigator (S.N.) remeasure all the data for 2 different randomly selected, non-native subjects. On average, both intra- and inter-judge vowel duration measurements differed by 4 ms (approximately 2%) or less and consonant durations differed by 2 ms (approximately 1%) or less. All of these reliability measurements demonstrated good agreement between the 2 investigators for the measures of interest in the study.

[1] The target words were produced in two positions within the sentence with the intent of studying the phenomenon of phrase-final lengthening. However, various native and non-native participants were not as consistent in their production of the phrasing related to nonfinal position as for final position. Therefore, to minimize variability in the results due to sentence-internal pausing, for example, only temporal data from sentence-final position are reported in the 'Results' section. Despite possible limitations due to inappropriate pausing by some participants, however, it is worth noting that the native English-speaking subjects, on average, lengthened final-syllable vowels by 11 versus 3% 'shortening' by the non-native speakers. In addition, native English speakers lengthened final consonants by 29% compared with 17% lengthening by the non-native speakers. Also, note that 'cop' was only included for analysis in the final versus nonfinal comparison condition and therefore, no data is reported on this word in the current paper.

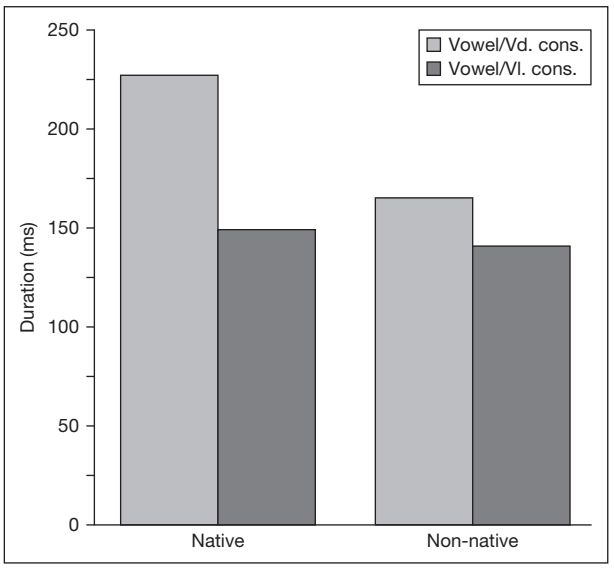

Fig. 1. Average vowel durations preceding voiced and voiceless consonants for the native and non-native speaker groups.

Results

The native English talkers and the Chinese-accented talkers differed in the extent to which they realized one, but not the other, of the two duration contrasts that were examined.

(a) *Duration of Tense relative to Lax Vowels*. The native talkers and the Chinese-accented talkers did not differ appreciably in the extent to which they realized this contrast. Tense vowels tended to be longer than lax vowels for both talker groups (Mann-Whitney $U = 36.5$; $p > 0.10$). The individual native talkers' tense/lax relative lengthening values ranged from 1.01 to 1.27, whereas the non-native talkers showed a somewhat greater range in this contrast with values from 1.00 to 1.60 lengthening (where 1.0 = no lengthening, values above 1.0 indicate longer tense than lax vowels and values below 1.0 indicate shortening where lengthening is expected).

(b) *Vowel Length before Voiced versus Voiceless Obstruents*. The native and non-native talkers differed for this contextually conditioned lengthening of vowel durations preceding voiced versus voiceless consonants. The native English talkers, as a group, had a considerably greater contrast for vowel durations preceding voiced versus voiceless consonants than the non-native group (fig. 1). On average, the native talkers' vowel durations before voiced consonants versus before voiceless consonants had a relative lengthening value of 1.54, whereas the non-native talkers showed only a 1.17 'lengthening effect' in this context. However, the duration differences for vowels preceding voiced versus voiceless consonants were significant within both the native and the non-native talker groups (Wilcoxon matched-pairs signed ranks = 55.0; $p < 0.005$ in both cases). Therefore, both groups of talkers produced significantly longer vowels before voiced versus voiceless obstruents, but the native talkers produced a much greater contrast than the non-native talkers did. Vowels before voiceless consonants were similar in duration across native and non-native talkers; however, the two groups

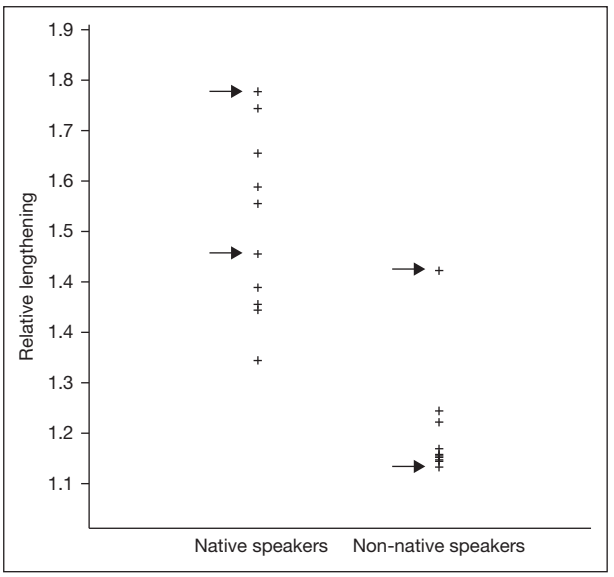

Fig. 2. Relative vowel lengthening averaged across lexical items before voiced versus voiceless consonants for individual, native and non-native speakers. The arrows indicate speakers selected for perception testing in experiment 2, who are described in more detail in the text.

