

RESEARCH ON SPOKEN LANGUAGE PROCESSING

Progress Report No. 23 (1999)

Indiana University

**The Effect of Linguistic Experience on Perceptual Similarity Among Nasal
Consonants: A Multidimensional Scaling Analysis¹**

James D. Harnsberger

Speech Research Laboratory

Department of Psychology

Indiana University

Bloomington, Indiana 47405

¹ This work was supported by NIH-NIDCD Training Grant DC00012 to Indiana University.

The Effect of Linguistic Experience on Perceptual Similarity Among Nasal Consonants: A Multidimensional Scaling Analysis

Abstract. In cross-language speech perception studies, the perceptual categories of a listener group are often assumed to be adequately represented by single, abstract labels, such as the position-dependent allophonic variant of a phoneme. However, listeners of different languages may also vary in their perceptual weighting of acoustic cues that signal a given phoneme, or its allophonic realization, in their language. These two competing units of analysis were evaluated in a cross-language perceptual similarity test employing a broad range of non-native stimuli and listener groups. In this experiment, an AXB classification test using nasal consonants from Malayalam (bilabial, interdental, alveolar, retroflex, palatal, velar) 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 similarity scores revealed language-specific differences that were not predictable from the test groups' nasal inventories, as represented by position-dependent allophones. The dental-retroflex and alveolar groups showed intra-group differences in their clustering of stimuli and their 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 the native perceptual categories of listeners must be made at the level of the individual acoustic cues that are used to match acoustic input to particular perceptual categories, rather than abstract labels.

Introduction

Cross-language speech perception research has demonstrated that specific linguistic experience can limit listener sensitivity to some non-native phonemic distinctions (Abramson and Lisker, 1970; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura, 1975; Werker, Gilbert, Humphrey, and Tees, 1981). Non-native consonant stimuli that correspond to, or are most similar to, a single phoneme in a listener's native language have often proven to be difficult for listeners to accurately discriminate and identify, such as /l-ɹ/ for Japanese listeners (MacKain, Best, and Strange, 1981), Hindi dental-retroflex stops for English listeners (Werker et al., 1981), and Salish /k'-q'/ for English and Farsi listeners (Werker and Tees, 1984a, b). However, non-native consonant contrasts have been shown to vary in their discriminability, from chance to near native level performance (Polka, 1991; Pruitt, 1995). 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 differ in their ability to identify and/or discriminate natural American English (AE) or Australian /l-ɹ/, despite the fact that all three languages have only one native liquid phoneme (Henly and Sheldon, 1986; Ingram and Park, 1998). 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 (Pruitt, 1995). The variability in sensitivity of listener groups to non-native, non-phonemic contrasts has led some researchers to focus on alternate descriptions of perceptual categories and models of perceptual similarity. Candidates for perceptual category descriptors, or units of analysis, include context-dependent allophones (Strange, 1995), phonetic features or cues (Bennett, 1968; Gottfried and Beddor, 1988), and distributions of individual exemplars (Pisoni and Lively, 1995; Yoneyama and Johnson, 1998). Perceptual similarity is frequently

assumed to be a transparent, linear mapping between a stimulus and a category that share a common descriptor, usually a phoneme or allophone.

This study was designed to address the unit of analysis and perceptual similarity issues in cross-language speech perception. The study addressed the following, related questions: Is the position-dependent allophone as the unit of analysis sufficient to describe listeners' perceptual categories in cross-language speech perception? Do listener groups with common phonetic inventories for a given non-native sound show any differences in their perception of such sounds? These questions were assessed by examining the perceived similarity among a set of non-native sounds by a large set of listener groups varying in their native inventories, which were described in terms of position-dependent allophones.

Six types of nasal consonants varying in place of articulation served as stimulus materials in this study. They included bilabial ([m]), interdental ([ɱ]), alveolar ([n]), retroflex ([ɳ]), palatal ([ɲ]), and velar ([ŋ]) nasals. Nasal consonants were chosen as stimuli because it was hoped that they would prove to be a perceptually challenging set for some or all of the non-native listener groups. In prior work, nasal consonants varying in place of articulation have been shown to be confusable relative to other contrasts (Mohr and Wang, 1968; Hura, Lindblom, and Diehl, 1992). Non-native sounds that are highly confusable may elicit significant cross-language differences in perceived similarity that would not be predictable from abstract descriptions of a listener's language, based on units such as the phoneme or allophone.

