



Published in final edited form as:

*Speech Commun.* 2010 ; 52(11-12): 930–942. doi:10.1016/j.specom.2010.06.003.

## A Perceptual Phonetic Similarity Space for Languages: Evidence from Five Native Language Listener Groups<sup>1</sup>

Ann Bradlow<sup>a</sup>, Cynthia Clopper<sup>b</sup>, Rajka Smiljanic<sup>c</sup>, and Mary Ann Walter<sup>d</sup>

<sup>a</sup> Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208, USA

<sup>b</sup> Department of Linguistics, The Ohio State University, USA

<sup>c</sup> Department of Linguistics, University of Texas at Austin, USA

<sup>d</sup> Middle East Technical University, Turkish Republic of Northern Cyprus

### Abstract

The goal of the present study was to devise a means of representing languages in a perceptual similarity space based on their overall phonetic similarity. In Experiment 1, native English listeners performed a free classification task in which they grouped 17 diverse languages based on their perceived phonetic 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. Experiment 3 repeated Experiment 2 with four groups of non-native English listeners: Dutch, Mandarin, Turkish and Korean listeners. Taken together, the results of these three experiments represent a step towards establishing an approach to assessing the overall phonetic similarity of languages. This approach could potentially provide the basis for developing predictions regarding foreign-accented speech intelligibility for various listener groups, and regarding speech perception accuracy in the context of background noise in various languages.

<sup>1</sup>Earlier versions of Experiments 1 and 2 of this study were presented at the XVI<sup>th</sup> International Congress of Phonetic Sciences, Saarbrücken, Germany and appear in the proceedings (Bradlow, Clopper and Smiljanic, 2007)

Correspondence: Ann R. Bradlow, Department of Linguistics, Northwestern University, 2016 Sheridan Rd., Evanston, IL 60208 USA, [abradlow@northwestern.edu](mailto:abradlow@northwestern.edu), 847-491-8054 (phone), 847-491-3770 (fax).

<sup>1</sup>ALSCAL multidimensional scaling analyses take a single matrix summed over all participants and return a multidimensional space for the objects. The resultant space can be rotated to find the best interpretation, however it does not have to be rotated for interpretation. In contrast, INDSCAL (individual differences scaling) multidimensional analyses take individual matrices from each participant and return a multidimensional space for the objects (here, languages) and dimension weights for each participant. An INDSCAL analysis cannot be rotated and must be interpreted with respect to the dimensions that are returned.

<sup>2</sup>Unlike with more commonly used hierarchical clustering methods, the terminal nodes (branches) of additive similarity trees are not equidistant from the root of the tree. This feature is desirable for capturing overall similarity in the space (because some objects can be more different from the root than others), but it makes interpretation of the location along the x-axis of the branching nodes less informative.

<sup>3</sup>The different numbers of participants in each group was due to recruitment constraints. A statistical consequence of this variation is that the means (e.g. in Table 2) from the smaller groups may be less stable (noisier) than the means from the larger groups.

<sup>5</sup>Following Derwing and Munro (1997) and Munro and Derwing (1995), we use the term “intelligibility” to refer to objective speech recognition accuracy as assessed by word or sentence transcription accuracy (in terms of percent correct recognition). This is in contrast to subjective judgments of “comprehensibility” and “accentedness,” which refer to assessments along a rating scale of a talker’s understandability and degree of foreign-accent, respectively.

**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.

## Keywords

phonetic similarity; cross-language speech intelligibility; language classification

---

## 1. Introduction

Decades of research on cross-language and second-language phonetics and phonology have established the tight connection between native language sound structure and patterns of non-native language speech production and perception. Moreover, models and theories of non-native speech production and perception (Flege, 1995; Best et al., 2001; Kuhl et al., 2008) have provided principled accounts and specific predictions regarding the relative ease or difficulty of perception and production that various sound contrasts present for various native language (L1) and target language (L2) pairings. A recurrent theme in this large literature is the important role played by phonetic distance or similarity between languages in determining the observed patterns of cross-language and second-language speech perception and production. Phonetic and phonological similarity is typically not formally defined in this literature and often is limited to segmental similarity. However, the overall phonetic similarity of two languages will be driven not only by their segmental properties, but also by their prosodic properties, including metrical and intonational structures, and their phonotactic properties, including permissible segment combinations and syllable shapes. Thus, a major remaining challenge is how exactly to determine the overall phonetic and phonological likeness of any two languages such that inter-language distances can be adequately quantified (e.g. see Strange, 2007 regarding issues related to assessing distances between vowel systems).

The challenge of assessing phonetic distance becomes all the more pressing when we consider overall speech intelligibility between native and non-native speakers of a target language because it requires us to consider multiple acoustic-phonetic dimensions over multiple linguistic units (e.g. segments, syllables, and phrases). That is, for the purpose of identifying mechanisms that underlie overall speech intelligibility between native and non-native talkers of a target language we need a multi-dimensional space that incorporates native-to-nonnative (mis)matchings across the sub-segmental, segmental and supra-segmental levels. This daunting challenge is important because the availability of a multidimensional similarity space for languages would allow us to generate predictions about the mutual intelligibility of various foreign accents. To the extent that the critical dimensions of this space can be specified, we would then be able to identify the specific features of various foreign accents that are most likely to cause intelligibility problems for native listeners and that could potentially be the most beneficial targets of automatic enhancement techniques or speech training. Accordingly, the present study presents a perceptual similarity approach to language classification as a step towards the goal of deriving perceptually motivated predictions regarding variability in cross-language speech intelligibility.

