The acoustic–perceptual salience of nasal place contrasts

Chandan R. Narayan*

Institute for Research in Cognitive Science, University of Pennsylvania, 3401 Walnut Street, Suite 400A, Philadelphia, PA 19104, USA

Abstract

This study examined the perception of the typologically frequent /m/-/n/ contrast and the less common /n/-/ŋ/ contrast in syllable-onset position. In an acoustic study of [ma], [na], and [ŋa] tokens as spoken by three native speakers of Filipino (Experiment 1), both static (F2 and F3 values measured at the nasal–vowel juncture) and dynamic (rms energy change from murmur to vowel onset) measures showed that [na] tokens are more similar to [ŋa] than to [ma]. To test whether the acoustic similarity led to corresponding perceptual effects, native English and Filipino listeners were presented Filipino [ma]-[na] and [na]-[ŋa] pairs in a discrimination test (Experiment 2). English listeners showed a non-native effect, accurately discriminating the [ma]-[na] distinction while performing at chance on [na]-[ŋa]. Interestingly, Filipino listeners showed the same pattern of performance, albeit at a more moderate level. The performance of Filipino listeners in Experiment 2 was interpreted as originating from the acoustic similarity between /na/ and /ŋa/. To draw out the effects of the varying perceptual salience of the two contrasts, in Experiments 3 and 4 Filipino listeners were presented with the same contrasts (with the addition of [ma]-[ŋa] in Experiment 4) for discrimination in three listening conditions: no additive noise, 0 and −5 dB SNR. Listeners' discriminations of the [ma]-[na] and [ma]-[ŋa] contrasts remained near ceiling levels across the three conditions, while performance on [na]-[ŋa] fell to near chance in the noisiest condition. Taken together, these results are suggestive of a role for acoustic–perceptual salience in the distribution of nasal place contrasts in the world's languages, reflecting acoustically robust and perceptually distinct contrasts over those that are acoustically similar and perceptually confusable.

1. Introduction

Phonologists and phoneticians share an interest in understanding why phonological systems in the world's languages pattern the way they do (e.g., Blevins, 2006; Flemming, 2001; Guion, 1996; Hume & Johnson, 2001; Maddieson, 1986; Ohala, 1974; Steriade, 2001a, 2001b; Stevens, 1989; among others). A widely recognized contribution of perception to the shape of sound systems is that they must reflect sufficiently distinct phonetic contrast between phones (e.g., Crothers, 1978; Diehl & Kluender, 1989; Disner, 1983; Jakobson, 1941, 1963; Passy, 1890). This view has been especially advocated by Lindblom (1986, 1990) under the well-known theory of dispersion, which proposes that phonological contrasts are sufficiently distinctive perceptually. Following on from Liljencrants and Lindblom's (1972) (and later Lindblom's, 1986) model for predicting the inventories

---

*Tel.: +1 215 573 6282.

E-mail address: cnarayan@sas.upenn.edu

of *n*-vowel systems, dispersion theory suggests that phonological systems tend towards satisfying the dual needs of maximizing the distinctiveness of contrasts for the listener and minimizing articulatory effort for the speaker. The Liljencrants–Lindblom model showed, based on well-known principles of acoustic theory and speech perception, that vowel contrasts in the world's languages bear similarities to those derived under the dispersion theory. Johnson, Flemming, and Wright (1993) directly tested the dispersion theory as an active perceptual process and found that listeners chose, as best exemplars of English vowel categories, stimuli that were more dispersed in $F1 \times F2$ space than vowel qualities produced in real speech. Such a "hyperspace effect" was taken as direct evidence that dispersion, or the preference for maximally distinct contrasts, operates within listeners (but see Johnson, Flemming, & Wright, 2004; Whalen, Magen, Pouplier, Kang, & Iskarous, 2004). More recently, the theory of dispersion has been incorporated into formal phonological analyses of inventory design. For example, Flemming (2004, 2006) showed that languages strike a balance between constraints (in optimality theoretic terms) specifying robust perceptual contrasts while maximizing the number of contrasts.

