



Perceptual adaptation to non-native speech

Ann R. Bradlow *, Tessa Bent ¹

Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208, USA

Received 27 October 2006; revised 4 April 2007; accepted 7 April 2007

Abstract

This study investigated talker-dependent and talker-independent perceptual adaptation to foreign-accent English. Experiment 1 investigated talker-dependent adaptation by comparing native English listeners' recognition accuracy for Chinese-accented English across single and multiple talker presentation conditions. Results showed that the native listeners adapted to the foreign-accented speech over the course of the single talker presentation condition with some variation in the rate and extent of this adaptation depending on the baseline sentence intelligibility of the foreign-accented talker. Experiment 2 investigated talker-independent perceptual adaptation to Chinese-accented English by exposing native English listeners to Chinese-accented English and then testing their perception of English produced by a novel Chinese-accented talker. Results showed that, if exposed to multiple talkers of Chinese-accented English during training, native English listeners could achieve talker-independent adaptation to Chinese-accented English. Taken together, these findings provide evidence for highly flexible speech perception processes that can adapt to speech that deviates substantially from the pronunciation norms in the native talker community along multiple acoustic-phonetic dimensions.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Speech perception; Perceptual learning; Foreign-accented speech

* Corresponding author. Tel.: +1 847 491 8054; fax: +1 847 491 3770.

E-mail address: abradlow@northwestern.edu (A.R. Bradlow).

¹ Present address: Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47408, USA.

1. Introduction

Common wisdom tells us that listeners with extensive contact with non-native speakers are more adept at understanding foreign-accented speech than listeners with little or no exposure to foreign-accented speech even if the initial contact occurred relatively late in life. How exactly does this listener adaptation to non-native speech occur? To what extent does adaptation to foreign-accented speech resemble perceptual learning for native accented speech? What changes from the level of the acoustic encoding of the speech signal to its cognitive and linguistic representation does the listener undergo during this process of adaptation? And, what exposure conditions best induce this listener adaptation to foreign-accented speech?

Recent work in perceptual learning for native accented speech (e.g. Eisner & McQueen, 2005, 2006; Kraljic & Samuel, 2005, 2006, 2007; Maye, Aslin, & Tanenhaus, 2003; Norris, McQueen, & Cutler, 2003) has demonstrated impressive cognitive flexibility as a fundamental listener strategy for handling variability in the speech signal. In the present study, we build on this general idea of perceptual learning for speech by focusing on the particular case of listener adaptation to foreign-accented speech. Foreign-accented speech provides a unique window into the processes that underlie listener adaptation to speech as it represents a variety of speech that differs markedly and along multiple acoustic-phonetic dimensions from the pronunciation norms in the native talker community. Yet, since the features of any given foreign accent arise primarily from the interaction of the phonological structures of the talker's native language and the target language, the phonetic characteristics of foreign-accented speech are highly systematic and quite consistent across talkers from the same native language background. Thus, whereas the literature on perceptual learning for native-accented speech has examined listener adaptation to speech with rather limited variation from the "standard" variety, the focus of the present study on adaptation to foreign-accented speech allowed us to investigate perceptual learning for a variety of speech with a high degree of naturally-produced, systematic variability.

There is some evidence in the previous speech perception literature for listener adaptation to foreign-accented speech. Clarke and Garrett (2004) showed that on a cross-modal word verification task (in which subjects must assess whether a visually presented word matches the final word of a previous, auditorally-presented sentence), exposure to less than 1 min of speech by a foreign-accented talker was sufficient for listeners to overcome the initial decrease in processing speed for foreign-accented versus native-accented speech. This rapid adaptation to the speech of an individual foreign-accented talker was demonstrated in the case of both Spanish- and Chinese-accented English. Furthermore, Weil (2001) showed that training with a single Marathi-accented talker of English facilitated perception of sentences produced by a novel Marathi-accented talker to the extent that post training sentence recognition accuracy with the novel Marathi-accented talker was equivalent to the post training performance with the trained Marathi-accented talker.

Perceptual adaptation has also been demonstrated for various other cases of speech by "special" talker populations including perceptual adaptation to speech

produced by talkers with hearing impairments (McGarr, 1983), to computer synthesized speech (Greenspan, Nusbaum, & Pisoni, 1988; Schwab, Nusbaum, & Pisoni, 1985), to time-compressed speech (e.g. Dupoux & Green, 1997; Pallier et al., 1998) and to noise noise-vocoded speech (Davis, Johnsrude, Hervais-Ademan, Taylor, & McGettigan, 2005). For example, McGarr (1983) showed that teachers of the deaf, speech-language pathologists and audiologists in schools for the deaf scored consistently higher than inexperienced listeners on word and sentence recognition tests with speech produced by children with hearing impairments, presumably due to their long-term adaptation to the speech patterns exhibited by this particular talker population. In a series of studies investigating the effect of laboratory-based training procedures on recognition accuracy of synthetic speech, Greenspan et al. (1988) and Schwab et al. (1985) demonstrated robust and long-lasting adaptation by listeners exposed to synthetic speech during training in contrast to no improvement in synthetic speech recognition by listeners who received no explicit training or by listeners who were exposed to natural speech during the training phase. Similarly, quite rapid and generalized listener adaptation to time-compressed speech has been demonstrated (e.g. Dupoux & Green, 1997; Pallier et al., 1998). These studies showed that adaptation to time-compressed speech occurs at a relatively abstract level rather than exclusively at a low level of tuning to the acoustic effects of time-compression as it can extend from a trained talker to a novel talker (Dupoux & Green, 1997), as well as from a trained language to a closely-related novel language (Pallier et al., 1998). Finally, using a signal manipulation technique (noise vocoding) that simulates the input to an electrical hearing device such as a cochlear implant, Davis et al. (2005) demonstrated dramatic improvements in perception of noise-vocoded sentences following relatively brief perceptual training. This laboratory demonstration of listener adaptation to noise-vocoded speech provided a highly effective model for studying the processes of auditory learning that are typically experienced by cochlear implant recipients in the period immediately following implantation or following any subsequent device adjustment. This study also showed the involvement of top-down, lexically driven mechanisms in listener adaptation to noise-vocoded sentences. Taken together, these studies indicate that intelligibility of speech that deviates substantially from the norms in the native, normal talker community is not necessarily persistently compromised, and that surprisingly high levels of perceptual accuracy can be achieved if the listener has sufficient experience listening to the “special” speech.

The recent literature on listener adaptation to native-accented speech has also yielded some important generalizations relevant to the present study of adaptation to foreign-accented speech. In a seminal study, Norris et al. (2003) exposed one group of Dutch listeners to Dutch words in which the final /f/ had been replaced by an ambiguous sound between /f/ and /s/ (the “/f/ group”), and exposed another group of comparable listeners to Dutch words in which the final /s/ had been replaced by the same ambiguous sound (the “/s/ group”). Following the exposure phase (during which listeners performed a lexical decision task with items including the modified items), the two groups differed in their categorization patterns along an /f/–/s/ continuum: listeners in the /f/ and /s/ groups tended to categorized ambiguous stimuli along the /f/–/s/ continuum as /f/ and /s/, respectively. An important

finding of this study was that this perceptual learning for speech was lexically-driven: the same categorization shift was not observed following exposure to non-words with the ambiguous sound in final position. A series of subsequent studies using the same paradigm showed that this perceptual learning for speech is quite resistant to decay over time (Eisner & McQueen, 2006; Kraljic & Samuel, 2005) and can generalize across items and talkers (Kraljic & Samuel, 2006, 2007). Furthermore, using a slightly different paradigm in which perceptual adaptation was measured on the basis of word judgments in a lexical decision task, Maye et al. (2003) demonstrated adaptation to a synthetic manipulation of vowels in natural speech that was modeled after a real dimension of dialectal variation in American English (i.e. vowel lowering as in “wetch” for “witch”). After exposure to the speech containing words with the synthetically lowered mid-vowel, listeners were more likely to accept novel words with lowered vowels as exemplars of words with /I/ in the unmodified speech.