The typical approach to language classification appeals to specific structural features of the relevant languages' sound systems such as the phoneme inventory, the phonotactics (e.g. preferred syllable shapes), rhythmic structures (e.g. stress-, mora- or syllable-timed) or prosodic structure (e.g. lexical tones, predictable word stress patterns etc.). For example, Dunn et al. (2005) applied biological cladistic methods (i.e. a system of classification based on evolutionary relationships) to structural features of the sound and grammar systems of the Papua languages of Island Melanesia. This structural approach, which included 11 sound structure features (8 phoneme inventory and 3 phonotactic) and 114 grammar-based features (including features related to ordering of noun phrase elements, the nominal classification

system, the verb system etc.) led these researchers to an account of the development of these languages that extended to impressive time depths. Similarly, in an example of a dialect classification study, Heeringa, Johnson and Gooskens (2009) showed good agreement between models of Norwegian dialect distances based on acoustic features (Levenshtein distances between productions of a standard passage), traditional dialectology based on a set of predetermined linguistic features (6 sound system, 4 grammatical) and native speakers' perceptions (perceived distances between a given dialect and the listener's native dialect). The language and dialect classification findings from studies such as these are highly informative. However, for the purposes of devising a similarity space for languages from which cross-language speech intelligibility can be determined, a different set of challenges is presented. In particular, the acoustic feature-based approach of Heeringa et al. (2009) requires productions of a standardized text, and the linguistic feature-based approach of Dunn et al. (2005) does not easily take into account variation in the relative salience of the individual features in foreign-accented speech, or the effect of the listener's language background (see also Meyer, Pellegrino, Barkat-Defradas and Meunier, 2003). For example, languages A and B, that have no known genetic relationship or history of population contact, may both have predominantly CV syllables, similar phoneme inventories and a prosodic system with lexical pitch accents. Yet, these two languages may sound less similar to a naïve observer 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. 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. That is, we need a language classification system whose parameters reveal the nature and functional implications of foreign-accented speech. Rather than the discovery of language history or the dialect landscape, the overall goal of this language classification enterprise is to predict which foreign accents will most or least impede speech intelligibility in cases of various target and source languages. By classifying languages in terms of their overall perceived 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 (cf. Bent and Bradlow, 2003).

The logic behind the perceptual similarity approach is that since the range of possible sound structures is limited by the anatomy and physiology of the human vocal tract as a sound source, and of the human auditory system as a sound receptor, languages that are genetically unrelated and that may never have come into contact within a population of speakers may share specific features of their sound structure that make them "sound similar" to naïve listeners. That is, languages may converge on similar sound structure features whether or not they are derived from a common ancestor or have a history of population contact. Sound structure similarity may then play an important role in determining the mutual foreign-accent intelligibility across speakers of two apparently unrelated languages especially in the context of communication via a third, shared language that functions as a lingua franca. However, 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.

In the present study, we attempted to create a language classification space with dimensions that are based on perception rather than on a priori phonetic or phonological constructs. A similar approach has been taken by researchers interested in improving and understanding the basis for language identification by humans and computers. In particular, a series of studies by Vasilescu and colleagues (Vasilescu, Candea and Adda-Decker, 2005; Barkat and Vasilescu, 2001; and Vasilescu, Pellegrino and Hombert, 2000) investigated the abilities of

listeners from various native language backgrounds to discriminate various language pairs (with a focus on the Romance, Arabic and Afro-Asiatic language families). The data showed that the ability to discriminate languages varied depending on the listener's prior exposure to the languages, and multi dimensional scaling analyses showed that the strategies employed in the language discrimination task was dependent on the listener's native language.

In the present study, we used a perceptual free classification experimental paradigm (Clopper and Pisoni, 2007) with digital speech samples from a highly diverse set of languages. In Experiment 1, native English listeners performed a free classification task with digital speech samples of 17 languages (excluding English). The classification data were submitted to clustering and multi-dimensional scaling (MDS) analyses. The MDS analysis provided some suggestion of the dimensions along which the 17 languages were perceptually organized, which in turn suggested a relative ranking of the 17 languages in terms of their distance from English. Accordingly, in Experiment 2, a separate group of native English listeners then ranked the same samples of 17 languages from Experiment 1 in terms of their perceived difference from English. These rankings were then correlated with the distances-from-English suggested by Experiment 1. Finally, the purpose of Experiment 3 was to investigate the native-language-dependence of the observed distance-from-English judgments from Experiment 2. Accordingly, four groups of non-native English listeners performed the distance-from-English ranking task (as in Experiment 2 with the native English listeners). These non-native listeners were native speakers of Dutch, Mandarin, Turkish or Korean.

## 2. Experiment 1

### 2.1 Method

**2.1.1 Materials**—Samples of seventeen languages were selected from the downloadable digital recordings on the website of the International Phonetic Association (IPA). The selected 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 taken from translations of “The North Wind and the Sun” passage in each language. The samples were selected to be sentence-final intonational units with no intonation breaks in the middle, easily separable from the rest of the utterance, and include the non-English segments or segment combinations as listed in the language descriptions that accompany transcriptions of the recordings in the IPA Handbook (International Phonetic Association, 1999). Transcriptions of the 17 selected samples are provided in the appendix. We excluded English, the native language of the participants in this experiment, and languages that are commonly studied by students at universities in the United States of America, such as French and German. The reason for including only languages that were expected to be unfamiliar to the participants was to ensure that they performed the free classification task based on perceived phonetic similarity rather than based on signal-independent knowledge about the languages and their genetic or geographical relationships.

**2.1.2 Phonetic analysis**—A set of phonetic parameters were examined for each language sample, as shown in Table 1. Parameters 1–4 capture overall timing and speaking rate characteristics. Total sample duration (1) was simply the duration of the digital speech file. Segments (2) and syllables (3) were counted based on the phonetic transcriptions in the IPA handbook. A syllable was defined as a peak in sonority (a vowel or vowels) separated from other peaks by less sonorous segments in the transcription. A syllable in Cantonese was defined as a segment or group of segments with a single tone in the transcription. Speech rate (4) was calculated as the number of syllables (2) divided by the total duration (1) in seconds.

Parameters 5–12 are based on proposed measures of typological rhythm class and its related phonotactic features (Ramus, Nespor, Mehler, 1999). For parameters 5, 6, 9, and 11, vocalic portion was defined as a portion of the speech signal that exhibited high amplitude relative to surrounding consonantal portions, periodicity in the waveform display, and clear formant structure in the spectrogram display. For parameters 7, 8, 10 and 12, consonantal portion was defined as a portion of the speech signal that was clearly aperiodic. With these definitions, all vowels were marked as vocalic intervals, and obstruents and sonorants that could be clearly separated from neighboring sounds were marked as consonantal portions. In the case of a sonorant-vowel or vowel-sonorant sequence where no clear boundary was observed (i.e. there was no clear amplitude difference or change in formant structure), the sonorant was included in the vocalic portion. A contiguous sequence of one or more vowels (or consonants) was marked as one vocalic (or consonantal) portion even when it spanned syllable and word boundaries.