This paper follows in the tradition of functional explanations in phonology by considering acoustic similarity between members of a phonetic contrast, its corresponding effects on perceptual salience, and their consequences within a language and across phonological systems. As a starting point consider the fact that phonological inventories are often not symmetrical, with some contrasts being typologically more common than others (along some linguistically significant dimension). Presence of the typologically less common contrast in a given language implies presence of the typologically frequent contrast in that language. But what are the perceptual salience relations in languages that have both common and uncommon contrasts? I argue that, in these languages, typologically less common contrasts are also underrepresented in spontaneous speech and lexical distinctions. This is suggested to result from their relative lack of perceptual salience, such that perceptually distinct contrasts (which are acoustically discrete) are preferred over perceptually fragile contrasts (which are acoustically similar and thus prone to be confusable) (Maddieson, 1986).

There is evidence in the literature that differences in degree of acoustic similarity between phonetic categories correspond to different degrees of perceptual confusion (Bladon & Lindblom, 1981; Carlson, Granström, & Klatt, 1979; Nord & Sventelius, 1979). These studies have shown that acoustic distances, calculated on the basis of various physical properties of the signal, are highly correlated with listeners' similarity judgments (with correlation coefficients generally between 0.80 and 0.90). Krull's (1990) study of naturally produced Swedish voiced stops showed that the size of acoustic differences corresponded to listener confusions in an identification task. For example, she found that [gɛ] and [ga] were close to [d] in $F2 \times F3$ space (measured at the CV boundary) and this acoustic distance led to a corresponding perceptual confusion.

Evenly spaced phones in acoustic space are perceptually ideal because they allow a maximal number of distinct contrasts. Since, however, acoustic spacing is often asymmetrical, languages may maximize perceptual salience by reducing the number of contrasts (e.g., Old Indo-Aryan /s/-/ʃ/-/ʂ/ > Western and Southern New Indo-Aryan /s/-/ʃ/ presumably due to the acoustic similarity of /ʃ/ and /ʂ/,[1] Masica 1991; Proto-Dravidian word-initial /m/-/n/-/ɲ/ > Kannada /m/-/n/, Zvelebil, 1970). The problem of acoustically similar phones in contrast may also be lessened by restricting the contrast to environments that enhance perceptual distinctiveness (Diehl & Kluender, 1989; Flemming, 2006; Kingston & Diehl, 1994; Steriade, 2001a, 2001b). Such phonological restrictions would allow a maximum number of contrasts to be maintained while maximizing their perceptual salience. For example, Middle Indo-Aryan (e.g., Maharashtri) allowed both /n/ and /ŋ/ in word-initial and medial positions (Bubenik, 1996), while modern Marathi (Western New Indo-Aryan) restricts /ŋ/ to word-medial position (Pandharipande, 2003). Presumably, the acoustic–perceptual cues to nasal place are more numerous in VN(C)V position than in #NV structures (here and elsewhere N is used to denote a nasal consonant at any place of articulation); thus allowing for more robust identification of place. Finally, the effects of acoustically similar contrasts are often reduced by minimizing the occurrence of the contrast in the lexicon and/or speech. For example, the palatal nasal in Catalan appears medially (in syllable-onset position) in many lexical items, but in less than 20 items in word-initial position (Wheeler, 2005). Again, the acoustic–perceptual cues for VNV identification are more robust than in NV contexts.

---

[1]The situation in Modern Indo-Aryan was most likely preceded by a stage in Middle Indo-Aryan that collapsed all sibilants to /s/. The palatal /ʃ/ was re-introduced with Sanskrit borrowings. This begs the question, why was /ʂ/ not borrowed as well?