In the present study, we build on this background and general idea of lexically-driven perceptual learning for speech by investigating both talker-dependent (Experiment 1) and talker-independent (Experiment 2) adaptation to foreign-accented speech. In an attempt to understand the mechanisms that underlie this perceptual adaptation, the two experiments presented in this study manipulated the extent to which listeners could rely on access to higher-level (i.e. lexical, prosodic and semantic-contextual) information in their task of adaptation to the foreign-accented talker(s). In particular, we investigated adaptation to foreign-accented English under conditions of listener exposure to foreign-accented speech that varied substantially in terms of its baseline level of overall sentence intelligibility for native listeners. The rationale behind this manipulation was that if access to higher-level information that is available in sentence-length stimuli facilitates perceptual adaptation to foreign-accented speech as it does in the case of lexically-driven adaptation to native-accented speech (Norris et al., 2003), then listeners should show faster and more extensive adaptation to foreign-accented speech with a relatively high baseline level of sentence intelligibility than to foreign-accented speech with a relatively low baseline level of sentence intelligibility. Alternatively, if accurate sentence recognition is not a particularly efficient or effective “wedge” into perceptual adaptation to foreign-accented speech, then the efficiency and/or extent of listener adaptation to various foreign-accented talkers should not be tied to variation in baseline sentence intelligibility of the foreign-accented speech samples.

2. Experiment 1

Experiment 1 investigated native listener adaptation to individual foreign-accented talkers that varied in their baseline sentence intelligibility score. We examined two measures of listener adaptation. First, we compared sentence recognition accuracy by native English listeners when presented with the speech of individual foreign-accented talkers in single- versus multiple-talker presentation formats. Additionally, we compared performance across the first, second, third and fourth quartiles of the single-talker test session as a means of tracking listener adaptation to

the talker over the course of exposure in the single-talker test session. The listeners in this experiment were all native speakers of American English while the talkers were non-native speakers of English with various levels of English production proficiency.

One possible outcome of this experiment is that the extent of native listener adaptation to a foreign-accented talker would depend on the amount of exposure to speech produced by that talker regardless of the talker's baseline sentence intelligibility. This possibility is consistent with the view that the processes of perceptual adaptation involve only (or primarily) adjustments at an early, pre-lexical stage of speech processing and that the recognition of word-, phrase- and sentence-sized units is not a source of feedback that facilitates the processes of perceptual adaptation. Note that this view does not predict that the baseline sentence intelligibility differences observed before listener adaptation will be neutralized after adaptation. Rather, this view predicts a constant amount of native listener speech perception improvement following a fixed period of exposure to non-native talkers with varying rates of sentence intelligibility. That is, the extent of native listener adaptation will not depend on the rate of initial sentence intelligibility across the various non-native talkers.

An alternative possible outcome is that the rate and/or extent of adaptation to a foreign-accented talker depend(s) on the talker's baseline sentence intelligibility, with the more likely scenario being faster and/or more extensive adaptation to a high-intelligibility talker than to a low-intelligibility talker. This result would suggest that the perceptual modifications and generalizations involved in listener adaptation are facilitated when the foreign-accented speech signal is less ambiguous with respect to linguistic category membership at the lexical and higher levels. An implication of this possible outcome is that the processes of perceptual adaptation to foreign-accented speech involve the integration of information across the acoustic, lexical and higher levels of linguistic structure rather than a more simple, unidirectional, low-level adjustment to the processing of incoming speech signals.

2.1. *Materials*

Materials for Experiment 1 came from the Northwestern University Foreign-Accented English Speech Database (NUFAESD). This database (described in detail in [Bent & Bradlow, 2003](#)) contains recordings of 64 sentences each produced by 32 non-native talkers of English from a variety of native language backgrounds. All talkers were graduate students at Northwestern University with an average age of 25.5 years (range 22–32 years). They had been studying English since an average of 12 years of age (range 5–18 years) for an average of 9.8 years (range 6–17 years). The talkers had been in the United States for an average of 2.7 months (range 0.25–24 months). The complete database includes recordings of four Bamford-Kowal-Bench (BKB) sentence lists by each non-native talker ([Bamford & Wilson, 1979](#); [Bench & Bamford, 1979](#)). Each list consists of 16 simple declarative sentences with 3 or 4 keywords each for a total of 50 keywords per list. The four BKB lists selected for inclusion in this database received equivalent intelligibility scores in a test with normal-hearing children conducted by the original test developers ([Bamford & Wilson, 1979](#)). The sentences were recorded in a sound-attenuated booth in the Phonet-

ics Laboratory in the Linguistics Department at Northwestern University on an Ariel Proport A/D sound board with a Shure SM81 microphone.

In addition to the sentence recordings, the database includes an intelligibility score for each talker based on results from sentence-in-noise recognition tests with native English listeners. For these intelligibility tests, the digital sentence recordings were equated in amplitude and embedded in white noise at a +5 dB signal-to-noise ratio. There were a total of 8 intelligibility test conditions each of which included all of the 64 sentences. In each of the 8 conditions, each of the 32 talkers in the database produced only 2 of the 64 sentences (32 talkers \times 2 sentences each = 64 sentences/condition). Each individual talker's intelligibility score was calculated on the basis of one set of 16 sentences (8 conditions \times 2 sentences/condition = 16 sentences). A total of 40 native listeners participated in these intelligibility tests (8 conditions \times 5 listeners/condition = 40 listeners). The listener's task was to listen to each sentence and transcribe it in standard English orthography. The intelligibility score for each talker was then calculated as the percentage of keywords correctly recognized by the listeners in response to all sentences produced by that talker across the full set of 8 test conditions (i.e. one of the four BKB lists). This score is therefore a measure of the talker's "baseline" intelligibility in sentences presented at a +5 dB signal-to-noise ratio in a multiple-talker presentation format when presented to adult native English listeners with normal speech and hearing.

The percent correct scores were converted to rationalized arcsine transform units (RAU). This transformation "stretches" out the upper and lower ends of the scale thereby allowing for valid comparisons of differences across the entire range of the

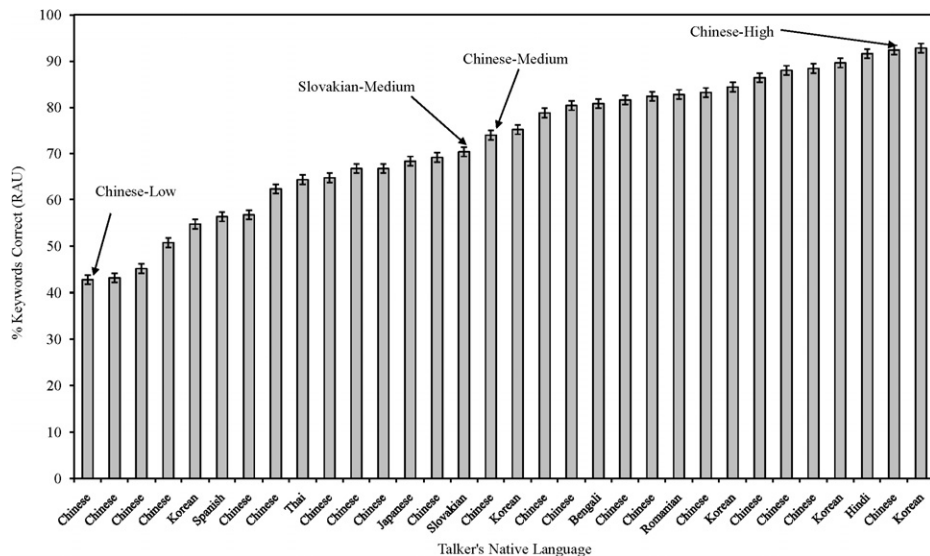


Fig. 1. Intelligibility scores for the group of 32 non-native talkers included in the Northwestern University Foreign-Accented English Speech Database (NUFAESD). Error bars indicate the standard deviation. Arrows indicate the four test talkers for Experiment 1.

scale (Studebaker, 1985). Scores on this scale range from -23 RAU (corresponding to 0% correct) to $+123$ RAU (corresponding to 100% correct). Intelligibility scores, in terms of keywords correctly recognized on the RAU scale, for the entire group of 32 talkers obtained from these tests with native English listeners ranged from 43 to 93 RAU with a distribution as shown in Fig. 1.

