PERCEPTUAL SIMILARITY AMONG NASALS VARYING IN PLACE OF ARTICULATION: A MULTIDIMENSIONAL SCALING ANALYSIS

James D. Harnsberger
Speech Research Laboratory, Indiana University, USA

ABSTRACT
To examine perceptual similarity in cross-language speech perception as a multidimensional metric, an AXB classification test using nasal consonants from Malayalam (\{\textit{m}, [\textipa{n}]), \textit{n}, [\textipa{n}]\}) was administered to three sets of listener groups with common coronal nasal inventories: dental-retroflex (Marathi, Punjabi), alveolar-retroflex (Tamil, Oriya), and alveolar (Bengali, American English). A two dimensional multidimensional scaling analysis of the results showed several interesting patterns. First, cross-linguistically, the stimuli frequently clustered into three groups, \{\textit{m}\}, \{\textit{n}\}, \{\textipa{n}\}, and \{\textipa{n}\}. Second, language-specific differences were not predictable from the test groups’ nasal inventories. The dental-retroflex and alveolar groups showed intra-group differences in their clustering of stimuli and weighting of both perceptual dimensions, leading to language-specific perceptual spaces. Only the alveolar-retroflex group spaces were similarly organized. The results demonstrate that descriptions of native categories must include language-specific weighting of acoustic cues.

1. INTRODUCTION
Cross-language speech perception research has demonstrated that specific linguistic experience can limit listener sensitivity to some non-native phonemic distinctions, particularly consonantal ones [1, 9, 16]. Non-native consonant stimuli that correspond, or are similar to, a single phoneme in a listener’s native language have often been observed to be difficult for listeners to accurately discriminate and identify, such as \textipa{lb}/ for Japanese listeners [8], Hindi dental-retroflex stops for English listeners [15, 16, 18], and Salish \textipa{kl’q’} for English and Farsi listeners [17, 18]. However, non-native consonant contrasts have been shown to vary in their discriminability, from chance to near native level performance [11, 12]. Moreover, listener groups varying in native language, but sharing a similar phonemic inventory, can differ in the extent to which they find a given non-native consonant contrast difficult to discriminate. For instance, Japanese and Cantonese, and Japanese and Korean, listeners have differed in their ability to identify and/or discriminate natural American English (AE) or Australian \textipa{lb}/, despite the fact that all three languages have only one native liquid phoneme [5, 6]. Japanese and AE listeners’ perception of the four dental-retroflex stop contrasts from Hindi differed significantly, with Japanese listeners on the whole finding the contrasts easier to discriminate [12].

The varying sensitivity of listener groups to non-native, non-phonemic contrasts has led researchers to focus on alternate description of perceptual categories and models of perceptual similarity. Candidates for perceptual category descriptors, or units of analysis, include context-dependent allophones [13], phonetic features or cues [2, 3], and distributions of individual exemplars [10, 19]. Perceptual similarity is frequently assumed to be a transparent, linear mapping between a stimulus and a category that bear common, abstract descriptions, most frequently phonemic.

This study was designed to address the unit of analysis and perceptual similarity issues in cross-language speech perception. The study employed a broad range of listener groups to elicit similarity judgments for a set of non-native nasal consonants. The perceptual spaces of these listener groups were mapped from the similarity judgments via multidimensional scaling for the purposes of cross-language comparison. The study addressed the following questions: Are there patterns of perceptual similarity that transcend native language experience? Is the phonemic or allophonic level of detail sufficient to describe listeners’ perceptual categories? Do listener groups with common phonological inventories for a given non-native contrast show any differences in the weighting of acoustic dimensions that cue the contrast?

2. METHODS
2.1. Stimulus Materials
The stimulus materials were restricted to two exemplars each of six types of nasal consonants varying in place of articulation (bilabial, interdental, alveolar, retroflex, palatal, and velar) from a single speaker of Malayalam, a Dravidian language spoken in southern India. The stimuli appeared as medial geminates in an \textipa{[i]} context. Every exemplar was not matched with every other in generating trials. Instead, arbitrarily, all trials consisted of only tokens from the first set of exemplars or the second set (e.g., \{\textipa{m}\}-[\textipa{n}], \{\textipa{m}\}-[\textipa{n}], but never \{\textipa{m}\}-[\textipa{n}]). Each set of exemplars was combined, resulting in 20 different kinds of triads. Both sets of 20 triads appeared in six orders (ABC, ACB, BAC, BCA, CAB, and CBA) for a total of 240 trials. The ISI for the classification test was 1s, the ITI 5s, and the IBI 6s, with 20 trials per block (total test time: 33.5 minutes). The use of relatively few stimuli from a single talker was necessitated by the length of two other tests administered along with the AXB classification test, given the amount of time available at testing facilities in India, and the issue of subject fatigue.