The listener groups tested with this stimulus set were speakers of Malayalam, Marathi, Punjabi, Tamil, Oriya, Bengali, and AE. Malayalam listeners served as a control group. The six non-native listener groups were chosen to represent three different types of coronal nasal consonant inventories: dental-retroflex (Marathi, Punjabi), alveolar-retroflex (Tamil, Oriya) and alveolar (Bengali, AE). Listener groups were selected on the basis of their *coronal* nasal consonant inventory because prior work on the perception of non-native place distinctions (Polka, 1991), as well as pilot testing (see Harnsberger, 1998), had indicated that dental, alveolar, and retroflex nasals may be more confusable as a group relative to other pairings of nasals (i.e. bilabial and alveolar, palatal and velar), and thus could elicit significant cross-language differences in perceived similarity.

The listener groups were presented with triads of nasal consonants varying in place of articulation in an AXB format, and were asked to choose which nasal consonant stimulus, A or B, was more similar to X. The frequency with which two stimulus types, such as dental and alveolar, were grouped together across all test trials served as a raw similarity score. A full set of these scores, for all possible pairings of the six types of nasal consonants, were then submitted to an ALSCAL multidimensional scaling analysis (MDS) for the purpose of mapping each listener group's perceptual "space" for this stimulus set, that is, the arrangement of the six stimulus types in an n-dimensional space illustrating their degree of similarity. The perceptual spaces of each listener group were compared to see if any cross-language differences emerged that could not be predicted from their nasal consonant inventories, described in terms of allophones. Specifically, the six non-native listener groups, paired off by their common coronal nasals into three test groups (dental-retroflex, alveolar-retroflex, and alveolar), were chosen to compare two hypotheses, the *allophonic category center* and *cue-weighting* hypotheses. The allophonic category center hypothesis states that listeners' native categories are abstract and sufficiently described by the context-dependent allophonic variants of phonemes. The hypothesis represents an assumption made by several cross-language speech perception models (Best, 1995; Flege, 1995). In contrast, the cue-weighting hypothesis maintains that listener groups (varying in native language) may systematically differ in their weighting of the acoustic cues for allophonically the "same" nasal consonants. For example, Bengali and AE listeners, two listener groups which both have native alveolar nasals, may differ in the degree to which they attend to particular

acoustic cues to alveolar nasals. Cue-weighting as a source of cross-linguistic variability has been suggested by the results of a number of studies, including work by Bennett (1968), Gottfried and Beddor (1988), and Rochet (1991). 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. For example, we would expect to see no differences in the perceptual spaces of Marathi and Punjabi listeners (both dental-retroflex groups), Tamil and Oriya listeners (both alveolar-retroflex) or Bengali and AE listeners (both alveolar). If instead, the cue-weighting hypothesis is correct, then we would expect to observe perceptual spaces that vary on a listener group by listener group basis. Such individual listener group variability would suggest language-specific weighting of the perceptual dimensions of a space, each of which corresponds to an acoustic cue or a complex of cues. That is, we might observe significant differences between the Marathi and Punjabi spaces, between the Tamil and Oriya spaces, and/or the Bengali and English spaces.

Methods

Stimulus Materials

The stimulus materials were restricted to two exemplars each of six types of nasal consonants varying in place of articulation ([m], [ɱ], [n], [ɳ], [ɲ], [ɳ]) from a single speaker of Malayalam, a Dravidian language spoken in southern India. The stimuli appeared as medial geminates in an [iNi] 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., [m₁]-[n₁]-[ɳ₁], [m₂]-[n₂]-[ɳ₂], but never [m₁]-[n₂]-[ɳ₁]). 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 interstimulus, intertrial, and interblock intervals for the similarity test were 1 s, 5 s, and 6 s, respectively, with 20 trials per block (total test time: 33.5 minutes). The use of just two exemplar sets for a total of 240 trials, 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 The stimuli from this single talker were consistently correctly judged as representative of Malayalam nasals in pilot tests, and elicited similar results as those of a second Malayalam talker in additional identification and discrimination tests (Harnsberger, 1998).