2.2. Talkers

Four test talkers were selected from the NUFAESD for this experiment (indicated with arrows on Fig. 1). Sentence productions from each of these four talkers were then used to obtain an intelligibility score in a single-talker test format. The native language of three of these four talkers was Chinese while the native language of the fourth talker was Slovakian. The three Chinese talkers covered the full range of baseline sentence intelligibility scores as determined from the multiple-talker test conducted at the time of the database compilation (as described above), and will be referred to as the Chinese-low, Chinese-medium and Chinese-high talkers. Their baseline sentence intelligibility scores were 43 RAU, 74 RAU and 93 RAU, respectively. The Slovakian talker's baseline sentence intelligibility was in the mid-range (70 RAU) and will therefore be referred to as the Slovakian-medium talker. All four talkers were male.

For each of the four test talkers, two single-talker test conditions were constructed. Within each condition, all four BKB lists were presented with the order of the first and fourth lists counter-balanced across conditions. For each talker, the list in the third quartile of both single-talker conditions was the list on which the baseline, multiple-talker intelligibility score was based (the multiple-talker score from the NUFAESD and shown in Fig. 1). As in the multiple-talker test, all sentences in the single-talker tests were embedded in white noise at a $+5$ dB signal-to-noise ratio.

2.3. Listeners

Four independent groups of listeners (20 each for the Chinese-low, Chinese-medium and Chinese-high talkers and 22 for the Slovakian-medium talker) participated in this experiment for a total of 82 listeners. All listeners were recruited from the Northwestern University Linguistics Department subject pool and received course credit for their participation. They were all native speakers of American English with normal speech and hearing at the time of testing. For each of the four talker conditions, half of the listeners participated in each of the two list order conditions. On each trial, the listener's task was to listen to the sentence stimulus and to transcribe it in standard English orthography on specially-prepared answer sheets. The sentence stimuli were played out through the computer soundcard (SoundBlaster Live) over headphones (Sennheiser HD 580). The listeners were seated in a sound-attenuated booth in front of a computer monitor. Stimulus presentation was controlled by experiment running software (Superlab Pro 2.01). The stimulus presentation rate was controlled by the subject but each file was presented only once with no possibility of repetition.

2.4. Results

Table 1 and Fig. 2 show average intelligibility scores (% keywords correctly recognized transformed on the RAU scale) across all listeners for all four talkers in the single-talker presentation format. Table 1 provides an overall score as well as scores by quartile. Also shown are the intelligibility scores from the multiple-talker test averaged across all talkers and listeners (i.e. across all eight conditions in the original NUFAESD test).

First, we compare intelligibility across the single- and multiple talker conditions for each of the four non-native talkers. This comparison should be taken with caution since the scores reported in the single-talker condition are based on listener exposure to 64 sentences by the talker in question, whereas the multiple-talker scores are based on listener exposure to just 16 sentences by that talker. For two of the four talkers the overall intelligibility score from the single-talker test was considerably higher than the baseline intelligibility score for that talker from the NUFAESD multiple-talker test (Chinese-low: 53 RAU vs. 43 RAU in the single- and multiple-talker tests, respectively; Chinese-medium: 81 RAU vs. 74 RAU in the single- and multiple-talker tests, respectively). There was no difference in intelligibility for the Chinese-high talker across the single- and multiple-talker tests, probably due to a ceiling effect at 93 RAU keywords correctly recognized. There was minimal improvement from the multiple- (70 RAU) to the single-talker (71 RAU) presentation format for the Slovakian-medium talker; however, as discussed further below, we did observe improvement across the quartiles of the single-talker test for this talker. The second major pattern in these data is the general upward progression of intelligibility scores across the four quartiles of the single-talker test for each of the four talkers. The lack of an upward trend in the multiple-talker test suggests that the upward trend in each of the single-talker tests is most likely due to listener adaptation to the talker rather than listener adaptation to the task.

The data were entered into a two-factor repeated measures ANOVA with Quartile (1, 2, 3 and 4) as a within-subjects factor and Talker (Chinese-low, Chinese-medium,

Table 1

Average intelligibility scores (% keyword correct transformed on the RAU scale) across all listeners for all four talkers in the multiple-talker presentation format and for each quartile in the single-talker presentation format

| | Condition | Quartile 1 | Quartile 2 | Quartile 3 | Quartile 4 | Overall |
|---------------|---|------------|------------|------------|------------|---------|
| Single-talker | Chinese-low (baseline = 43 RAU) | 51 (10) | 53 (7) | 49 (5) | 58 (9) | 53 (5) |
| | Slovakian-medium (baseline = 70 RAU) | 66 (11) | 71 (12) | 70 (10) | 77 (15) | 71 (9) |
| | Chinese-medium (baseline = 74 RAU) | 78 (14) | 73 (7) | 88 (10) | 90 (9) | 81 (8) |
| | Chinese-high (baseline = 93 RAU) | 85 (9) | 95 (11) | 100 (8) | 94 (9) | 93 (3) |
| | Multiple-talker | 80 (10) | 58 (10) | 76 (8) | 75 (11) | 72 (7) |

Standard deviations are shown in parentheses.

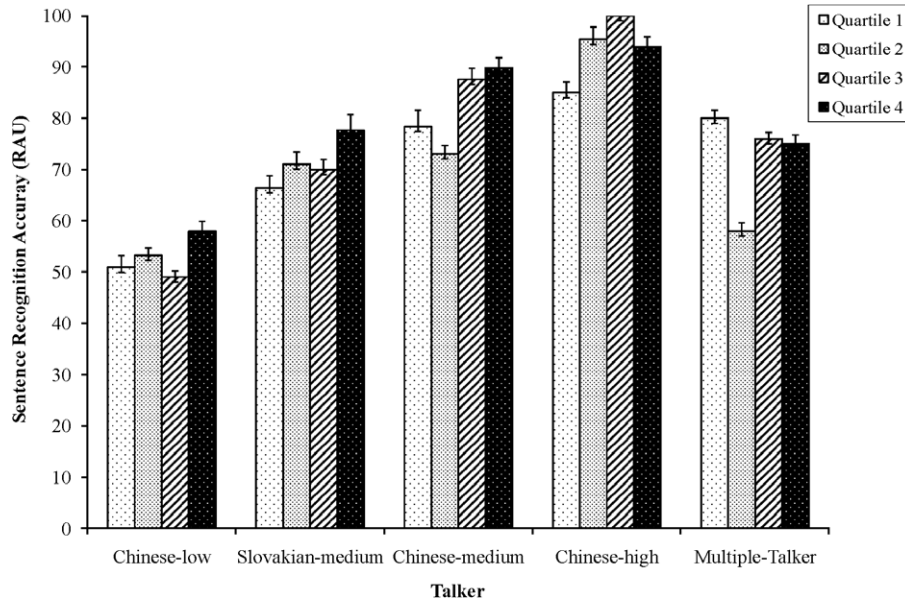


Fig. 2. Average intelligibility scores (% keyword correct transformed on the RAU scale) across all listeners for all four talkers in the multiple-talker presentation format and for each quartile in the single-talker presentation format. Error bars indicate standard errors.