differed for vowels preceding voiced consonants, where vowel durations were substantially longer for native than non-native talkers. That is, the native and non-native talkers' vowels preceding voiceless consonants were not significantly different (Mann-Whitney $U = 46.0$; $p > 0.10$). However, before voiced consonants, the native talkers' vowels were significantly longer than those of the non-native talkers (Mann-Whitney $U = 9.0$; $p < 0.005$).

In addition to considering group findings, it is also of interest to examine the performance of the individual talkers. As can be seen in figure 2, there was some, but quite limited, overlap among the individual native and non-native talkers for the relative lengthening effect of voiced consonants on preceding vowels. The range of relative vowel lengthening was from 1.30 to 1.78 for the native English speakers (mean = 1.54). In contrast, other than one 'outlier' (with a 1.46 relative lengthening value), the individual non-native talkers ranged from 1.11 to 1.21 relative lengthening (mean = 1.17). The differences in relative lengthening values between the two groups of subjects were significant (Mann-Whitney $U = 4.0$; $p < 0.001$).

Discussion

Two temporal patterns were examined in the speech of 10 native talkers of English and 10 non-native talkers of English. The native and non-native talkers differed in their productions of the contextually conditioned pattern of vowel lengthening preceding voiced consonants, which averaged 1.54 for the native English talkers compared to an average of 1.17 for the non-native talkers. One possible reason that Mandarin speakers did not produce as large a difference in this contrast is the lack of a final voiced-voiceless distinction in Mandarin; therefore, this particular vowel lengthening pattern of English must be learned by the Chinese talkers and is not subject to positive transfer

from their native language. Given that most of the non-native subjects were adults who had been in the United States only a short time and were, therefore, reasonably inexperienced talkers of English, they may have focused more on producing the word-final consonants themselves and may not have focused as much on the contextually conditioned vowel lengthening that tends to occur prior to voiced obstruents among native talkers of English (see evidence for this proposal in the 'Discussion' of experiment 2). These results are in accord with previous studies in which non-native talkers have been found to produce this temporal contrast to a lesser degree than native English talkers [Flege, 1993; Flege et al., 1992; Mack, 1982]. However, even though both groups exhibited considerable inter-talker differences, there was very little overlap between the two groups with the native talkers showing greater between-category duration differences than the non-native talkers.

In contrast to the findings for the vowel length before voiced versus voiceless consonants, there were no differences between the two talker groups in terms of the durations for tense versus lax vowels. This result is in accord with the recent findings of Chen [2006], who also found that Mandarin talkers produced the duration difference between tense-lax vowel pairs to the same or a greater extent than native English talkers. Chen [2006] also found that the non-native (Mandarin) talkers failed to accurately produce the spectral differences between the tested English tense-lax vowel pairs. In the present study, our focus is on duration differences only, and thus no vowel spectral features are included.

Experiment 2: Perception of Minimal Pairs Differing in the Voicing of Coda Consonants

An important question in speech production is whether between-category differences have any appreciable effect on speech intelligibility. Therefore, the aim of this experiment was to investigate perceptual consequences of between-category differences across speakers for the temporal feature that showed a difference between the native and non-native speakers in experiment 1 (namely, vowel length before voiced versus voiceless consonants). Specifically, we sought to determine whether words produced by talkers who manifest a relatively large contrast in vowel duration before final voiced versus voiceless consonants are more accurately recognized than words with a smaller vowel duration contrast in this phonetic context. That is, do phonetic details that differ across talkers influence word recognition, or are these details essentially inconsequential for phonemic categorization and lexical access?

Previous work has shown that variability in overall speech intelligibility between native talkers and listeners can be accounted for to some extent by differences in articulatory precision. For example, studies have shown that variability in intelligibility can be related to differences in vowel space area (greater area for talkers with higher overall intelligibility) and in precision of intersegmental timing such as closure duration, voicing during closure and relative segment duration [Bradlow et al., 1996; Hazan and Markham, 2004]. Moreover, other work has shown that listeners are sensitive to talker-specific pronunciation patterns in that experience with a talker's voice and articulation patterns results in improved recognition of that talker's speech [Allen and Miller, 2004; Nygaard et al., 1994; Nygaard and Pisoni, 1998]. The present experiment extends these previous findings by assessing the extent to which the realization of the vowel duration