2.2. Participants
Speakers of Malayalam (N=18), Marathi (N=18), Punjabi (N=14), Tamil (N=14), Oriya (N=16), Bengali (N=17), and AE (N=18) participated in this study. All but the AE listeners were tested in India, in order to recruit subjects who varied little in terms of age, dialect, and overall linguistic experience. The Malayalam listener group was recruited to serve as a control group. The six non-native listener groups served in three test groups representing different types of native nasal consonant inventories, defined specifically in terms of coronal nasal consonants: a test group with a native dental-retroflex nasal consonant contrast (Marathi, Punjabi1), an alveolar-retroflex test group (Tamil, Oriya2), and an alveolar test group (Bengali, AE).
These three test groups were chosen to compare two hypotheses, the allophonic category center and cue-weighting hypotheses. The allophonic category center hypothesis refers to the assumption that listeners’ native categories are abstract and sufficiently described by the context dependent allophonic variants of phonemes. In contrast, listener groups may individually vary in their weighting of the acoustic cues for nasal consonants. If the allophonic category center hypothesis is correct, we would expect to see internal consistency in the perceptual spaces of all three groups, and systematic differences between each group. If instead, the cue-weighting hypothesis is correct, then we would expect to observe perceptual spaces that suggest differential weighting of perceptual dimensions on a listener group by listener group basis.

2.3. Procedure
A forced-choice AXB classification test was administered, in which participants decided whether A or B was more similar to X. Unlike AXB discrimination, A, B, and X in a classification test are tokens from three rather than two place categories. Subjects were instructed to decide which nasal consonant was more similar to that of middle word, the nasal consonant of the first or the third word. Subjects were told that, while all three nasal consonants may sound quite different from one another, they were to judge which, generally speaking, was more similar to X, A or B. Prior to the test, the subjects listened to 10 randomly chosen trials to familiarize them with the test format and the kinds of stimuli being compared, with no feedback provided by the investigator.

2.3.1. Testing Conditions. All listener groups, except for AE, were tested in sound-attenuated chambers affiliated with private studios in New Delhi (Malayalam, Marathi, Tamil, Oriya) and Calcutta (Bengali) in India, except for the Punjabi participants, who were tested in a quiet room in Amritsar, India. AE subjects were tested in a sound-attenuated chamber in the Phonetics Lab at the University of Michigan. Within a single session in India, up to eight subjects were tested, with instructions provided in English by a native speaker of Indian English. At the University of Michigan, up to four subjects were tested at one time, with instructions provided in American English.

3. Predictions
Two sets of predictions were generated by the hypotheses tested in this study, allophonic category center and cue-weighting. According to the allophonic category center hypothesis, listeners’ perceptual categories possess category centers that correspond to a phoneme’s rule-governed phonetic manifestation in a given context, with context defined in terms of position within a syllable or word, any proximate conditioning vowels or consonants, or position in prosodic structure. Listeners with common allophonic category centers would map stimuli to those categories in a common manner, which in turn would affect their similarity judgments of stimuli that fall into these categories, as well as their overall similarity space for a class of speech sounds. In contrast, the cue-weighting hypothesis predicts that languages with common phoneme inventories, and even common allophonic distributions of those phonemes, can nevertheless differ in their perceptual weighting of critical cues to those phonemes or allophones, which in turn would determine their overall perceptual space for a class of speech sounds.

4. RESULTS
The results of the AXB classification test were analyzed by listener group. Similarity scores for each stimulus pair (e.g. [n]-[n], [n]-[n]) were calculated by averaging individual subjects’ proportion of responses of matching two stimuli, across all orders and all pairings with a third stimulus. For instance, the [n]-[n] similarity score for the Punjabi group was based on the number of times [n] was selected as similar to [n] when either [n] or [n] was X in an AXB triad and when either [n] or [n] was either A or B. Each listener group’s mean similarity scores were submitted separately to both a one- and two-dimensional (2D) ALSCAL multidimensional scaling analysis (MDS). In all cases, the 2D analysis provided a substantially better fit to the similarity scores than the one-dimensional analysis. Table 1 lists the stress and proportion of variation (R²) values for each listener group.