Parameters 13–16 focused on F0, which was extracted from the digital speech file using the standard pitch extraction function in Praat (Boersma and Weenink, 2009). Finally, parameters 17–20 capture other features in the samples at the segmental and suprasegmental levels. All were based on information provided in the IPA Handbook transcription and accompanying text. (It is important to note that these 20 parameters do not capture all possible acoustic-phonetic features of the speech samples. A full acoustic analysis of the samples, including both static and dynamic features, at the sub-segmental, segmental and suprasegmental levels remains to be done.)

**2.1.3 Participants**—Twenty-five native speakers of American English participated in Experiment 1. They were recruited from the Northwestern University Linguistics Department subject pool (age range 17–30 years, 16 females, 9 males). All participants received course credit for their participation. None reported any speech or hearing impairment at the time of testing.

**2.1.4 Free classification procedure**—The procedure for this experiment followed that of Clopper and Pisoni (2007) very closely. Participants were seated in individual sound treated booths in front of a computer. In the center of the screen was a 16×16 grid. On the left of the screen were 17 rectangles (the “language icons”) with arbitrary labels, e.g. BB, VV etc. Double-clicking on one of these language icons caused the speech sample for that language to be played out over headphones. The participants were asked to “group the languages by their sound similarity,” and were reminded that languages of the world can differ in terms of phonetic inventory, phonotactics, and prosody. They were instructed 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. Participants 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. Subjects were always given the opportunity to ask questions about this task if they found the instructions unclear or confusing. Since no subject expressed any doubt about the task we believe that they were able to perform it adequately. Typically, the language classification task took approximately 10–15 minutes.

Following the free classification task described above, the participants 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 verify that the participants based their judgments on phonetic similarity rather signal-independent knowledge.



**2.1.5 Free classification data analysis**—Data from the free classification task were submitted to three separate analyses (see Clopper, 2008 for more details). 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, additive similarity tree clustering (Corter, 1982) and ALSCAL multidimensional scaling (Takane, Young, & de Leeuw, 1977; Kruskal & Wish, 1978) 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 the data set (in this case, languages). At each iteration, the two most similar objects or clusters of objects from a previous iteration are joined, and the similarity matrix is recalculated with each cluster treated as a single object. The resulting structure is represented visually as a tree that reveals clusters of similar objects. The multidimensional scaling (MDS) analysis fits the entire similarity matrix to a model with a specified number of orthogonal dimensions. At each iteration, the function relating model distances to data distances is made increasingly monotonic. The resulting MDS representation facilitates the identification of the physical dimensions that underlie the perceptual similarity space. The two models (clustering and MDS) may produce somewhat different visual representations of the perceptual similarity space, given the different methods of model-fitting and different visual representations of distance.

## 2.2 Results

The language identification 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 language scores across all listeners ranged from 0.08 for Hausa to 1.16 for Cantonese and Hebrew, with a mean and median of 0.67 and 0.72, respectively. Average listener scores across all languages ranged from 0.18 to 1.0 with mean and median scores of 0.67 and 0.71, respectively. Thus, none of the languages was easily identified and none of the listeners was particularly familiar with the languages.

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

The clustering solution reveals three main clusters: the top cluster including Amharic, Arabic, Hebrew, Persian, Cantonese, Sindhi, and Japanese; the middle cluster including Croatian, Catalan, Galician, Hausa, and Korean; and the bottom cluster including Dutch, Swedish, Hungarian, Turkish, and Slovene. The top cluster is further divided into two sub-clusters, separating the Semitic languages and Persian from Cantonese, Japanese, and Sindhi. The middle cluster is also further divided into two sub-clusters separating Hausa and Korean from the Indo-European (Slavic and Romance) languages.

The circles and squares indicate the six language pairs that were judged to sound the most similar to each other. Of these six pairs, the three 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 grouping of these three pairs 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 three pairs (circles), Cantonese-Sindhi, Hausa-Korean, and Hungarian-Turkish do not represent languages with known genetic relationships, and thus suggest that the free classification task was sensitive to their “genetically unexpected” sound similarity.

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) and was interpretable without rotation. The three clusters that were observed in the additive similarity tree analysis (Figure 1) are also visible in the MDS solution. The languages in the top cluster are all located in the top right quadrant of the MDS solution, except for Japanese which is in the top left quadrant. The languages in the middle cluster are all located in the left central section of the MDS solution, and the languages in the bottom cluster are all located at the bottom of the MDS solution. Thus, both visual representations of the perceptual similarity structure of these languages revealed perceptual groups of genetically similar languages, such as the Semitic or Romance languages, as well as perceptual groups of genetically more distant languages, such as Cantonese and Sindhi, or Hungarian, Turkish, and Slovene.

Our interpretation of the dimensions that underlie the space in Figure 2 was based on the series of phonetic analyses of the 17 language samples shown in Table 1. For the parameters with continuous scales (i.e. parameters 1–17), we examined the correlation between the values along the parameter scale and the values along each of the MDS dimensions. None of these correlations turned out to be significant. For the categorical scales (i.e. parameters 18–20), the examination was qualitative rather than quantitative (do the MDS dimensions separate the languages in terms of the parameter categories?). Dimension 1 (horizontal) in the MDS space shown in Figure 2 appeared to sort languages according to the presence or absence of marked back segments in the sample (defined as dorsal consonants other than /g/, /k/, or /ŋ/). In particular, the languages on the right hand side of the space are languages with more than just /k/, /g/ or /ŋ/ in their sample, including /x/ and other sounds produced further back in the vocal tract. Dimension 2 (vertical) showed a tendency to divide languages into those with marked front vowel rounding (bottom quadrants) versus those with more typical unrounded front vowels (top quadrants) in the sample. However, this interpretation of Dimension 2 does not account for the placement of Cantonese and Slovene. The Cantonese sample included one front rounded vowel, yet Cantonese appears at the top end of Dimension 2 in Figure 2 (i.e. the “typical front vowel rounding” end of the dimension). The Slovene sample did not include front rounded vowels, yet it appears at the bottom end of Dimension 2 in Figure 2 (i.e. the “marked front vowel rounding” end of the dimension). An alternative interpretation of Dimension 2 (vertical) is that it divides 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. An examination of rotations of the space in five degree increments did not reveal any other dimensions that were more interpretable or strongly correlated with any of the other phonetic properties in Table 1.