The goal of this paper is to show that typological asymmetries across languages are mirrored within a language and may result from asymmetrical acoustic spacing of distinctive sounds and corresponding asymmetries in their perception. I approach this through a study of the perception of syllable onset nasal place of articulation. Nasals as a class are interesting for an inquiry into the relative salience of contrast because their place characteristics are notoriously difficult to perceive (House, 1957; Hura, Lindblom, & Diehl, 1992; Malécot, 1956; Mohr & Wang, 1968) yet certain nasal onset places of articulation are found in nearly every language of the world (Maddieson, 1984).

The acoustic consequences of the resonating nasal cavity are highly variable. Studies focused on identifying the place-specific characteristics of the nasal murmur have yielded inconsistent results, presumably due to variability between and within speakers' productions (Fujimura, 1962; Qi & Fox, 1992). Place characteristics have been identified, however, at the nasal–vowel (NV) juncture. Following Repp (1986), Kurowski and Blumstein (1987) investigated the dynamic aspect of the NV juncture and showed that by transforming the units of analysis from Hertz to Bark (i.e., units based on the psychoacoustic characteristics of the human auditory system), general place of articulation characteristics for [m] and [n] in various vowel contexts emerge. Their analysis identified two spectral regions (in Bark 5–7 and Bark 11–14) in which rms energy changes from the nasal murmur to the vocalic transition allowed for successful classification of alveolar and labial places of articulation. Relative to the high-frequency band, labials were found to show greatest rms energy change in the region of Bark 5–7. Alveolars showed greatest change in the higher band. Seitz, McCormick, Watson, and Bladon (1990) followed Kurowski and Blumstein's (1987) general method but examined relative, rather than absolute, frequencies. Their metric performed slightly better (in classification) than Kurowski and Blumstein's (1987) method on the same stimuli. Harrington (1994), in a large study of continuous speech data, showed that the dynamic method, in conjunction with static measures at various locations in the murmur and vowel, resulted in the highest classification scores for labials and alveolars.

Similar to the dynamic approach, locus equations, which specify a best fit line from the F2 frequency of the post-consonantal vowel at the release of the consonant to the midpoint of the vowel have been used to characterize consonantal place. Locus equations have been found to be consistent within a place and across manner (Sussman and Shore, 1996), suggesting that the acoustic characteristics at CV and NV junctures are roughly equivalent within place. More generally, given the theoretical similarity in the articulatory configuration of the oral components of both nasal and non-nasal manners, the acoustic correlates for nasal places of articulation are considered similar to their homorganic oral counterparts, with gross spectral information at the CV juncture providing a general characterization (Blumstein and Stevens, 1979; Stevens, 1989).

Unlike the acoustics literature on nasals which has been predominantly concerned with syllable-initial /m/ and /n/, the literature on nasal place perception in syllable-initial position has considered bilabial, alveolar/dental places, and post-alveolar places like the velar (Delattre, Liberman, & Cooper, 1955; Larkey, Wald, & Strange, 1978; Liberman, Delattre, Cooper, & Gerstman, 1954; Nakata, 1959). The majority of these studies were conducted using synthetic stimuli in order to identify the relative contributions of the nasal murmur and formant transitions into the following vowel, both of which were considered to provide place information. The results of these early studies suggested that place information is best conveyed by the vocalic transitions, while the murmur provides manner information. This general result held true for real speech stimuli as well (Malécot, 1956; Repp, 1986). The perceptual literature also confirmed that the formant loci at the NV juncture serving as cues for nasal place perception are similar to the loci serving oral stop perception (Larkey et al., 1978; Liberman et al., 1954). This finding is consistent with the acoustic similarity between homorganic syllable-onset nasal and oral stops, which show similar constrictions and release into the following vowel (Stevens, 1998).

