Listener adaptation to foreign-accented English

Ann R. Bradlow[†] and Tessa Bent[†]

† Department of Linguistics, Northwestern University,

2016 Sheridan Road, Evanston, IL 60208, United States of America

E-mail: abradlow@northwestern.edu, t-bent@northwestern.edu

ABSTRACT

This study investigated talker-independent perceptual learning of a foreign accent by exposing native English listeners to Chinese-accented English and then testing their perception of English produced by a novel Chinese-accented talker. Training conditions involved exposure to either a single or multiple Chinese-accented talkers. Control conditions involved either no training or task training with native English talkers. A talker-specific training condition in which the training talker was the same as the Chinese-accented test talker was also included. Results showed better performance on the post-training test with the Chinese-accented talker after both the single and multiple talker training conditions relative to the no-training control. However, test performance following single talker training was equivalent to post-test performance following the task training condition (training with native English talkers), demonstrating a strong task familiarity effect. Performance following the multiple talker training was significantly better than following either the single talker or the task training conditions, and was equivalent to performance following the talker-specific training. These findings suggest that, with exposure to sufficient variability, listeners can develop highly generalized cognitive representations of Chinese-accented speech.

1. INTRODUCTION

There is much anecdotal evidence that perception of foreign-accented speech improves with increased familiarity with the talker and that perception of speech by a novel non-native talker is facilitated by previous experience with non-native talkers from the same native language background. The overall goal of the present study was to establish whether listeners can achieve both talker-specific and talker-independent adaptation to foreign-accented speech under controlled laboratory conditions. Evidence in favor of talker-specific adaptation would indicate that listeners develop and make use of highly detailed long-term cognitive representations of foreign-accented speech in much the same way as they do for native accented speech. Evidence in favor of talker-independent adaptation would suggest that these detailed cognitive representations are organized in a manner that allows for generalization across talkers with a shared foreign accent.

Previous research with native talkers and listeners has provided strong evidence for listener adaptation to an individual talker. For example, experiments have shown an intelligibility advantage for words presented in a single talker format over words presented in a multiple talker format [1], for novel words produced by a trained talker (in a voice recognition task) over novel words produced by an untrained talker [2, 3], for repeated words when the voice producing the second presentation of the word was judged to be similar in quality to the voice that produced the first presentation of the word [4], and for lexically "hard" words (i.e. words with many similar sounding neighbors with which they can easily be confused) presented in the last quarter of a testing session over words presented in the first quarter [5]. Evidence in favor of listener adaptation to an "accent" has been suggestive though relatively inconclusive. For example, when listening to speech produced by children with congenital and profound hearing losses, listeners with at least 1 year of experience listening to the speech of speakers with hearing losses had an advantage over inexperienced listeners [6], when listening to synthetic speech, listeners who received training with synthetic speech had an advantage over listeners with no prior exposure to synthetic speech [7, 8], and when listening to sentences produced by a novel Marathi-accented talker of English, listeners trained with words, sentences and short passages as produced by a Marathi-accented talker of English had an advantage over untrained listeners [9]. However, a recent study failed to find an intelligibility advantage for novel talkers of a trained foreign accent when the listeners were trained with a voice recognition task [10].

The overall goal of the present experiment was to shed additional light on this issue by investigating the effects of various training regimes for the purposes of promoting both talker-specific and talker-independent perceptual adaptation to foreign-accented English. Specifically, we investigated whether the high variability training approach that has proved to be highly effective for novel phoneme contrast learning would also be effective for novel accent learning.

2. METHOD

The training and test task was an open-set transcription task with recorded sentences that had been mixed with

white noise (+5 dB SNR). During two training sessions administered over two days listeners transcribed five repetitions of two sets of 16 simple English sentences (five repetitions of one set per day). Training was followed by a post-test in which the listeners transcribed a third set of 16 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 sentence lists were slightly modified versions of the lists in the Revised Bamford-Kowal-Bench Standard Sentence Test, which was developed for use with British listeners [11]. The revised test was developed by the Cochlear Corporation for use with American listeners. Each list consists of 16 sentences incorporating 50 keywords per list. Performance was measured as percent keywords correctly transcribed on the two post-tests. Percent correct scores were converted to rationalized arcsine transform units (rau) [12] which places the scores on a linear and additive scale, thus facilitating meaningful statistical comparisons across the entire range of the scale.