Chinese-high, Slovakian-medium and multiple) as a between-subjects factor. Both main effects (Quartile: $F(3, 351) = 28.49$; Talker: $F(4, 351) = 92.49$) and the Quartile-by-Talker interaction ($F(12, 351) = 17.20$) were highly significant at the $p < .0001$ level. Two separate ANOVAs were then performed: a one-factor repeated-measures ANOVA for the multiple-talker test with Quartile as the within-subjects factor, and a two-factor repeated-measures ANOVA for the single-talker tests with Quartile as a within-subjects factor and Talker (Chinese-low, Chinese-medium, Chinese-high, Slovakian-medium) as a between-subjects factor. In the multiple-talker analysis, the effect of Quartile was highly significant ($F(3, 117) = 62.01$, $p < .0001$). Pair-wise comparisons showed that all quartiles were significantly different from each other (at the $p < .008$ level) except for quartiles 3 and 4 where there was no significant difference. However, as shown in Table 1 and Fig. 2, there is no upward trend as we go from the first to the fourth quartiles as would be expected if the listeners were showing adaptation to the transcription task. Instead the fluctuations across quartiles reflect idiosyncratic differences across the stimuli (particular sentences by particular talkers) in each of the four quartiles.

In the single-talker analysis, the main effects of Quartile and Talker were both significant (Quartile: $F(3, 234) = 22.96$; Talker: $F(3, 234) = 123.35$), as was the Quartile-by-Talker interaction ($F(9, 234) = 7.16$) at the $p < .0001$ level. We then performed separate analyses on the data for each talker in the single-talker tests with Quartile as a within-subjects factor. In each case, the effect of Quartile was highly

significant at the $p < .002$ level (Chinese-low: $F(3, 57) = 6.07$; Chinese-medium: $F(3, 57) = 15.44$; Chinese-high: $F(3, 57) = 13.15$; Slovakian-medium: $F(3, 63) = 7.61$).

Within each talker condition, pair-wise comparisons across the four quartiles showed that the Quartile-by-Talker interaction was due to a difference in the time it took for listeners to adapt to the different talkers. The overall pattern was that adaptation to relatively low-intelligibility talkers such as the Chinese-low and Slovakian-medium talkers required exposure to more speech by the talker than adaptation to relatively high-intelligibility talkers such as the Chinese-medium and Chinese-high talkers. Given the large number of pair-wise comparisons in each talker condition (6 comparisons), we set the significance level (Bonferroni/Dunn tests) at $p < .0083$. For the Chinese-low talker, the only significant gain was from the third to the fourth quartile and for the Slovakian-medium talker, the only significant gain was from the first to the fourth quartile. Thus, it appears that the listeners had some trouble adapting to the speech of these two relatively low-intelligibility talkers over the course of exposure to these four sentence lists. For the Chinese-medium talker, intelligibility generally improved by the third quartile (intelligibility in quartile 3 was significantly better than in quartile 2) with no additional improvement by the fourth quartile. However, the difference in intelligibility between the first and third quartiles for the Chinese-medium talker failed to reach significance at the specified level of $p < .0083$, (Quartile 1 vs. Quartile 3: $p = .0172$). For the Chinese-high talker, by the second quartile there was a significant improvement in intelligibility with no further improvement in the third and fourth quartiles.

The analyses discussed above compared the amount of exposure required for the listeners to show significant improvement in recognizing the speech of an individual foreign-accented talker. In other words, those analyses dealt with the rate of listener adaptation to the four test talkers within the span of a fixed amount of exposure (i.e. to four lists of 16 sentences for a total of 64 sentences). Given that the listeners showed adaptation to all four talkers by the end of the test session, we then compared the proportional improvement from the first to fourth quartiles across the four talkers to see whether the extent of improvement over the course of exposure to the four lists differed across the four talkers in the single-talker intelligibility tests. In this analysis, the dependent measure was proportional improvement which we defined as the difference in intelligibility between the first and fourth quartiles divided by the intelligibility score in the first quartile (i.e. $(Q4-Q1)/Q1$). A one-factor ANOVA showed no effect of Talker (Chinese-low, Chinese-medium, Chinese-high versus Slovakian-medium) on proportional improvement indicating that the amount of improvement from the beginning to the end of the test session was equivalent across the four talkers. The average amounts of improvement from the first to fourth quartiles for the Chinese-low, Chinese-medium, Chinese-high and Slovakian-medium talkers were 18.0 RAU, 17.8 RAU, 11.6 RAU and 18.3 RAU, respectively. (Note that these percentages differ from those calculated from the scores shown in Table 1 because they are calculated on the basis of individual listener scores within quartiles rather than on the average listener score for each quartile).

Since the lowest and highest intelligibility scores were not always obtained in the first and fourth quartiles, respectively, we also calculated the proportion improve-

ment across the lowest and highest quartiles for each individual listener and then compared this measure of proportion improvement across the four talkers. As in the (Q4–Q1)/Q1 analysis, this (max–min)/min analysis showed no effect of Talker. The average amounts of improvement from the quartile with the lowest intelligibility (min) to the quartile with the highest intelligibility (max) for the Chinese-low, Chinese-medium, Chinese-high and Slovakian-medium talkers were 38.0 RAU, 34.3 RAU, 26.3 RAU and 30.9 RAU, respectively. These analyses therefore indicate that the extent of improvement within the span of just four sentence lists is equivalent across the four talkers. However, it remains for future research to determine whether additional intelligibility improvements could be obtained for the talkers with relatively low baseline intelligibility if the listeners were given additional exposure.

2.5. Discussion

The purpose of this experiment was to investigate listener adaptation to individual talkers of foreign-accented English as a function of the talker's baseline level of English sentence intelligibility. The overall pattern of results showed that intelligibility of foreign-accented English improved with exposure to each of the individual talkers regardless of baseline intelligibility, but the amount of exposure required in order to achieve significant improvement in intelligibility increased as baseline intelligibility decreased. This pattern of perceptual learning is generally consistent with the claim that the process of “tuning” to a novel talker is facilitated by the integration of information across levels of representation. However, the fact that significant perceptual learning occurred for all of the talkers including the lowest intelligibility talker, Chinese-low for whom the baseline sentence intelligibility score was just 43 RAU, indicates that a high rate of words-in-sentence recognition is not a necessary condition for significant perceptual learning to occur and that native listeners can adapt to foreign-accented speech even under conditions where the initial speech recognition accuracy rate is quite low (under 50% correct recognition of keywords in sentences). It is also possible that the extent to which native listeners can adapt to a specific foreign-accented talker is determined by the quality rather than simply the quantity of exposure to the talker's speech. That is, faster adaptation may occur if initial exposure includes utterances that provide an adequate sampling of the individual talker's articulatory inventory than if the initial exposure provides only limited experience with the relevant acoustic-phonetic patterns.