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

² For example, if instead of two exemplar sets (A₁-X₁-B₁ and A₂-X₂-B₂), all of the twelve stimuli had been combined in all possible orders, the number of test trials would have been 960 (8 exemplar sets * 6 order * 20 types of triads), which would have required approximately 2.5 hours of testing. Unfortunately, testing in India did not allow for the use of testing facilities for the number of sessions necessary to run such a lengthy test, in combination with several other tests that were administered (Harnsberger, 1998).

consonant contrast (Marathi, Punjabi³), an alveolar-retroflex test group (Tamil, Oriya⁴), and an alveolar test group (Bengali, AE)

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 categories.⁵ 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.

All listener groups, except for Punjabi and AE listeners, were tested in sound-attenuated chambers affiliated with private studios in New Delhi (Malayalam, Marathi, Tamil, Oriya) and Calcutta (Bengali) in India. Native Punjabi speakers 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.

Predictions

Two general 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. The report of the results will focus in particular on the perceived similarity for the coronal nasals, as the dental-retroflex and alveolar-retroflex groups show possible intragroup differences in their noncoronal nasal consonant inventory (see footnotes 3 and 4). However, an examination of the complete perceptual spaces of the six

³ 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/, /ŋ/ and /ŋ̠/ 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 [iŋ^hi] stimuli in perceptual space (see Results and Discussion).

⁴ 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 Harnsberger (1998) 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 [iŋ^hi] stimuli in perceptual space (see Results and Discussion).

⁵ See Goldinger (1998) for another example of AXB classification.

listener groups individually provides important descriptive information concerning the relationship between the nasal consonant inventory of a listener group and its perception of non-native nasal contrasts.

Results and Discussion

Cross-Language Differences

The results of the AXB classification test were analyzed by listener group. The raw similarity scores for all possible stimulus pairs (e.g. bilabial and alveolar, retroflex and palatal) were calculated by determining the frequency with which a stimulus type that appeared as X was judged as similar to another stimulus type, across all triads in which they appeared together. For example, the frequency with which [n] was judged as similar to [m] was calculated over all triads in which [m] was X and [n] was either A or B. These similarity scores were calculated for each individual subject, and the scores were averaged together to calculate the similarity scores for an entire listener group. In addition, the scores for similar stimulus pairs were geometrically averaged for submission to an ALSCAL MDS analysis. For example, the score for [m]-[n] when [m] was X was geometrically averaged with the score for [n]-[m] when [n] was X. Thus, a single similarity score for two stimulus types, such as [m] and [n], was used in the MDS analysis, as the analysis requires.

Each listener group's mean similarity scores were then submitted separately to both a one- and two-dimensional (2D) ALSCAL 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 values for each listener group for the 2D analyses. The perceptual spaces derived from the analyses for the control and test groups appear in Figures 1-4, with each axis representing similarity on some undefined, perceptually relevant dimension. Figure 1 provides the perceptual space of the control group for this experiment, the Malayalam listeners. For this task, the Malayalam listeners were expected to show a relative lack of clustering among all six stimulus types, given that all six nasals are used contrastively in the language. This expectation was upheld in all interstimulus distances, with the exception of the interdental and alveolar nasals, which were judged as significantly more similar to one another than either was to any other nasal in the perceptual space in a factorial ANOVA ($df = 14$, $F = 13.9$, $p \leq 0.0001$). Such close similarity relationships are indicated in Figure 1, and in Figures 2-4 as well, with a solid-line circle encompassing the similar stimulus types. While the close similarity between the interdental and alveolar nasals was somewhat surprising, these distances are measures of relative similarity, which do not necessarily entail that the stimuli were similar enough to be confusable for Malayalam listeners. In fact, these stimuli were successfully discriminated by the same Malayalam listeners in a prior experiment, a categorial AXB discrimination test in 96% of all test trials (Harnsberger, 1998). Overall, the Malayalam listeners produced a perceptual space in which most nasals occupied their own region of perceptual space, consistent with their phonemic status in the language.

