

Accent-independent adaptation to foreign accented speech

Melissa M. Baese-Berk

*Department of Communicative Sciences and Disorders, Michigan State University,
Oyer Center B-7, East Lansing, Michigan 48824
mbaese@msu.edu*

Ann R. Bradlow

*Department of Linguistics and Northwestern University Knowles Hearing Center,
Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208
abradlow@northwestern.edu*

Beverly A. Wright

*Department of Communication Sciences and Disorders and Northwestern University
Knowles Hearing Center, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208
b-wright@northwestern.edu*

Abstract: Foreign-accented speech can be difficult to understand but listeners can adapt to novel talkers and accents with appropriate experience. Previous studies have demonstrated talker-independent but accent-dependent learning after training on multiple talkers from a single language background. Here, listeners instead were exposed to talkers from five language backgrounds during training. After training, listeners generalized their learning to novel talkers from language backgrounds both included and not included in the training set. These findings suggest that generalization of foreign-accent adaptation is the result of exposure to systematic variability in accented speech that is similar across talkers from multiple language backgrounds.

© 2013 Acoustical Society of America

PACS numbers: 43.71.Hw, 43.71.Sy [SGS]

Date Received: November 12, 2012 **Date Accepted:** January 16, 2013

1. Introduction

The need to communicate with people who speak with a foreign accent is a frequent occurrence. For example, there are now more non-native than native talkers of English (e.g., Graddol, 1997, 2006; Jenkins, 2000), and interactions between these groups are increasingly common. While some foreign-accented talkers are immediately intelligible to native listeners even if they are easily identifiable as non-native talkers, others require considerable listener effort in order to be understood. Nevertheless, with appropriate experience, listeners are usually able to adapt to novel talkers and accents (Clarke and Garrett, 2004; Bradlow and Bent, 2008; Sidaras *et al.*, 2009). The broad question of interest in the current study is what circumstances during training facilitate highly generalized adaptation to foreign-accented speech. That is, what types of exposure during training enable listeners to better understand foreign-accented talkers whom they have not encountered before? Our hypothesis is that generalization of adaptation to foreign-accented speech in a given target language (in our case, English) is facilitated by exposure to systematic variability along the dimension(s) of desired generalization. We test this idea here through an examination of generalization to talkers who speak with a novel accent (i.e., who come from a native language background that is different than previously encountered foreign accented-talkers).

Consistent with the exposure-to-variability hypothesis, previous work demonstrated generalization of adaptation to foreign-accented speech to novel talkers within an accent (i.e., within a group of talkers who share a native language background) following exposure to multiple talkers (but not to a single talker) of that particular accent (Bradlow and Bent, 2008; Sidaras *et al.*, 2009). Thus, some aspect of exposure to multiple talkers with a common accent facilitates talker-independent learning. A possible candidate is the similarity in target language (*L2*) production across talkers from the same native language (*L1*) background that arises from *L1-L2* interactions at the segmental (e.g., Best *et al.*, 2001), phonotactic (e.g., Weber and Cutler, 2006), and prosodic (e.g., Baker, 2010) levels. For example, a common feature of Spanish-accented English is the insertion of a vowel before word-initial consonant clusters (e.g., “*esky*” for “*sky*”) due to the prohibition of such clusters in Spanish (Carlisle, 1991). Similarly, Dutch talkers frequently have difficulty producing the English vowel contrast in “*pen*” and “*pan*” because the vowel system in Dutch does not contain this contrast (Cutler *et al.*, 2004). Thus, listeners may learn to recognize the systematicity of these patterns across talkers from a specific language background, and hence to perceive these patterns in novel talkers from that language background. Moreover, this hypothesis also predicts that while exposure to multiple talkers from a given language background should result in generalization to a novel talker from that language background, it should not benefit recognition of speech in a novel accent because the dimension(s) of variability that are key to recognition of the novel accent likely will differ from those to which the listener has already been exposed. Bradlow and Bent (2008) reported just this result (adaptation to a novel talker of Mandarin-accented English but not to a talker of Slovakian-accented English following exposure to multiple talkers of Mandarin-accented English).