The interpretation of the MDS dimensions in terms of marked back consonants, marked front rounded vowels, and geography is also consistent with the interpretation of the clustering solution as revealing three main clusters. The top cluster includes eastern languages with marked back consonants and without front rounded vowels. The bottom

cluster includes western languages with marked back consonants and front rounded vowels. The middle cluster includes languages without marked back consonants.

While the second dimension of this space is difficult to uniquely interpret, we can provisionally and speculatively locate English in the space shown in Figure 2. Specifically, as a western language without “marked back segments” (which we define rather broadly here as back segments other than /k/, /g/ or /ŋ/), 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 marked back consonants are more salient for judgments of language similarity, 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 (despite the genetic and structural similarities of Dutch and English). If front rounded vowels are more salient for judgments of language similarity, then Swedish, Dutch, and Turkish should be judged to sound most similar to English, and Cantonese and Sindhi should be most different from English. If marked back consonants and front rounded vowels combine perceptually for judgments of language similarity, then distance from English may be best represented by Euclidean distance in the space from the bottom left corner, in which case Hungarian should be closest, and Sindhi and Arabic should be farthest from English. As a way of verifying our interpretation of the MDS space in Figure 2, Experiment 2 tested these predictions using an explicit distance-from-English task.

### 3. Experiment 2

#### 3.1 Method

The stimuli were identical to those of Experiment 1. A new group of twenty-three native American English listeners (age range 18–22 years, 20 females, 3 males) 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

As for Experiment 1, 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 language identification scores across all listeners ranged from 0.04 for Hausa to 1.30 for Arabic. Average scores for the listeners across all languages ranged from 0.29 to 1.0 with mean and median scores of 0.67 and 0.71, respectively. Thus, as for Experiment 1, none of the languages was easily identified and none of the listeners was particularly familiar with the languages.

Table 2 (first data column) shows the mean distances from English for each of the languages listed in alphabetical order as rated by the 23 native English listeners (see Table 4 below for a listing of the languages in order of increasing distance from English). The proximity of Dutch to English is not particularly surprising based on the known genetic and structural similarities of these two languages and suggests that the role of marked back segments (which Dutch includes but English does not) may be less salient than other features of similarity between these two languages when the task involves explicit similarity judgments to English. However, some rather 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.

The distances from English were then correlated with the ordering of languages along Dimension 1 (horizontal), Dimension 2 (vertical) and the Euclidean distance from the bottom left corner (i.e. the point with coordinates -2,-2) in the MDS solution from Experiment 1 (see Table 3 below). These correlations (Pearson product-moment) were calculated with scores on the continuous distance-from-English and MDS dimension scales. These correlations indicated that Experiments 1 and 2 converge in establishing the east-west dimension and the integrated two-dimensional space in the MDS solution as salient dimensions of sound similarity for a diverse set of languages for native English listeners..

Taken together Experiments 1 and 2 represent a first step towards establishing the general feasibility of language classification based on perceived phonetic similarity. The predictions based on the free classification task were partly confirmed by the distance-from-English task: both the east-west dimension and the distance from the bottom left corner of the MDS space shown in Figure 2 were significantly correlated with the perceived distances from English. The task demands were somewhat different in the two tasks, however, which may explain the differences that were observed, particularly with respect to the attention that was paid to marked back consonants in judging language similarity. In the free classification task, the listeners were asked to make discrete groups of languages and no language was provided as a reference. In the distance-from-English task, the listeners were asked to arrange languages along a continuum from most to least similar to English. Perceptually similar languages would be grouped together in the free classification task and would likely be ranked as equally distant from English in the distance-from-English task. Perceptually different languages would be grouped separately in the free classification task, but could potentially still be ranked as equally distant from English in the distance-from-English task. The explicit reference to English or the intrinsically uni-dimensional nature of the distance-from-English task may have encouraged the listeners to attend to a somewhat different set of properties (or to weight the same properties differently) than when they were asked to sort the languages by overall similarity without explicit reference to English.

The present data also have several important limitations. First, the present analyses are based on a very small sample of languages and each language was represented by just one brief audio sample produced by just one native talker of the language. Obviously, this extremely limited sample is not sufficient for drawing strong conclusions about the perceived sound similarity of the languages. In a study of language discrimination that avoided a confound of talker and language by having pairs of languages produced by bilingual speakers, Stockmal, Moates and Bond (2000) report that listeners were able to abstract from talker to language characteristics. Thus, in a future, “scaled-up” version of the present study that includes multiple talkers from various languages, we may be able to assess both intra- and inter-language sound similarity, thereby making it possible to confidently identify the physical dimensions that underlie the perceptual dimensions of overall language sound structure similarity as distinct from individual talker distances.

A second major limitation of the data from Experiments 1 and 2 is that they are based on classification patterns from native English listeners. It is likely that perceptual salience interacts with experience- dependent learning in such a way that rather different classification patterns would be observed in data from listeners from different native language backgrounds (cf. Vasilescu, Candea and Adda-Decker, 2005; Barkat and Vasilescu, 2001; and Vasilescu, Pellegrino and Hombert, 2000; Meyer, Pellegrino, Barkat-Defradas and Meunier, 2003). For example, while speakers of a language without marked back 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, resulting in a rather different perceptual similarity space. In order to address this issue of language-background dependence, we conducted Experiment 3 in which listeners from different native language backgrounds performed the ladder task (ranking of languages in terms of the overall phonetic distance from English as in Experiment 2). Recall that the free classification task was designed to exclude English and other languages that would be highly familiar to American students to make the stimulus materials equally unfamiliar. The task would be fundamentally different if the native language of the listener were included, and results would not be comparable across listener groups from different native language backgrounds. Thus, we present only distance-from-English data from the non-English listeners due to the fact that for each of the four groups of non-native English listeners, their native language or a highly familiar, related language was included in the set of 17 language samples. Given that English was included as the common baseline for judgments of language distance in the ladder task, the inclusion of the native languages of non-English listeners should be less disruptive in this task than in the free classification task.

