A PERCEPTUAL SIMILARITY SPACE FOR LANGUAGES

Ann Bradlow, Cynthia Clopper and Rajka Smiljanic

Northwestern University, Ohio State University
abradlow@northwestern.edu, clopper.1@osu.edu, rajka@babel.ling.northwestern.edu

ABSTRACT

The goal of the present study was to devise a means of representing languages in a perceptual similarity space based on their overall sound structures. In Experiment 1, native English listeners performed a free classification task in which they grouped 17 diverse languages based on their sound similarity. A similarity matrix of the grouping patterns was then submitted to clustering and multidimensional scaling analyses. In Experiment 2, an independent group of native English listeners sorted the group of 17 languages in terms of their distance from English. Taken together, the results of the two experiments provide the basis for developing predictions regarding foreign-accented speech intelligibility.

Keywords: universals and typology, speech perception, language classification

1. INTRODUCTION

An important goal of research on speech communication in a global context is to understand, and ultimately enhance, mutual intelligibility between speakers who communicate through the medium of a language that is the "mother tongue" of just one or neither of the conversation participants. In the present study, we approached this goal by attempting to devise a language classification system whose parameters might reveal the nature and functional implications of foreign-accented speech. Rather than the discovery of language history or the modeling of cross-language segmental perception and production assimilation patterns [2, 4], the overall goal of this language classification enterprise is to predict which foreign accents will most or least impede overall speech intelligibility in cases of various target and source languages. For example, by classifying languages in terms of their overall sound similarity we may be able to explain why native English listeners often find Chinese-accented English harder to understand than Korean-accented English, and why native Chinese listeners (with some knowledge of English) can find Korean-accented English about as intelligible as Chinese-accented or native-accented English [1].

Since overall sound similarity is based on the perceptual integration of multiple acoustic-phonetic dimensions, it cannot easily be determined on the basis of structural analysis. For example, languages A and B, that have no known genetic relationship or known history of population contact, may both have predominantly CV syllable shapes, similarly sized and structured phoneme inventories and a prosodic system with lexical pitch accents. Yet, these two languages may sound to a naïve observer less similar than two languages, C and D, that both have lexical tone systems with both level and contour tones, but have widely differing phoneme inventories and phonotactics. Therefore, in order to capture the possibility that A-accented B may be less intelligible to native B listeners than C-accented D is to native D listeners, we need a language classification system that is based on overall perceived sound similarity.

In the present study, we used a perceptual free classification experimental paradigm [3] with digital speech samples from several natural languages in an attempt to develop a perceptual similarity space for languages. That is, we attempted to create a language classification space with parameters that are based on perception rather than on a priori phonetic or phonological constructs.

2. EXPERIMENT 1

2.1 Method

Samples of 17 languages were selected from the downloadable digital recordings on the IPA website. The samples were all produced by a male native speaker of the language and were between 1.5 and 2 seconds in duration with no disfluencies. The samples were all sentence-final with no intonation breaks in the middle, were easily separable from the rest of the utterance, and included the non-English segments or segment combinations as listed in the language descriptions that accompany the recordings in the IPA Handbook [5].

25 native American English listeners were recruited from the Northwestern University Linguistics Department subject pool (age range 17-30 years) and received course credit for their participation. Listeners were seated in individual sound treated booths in front of a computer. In the center of the screen was a 16x16 grid. On the left of the screen were 17 rectangles with arbitrary labels, e.g. AA, BB etc. Double-clicking on one of these "language icons" caused the speech sample for that language to be played out over headphones.

The listeners were instructed to "group the languages by how they sound." They were to perform this task by dragging the language icons onto the grid in an arrangement that reflected their judgments of how similar the languages sounded: languages that sounded similar should be grouped together on the grid. Languages that sounded different should be in separate groups on the grid. Subjects could take as long as they liked to form their language groups and could form as many groups as they wished. They could listen to each language sample as often as they liked.

Following the free classification task described above, subjects performed a language identification task in which they were asked to listen to the languages and to identify them by name, by geographical region where the language is spoken or by language family to which the language belongs. The purpose of this questionnaire was to ensure that the subjects performed the free classification task based on sound similarity rather than forming groups based on signal-independent knowledge about the languages and their genetic or geographical relationships.

Data from the free classification task were submitted to 3 separate analyses. First, simple descriptive statistics on the grouping patterns were compiled. Second, based on a similarity matrix representing the frequency that each language was grouped together with every other language, clustering (additive similarity tree) and multidimensional scaling (MDS) analyses were performed. The clustering analysis involves an iterative pair-wise distance calculation that provides a means of quantifying the average pair-wise distances across all listeners for the objects (in this case, languages) in the data set. The MDS analysis fits the entire similarity matrix to a model with a specified number of orthogonal dimensions. It therefore provides a representation that facilitates the identification of the physical dimensions that underlie the perceptual similarity space.