Here we test another prediction of the general hypothesis that exposure to systematic variability along the relevant dimension(s) facilitates generalization of adaptation to foreign-accented speech, namely, that listeners will generalize their adaptation to foreign accents to both untrained talkers and untrained accents if exposed to sufficient systematic variability across multiple accents. Systematic variation across accents results from typologically rare features of the target language sound structure. For example, the English vowel in the word “*but*” is difficult to produce for talkers from a wide range of native languages including Italian (Flege *et al.*, 1999), Japanese (Oh, *et al.*, 2011), and Spanish (Fox *et al.*, 1995) and probably many other languages since this vowel is absent from the vowel inventory of many languages. Similarly, non-native talkers of English from a variety of language backgrounds fail to reduce unstressed syllables to the same extent as native English talkers, thereby giving foreign-accented English a deviant overall temporal structure (Baker *et al.*, 2011). Another source of similarity across *L2* talkers from various *L1* backgrounds is the general difficulty of *L2* speech production encountered by *L2* talkers regardless of the particular *L1* and *L2* involved. Speaking a foreign language is more effortful and less automatic than speaking a native language, which typically results in a markedly slower speech rate for foreign-accented speech (e.g., Guion *et al.*, 2000). Therefore, it is possible that these sources of variability will result in systematicity in foreign-accented speech that provides listeners with the leverage they need in order to achieve accent-independent learning. To test this possibility, we exposed participants to the speech of multiple talkers who each came from a different language background. We then examined their recognition of foreign-accented speech by novel talkers from accent backgrounds both included and not included in the initial exposure set. If listeners are able to adapt to variability across individual talkers from multiple foreign accents (one talker per accent), they should adapt to both a novel accent and to a novel talker from a previously encountered accent. However, if listeners are only able to adapt to variability resulting from multiple talkers of a single-foreign-accent background, they should not adapt to a novel accent. Simultaneously, adaptation to novel talkers from previously encountered language backgrounds should be inhibited.

2. Methods

A total of 30 native, monolingual English listeners between 18 and 34 yrs old served as participants. None of the participants reported speech or hearing deficits, or had studied the native languages of the talkers used in either the exposure or test materials presented here. Furthermore, participants were screened for previous significant exposure to foreign-accented speech in general (beyond the particular foreign-accent represented in the materials used in this study), and had not participated in previous studies of adaptation to foreign-accented speech.

The overall design of this study closely follows the design of the training study from [Bradlow and Bent \(2008\)](#). Materials came from the Northwestern University Foreign-Accented English Speech Database (NUFAESD), which includes four Bamford–Kowal–Bench (BKB) sentence lists ([Bamford and Wilson, 1979](#); [Bench and Bamford, 1979](#)) recorded by a total of 32 non-native talkers of English from various native language backgrounds. Each list contained 16 simple declarative sentences with 3 or 4 keywords for a total of 50 keywords per list. The contents and design of this database are discussed at length in [Bent and Bradlow \(2003\)](#) and [Bradlow and Bent \(2008\)](#).

The present study involved two training sessions conducted over two consecutive days followed by two post-tests which were administered on the second day immediately following the second training session. During each of the two training sessions, the trainees were exposed to five repetitions of one list of BKB sentences ($5 \times 16 = 80$ sentences/day; 160 sentences total). Each repetition was produced by a talker from a different native language background (Thai, Korean, Hindi, Romanian, and Mandarin). Thus the trainees ($n = 10$) were exposed to five foreign-accented English talkers, each with a different accent. The training and test talkers were all males with mid-range intelligibility as assessed by native English listeners during the development of the NUFAESD. The sentences presented during the training phase were blocked by talker, and the order of the 16 sentences within each block was the same. In the post-tests, trainees were exposed to a third set of BKB sentences produced by a novel Mandarin-accented English talker (an accent encountered during training) and a fourth set produced by a novel Slovakian-accented English talker (a novel accent). Neither test talker was included in any training set. All sentences were mixed with white noise at a signal-to-noise ratio of +5 dB. The noise began 500 ms before and ended 500 ms after each sentence presentation. Sentences were presented over headphones at a comfortable listening level.

The listeners' task during both the training and the post-tests was to write down the sentences they heard. No feedback was given in either phase. In the post-test, the three to four keywords in each sentence were scored as either correct or incorrect, yielding an overall recognition accuracy score (proportion words correctly recognized).