The cross-linguistic literature has also provided important information regarding nasal place perception (e.g., Harnsberger, 2000). The non-native effect shown in numerous vowel and consonant contrasts (e.g., /r-l/ perception by Japanese listeners in Miyawaki et al., 1975; see Strange, 1995 for a full review) has been shown in nasal consonant perception as well. Larkey et al. (1978) investigated English speakers' perception of synthetically generated onset and coda nasal place stimuli, [næ], [mæ], [ŋæ], [æn], [æm], [æŋ] along a continuum varying in the F2 and F3 frequencies of the pre- and post-nasal vowel. Using an AXB task, they found that English-speaking listeners' identification consistency and discrimination accuracy were inferior for the syllable-onset [næ]-[ŋæ] contrast compared to the coda [æn]-[æŋ] contrast. (In comparison, listeners in that

same study accurately identified and discriminated the [b]-[g] contrast in both syllable-onset and coda positions.) In a more recent cross-language, AXB experiment using natural speech stimuli from Cantonese, Narayan (2004) found that English-speaking listeners were able to discriminate [ma]-[na] at rates above 90% correct, but were less than 70% accurate on the [na]-[ŋa] contrast. Cantonese-speaking listeners, for whom the velar nasal is phonemic in syllable-onset position, were able to successfully discriminate both the [ma]-[na] contrast (at 98% correct) and [na]-[ŋa] contrasts (at 94% correct).[2] In both the experiments by Larkey et al. (1978) and Narayan (2004), English-listeners' relatively poor performance in discriminating the non-native contrast was attributed to the lack of an onset /n/-/ŋ/ contrast in English.

2. Nasal place typology

The typology of nasal place presents a clear tendency for a preference for certain places of articulation over others. Ferguson (1963) claims that if a language has only one nasal it is /n/ and if a language has only two nasals they are /n/ and /m/. This asymmetry towards /n/ and /m/ over post-alveolar places of articulation is empirically supported by Maddieson's (1984) survey of the sound patterns of 317 of the world's languages in the UCLA Phonological Segment Inventory Database (UPSID). In that database, 316 languages (or over 99%) have apical nasals (at dental, dental/alveolar, or alveolar places of articulation), 299 (94%) have the bilabial nasal and 167 (52%) have the velar. Maddieson noted that the presence of an oral stop at a particular place of articulation does not necessarily imply a homorganic nasal. He found that while 283 languages in the database exhibited the voiceless oral velar obstruent /k/, only 167 showed the homorganic nasal stop /ŋ/.

The typology becomes more complex when we consider the distribution of /ŋ/ according to syllable position. In a recent survey of 468 genetically diverse languages, Anderson (2005) found that half lacked /ŋ/ altogether (in both syllable-onset and coda positions). Of the 234 languages that had the velar nasal, 88 had it in syllable-coda position only, while the remaining 146 had it in *both* onsets and codas. There was one language in Anderson's survey, a Grebo language (Kru, Niger-Congo; Liberia), in which /ŋ/ occurs in *only* syllable-onset position. Taken together, the typological asymmetry that emerges is two-fold: (1) overall, alveolar/dental and bilabials nasals are near universal while velar nasals occur in fewer than half of the world's languages and (2) languages with velar nasals tend to distribute them in either coda position only or in both onset and coda positions.

2.1. Intra-language nasal place typology: Filipino (Austronesian)[3] and Thai (Tai)

That cross-linguistic typology shows the favored status of /n/ and /m/ (especially in syllable onsets) in the world's phonological systems should not take away from the fact that many languages of the world *do* contrast syllable-onset /ŋ/ with other nasal places of articulation. We might ask, then, whether languages that exhibit a /m/-/n/-/ŋ/ contrast in syllable onsets (an uncommon pattern) nonetheless show an asymmetry in their frequency distribution—that is, is syllable onset /ŋ/ used less often than /m/ and /n/ by speakers whose phonology shows an /m/-/n/-/ŋ/ contrast? To answer this question, transcribed corpora of the spontaneous speech (story telling and candid questionnaire answers) of 60 Filipino speakers (made available from the Digital Signal Processing group at the University of Diliman, Philippines) and 24 speakers of Thai (Cotsomrong, Sunpetchniyom, Kasuriya, Thatphithakku, & Wutiwiwatchai, 2005) were analyzed. Filipino and Thai exhibit a three-way phonological contrast between syllable onset /m/, /n/, and /ŋ/ (Schachter & Otanes, 1972; Smyth, 2002). The corpora were examined using a script that searched transcribed text for instances of syllable-initial nasal consonants across varying vowel contexts. The resulting distributions are given in Fig. 1.