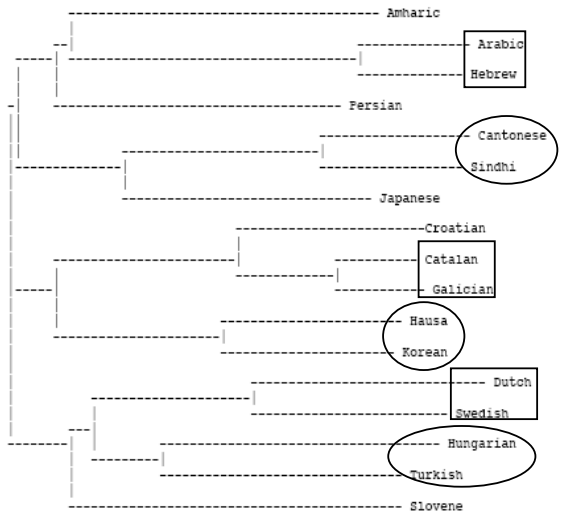
2.2 Results

The questionnaire confirmed that the subjects were generally unable to identify the languages. Responses were scored on a 3 point scale: 0 = incorrect or blank, 1 = correct geographical region (broadly construed) or language family, 2 = correct language. Average scores (across all listeners) ranged from 0.08 for Hausa to 1.16 for Cantonese and Hebrew, with a mean of 0.67.

In the free-classification task, the listeners formed an average of 6.96 groups with 2.57 languages/group. The median and range for the number of groups were 7 and 4-11, respectively. For the number of languages/group, the median and range were 2 and 1-7, respectively.

Figure 1 shows the results of the clustering analysis. The distance between any two languages can be determined by summing the lengths of the horizontal lines that must be traversed to get from the position of one language to the other. This clustering analysis was able to capture a substantial portion of the variance associated with the task (R^2 value of .80).

Figure 1. Clustering analysis.

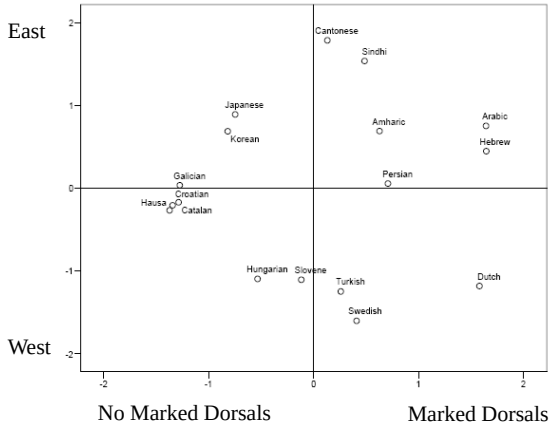


The circles and squares indicate the 6 language pairs that were judged to sound the most similar to each other. Of these 6 pairs, the 3 in squares are languages that are known to be closely related genetically and to share numerous sound structure features: Arabic - Hebrew, Catalan - Galician, Dutch - Swedish. The fact that these 3 pairs were grouped together provides some confirmation of the validity of the technique in terms of its sensitivity to sound-based perceptual similarity of languages to naïve listeners with short speech samples. The

other 3 pairs (in circles), Cantonese - Sindhi, Hausa - Korean, and Hungarian - Turkish do not represent languages with known genetic relationships.

In order to gain some insight into the perceptual dimensions that underlie the patterns of similarity judgment, we then examined the 2-dimensional MDS solution (Figure 2). This solution provided the best fit for these data as indicated by the "elbow" in stress values: for the 1-, 2- and 3-dimensional solutions, the stress values were .49, .22 and .15, respectively.

Figure 2: MDS analysis.



Dimension 1 (horizontal) in the MDS space shown in Figure 2 appeared to sort languages according to the presence or absence of marked dorsal segments in the consonant inventory. In particular, the languages on the right hand side of the space are languages with more than just /k/ or /g/ in their inventory of dorsals, including /x/ and other sounds produced further back in the vocal tract. Dimension 2 (vertical) appeared to divide languages along a geographical east-west dimension. While this is not an acoustic dimension per se, it may reflect some combination of sound structure features that spread (due to either contact or genetic relationship) across the globe according to a geographically defined pattern.

To further interpret the dimensions that underlie the space in Figure 2, we conducted a series of phonetic analyses of the 17 language samples and correlated these parameters with each of the MDS dimensions. These parameters included sample duration, number of syllables in sample, number of segments in sample, speech rate (syllables/second), number and duration of vocalic portions (#V), number and duration of consonantal portions (#C), %V, %C, Std. Dev. V, Std. Dev. C, F0 mean, min, max, range, maximum number of consonants in a row, number of miscellaneous notable segments (/x/, /R/ etc). The only parameter that showed some relationship to any dimension was the last parameter listed, which related to the presence or absence of "notable" segments (typically marked dorsals).

While we do not yet have a clear notion of the physical dimensions that underlie this similarity space of languages, we can locate English in the space shown in Figure 2. Specifically, as a western language without marked dorsals, English should be located towards the bottom left corner of the language space. This then sets up some predictions regarding the perceptual distance from English of each of the 17 languages. If Dimension 1 is the most salient dimension, then Galician, Catalan, Croatian and Hausa should be judged to sound most similar to English and Arabic, Hebrew and Dutch should be most different from English. If Dimension 2 is the most salient dimension, then Swedish should be judged to sound most similar to English and Cantonese should be most different from English. If the two dimensions combine perceptually then distance from English may be best represented by distance along the diagonal, in which case Hungarian should be closest and Sindhi and Arabic should be farthest from English. Experiment 2 tested these predictions.