## 4. Experiment 3

### 4.1 Method

The stimuli and procedure were identical to those of Experiment 2. As in the experiments with the native listeners, three of the four groups of non-native listeners (Dutch, Mandarin and Korean) ended the test session by performing the language identification questionnaire. The Turkish listeners did not perform this language identification questionnaire due to practical constraints on the collection of this data set. Nevertheless, at the conclusion of the data collection sessions, the Turkish listeners were asked informally whether they could identify any of the languages.

Four groups of listeners participated: 20 native speakers of Dutch (17 females, 3 males, age range 19–27 years), 16 native speakers of Mandarin (8 males, 8 females, age range 22–28 years), 13 native speakers of Turkish (8 females, 5 males, age range 20–26 years), and 6 native speakers of Korean (4 females, 2 males, age range 23–30). The Mandarin and Korean participants were recruited from Northwestern University's International Summer Institute 2007 (ISI 2007) and tested in the Phonetics Laboratory in the Department of Linguistics (same as for the native English speaking participants in Experiments 1 and 2 above). ISI is an annual, month-long program that provides intensive English language training and an introduction to academic life in the USA to incoming graduate students from across Northwestern University. The institute takes place in the month before the start of the academic year. ISI participants have all been admitted into graduate programs at Northwestern and are nominated by their departments for participation in the program. All ISI participants have met the university's English language requirement for admission (TOEFL score of 600 or higher on the paper-based test, 250 or higher on the computer-based test, 100 or higher on the internet based test), but have no or very limited experience functioning in an English language environment. At the time of testing for this study, the ISI participants had been living in the USA for less than 1 month. Both the Turkish and Dutch groups were tested on the same type of equipment (computers and headphones) as used in the Northwestern University Phonetics Laboratories. The Dutch participants were all students in Nijmegen, Netherlands and were tested at the Max Planck Institute for Psycholinguistics. As is typical for Dutch university students, their proficiency in English was relatively high. The Turkish participants were all tested at the Middle East Technical University, in the Turkish Republic of Northern Cyprus. These participants were all students at the university, with a relatively high level of English proficiency as evidenced by the English language medium of the university. All non-native listeners were paid for participating.

### 3.2 Results

The post-test questionnaire showed that the Dutch listeners were all able to identify their own native language, and as a group they were moderately accurate at identifying Arabic and Cantonese. On the 3 point scale (0 = incorrect or blank, 1 = correct geographical region or family, 2 = correct language), the average language identification scores for Arabic and Cantonese by the Dutch listeners were 1.3 and 1.4, respectively. The Mandarin listeners were generally able to identify two of the 17 languages, namely Cantonese and Japanese. On the 3 point scale, the average language identification scores for Cantonese and Japanese by the Mandarin listeners were 1.5 and 1.8, respectively. The Korean listeners were all able to identify their own native language and Japanese, and were also generally able to identify Cantonese (average score of 1.3). The Turkish listeners did not perform the language identification questionnaire, however post-test informal debriefing about strategies used and languages recognized indicated that they all recognized Turkish (their own language) amongst the set of 17 languages. Additionally, one Turkish listener mentioned identifying Arabic and Chinese in the set of languages, and one other mentioned Japanese. Thus, as expected, for the non-native listeners the set of 17 languages included some highly familiar languages.

Included in Table 2 above are the mean distances from English as rated by the Dutch, Mandarin, Turkish and Korean listeners. Table 4 below shows the distance-from-English “ladders” for all four listener groups. Table 5 shows the distance-from-English correlations (Pearson product moment with scores on the continuous rung scale) between the four listener groups. Also, shown in Table 5 are the average rank differences across each of the pairs of listener groups. These average rank differences were calculated from absolute values of the differences in rankings across listener group pairs for each of the 17 languages. For example, an average rank difference score of 2.3 rungs for the Mandarin and English listeners indicates that on average the English listener distance-from-English rankings for each of the 17 languages was 2.3 rungs from the Mandarin distance-from-English rankings<sup>4</sup>.

As shown in Table 5, the rankings from the Dutch listeners were highly correlated with the English, Mandarin and Korean listeners’ rankings. Similarly, the Mandarin and Korean listeners’ rankings were highly correlated. The English and Mandarin listeners’ ranking were also significantly correlated, though less strongly than the other correlated pairs. Parallel to these correlation coefficients, the average English-Dutch, Dutch-Mandarin, and Mandarin-Korean rank differences were relatively small (less than an average of 2 rungs apart).

The overall picture from the correlational analysis in Table 5 is one in which the distance-from-English ladders from the five groups of listeners are generally significantly positively correlated with each other. The language ladders from the five different groups of listeners showed some generally consistent trends. For example, all listener groups placed Swedish and Dutch within the first five rungs (i.e. quite close to English) and Cantonese within the last five rungs (quite far from English). Two pairs of languages, Galician-Catalan and Hebrew-Japanese, were consistently ranked close to each other on each of the five ladders shown in Table 4. Overall, amongst the four groups of non-native English listeners, three produced distance-from-English “ladders” that were highly correlated ( $p < .002$ ) with each other (the Dutch, Mandarin and Korean listeners), and two of these non-native listener groups produced distance-from-English ladders that were highly significantly correlated with the native English listeners’ ladder (the Dutch and Mandarin listeners). The Turkish

<sup>4</sup>Please note that the numbers in Table 2 are given with precision up to 1 decimal point. The average rank distances provided in Table 5 were calculated with more precise numbers. Thus, there may be some small degree of “rounding error” between averages calculated on the basis of the numbers in Table 2 and the numbers provided in Table 5.

group's distance-from-English ladder was significantly correlated with that of three of the other groups (English, Dutch, and Korean) though less strongly than the other correlations. The one exception to this pattern is the correlation between the Turkish and Mandarin distance-from-English ladders which failed to reach significance at the  $p < .05$  level. This relatively consistent pattern of distance-from-English rankings suggests that native language background may play a rather limited role in the perception of overall phonetic similarity of languages, at least in a one-dimensional classification task that focuses attention on perceived distance from an "anchor" (in this case, English).