contrast before voiced versus voiceless obstruents affected word recognition when the talkers and listeners either shared or did not share the same native language. The current experiment extends previous work by testing both non-native talkers and listeners in addition to native talkers and listeners and assesses a phoneme contrast rather than broadly assessing intelligibility and acoustic-phonetic parameters. Based on previous results, it is expected that talkers who produce a larger difference in vowel duration before voiced versus voiceless consonants will be more intelligible. However, the effect of contrast enhancement may be mediated by a native language background match between the talker and listener [e.g. Bent and Bradlow, 2003]. Practically, this issue is important since details of pronunciation that influence intelligibility should potentially be targeted in second-language learning environments, whereas production patterns that do not influence perception are not as important to include in pronunciation curricula. Accordingly, this experiment tested the abilities of native and non-native listeners to identify words produced by native English talkers and by Chinese-accented talkers with different amounts of vowel lengthening before voiced versus voiceless consonants.

While the main purpose of the experiment was to compare performance across native English and native Chinese listeners on productions by native English and native Chinese talkers, as a secondary goal, we wanted to include a group of non-Chinese, non-native listeners as a means of de-confounding the native language match and the status of the listener as non-native speakers of English for the conditions that combined non-native talkers and non-native listeners. In this case, we included a highly heterogeneous group of non-Chinese non-natives primarily for a practical reason (i.e. this was the most readily available group of subjects at the time due to the ESL program from which we recruited our subjects). Including this non-native listener group allowed us to assess performance across a group of listeners where any particular interaction between the listeners' native language sound structure and the sound structure of Chinese would, on average, be 'washed out'. In future research, specific native languages comparisons should be investigated.

Methods

Listeners

The listeners in this study included both native and non-native speakers of English. None of these listeners had participated in experiment 1. Twenty native English listeners with a mean age of 19 years (16 females and 4 males) participated. All subjects were undergraduates at Northwestern University. Thirty-five non-native English listeners with a mean age of 24 years (14 females and 21 males) also participated. Of the non-native listeners, 27 were speakers of Mandarin or another Chinese dialect[2], and 8 were speakers of various other Western or Non-Western languages including 1 speaker each of German, Hebrew, Hindi, Italian, Korean, Tamil, Telugu, and Turkish. The participants in this study were from the same general population as the participants in experiment 1 (graduate students at Northwestern University).

Stimuli

The stimuli were six voiced/voiceless pairs (cab/cap, cub/cup, phase/face, peas/peace, peg/peck, and pig/pick) extracted from sentence-final position of the production study recordings of experiment 1. Four talkers' recordings were included in this perception study, 2 native and 2 non-native talkers selected from the 20 participants in experiment 1. These particular talkers were chosen on

[2] All of the Chinese listeners were fluent in Mandarin as all their schooling including University had been conducted in Mandarin. Of the 27 Chinese listeners, 19 reported Mandarin as their home dialect, 4 reported a Chinese dialect other than Mandarin, and 4 did not report a specific dialect.

the basis of characteristics they exhibited in the production of vowels before voiced versus voiceless consonants (fig. 2). The talkers were chosen to represent the entire range of values observed among the 20 subjects in experiment 1 for relative vowel lengthening before voiced versus voiceless consonants.[3] Specifically, one of the native talkers exhibited the greatest amount of relative vowel lengthening before voiced versus voiceless consonants ('Nat. Max.' = 1.78), whereas the other native talker showed an amount of lengthening that was closest to the average value for the group of native talkers ('Nat. Avg.' = 1.43). One of the non-native talkers was chosen because he showed the greatest amount of vowel lengthening before voiced consonants of any of the 10 non-native talkers ('Non-Nat. Max.' = 1.46), which was approximately the same amount of lengthening as the 'average' native speaker (Nat. Avg.). Although this talker was an outlier in the non-native group, as he showed much more lengthening than the other non-native talkers, he was selected to compare intelligibility for a native and a non-native talker who produced this cue to approximately the same extent. While the Non-Nat. Max. talker was an outlier in the non-native group, it seems likely that with a larger group of non-native participants, particularly ones who have spent longer amounts of time interacting with native speakers, more non-native talkers would exhibit values close to the native talker average. The other non-native talker was selected because he had the least amount of relative vowel lengthening of any of the non-native (or native) talkers ('Non-Nat. Min.' = 1.11). Three of these talkers were male (Nat. Avg., Non-Nat. Max., Non-Nat. Min.) and 1 was female (Nat. Max.). It should be noted that listeners in the current experiment never heard the vowel duration difference as a direct contrast as they heard the words one at a time rather than in pairs (see task description below).

