

Published in final edited form as:

J Phon. 2013 November 1; 41(6): . doi:10.1016/j.wocn.2013.09.003.

Auditory free classification of nonnative speech

Eriko Atagi^{a,*} and Tessa Bent^a

^aDepartment of Speech and Hearing Sciences, Indiana University, 200 S Jordan Ave.,
Bloomington, IN 47405, U.S.A

Abstract

Through experience with speech variability, listeners build categories of indexical speech characteristics including categories for talker, gender, and dialect. The auditory free classification task—a task in which listeners freely group talkers based on audio samples—has been a useful tool for examining listeners' representations of some of these characteristics including regional dialects and different languages. The free classification task was employed in the current study to examine the perceptual representation of nonnative speech. The category structure and salient perceptual dimensions of nonnative speech were investigated from two perspectives: general similarity and perceived native language background. Talker intelligibility and whether native talkers were included were manipulated to test stimulus set effects. Results showed that degree of accent was a highly salient feature of nonnative speech for classification based on general similarity and on perceived native language background. This salience, however, was attenuated when listeners were listening to highly intelligible stimuli and attending to the talkers' native language backgrounds. These results suggest that the context in which nonnative speech stimuli are presented—such as the listeners' attention to the talkers' native language and the variability of stimulus intelligibility—can influence listeners' perceptual organization of nonnative speech.

Keywords

foreign-accented speech; classification; perceptual similarity; native language background

1.0 Introduction

The speech signal simultaneously presents listeners with two streams of information: linguistic and indexical. Traditionally, indexical information in the speech signal was thought of as noise that must undergo normalization during speech perception (e.g., Joos, 1948). More recently, indexical information has been reconceived as information that is stored by the listener and is integrated with linguistic information during speech perception (Luce & Lyons, 1998; Nygaard & Pisoni, 1998; Palmeri, Goldinger, & Pisoni, 1993). Under this view, in addition to building representations for linguistic categories such as phoneme or semantic classes, listeners build categories of indexical variables such as talker, gender, or dialect (Clopper & Bradlow, 2009; Clopper, Conrey, & Pisoni, 2005; Clopper & Pisoni, 2004a, 2004b; Nygaard & Pisoni, 1998). A listener's experience building these indexical

© 2013 Elsevier Ltd. All rights reserved.

*Corresponding author: eatagi@indiana.edu, Tel: +1 512 587 4748.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

categories can further facilitate retrieval of linguistic information (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994).

One method that has recently been applied to investigating the perceptual representation of indexical characteristics is the auditory free classification task (Bradlow, Clopper, Smiljanic, & Walter, 2010; Clopper & Bradlow, 2008, 2009; Clopper & Pisoni, 2007). The auditory free classification task is a method that allows listeners to categorize speech samples without experimenter-constrained categories. In addition to the analysis of the resulting category structure, the data can be interpreted as indicating perceptual distances among the stimuli. From these distances, salient perceptual dimensions for the listener population can be derived using analytic tools such as multidimensional scaling (MDS). For example, Clopper and Pisoni (2007) used this method with regional American English dialects and showed that listeners make even more fine-grained distinctions between dialects than can be predicted by forced-choice tasks. Furthermore, the perceptual dimensions most salient to listeners were gender, geography (i.e., north-south), and the presence of phonetic characteristics that are less common across dialects (e.g., r-lessness is a relatively infrequent characteristic across American English dialects; therefore, speakers from r-less dialect regions are perceptually distinct from the majority of American English speakers). The auditory free classification method has also been applied to examining the perceptual similarity of different languages (Bradlow et al., 2010). For different languages, three salient dimensions were observed: the presence of dorsal consonants that are not /k, ŋ/; the presence of rounded front vowels; and the geographical location of the language (i.e., east-west). The particular dorsal consonants and rounded front vowels are sounds that are cross-linguistically infrequent and thus make languages with these phonemes perceptually distinct compared to languages without these phonemes. These results demonstrate that the auditory free classification task is a useful method for examining how listeners represent speech variability. Further, these studies suggest that the presence of uncommon phonetic or phonemic characteristics, as well as geography, may be key features in the categorization and representation of distinct forms of speech and language variability. The current study also used the free classification task to investigate the representation of another source of variability in the speech signal: nonnative accents.

Nonnative accents are a type of across-talker variability that index speaker origin, similar to regional dialect variation or different languages. Similar to the systematic acoustic-phonetic variability seen in regional accents, nonnative talkers who share a native language will tend to deviate from target language norms in similar ways (e.g., Flege, 1995; Goslin, Duffy, & Floccia, 2012; McLennan & Gonzalez, 2012; Wells, 1982). To investigate how native listeners perceptually represent the variability present in nonnative speech, Atagi and Bent (2011) used the auditory free classification task to investigate the perceptual similarity space for nonnative accents. MDS analysis revealed that the first perceptual dimension was gender, which is well documented as a consistently salient feature in speech perception across various task types such as categorization (Clopper et al., 2005), similarity judgments (Clopper, Levi, & Pisoni, 2006), free classification (Clopper & Bradlow, 2008; Clopper & Pisoni, 2007), and speeded classification (Mullennix & Pisoni, 1990). The coordinates of the second dimension in the MDS output were interpreted to be degree of foreign accent. Listeners, therefore, may process nonnative speech samples similarly to how listeners in Clopper and Pisoni (2007) perceived regional dialects—by noting the extent of phonological deviation from their representation of native language norms. Taken together, the results from nonnative speech (Atagi & Bent, 2011), regional accents (Clopper & Pisoni, 2007), and different languages (Bradlow et al., 2010) suggest that the perceived magnitude of phonological deviation from a listener's speech norms may be a salient perceptual dimension for multiple forms of speech variability.

The similarities in the results across these three previous studies were found despite some methodological differences. Specifically, listeners in Atagi and Bent (2011) were instructed to group talkers based on general similarity. Clopper and Pisoni (2007), in contrast, instructed listeners to group the talkers by perceived region of origin. Their finding that gender, geography, and uncommon phonetic characteristics were salient features of regional dialect variation was extracted in the context of a classification task based on categorization. That is, listeners were attempting to classify talkers by preexisting notions of regional dialect categories—though the exact category labels or number of categories may have differed across listeners. In their classification task of different languages, Bradlow et al. (2010) instructed their listeners to “group the languages by their sound similarity” (p. 933), which more directly invokes classification by general perceptual similarity. Furthermore, in Bradlow et al. (2010), each language was represented by a single talker; thus unlike the task in Clopper and Pisoni (2007), there were no presumed categories into which the stimuli would be classified. The task procedure in Atagi and Bent (2011) was similar to that of Bradlow et al. (2010), but the stimulus items were similar to that of Clopper and Pisoni (2007). Specifically, listeners in Atagi and Bent (2011) were instructed to group by perceptual similarity. They could have selected to classify talkers by shared origin (i.e., native language background) as their grouping strategy. Talker native language background, however, was not found to be one of the primary, salient dimensions. Nevertheless, it is unclear whether a change in instructions would significantly affect listeners’ classification behavior. Thus, in addition to the classification of nonnative talkers by general similarity, the current study further investigated listeners’ grouping of talkers by native language background during classification tasks in which the listener’s attention is focused explicitly on this categorical feature.

In addition, the current study investigated whether the perceptual salience of speech properties are fixed. Every utterance consists of indexical properties that are invariant within a talker—for example, the talker’s gender and place of origin—as well as those that could vary from one utterance to the next, such as the level of intelligibility of a given sentence. For nonnative talkers especially, such within-talker, across-item variability could be significant, and could depend on the relative difficulty posed by the phonological and phonetic characteristics of a particular sentence (Best, McRoberts, & Goodell, 2001; Flege, 1995). Therefore, some nonnative talkers may vary significantly in intelligibility from one sentence to another. In the current study, the effects of varying the levels of intelligibility were examined by comparing stimulus sets where the selected sentence was highly intelligible for all talkers and where the sentence was low in intelligibility for some of the talkers. It was hypothesized that the inclusion of lower intelligibility sentences would present listeners with greater variability within the stimulus set, and introduce more deviations from native language norms than high intelligibility sentences. Listeners would have more accent-specific acoustic-phonetic cues available to guide classification by native language background in low intelligibility conditions.