Figures 2-4 show the perceptual spaces of the alveolar, dental-retroflex, and alveolar-retroflex listener groups. As predicted by the cue-weighting hypothesis, listener groups with similar nasal consonant inventories at the allophonic level differed from one another in several cases. For instance, the alveolar test group, composed of Bengali and AE listeners with identical nasal consonant inventories, differed from one

another in their apparent attention to the two perceptual dimensions and in the distances between different stimulus types. 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. Moreover, Bengali listeners also judged the coronal stimuli to be highly similar to one another relative to the AE listeners. All of the English coronal interstimulus distances (interdental-alveolar, alveolar-retroflex, dental-retroflex) were significantly greater ($p \leq 0.05$) than the corresponding Bengali distances in post-hoc t-tests in a repeated measures mixed model ANOVA, in which Native Language ($F(1,33) = 5.44, p \leq 0.05$), Stimulus Pair ($F(14,462) = 41.89, p \leq 0.0001$), and their interaction ($F(14,462) = 3.34, p \leq 0.0001$), were all significant. In addition, the coronal interstimulus distances in the Bengali space were significantly shorter than all other distances between coronal and noncoronal stimulus types in the Bengali space. In other words, Bengali listeners judged the non-native coronal stimuli to be highly similar to one another and relatively dissimilar to [m], [n], and [ŋ]. Overall, the Bengali space can be apportioned into three sets of sounds: the coronal stimulus types, the palatal and velar stimulus types, and the bilabial stimulus type. These clusters of similar sounds are indicated in Figure 2 by circles around stimulus types that were judged to be similar to one another relative to their similarity to stimulus types outside of a given circle.

Group	Stress ⁶	R ²
Malayalam	0.07129	0.96907
Marathi	0.03193	0.99476
Punjabi	0.04794	0.98985
Tamil	0.00486	0.9999
Oriya	0.01615	0.99882
AE	0.05194	0.9822
Bengali	0.03993	0.99133

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

In contrast to the Bengali listeners, 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 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

⁶ "Stress" in an MDS analysis is a measure of the fit of the derived distances to the raw similarity scores submitted to the analysis.

stimuli. The resulting distances between coronal stimulus types did not differ significantly from the distances of individual coronal stimulus types and near-neighbors such as [ɲ] for [n] or [m] for [ŋ], as indicated by dotted-circle lines in Figure 2. Thus, the AE space could not be nicely apportioned in the same manner as the Bengali space, a result that could not have been predicted by the allophonic category center hypothesis and one that was congruent with the cue-weighting hypothesis.

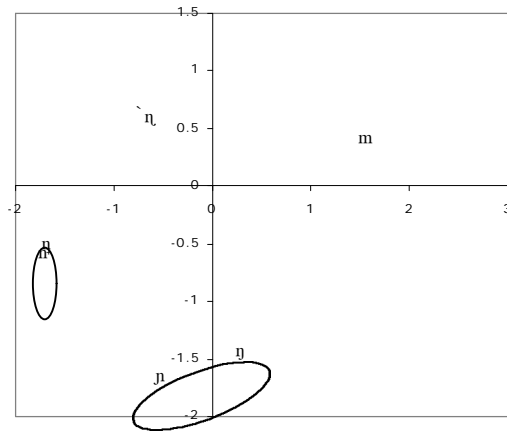


Figure 1. 2D MDS analysis of the Malayalam similarity scores. Solid-line circles encompass nasals whose distances were not significantly different from one another *and* whose distances were significantly different from the distance between them and all other nasals outside of the circle.

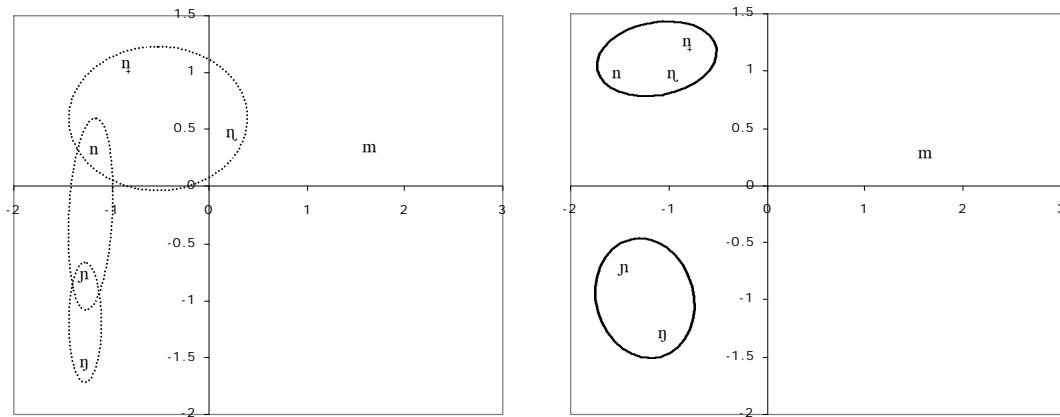


Figure 2. 2D MDS analysis of the alveolar test group’ similarity scores (Left: AE, Right: Bengali). Solid-line circles encompass nasals whose distances were not significantly different from one another *and* whose distances were significantly different from the distance between them and all other nasals outside of the circle. Dotted-line circles indicate less exclusive, partially overlapping clusters of nasals.