Task

The task was a two-alternative, forced-choice word identification task (chance = 50%) involving the voiced/voiceless contrast in word-final position. Each listener heard five randomized repetitions of each of the 12 stimulus words produced by the 4 different talkers, presented in a blocked design (total = 240 trials). Trials were blocked by talker with the 2 native English talkers' productions presented first followed by the 2 non-native talkers' productions. The talker with the greater vowel duration difference for each language group was presented first (i.e. Nat. Max., Nat. Avg., Non-Nat. Max., Non-Nat. Min.). This order gave the listeners the opportunity to adapt to the task before being presented with stimuli from the talkers that were expected to be less intelligible (i.e. the non-native talkers with less extensive vowel duration contrasts), thereby ensuring that the possibility of lower word identification scores for the non-native talkers could not be due to a lack of familiarity with the task. On each trial the listeners heard one word and were presented with the two possibilities of a minimal pair on a computer screen; they had to identify which word of the minimal pair they heard, e.g., hear 'cap', and identify it as 'cap' or 'cab'. Stimuli were played through headphones at a comfortable listening level. The experiment was controlled by experiment-running software (SuperLab Pro 2.01), and listeners entered their responses on a specially designed response box.

Results

Results of the word identification task are shown in figure 3 for the 4 different talkers and for the native and two different non-native listener groups. That is, some non-native listeners shared the native language background of the non-native talkers and some did not. We, therefore, compared performance on this word recognition test between the Chinese and the non-Chinese non-native listeners. While this comparison can help identify possible effects of native language match versus mismatch between

[3] We acknowledge that the values for vowel lengthening before voiced versus voiceless consonants are confounded with the language backgrounds of the speakers. That is, because the two participant groups showed very little overlap, it was impossible to select a non-native talker with a value comparable to the Nat. Max. talker's values and likewise it was impossible to select a native talker with a value comparable to the Non-Nat. Min. talker's values. However, due to the values for the speakers on which we collected data, the decision to choose speakers with extreme values allowed us to test perception of speakers across the entire range of the scale.

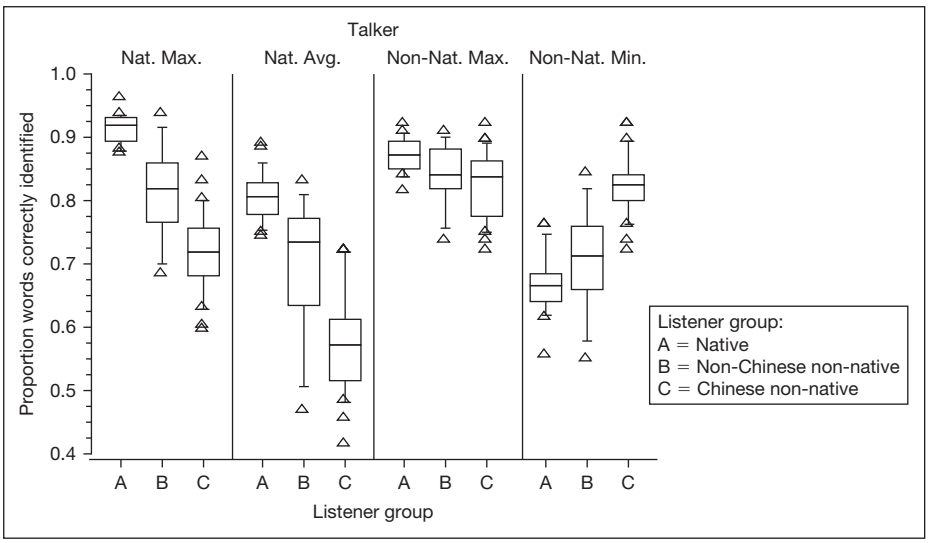

Fig. 3. Proportion correct word identification by (from left to right) native, non-Chinese non-native, and Chinese non-native listeners on a one-interval, two-alternative forced-choice task for final consonant voiced/voiceless contrasts produced by native and non-native (Chinese) talkers. In these box plots, the box shows the 25–75th percentiles with the median indicated by the line inside the box, the error bars show the 10th and 90th percentiles, and the triangles below and above the error bars indicate data points that fell outside of the 10th and 90th percentiles, respectively.

non-native listeners and talkers, recall that the number of non-Chinese participants available at the time of testing was substantially smaller than the number of Chinese participants (Chinese n = 27, non-Chinese n = 8). Also, while the Chinese group was quite homogeneous with respect to language background, the non-Chinese group included 1 talker from a variety of different language backgrounds.

Overall, the patterns observed in this experiment indicate that non-native listeners performed equivalently to or outperformed the native listeners for the non-native talkers. This is evidenced by (1) the increase in performance for both groups of non-native listeners when presented with words produced by the more native-like non-native talker (Non-Nat. Max.) relative to their level of performance with the native talkers and, (2) by the superior performance of the Chinese non-native listeners over the native and other non-native listeners for the less native-like non-native talker (Non-Nat. Min.). The native listeners were also sensitive to inter-talker differences in the production of the contrast in both native and non-native speech: for the native listeners, the talker with greater lengthening from each language background was more intelligible than the one with less lengthening (i.e. Nat. Max. > Nat. Avg.; Non-Nat. Max. > Non-Nat. Min.).