---

[2]The syllable-onset /n/-/ŋ/ contrast is all but lost in the speech of Cantonese speakers from Hong Kong. The listeners in Narayan's (2004) study were from the Guangdong province of mainland China, where the contrast is retained.

[3]Filipino is based on Tagalog (Austronesian, Philippines) and is the national language of the Philippines. Filipino tends to borrow lexical items from the other Austronesian languages of the Philippines (such as Cebuano, Ilocano, Kampanpangan, etc.) and serves as a *lingua franca* among the various ethnic communities of the Philippine islands. The term "Filipino" is used rather than "Tagalog" in this paper as the latter term is also used to describe the Tagalog ethnic group. "Filipino" is considered a neutral designation as many of the speakers and listeners who participated in these studies belong to other, non-Tagalog, ethnic communities.
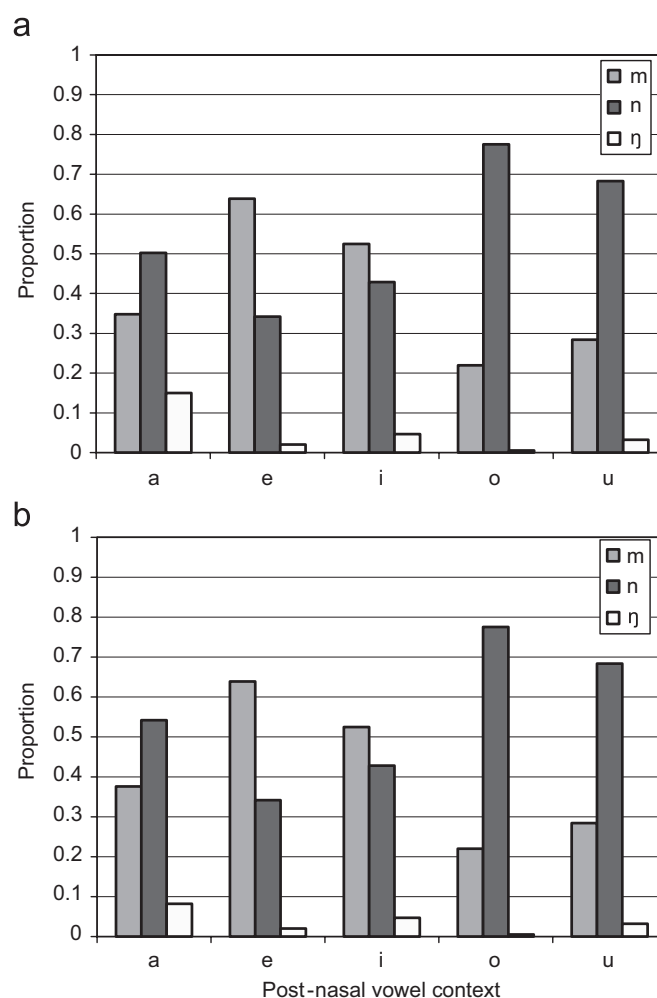
a

b

Fig. 1. Proportion of nasals in spontaneous (a) Filipino and (b) Thai speech. Data based on the automatic transcription of the spontaneous speech of 60 native speakers of Filipino and 24 native speakers of Thai.

The figure shows that in both languages /n/ is the most frequent nasal when followed by the central (/a/) and back (/o/ and /u/) vowels, while /m/ is most frequent before the front vowels (/i/ and /e/). Across all vowel contexts, the least frequent nasal is /ŋ/. Overall, the most frequently occurring syllable-onset nasal is coronal, followed by the bilabial then the velar. Of the 12,608 onset nasals in the Filipino corpus, /ŋ/ constitutes only 11% of tokens, while /n/ constitutes 52% and /m/ 37%. The distribution is a bit more asymmetrical in the 11,996 onset nasals from Thai corpus, where /ŋ/ constitutes only 6% of the tokens, while /n/ constitutes 55% and /m/ 39%. These intra-language asymmetries in nasal place distribution resemble the cross-linguistic typology of nasal place.