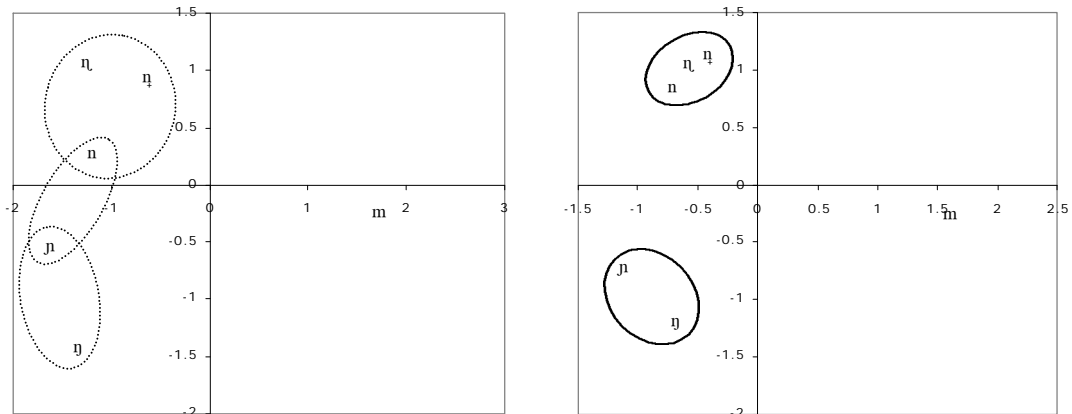


Figure 3. 2D MDS analysis of the dental-retroflex test groups' similarity scores (Left: Marathi, Right: Punjabi). Solid-line circles encompass nasals whose distances were not significantly different from one another *and* whose distances were significantly different from the distance between them and all other nasals outside of the circle. Dotted-line circles indicate less exclusive, partially overlapping clusters of nasals.

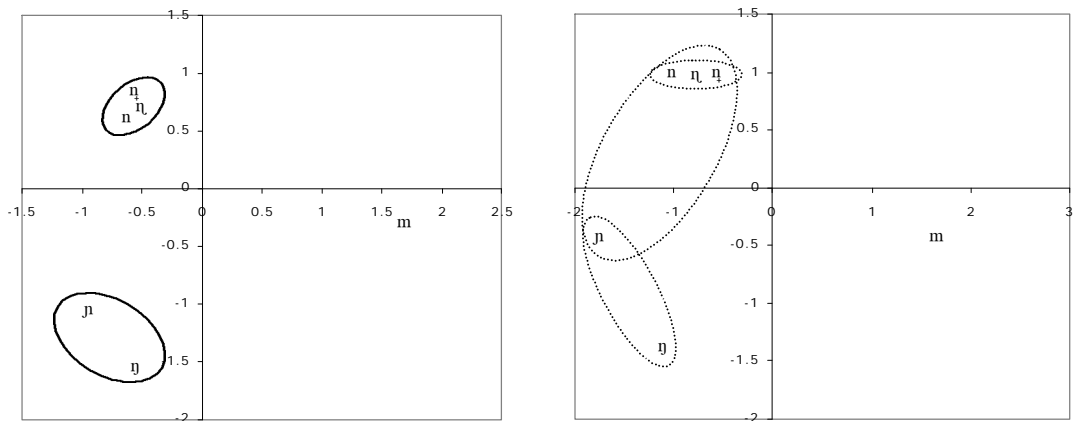


Figure 4. 2D MDS analysis of the alveolar-retroflex test groups' similarity scores (Left: Tamil, Right: Oriya). Solid-line circles encompass nasals whose distances were not significantly different from one another *and* whose distances were significantly different from the distance between them and all other nasals outside of the circle. Dotted-line circles indicate less exclusive, partially overlapping clusters of nasals.