The perceptual learning that underlies the native listener adaptation to individual foreign-accented talkers observed in this experiment involves some degree of generalization since the listeners had no prior experience with the particular sentence stimuli in the latter part of the test session (for which they showed improved perception relative to sentences presented in the earlier part of the test session). Thus the listeners learned something general about the individual talker's voice and articulatory patterns when producing English; however, the generalization was limited since the listeners were never required to transcribe speech that deviates substantially from previously experienced samples. That is, the range of variability over which the listeners were asked to generalize was rather narrow in terms of both talker- and utter-

ance-related features (the sentences were all of a similar type and structure). In Experiment 2 we extended this range by investigating talker-independent adaptation to a foreign-accent. That is, we attempted to train listeners on a foreign-accent that is shared across non-native talkers from the same native language background.

3. Experiment 2

Experiment 2 investigated the extent to which native listeners exhibit talker-independent adaptation to a particular foreign-accent, namely Chinese-accented English. As in Experiment 1, this experiment explored the prediction that foreign-accented talkers whose speech is more closely aligned with native-accented speech (i.e. has higher baseline sentence intelligibility) would promote more efficient talker-independent, accent-dependent adaptation than foreign-accented talkers to whose speech native listeners responded with a relatively low rate of sentence recognition.

3.1. Method

This training experiment involved a training phase followed by a post-test phase. The task in both phases was a sentence-in-noise recognition task in which recorded sentences were mixed with white noise at a +5 dB signal-to-noise ratio and then presented to the listeners for transcription in standard English orthography. The training phase involved two sessions administered over two consecutive days. Over the course of the two training sessions, the listeners transcribed five repetitions of two sets (one set per day) of BKB sentences (from the NUFAESD as described above). The second training session, was immediately followed by the post-test in which the listeners transcribed a third set of BKB sentences produced by a Chinese-accented talker (post-test 1) and a fourth set of 16 sentences produced by a Slovakian-accented talker (post-test 2). The talkers in post-tests 1 and 2 were the Chinese-medium talker and the Slovakian-medium talker from Experiment 1, respectively (indicated with arrows on Fig. 1). As shown in Fig. 1, the Chinese-medium and the Slovakian-medium talkers had intelligibility scores as assessed in the multiple-talker presentation format of 74 and 70 RAU, respectively. Performance in this training study was measured as percent keywords correctly transcribed on the two post-tests. Percent correct scores were converted to rationalized arcsine transform units (RAU) as described above.

As shown in Table 2, the overall design of this experiment involved three test conditions and two control conditions. The test conditions involved either five Chinese-accented training talkers (“Multiple-Talker” training, Condition 1) or a single Chinese-accented training talker (“Single-Talker” training, Conditions 2 and 3). In the Multiple-talker training condition (Condition 1), the intelligibility scores for the five training talkers were 79, 83, 86, 88 and 88 RAU (as assessed in the multiple-talker presentation format in the original compilation of the NUFAESD described above). The single-talker training conditions included a “Talker-Specific” training condition (Condition 2) in which the single training talker was the same as the Chinese-

Table 2
Talkers and listeners in the training conditions for Experiment 2

| | Condition | Training talker(s) | Baseline intelligibility (RAU) |
|---|--------------------------------|--|--------------------------------|
| 1 | Multiple-talker ($n = 10$) | 5 Chinese-accented | 79, 83, 86, 88, 88 |
| 2 | Talker-specific ($n = 10$) | Chinese-medium (same as in post-test 1 and Exp.1) | 74 |
| 3 | (a) Single-talker ($n = 10$) | Chinese-accented | 92 |
| | (b) Single-talker ($n = 10$) | Chinese-accented (also in condition 1) | 88 |
| | (c) Single-talker ($n = 10$) | Chinese-accented (also in condition 1) | 79 |
| | (d) Single-talker ($n = 10$) | Chinese-accented | 43 |
| 4 | Task control ($n = 10$) | 5 Native American English | – |
| 5 | No training ($n = 17$) | – | – |

accented talker in post-test 1 (i.e. Chinese-medium, also tested in Experiment 1 and indicated in Fig. 1). The other Single-talker training conditions (including Conditions 3a–d) varied with respect to the baseline intelligibility of the training talker as determined from the multiple-talker test conducted at the time of the database compilation (see NUFAESD description above). Intelligibility scores for the talkers in Conditions 3a, 3b, 3c and 3d were 92, 88, 79 and 43 RAU, respectively. The talkers in Conditions 3b and 3c were also talkers in the Multiple-talker training condition (Condition 1). The two control conditions involved “training” with five native American English male talkers to control for adaptation to the task (Condition 4) and no training (Condition 5).

Ten native American English listeners, recruited from the Northwestern University community, participated in each of the training conditions listed in Table 2. Seventeen listeners from the same population participated in the untrained control condition (Condition 5) for a total of 87 listeners. All listeners reported normal speech and hearing, and all were either paid or received course credit for their participation.

3.2. Results

Fig. 3 shows performance in RAU on the two post-tests in each of the training conditions. On each of the two post-tests, performance was equivalent following all of the single-talker training conditions (Conditions 3a–d). We therefore combined data from these two conditions into one single-talker training condition (Condition 3) in the figure and for the purposes of all further analyses. On the test with the Chinese-accented talker (left panel of Fig. 3), performance in the various training conditions fell into three groups. Performance was worst (65 RAU) in the no training control condition (Condition 5), intermediate (~ 82 RAU) in the single talker training conditions (Conditions 3a–d combined) and in the task control condition (“training” with five American English talkers, Condition 4), and best (~ 90 RAU) in the multiple talker and talker-specific training conditions (Conditions 1 and 2). Performance on post-test 2 with the Slovakian-accented talker was poor in the no training control condition (59 RAU) and only fair in all of the other training conditions (~ 72 RAU).

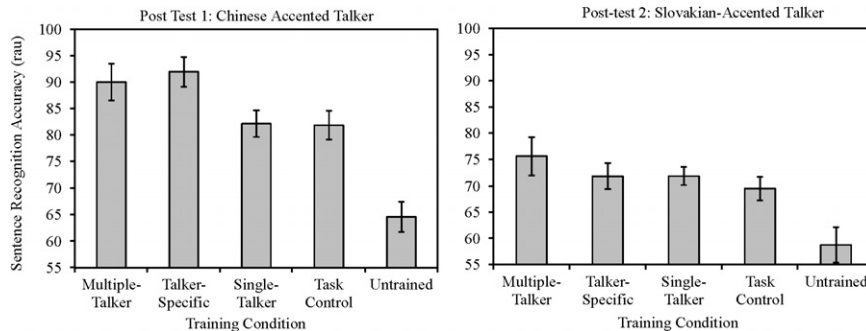


Fig. 3. Performance in RAU on the two post-tests in each of the training conditions in Experiment 2. Error bars represent the standard error of the mean. Training conditions are explained in Table 2 and in the accompanying text.

A two-way repeated measures ANOVA with post-test (Chinese-accented vs. Slovakian-accented) as a within-subjects factor and training condition as a between-subjects factor (Conditions 1, 2, 3, 4 and 5 as shown in Table 2) showed significant main effects of both post-test ($F(1, 82) = 92.90, p < .0001$) and training condition ($F(4, 82) = 14.036, p < .0001$). The post-test \times training condition interaction was also significant ($F(4, 82) = 2.860, p < .03$). Separate one-factor ANOVAs for each of the two post-tests showed significant effects of training condition (Chinese-accented talker: $F(4, 82) = 14.892, p < .0001$); (Slovakian-accented talker: $F(4, 82) = 6.211, p < .001$). For the post-test with the Chinese-accented talker (post-test 1), pair-wise comparisons (1-tailed, t -tests), showed that performance following the no training control condition (Condition 5) was significantly worse (at the $p < .0001$ level) than performance following all of the other conditions. Performance on post-test 1 following the task control condition (Condition 4) did not differ significantly from performance following the single talker training condition (Condition 3), but was poorer than performance following the talker-specific training condition ($p < .02$) and the multiple-talker training condition ($p < .05$). Similarly, performance on post-test 1 following the single talker training condition (Condition 3) was poorer than performance following the talker-specific training condition ($p < .02$) and the multiple-talker training condition ($p < .05$). Performance following the multiple talker training condition (Condition 1) and the talker-specific training condition (Condition 2) did not differ significantly from each other. For the post-test with the Slovakian-accented talker, pair-wise comparisons (1-tailed, t -tests) showed significantly worse performance (at the $p < .02$ level) in the no training control condition (Condition 5) relative to all of the other conditions. There were no significant differences across any of the other training conditions (Conditions 1–4).