## 5. General 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 type of data from the present study we could potentially 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). And, Turkish speakers may find Slovenian-accented English easier to understand than Hungarian-accented English, whereas Mandarin and Korean speakers may find Hungarian-accented English easier to understand than Slovenian-accented English (Experiment 3). As discussed in detail in Derwing and Munro (1997) and Munro and Derwing (1995), there are in fact three related but independent dimensions of language distance that are likely to influence speech communication: objective intelligibility, subjective comprehensibility, and degree of perceived foreign accent. It is possible that with a well-defined perceptual similarity space for languages, we could develop a principled analysis of all of these parameters of language similarity.

While these specific predictions are suggested by the present data, it is important that the present study be viewed more as a demonstration of an approach than as the presentation of a well-founded, actual perceptual similarity space for languages. The limitations of the present data are substantial, particularly with respect to the constraints imposed by the limited set of languages included and by the single speech sample per language that was presented to the listeners. Without a broader set of languages and a much more varied set of samples per language, we can conclude little about the true overall sound structure of the global system of modern languages. Nevertheless, based on the present study we can draw three general conclusions regarding the development of a perceptual similarity space for languages.

First, the free classification and "ladder" techniques hold great promise as a means of assessing the overall phonetic similarity of various languages. A major asset of these techniques is that they allow the inclusion of a large number of language samples without requiring judgments of similarity for each possible pair of language samples. Second, any perceptual similarity space for languages may have to be considered with respect to a particular listener population. However, the results of Experiment 3 suggest that listeners from a variety of native language backgrounds may display similar patterns of perception of overall phonetic similarity between languages. This suggests that there may be some sound

structure features that have general salience and that may play important roles in the distance relationships between languages regardless of the listeners' language backgrounds. Finally, while the prospect of devising a highly generalized, perceptual similarity space for languages may be highly ambitious and somewhat daunting, a "scaled-down" rather than scaled-up version of the present study may be a profitable direction to follow for many practical purposes. For example, several recent studies have demonstrated that speech-in-speech perception is more accurate when the language of the target and background speech are different than when the target and background speech are in the same language (Rhebergen et al., 2005; Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Calandruccio et al., 2008). However, it is not yet known whether this release from masking due to a target and background language mismatch is modulated by the phonetic similarity of the two languages in question. An independent measure of phonetic similarity of a relatively small set of languages would allow us to address this question, which would in turn allow us to make progress towards identifying a mechanism that may underlie the observed patterns of speech-in-speech perception. Similarly, work on speech intelligibility between non-native speakers of English has demonstrated that non-native listeners can find non-native (i.e. foreign-accented) speech as or more intelligible than native-accented speech (e.g. Bent, Bradlow, & Smith, 2008; Hayes-Harb, Smith, Bent, & Bradlow, 2008; Imai, Walley, & Flege, 2005; Munro, Derwing, & Morton, 2006; Stibbard & Lee, 2006; Bent and Bradlow, 2003; van Wijngaarden, 2001; van Wijngaarden, Steeneken, & Houtgast, 2002 among others). Target language proficiency appears to play an important role in this "interlanguage intelligibility benefit" (e.g. see Hayes-Harb, Smith, Bent, & Bradlow, 2008, and Smiljanic and Bradlow, 2007), nevertheless it has also been observed to extend to non-native listeners from different native language backgrounds (e.g. Bent and Bradlow, 2003). However, it remains to be determined exactly which native language backgrounds produce relatively intelligible foreign-accented English for exactly which groups of non-native English listeners. Here too, progress could potentially be made by pursuing a free-classification and/or ladder approach to assessing the overall phonetic similarity of a particular, relatively small set of languages.

## Acknowledgments

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

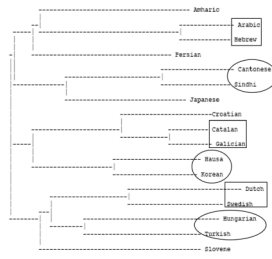
## References

- Barkat M, Vasilescu I. From perceptual designs to linguistic typology and automatic language identification: overview and perspectives. *proceedings of Eurospeech 2001*;2001:1065–1068.
- Bent T, Bradlow AR. The interlanguage speech intelligibility benefit. *J Acoust Soc Am* 2003;114:1600–1610. [PubMed: 14514213]
- Best C, McRoberts G, Goodell E. American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J Acoust Soc Am* 2001;109:775–794. [PubMed: 11248981]
- Boersma, P.; Weenink, D. Praat: doing phonetics by computer (Version 5.1.22) [Computer program]. 2009. Retrieved from <http://www.praat.org/>
- Calandruccio, L.; Van Engen, KJ.; Yuen, C.; Dhar, S.; Bradlow, AR. Assessing the clear speech benefit with competing speech maskers. *American Speech Language Hearing Association Convention*; Chicago, IL: 2008.
- Clopper CG. Auditory free classification: Methods and analysis. *Behavior Research Methods* 2008;40:575–581. [PubMed: 18522069]
- Clopper CG, Pisoni DB. Free classification of regional dialects of American English. *J Phonetics* 2007;35:421–438.

