Relationship between listeners' nonnative speech recognition and categorization abilities

Eriko Atagi^{a)} and Tessa Bent

*Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue,
Bloomington, Indiana 47405
eatagi@brandeis.edu, tbent@indiana.edu*

Abstract: Enhancement of the perceptual encoding of talker characteristics (indexical information) in speech can facilitate listeners' recognition of linguistic content. The present study explored this indexical-linguistic relationship in nonnative speech processing by examining listeners' performance on two tasks: nonnative accent categorization and nonnative speech-in-noise recognition. Results indicated substantial variability across listeners in their performance on both the accent categorization and nonnative speech recognition tasks. Moreover, listeners' accent categorization performance correlated with their nonnative speech-in-noise recognition performance. These results suggest that having more robust indexical representations for nonnative accents may allow listeners to more accurately recognize the linguistic content of nonnative speech.

© 2014 Acoustical Society of America

[AC]

Date Received: August 7, 2014 **Date Accepted:** November 30, 2014

1. Introduction

Listeners extract multiple types of information from a speech signal. Some types of information guide the listener's recognition of the linguistic message, while others inform the listener about the identity and background of the talker, including their age, gender, and location of origin (Abercrombie, 1967). The information in speech that indexes such talker characteristics—known as indexical information—has been found to play an important role in listeners' spoken language processing. Specifically, recent perspectives on speech perception view indexical information as an important type of information in speech that is encoded and represented with linguistic information (for a review of the literature, see Pisoni and Levi, 2007). Supporting this perspective is a number of studies showing that encoding of indexical characteristics facilitates spoken word recall and recognition (e.g., Palmeri *et al.*, 1993). In addition, perceptual learning studies demonstrate that listeners who learn to identify novel talkers by voice show improvements in speech-in-noise intelligibility for those talkers (e.g., Nygaard *et al.*, 1994). Consequently, these studies propose that encoding talker-specific indexical characteristics benefits linguistic processing.

A beneficial indexical-linguistic relationship for listeners' speech recognition ability for degraded speech signals has also been observed, such as for cochlear implant (CI)-simulated speech (Loebach *et al.*, 2008; Krull *et al.*, 2012) and dysarthric speech (Borrie *et al.*, 2013). These studies illustrated that perceptual training through indexical tasks, such as talker identification (Loebach *et al.*, 2008; Krull *et al.*, 2012; Borrie *et al.*, 2013) and gender identification (Loebach *et al.*, 2008), can lead to improved performance on speech recognition tasks with degraded speech signals. As a result of perceptual training through these indexical tasks, listeners showed an intelligibility benefit for speech that is highly variable and acoustically inconsistent (i.e., dysarthric speech;

^{a)}Current address: Volen National Center for Complex Systems (MS 013), Brandeis University, 415 South Street, Waltham, MA 02454-9110.

Borrie *et al.*, 2013), as well as demonstrated generalization to new talkers under the same type of signal degradation (i.e., CI-simulated speech; Krull *et al.*, 2012).

Another type of speech that can cause decrements in intelligibility is nonnative speech. Although overall word recognition with nonnative speech tends to be reduced relative to native speech, it is characterized by substantial across-listener variability. For example, Derwing and Munro (1997) reported an across-listener range of 68%–92% in their nonnative speech transcription task. Very few studies have systematically investigated possible factors that account for such across-listener individual differences in accented speech recognition (cf. individual differences in older adults' adaptation to accented speech, see Janse and Adank, 2012). Our study addresses this gap by exploring listeners' indexical representation for nonnative speech as a factor that may be contributing to the considerable variability that exists across listeners in their accented speech recognition performance (Derwing and Munro, 1997).

The present study examines listeners' accent categorization abilities as one way to quantify the robustness of their indexical representation for nonnative speech. Nonnative speech contains a type of indexical information that cues listeners about the talker's native language background and location of origin. Prior studies have investigated how well naive listeners represent nonnative accent categories (Derwing and Munro, 1997; Vieru *et al.*, 2011). In a six-alternative forced-choice (6AFC) task, native French listeners categorized foreign-accented French at above chance levels (talker L1: Arabic, English, German, Italian, Portuguese, and Spanish; Vieru *et al.*, 2011). Similarly, American listeners' identification of four nonnative accents (talker L1: Cantonese, Japanese, Polish, and Spanish) was above chance in a four-alternative forced-choice accent categorization task (Derwing and Munro, 1997). Furthermore, listeners' performance on a nonnative accent categorization task was predicted by the listeners' familiarity with the native languages of the nonnative talkers (Derwing and Munro, 1997). This relationship between foreign language familiarity and accent categorization suggests that familiarity with the relevant native language may play a role in shaping a listener's perception of nonnative accents.