Perception of a particular stimulus also appears to be highly dependent on the context in which the stimulus is presented (Tversky, 1977). For example, with regional dialects, source and environmental characteristics of the stimulus set have been shown to affect listeners’ classification and categorization behavior (Clopper & Bradlow, 2008; Clopper et al., 2005). Dialect classification under moderate amounts of noise is poor even when intelligibility at the same level of noise is well above floor (Clopper & Bradlow, 2008). Further, conditions that include both male and female talkers compared to conditions with a single gender show different perceptual similarity structures of regional accent groups, although classification accuracy is not affected (Clopper et al., 2005). In addition, foreign accent ratings have been shown to fluctuate depending on the proportion of native talkers included in the stimulus set, such that the same nonnative talkers are perceived to have a weaker accent when no native

speakers are included in the set of stimulus sentences (Flege & Fletcher, 1992). The current study, therefore, included a second manipulation that expands on these earlier studies of stimulus set factors.

The second manipulation compared stimulus sets that consist of nonnative talkers only or consist of both nonnative and native talkers. The inclusion of native talkers in the stimulus set could have one of three possible effects. First, the perceptual space of accents could be stretched to accommodate talkers who have no foreign accent. In this case, the availability of native speech exemplars in the stimulus set would not affect the perceptual similarity of the nonnative speech exemplars—listeners would simply expand the perceptual space to include the native talkers. Such an outcome would suggest that listeners are able to clearly distinguish between nonnative and native talkers, and the availability of native talkers does not affect their perception of nonnative accents. Conversely, the availability of native speech could result in a complete reorganization of the perceptual space to accommodate native talkers. A complete reorganization would indicate that listeners employed two different classification strategies depending on whether native talkers were included. Finally, some nonnative accents could be perceived as being more similar to native accents than other nonnative accents, resulting in a partial rearrangement of the perceptual space. A partial rearrangement would occur if the organizational schema for accents were dependent on accent strength. Specifically, the inclusion of native talkers into the perceptual space would only implicate the nonnative accents that are perceptually similar to the native accent. Furthermore, this third scenario would suggest that not every nonnative talker could be reliably distinguished from native talkers.

The current study aimed to: (1) replicate the results from Atagi and Bent (2011) that found gender and degree of accent as the two most salient dimensions of nonnative speech; (2) examine the categorization of nonnative accents in conditions in which listeners group talkers by perceived native language background; and (3) investigate how manipulating properties of the stimulus set, including levels of intelligibility and whether native talkers are included, influences categorization behavior.

2.0 Hoosier Database of Native and Non-native Speech for Children

All stimuli used in the current study were sentences selected from the Hoosier Database of Native and Non-native Speech for Children (Bent, 2010). This database contains digital audio recording of native and nonnative speakers of English reading words, sentences, and paragraphs. All materials in this database are in English and are lexically, syntactically, and semantically appropriate for use with children. Speech samples from 28 talkers comprised of two adult females and two adult males from each of the following native language backgrounds: 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). In addition, all talkers had pure-tone hearing thresholds of ≤ 20 dB at 500, 1000, 2000, and 4000 Hz and reported no history of speech or hearing impairment. The talkers were recorded in a sound-attenuated booth using a Marantz PDM670 digital recorder and a Shure Dynamic WH20XLR headset microphone.

Several baseline measurements have been collected for this database (Appendix A). Following the terminology established by Derwing and Munro (Derwing & Munro, 1997; Munro & Derwing, 1995a, 1995b), intelligibility is an objective measure of communicability, calculated as percent accurate transcription by native listeners. Objective intelligibility scores for each of the 28 talkers at both the word and sentence (keyword

correct) levels were collected previously for 550 words and 250 sentences per talker (Bent, 2010). In contrast, comprehensibility is defined as a subjective measure of perceived ease of understanding speech. Perceived degree of foreign accent is also considered a subjective measure. Ratings for these two subjective measures were collected previously (Atagi & Bent, 2011). Nine-point scales were used. For comprehensibility, the scale ranged from 1 = very easy to understand to 9 = very difficult to understand; and for accent, the scale ranged from 1 = no foreign accent to 9 = very strong foreign accent (Derwing & Munro, 1997; Munro & Derwing, 1995a, 1995b).

3.0 Experiment 1: Classification by general similarity

The first goal was to replicate the results from Atagi and Bent (2011) in two conditions: one condition with the same set of stimuli as (Atagi & Bent, 2011), in which the selected stimuli sentences were different across all talkers (Multi-sentence condition); and another condition with a new set of stimuli, in which the same stimulus sentence was selected for all talkers (Same-sentence condition). The Multi-sentence condition was an exact replication and was conducted in part to test whether listeners in the earlier study were biased toward attending to degree of accent due to their completion of accent and comprehensibility ratings prior to completing the free classification task.

The replication in the Same-sentence condition in the current experiment was conducted with a new set of stimuli with the second goal in mind: to investigate whether listening to the same sentence from all talkers would differentially affect the salience of degree of accent or comprehensibility. The coordinates of the second dimension in the MDS output from Atagi and Bent (2011) were interpreted as talkers' degrees of accent; however, accent and comprehensibility ratings were highly collinear and both measures were also correlated with the second dimension of the MDS output. The sentences used in Atagi and Bent (2011) and the Multi-sentence condition here ranged widely in degrees of intelligibility and, as no two talkers repeated the same sentence, there were no additional cues to increase comprehensibility (i.e., ease of understanding). By having all talkers produce the same, highly intelligible sentence, listeners could easily infer for all talkers what the intended sentence was, alleviating difficulties with comprehensibility. The manipulation of having all talkers produce the same sentence was therefore included in an attempt to differentiate degree of accent from comprehensibility. If the second dimension were more strongly correlated to degree of accent in the Same-sentence condition than in the Multi-sentence condition, there would be greater support for the interpretation of the second dimension as degree of accent rather than comprehensibility.

3.1 Listeners

Twenty-eight adult monolingual native English speakers, 21 females and 7 males, were recruited from the Indiana University campus. Their mean age was 22.6 years with a range of 19 – 38 years. All listeners were born in the United States; most of the listeners had not studied or lived abroad. Most of the listeners had studied at least one other language, but all reported not being able to conversationally speak or comprehend a language other than English. All listeners had normal hearing thresholds of 25 dB at 250 Hz, 20 dB at 500, 1000, 2000, 4000, and 8000 Hz, and reported no history of speech or hearing impairment. They also had limited to no instruction in linguistics, speech and hearing sciences, or any other related fields. Listeners were paid for their participation.

3.2 Methods

In Experiment 1, listeners completed two free classification tasks in which they grouped nonnative talkers based on general perceptual similarity (Multi-sentence condition and

Same-sentence condition). Listeners were presented with a 12×12 grid on the computer screen. To the left of this grid were 24 “icons” of the same size as one cell on the grid, labeled with arbitrary, two-letter sequences randomly generated for each listener. Each icon was linked to a speech sample, which played while the listener clicked on it. The listeners were instructed to drag each icon onto the grid and to group the icons, so that speakers who sounded similar should be grouped together. Listeners were further instructed to pay no attention to the meaning of the sentences in making these groups. Listeners could take as long as they liked to complete this task and could form as many groups and as many speakers in each group as they wished. They were able to listen to the speech samples as many times as they needed. (For further description of the auditory free classification task and related analyses, see Clopper, 2008.)