In summary, results of this training study support four major conclusions. First, perception of sentences produced by a foreign-accented talker was better after practice with the task regardless of whether the training talker(s) came from the same language background as the test talker. This conclusion is supported by the finding that performance on the test with the Slovakian-accented talker (post-test 2) was better in

all conditions that involved practice with the task (Conditions 1–4) than in the no training condition (Condition 5), and by the finding that performance on the test with the Chinese-accented talker (post-test 1) was better following training with native American–English accented talkers (Condition 4) than in the no-training condition (Condition 5).

Second, no additional improvement in perception of sentences produced by a foreign-accented talker (over and above the task familiarity benefit) was gained through training with a single talker from the same language background as the test talker. This finding is demonstrated by the fact that performance on the test with the Chinese-accented talker in the single talker training condition (Condition 3) did not differ significantly from performance in the task training control condition (Condition 4). It is also noteworthy that performance did not differ across the various sub-conditions of the single talker training condition (Conditions 3a–d) which varied according to the baseline intelligibility of the training talker.

Third, training with multiple talkers from the same language background as the test talker resulted in a significant improvement in the perception of sentences produced by the test talker over and above the improvement gained through practice with the task. This is demonstrated by the significantly better performance in the multiple talker training condition (Condition 1) relative to the single talker training condition (Condition 3) and to the task control condition (Condition 4).

Finally, listener adaptation to Chinese-accented English following training with multiple Chinese-accented talkers was equivalent to the listener adaptation following training with the test talker himself. This is demonstrated by the equivalent levels of performance on the test with the Chinese-accented talker in the multiple talker training condition (Condition 1) and the talker-specific training condition (Condition 2).

3.3. Discussion

Taken together the findings of this training experiment indicate that, if exposed to multiple talkers of a Chinese-accented English, native listeners of American English can achieve talker-independent adaptation to Chinese-accented speech. Moreover, as shown by the post-test 1 accuracy score for the talker-specific training condition from Experiment 2 and the 4th quartile accuracy score for this talker in Experiment 1, exposure to five talkers of Chinese-accented English was as effective at enhancing sentence recognition accuracy for this talker (Chinese-medium) as exposure to that talker himself. Of course, this equivalence of performance does not necessarily imply that the talker-dependent adaptation on the one hand (talker-specific training from Experiment 2 and the single-talker condition from Experiment 1) and the talker-independent adaptation on the other hand (multiple-talker training in Experiment 2) involve the same underlying processes of perceptual learning. Nevertheless, it provides a convenient means of assessing the relative extents of talker-dependent and talker-independent listener adaptation to foreign accented speech and, from a practical point of view, suggests that exposure to multiple-talkers of a foreign-accent may be a highly effective means of enhancing speech communication between native and non-native speakers.

The results of post-test 2 with the Slovakian-accented talker demonstrate a noteworthy limit on listener adaptation to foreign-accented speech. Specifically, it does not appear that exposure to any one foreign accent (e.g. Chinese-accented English) promotes a degree of perceptual flexibility that facilitates recognition of any other foreign accent (such as Slovakian-accented English). However, it remains for future research to determine whether exposure to multiple foreign-accents (e.g. Chinese-, Japanese-, Korean-, Spanish- and Arabic-accented English) would promote accent-independent listener adaptation (e.g. to a novel, untrained accent such as Slovakian-accented English), or whether exposure to one accent would generalize to a typologically-related novel accent (e.g. would exposure to Spanish-accented English generalize to French-accented English?).

The finding that talker-independent adaptation to Chinese-accented English required exposure to multiple talkers of the accent is consistent with studies demonstrating the efficacy of a high-variability approach to non-native phoneme contrast training. For example, highly successful learning of the English /r-/l/ contrast by Japanese listeners has been demonstrated following training with words that placed /r/ and /l/ in various phonetic environments as produced by various native talkers of English. This learning has been shown to generalize to untrained words and to novel talkers, to transfer to improved production of the contrast and to be retained for at least 6 months following initial training (Bradlow, Pisoni, Yamada, & Tohkura, 1997, 1999; Lively, Logan, & Pisoni, 1993, 1994; Logan, Lively, & Pisoni, 1991; Yamada, 1993). Similarly successful learning has been demonstrated for other non-native contrasts following similar high-variability training procedures including training of English listeners on Chinese lexical tone contrasts (Wang, Jongman, & Sereno, 2003; Wang, Spence, Jongman, & Sereno, 1999), training of English and Japanese listeners on Hindi dental and retroflex stops (Pruitt, 1995), training English listeners on Japanese vowel length contrasts (Yamada, Yamada, & Strange, 1996), training Chinese listeners on English word-final /t/ and /d/ (Flege, 1995), and training English listeners on various German vowel contrasts (Kingston, 2003). Furthermore, in a recent study of perceptual learning of American English regional dialects, Clopper and Pisoni (2004) found that listeners who were exposed to multiple talkers of various regional American English dialects were significantly more accurate at classifying novel talkers from the trained dialect regions than listeners who were exposed to just one talker from each of the trained dialects. Taken together then, there appears to be ever-mounting evidence that exposure to highly variable training stimuli promotes, rather than interferes with, perceptual learning for speech be it at the level of phoneme or dialect/accent category representation.

In the present study we found that training with a single talker of Chinese-accented English, regardless of the training talker's baseline level of sentence intelligibility, was not effective at promoting accurate recognition of a novel Chinese-accented English talker. This finding contrasts with the previous finding of Weil (2001) that training with a single Marathi-accented talker of English facilitated perception of sentences produced by a novel Marathi-accented talker. Interestingly, in that study, post training performance with the novel Marathi-accented talker was equivalent to the post training performance with the trained Marathi-accented talker

(equivalent to the talker-specific training condition in the present study). A significant difference between the present study and the study of Marathi-accented English, in addition to the specific talkers and foreign accents involved, was the amount of training given to the listeners. In the present study, listeners were exposed to a total of 160 sentences consisting of 500 keywords over the course of two training sessions. In the study of listener adaptation to Marathi-accented English, listeners were exposed to a large amount of highly variable speech materials including isolated words, sentences, and passages over the course of three training sessions. Taken together, the results of these two studies may indicate that exposure to a wide range of stimuli as produced by a single talker and exposure to more limited stimuli as produced by multiple talkers may offer alternative means of achieving highly generalized perceptual adaptation to foreign-accented English.