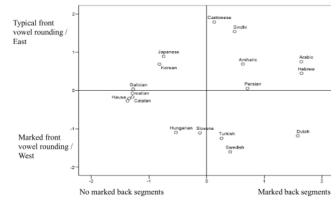


- Corter JE. ADDTREE/P: A PASCAL program for fitting additive trees based on Sattath and Tversky's ADDTREE algorithm. *Behavior Research Methods and Instrumentation* 1982;14:353–354.
- Derwing TM, Munro MJ. Accent, intelligibility and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition* 1997;20:1–16.
- Dunn M, Terrill A, Reesink G, Foley RA, Levinson SC. Structural Phylogenetics and the Reconstruction of Ancient Language History. *Science* 23 2005;309(5743):2072–2075.
- Flege, JE. Second language speech learning: Theory, findings, and problems. In: Strange, W., editor. *Speech perception and linguistics experience: Issues in cross-language research*. Baltimore, MD: York Press; 1995. p. 233-277.
- Garcia Lecumberri ML, Cooke M. Effect of masker type on native and non-native consonant perception in noise. *J Acoust Soc Am* 2006;119:2445–2454. [PubMed: 16642857]
- Handbook of The International Phonetic Association. Cambridge, UK: Cambridge University Press; 1999.
- Hayes-Harb R, Smith BL, Bent T, Bradlow AR. Production and perception of final voiced and voiceless consonants by native English and native Mandarin speakers: Implications regarding the inter-language speech intelligibility benefit. *J Phonetics* 2008;36:664–679.
- Heeringa W, Johnson K, Gooskens C. Measuring Norwegian dialect distances using acoustic features. *Speech Communication* 2009;51:167–183.
- Imai S, Walley AC, Flege JE. Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *J Acoust Soc Am* 2005;117:896–907. [PubMed: 15759709]
- Kruskal, JB.; Wish, M. *Multidimensional Scaling*. Newbury Park, CA: Sage; 1978.
- Kuhl PK, Conboy BT, Coffey-Corina S, Padden D, Rivera-Gaxiola M, Nelson T. Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B* 2008;363:979–1000.
- Meyer J, Pellegrino F, Barkat M, Meunier F. The notion of perceptual distance: the case of Afroasiatic languages. *proceedings of International Congress of Phonetic Sciences*. 2003
- Munro MJ, Derwing TM. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 1995;45:73–97.
- Munro MJ, Derwing TM, Morton SL. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 2006;28:111–131.
- Ramus F, Nespore M, Mehler J. Correlates of linguistic rhythm in the speech signal. *Cognition* 1999;72:265–292. [PubMed: 10585517]
- Rhebergen KS, Versfeld NJ, Dreschler WA. Release from informational masking by time reversal of native and non-native interfering speech. *J Acoust Soc Am* 2005;118:1–4.
- Smiljanic, R.; Bradlow, AR. Clear speech intelligibility: Listener and talker effects. *Proceedings of the XVIth International Congress of Phonetic Sciences*; Saarbrücken, Germany. 2007.
- Stibbard RM, Lee JI. Evidence against the mismatched interlanguage intelligibility benefit hypothesis. *J Acoust Soc Am* 2006;120:433–442. [PubMed: 16875239]
- Stockmal V, Moates DR, Bond ZS. Same talker, different language. *Applied Psycholinguistics* 2000;21:383–393.
- Strange, W. Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In: Bohn, O-S.; Munro, MJ., editors. *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam and Philadelphia: John Benjamins; 2007. p. 35-55.
- Takane Y, Young FW, de Leeuw J. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika* 1977;42:7–67.
- Van Engen KJ, Bradlow AR. Sentence recognition in native- and foreign-language multi-talker background noise. *J Acoust Soc Am* 2007;121(1):519–526. [PubMed: 17297805]
- van Wijngaarden SJ. Intelligibility of native and non-native Dutch speech. *Speech Communication* 2001;35:103–113.
- van Wijngaarden SJ, Steeneken HJM, Houtgast T. Quantifying the intelligibility of speech in noise for non-native listeners. *J Acoust Soc Am* 2002;111:1906–1916. [PubMed: 12002873]

- Vasilescu I, Pellegrino F, Hombert J-M. Perceptual features for the identification of Romance languages. *proceedings of the International Congress of Spoken Language Processing 2000*;2:543–546.
- Vasilescu I, Candea M, Adda-Decker M. Perceptual salience of language-specific acoustic differences in autonomous fillers across eight languages. *proceedings of Interspeech 2005*:1773–1776.



**Figure 1.**  
Experiment 1 clustering analysis.



**Figure 2.**  
Experiment 1 multi-dimensional scaling solution.

**Table 1**

List of phonetic parameters examined in each of the 17 language samples.

---

1	Total sample duration in seconds
2	Number of syllables in sample
3	Number of segments in sample
4	Speech rate (syllables/second)
5	Number of vocalic portions
6	Duration of vocalic portions
7	Number of consonantal portions
8	Duration of consonantal portions
9	Percentage of sample duration devoted to vocalic portions
10	Percentage of sample duration devoted to consonantal portions
11	Standard deviation of vocalic portion durations
12	Standard deviation of consonantal portion durations
13	F0 mean
14	F0 minimum
15	F0 maximum
16	F0 range (F0 maximum – F0 minimum)
17	Maximum number of consonants in a row
18	Presence/absence of dorsal consonants aside from /g, k/, or /ŋ/ (“marked back segments”)
19	Presence/absence of front rounded vowels (“marked front vowel rounding”)
20	Presence/absence of other “notable/non-English” phonetic features (e.g. geminate consonants, bilabial fricatives, lexical tone)

---



**Table 2**

Mean distances from English (standard deviations are given in parentheses) for an alphabetical listing of the 17 languages in Experiment 2 (English listeners) and Experiment 3 (Dutch, Mandarin, Turkish, and Korean listeners).