Before the start of the experiment, listeners completed a language background questionnaire and a hearing screening. Instructions for all tasks were given verbally by the experimenter and on a computer screen. Custom software designed in Python (<http://www.python.org>) and presented through a Mac mini was used for stimulus presentation and data collection. Listeners were seated in front of a ViewSonic 20" VX2033wm widescreen LCD monitor in a sound treated booth. Stimulus items were presented binaurally in quiet at approximately 68 dB through Sennheiser HD 280 pro headphones.

3.4 Stimuli

For the Multi-sentence condition, 24 different sentences, one for each of the 24 nonnative talkers, were selected as stimuli from the Hoosier Database (Appendix B). All sentences were from the Hearing in Noise Test – Children’s Version (HINT-C) (Nilsson, Soli, & Gelnett, 1996). In the Same-sentence condition, all 24 nonnative talkers produced the same sentence: “Strawberry jam is sweet.” Intelligibility across the 24 talkers for this sentence was high, ranging from 92.5% to 100% (mean: 99.5%) in keywords correct (Bent, 2010).

3.5 Results and discussion

Listeners created 7.1 and 6.8 groups on average in the Multi-sentence and Same-sentence conditions, respectively. The number of groups created was not significantly different across the two conditions ($t = 0.63$; ns; $df = 27$). To analyze the salient perceptual dimensions, a 24×24 symmetric similarity matrix was computed. The number of times a pair of talkers was grouped together across all 28 listeners was summed so that two talkers who were never grouped together would receive a score of 0 and those who were grouped together by all listeners would receive a score of 28. This matrix was then submitted to a multidimensional scaling analysis in SPSS 19.0 (Euclidean distance algorithm with ordinal similarity data).

The two-dimensional MDS solutions gave the best fit in both conditions, indicated by an “elbow” in the stress plots. The two-dimensional solutions for both of these conditions (Figure 1) provided two interpretable dimensions: a binary differentiation of talker gender in the first dimension, and a second dimension that corresponds to the talkers’ degrees of foreign accent. Correlations between the first dimension (gender) and the talkers’ mean fundamental frequencies—calculated separately for male and female talkers—indicated a moderate, significant correlation for the female talkers in the Same-sentence condition only ($r = 0.59$; $p = 0.04$). Correlations were not significant for the Multi-sentence condition (male: $r = 0.18$; $p = 0.58$; female: $r = 0.05$; $p = 0.88$) or for the male talkers in the Same-sentence condition ($r = -0.01$; $p = 0.98$). These correlations suggest that the gender dimension does not appear to be a scaling of the talkers’ fundamental frequencies. Correlation between the coordinate values of the second dimension and the talkers’ overall degrees of foreign accent was calculated, revealing the second dimension to be strongly correlated with the talkers’ overall foreign accent in both the Multi-sentence condition

(accent: $r = 0.74$; $p < 0.0001$) and the Same-sentence condition (accent: $r = 0.94$; $p < 0.0001$). Even when effects of comprehensibility were reduced in the Same-sentence condition, the second dimension not only remained significantly correlated to degree of foreign accent, the correlation strength increased. This increase in strength of correlation suggests that degree of foreign accent is the more central feature, rather than comprehensibility.

These results in which gender and degree of foreign accent are the two most salient dimensions replicate the findings in Atagi and Bent (2011). Therefore, the accent and comprehensibility ratings completed by the listeners in Atagi and Bent (2011) prior to the free classification task did not appear to significantly affect their classification strategy.

4.0 Experiment 2: Classification by perceived native language

In addition to the tasks in which listeners grouped by general similarity, the current study investigated listeners' abilities to accurately classify talkers based on perceived native language background when explicitly instructed to do so. In two previous studies that have employed forced-choice tasks, listeners were able to identify the native language backgrounds of nonnative talkers with above-chance accuracy. In a four-alternative forced-choice native language identification task with nonnative speakers of English (Derwing & Munro, 1997), native listeners correctly identified the native language at an average rate of 52% (ranging 41 – 63% depending on the native language). A six-alternative forced-choice accent identification study with nonnative speakers of French (Vieru, Boula de Mareuil, & Adda-Decker, 2011) also found that native listeners identified the native language background of nonnative talkers with above chance accuracy at an average of 52% (ranging 25 – 77% depending on the native language). However, forced-choice tasks are limited by restricting listeners' responses to categories that are provided by the experimenter rather than expressing their own perceptual representations (Clopper & Pisoni, 2007). Results from free classification tasks suggest that listeners make more fine-grained distinctions of dialect variation than forced-choice tasks indicate (Clopper & Pisoni, 2004b, 2006, 2007). To reduce the possible response biases introduced by forced-choice tasks, a free classification task was used in the current study to investigate listeners' classification of talkers based specifically on perceived native language background. The current experimental design also allowed for an investigation of how two properties of the stimulus set—degree of intelligibility and inclusion of native talkers—influenced classification accuracy and perceptual organization.

4.1 Listeners

The same listeners as Experiment 1 participated in Experiment 2.

4.2 Methods

Experiment 2 was administered approximately 30 minutes after the completion of Experiment 1. Listeners participated in two other free classification tasks between Experiments 1 and 2. For Experiment 2, listeners completed four free classifications, which were administered with the same set up and equipment as described in section 3.2. In Experiment 2, however, listeners were instructed to group talkers by perceived native language. The final two classifications included four native talkers, for a total of 28 talkers (instead of 24); therefore, 28 icons were presented on the screen.

4.3 Stimuli

Across the four classification tasks, four different HINT-C sentences were used as stimuli (Table 1). Again, all stimuli were selected from the Hoosier Database. Within each

classification task, the same sentence was used for all talkers, similar to the Same-sentence condition in Experiment 1. The four tasks were manipulated such that the average level of intelligibility (calculated as percent keywords correct) across the nonnative talkers were high for two of the sentences (99.6% and 99.7%) and low for the other two (77.2% and 75.6%) (Table 1; see Appendix C for more details). Additionally, in the two final classification tasks (one with high and one with low intelligibility), the talker set included an additional four native talkers, for a total of 28 talkers. All listeners completed the two classifications with only the nonnative talkers (Hi/Lo-NonNatOnly) before the conditions in which native talkers were included (Hi/Lo-NativeIncl). The presentation order of the high and low intelligibility conditions was counterbalanced across listeners.

4.4 Results and discussion

On average, listeners created 6.9 groups in the Hi-NonNatOnly condition; 7.8 in Lo-NonNatOnly; 6.5 in Hi-NativeIncl; and 8.3 in Lo-NativeIncl. Listeners created more groups than the actual number of native language groups presented (Hi-NonNatOnly: $t = 2.42$; $p < 0.05$; $df = 27$; Lo-NonNatOnly: $t = 3.79$; $p < 0.005$; $df = 27$; Lo-NativeIncl: $t = 2.24$; $p < 0.05$; $df = 27$), except in one condition (Hi-NativeIncl: $t = -1.31$; ns; $df = 27$). The results were first analyzed for accuracy of classification by native language background. Then, to examine listeners' perceptual space for these talkers when instructed to classify by perceived native language, the results of each condition was submitted to MDS analysis. An additive clustering analysis was also conducted to model the perceived similarities between the native language groups.

4.4.1 Accuracy of classification by perceived native language—Unlike

Experiment 1 where the task was a subjective classification of perceived similarity, the task in Experiment 2 allows for measurements of objective accuracy—that is, listeners can make correct and incorrect pairings. Classification accuracy was measured using an analysis that is similar to a signal detection analysis (Macmillan, 1993), where a difference score, which is similar to d -prime, accounts for the number of talker groups that each listener created (Clopper & Bradlow, 2009). First, proportion of correct pairings (“hits”) is calculated by dividing the number of correct talker pairings by native language background out of the total possible number of correct talker pairings. Talker pairings were considered correct when two talkers with the same native language background were grouped together. Then, proportion of pairwise errors (“false alarms”) was calculated based on the number of pairwise talker language background mismatches out of the total number of incorrect pairings. Talker pairings were considered an error when two talkers with different native language backgrounds were grouped together. Finally, pairwise errors were subtracted from correct pairings for each listener to calculate the difference score (d -prime = hits – false alarms).

The difference score across the four conditions were analyzed with two-way repeated-measures ANOVAs with level of intelligibility (high or low) and inclusion of native talkers (nonnative talkers only or native talkers included) as the two within-subject variables. There was a main effect of intelligibility ($F(1,27) = 6.5$; $p < 0.05$). Classification was overall more accurate for the high intelligibility conditions than the low intelligibility conditions. There was also a main effect of native talker inclusion ($F(1,27) = 50.4$; $p < 0.0001$). Overall classification accuracy was higher when native talkers were included in the stimulus set. Interaction between intelligibility and native talker inclusion was not significant.

Figure 3 shows rates with which listeners grouped talkers from each of the native language backgrounds, both for correct pairings (on the diagonal) and for pairwise errors (off the diagonal). The accuracy rates for the correct pairing of nonnative talkers were not high

overall, but were at above-chance levels for almost all language backgrounds, with only a few exceptions. These exceptions were the French talkers in all but the Hi-NativeIncl condition, and the Spanish talkers in the Lo-NonNatOnly condition. Listeners were strikingly accurate at classifying the native talkers with each other, as indicated by the nearly perfect scores (100 and 98) in Figures 3C and 3D. However, rates for incorrect grouping of nonnative talkers with native talkers were higher than would be expected based on previous literature on the high discriminability between native and nonnative speech (Flege, 1984; Park, 2008).

The increased overall classification accuracy when native talkers were included is at least partially due to the listeners' ability to accurately group the four native talkers together consistently. Nevertheless, native talker grouping alone does not account for the increase in classification accuracy. Several language backgrounds increased in classification accuracy by at least 5% when native talkers were included (i.e., Japanese and Mandarin in the high intelligibility conditions; Japanese and Korean in the low intelligibility conditions). Moreover, the grouping of some nonnative talkers with the native talkers contributed to the rates of pairwise error, but the difference scores increased nonetheless.

4.4.2 Multidimensional scaling—To examine the perceptual organization of all talkers as in Experiment 2, a talker-by-talker confusion matrix summed across all listeners was computed for each condition. The matrices for the conditions with nonnative talkers only were 24×24 in size; for the conditions with native talkers included, 28×28. Each matrix was then submitted to the same multidimensional scaling analysis used in Experiment 1. Figure 4 shows the resulting two-dimensional MDS outputs. Across all four conditions, one dimension was found to significantly correlate with the talkers' overall degrees of foreign accent. The correlation coefficients were higher for the low intelligibility conditions (Lo-NonNatOnly: $r = 0.79$; $p < 0.0001$; Lo-NativeIncl: $r = 0.88$; $p < 0.0001$) than the high intelligibility conditions (Hi-NativeIncl: $r = 0.77$; $p < 0.0001$; Hi-NonNatOnly: $r = 0.63$; $p < 0.001$).

The MDS solutions for the high intelligibility conditions (Figures 4A and 4C) show clusters of talkers that appear to be based on the talkers' native language backgrounds. In both of these MDS solutions, Japanese, Korean, and Mandarin talkers make up one cluster; another cluster consists of all four of the German talkers and two of the French talkers; and a third cluster comprises of Spanish talkers and one or two French talkers. In the low intelligibility conditions (Figures 4B and 4D), there are less defined clusters, and groups based on native language backgrounds are harder to discern. In these conditions, there are many instances of longer intergroup distances than intra-group distances. This difference between the high and low intelligibility conditions parallel the classification accuracy results discussed above, in which listeners' classification accuracy in the low intelligibility conditions were significantly worse. The nearly perfect classification of the native talkers with each other (as discussed in the previous section) results in the convergence of all four native talkers onto one coordinate point in Figures 4C and 4D.

The second dimension is difficult to interpret in Experiment 2, whereas in Experiment 1 the second dimension was clearly related to gender. Although there is some segregation of the male and female talkers in three of the MDS solutions, there is not a clear divide between the genders as was observed in Experiment 1.

4.4.3 Additive clustering analyses—Listeners' groupings of talkers by perceived native language were also analyzed using ADDTREE, an additive cluster analysis (Sattath & Tversky, 1977). Unlike ultrametric trees such as those produced by hierarchical clustering, the intra-cluster distances may exceed inter-cluster distances in an additive tree. This feature

of additive trees allow for a correspondence to a two-dimensional spatial solution for psychological distances (such as a two-dimensional MDS) in the number of parameters incorporated in the representation (Sattath & Tversky, 1977). In the tree representations produced by an additive clustering model, the total length of the vertical branches between any two nodes represents degree of perceptual similarity; horizontal distance and position are irrelevant.

Across the four classification conditions, two 6×6 (Hi-NonNatOnly and Lo-NonNatOnly) and two 7×7 (Hi-NativeIncl and Lo-NativeIncl) similarity matrices of native language backgrounds were computed for each condition. Each matrix represents the similarity of the native language background categories. Each cell in one of these matrices was calculated by summing the total number of times each listener grouped together any of the four talkers with the same native language background (Clopper & Pisoni, 2007). Each of these similarity matrices was submitted to ADDTREE to produce graphical tree models of the perceptual similarity of the native language backgrounds.

The additive trees in Figure 5 reveal a similarity structure of the native language backgrounds that remain consistent within high or low intelligibility conditions. In the high intelligibility conditions, the three Asian language backgrounds cluster closely with each other, with Korean and Mandarin clustering most closely (Figures 5A and 5C). In contrast, in the low intelligibility conditions, Japanese and Korean cluster together, which then cluster with Spanish (Figures 5B and 5D). French and German cluster together in all four conditions.

In the conditions where native talkers were included (Figures 5C and 5D), the American English (native) background is embedded into the organizational structure—clustering with German and French—rather than simply added on as a separate branch of their own. This result parallels that of the classification accuracy discussed earlier: some nonnative talkers were classified with the native talkers in both Hi-NativeIncl and Lo-NativeIncl conditions. The inclusion of native talkers, however, did not significantly rearrange the overall organization of the nonnative talkers' native language backgrounds. That is, other than the addition of a node for American English, the structures for Hi-NonNatOnly and Hi-NativeIncl (Figures 5A and 5C) are very similar to each other, as are the structures for Lo-NonNatOnly and Lo-NativeIncl (Figures 5B and 5D).

5.0 Discussion

Taken together, the results from the current study suggest that degree of foreign accent is a central perceptual dimension when listeners classify nonnative talkers by either general similarity or perceived native language. However, manipulating the overall intelligibility of the stimulus set and focusing the listeners' attention on talker language background attenuated the salience of accent strength. Specifically, the perceptual salience of accentedness was reduced when all talkers were highly intelligible and listeners were instructed to classify talkers by perceived native language. In classifying by perceived native language, listeners consistently found French and German accents to be similar. In contrast, Korean and Japanese accents were perceptually similar to the Mandarin accent in the high intelligibility conditions, but were more similar to the Spanish accent in the low intelligibility conditions. Additionally, the inclusion of native speech did not significantly change the perceptual similarity space for nonnative speech.

Previous auditory free classification studies with regional dialects (e.g., Clopper & Pisoni, 2007) and different languages (Bradlow et al., 2010) both claimed the presence of uncommon phonetic or phonemic characteristics as a salient perceptual dimension. The salience of foreign accent strength found in the current study is similar to the salience of

uncommon phonetic and phonemic characteristics. That is, nonnative talkers' productions may be principally judged by the amount of phonological variants that were present in the speech signal and detected by the listener as being different from native speech, resulting in the perception of varying degrees of nonnative accent.