Both of these strategies – exposure to a wide range of stimuli from a single talker or to a smaller range of utterance-types by multiple talkers – are consistent with the idea that robust and highly generalized perceptual learning for speech involves integration of information across levels of representation and processing as suggested by the literature on perceptual adaptation to native accented speech (Eisner & McQueen, 2005, 2006; Kraljic & Samuel, 2005, 2006, 2007; Maye et al., 2003; Norris et al., 2003). However, it should be noted that the work on perceptual learning for native-accented speech has indicated some talker- and phoneme-specificity to the observed perceptual learning. For example, Kraljic and Samuel (2007) found that training on a particular talker's voice onset times for the English /t-/d/ contrast resulted in an adjustment of voice onset time perception that generalized to a different place of articulation, such as for the /p-/b/ contrast, and to novel talkers. On the contrary, comparable training on the spectrally-cued /s-/ʃ/ contrast did not generalize to a novel talker. This pattern of perceptual learning suggests that generalization may be constrained by local, acoustic-phonetic features of the stimuli. In the talker-independent perceptual learning demonstrated in the present experiment, one can imagine that the observed generalization to a novel talker could be due to specific segment-level adjustments such as demonstrated by Kraljic and Samuel (2007) for voice onset time. Alternatively, the listeners in this experiment could have used higher-level, semantic-contextual information available from the sentences to assist in word recognition, which could have in turn fed backwards to more accurate identification at the phoneme level for all talkers who share an accent with the trained talkers. An important direction for future research will be to investigate these two possibilities further with regard to the specific case of native listener adaptation to a foreign accent, perhaps by varying the levels of linguistic representation made available to listeners in the training materials (e.g. would training with semantically anomalous sentences, phonotactically legal and illegal non-words show similar perceptual learning and generalization patterns?).

As a final methodological comment on this training experiment, we emphasize the value of including a task training condition as a baseline measure. In this study, listeners in the control condition who were exposed to native-accented English showed significantly better performance in the post-tests than the listeners in the untrained control condition presumably due to their training on the task of speech-in-noise

recognition. In fact, these listeners' experience with the task resulted in equivalent performance to the single-talker trained listeners on the post-test with the Chinese-accented talker and to all of the trained listeners on the post-test with the Slovakian-accented talker. While the sources of the equivalent performance may be different – experience with just the task in one case and experience with both the task and a foreign-accent in the other case – it gives us reason to be cautious regarding conclusions from training studies made on the basis of comparisons across trained and untrained groups who do not have equivalent experience with the test task (see also Reber & Perruchet, 2003 and subsequent commentaries for discussion of this issue in the literature on artificial grammar learning).

4. General Discussion

The overall goal of the present study was to investigate native listener adaptation to foreign-accented speech. Perceptual learning of speech has been quite widely demonstrated in the case of native-accented speech for native listeners (e.g. Eisner & McQueen, 2005, 2006; Kraljic & Samuel, 2005, 2006, 2007; Maye et al., 2003; Norris et al., 2003), and in other cases of speech by “special” talkers (for speech by children with hearing impairments, McGarr, 1983; for speech by computers, Greenspan et al., 1988; Schwab et al., 1985; for time-compressed speech, Dupoux & Green, 1997; Pallier et al., 1998 and for noise-vocoded speech, Davis et al., 2005). In this study we sought to broaden the empirical base for our understanding of the processes of listener adaptation to speech variability by focusing specifically on foreign-accented speech. Experiment 1 indicated that native listeners can adapt to the foreign-accented speech of a particular individual talker (talker-dependent adaptation) with some variation in the rate and extent of this adaptation depending on the baseline sentence intelligibility of the foreign-accented talker. Moreover, Experiment 2 indicated that native listeners can adapt to a foreign-accent as it extends across non-native talkers from the same native language background (talker-independent adaptation) provided they are exposed to multiple exemplars of the particular foreign-accent (e.g. Chinese-accented English). How exactly does this listener adaptation to foreign-accented speech occur? What changes from the level of the acoustic encoding of the speech signal to its cognitive and linguistic representation does the listener undergo during this process of adaptation?

In the case of talker-dependent adaptation to foreign-accented speech as demonstrated in Experiment 1, the present findings are consistent with the general idea of lexically-driven perceptual learning for speech as described in Norris et al. (2003) and subsequent studies. Our key finding in this regard was faster adaptation to relatively high intelligibility foreign-accented speech than to relatively low intelligibility foreign-accented speech, suggesting that perceptual learning was facilitated by the more consistent and accurate feedback from higher levels of linguistic structure in the case of the high intelligibility foreign-accented speech. It is important to note that since the foreign-accented speech stimuli in this study varied from native-accented speech and from each other along multiple

acoustic-phonetic dimensions, including segmental and supra-segmental levels, they do not allow us to draw firm conclusions regarding the levels of representation and processing involved in the perceptual adaptation by native listeners. Unlike previous studies that have typically involved synthetic manipulation along a single acoustic-phonetic dimension allowing for the isolation of the locus of perceptual adaptation, our foreign-accented speech stimuli consisted of unmodified natural speech allowing for adaptation along numerous interacting dimensions. It remains for future research to determine which acoustic-phonetic aspects of foreign-accented speech are easiest or hardest for listeners to adapt to and whether rapid adaptation to foreign-accented speech requires involvement of the lexicon and higher levels of linguistic structure available in sentence-length stimuli, or whether similarly rapid adaptation can occur with non-word foreign-accented stimuli that vary with respect to their segmental and supra-segmental accuracy relative to the native talker norms.

In addition to the role of variation in feedback from higher-levels of category structure in accounting for variation in the extent and efficiency of native listener adaptation to foreign-accented speech, it is possible that variation in adaptation is directly related to the extent of intra-talker variability within the individual non-native talkers' foreign-accented speech productions. That is, intra-talker variability within linguistic categories at various levels of representation (segments, syllables, words, and larger units involving phrase-level prosody) may be a significant factor in determining the initial intelligibility of a given non-native talker's speech and the ease with which native listeners can adapt to the talker's speech. Extensive intra-talker variability may be a feature of low proficiency foreign-accented talkers due to the instability inherent in the process of learning a foreign language. Moreover, this intra-talker variability may diminish as the individual converges on a more stable state of target language production that is highly functional in the community of language users. It remains for future research to verify these predicted relationships between intra-talker variability in speech production, initial intelligibility and ease of native listener adaptation.

In the case of talker-independent adaptation to foreign-accented speech as demonstrated in Experiment 2, it is possible that the underlying processes of perceptual learning in response to speech variability due to a foreign-accent are no different to the underlying processes of adaptation to naturally occurring variability across talkers and communicative settings in native accented speech. That is, long-term linguistic representations are updated in response to novel input, and, the broader the base of similarly novel input (e.g. in a multiple talker training condition) the more robust and generalizable are the long-term adjustments. Under this view, talker-independent perceptual learning of a novel foreign accent such as Chinese-accented English is just a particular case of the general phenomenon of perceptual adaptation to a group of similar sounding individuals who may or may not share any socio-linguistic category affiliation.

An alternative possibility is that talker-independent adaptation to a particular foreign-accent as it exists over a group of talkers involves the establishment of a novel and well-specified category label at a level of representation that is distinct from

linguistic category representational levels (where the categories are phonemes, words, etc.) Under this possibility, the listeners in the multiple talker training condition (Condition 1) and the talker specific training condition (Condition 2) of Experiment 2 may have achieved the same level of speech recognition accuracy in the post-test with the Chinese-accented talker through different means. The listeners in Condition 1 (multiple talker training) could have used their exposure to multiple Chinese-accented talkers to develop a Chinese-accented English category representation which could then guide their subsequent processing of incoming Chinese-accented speech. In contrast, the listeners in Condition 2 (talker specific training) could have performed their adjustments without the mediation of a category label at the level of accent/dialect representation. Under this view, we predict that these two groups of listeners will perform differently in a task of accent/dialect classification. That is, since the listeners in Condition 1 presumably developed a Chinese-accented category label they should be fast and accurate at classifying novel foreign-accented talkers as members of that category or not. In contrast, the listeners in Condition 2 presumably did not engage a level of dialect category representation during the speech recognition training procedure and therefore should be less able to perform an explicit dialect classification task. The relationship between talker and dialect identification training and subsequent speech recognition abilities has been investigated in previous work (e.g. Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994) and recent work has demonstrated perceptual learning for American English regional dialect classification (e.g. Clopper & Pisoni, 2004); however, it remains for future research to investigate the reverse situation (transfer of speech recognition training to dialect classification) in the specific case of foreign-accented speech.