An ANOVA showed a significant main effect of Talker [$F(3, 52) = 65.09$, $p < 0.0001$] and Listener [$F(2, 52) = 31.12$, $p < 0.0001$]. The two-way Talker-by-Listener interaction was also significant [$F(3, 156) = 56.54$, $p < 0.0001$] due to different patterns for the three listener groups across the 4 talkers. Pairwise comparisons (paired t tests) showed that for the native listeners, all differences between talkers were significant ($p < 0.0001$). For the Chinese listeners, all differences between talkers were significant

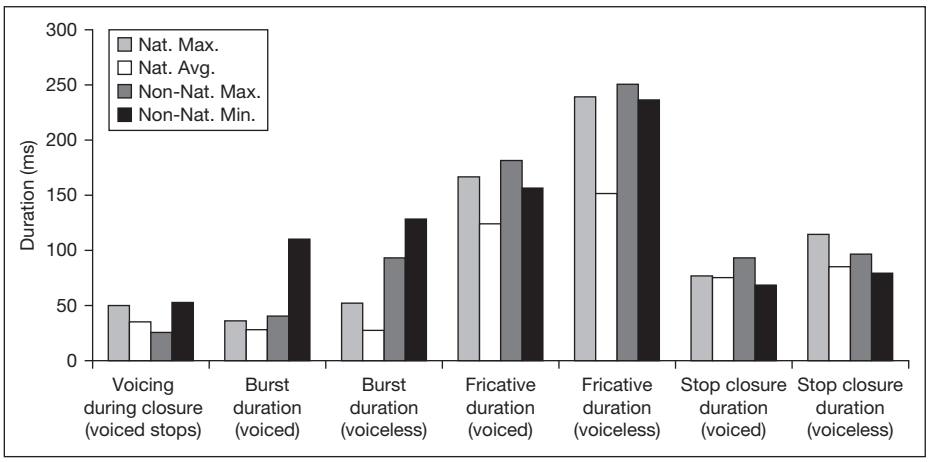

Fig. 4. Coda consonant measurements for the 2 native and 2 non-native talkers' productions presented to listeners in the perception experiment (experiment 2).

($p < 0.0001$) except between the 2 non-native talkers. For the non-Chinese non-native listeners, the only significant differences were between the 2 native talkers and between Nat. Avg. and Non-Nat. Max.

Comparisons across listener groups within each talker showed that for the native talkers, all listener group differences were significant ($p < 0.0025$) with the native listeners performing best (Nat. Max. = 91% and Nat. Min. = 81%), followed by the non-Chinese non-native listeners (81 and 70%) and finally the Chinese non-native listeners (72 and 58%). This pattern suggests that, on average, the other non-native listeners were probably more proficient in English word recognition than the Chinese listeners or that they had more experience with vowel duration differences before voiced versus voiceless consonants either in their native language or through study of other foreign languages. The results are quite different for the Chinese talkers. For the more native-like non-native talker (Non-Nat. Max.), all listeners recognized approximately the same proportion of words (87, 84 and 82% for the native, non-Chinese non-native, and Chinese non-native groups, respectively; no group differences were statistically significant). However, for the less native-like non-native talker (Non-Nat. Min.), the Chinese listeners recognized significantly more words correctly (83%) than either the native (67%) or the other non-native listeners (71%; $p < 0.0001$ for both comparisons). For the Chinese listeners, word identification accuracy did not differ across the 2 Chinese talkers (82 and 83%), whereas differences between these talkers were significant for the native (87 and 67%) and other non-native listeners (84 and 71%).

Discussion

Experiment 2 demonstrated that there were perceptual consequences for the range of between-category differences in production of the vowel length contrast before voiced versus voiceless consonants for both native and non-native talkers and

listeners. That is, both native and non-native listeners, as groups, more accurately identified words produced by a native talker with a relatively large amount of vowel lengthening before voiced versus voiceless consonants (Nat. Max.) than by a native talker with an average amount of vowel lengthening in this phonetic context (Nat. Avg.). Furthermore, native listeners were also sensitive to the vowel length difference between the 2 non-native talkers and were more accurate at identifying the productions of the non-native talker with a relatively large amount of vowel length difference (Non-Nat. Max.) than the non-native talker with a relatively small amount of vowel length difference (Non-Nat. Min.). The finding that between-category differences among talkers influence word intelligibility may partially account for previous findings that some talkers are more intelligible than others. Furthermore, these findings may partially account for native listeners' perception of foreign accents and decrements in intelligibility for non-native talkers. Since native listeners were least accurate in identifying words produced by Non-Nat. Min., it is possible that the perception of the voiced/voiceless contrast would also have been difficult for native listeners hearing the speech of some of the other non-native talkers in experiment 1, given that they had quite similar patterns of vowel lengthening before voiced versus voiceless obstruents. Thus, it appears reasonable to conclude that part of the reduced intelligibility associated with at least some of these non-native talkers was related to their relatively minimal distinction between vowel durations preceding voiced versus voiceless obstruents. It is also important to realize, of course, that in normal speaking circumstances versus 'isolated word' situations (as was the case in this study), context often helps with the recognition of words in which phonemic contrasts may be partially or even completely neutralized acoustically.