Language	English listeners (n=23)	Dutch listeners (n=20)	Mandarin listeners (n=16)	Turkish listeners (n=13)	Korean listeners (n=6)
Amharic	11.1 (4.7)	11.9 (3.4)	11.1(4.2)	11.2(3.5)	10.0(4.8)
Arabic	13.4 (3.5)	14.0 (2.3)	12.8(3.8)	13.4(3.8)	10.5(4.2)
Cantonese	14.8 (2.4)	15.5 (2.2)	15.3(2.6)	9.5(3.9)	12.5(1.9)
Catalan	4.4 (3.4)	8.2 (3.1)	8.0(3.0)	6.8(3.5)	7.0(3.9)
Croatian	6.3 (2.8)	9.9 (3.3)	10.3(2.5)	6.5(3.0)	9.2(4.0)
Dutch	4.2 (3.5)	5.1 (3.5)	6.9(3.6)	4.5(3.9)	4.8(2.5)
Galician	5.2 (3.2)	7.1 (2.8)	6.8(3.3)	7.3(2.6)	7.7(3.7)
Hausa	7.0 (3.1)	9.6 (3.1)	11.7(3.9)	7.8(3.9)	10.7(3.5)
Hebrew	10.6 (3.8)	13.2 (2.2)	13.4(3.4)	6.0(3.6)	11.3(3.6)
Hungarian	8.6 (3.3)	9.4 (4.2)	5.2(4.2)	8.1(4.2)	6.2(4.7)
Japanese	9.0 (3.8)	12.4 (3.9)	14.3(2.4)	6.0(4.8)	13.7(1.8)
Korean	8.9 (4.1)	10.0 (4.6)	13.3(3.4)	10.0(5.0)	13.7(2.1)
Persian	9.7 (3.1)	11.9 (3.6)	8.6(3.9)	10.2(4.0)	8.8(3.4)
Sindhi	11.9 (4.4)	11.6 (4.2)	10.7(3.7)	7.2(4.4)	9.0(5.0)
Slovene	9.1 (4.5)	8.6 (4.9)	8.0(4.6)	3.2(2.5)	7.2(3.8)
Swedish	6.5 (3.3)	7.9 (3.9)	5.3(3.2)	3.9(3.6)	3.8(2.3)
Turkish	7.9 (3.4)	11.0 (3.0)	6.3(2.8)	6.7(3.8)	6.7(3.7)

**Table 3**

MDS (Experiment 1) and ladder (Experiment 2) correlations.

<b>Correlation with distance on ladder</b>	<b>Parameters (Pearson product-moment correlation coefficient with scores on the continuous scales)</b>
Dim. 1: marked dorsals	0.44
Dim. 2: "East-West"	0.69 **
Euclidian distance from -2, -2	0.72 **

\*\* indicates significance at the  $p < .001$  level.

**Table 4**

The distance-from-English “ladders” from all four listener groups in Experiments 2 and 3. Ladder rung “1” is closest to English, “17” is farthest from English.

Ladder rung	Listener Group						
	English	Dutch	Mandarin	Turkish	Korean	English	Korean
1	Dutch	Dutch	Hungarian	Slovene	Slovene	Swedish	Swedish
2	Catalan	Galician	Swedish	Swedish	Swedish	Dutch	Dutch
3	Galician	Swedish	Turkish	Dutch	Dutch	Hungarian	Hungarian
4	Croatian	Catalan	Galician	Hebrew	Hebrew	Turkish	Turkish
5	Swedish	Slovene	Dutch	Japanese	Japanese	Catalan	Catalan
6	Hausa	Hungarian	Catalan	Croatian	Croatian	Slovene	Slovene
7	Turkish	Hausa	Slovene	Turkish	Turkish	Galician	Galician
8	Hungarian	Croatian	Persian	Catalan	Catalan	Persian	Persian
9	Korean	Korean	Croatian	Sindhi	Sindhi	Sindhi	Sindhi
10	Japanese	Turkish	Sindhi	Galician	Galician	Croatian	Croatian
11	Slovene	Sindhi	Amharic	Hausa	Hausa	Amharic	Amharic
12	Persian	Amharic	Hausa	Hungarian	Hungarian	Arabic	Arabic
13	Hebrew	Persian	Arabic	Cantonese	Cantonese	Hausa	Hausa
14	Amharic	Japanese	Korean	Korean	Korean	Hebrew	Hebrew
15	Sindhi	Hebrew	Hebrew	Persian	Persian	Cantonese	Cantonese
16	Arabic	Arabic	Japanese	Amharic	Amharic	Japanese	Japanese
17	Cantonese	Cantonese	Cantonese	Arabic	Arabic	Korean	Korean

**Table 5**

Distance-from-English correlations (Pearson product moment correlations with scores on the continuous rung scale) and average rank differences (in parentheses) between the 5 listener groups.

<b>Listeners</b>	<b>Dutch (n=20)</b>	<b>Turkish (n=13)</b>	<b>Mandarin (n=16)</b>	<b>Korean (n=6)</b>
English (n=23)	0.90 p<.001 (1.8 rungs)	0.57 p<.05 (2.1 rungs)	0.64 p<.01 (2.3 rungs)	0.57 p<.05 (2.4 rungs)
Dutch (n=20)		0.59 p<.05 (2.9 rungs)	0.74 p<.001 (1.7 rungs)	0.70 p<.002 (2.2 rungs)
Turkish (n=13)			0.43 p>.05 (3.1 rungs)	0.50 p<.05 (2.3 rungs)
Mandarin (n=16)				0.94 p<.001 (1.2 rungs)

## Appendix

Language	Language Sample (IPA)
Amharic	jeləbbəsə l lɨbs a əllək'ə
Arabic	xalaʕa ʕabaaʕataʕu ʕalaa ttau
Cantonese	pək1 fɔŋ1 tʰɔŋ1 tʰai1 jœŋ1 ɦei1 tou1 au1 ken1 pin1 kɔ1 lek1 ti1
Catalan	lə tɾəmuntənə j əl səl əz ðisputaβən
Croatian	i ne poʒe: naku:pa:ne u ri:je:ku tekutɔitsu
Dutch	ə nɔ:rəʊvɪnt bəχən œyt ələ mɑχ tə blɑ:zə
Galician	o βento ðo nɔ:te aβanaono o se empe:ɲo
Hausa	ja zo: sə:ɲe? ə ɾi:ɡər sə:ɲii
Hebrew	mi meʕem hazak joter
Hungarian	minel ɛθ ɛb:ɛŋ fujt
Japanese	komao a taijo: no ban ni naɾimasita
Korean	anɯl su əbsʌsɯmŋi:sa
Persian	əo re xoə əzæm? kærə
Sindhi	ɡərəm koʈo pae uta əci ləŋɡʱjo
Slovene	əa je so:mtse mət nei ə ɛ nje:ga
Swedish	ɡa nu:ɖan inəən ep fœsø:kət
Turkish	pojraz ar ɣyɔzyle esmije ba ʕa:ɯ