Together, the findings across these studies suggest that the perception of variability across several sources of speech variation—including regional dialects, nonnative accents, as well as different languages—could all involve a scaling between the listener's own linguistic representation and speech samples that deviate from that representation. A scaling of all of these varieties would place regional dialects on one side of a spectrum, with phonologically similar regional dialects situated short distances away from each other. Foreign languages would be located on the opposite end of the spectrum. Nonnative speech would be located in the middle. To capture how listeners represent the full spectrum of within-language variation, future studies will need to incorporate non-American varieties of English along with regional dialects and nonnative speech from a wider variety of native language backgrounds.

Gender was also salient when listeners classified talkers by general perceptual similarity. With the change in the instruction to group the talkers by native language, the salience of gender disappeared completely or became significantly diminished. In the condition where the talkers were all highly intelligible and native talkers were included in the stimulus set, male and female talkers were not at all distinguishable in the two-dimensional MDS solution. In this condition, listeners were also most accurate in their native language classification, suggesting that listeners were more successful in this condition at ignoring the irrelevant features for accent classification, such as gender. Listeners may have been less successful at ignoring gender in the other three conditions—female talkers appeared to one side of the MDS space; male talkers, however, were dispersed all across the space. The change in instruction, therefore, significantly weakened the salience of gender, especially for the male talkers.

When listeners classified talkers by perceived native language, two organizational structures emerged: one for the low intelligibility conditions and another for the high intelligibility conditions. While the French and German accents were closest in perceptual distance to each other across all conditions, the organizational structure of the four other nonnative accents changed across the two intelligibility conditions. The Japanese and Korean accents were most perceptually similar to the Spanish accent in the low intelligibility conditions, but most similar to the Mandarin accent in the high intelligibility conditions. The structure found with the low intelligibility conditions appears to be driven by the talkers' strengths of foreign accent. Specifically, with the stimuli used here, Spanish, Japanese, and Korean native language groups are on average the more accented native language backgrounds (Appendix A). This accent-based structure for the low intelligibility conditions—but not in the high intelligibility conditions—is also consistent with the results of the two dimensional MDS solutions showing that degree of foreign accent was more strongly correlated with the first dimension in the low intelligibility conditions than the high intelligibility conditions. A limitation of the current design is that one unique sentence was used for each condition. Thus, the current results could have been influenced by the specific sentence that was assigned to each condition. Future work should address this limitation by including multiple sentences when examining the perceptual similarities of nonnative accents.

The organizational structure in the high intelligibility conditions appears to be based on a factor other than accent strength. In the high intelligibility conditions, both the MDS and clustering solutions revealed that all three Asian language backgrounds were perceived to be very similar to each other. In contrast, the MDS solutions for the low intelligibility

conditions show that there was less agreement across listeners regarding whether the Mandarin talkers are classified with the German talkers or with the Japanese and Korean talkers. To account for the perceptual similarity of the Asian language backgrounds in the high intelligibility classification schema, future research should explore several possible explanations.

The talkers' voice qualities or use of prosody could have influenced listeners' nonnative speech classification performance. Munro, Derwing, and Burgess (2010) found that listeners' ability to detect a nonnative accent remained at above-chance levels even with speech that has been content-masked, temporally disrupted, or monotonized in pitch variation. Although it is often assumed that nonnative accents stem from native language interference on L2 production, the results reported by Munro et al. (2010) suggest that non-segmental features of speech, such as voice quality and stylistic use of prosody, are likely to be important factors in nonnative speech perception. The perceptual similarity of the three Asian language backgrounds could have been due to such suprasegmental similarities that have not yet been thoroughly explored. Future work should examine listeners' perceptual similarity judgments of the current stimuli's suprasegmental features to investigate the roles played by voice quality and prosody during nonnative speech classification.

Another factor that may lead to similarities in the nonnative talkers' production of English is phonological similarities between the talkers' native languages. The phonological structures of Japanese, Korean, and Mandarin, however, are quite distinct from each other, both in their phonemic inventories and phonotactic constraints. Nevertheless, relative to the European languages, the Asian languages included in the current study have more restricted syllable structures. Though none of the stimuli used in this study were characterized by extensive deletions or epenthesis, such restrictions may have led to similarities in the Japanese, Korean, and Mandarin talkers' productions of English. Future work should include more constrained stimuli to facilitate the examination of the effects that native language phonemic inventory and phonotactics have on the perceptual similarities of nonnative accents found in the current study.

The perceptual similarity of the Asian language backgrounds could also be due to listeners' familiarity and sociolinguistic experiences (Clopper, 2004; Clopper & Pisoni, 2007) with nonnative accents. Specifically, while listeners may not always be aware of the native language of a nonnative talker, visual appearance may be a consistent cue that broadly indicates the talker's geographical origin (e.g., Asian or European). Across multiple interactions with nonnative talkers, listeners may encode several types of information in long-term memory. Some of the information from these experiences could be visual features of talkers. That is, listeners may store and integrate their experiences with accent characteristics and visual appearance into memory, then draw on these past interactions when classifying talkers by native language background. Listeners may therefore be more likely to group talkers with Asian first languages separately from those talkers with European first languages. Future studies will need to systematically investigate the relationship between listeners' nonnative speech classification and their past experiences with nonnative accents, as well as the effects of visual cues on nonnative speech classification.

In addition to models of perceptual similarity, when the instruction was to group the talkers by native language, listeners' classification behavior could be characterized by an objective measure of classification accuracy. Classification accuracy was higher when the stimulus set consisted of high intelligibility sentences compared to lower intelligibility sentences. This result was unexpected, as it would seem that lower intelligibility sentences would contain more phonetic cues to indicate the talkers' native languages. Instead, attending to talkers'

native language backgrounds in the low intelligibility conditions, in which there was greater variability across stimuli, may have increased the task difficulty, leaving fewer cognitive resources available for accurate classification. With this introduction of additional complexity to the task, listeners may no longer have been able to determine which phonetic cues were indicative of a specific native language background and which were irrelevant indexical or talker-specific cues. Under these more challenging conditions, listeners may have defaulted to attending to a more salient feature of nonnative speech (i.e., foreign accent strength). These results suggest that instructing listeners to attend to a different dimension (i.e., talker native language) can attenuate the perceptual salience of both accent strength and gender, but only when adequate attentional resources are available.

Finally, when native talkers were included in the stimulus set, listeners did not reliably distinguish between native and nonnative talkers. This result was reflected in the non-zero off-diagonal scores in the classification matrices (Figure 3), as well as the lack of an independent native accent branch in the clustering analysis (Figure 5). Further, the presence of native stimuli did not rearrange the perceptual organization of nonnative accents, indicating that the inclusion of native talkers largely did not affect the perceptual similarity of nonnative talkers to other nonnative talkers. These results are consistent with the finding of the salience of accentedness in nonnative speech classification—i.e., whether or not native speech was present in the stimulus set, listeners were highly sensitive to the talkers' degrees of foreign accent. Nevertheless, when native talkers were included, listeners classified some of the nonnative talkers with native talkers, especially in the condition where all of the talkers were highly intelligible. In other words, although listeners demonstrated extremely high accuracy for classifying the four native talkers together, they also frequently included nonnative talkers with the native talker group. The additive clustering models of native language group similarities therefore fit the native language groups with the least accented talkers—i.e., French and German. These results contrast with those of Flege (1984) and Park (2008) who reported high sensitivity to nonnative accents in accent detection tasks for native listeners. The difference between the current study and previous studies could have resulted from task differences—accent detection versus free classification—or from the number of talker native language backgrounds included in the experiments. Both Flege (1984) and Park (2008) used nonnative talkers from one native language background, whereas nonnative talkers from six different native languages were included in the current study. Therefore, for highly proficient nonnative talkers, the task and the number of talker language backgrounds presented in the experiment may influence whether listeners can accurately identify the talkers as nonnative. Future accent detection experiments with talkers from multiple native languages will be necessary to tease apart the contribution of these two factors.