The dental-retroflex test group, whose spaces appear in Figure 3, also showed intra-group differences not predicted by the allophonic category center hypothesis. The Marathi listeners, like the Bengali, appeared to use a single perceptual dimension in judging the [n], [ɲ], [ɳ], [ŋ], and, to a lesser extent, [ɳ̥] stimuli. And like AE, Marathi listeners demonstrated relatively less clustering among the coronal nasals. In contrast, Punjabi listeners resembled Bengali, and not Marathi listeners, in tightly clustering the [ɳ̥], [n], and [ɳ], reflecting perhaps less sensitivity to phonetic cues that are not contrastive in Punjabi. These general observations were confirmed in a repeated measures mixed model ANOVA, with interstimulus distances compared in post-hoc t-tests. While the Native Language of the listeners was not a significant factor ($F(1,30) = 0.79$, n.s.), Native Language did interact significantly with Stimulus Pair ($F(14,240) = 1.77$, $p \leq 0.05$). Stimulus Pair alone was also significant ($F(14,240) = 91.02$, $p \leq 0.0001$). In post-hoc t-tests, the two groups differed in their judged distances between alveolar and palatal nasals, and between interdental and retroflex nasals. Both of these differences were a product of the relative clustering of the coronal nasals by the two groups. With tight clustering in the Punjabi space, the coronal nasals were judged as highly similar to one another, and highly dissimilar to the palatal nasal. In post-hoc t-tests, the [ɳ̥]-[n], [ɳ̥]-[ɳ], and [n]-[ɳ] distances were not significantly different from one another, and all were significantly different from all coronal-noncoronal distances. In contrast, the relative spread of the coronal nasals in the Marathi space resulted in relatively a more dissimilar interdental-retroflex pair, and in a relatively similar alveolar-palatal pair. Post-hoc t-tests confirm these aspects of the Marathi space: the alveolar-palatal distance was not significantly different ($p \leq 0.05$) from the distance between the alveolar and retroflex nasal. This similarity relationship is indicated in Figure 3 by a dotted circle encompassing the alveolar and palatal nasal symbols. These results, like those of the alveolar group, cannot be accounted for by the allophonic category center hypothesis, and illustrate the need for a description of the cues that differentiate nasal consonants in these languages.

Of the three test groups, the alveolar-retroflex group showed the greatest similarities, and thus their data strongly supported the allophonic category center hypothesis. Both Tamil and Oriya listeners tightly clustered the [ɳ̥]-[n]-[ɳ] series, and both appeared to judge the non-bilabial stimuli along a single dimension. The similarities between these two spaces were supported by comparing the interstimulus distances of both groups in a repeated measures mixed model ANOVA. In this analysis, Native Language was not significant ($F(1,28) = 2.72$, n.s.), nor was its interaction with Stimulus Pair ($F(14,392) = 1.39$, n.s.). Stimulus Pair was itself significant ($F(14,392) = 88.42$, n.s.), as expected. In both groups' spaces, the coronal nasals clustered together, and had interstimulus distances that were significantly shorter than distances involving noncoronal stimuli ($p \leq 0.05$). In the Tamil space, the palatal and velar nasals also clustered together, as the solid circle in Figure 4 indicates. However, in the Oriya space, the distance between the palatal and velar nasal was large enough to insure that the palatal-alveolar and palatal-velar distances were not significantly different. This small difference was the only one observed between the two spaces. Interestingly, the distance between the palatal nasal and the coronal nasals was not significantly greater in the Tamil space than the Oriya space, despite the fact that Tamil possesses a palatal nasal phoneme along with their alveolar and retroflex nasals. Overall, the alveolar-retroflex spaces supported the allophonic category center hypothesis.

Language-General Patterns

Finally, in addition to cross-language differences, several general patterns emerged across most of the listener groups that were not anticipated prior to the experiment. First, the stimulus types formed three groups in terms of similarity, {[m]} {[ɳ̥]-[n]-[ɳ]} and {[ɳ]-[ɳ̥]}. 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. In addition, 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 [ɲ]-[ɳ]-[ŋ] series. These general groupings, made by a diverse set of listener groups spanning a range of nasal consonant inventories, may reflect general psychoacoustic differences between the stimuli that are robust enough to be perceived despite the “filtering” effect of the listeners' native nasal consonant inventory. Such language-general similarity patterns could serve as benchmarks for models of perceptual similarity that attempt to predict the identification of a non-native sound based on the native phonetic inventory of a listener's language, assuming that these patterns hold in similar experiments with a greater number of stimuli from multiple talkers.