Previous studies on accent categorization have not addressed whether having robust indexical representations of nonnative accents is related to listeners' abilities to extract the linguistic meaning from nonnative speech. Nonnative speech can be conceptualized similarly to dysarthric speech, in that both are types of adverse listening condition caused by source degradation (Mattys *et al.*, 2012). Drawing on the intelligibility benefit gained through training on indexical tasks with degraded speech signals including dysarthric speech (Loebach *et al.*, 2008; Krull *et al.*, 2012; Borrie *et al.*, 2013), the present study explores the possibility of such an indexical-linguistic processing relationship with nonnative speech. Specifically, this study employed a 6AFC nonnative accent categorization task to measure listeners' indexical processing of nonnative speech and a nonnative speech-in-noise transcription task to measure their linguistic processing of nonnative speech. It was hypothesized that listeners' nonnative accent categorization performance would correlate with their accuracy on the nonnative speech-in-noise recognition task, similar to the speech-in-noise intelligibility benefit found when listeners are familiar with the talker's voice (Nygaard *et al.*, 1994). Finally, familiarity with a particular foreign language has been shown to afford listeners with greater sensitivity to and familiarity with features of foreign accented speech (Derwing and Munro, 1997). Therefore, in a secondary hypothesis, amount of experience with foreign language study was also expected to contribute to individual differences in listeners' nonnative speech recognition abilities.

2. Methods

Fifty adult monolingual native English speakers between the ages of 18 and 34 years (mean: 20.8 years) served as participants. All listeners were born and raised in the United States; 42 of the participants had not studied or lived abroad for a total of more than a month. Of the remaining eight, one had spent a total of 15 months

abroad across a span of 5 years; the rest had spent six months or less abroad. Though none of these listeners were conversationally proficient in any language other than English, 44 reported having studied one to three of the following spoken languages: Spanish ($N=35$), French ($N=8$), German ($N=6$), Japanese ($N=4$), Mandarin ($N=2$), Korean, Russian, Portuguese, Italian, and Modern Hebrew ($N=1$ for each). On average, listeners had 4.5 years of total foreign language study (range: 0–13 years). All listeners were screened for normal hearing thresholds, and reported no history of speech or hearing impairment.

Stimuli for this study were selected from the Hoosier Database of Native and Non-native Speech for Children (Hoosier Database; [Atagi and Bent, 2013](#); [Bent, 2014](#)). The stimuli were read by a total of 24 talkers. These talkers were 24 nonnative talkers from six native language backgrounds (two females and two males from each language background). The native language backgrounds of these talkers were Spanish, French, German, Japanese, Korean, and Mandarin. From the Hoosier Database, 120 sentences were selected from the Bamford–Kowal–Bench (BKB) Standard Sentence Test ([Bamford and Wilson, 1979](#)) for use in the categorization task. Another 120 sentences were selected from the Hearing in Noise Test—Children’s Version (HINT-C) ([Nilsson et al., 1996](#)) for use in the nonnative speech recognition task.

Before the start of the experiment, listeners completed the language background questionnaire and a hearing screening. Listeners were then seated in a sound-treated booth for the experimental tasks, where stimulus items were presented binaurally at approximately 68 dB sound pressure level. The order of presentation of the two tasks—nonnative speech recognition (transcription) and accent categorization—was counterbalanced across listeners. Order of presentation for the sentences was randomized in both the accent categorization and transcription tasks. Each sentence was played once and never repeated in any other trial across both tasks. Within each task, order of talker presentation was randomized, such that each sentence would be presented in the voice of any one of the 24 talkers. Each talker contributed five unique sentences to each task.

In the nonnative speech recognition task, listeners were presented with 120 trials. On each trial, listeners heard one HINT-C sentence in the voice of one of the 24 talkers, which were embedded within -2 dB signal-to-noise ratio speech-shaped noise with the noise extending 450 ms before and after the sentence. Immediately after each sentence finished playing, the listeners were prompted to type the sentence in standard English orthography.