The present study has provided evidence for the basic phenomenon of native listener adaptation to foreign-accented speech. However, there are numerous additional open questions that remain to be addressed for a comprehensive understanding of the mechanisms that underlie this remarkable flexibility of the normal speech perception system. First, it remains to be determined whether the perceptual learning involved in native listener adaptation to foreign-accented speech is vulnerable to decay. Eisner and McQueen (2006) reported retention of learning for at least 12 h by listeners who were trained to recognize an ambiguous sound as either /f/ or /s/ through exposure to natural speech in which words with the target phoneme were replaced by the ambiguous sound (lexically-biased adaptation along an /f-/s/ continuum). In contrast, Fenn, Nusbaum, and Margoliash (2003) reported that listeners who were trained to recognize synthetic speech exhibited some decay of learning over a 12-h period during which they were awake and exposed only to natural speech; however, there was a total recovery of learning following a night's sleep. Eisner and McQueen (2006) suggest that adaptation to synthetic speech may be more subject to decay than adaptation along a natural /f-/s/ continuum because it involves perceptual adjustments that are more drastic, occurring along multiple acoustic-phonetic dimensions, and require more effortful training with feedback. Investigations of retention of perceptual adaptation to foreign-accented speech, which also requires adjustments along multiple acoustic-phonetic dimensions, may shed light on the nature of this

general phenomenon of perceptual learning for speech with regard to its vulnerability to decay over time.

The present study has demonstrated remarkable flexibility on the part of the speech perception system; however, we can still wonder to what extent this flexibility is input-dependent and to what extent flexibility itself can be acquired. That is, can listeners who are exposed to multiple foreign accents acquire a general flexibility of speech perception such that they become highly proficient at recognizing speech from multiple, highly variable foreign-accented speech, or is the observed flexibility tied to the specific acoustic-phonetic features of the input? In a study of adaptation to a synthetic model of a novel dialect of American English, [Maye et al. \(2003\)](#) showed that the perceptual learning was tied to the specific vowel shift incorporated into their training stimuli. The listeners did not show a general vowel space shift that affected other vowels in the system suggesting that the learning was input-dependent and did not involve acquisition of a general perceptual flexibility. In the case of adaptation to foreign-accented speech, which involves adjustments along multiple acoustic-phonetic dimensions, one might imagine that listeners may be able to develop enhanced general flexibility in addition to accent-specific learning.

Recent work has suggested that auditory perceptual learning may be able to proceed without constant active, focused attention to the task. For example, [Sabin and Wright \(2006\)](#) reported that, on a task of perceptual learning with simple auditory stimuli (frequency discrimination), passive exposure to the stimuli following a brief period of active attention augmented learning. In fact, a combination of active and passive exposure during training resulted in equivalent discrimination threshold improvements to the gains made on the basis of exclusively active exposure during training. The idea of combined active and passive exposure as an effective speech training procedure is appealing because it would provide a conceptually “clean” link between laboratory-based speech training and real-world speech learning. It is very possible that real-world perceptual adaptation to foreign-accented speech occurs in a setting that is more like a combined active-passive training condition than a fully active laboratory-based training condition. That is, listeners are not always exposed to foreign-accented speech through the medium of active participation in a conversation but sometimes through more passive (over)hearing of foreign-accented speech in the general environment and media of mass communication.

Finally, an important issue that remains to be investigated is the extent to which native listener perceptual adaptation to foreign-accented speech transfers to native listener adjustments in speech production. Transfer of perceptual learning to the production domain has been demonstrated in the literature on non-native phoneme contrast learning (e.g. [Bradlow et al., 1997](#), [Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999](#); [Wang et al., 2003](#)) in response to high variability perceptual training procedures of the kind used in Experiment 2 of the present study. If the processes of speech perception are as highly flexible, dynamic and responsive to current input as the present study and related literature on perceptual learning for speech suggest, then we might imagine that if given sufficiently extensive and intensive exposure to foreign-accented speech, native talkers may begin to shift their pronunciations in the direction of the ambient foreign-accented speech. That is, we may begin to

observe large-scale shifts in production patterns across a population of language users in response to perceptual adaptations by individual listeners to a newly-encountered, contact variety of the target language. Under this scenario, we are likely to witness massive changes in English even as spoken by predominantly monolingual populations in response to increasing contact with non-native talkers.

Acknowledgements

We are grateful to L. Gebhardt, Y. Lee, L. Miller, A. Phillips, M. Shieh and K. Wong for data collection. This work was supported by NIH-NIDCD Grants. DC 03762 and R01-DC005794.

References

- Bamford, J., & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In J. Bench & J. Bamford (Eds.), *Speech-hearing tests and the spoken language of hearing-impaired children* (pp. 148–187). London: Academic Press.
- Bench, J., & Bamford, J. (Eds.). (1979). *Speech-hearing tests and the spoken language of hearing-impaired children*. London: Academic Press.
- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, *114*(3), 1600–1610.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985.
- Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*(4), 2299–2310.
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658.
- Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, *47*, 207–239.
- Davis, M. H., Johnsrude, I. S., Hervais-Ademan, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-coded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241.
- Dupoux, E., & Green, K. P. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 914–927.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics*, *67*(2), 224–238.
- Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability overtime. *Journal of the Acoustical Society of America*, *119*(4), 1950–1953.
- Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, *425*, 614–616.
- Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*, 425–442.
- Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 421–433.
- Kingston, J. (2003). Learning foreign vowels. *Language and Speech*, *46*(2–3), 295–349.

- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*, 141–178.
- Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Cognitive Psychonomic Bulletin & Review*, *13*(2), 262–268.
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*, 1–15.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242–1255.
- Lively, S. E., Pisoni, D. B., Yamada, R. A., Tokhura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*, 2076–2087.
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.
- Maye, J., Aslin, R., & Tanenhaus, M. (2003). In search of the weckud wetch: Online adaptation to speaker accent. In *Proceedings of the 16th Annual CUNY Conference on Human Sentence Processing*, March 27–19, Cambridge, MA.
- McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research*, *26*, 451–458.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, *60*, 355–376.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.
- Pallier, C., Sebastian Gallés, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition*, *26*(4), 844–851.
- Pruitt, J. S. (1995). Perceptual training on Hindi dental and retroflex consonants by native English and Japanese speakers. *Journal of the Acoustical Society of America*, *97*, 3417.
- Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *The Quarterly Journal of Experimental Psychology*, *56A*(1), 97–115.
- Sabin, A. T., & Wright, B. A. (2006). Contribution of passive stimulus exposures to learning on auditory frequency discrimination. *Association for Research in Otolaryngology Abstracts*, *29*, 6–7. A.
- Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors*, *27*, 395–408.
- Studebaker, G. A. (1985). A “Rationalized” arcsine transform. *Journal of Speech and Hearing Research*, *28*, 455–462.
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*, 1033–1043.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649–3658.
- Weil, S. A. (2001). Foreign-accented speech: Encoding and generalization. *Journal of the Acoustical Society of America*, *109*, 2473 (A).
- Yamada, R. A. (1993). Effect of extended training on /r/ and /l/ identification by native speakers of Japanese. *Journal of the Acoustical Society of America*, *93*, 2391.
- Yamada, T., Yamada, R. A. & Strange, W. (1996). Perceptual learning of Japanese mora syllables by native speakers of American English. In *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan.