PERCEPTUAL DIMENSIONS OF NONNATIVE SPEECH

Eriko Atagi & Tessa Bent

Department of Speech and Hearing Sciences, Cognitive Science Program, Indiana University, USA
eatagi@indiana.edu; tbent@indiana.edu

ABSTRACT

Foreign-accented speech has most commonly been characterized across three related, but independent dimensions: intelligibility, comprehensibility, and accent [6]. The present study applied an auditory free classification task, which has been used to test listeners' perceptual representations of regional dialects [5] and different languages [3], to further investigate the salient perceptual dimensions of foreign-accented speech. Participants first rated recordings of 24 nonnative speakers for comprehensibility and degree of accent. The speakers' intelligibility was measured in a separate study. These three dimensions were significantly correlated; comprehensibility and degree of accent were the most highly correlated. Multidimensional scaling analyses of the auditory free classification data revealed two dimensions of perceptual salience: gender and degree of accent. The current results, together with previous findings [3, 5], suggest that listeners' perceptual representation of speech variability centrally involve a scaling of the distance between a speech stimulus and the listener's own linguistic system.

Keywords: speech perception, sociophonetics, speech variability

1. INTRODUCTION

Variability in speech can arise from individual speaker differences such as age and gender; it can also indicate where a speaker is from, based on the speaker's regional dialect, native language, or foreign accent. The information that is extracted and represented by listeners when perceiving such talker-specific variation in the speech signal is key to understanding the speech communication process. Recently, the auditory free classification task has been used to investigate perceptual representations of speech variability including regional dialects [5] and different languages [3]. The auditory free classification task is a method that allows listeners to categorize speech samples in which categorization is not constrained. That is, listeners indicate their perception of similarity among speech samples, rather than having labels or dimensions specified by the experimenter. Using this paradigm, Clopper and Pisoni [5] showed that the perceptual dimensions most salient to listeners for regional dialects of American English were gender, geography (i.e., north-south) and markedness (i.e., the degree to which a speaker's dialect contains marked phonological variants) [5]. Similarly, when the same method was used to study the perceptual similarity space of 17 languages, Bradlow, et al. [3] found three salient dimensions: the presence of marked back consonants (i.e., dorsal consonants that are not /k, ɡ, ŋ/); the presence of marked front vowels; and the geographical location of the language (i.e., east-west). These studies demonstrate how listeners cognitively represent variability stemming from two sources.

Foreign accents are another source of variability in speech. The presence of a foreign accent is highly salient to listeners and has been shown to be detectable by native listeners, even in speech samples as small as one phoneme [12]. Which perceptual features of foreign accented speech are most salient to listeners, however, remains an open question.

Speech produced by nonnative speakers is most commonly characterized by the influence of the first language [2, 7, 8] and on three perceptual dimensions: intelligibility (i.e., listener's ability to accurately report the words that a talker has produced), comprehensibility (i.e., the subjective, perceived ease with which listeners understand speech), and strength of foreign accent [6, 9, 10]. However, characteristics of speech other than these four may be more or equally salient to listeners. To address the issue of how listeners cognitively represent variability present in foreign-accented speech, the current study used the free classification task. This task allowed a direct assessment of native listeners' classification of foreign-accented speech. The parameters were determined by the listeners themselves rather than set by the experimenter.

2. METHODS

2.1. Talkers

Speech samples from 24 nonnative speakers of English with six different native language backgrounds were selected from the Hoosier Database of Native and Non-native Speech for Children [1]. The native languages of these speakers were French, German, Spanish, Japanese, Korean, and Mandarin. There were two female and two male talkers representing each language background. Intelligibility scores of each talker in the database had been measured in a separate study and were calculated based on the proportion of keywords correctly transcribed by ten native listeners per talker [1].

2.2. Listeners

Twenty-seven monolingual, native speakers of American English (18 female, 9 male) were recruited on the Indiana University campus in Bloomington, Indiana, to participate in the current study. Listeners' mean age was 21 (range: 18-37).

2.3. Procedure

Listeners first completed comprehensibility and degree of accent ratings on three sentences selected from the Hearing in Noise Test – Children's Version (HINT-C) [11]. For each sentence, listeners assigned a rating from 1 to 9. For the comprehensibility ratings, the listeners were instructed to indicate how easy or difficult it was to understand the sentence where 1 = "very easy to understand" and 9 = "very difficult to understand"; for the degree of accent ratings, they were instructed to indicate how strong the speaker's accent was where 1 = "no foreign accent" and 9 = "very strong foreign accent" [6, 9, 10, 13].

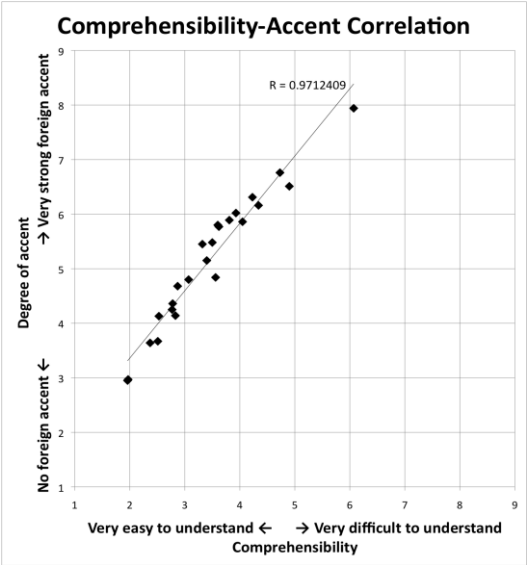
Next, listeners were presented with a 12x12 grid on a computer screen. To the left of this grid were 24 square "icons" labeled with arbitrary, two-letter sequences that were randomly generated for each listener. When the listener clicked on an icon, a speech sample of a talker played over headphones. These speech samples consisted of 24 unique sentences from the HINT-C [11]. The listeners were instructed to drag each icon onto the grid and to group the icons so that speakers who sound similar were 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 also able to listen to the speech samples as many times as they needed. After completing the free classification task, listeners reported their classification strategies in a written, free response.

3. RESULTS

3.1. Intelligibility, comprehensibility, accent

The correlations among the measurements of intelligibility, comprehensibility, and degree of accent were assessed. The correlation between comprehensibility and degree of accent was particularly strong ($r = 0.97$; $p<0.0001$) (Figure 1). Although these measurements were highly correlated, listeners utilized a greater range of the rating scale when rating degree of accent. This difference in range indicates that listeners generally gave lower scores when rating for comprehensibility than for accent, suggesting that even a strongly accented talker was only somewhat difficult to understand.

Figure 1: Correlation of comprehensibility and accent.



Sentence intelligibility was significantly correlated with both comprehensibility ($r = -0.66$, $p < 0.001$) and degree of accent ($r = -0.67$, $p < 0.001$). Talkers with higher comprehensibility scores (i.e., more difficult to understand) and higher degree of accent (i.e., more strongly accented) were more likely to have lower intelligibility scores (i.e., lower proportion of keywords correctly transcribed). However, the

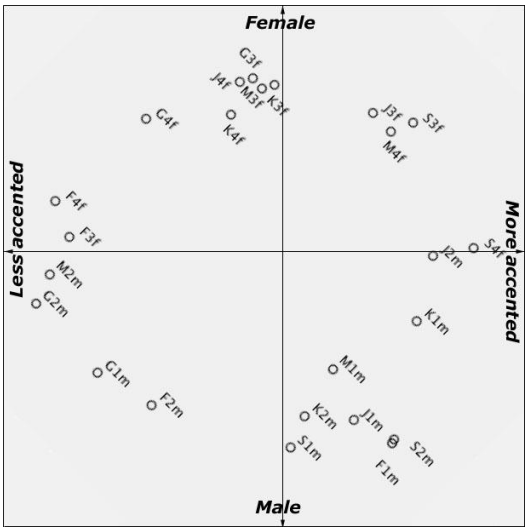
relationship between the subjective measures and intelligibility was not as strong as between the two subjective measures themselves. These results confirm Derwing and Munro's [6] finding that some features of accent may be highly salient and thus have an effect on how listeners perceive accent and comprehensibility, but do not interfere with intelligibility.

3.2. Auditory free classification

In the auditory free classification task, listeners on average created 6.2 groups (range = 4 – 13) with 4.2 talkers per group (range = 1 – 14).

The data from the auditory free classification task was submitted to a multidimensional scaling (MDS) analysis. A 24x24 symmetric similarity matrix was first computed by summing the number of times a pair of talkers was grouped together across all 27 listeners. The matrix thus represents the perceptual similarity of all 24 talkers where 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 27. The MDS analysis was then performed on this talker similarity matrix.

Figure 2: MDS analysis.



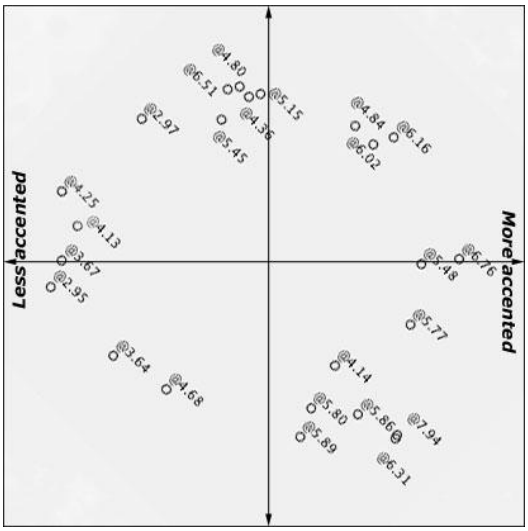
The two-dimensional MDS solution (Figure 2) provided the best fit for the current data; that is, stress was most reduced from the one- to two-dimensional solution and the three- and four-dimensional solutions did not significantly reduce the stress values. Additionally, the two-dimensional MDS solution in Figure 2 has been rotated 45 degrees clockwise to facilitate interpretation along the horizontal-vertical axes.