The significant interactions between talker and listener groups suggest that both production and perception strategies may have differed between the native and non-native participants. That is, while the vowel length difference between voiced versus voiceless consonants clearly provides an important cue to word identity, the production of other cues (e.g. final consonant duration or burst release duration) or the perceptual weighting of other cues can influence listeners' perceptual accuracy. In terms of production strategies, the finding that the native and non-native listeners were 6–24% more accurate in identifying words with final voiced versus voiceless consonants produced by Non-Nat. Max. than those produced by Nat. Avg. (both of whom had similar amounts of vowel lengthening) suggests that the non-native talker may have provided additional acoustic cues in his productions that the native talker did not. To investigate this possibility, additional acoustic analyses of the stimuli used in this experiment were conducted. These post-hoc analyses are certainly not conclusive but support the idea that the non-native speakers (particularly Non-Nat. Min.) may have produced some cues to a greater extent than the native speakers. These cues may have been used by the non-native listeners during the word identification task, while the native listeners may have been attending more to the vowel durations which Non-Nat. Min did not produce extensively enough to be used for effective word identification. For these analyses, frequency of occurrence of final stop release, voicing during closure for voiced final stops, burst duration, fricative duration, and stop closure duration were measured. Measures other than frequency of final stop release are shown in figure 4 and are reported as durations in milliseconds.

For these follow-up measures, the native and non-native talkers showed differences in the production of some but not all of the consonantal features. Final stops

were released 100% of the time for both non-native talkers, whereas the native talkers had somewhat lower rates of final stop release with 88% for Nat. Max. and 96% for Nat. Avg. In addition to the non-native talkers more frequently releasing final stops compared to Nat. Max., the non-native talkers also produced longer final consonants than Nat. Avg. Specifically, Non-Nat. Min. (for both voiced and voiceless final stops) and Non-Nat. Max. (for voiceless stops) produced longer burst durations (53–89 ms) than the native talkers. Furthermore, the durations of final fricatives for both non-native talkers were longer than Nat. Avg. while they were about the same as Nat. Max. This analysis suggests that the non-native listeners, especially the Chinese non-native listeners, may have been attending more to the information present in the final consonants to make their word identifications for the productions by Non-Nat. Min. while the native listeners were attending more to vowel duration differences. However, one consonant cue that the non-native speakers did not produce more extensively than the native speakers was the closure duration difference between voiced versus voiceless stops. Talker Nat. Max. produced the largest relative difference (her voiceless stop closure durations were 1.49 ms longer than her voiced stop closure durations) while the other 3 talkers produced much smaller differences (Nat. Avg. = 1.13, Non-Nat. Max. = 1.04; Non-Nat. Min. = 1.16). While this analysis is suggestive, there are other cues that the Chinese non-native listeners may have been attending to including vowel quality, formant transition duration, rate of energy decay, and/or F1 offset frequency. Since this experiment was not designed to test all possible cues, this issue will need to be examined in future research. In particular, it may be useful to synthetically manipulate the different possible cues to final consonant voicing and determine their effects on perception by native and non-native listeners.

On average, the Chinese non-native listeners were more accurate in identifying the voiced/voiceless contrast in the non-native talkers' productions than they were in identifying them in the native talkers' productions. The non-Chinese non-native listeners showed slightly higher accuracy for the Non-Nat. Max. compared to the native talkers but performed more accurately on the native talkers than Non-Nat. Min. These findings are consistent with results obtained by Bent and Bradlow [2003] and by Imai et al. [2005]. In the study by Bent and Bradlow [2003], non-native listeners found sentences produced by high-proficiency non-native talkers equally intelligible to sentences produced by native talkers (even when the talker and listener had different native languages), which was referred to as the matched (shared native language) or mismatched (different native languages) 'interlanguage intelligibility benefit'. Imai et al. [2005] found that Spanish-accented English words from dense lexical neighborhoods (i.e. words having many similar-sounding lexical neighbors with which they could easily be confused as defined by words typically in the lexicons of native speakers) were more accurately recognized by native Spanish than native English listeners. These items require fine-grained speech sound discrimination at the segmental level since lexical neighbors are defined in terms of single phoneme differences from the target word. Therefore, this finding is consistent with the phonological mismatch hypothesis of Imai et al. [2005], which states that differences at the segmental level between English words and listeners' lexical representations will lead to decreased word identification accuracy. The present findings are consistent with these previous studies in showing that word recognition accuracy depends (at least in part) on a match between the talkers' and listeners' phonological systems.