As shown in Table 1, the overall design involved five test conditions and two control conditions. The test conditions involved either five Chinese-accented training talkers (multiple talker training) or a single Chinese-accented training talker (single talker training). The two control conditions involved "training" with five native American English talkers (to control for adaptation to the task) and no training. The single talker training conditions included a talker-specific training condition in which the single training talker was the same as the Chinese-accented talker in post-test 1. The other single talker training conditions varied with respect to the intelligibility of the training talker as measured in a separate test of non-native talker intelligibility in which sentences from 32 non-native talkers were transcribed by native English listeners (n=40) in a multiple talker format.

Condition (# listeners)	Training Talker(s)
1: Multiple Talker (n=10)	5 Chinese-accented
2: Talker-specific training (n=10)	1 Chinese-accented (same as in post-test 1)
3: Single Talker (n=10)	1 Chinese-accented (also in condition 1)
4: Single Talker (n=10)	1 Chinese-accented (also in condition 1)
5: Single Talker (n=10)	1 Chinese-accented
6: Task control (n=10)	5 native American English
7: No training (n=17)	-----

**Table 1:** Talkers and listeners in the various conditions.

In the multiple talker training condition (condition 1), the intelligibility scores for the five training talkers were 79, 83, 86, 88 and 88 rau. Intelligibility scores for the talkers in single talker training conditions 3-5 were 79, 88, and 43 rau, respectively. Two of these talkers were also talkers in condition 1. The test talkers had intelligibility scores of

74 rau (Chinese-accented talker in post-test 1 and the talker-specific training condition) and 70 rau (Slovakian-accented talker in post-test 2).

All of the training and test talkers were male graduate students at Northwestern University. Listeners were recruited from the Northwestern University community. All talkers and listeners reported normal speech and hearing. The distribution of listeners across conditions is shown in Table 1. All talkers and listeners were either paid or received course credit for their participation.

3. RESULTS

Figure 1 shows performance in rau on the two post-tests in each of the training conditions.
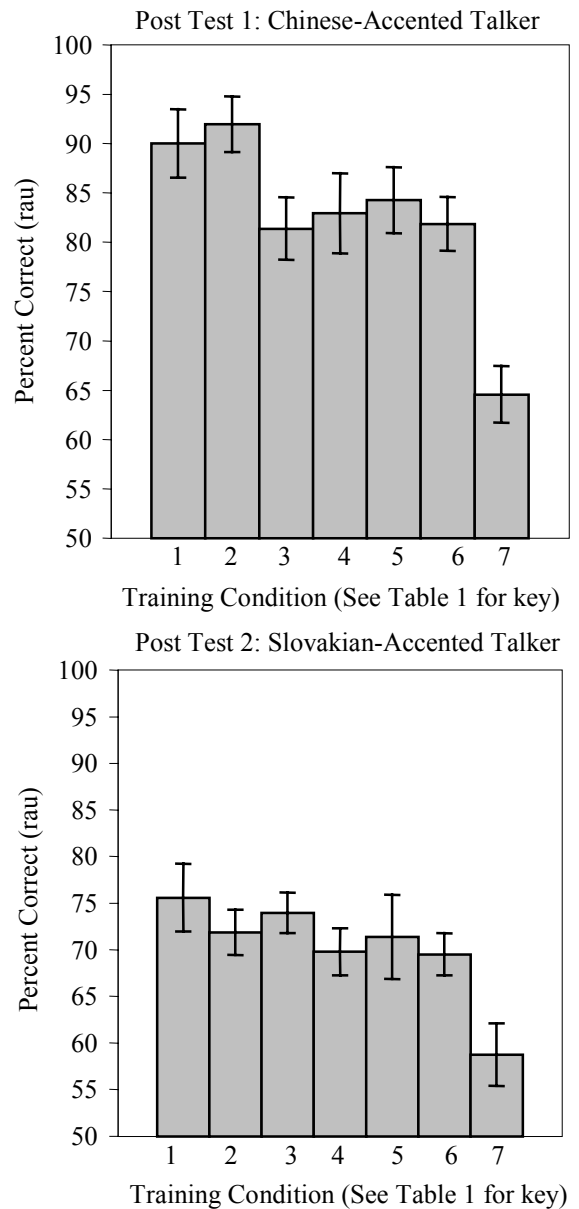
Figure 1: Post-test performance across training conditions. Error bars show the standard error.