Each point on the MDS solution represents a talker and is labeled with a unique talker ID. Each talker ID includes information about the talker's native language background (indicated with the first letter of the language name) and gender ("m" for male talkers or "f" for female talkers). For example, "J2m" is a male Japanese speaker and "S3f" is a female Spanish speaker.

The most evident pattern on the MDS solution is a divide of the talkers' genders, apparent along the vertical axis in the MDS solution in Figure 2. There is an unambiguous split between the genders.

The correlations between the coordinates of the horizontal dimension of the rotated MDS solution with comprehensibility ($r = 0.68$, $p < 0.001$), degree of accent ($r = 0.71$, $p < 0.001$), and intelligibility ($r = -0.63$, $p < 0.001$) were significant. Thus, the second dimension could be interpreted as comprehensibility, degree of accent or intelligibility. However, degree of accent provided the strongest correlation and thus the best interpretation. In Figure 3, the same MDS solution is shown again with each point labeled with the talkers' degrees of accent. The difference in goodness of fit between accent and comprehensibility appears to be an effect of the difference in the range of ratings utilized by the listeners as mentioned earlier. The wider range with which listeners perceive degree of accent relative to comprehensibility provides a larger distinction — i.e., perceptual space — between individual talkers.

Figure 3: MDS analysis with each talker's accent.



4. DISCUSSION

The current study aimed to explore an intuitive classification of foreign-accented speech by native listeners and to examine which features of foreign-accented speech are most perceptually salient to native listeners. The auditory free classification task revealed two salient dimensions: gender and degree of accent.

Gender was a clearly very salient dimension. This result is not surprising, as gender is well documented as a consistently salient feature in speech perception [4, 5]. The current study thus confirmed the salience of gender as a feature of speech that remains constant even with foreign-accented speech stimuli.

The native language backgrounds of the talkers did not appear to be one of the most perceptually salient features of the talkers in the current study. The MDS results in Figure 2 indicate no clustering based on language backgrounds (e.g., Mandarin speakers can be found in all four quadrants). Given the abundance of literature in second language phonetics that demonstrates a significant influence of native language phonetics and phonology on second language speech production, e.g., [2, 7, 8], native language background would seem like a possible salient feature of nonnative speech. In fact, 13 of the listeners indicated that they had intended to group the talkers based on their native languages or origins. This strategy for grouping the talkers, however, was often inaccurate. That is, listeners often grouped together talkers with varying native language backgrounds and often reported having groups of talkers from language backgrounds that were not one of the language backgrounds included (e.g., Russian). However, listeners may be able to identify native language background in a task in which they are explicitly asked to do so [14].

5. CONCLUSION

The current study suggests that degree of accent is one of the most salient features of foreign-accented speech for native listeners. One possible interpretation of the accent dimension is to define accent as a type of markedness. Speech containing phonological and phonetic variants that differ substantially from the listener's representation of their own native language are "marked". In studies examining the perceptual representations of other types of speech variability, including American English regional dialects [5] and different languages [3], markedness also appears as a salient perceptual dimension. Taken together with the current results, these studies suggest that there are common cognitive representations of speech variation across distinct sources of variability.

6. ACKNOWLEDGEMENTS

This research would not have been possible without the technical assistance from Charles Brandt and research funding from NIDCD (Grant #1R21DC010021).

7. REFERENCES

[1] Bent, T. 2010. Native and non-native speech database for children. *J. Acoust. Soc. Am.* 127, 1905.

[2] Best, C., McRoberts, G., Goodell, E. 2001. American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoust. Soc. Am.* 109, 775-794.

[3] Bradlow, A.R., Clopper, C.G., Smiljanic, R., Walter, M.A. 2010. A perceptual similarity space for languages: Evidence from five native language listener groups. *Speech Communication* 52, 930-942.

[4] Clopper, C.G., Conrey, B., Pisoni, D.B. 2005. Effects of talker gender on dialect categorization. *J. Language and Social Psychology* 24, 182-206.

[5] Clopper, C.G., Pisoni, D.B. 2007. Free classification of regional dialects of American English. *J Phonetics* 35 421-438.

[6] Derwing, T.M., Munro, M.J. 1997. Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition* 19, 1-16.

[7] Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore: York Press, 233-277.

[8] Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T. 2008. Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NML-e). *Phil. Trans. Roy. Soc.* B363, 979-1000.

[9] Munro, M.J., Derwing, T.M. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45, 73-97.

[10] Munro, M.J., Derwing, T.M. 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38, 289-306.

[11] Nilsson, M.J., Soli, S.D., Gelnett, D.J. 1996. *Development of the Hearing in Noise Test for Children (HINT-C).* Los Angeles: House Ear Institute.

[12] Park, H. 2008. *Phonological Information and Linguistic Experience in Foreign Accent Detection.* Unpublished dissertation, Indiana University.

[13] Southwood, M.H., Flege, J.E. 1999. Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics* 13, 335-349.

[14] Vieru, B., Boula de Mareüil, P., Adda-Decker, M. 2011. Characterisation and identification of non-native French accents. *Speech Communication* 53, 292-310.