In the 6AFC accent categorization task, listeners were also presented with 120 trials. Listeners were instructed to “select the native language of the speaker you hear,” and to press the corresponding key on the keyboard, which was the initial letter of each response option (e.g., S for Spanish, F for French). In each trial, listeners first saw a fixation cross in the middle of the computer for 500 ms, then heard one BKB sentence in the voice of one of the 24 talkers. The sentence was presented in quiet; immediately after presentation, six response choices appeared on the computer screen: Spanish, French, German, Japanese, Korean, and Mandarin Chinese.

3. Results

3.1 Variability across talkers and accents

In the accent categorization task, listeners as a group were above chance for each native language category (diagonal in Table 1; chance = 0.17, $p < 0.01$). There was significant variability in categorization performance across the native language categories, $F(5,45) = 11.23$, $p < 0.001$. As a group, the Mandarin talkers were least accurately categorized (0.22); the German talkers were most accurately categorized (0.37). The three Asian language categories were the least accurately categorized. The confusion patterns in Table 1 reveal that listeners frequently miscategorized the Japanese, Korean, and Mandarin talkers into one of the other two Asian language categories. In contrast,

Table 1. Confusion matrix of the proportion of responses by listeners, where each column represents listeners' response when presented with a talker from one of the accent categories indicated in the given row. The sum across each row is 1. The diagonal corresponds to proportion of correct responses.

Stimulus	Response					
	French	German	Japanese	Korean	Mandarin	Spanish
French	0.34	0.19	0.12	0.12	0.11	0.13
German	0.23	0.37	0.09	0.10	0.06	0.15
Japanese	0.11	0.13	0.23	0.22	0.23	0.07
Korean	0.10	0.11	0.22	0.24	0.24	0.11
Mandarin	0.11	0.13	0.18	0.24	0.22	0.12
Spanish	0.17	0.17	0.12	0.13	0.11	0.30

German, French, and Spanish talkers were rarely miscategorized into one of the Asian language categories. There was some confusion between the German and French accent categories.

Furthermore, at the level of individual talkers, some talkers were more accurately categorized than others [Fig. 1(A)]. The least accurately categorized talker was a Korean male talker (K2: 0.13); the most accurately categorized was a French male talker (F1: 0.61).

For the nonnative speech-in-noise recognition task, transcription accuracy was scored as the proportion of every word correctly transcribed. Words with added or deleted morphemes were counted as incorrect. There was a wide range of intelligibility scores across talkers, with the most intelligible talker (a Mandarin female talker, M4) at 0.81 and the least intelligible talker (a Spanish male talker, S1) at 0.37 [Fig. 1(A)]. Intelligibility levels also varied significantly by native language background, $F(5,45) = 68.25$, $p < 0.001$.

3.2 Relationship between listeners' accent categorization and nonnative speech recognition

Individual listeners displayed a wide range of accuracy in overall nonnative accent categorization [across all accents; mean = 0.28, range = 0.14–0.46; Fig. 1(B)]. However, all but two listeners performed above chance. Listeners were also variable in their

Fig. 1. Mean accuracies for the accent categorization and nonnative speech-in-noise transcription tasks, plotted for each talker (A on left) and for each listener (B on right). In (A), each point represents a talker and is identified with the initial letter of the talker's native language followed by a number (1–4). Talkers identified by number 1 or 2 are male talkers; talkers identified by number 3 or 4 are female talkers. (Examples: F1 is a French male; M4 is a Mandarin female.) In (B), each point is a listener.

transcription of nonnative speech in noise (mean = 0.60, range = 0.44–0.74). Performance on the two tasks revealed a significant, moderate correlation ($r = 0.34$, $p = 0.02$).

The role of foreign language study in nonnative speech recognition was examined by conducting an analysis of covariance with number of foreign languages studied as a categorical variable (experience with no, one, two, or three foreign languages), and total duration of foreign language study (summed across all languages studied by a listener) and accent categorization accuracy as continuous independent variables. Performance on the accent categorization task had an independent significant effect on performance on the speech recognition task, $F(1,38) = 8.95$, $p = 0.005$, $\eta_p^2 = 0.17$. However, neither the effects of number of languages studied, $F(3,38) = 2.08$, $p = 0.12$, nor total duration of foreign language study, $F(1,38) = 0.62$, $p = 0.44$, were significant. Further, there were no significant interactions, $p > 0.1$.