Conclusions

The results of the 2D MDS analysis of the similarity scores from the classification test revealed substantial effects of language experience in the organization of perceptual spaces that cannot be accounted for by reference to abstract category centers such as 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 can be seen as a conclusion driven by 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 (Summerfield, 1987).

What appears to be cue weighting could also be the product of a language-specific distribution of individual exemplars of a nasal consonant, each of which corresponds to an instance or episode stored in long-term memory (Pisoni and Lively, 1995). 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 episodic-based account might take the following form: Bengali and AE listeners show, for instance, a difference in their relative clustering of the [ɲ]-[ɳ]-[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 [ɲ], [ɳ], and [ŋ].

The results of this experiment provide no support for a model of language-specific cue weighting, over an episodic-based model of speech perception. The differences between listener groups that share the same nasal consonant inventory only highlight the need to go 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 study raises is the need for more detailed descriptions of the phonetics of languages than exist for most of the world's languages. To test a cue-weighting model, we will need to describe how the phonetic cues for a given phoneme or allophone are weighted in a linguistic community. 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.

In summary, this study evaluated two competing units of analysis in cross-language speech perception, the position-dependent allophone and the phonetic cue, in a perceptual similarity test employing a broad range of nasal consonant stimuli and listener groups varying in their nasal consonant inventory. A two-dimensional multidimensional scaling analysis of the similarity scores revealed language-specific differences that were not predictable from the test groups' nasal inventories, as represented by position-dependent allophones. The two sets of listener groups, representing dental-retroflex and alveolar nasal inventories, showed intra-group differences in their clustering of stimuli and their weighting of perceptual dimensions, leading to language-specific perceptual spaces. Only the spaces of the two alveolar-retroflex were organized in a similar manner. This study, even with its inherent limitations in terms of stimuli (only two exemplars of each speech sound from one talker), indicated that abstract phonological or phonetic models of the perceptual category do not capture important and significant variability in the cross-language data, and that new techniques and new theoretical assumptions are warranted in our study of cross-language speech perception.

References

- Abramson, A. S., and Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In Hála, B., Romportl, M. and Janota, P. (ed.), *Proceedings of the Sixth International Congress of Phonetic Sciences*, pp. 569-573. Prague: Academia.
- Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech*, 11, 65-85.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in Cross-language speech perception*, pp. 171-204. Baltimore: York Press.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in Cross-language speech perception*, pp. 233-277. Baltimore: York Press.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.
- Gottfried, T. L. and Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech*, 31, 57-75.
- Harnsberger, J. D. (1998). *The perception of non-native nasal contrasts: a cross-linguistic perspective*. Unpublished doctoral dissertation, University of Michigan.
- Henly, E. and Sheldon, A. (1986). Duration and context effects on the perception of English /r/ and /l/: A comparison Cantonese and Japanese speakers. *Language Learning*, 36, 505-521.
- Hura, S.L., Lindblom, B., and Diehl, R. L. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech*, 35, 59-72.

- Ingram, J. and Park, S-G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America*, 103, 1161-74.
- MacKain, K. S., Best, C. T., and Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 369-90.
- Miyawaki, K., Strange W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophysics*, 18, 331-65.
- Mohr, B. and Wang, W. S.-Y. (1968). Perceptual distance and the specification of phonological features. *Phonetica*, 18, 31-45.
- Pisoni, D. and Lively, S. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, pp. 433-458. Baltimore: York Press.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, 89, 2961-777.
- Pruitt, J. S. (1995). *The perception of Hindi dental and retroflex stop consonants by native speakers of Japanese and AE*. Unpublished doctoral dissertation, University of South Florida, Tampa.
- Rochet, B. L. (1991). Perception of the high vowel continuum: A cross-language study. *Proceedings of the International Congress of Phonetic Sciences*, 4, 94-97.
- Strange, W. (1995). Cross-language studies of speech perception: A historical review. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, pp. 3-45. Baltimore: York Press.
- Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In Dodd, B. and Campbell, R. (ed.), *Hearing by Eye: The Psychology of Lip-Reading*, pp. 3-51. Hillsdale, NJ: LEA.
- Werker, J. F., Gilbert, J. H. V., Humphrey, K. and Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-53.
- Werker, J. F. and Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization in the first year of life. *Infant Behavior and Development*, 7, 49-63.
- Werker, J. F. and Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75, 1866-78.
- Yoneyama, K. and Johnson, K. (1998). An instance-based model of Japanese speech recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 103, 3090.