We compared the performance of the multiple-foreign-accent training group to that of the no-foreign-accent and single-foreign-accent training groups from [Bradlow and Bent \(2008\)](#) (their task control and multiple-talker groups). The no-foreign-accent group ($n = 10$) served as a control group. Those trainees were exposed to five different native, male talkers of American English producing the BKB sentences. That is, they were trained with the same sentence list and with multiple talkers like all of the other groups but they were not exposed to any foreign-accented speech before the test. The single-foreign-accent group ($n = 10$) was exposed to a set of five male, Mandarin-accented English talkers. Thus, those listeners were exposed to a number of different talkers, all of whom had the same foreign accent. A direct comparison across all of these groups was possible because both the training and test procedures, and the talkers and sentences in the post-tests, were identical across all training groups in the present study and in [Bradlow and Bent \(2008\)](#).

To analyze the data, logistic mixed effects regression models were implemented in R package lme4 ([Bates and Mächeler, 2009](#)), with performance on the

post-tests as the dependent variable. The training group (no-foreign-accent, single-foreign-accent, and multiple-foreign-accent) was entered as a fixed effect. Random intercepts were included for listener and sentence, and random slopes were included for the training group (by items) and post-test talker (by subjects). The inclusion of random intercepts and slopes allowed for simultaneous modeling of participant and item effects within a single analysis. Separate regressions were performed for each post-test (i.e., proportion of words correct for the Mandarin and Slovakian post-tests were the dependent variables in two separate regression analyses). In order to assess the full set of comparisons among the three levels of the training group we re-ordered the dummy variables within the model twice so each level became the referent level for comparison.

3. Results

3.1 Generalization to a novel talker with a novel accent

Training on multiple accents, excluding Slovakian-accented English, resulted in generalization to a novel Slovakian-accented talker while training on multiple talkers of Mandarin-accented English did not (see Fig. 1). The logistic mixed effects regression model revealed that performance on the novel Slovakian-accented talker for the multiple-foreign-accent training group (current data) was significantly better than that for both the no-foreign-accent (data from Bradlow and Bent, 2008) group ($\beta = 0.94$, $z = 2.39$, $p < 0.02$) and the single-foreign-accent (data from Bradlow and Bent, 2008) group ($\beta = -0.58$, $z = -2.99$, $p < 0.005$). In contrast, performance on the Slovakian-accented talker did not differ significantly between the single-foreign-accent and the no-foreign-accent training groups ($z < 1$). Thus, after training on multiple accents, listeners generalized their learning to a novel talker from a language background that was not included in the training set.

3.2 Generalization to a novel talker with a familiar accent

Training on multiple different accents, one of which was Mandarin-accented English, resulted in as much generalization to a novel talker of Mandarin-accented English as did training on multiple talkers of Mandarin-accented English (see Fig. 2). The logistic

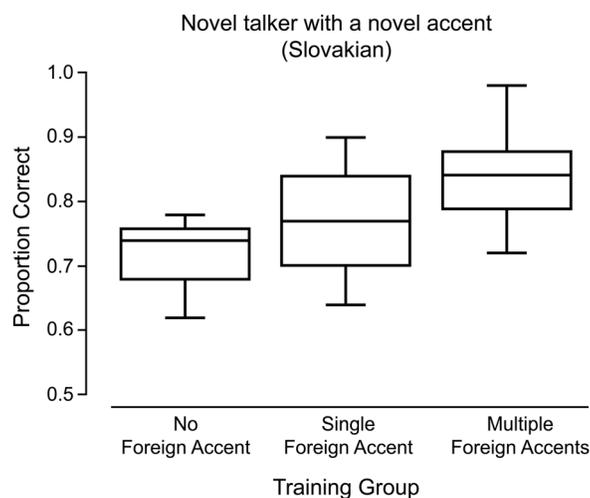


Fig. 1. Performance (proportion of keywords correct) on the post-test for a novel talker with a novel accent (Slovakian-accented English). Participants were trained on native English talkers (“No Foreign Accent”), five Mandarin-accented talkers (“Single Foreign Accent”), or five talkers from different native language backgrounds (“Multiple Foreign Accent”). Only participants in the Multiple-Foreign-Accent training group demonstrate generalization to a novel talker from a novel accent background.