On the test with the Chinese-accented talker, performance in the various training conditions fell into three groups. Performance was worst (65 rau) in the no training control condition, intermediate (approximately 82 rau) in the three single talker training conditions and the task control condition ("training" with five American English talkers), and best (approximately 90 rau) in the multiple talker and talker-specific training conditions. 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 (approximately 72 rau).

A two-way repeated measures ANOVA with post-test (Chinese-accented vs. Slovakian-accented) as the within-subjects factor and training condition as the between-subjects factor showed significant main effects of both post-test ($F(1,70)=98.56$, $p<.0001$) and training condition ($F(6,70)=8.14$, $p<.0001$). The post-test x training interaction was also significant ($F(6,70)=2.34$, $p<.05$). Separate one-factor ANOVAs for each of the two post-tests showed significant effects of training condition (Chinese-accented talker: $F(6,70)=9.42$, $p<.0001$; Slovakian-accented talker: $F(6,70)=3.85$, $p<.002$).

For the post-test with the Chinese-accented talker, pair-wise comparisons (1-tailed t-tests) showed no significant differences between any of the single talker training conditions. Furthermore, the task control condition did not differ significantly from any of the single talker training conditions. Performance in the no training control condition was significantly worse (at the $p<.001$ level) than in all of the other conditions. The multiple talker training condition and the talker-specific training condition did not differ significantly from each other. Moreover, performance in each of these two conditions was significantly better than performance in either the task control condition or in a combined single talker training condition (conditions 3, 4 and 5 combined) at the $p<.05$ level. For the post-test with the Slovakian-accented talker, pair-wise comparisons (1 tailed t-tests) showed significantly worse performance (at the $p<.01$ level) in the no training control condition relative to all of the other conditions. There were no significant differences across any of the other training conditions (conditions 1-6).

4. CONCLUSIONS AND DISCUSSION

The results of this training study showed that:

1. 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 than in the no training condition.

2. No additional improvement in perception of sentences produced by a foreign-accented talker (over and above the task familiarity benefit described above) was gained through training with a single talker from the same language background as the test talker. This is demonstrated by the fact that performance on the test with the Chinese-accented talker in the single talker training conditions did not differ significantly from performance in the task training control condition (i.e. training with native American English accented talkers).

3. 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 relative to the single talker training conditions and to the task control 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 and the talker-specific training condition.

Taken together these findings indicate that, if given appropriate exposure to a foreign accent, listeners can achieve talker-independent adaptation to foreign-accented speech. This finding is consistent with the previous finding that training with a single Marathi-accented talker of English facilitated perception of sentences produced by a novel Marathi-accented talker [9]. 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 (the talker-specific training condition). Recall that in the present study, only multiple talker training resulted in the same amount of learning as in the talker-specific training condition. The unexpectedly large amount of talker-independent learning obtained in the study with Marathi-accented English after single talker training may be specific to the talkers or particular foreign accent included in that study, or it may be due to the large amount of exposure given to the listeners in that study. Whereas the present study involved only two relatively brief training sessions with exposure to a total of 160 sentences (500 keywords), the Marathi-accented English training study involved three training sessions each of which involved considerably more exposure to the talker's productions than the present study's training sessions. Furthermore, the training materials in that study were highly variable in terms of stimulus type (isolated words, sentences, passages). If it can be established that the large amount of learning observed in the Marathi-accented English study is not specific to either the particular talkers or to Marathi-accented English, then the results of that study and of the present study together suggest that a relatively large amount of exposure to a single talker and a relatively small amount of exposure to multiple talkers can be traded off each other for the purpose of highly generalized perceptual adaptation to foreign-accented English. The

importance of talker variability for achieving highly generalized and rapid speech perception learning has been demonstrated conclusively in studies of novel phoneme acquisition [13]. The present study demonstrates that exposure to talker variability also facilitates rapid, talker-independent perceptual learning of a foreign accent which involves a wide range of acoustic-phonetic features.