In summary, a scaling of similarity to one's own representation for speech appears to be a central organizational schema for listeners' perception of nonnative speech and other forms of speech variability. This scaling can be affected by context changes, such as by focusing the listeners' attention on certain stimulus features through differential instructions and presenting stimuli that are easier for listeners to process. However, other context changes, such as the presence of native speech, may not significantly affect the perceptual organization of nonnative speech stimuli. Future studies should continue to investigate how stimulus set features, task difficulty, and cognitive processes such as attention influence perceptual behavior towards speech variability.

Acknowledgments

This research was supported by NIH-NIDCD R21 DC010027 and NIH-NIDCD T32 DC00012 to Indiana University. Portions of this work were presented at the 162nd Meeting of the Acoustical Society of America in

2011. The authors would like to thank Charles Brandt for programming the experiment and Rob Nosofsky for providing the programs to conduct some of the analyses.

References

- Atagi, E.; Bent, T. Perceptual dimensions of nonnative speech. In: Lee, W-S.; Zee, E., editors. International Congress of Phonetic Sciences XVII. Hong Kong: City University of Hong Kong; 2011. p. 260-263.
- Bent T. Native and non-native speech database for children. *Journal of the Acoustical Society of America*. 2010; 127:1905.
- Best CT, McRoberts GW, Goodell E. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*. 2001; 109:775–794. [PubMed: 11248981]
- Bradlow AR, Clopper CG, Smiljanic R, Walter MA. A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication*. 2010; 52:930–942. [PubMed: 21179563]
- Clopper, CG. Doctoral dissertation. Indiana University Bloomington; 2004. Linguistic experience and the perceptual classification of dialect variation.
- Clopper CG. Auditory free classification: Methods and analysis. *Behavior Research Methods*. 2008; 40:575–581. [PubMed: 18522069]
- Clopper CG, Bradlow AR. Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech*. 2008; 51:175–198. [PubMed: 19626923]
- Clopper CG, Bradlow AR. Free classification of American English dialects by native and non-native listeners. *Journal of Phonetics*. 2009; 37:436–451. [PubMed: 20161400]
- Clopper CG, Conrey B, Pisoni DB. Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology*. 2005; 24:182–206. [PubMed: 21423866]
- Clopper CG, Levi SV, Pisoni DB. Perceptual similarity of regional dialects of American English. *Journal of the Acoustical Society of America*. 2006; 119:566–574. [PubMed: 16454310]
- Clopper CG, Pisoni DB. Effects of talker variability on perceptual learning of dialects. *Language and Speech*. 2004a; 47:207–239. [PubMed: 15697151]
- Clopper CG, Pisoni DB. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*. 2004b; 32:111–140. [PubMed: 21451736]
- Clopper CG, Pisoni DB. Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change*. 2006; 18:193–221. [PubMed: 21423820]
- Clopper CG, Pisoni DB. Free classification of regional dialects of American English. *Journal of Phonetics*. 2007; 35:421–438. [PubMed: 21423862]
- Derwing TM, Munro MJ. Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*. 1997; 19:1–16.
- Flege JE. The detection of French accent by American listeners. *Journal of the Acoustical Society of America*. 1984; 76:692–707. [PubMed: 6491043]
- Flege, JE. Second Language Speech Learning Theory, Findings, and Problems. In: Strange, W., editor. *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press; 1995. p. 233-277.
- Flege JE, Fletcher KL. Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*. 1992; 91:370–389. [PubMed: 1737886]
- Goslin J, Duffy H, Floccia C. An ERP investigation of regional and foreign accent processing. *Brain and Language*. 2012; 122:92–102. [PubMed: 22694999]
- Joos, M. *Acoustic phonetics*. Baltimore: Linguistic Society of America; 1948.
- Luce PA, Lyons EA. Specificity of memory representations for spoken words. *Memory & Cognition*. 1998; 26:708–715. [PubMed: 9701963]
- Macmillan, NA. Signal detection theory as data analysis method and psychological decision model. In: Keren, G.; Lewis, C., editors. *A Handbook for data analysis in the behavioral sciences: Methodological issues*. Hillsdale NJ: Lawrence Erlbaum Associates Inc; 1993. p. 21-57.

- McLennan CT, Gonzalez J. Examining talker effects in the perception of native- and foreign-accented speech. *Attention Perception & Psychophysics*. 2012; 74:824–830.
- Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech-perception. *Perception & Psychophysics*. 1990; 47:379–390. [PubMed: 2345691]
- Munro MJ, Derwing TM. Foreign accent, comprehensibility and intelligibility in the speech of 2nd language learners. *Language Learning*. 1995a; 45:73–97.
- Munro MJ, Derwing TM. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*. 1995b; 38:289–306. [PubMed: 8816082]
- Munro MJ, Derwing TM, Burgess CS. Detection of nonnative speaker status from content-masked speech. *Speech Communication*. 2010; 52:626–637.
- Nilsson, M.; Soli, SD.; Gelnett, DJ. *Development of the Hearing in Noise Test for Children (HINT-C)*. Los Angeles, CA: House Ear Institute; 1996.
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Perception & Psychophysics*. 1998; 60:355–376. [PubMed: 9599989]
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychological Science*. 1994; 5:42–46. [PubMed: 21526138]
- Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology-Learning Memory and Cognition*. 1993; 19:309–328.
- Park, H. Doctoral dissertation. Indiana University; 2008. Phonological information and linguistic experience in foreign accent detection.
- Sattath S, Tversky A. Additive similarity trees. *Psychometrika*. 1977; 42:319–345.
- Tversky A. Features of Similarity. *Psychological Review*. 1977; 84:327–352.
- Vieru B, Boula de Mareuil P, Adda-Decker M. Characterisation and identification of non-native French accents. *Speech Communication*. 2011; 53:292–310.
- Wells, JC. *Accents of English: An Introduction*. Cambridge: Cambridge University Press; 1982.

Appendix A

Characteristics of the talkers in the Hoosier Database: overall degrees of foreign accent and levels of comprehensibility as measured in Atagi & Bent (2011) on a nine-point scale; overall levels of intelligibility as measured in Bent (2010) (percent keywords correct, in quiet, across 250 sentences).

Talker	Native Language	Gender	Overall degree of accent	Overall level of comprehensibility	Overall level of intelligibility (sentence level)
E1	English	male	1.1	1.0	99.5
E2	English	male	1.4	1.2	99.7
E3	English	female	1.1	1.1	99.4
E4	English	female	1.1	1.1	99.1
F1	French	male	6.3	4.2	93.4
F2	French	male	4.7	2.9	95.9
F3	French	female	4.1	2.5	97.2
F4	French	female	4.3	2.8	97.5
G1	German	male	3.6	2.4	97.7
G2	German	male	3.0	2.0	96.8
G3	German	female	4.8	3.0	97.1

Talker	Native Language	Gender	Overall degree of accent	Overall level of comprehensibility	Overall level of intelligibility (sentence level)
G4	German	female	3.0	2.0	97.9
J1	Japanese	male	5.9	4.1	88.7
J2	Japanese	male	5.5	3.5	91.7
J3	Japanese	female	6.0	3.9	91.0
J4	Japanese	female	6.5	4.9	95.0
K1	Korean	male	5.8	3.6	93.7
K2	Korean	male	5.8	3.6	95.1
K3	Korean	female	5.1	3.4	96.2
K4	Korean	female	5.5	3.3	96.5
M1	Mandarin	male	4.1	2.8	95.6
M2	Mandarin	male	3.7	2.5	96.5
M3	Mandarin	female	4.4	2.8	96.9
M4	Mandarin	female	4.8	3.6	96.6
S1	Spanish	male	5.9	3.8	92.0
S2	Spanish	male	7.9	6.1	92.5
S3	Spanish	female	6.2	4.3	90.9
S4	Spanish	female	6.8	4.7	92.2

Appendix B

Stimuli sentences for each talker used in the Multi-sentence condition, and their respective levels of intelligibility (percent keywords correct) as measured in Bent (2010).

Talker	Sentence used in the Multi-sentence condition	Intelligibility of sentence
F1	Men normally wear long pants.	78
F2	They went on vacation.	98
F3	Children like strawberries.	97
F4	They rode their bicycles.	95
G1	The ground was very hard.	98
G2	He found his brother hiding.	98
G3	Father paid at the gate.	98
G4	They heard a funny noise.	100
J1	The kitchen clock was wrong.	100
J2	She writes to her friend daily.	62
J3	The park is near a road.	95
J4	The sky was very blue.	90
K1	The police cleared the road.	83
K2	Her sister stayed for lunch.	100

Talker	Sentence used in the Multi-sentence condition	Intelligibility of sentence
K3	The cow was milked everyday.	90
K4	The house has nine bedrooms.	98
M1	She made her bed and left.	100
M2	A child ripped open the bag.	100
M3	Father forgot the bread.	100
M4	He closed his eyes and jumped.	98
S1	The three girls were listening.	100
S2	The yellow pears taste good.	50
S3	School got out early today.	96
S4	The bus stopped suddenly.	87

Appendix C

Each talker's intelligibility levels for the sentences used in the four conditions of Experiment 2.

Talker	HiNonNatOnly "A tree fell on the house"	LoNonNatOnly "The tub faucet was leaking"	HiNativeIncl "The ice cream is melting"	LoNativeIncl "The grocer sells butter"
E1	--	--	100	87
E2	--	--	100	97
E3	--	--	100	100
E4	--	--	100	100
F1	98	83	100	80
F2	100	78	100	53
F3	98	93	100	83
F4	100	88	100	100
G1	100	75	100	100
G2	100	83	100	90
G3	98	73	100	97
G4	100	90	100	100
J1	100	45	100	47
J2	100	83	100	90
J3	100	70	98	27
J4	100	70	100	87
K1	100	80	100	47
K2	100	73	98	63
K3	100	73	100	70
K4	100	80	100	70
M1	100	73	100	77
M2	100	73	98	93
M3	100	73	100	83

Talker	HiNonNatOnly “A tree fell on the house”	LoNonNatOnly “The tub faucet was leaking”	HiNativeIncl “The ice cream is melting”	LoNativeIncl “The grocer sells butter”
M4	100	78	100	73
S1	100	58	98	80
S2	100	63	100	100
S3	100	78	100	67
S4	100	90	100	77

Highlights

- Listeners classified nonnative accents by overall similarity and native language.
- Level of intelligibility and inclusion of native speakers were manipulated.
- Degree of accent was highly salient in classification of nonnative speech.
- Attention to native language background weakened the salience of accent.
- Perception of phonological deviation is central to representing speech variability.

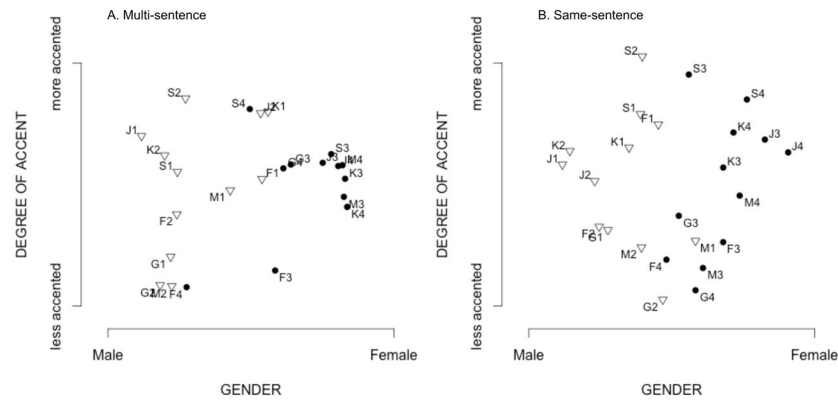


Figure 1. Two-dimensional MDS solutions for the Multi-sentence (A) and Same-sentence (B) conditions in which listeners grouped talkers by overall perceived similarity. Each point on the MDS solution represents a talker and is labeled with a unique talker ID as indicated in Appendix A. Each talker ID includes information about the talker's native language background (indicated with the first letter of the language name). Male talkers are indicated by open triangles; filled circles indicate female talkers.

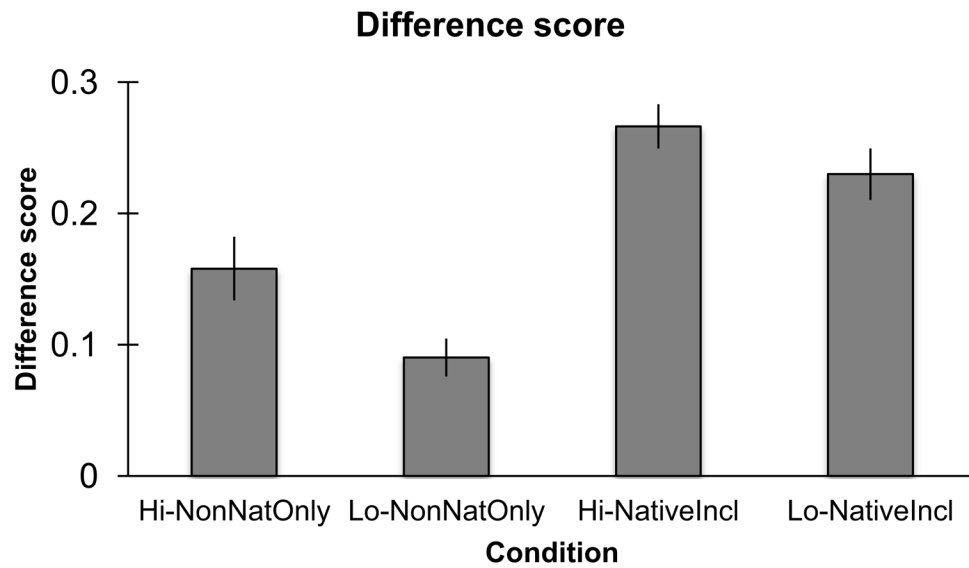


Figure 2. Classification performance by perceived native language background across all four conditions (Hi-NonNatOnly, Lo-NonNatOnly, Hi-NativeIncl, Lo-NativeIncl), in difference scores. “Hi/Lo” indicate the condition’s overall intelligibility level; “NativeIncl” conditions are where native talkers were included, and “NonNatOnly” conditions are when the stimuli set consisted only of nonnative talkers. Error bars indicate standard error.

A. Hi-NonNatOnly							C. Hi-NativeIncl						
	French	German	Japanese	Korean	Mandarin	Spanish	English	French	German	Japanese	Korean	Mandarin	Spanish
French	14						98						
German	19	44					19	16					
Japanese	16	11	23				47	22	45				
Korean	11	4	29	38			0	15	4	35			
Mandarin	16	13	24	31	33		1	14	6	38	38		
Spanish	11	4	8	11	9	31	2	10	6	36	40	42	
							0	19	5	19	17	13	24
B. Lo-NonNatOnly							D. Lo-NativeIncl						
	French	German	Japanese	Korean	Mandarin	Spanish	English	French	German	Japanese	Korean	Mandarin	Spanish
French	16						100						
German	18	30					15	10					
Japanese	10	11	21				10	18	33				
Korean	8	7	26	34			0	12	13	35			
Mandarin	14	21	19	17	24		2	12	10	25	39		
Spanish	14	6	13	13	9	12	18	15	14	25	16	17	
							1	12	5	14	19	10	15

Figure 3.

Classification matrices indicating rates of correct pairings (cells on the diagonal) and of incorrect pairings (cells off of the diagonal) between each of the native language backgrounds presented to listeners in each condition. Each cell indicates the rate at which talkers with the two language backgrounds of the corresponding row and column were classified together. Each cell contains a score between 0 and 100, where 0 indicates that no talker of a given native language background was paired with any talker from the other native language background. The diagonal indicates rates of correct native language grouping (i.e., “hits”); the off-diagonal cells are the rates of incorrect pairings (i.e., “false alarms”). An off-diagonal score of 100 would indicate that all listeners grouped all four talkers from one native language background with all four talkers from the other native language background. Numbers in bold indicate native language pairings that were numerically above chance (i.e., greater than 16.7 for Hi-NonNatOnly and Lo-NonNatOnly; greater than 14.3 for Hi-NativeIncl and Lo-NativeIncl).

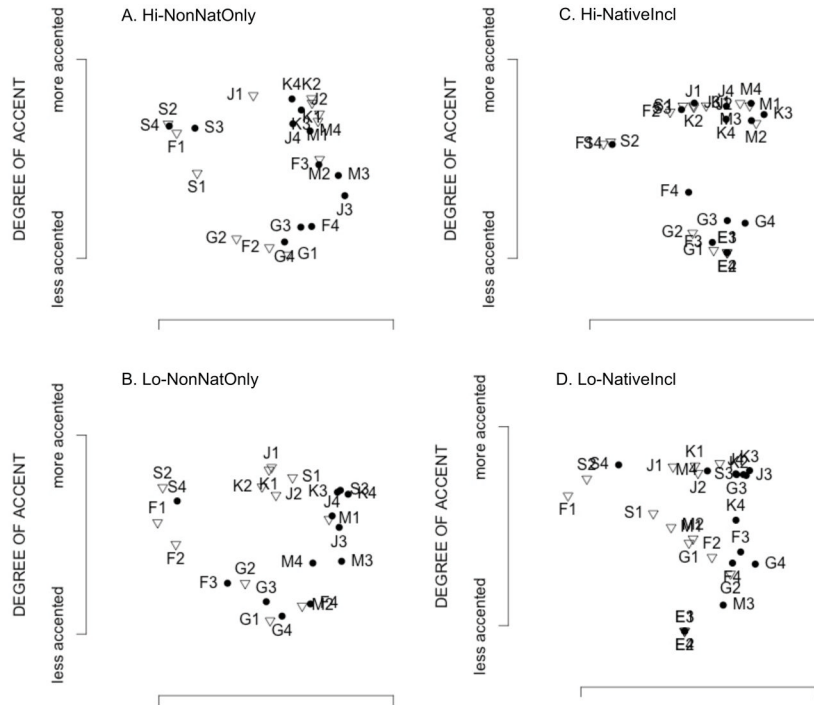


Figure 4. Two-dimensional MDS solutions for four conditions in Experiment 2, in which listeners grouped talkers by perceived native language background. Each point on the MDS solution represents a talker and is labeled with a unique talker ID as indicated in Appendix A. Each talker ID includes information about the talker’s native language background (indicated with the first letter of the language name). Male talkers are indicated by open triangles; filled circles indicate female talkers.

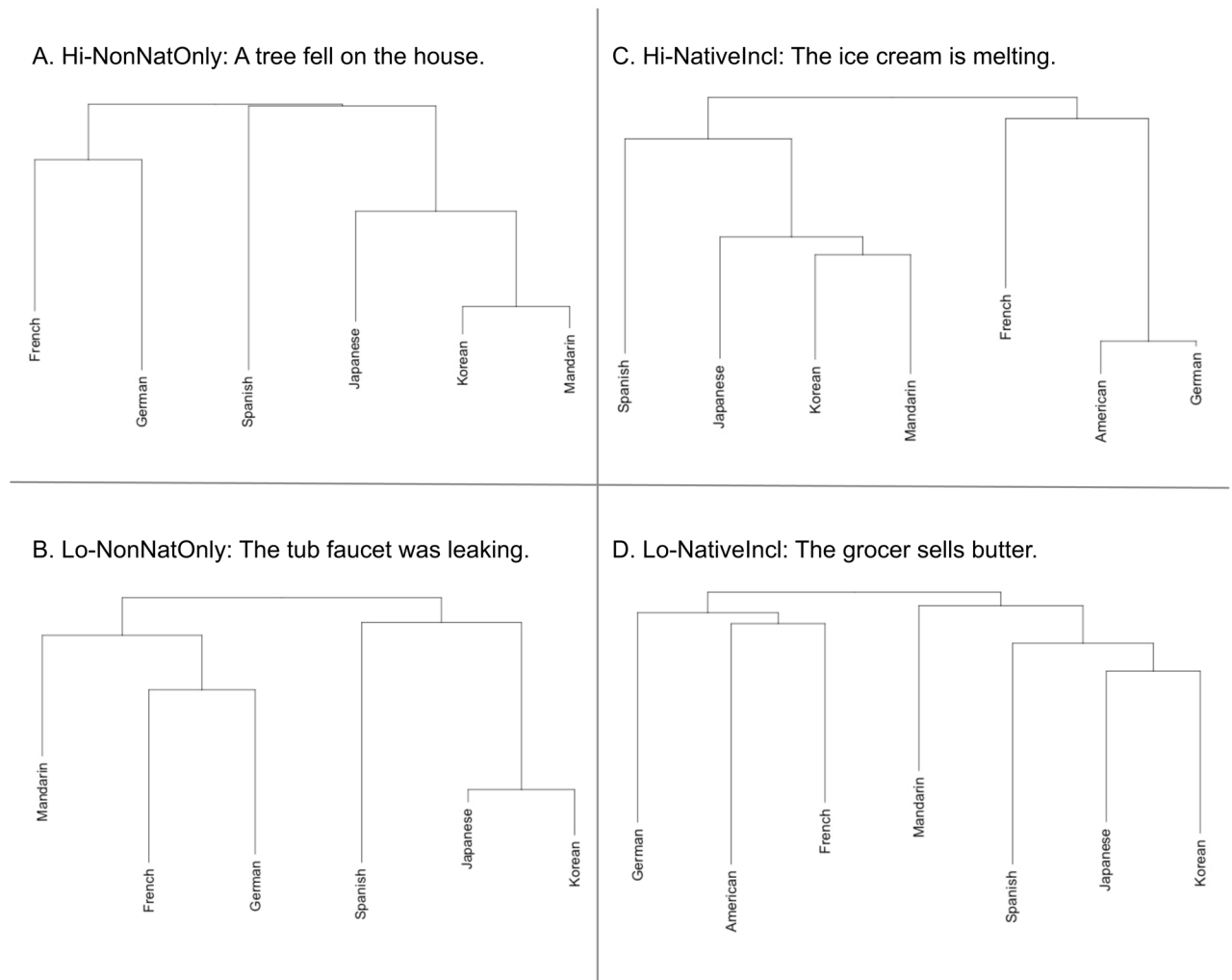


Figure 5.
ADDTREE clustering solutions across all four conditions.

Table 1

Characteristics of stimuli in the perceived native language classification tasks.

Talkers	Intelligibility	Intelligibility of nonnative talkers		Sentence	Condition Name
		Average	Range		
24 nonnative	High	99.7%	98 – 100%	A tree fell on the house.	Hi-NonNatOnly
24 nonnative	Low	75.6%	45 – 93%	The tub faucet was leaking.	Lo-NonNatOnly
4 native and 24 nonnative	High	99.6%	98 – 100%	The ice cream is melting.	Hi-NativeIncl
4 native and 24 nonnative	Low	77.2%	27 – 100%	The grocer sells butter.	Lo-NativeIncl