Given the inter- and intra-language typology of nasal place above, the present set of experiments asks whether velar nasals are acoustically similar to alveolar or bilabial nasals and if so, whether there are corresponding perceptual effects in listeners whose phonology reflects the three-way onset contrast. A two-step approach tested the role of perceptual salience in the distribution of common and uncommon nasal contrasts. First, the acoustics of nasal place in Filipino was investigated in order to assess the relative acoustic similarity between /m/, /n/, and /ŋ/ (Experiment 1). Findings from the acoustic study were then used to select real-speech stimuli for three perception studies assessing the relative salience of an /m/-/n/ contrast versus /n/-/ŋ/.

3. Experiment 1: acoustic analysis of Filipino stimuli

Experiment 1 was conducted to characterize the acoustic similarities between /m/, /n/, and /ŋ/ in Filipino and to guide stimuli selection for subsequent perceptual experiments. Two measures were analyzed, static and

dynamic, in characterizing the acoustic cues for /m/, /n/, and /ŋ/ in syllable-onset position followed by the central vowel /a/. The static measurement was focused on the F2 and F3 characteristics as measured at the NV juncture, which was guided by general acoustic properties of CV syllables as well as perceptual literature showing the effects of F2 and F3 values for place perception. The acoustics literature, reviewed above, suggests that the characteristics of syllable onset nasal place at the NV juncture are similar to the acoustics of their homorganic oral counterparts (Blumstein & Stevens, 1979; Cooper, Delattre, Liberman, Borst, & Gesterman, 1952; Stevens, 1989). The acoustic features of the velar nasal at the NV juncture are therefore predicted to be similar to those of /g/, with the post-nasal vowel onset exhibiting F2 values similar to those of alveolar tokens and F3 values lower than those of alveolars, revealing the so-called "velar pinch" (Stevens, 1998). Owing to the lack of a front-cavity resonance for labials, post-bilabial vowel onsets are predicted to have F2 values considerably lower in frequency than those of vowel onsets following either alveolar or velar closures. The dynamic measurement followed closely the acoustic analysis of nasal place by Kurowski and Blumstein (1987) and more generally the analyses by Seitz et al. (1990) and Harrington (1994), all of which show place characteristics emerging from a change-over-time metric from nasal murmur to vowel in specific frequency bands.

3.1. Methods

3.1.1. Materials

Speech materials consisted of Filipino monosyllables read in isolation. Randomized lists containing the three *target* syllables (which correspond to the stimuli used in subsequent perception studies), /ma/, /na/, and /ŋa/, were created with equal numbers of *distractor* syllables—/mi/, /ni/, /nu/, /ba/, /pa/, /ga/ to prevent repetition fatigue. Also included in the materials list were multiple repetitions of the nasal coda syllables /am/, /an/, and /aŋ/, along with the distractor syllables /im/, /in/, /un/, /ab/, /ap/, and /ag/. Each target syllable (syllable-onset nasal) occurred 34 times, with distractor syllables concluding each list to avoid analysis of target syllables subject to list effects. The entire list contained 400 monosyllables including nasal-onset distractors, and nasal-coda syllables. The syllables were written in Roman orthography using conventional Filipino spelling, e.g., "ma," "na," "nga," "ma," "an," "ang." The /a/ context was chosen as it is the most frequent post-nasal vocalic context in spontaneous Filipino speech (Fig. 1). The three target syllables are meaningful in Filipino: *ma* indicates the possessor of a quality, *na* means "now" or "already," and *nga* is an emphatic particle meaning "indeed." (Rubino, 1998).