3. EXPERIMENT 2

3.1 Method

The stimuli were the same as Experiment 1. 17 native American English listeners (age range 18-22 years) were recruited from the same population as Experiment 1. The task for Experiment 2 was similar to the free classification task (Experiment 1) except the display was a "ladder" instead of a grid (a series of rows in just one column) with the word "English" on the bottom "rung" of the ladder. The listeners were instructed to "rank the languages according to their distance from English." Subjects could put more than one language on the same "rung" if they thought they were equally different from English. Following this "ladder task", subjects performed the same language identification task as in Experiment 1.

3.2 Results

The post-test questionnaire confirmed that the subjects were generally unable to identify the languages. On the 3 point scale (0 = incorrect or

blank, 1 = correct geographical region or family, 2 = correct language), average scores ranged from 0.04 (Hausa) to 1.3 (Arabic).

Table 1 shows the mean distances from English. The proximity of Dutch to English is expected based on known genetic and structural similarities; however, some unexpected ratings emerged in this task too. For example, Croatian was judged to be about as close to English as Hausa and Turkish, and much closer to English than Cantonese or Arabic.

Table 1: Mean distances from English (standard deviations are given in parentheses).

Language	Mean dist. from English	Language	Mean dist. from English
Dutch	3.7 (3.1)	Korean	9.1 (4.5)
Galician	4.2 (2.7)	Persian	9.7 (3.1)
Catalan	4.4 (3.0)	Hebrew	10.6 (3.4)
Swedish	5.3 (3.3)	Japanese	10.7 (4.1)
Hausa	6.2 (3.1)	Amharic	11.6 (3.9)
Croatian	6.8 (3.1)	Arabic	12.5 (3.1)
Turkish	7.2 (2.6)	Sindhi	12.6 (3.6)
Hungarian	7.4 (2.9)	Cantonese	14.4 (2.7)
Slovene	8.2 (4.0)		

The distances from English were then correlated with the ordering of languages along Dimension 1 (horizontal), Dimension 2 (vertical) and the diagonal (bottom left to top right) in the MDS solution. Table 2 shows the rank order (Spearman rho) and parametric (Pearson R) correlation coefficients. These correlations indicate that Experiments 1 and 2 converge in establishing Dimension 2 (and possibly the diagonal) in the MDS solution as a salient dimension of sound similarity for a diverse set of languages.

Table 2: MDS and ladder correlations.

Correlation with distance on ladder	Rank Order (Spearman)	Parameters (Pearson)
Dim. 1: marked dorsals	0.41	0.38
Dim. 2: "East-West"	0.80	0.79
Diagonal	0.66	0.74

CONCLUSIONS AND DISCUSSION

The overall goal of this study was to devise a means of representing natural languages in a perceptual similarity space. This means of classifying languages could then be used to predict generalized spoken language intelligibility between speakers of various languages when communicating in the native language of one of the talkers (intelligibility of foreign-accented speech for native listeners of the target language) or when communicating via a third language.

Based on the data from the present study we can develop and test predictions such as the following: Cantonese-accented Sindhi and Sindhi-accented Cantonese should be relatively intelligible to native speakers of Sindhi and Cantonese, respectively. (Experiment 1); Native speakers of English should find Hausa-accented English easier to understand than Cantonese-accented English. (Experiment 2); Cantonese-accented English and Sindhi-accented English should be relatively intelligible to native speakers of Sindhi and Cantonese, respectively even though they may both be quite difficult for native English listeners to understand (Experiments 1 and 2).

The present study has established the general feasibility of language classification based on perceptual similarity; however, it has several limitations. First, future studies should include more languages and more samples per language so that both inter- and intra-language comparisons can be assessed. Second, the physical dimensions that underlie the perceptual dimensions of the MDS solution have not yet been adequately identified. Additional analyses should include source characteristics other than pitch and various other sub- and supra-segmental features. Finally, listener characteristics could be varied to investigate how perceptual salience interacts with experience-dependent learning. For example, while speakers of a language without marked dorsal consonants may find the presence of such consonants highly salient, native listeners of a language with these consonants may not find their absence salient at all.

ACKNOWLEDGEMENTS

We are grateful to Rachel Baker and Arim Choi for research assistance. This work was supported by NIH grants F32 DC007237 and R01 DC005794.

REFERENCES

[1] Bent, T, Bradlow, AR 2003. The interlanguage speech intelligibility benefit. *J. Acoust. Soc. Am.*, 114, 1600-1610.

[2] Best, CT, McRoberts, GW, Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoust. Soc. Am.*, 109, 775-794.

[3] Clopper, CG., Pisoni, DB. In press. Free classification of regional dialects of American English. *J. Phonetics.*

[4] Flege, JE 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistics experience: Issues in cross-language research.* Baltimore, MD, York Press, 233-277.

[5] Handbook of The International Phonetic Association. 1999. Cambridge, UK: Cambridge University Press.