ICA 2013 Montreal  
Montreal, Canada  
2 - 7 June 2013

Speech Communication  
Session 4pSCb: Production and Perception I: Beyond the Speech Segment (Poster Session)

4pSCb7. Auditory free classification of nonnative speech by nonnative listeners  
Eriko Atagi* and Tessa Bent

*Corresponding author’s address: Department of Speech and Hearing Sciences, Indiana University, 200 S Jordan Ave, Bloomington, Indiana 47405, eatagi@indiana.edu

Nonnative listeners are less accurate than native listeners at classifying talkers by regional dialect (Clopper & Bradlow, 2009). This decrement may be due to less robust knowledge about the underlying sound structure of the target language or less extensive experience with socio-cultural phonetic variation in the target language. To disentangle the contribution of these two factors, this study examined native and nonnative listeners’ abilities to classify talkers who varied on another sociophonetic dimension: foreign accent. Unlike regional dialect variation, nonnative listeners typically have more experience with nonnative speech than native listeners, particularly for talkers with the same native language background. Using auditory free classification, native listeners of English and native Korean listeners classified talkers by perceived native language. Talkers consisted of nonnative talkers from six native language backgrounds and native talkers. Results demonstrated that native listeners were nearly perfect at grouping the native talkers together, but Korean listeners were much less accurate. Further, Korean listeners did not show an advantage for grouping Korean-accented talkers together. These results suggest that nonnative listeners' less robust linguistic representations of the target language can hinder their abilities to attend to the acoustic-phonetic features that index dialect and accent categories. [Supported by NIH-NIDCD Grant R21DC010027.]

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

There are two properties of speech, linguistic and indexical, that listeners concurrently process during speech perception. These two properties continuously interact to facilitate processes such as lexical retrieval (Nygaard and Pisoni, 1998). Thus, in addition to building linguistic categories, the development of indexical categories, such as gender and dialect, is necessary to ensure fast and accurate spoken language processing. For late learners of a second language, acquiring native-like linguistic (e.g., phonological) categories, such as phonemes, is a complex and often difficult process (e.g., Davidson, 2011; Strange, 1995; Strange and Shafer, 2008). Compared to the amount of research that has been conducted on the acquisition of second language linguistic properties, relatively few studies have investigated second language learners’ abilities to process and encode indexical speech variables in the second language.

The indexical feature that has most frequently been investigated for second language learners is regional dialect perception. In one such study, Eisenstein (1982) investigated beginning, intermediate, and advanced English language learners’ discrimination of five English dialects: Standard American English, New York English, African American English, Hawaiian Pidgin English, and Irish English. Beginning and intermediate learners were significantly worse at dialect discrimination than advanced learners; however, advanced learners performed as well as native listeners. In contrast, Scales et al. (2006) found that, compared to native listeners, intermediate to advanced English learners from a variety of native language backgrounds performed more poorly at an open-set dialect identification task of two native (British and American) and two nonnative (Chinese and Spanish) accents. Moreover, many of these English learners did not correctly identify the American accent, and misidentified the other accents as American, even though the American accent was the learners’ “target” accent and matched the local accent. In another study that used the auditory free classification task, Clopper and Bradlow (2009) examined nonnative listeners’ classification of four American regional dialects (New England, North, Midland, and South) and found that the nonnative listeners were less accurate than the native listeners.

These generally observed decrements in nonnative listeners’ abilities to accurately perceive regional dialect differences could be due to two factors regarding nonnative listeners’ perceptual representations and experiences in their second language. First, nonnative listeners have less robust knowledge about the underlying sound structure of the second language. Therefore, nonnative listeners may not have perceptual categories and boundaries that allow them to correctly determine whether specific phonetic variations indicate linguistically meaningful distinctions or index the talker’s location of origin. Second, nonnative listeners generally have less experience with socio-cultural phonetic variation in the second language. Nonnative listeners may not have enough exposure to dialectal variation to have created cognitive representations of multiple dialect groups.

To disentangle the contribution of these two factors, the current study compared native and nonnative listeners’ abilities to classify talkers who varied on a sociophonetic dimension with which nonnative listeners would be more familiar: foreign accent. Unlike regional dialect variation, nonnative listeners typically have more experience with nonnative speech than native listeners, particularly of talkers with whom they share a native language background.