4. Discussion

The goal of this study was to examine the relationship between listeners' indexical and linguistic processing of nonnative speech. To this end, listeners' performance on a nonnative accent categorization task was hypothesized to correlate with their nonnative speech-in-noise recognition accuracy. The results revealed substantial variability across listeners in both their abilities to categorize nonnative talkers by language background and their performance in recognizing nonnative speech in noise. Moreover, the results of listeners' performance on these two tasks were moderately correlated with each other.

Overall, listeners categorized the nonnative talkers by accent at above chance accuracy. Rates of correct categorization, however, varied across accent categories, as well as across individual talkers within each accent category. Specifically, listeners were least accurate at categorizing the talkers from the three Asian language backgrounds. Higher rates of confusion among the Asian accent categories were observed, and Asian accents were unlikely to be confused for non-Asian accents. Previous work has also found high confusability among accented English with Asian native language backgrounds (Derwing and Munro, 1997; Atagi and Bent, 2013). Listeners' accent categories thus may not be specified at the level of particular languages; instead, listeners may have more broadly defined accent categories, such as Asiatic and non-Asiatic/European. A similar pattern of perceptual organization has been observed in a free classification study of different languages, in which one aspect of listeners' perceptual similarities among languages could be interpreted as mapping onto an "East to West" dimension (Bradlow *et al.*, 2010).

While past nonnative accent categorization studies only examined categorization performance by listeners as a group (Derwing and Munro, 1997; Vieru *et al.*, 2011), our study utilized the variability in categorization performance across individual listeners as one measure of the robustness of listeners' indexical representations for nonnative speech. Results revealed substantial variability across individual listeners' abilities to categorize accents, including two listeners (out of 50) who were below chance. This variability was found even though listeners' linguistic experiences were relatively homogenous (i.e., monolingual speakers of English with no extensive time spent outside of the U.S.). Moreover, this variability observed in accent categorization performance correlated with the variability observed across listeners' nonnative speech recognition abilities.

Indexical characteristics of speech are encoded automatically during linguistic processing (e.g., Pisoni and Levi, 2007) and become part of the representation for a given lexical item (Pierrehumbert, 2001). Knowledge of accent categories is one possible categorization scheme that could be associated with listeners' representations of lexical items. These categories could contain information about phonetic and phonological accent characteristics. A nonnative accent can lead to a mismatch between an incoming signal and representations of the lexical item stored in the listener's long-term memory, resulting in increased processing costs and reduced speech recognition

accuracy (Van Engen and Peelle, 2014). However, for listeners whose lexical representations are linked to accent categories, the tightly coupled processing of indexical and linguistic characteristics would guide the match between an incoming nonnative speech signal and an item in the listener's lexicon. This use of indexical information to guide linguistic processing would also reduce the amount of cognitive effort necessary for understanding the speech signal. More executive cognitive resources would thus be available to support the linguistic processing of nonnative speech for listeners with robust representations of nonnative accents than for listeners with less-developed representations of nonnative accents (Van Engen and Peelle, 2014).

In contrast to the previous studies on the relationship between indexical and linguistic processing, the listeners in this study did not receive any explicit training on accent categorization during the experiment. Instead, listeners' performance on the accent categorization task was taken to be a measure of their knowledge about nonnative accents, stored in their long-term memory. Our study thus cannot determine a causal relationship between the indexical and linguistic tasks. An alternative explanation for the correlation between the accent categorization and nonnative speech-in-noise recognition would be that the listeners who performed well on both of these tasks could also have had good general cognitive-perceptual abilities (e.g., attention, working memory capacity). Whereas indexical and linguistic information in familiar accents are automatically processed together and require fewer executive cognitive resources, unfamiliar accents require listeners to perform additional processing of the signal (e.g., Rönnerberg *et al.*, 2013). Listeners with better cognitive-perceptual abilities may be more adept at the explicit, controlled processing necessary to identify the indexical characteristics and understand the linguistic content of unfamiliar nonnative accents. The amount of across-listener variance in nonnative speech recognition that could be accounted for by differences in general cognitive abilities requires further study.