Summary and General Discussion

This study included a detailed look at inter-talker differences in the production of two duration features in native and non-native speech (experiment 1) and provided basic data about the consequences of inter-talker differences for native and non-native speech perception for one temporal pattern (experiment 2). The results of these two experiments have both theoretical and pedagogical ramifications. From a pedagogical point of view, the present study has implications for second-language learners and teachers. A somewhat surprising finding of experiment 2 was that native listeners actually performed better, on average, on the word identification task for Non-Nat. Max. than for Nat. Avg. While these talkers' productions were similar with respect to the measured acoustic feature of vowel lengthening before voiced versus voiceless consonants, there were other acoustic parameters that differed across these 2 talkers, and there was a strong auditory impression of a foreign-accent for the non-native talker (but, of course, not for the native talker). In other words, the non-native talker was quite effective in conveying a certain lexical contrast despite the fact he was doing so on the basis of a different combination of acoustic-phonetic features than the native talker including longer voiceless burst durations and longer coda fricatives. This comparison, therefore, provides an example of a non-native talker bringing a kind of 'richness' to the task of English speech production that actually worked quite well for native and non-native listeners. Thus, second-language learners and teachers should probably provide exposure to a wide range of target language model talkers, perhaps even including some very proficient (but clearly foreign-accented) non-native talkers.

As with other aspects of non-native speech production, the present data indicated that certain fine-grained temporal patterns may be quite easily acquired by non-native speakers (e.g., tense versus lax vowel duration contrast). In contrast, other temporal patterns are apparently less easily acquired in a second language (e.g. relative vowel lengthening preceding voiced versus voiceless obstruents). The acoustic analyses showed that native and non-native speakers tended to exhibit quite large ranges of performance in producing various temporal patterns [see Smith, 2000, 2002 for similar data with native talkers]. Furthermore, these inter-talker between-category differences in native talker productions had perceptual consequences for both native and non-native listeners; that is, the native talker who produced a greater vowel duration difference between final voiced versus voiceless consonants was generally more intelligible to native and non-native listeners than the talker with a smaller duration difference. In the cases of Chinese non-native talkers and listeners, the production of and attention to other cues, in addition to the vowel duration difference between final voiced versus voiceless consonants, appears to have also played an important role. These perceptual results therefore provide an additional demonstration of the previously observed finding that non-native listeners may be more accurate at identifying words produced by non-native talkers than words produced by native talkers.

Models of speech perception and production often consider 'typical' patterns and may not take individual, contextual, and other sources of variation into account. 'Statistical' learning models [e.g. Maye et al., 2002; Pierrehumbert, 2003] have provided new insights into how non-native learners (and children) may accumulate knowledge about production differences between linguistically contrastive categories in any given language and how they can use this knowledge to build the complex cognitive structures that support speech communication. In line with statistical learning models,

experimental work on training the perception of novel non-native segmental contrasts has found that exposure to within-category variation and between-category differences are beneficial for the acquisition of more accurate phonemic category representations [for a general review see Bradlow, in press]. These current views of learning in perception emphasize a direct connection between variety in the input and speech sound learning; and, as such they are completely consistent with the present findings of a connection between differences in production and variability in intelligibility for both native and non-native talkers and listeners. Taken together, these findings support a general understanding of speech communication as a process that involves talker-listener alignment.

Acknowledgments

Earlier versions of this work were presented at the 144th (Fall 2002 in Cancun, Mexico) and 147th (Spring 2004 in New York, N.Y.) meetings of the Acoustical Society of America, and at the 15th International Congress of Phonetic Sciences (2003 in Barcelona, Spain). We are grateful to Mengting Shieh for assistance with recording and testing, and to the Northwestern University International Summer Institute and ESL Program for facilitating access to the non-native participants in this study. We are also grateful to Shawn Nissen for performing measurements for the reliability checking in experiment 1 and to Amy Hamilton for performing the additional measurements in experiment 2. This work was supported by NIH grants R01-DC005794 and R03-DC003762, by a seed grant from the Northwestern University Program in Culture, Language and Cognition, and by a Northwestern University Individual Research Grant.

References

Allen, J.S.; Miller, J.L.: Listener sensitivity to individual talker differences in voice-onset-time. J. acoust. Soc. Am. *115:* 3171–3183 (2004).

Bent, T.; Bradlow, A.R.: The interlanguage speech intelligibility benefit. J. acoust. Soc. Am. *114:* 1600–1610 (2003).

Bent, T.; Bradlow, A.R.; Smith, B.: Segmental errors in different word positions and their effects on intelligibility of non-native speech: all's well that begins well; in Munro, Bohn, Festschrift for James E. Flege (Benjamins, Amsterdam 2007).

Bradlow, A.R.: Training non-native language sound patterns: lessons from training Japanese adults on the English /r/-/l/ contrast; in Hansen, Zampini, State-of-the-art issues in second language phonology (in press).

Bradlow, A.R.; Bent, T.: Listener adaptation to foreign-accented English. Cognition *106:* 707–729 (2008).

Bradlow, A.R.; Torretta, G.M.; Pisoni, D.B.: Intelligibility of normal speech. I: Global and fine-grained acoustic-phonetic talker characteristics. Speech Commun. *20:* 255–272 (1996).

Chen, M.: Vowel length variation as a function of the voicing of the consonant environment. Phonetica *22:* 129–159 (1970).