METHODS

Listeners

Twenty-eight monolingual native English speakers (native listeners) and 25 native speakers of Korean (Korean listeners) with a second language of English were recruited from the Indiana University campus. Data from the native listeners were collected as part of an earlier study (Atagi and Bent, 2011). The mean age of the native listeners was 22.6 years with a range of 19 – 38 years. All native listeners were born and raised in the United States, and had not had extensive experience with nonnative speakers of English. The mean age of the Korean listeners was 25.0 years with a range of 20 – 33 years. Korean listeners had on average resided in English-speaking countries for 1.9 years (range: <1 – 7 years), and none of the Korean listeners reported proficiency in a language other than Korean and English. All listeners had normal hearing and reported no history of speech or hearing impairment.

Stimuli

The stimuli for the current study were selected from the Hoosier Database of Native and Non-native Speech for Children (Bent, 2010). Two sentences from the Hearing in Noise Test – Children’s Version (HINT-C) (Nilsson et
Talkers consisted of 24 nonnative talkers from six native language backgrounds and four native talkers for a total of 28 talkers, with two females and two males from each language background. The native language backgrounds of these talkers were American English (midland dialect), French (from France), German (from Germany), Spanish (from Colombia), Japanese, Korean (Seoul dialect), and Mandarin (Beijing dialect). At the time of recording, all of the nonnative talkers had spent four years or less in the U.S.A. or any other English-speaking country, and were only proficient in their native language and English (as a second language).

**Procedure**

Listeners completed a total of eight (native listeners) or six (Korean listeners) auditory free classification tasks. The current paper reports on the results of the final two free classification tasks. For these free classification tasks, listeners were presented with a 12x12 grid on the computer screen. To the left of this grid were 28 “icons” which, when clicked, played a speech sample by one of the 28 talkers producing the stimulus sentence. Within each free classification task, all talkers produced the same sentence. Listeners were instructed to group talkers by perceived native language by dragging each icon onto the grid and making groups of icons on the grid. Listeners took as long as they liked to complete each classification task and were able to form as many groups with as many speakers in each group as they wished. They also listened to the speech samples as many times as they needed. Before the start of the experiment, listeners completed a language background questionnaire and a hearing screening. In addition, the language proficiency of the Korean listeners was assessed with a commercial English proficiency test (Pearson’s VERSANT test), after the completion of the experiment.

**RESULTS**

Across the two free classification tasks, the two listener groups did not significantly differ in the number of talker groups created. Native listeners created an average of 7.43 groups and Korean listeners created an average of 8.26 groups ($t(48.1) = -1.17, p = 0.25$). Table 1 summarizes classification accuracy by talker native language background for the native and Korean listeners. Classification accuracy was calculated by first summing the number of times listeners grouped together talkers of a given language group, then dividing by the total number of possible correct classifications for one language group. A score of 1.00 would indicate that all talkers from the same language group were grouped together by all listeners. Native listeners were nearly perfect at grouping the native American English talkers together, while Korean listeners were significantly less accurate ($t(26.9) = 3.32, p < 0.01$). Furthermore, Korean listeners were not more accurate than native listeners at grouping Korean talkers together ($t(50.7) = 0.56, p = 0.58$). Korean listeners were, however, more accurate at grouping Japanese talkers together ($t(42.2) = 2.86, p < 0.01$). Classification accuracy for the other language backgrounds did not differ between the two listener groups.

<table>
<thead>
<tr>
<th>Talker Language Group</th>
<th>Native Listeners</th>
<th>Korean Listeners</th>
</tr>
</thead>
<tbody>
<tr>
<td>English **</td>
<td>.99</td>
<td>.82</td>
</tr>
<tr>
<td>French</td>
<td>.13</td>
<td>.14</td>
</tr>
<tr>
<td>German</td>
<td>.39</td>
<td>.39</td>
</tr>
<tr>
<td>Japanese **</td>
<td>.35</td>
<td>.20</td>
</tr>
<tr>
<td>Korean</td>
<td>.38</td>
<td>.35</td>
</tr>
<tr>
<td>Mandarin</td>
<td>.30</td>
<td>.32</td>
</tr>
<tr>
<td>Spanish</td>
<td>.20</td>
<td>.16</td>
</tr>
</tbody>
</table>

In a classification paradigm, however, to achieve highly accurate performance high rates of accurate categorization and low rates of misclassification (similar to high “hit” and low “false alarm” rates in signal detection theory) are both required. Table 2 summarizes the proportion of misclassification by talker native language background for the native and Korean listeners. Misclassification was determined by calculating the number of times listeners classified talkers of a given language background with talkers from different native language, and dividing by the total number of possible misclassifications for one language group. A score of 0 for a given language background.
group would indicate that no talker from a different language group was grouped together with that language group. Korean listeners were less likely than native listeners to group Korean talkers with talkers from other native language backgrounds ($t(43.3) = 3.34, p < 0.01$). This lower rate of misclassification by Korean listeners was also true for the other two Asian talker language backgrounds: Japanese ($t(44.6) = 3.23, p < 0.01$) and Mandarin ($t(46.3) = 2.76, p < 0.01$). Additionally, Korean listeners were significantly more likely to classify nonnative talkers with native American English talkers ($t(49.8) = -2.04, p < 0.05$). Thus, rates of misclassification for the native English talkers were higher for the Korean listeners than the native listeners.

**TABLE 2.** Average proportions of misclassifications for each talker native language group by native and Korean listeners. An asterisk (*) next to the talker language group indicates a statistically significant difference between the rate of misclassification by native listeners and Korean listeners (*$p < 0.05$; **$p < 0.01$).

<table>
<thead>
<tr>
<th>Talker Language Group</th>
<th>Native Listeners</th>
<th>Korean Listeners</th>
</tr>
</thead>
<tbody>
<tr>
<td>English **</td>
<td>.10</td>
<td>.14</td>
</tr>
<tr>
<td>French</td>
<td>.15</td>
<td>.15</td>
</tr>
<tr>
<td>German</td>
<td>.13</td>
<td>.13</td>
</tr>
<tr>
<td>Japanese **</td>
<td>.17</td>
<td>.10</td>
</tr>
<tr>
<td>Korean **</td>
<td>.17</td>
<td>.10</td>
</tr>
<tr>
<td>Mandarin **</td>
<td>.17</td>
<td>.12</td>
</tr>
<tr>
<td>Spanish</td>
<td>.11</td>
<td>.09</td>
</tr>
</tbody>
</table>

**DISCUSSION & CONCLUSION**

In the current study, nonnative listeners were significantly less accurate than native listeners at grouping native talkers. This result adds to the finding of decreased sensitivity to regional dialect differences for second language listeners (Clopper and Bradlow, 2009). Such perceptual decrements of classifying native talkers could be due to either a less robust knowledge of the sound structure of the second language, or less exposure to and experience with native talkers. Korean listeners, however, are likely to have had more experience with Korean-accented English than native listeners. Nevertheless, Korean listeners were not more likely to accurately categorize Korean talkers with each other. This result indicates that the difference in the ability to group listeners by regional dialect or foreign accent cannot be explained by the differences in amount of exposure to the varieties alone. The patterns of categorization accuracy suggest that nonnative listeners’ less robust representations of the second language sound structure may be the underlying cause for their difficulties with classifying varieties of the target language.

Comparison of the native and Korean listeners’ misclassification rates and patterns, however, indicated that the Korean listeners’ knowledge about the Asian languages included in the current study may have provided some advantage in distinguishing the Asian language backgrounds. Compared to native listeners, Korean listeners were less likely to group Korean talkers with talkers from other language backgrounds. Moreover, this lower rate of misclassification by the Korean listeners was also true of the Japanese and Mandarin talkers. Many of the Korean listeners, however, reported having studied some Japanese ($n = 7$) or Chinese ($n = 11$), which may have contributed to their ability to distinguish talkers from the Asian language backgrounds. In contrast, only two of the native listeners had studied an Asian language (i.e., Japanese). Knowledge of the sound structure of the three Asian languages did not necessarily help Korean listeners to more accurately group together the native talkers of the Asian languages. Yet, this knowledge may have led Korean listeners to be less likely to confuse the Asian language backgrounds for each other or for any other language background.

In summary, in addition to nonnative listeners’ reduced sensitivity to the acoustic-phonetic features that index regional dialect differences in their second language, nonnative listeners are less able to distinguish native from nonnative speakers. Furthermore, the nonnative speakers’ experience with a specific accent—Korean-accented English—did not provide a classification accuracy benefit compared to the native listeners. This result suggests that the nonnative listeners less robust linguistic representations of the target language, rather than inadequate exposure, may be responsible for their decreased sensitivity to certain sociophonetic features and may restrict their ability to take advantage of their extensive experience with a specific language variety (i.e., Korean-accented English in the current experiment). However, having knowledge of the first languages of nonnative talkers helped nonnative listeners to perceptually distinguish among accent categories by lowering rates of misclassification. Highly accurate accent categorization, therefore, may require listeners to have deep knowledge of both the first and second languages of nonnative talkers. Assessment of early bilingual listeners who are highly proficient in two languages would allow for the testing of this hypothesis.
ACKNOWLEDGMENTS

This research was supported by NIH-NIDCD Grant R21DC010027 to Indiana University. We would like to thank Charles Brandt for his programming assistance.

REFERENCES