Listeners' performance on the accent categorization task accounted for some of the variance in their nonnative speech-in-noise recognition abilities. Measures of listeners' foreign language experience, in contrast, were not a significant factor for their performance on the transcription task. Foreign language study reported by listeners covered a wide range, not only in the duration and languages studied, but also in the age at which listeners had experienced them and the environments in which they had those experiences. The lack of significant contribution of foreign language study to listeners' nonnative speech recognition abilities may indicate that the measures of foreign language experience used here were not sensitive measures of prior experience with linguistic variability.

In summary, the present study demonstrated that there is considerable individual variation across listeners in their abilities to categorize nonnative speech by native language and to recognize nonnative speech-in-noise. Moreover, these two abilities were correlated with each other, demonstrating a link between listeners' indexical and linguistic processing mechanisms for nonnative speech. Findings from this study encourage further study of the mechanisms behind the relationship between listeners' indexical and linguistic processing skills for nonnative speech, and of possible factors that contribute to listeners' development of these skills.

Acknowledgments

This work was supported by NIH Grants No. R21 DC010027 (T.B.) and No. T32 DC00012 (E.A.) from National Institute on Deafness and Other Communication Disorders, and Grant No. T32 NS007292 from National Institute of Neurological Disorders and Stroke (E.A.). An earlier version of this work was presented at the Annual Meeting of the Psychonomic Society in the Fall of 2013.

References and links

- Abercrombie, D. (1967). *Elements of General Phonetics* (Aldine Publishing Company, Chicago).
Atagi, E., and Bent, T. (2013). "Auditory free classification of nonnative speech," *J. Phon.* **41**, 509–519.

- 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, London), pp. 148–187.
- Bent, T. (2014). "Children's perception of foreign-accented words," *J. Child Lang.* **41**, 1334–1355.
- Borrie, S. A., McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., and Anderson, T. J. (2013). "The role of linguistic and indexical information in improved recognition of dysarthric speech," *J. Acoust. Soc. Am.* **133**, 474–482.
- Bradlow, A. R., Clopper, C. G., Smiljanic, R., and Walter, M. A. (2010). "A perceptual phonetic similarity space for languages: Evidence from five native language listener groups," *Speech Commun.* **52**, 930–942.
- Derwing, T. M., and Munro, M. J. (1997). "Accent, intelligibility, and comprehensibility: Evidence from four L1s," *Stud. Sec. Lang. Acquis.* **19**, 1–16.
- Janse, E., and Adank, P. (2012). "Predicting foreign-accent adaptation in older adults," *Q. J. Exp. Psychol.* **65**, 1563–1585.
- Krull, V., Luo, X., and Kirk, K. I. (2012). "Talker-identification training using simulations of binaurally combined electric and acoustic hearing: Generalization to speech and emotion recognition," *J. Acoust. Soc. Am.* **131**, 3069–3078.
- Loebach, J. L., Bent, T., and Pisoni, D. B. (2008). "Multiple routes to the perceptual learning of speech," *J. Acoust. Soc. Am.* **124**, 552–561.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). "Speech recognition in adverse conditions: A review," *Lang. Cogn. Process.* **27**, 953–978.
- Nilsson, M., Soli, S. D., and Gelnett, D. J. (1996). *Development of the Hearing in Noise Test for Children (HINT-C)* (House Ear Institute, Los Angeles).
- Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (1994). "Speech perception as a talker-contingent process," *Psychol. Sci.* **5**, 42–46.
- Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (1993). "Episodic encoding of voice attributes and recognition memory for spoken words," *J. Exp. Psychol. Learn. Mem. Cogn.* **19**, 309–328.
- Pierrehumbert, J. (2001). "Exemplar dynamics: Word frequency, lenition, and contrast," in *Frequency Effects and the Emergence of Lexical Structure*, edited by J. Bybee and P. Hopper (John Benjamins, Amsterdam), pp. 137–157.
- Pisoni, D. B., and Levi, S. V. (2007). "Representations and representational specificity in speech perception and spoken word recognition," in *The Oxford Handbook of Psycholinguistics*, edited by G. Gaskell (Oxford University Press, Oxford), pp. 3–18.
- Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., and Pichora-Fuller, M. K. (2013). "The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances," *Front. Syst. Neurosci.* **7**, 31.
- Van Engen, K. J., and Peelle, J. E. (2014). "Listening effort and accented speech," *Front. Hum. Neurosci.* **8**, 577.
- Vieru, B., Boula de Mareuil, P., and Adda-Decker, M. (2011). "Characterisation and identification of non-native French accents," *Speech Commun.* **53**, 292–310.