The findings of the present study contrast with the previously reported failure to find talker-independent adaptation to a foreign accent following voice recognition training [10]. The task specificity of the training required to achieve talker-independent adaptation to a foreign accent (sentence recognition rather than voice recognition) suggests that during training listeners must be engaged in a speech recognition task that focuses attention on linguistic information (i.e. at the level of category contrasts), rather than in a voice recognition task that focuses attention on indexical information. This feature of listener adaptation to foreign-accented speech distinguishes it from listener adaptation to an individual native accented talker, as shown by studies that demonstrated a clear word recognition advantage for a trained voice following voice recognition training, [2, 3]. This contrast in training task required for listener adaptation to an individual native talker versus for talker-independent adaptation to an accent suggests that there is a qualitative difference between the ways that listeners handle variability across talkers within native-accented speech and variability across foreign-accented talkers. From the listener's perspective, talker variability that also involves native versus foreign accent variability is not merely an extreme case of talker variability within native-accented speech. Talker variability within native accented speech is reflected primarily in indexical properties which carry instance specific information, yet there is a high degree of consistency across native talkers with respect to language-specific phonetics and phonology. In contrast, foreign-accented speech deviates markedly from native accented speech in terms of patterns of phonetic implementation and with respect to underlying phonological structure. Therefore, while the process of adaptation to an individual native talker involves adaptation to idiosyncratic speech characteristics of the individual, the process of talker-independent adaptation to a foreign accent involves some modification to the cognitive representation of the structure of the sound system of the target language such that it can now accommodate the "deviant" patterns of the novel accent/dialect in question. Thus, this latter task requires attention to the linguistic content of the utterance and not just to the indexical properties of the utterance.

In conclusion, the present study has demonstrated that relatively rapid talker-independent perceptual adaptation to foreign-accented speech can be achieved under controlled laboratory conditions provided that during training, the trainees are exposed to multiple talkers and are engaged in a sentence recognition task. Under these training conditions, listeners can apparently develop rich and highly generalized cognitive representations of a novel foreign accent,

REFERENCES

[1] J. Mullennix, D. B. Pisoni, and C. S. Martin, "Some effects of talker variability on spoken word recognition", *Journal of the Acoustical Society of America*, **85**, pp. 365-378, 1989.

[2] L. C. Nygaard, M. S. Sommers, and D. B. Pisoni, "Speech perception as a talker contingent process", *Psychological Science*, **5**, pp. 42-46, 1994.

[3] L. C. Nygaard and D. B. Pisoni, "Talker-specific learning in speech perception", *Perception & Psychophysics*, **60**, pp. 355-376, 1998.

[4] S. D. Goldinger, "Words and voices: Episodic traces in spoken word identification and recognition memory", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22**, pp. 1166-1183, 1996

[5] A. R. Bradlow and D. B. Pisoni, "Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors", *Journal of the Acoustical Society of America*, **106** , pp. 2074-2085, 1999.

[6] N. S. McGarr, "The intelligibility of deaf speech to experienced and inexperienced listeners", *Journal of Speech and Hearing Research*, **26**, pp. 451-458, 1983.

[7] E. C. Schwab, H. C. Nusbaum, and D. B. Pisoni, "Effects of training on the perception of synthetic speech", *Human Factors,* **27**, pp. 395-408, 1985.

[8] S. L. Greenspan, H. C. Nusbaum, and D. B. Pisoni, "Perceptual learning of synthetic speech produced by rule", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, pp. 421-433, 1988.

[9] S. A. Weil, "Foreign-accented speech: Adaptation and generalization," MA thesis, The Ohio State University, 2001.

[10] C. M. Clarke, "Perceptual adjustment to foreign-accented English," *Journal of the Acoustical Society of America*, **107,** p. 2856 , 2000.

[11] J. Bench, and J. Bamford, Eds. *Speech-hearing tests and the spoken language of hearing-impaired children*, London, Academic Press, 1979.

[12] G. A. Studebaker, "A "rationalized" arcsine transform", *Journal of Speech and Hearing Research*, **28**, pp. 455-462, 1985.

[13] S. E. Lively, J. D. Logan, and D. B. Pisoni, "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**, pp. 1242-1255, 1993.