Chen, Y.: Production of the tense-lax contrast by Mandarin speakers of English. Folia phoniat. logop. *58:* 240–249 (2006).

Clark, C.M.; Garrett, M.F.: Rapid adaptation to foreign-accented English. J. acoust. Soc. Am. *116:* 3647–3658 (2004).

Crowther, C.S.; Mann, V.: Native language factors affecting use of vocalic cues to final consonant voicing in English. J. acoust. Soc. Am. *92:* 711–722 (1992).

Crystal, T.H.; House, A.S.: The duration of American-English vowels: an overview. J. Phonet. *16:* 263–284 (1988).

Denes, P.: Effect of duration on perception of voicing. J. acoust. Soc. Am. *27:* 761–764 (1955).

Derwing, T.; Munro, M.: Accent, intelligibility, and comprehensibility: evidence from four L1s. Stud. Second Lang. Acquis. *19:* 1–16 (1997).

Escudero, P.: The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels; in Do, Dominguez, Johansen, Proc. 25th Annu. Boston Univ. Conf. on Lang. Dev., pp. 250–261 (2001).

Flege, J.E.: Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. J. acoust. Soc. Am. *89:* 395–411 (1991).

Flege, J.E.: Production and perception of a novel, second language phonetic contrast. J. acoust. Soc. Am. *93:* 1589–1608 (1993).
Flege, J.E.; Eefting, W.: Linguistic and developmental effects on the production and perception of stop consonants. Phonetica *43:* 155–171 (1986).
Flege, J.E.; Hillenbrand, J.: Differential use of temporal cues to the /s/–/z/ contrast by native and non-native speakers of English. J. acoust. Soc. Am. *79:* 508–517 (1986).
Flege, J.E.; Munro, M.; Skelton, L.: Production of the English word-final /t/-/d/ contrast by native speakers of Mandarin and Spanish. J. acoust. Soc. Am. *92:* 128–143 (1992).
Flege, J.E.; Port, R.: Cross-language phonetic interference: Arabic to English. Lang. Speech *24:* 125–146 (1981).
Hazan, V.; Markham, D.: Acoustic-phonetic correlates of talker intelligibility for adults and children. J. acoust. Soc. Am. *116:* 3108–3118 (2004).
Imai, S.; Walley, A.C.; Flege, J.E.: Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. J. acoust. Soc. Am. *117:* 896–907 (2005).
Klatt, D.H.: Linguistics uses of segmental duration in English: acoustic and perceptual evidence. J. acoust. Soc. Am. *59:* 1208–1221 (1976).
Laeufer, C.: Patterns of voicing-conditioned vowel duration in French and English. J. Phonet. *20:* 411–440 (1992).
Mack, M.: Voicing-dependent vowel duration in English and French: monolingual and bilingual production. J. acoust. Soc. Am. *71:* 173–178 (1982).
Maye, J.; Werker, J.; Gerken, L.: Infant sensitivity to distributional information can affect phonetic discrimination. Cognition *82:* B101–B111 (2002).
Munro, M.; Derwing, T.: Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. Lang. Learn. *45:* 73–97 (1995).
Nygaard, L.C.; Pisoni, D.B.: Talker-specific learning in speech perception. Perception Psychophysics *60:* 355–376 (1998).
Nygaard, L.C.; Sommers, M.S.; Pisoni, D.B.: Speech perception as a talker contingent process. Psychol. Sci. *5:* 42–46 (1994).
Peterson, G.E.; Lehiste, I.: Duration of syllable nuclei in English. J. acoust. Soc. Am. *32:* 693–703 (1960).
Pierrehumbert, J.: Probabilistic phonology: discrimination and robustness; in Bod, Hay, Jannedy, Probability Theory in Linguistics (MIT Press, Cambridge 2003).
Rogers, C.L.: Intelligibility of Chinese-accented English; doct. diss. Indiana University, Bloomington (unpublished, 1997).
Schmidt, A.M.; Flege, J.E.: Effects of speaking rate changes on native and nonnative speech production. Phonetica *52:* 41–54 (1995).
Smith, B.L.: Variations in temporal patterns of speech production among speakers of English. J. acoust. Soc. Am. *108:* 2438–2442 (2000).
Smith, B.L.: Effects of speaking rate on temporal patterns of English. Phonetica *59:* 232–244 (2002).
Smith, B.L.; Hillenbrand, J.; Ingrisano, D.: A comparison of temporal measures of speech using spectrograms and digital oscillograms. J. Speech Hear. Res. *29:* 270–274 (1986).
Stevens, K.N.: Acoustic phonetics (MIT Press, Cambridge 1998).
Strange, W.; Bohn, O.-S.: Dynamic specification of coarticulated German vowels: perceptual and acoustic studies. J. acoust. Soc. Am. *104:* 488–504 (1998).
Tajima, K.; Port, R.; Dalby, J.: Effects of temporal correction on intelligibility of foreign-accented English. J. Phonet. *25:* 1–24 (1997).