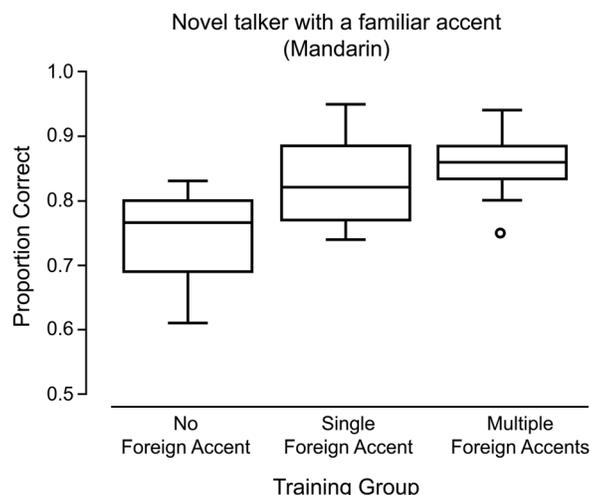


Fig. 2. Performance (proportion of keywords correct) on the post-test for a novel talker with a familiar accent (Mandarin-accented English). Otherwise as in Fig. 1. Participants in both the Single-Foreign-Accent and Multiple-Foreign-Accent groups demonstrate generalization to a novel talker from a familiar accent background.

mixed effects regression revealed that performance on the novel Mandarin-accented talker for both the single-foreign-accent and the multiple-foreign-accent training groups was significantly better than that of the no-foreign-accent training group ($\beta = 0.66$, $z = 2.1$, $p < 0.05$ and $\beta = 0.7$, $z = 2.25$, $p < 0.03$, respectively). Furthermore, performance did not differ between the single-foreign-accent and multiple-foreign-accent training groups ($z < 1$). Thus, after training on multiple accents, listeners generalized their learning to a novel talker from a language background that was included in the training set.

4. Discussion

In this study we examined what circumstances facilitate generalization of perceptual adaptation to foreign-accented speech. We hypothesized that exposure to systematic variability along specific dimensions results in generalization along those dimensions. Specifically, we examined whether exposure to talkers from a variety of native language backgrounds would allow listeners to generalize their adaptation to novel talkers from language backgrounds they had experience with, as well as to talkers from novel language backgrounds. In previous research, we have reported talker-independent adaptation to foreign-accented speech following exposure during training to multiple talkers of a single foreign accent (Bradlow and Bent, 2008; see also Sidaras *et al.*, 2009). Whereas talker variability in the earlier studies allowed for generalization of learning across talkers from a shared L1 background to a novel talker from that same L1 background (i.e., talker-independent but accent-specific generalization), accent variability in the present study allowed for generalization of learning across talkers from various accents to a novel talker with a novel accent (i.e., both talker- and accent-independent generalization).

We suggest that exposure to multiple foreign accents highlights commonalities across various foreign accents of (in this case) English. These commonalities may derive from typological peculiarities of the sound structure of English that provoke similar production difficulties for talkers from multiple language backgrounds and/or from the general difficulty of speaking in a non-native language, which results in accent-general features such as a slow speaking rate. The idea is that these commonalities lend systematicity to the variation across the accents, and that listeners use this systematic variability to generalize what they learned in training to a novel talker from

an unfamiliar language background. It should be noted that the specific similarities among the multiple foreign accents used in this study have not been identified in the materials used here. However, as mentioned in Sec. 1, there are some global properties that are similar across many foreign-accented talkers of English (e.g., difficulties with the crowded vowel space, diminished unstressed vowel reduction, and reduced speech rate) and the materials of the present study would presumably have represented all of these (and possibly other additional) general features of foreign-accented English.

This study also demonstrated that listeners generalize to a novel talker with an accent included in the multiple-accent training set (i.e., Mandarin-accented English), indicating that talker-independent/accent-dependent learning was not sacrificed in order for increased generalization of adaptation to occur in the multiple-accent group. That is, performance on the Mandarin-accented English post-test was not compromised by exposure to multiple foreign accents relative to exposure to multiple talkers of Mandarin-accented English. It is unlikely that the Mandarin-accented talker included in the training set is a “golden talker” whose inherent range of systematic variability allows generalization to a novel talker of Mandarin-accented English as well as to a novel accent even though the listeners were only exposed to this single talker; Bradlow and Bent (2008) reported that exposure to this exact talker alone did not generalize to a novel talker of Mandarin-accented English. The difference between the single-talker training from the earlier study and the multiple-foreign-accent training from the present study is the inclusion in the present study of accent variability from the talkers of four other *L1* backgrounds (i.e., the training set included five distinct foreign-accents, not just five talkers of a consistent foreign-accent).

A potential alternative explanation for the present results is that generally looser criteria for the match between speech input and lexical representations are employed when listeners are faced with non-optimal speech signals. This alternative has been offered for eye-tracking data demonstrating that listeners adjust their expectations regarding viable competitors for words presented in the context of other words that are partially masked by background noise (McQueen and Huettig, 2012) or subject to phonetic reduction (Brouwer *et al.*, 2012). Further investigation is required to help disambiguate this hypothesis from our proposal based on adaptation to systematicity (e.g., see Wade *et al.*, 2007).