<table>
<thead>
<tr>
<th>Listener Group</th>
<th>Stress</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malayalam</td>
<td>0.07129</td>
<td>0.96907</td>
</tr>
<tr>
<td>Marathi</td>
<td>0.03193</td>
<td>0.99476</td>
</tr>
<tr>
<td>Punjabi</td>
<td>0.04794</td>
<td>0.98985</td>
</tr>
<tr>
<td>Tamil</td>
<td>0.00486</td>
<td>0.9999</td>
</tr>
<tr>
<td>Oriya</td>
<td>0.01615</td>
<td>0.99882</td>
</tr>
<tr>
<td>AE</td>
<td>0.05194</td>
<td>0.9822</td>
</tr>
<tr>
<td>Bengali</td>
<td>0.03993</td>
<td>0.99133</td>
</tr>
</tbody>
</table>

Table 1. Fits of the 2D MDS analyses to the similarity scores

The 2D MDS analyses for the control and test groups appear in Figures 1-4, with each axis representing similarity on some perceptually relevant dimension. They show a language-general pattern, as well as language-specific differences in the perceptual space that suggest listeners were weighting cues rather simply filtering the stimuli via abstract allophonic category centers. First, a broad pattern was observed across most or all of the listener groups: the stimuli formed three groups in terms of similarity, {[/m]},{[/n],[n]-[n]},{[/n]-[n]}. The bilabial stimuli were typically judged as highly dissimilar from the other stimuli in the experiment, and occupied an extreme corner of perceptual space. In contrast, the interdental, alveolar, and retroflex stimuli clustered together for five out of seven listener groups, including the dental-retroflex and alveolar-retroflex test groups. The Malayalam listener group separated the retroflex stimuli out from the interdental and alveolar (Figure 1), and perhaps surprisingly, judged the latter two stimulus types to be highly similar, though they represented a native contrast. Unlike all other listener groups, the AE subjects separated out the interdental, alveolar, and retroflex stimuli, perhaps showing a sensitivity to their distinguishing cues that was not predictable from the AE nasal consonant inventory (Figure 2, left). Finally, all listener groups placed the palatal and velar stimuli in their own region of perceptual space, with these stimuli showing less clustering relative to the [/n],[n]-[n] series. In addition to the cross-linguistic similarity groupings, listener groups with similar nasal consonant inventories at the allophonic level differed dramatically in several cases. For instance, the alveolar test group, composed of Bengali and AE listeners, differed from one another in their apparent attention to the two perceptual dimensions and in their sensitivity to low level phonetic differences between the stimuli. On one hand, the Bengali group’s space showed an attention to both perceptual dimensions only in the differentiation of the bilabial stimuli from
the other five stimulus types. The remaining stimuli varied on roughly a single dimension, represented by the line in Figure 2. Moreover, relatively tight clustering of the interdental, alveolar, and retroflex stimuli was observed in the Bengali perceptual space. In contrast, the AE listeners weighted the two dimensions differently to produce an alternate set of similarity relations and relative lack of clustering, despite the fact that AE and Bengali share a common nasal consonant inventory, even at the allophonic level, for these stimuli. AE listeners used both dimensions to not only differentiate the bilabial but also the retroflex stimuli from all other stimuli.

The dental-retroflex test group (Figure 3) also showed intra-group differences not predicted by the allophonic category center hypothesis. The Marathi listeners, like Bengali, appeared to use a single perceptual dimension in judging the [n], [ŋ], [p], [ŋ], and, to a lesser extent, [ŋ] stimuli. And like AE, Marathi listeners demonstrated relatively less clustering among the highly similar [ŋ]-[ŋ]-[ŋ] series. On the other hand, Punjabi listeners resembled Bengali, and not Marathi listeners, in tightly clustering the [ŋ]-[ŋ]-[ŋ] series, reflecting perhaps less sensitivity to low level phonetic detail that is not contrastive in Punjabi.

Of the three test groups, the alveolar-retroflex group showed the greatest similarities. Both Tamil and Oriya listeners tightly clustered the [ŋ]-[ŋ]-[ŋ] series, and both judged the non-bilabial stimuli along a single dimension, represented in Figure 4 by single lines. One possible difference between the two listener groups was the extent to which a second dimension was employed in differentiating the non-bilabial stimuli. If we consider the extent to which the stimuli deviate from the lines in Figure 4, we might conclude that Tamil judgments adhered more closely to that single dimension than those of Oriya listeners. However, minus an appropriate statistical test to compare these two spaces, only general descriptive comments are possible.

5. DISCUSSION AND CONCLUSION

The results of the two dimensional MDS analysis of the similarity scores from the classification test reveal substantial effects of linguistic experience on the organization of perceptual spaces that cannot be accounted for by referencing abstract category centers such as phonemes or allophonic variants. Listener groups, instead of being easily classified by their nasal consonant inventory, were instead defined by their weighting of the cues or dimensions that primarily signal the contrast between the non-native stimuli. Cue-weighting itself, of course, involves abstraction, reducing the rich, highly redundant, signal into a few key components. Moreover, a two-dimensional cue- or dimension- weighting model of the perceptual category could be seen as a conclusion driven by of this particular form of analysis. Undoubtedly, listeners use more than just two cues, however defined, in perceiving speech under a variety of adverse environments for communication, including visual as well as auditory information [14].

What appears to be cue weighting could also be the product of a language-specific distribution of individual exemplars of nasal consonant, instances or episodes in long-term memory. That is, language-specific differences might fall directly out of the sum total of experiences a given listener group (or a given listener) has with a given stimulus type, such as nasal consonants. An account along these lines might take the following form: Bengali and AE listeners show, for instance, a difference in their relative clustering of the [ŋ]-[ŋ]-[ŋ] series. Bengali listeners cluster these more, and their clustering may reflect their greater experience hearing dental and retroflex nasals. Bengali listeners encounter dental and retroflex nasals more frequently than AE listeners as allophonic variants of their /n/, due to place assimilation before dental and retroflex oral stops, which are much more frequent in Bengali than AE. Thus, Bengali listeners may have more experience than AE listeners in classifying dental and retroflex
nasals as alveolar, accounting for their tight clustering of non-native [n], [N], and [ŋ].

The results of this experiment cannot provide support for a model of language-specific cue-weighting over an exemplar model of speech perception. The differences between listener groups that share the same nasal consonant inventory only highlight the need to go “below” the phoneme, “beyond” the allophone, in our search for the proper unit of analysis in cross-language speech perception. This is not a new idea, but it is one that has not been taken to heart in much cross-language research. One issue this idea raises is the need for much more detailed descriptions of the phonetics of languages than exist for most of the world’s languages. To test a cue-weighting model, we need to describe how phonetic cues for a given phoneme or allophone are weighted in a linguistic community, perhaps through the use of multiple one-dimensional synthetic continua in standard identification tests, or multidimensional synthetic spaces as in the method of adjustment technique [7]. Testing an exemplar-based model would require even more descriptive work on a language entailing the development of large phonetic databases from which distributions of speech sounds along given dimensions could be extracted. However, without such descriptive work, we may never be able to develop models of cross-language speech perception that are capable of making quantitative predictions of the magnitude of cross-language differences. This study, even with its inherent limitations in terms of stimuli (only one talker, only two exemplars of each speech sound), indicates that abstract phonological or phonetic models of the perceptual category do not capture important and significant variance in the cross-language data, and that a new approach, new techniques, and new assumptions are warranted.

NOTES
1. Marathi and Punjabi only differ in terms of the places of articulation exploited in their nasal series in one case: Marathi is sometimes described as having a velar nasal which contrasts with /m/, /n/ and /ŋ/ in final position. This velar nasal is a product of the reduction of an /ŋg/ to [ŋ] in casual speech. See Harnsberger (1998) for a summary of the literature on the phonetics of Marathi nasals. This difference between Marathi and Punjabi could have manifested itself in the position of the non-native [ŋː1] stimuli in perceptual space (see 4.0).
2. Tamil and Oriya only differ in terms of the places of articulation exploited in their nasal series in one case: Tamil, unlike Oriya, has a contrastive palatal nasal in all but final position. See [4] for a summary of the literature on the phonetics of Tamil nasals. This difference between Tamil and Oriya could have manifested itself in the position of the non-native [ŋː1] stimuli in perceptual space, but did not (see 4.0).
3. Malayalam listeners, in a categorial AXB discrimination test employing the exact stimuli, successfully discriminated the interdental and alveolar stimuli, with a mean percent correct score of 95.5% [4].
4. In an ALSCAL analysis, the absolute values of points in perceptual space are set to an arbitrary reference point. Thus, any whole set of points in these 2D graphs can be freely rotated, or shifted to any region in the space, so long as the interpoint distances remain unchanged.

REFERENCES