In summary, the results of the current study demonstrate that accent-independent adaptation to foreign-accented speech is possible after exposure to multiple talkers with different foreign accents. We show that listeners are able to generalize to novel talkers from a novel language background as well as from a language background included in training. We suggest that this generalization is a result of exposure to systematic variation during training.

Acknowledgments

The work was supported by NIH-NIDCD (No. R01-DC005794 to A.R.B.; No. R01-DC004453 to B.A.W.) and a Northwestern University Cognitive Science Graduate Fellowship for Interdisciplinary Research Projects to M.M.B.-B. We thank Dan Sanes for his help in preparing the final figures.

References and links

- Baker, R. E. (2010). “The acquisition of English focus marking by non-native speakers,” Unpublished doctoral dissertation, Northwestern University, Evanston, IL.
- Baker, R. E., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., and Bradlow, A. R. (2011). “Word durations in non-native English,” *J. Phonetics* **39**(1), 1–17.
- Bamford, J., and Wilson, I. (1979). “Methodological considerations and practical aspects of the BKB sentence lists,” in *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children*, edited by J. Bench and J. Bamford (Academic Press, London).
- Bates, D. M., and Mächler, M. (2009). lme4: Linear mixed-effects models using Eigen and R syntax, R package version 0.999375-42.

- Bench, J., and Bamford, J., editors. (1979). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic Press, London).
- Bent, T., and Bradlow, A. R. (2003). "The interlanguage speech intelligibility benefit," *J. Acoust. Soc. Am.* **114**(3), 1600–1610.
- Best, C., McRoberts, G. W., and Goodell, E. (2001). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," *J. Acoust. Soc. Am.* **109**(2), 775–794.
- Bradlow, A., and Bent, T. (2008). "Perceptual adaptation to non-native speech," *Cognition* **106**(2), 707–729.
- Brouwer, S., Mitterer, H., and Huettig, F. (2012). "Speech reductions change the dynamics in competition during spoken word recognition," *Lang. Cognit. Processes* **27**(4), 539–571.
- Carlisle, R. (1991). "The influence of environment on vowel epenthesis in Spanish/English interphonology," *Appl. Linguist.* **12**(1), 76–95.
- Clarke, C. M., and Garrett, M. F. (2004). "Rapid adaptation to foreign accented English," *J. Acoust. Soc. Am.* **116**(6), 3647–3658.
- Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). "Patterns of English phoneme confusions by native and non-native listeners," *J. Acoust. Soc. Am.* **116**(6), 3668–3678.
- Flege, J. E., MacKay, I., and Meador, D. (1999). "Native Italian speakers' production and perception of English vowels," *J. Acoust. Soc. Am.* **106**(5), 2973–2987.
- Fox, R., Flege, J. E., and Munro, M. (1995). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," *J. Acoust. Soc. Am.* **97**(4), 2540–2551.
- Graddol, D. (1997). *The Future of English?* (The British Council, London).
- Graddol, D. (2006). *English Next* (The British Council, London).
- Guion, S., Flege, J. E., Liu, H., and Yeni-Komshian, G. (2000). "Age of learning effects on the duration of sentences produced in a second language," *Appl. Psycholinguist.* **21**(2), 205–228.
- Jenkins, J. (2000). *The Phonology of English as an International Language* (Oxford University Press, Oxford).
- McQueen, J., and Huettig, F. (2012). "Changing only the probability that spoken words will be distorted changes how they are recognized," *J. Acoust. Soc. Am.* **131**(1), 509–517.
- Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., and Yamada, T. (2011). "A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting," *J. Phonetics* **39**(2), 156–167.
- Sidas, S. K., Alexander, J. E. D., and Nygaard, L. C. (2009). "Perceptual learning of systematic variation in Spanish accented speech," *J. Acoust. Soc. Am.* **125**(5), 3306–3316.
- Wade, T., Jongman, A., and Sereno, J. (2007). "Effects of acoustic variability in the perceptual learning of non-native accented speech sounds," *Phonetica* **64**(2–3), 122–144.
- Weber, A., and Cutler, A. (2006). "First-language phonotactics in second-language listening," *J. Acoust. Soc. Am.* **119**(1), 597–607.