3.1.2. Speakers and setup

Three female native speakers of standard Filipino were recorded. All three speakers were born in the Philippines and spoke Filipino as a first language. The speakers were either students or staff at the University of British Columbia, Vancouver, at the time of the recording. Their length of stay in Canada ranged from 3 to 10 years. The speaker who had lived in Canada the longest reported frequent visits to the Philippines. Although all three speakers were fluent in Canadian English, they all continued to use Filipino at home and with friends. Two speakers also reported fluency in the Hokkien dialect of Cantonese, and one speaker grew up speaking both Ilocano (Austronesian) and Filipino. None of the speakers reported any speech or hearing problems and all were phonetically naïve. Speakers were told that they were recording random Filipino monosyllables and none was aware of the precise nature of the study. All three speakers were recorded individually in different recording sessions in a sound-proof booth (Industrial Acoustics Company, Inc.) at the Interdisciplinary Speech Research Laboratory in the Department of Linguistics at the University of British Columbia. Speech materials were recorded through an external pre-amplifier directly onto the hard drive of an Apple Macintosh G4 computer (housed in an adjacent room) using the SoundEdit audio program with an AKG condenser microphone approximately 8" from the speaker's lips. Materials were recorded in mono at a sampling rate of 44.1 kHz and transferred to CD-R as .wav files for analysis.

Speakers were asked to produce the monosyllables at a comfortable speaking rate. The experimenter sat in the recording booth with the speaker, prompting the subject, if necessary, to repeat tokens. Each recording session lasted approximately half an hour (with an additional half hour for explanation and completion of a language background questionnaire) for which speakers were paid $20 CDN.

3.1.3. Acoustic analysis

Only syllable-onset nasal tokens were considered for analysis. Speech materials were analyzed using the Praat (v. 4.2.22 and 4.3.31) (Boersma & Weenink, 2005) software package for Microsoft Windows. Both spectral and temporal characteristics of the target nasal-onset tokens were measured. The temporal measures were nasal murmur duration and vowel duration. The spectral measures were F1, F2, and F3 taken at the juncture between nasal murmur and vowel. The higher-frequency spectral resonances and antiresonances of the nasal murmur, which undoubtedly contribute to the manner percept but whose specific place characteristics are highly variable in production (Fujimura, 1962), were not measured in the present analysis. Rather, the spectral measurements were focused on the NV juncture, which were expected to show clear effects of place of articulation.

The duration of the entire /Na/ syllable was measured by placing cursors at the beginning and end of the token based on waveform and wide-band spectrographic displays. The onset of the syllable was taken as the first zero-crossing of the first periodic pulse of the nasal murmur. The offset of the nasal murmur and onset of the vowel was taken at the zero crossing of the first high-amplitude period of the post-nasal vowel. The location of the NV juncture was confirmed by examining the spectrographic display which showed an abrupt increase in spectral energy at higher formants (F2 and F3) indicating the acoustic onset of the vowel. The placement of the nasal offset cursor in the waveform display corresponded to the onset of high-energy formant patterning in the spectrographic display. An additional visual cue to the transition from nasal murmur to the vowel appeared in the waveform display, with the vocalic portion being marked by a more complex periodicity than the nasal murmur. In some cases, the increase in amplitude between the nasal murmur and vowel was gradual, thus precluding determination of the NV juncture based on the usual abrupt amplitude jump between the two regions. In those cases, the difference between the shape of the murmur and vowel waveforms was the determining factor in cursor placement. As the offset of the vowel was often accompanied by high-frequency breathy noise, the offset cursor was placed at the offset of periodicity of the vowel based on both the wideband spectrographic display as well as the waveform display.

The formant measurements (F1, F2, and F3) were based on FFT spectra calculated over a 10 ms Hamming window centered at the NV juncture and on a dB/Hz scale. The 10 ms window length was chosen to ensure that only frequency information local to the NV juncture (and not from nasal-murmur pulses or the adjacent vowel) was captured in the analysis. In ambiguous cases, the cursor placement in the FFT display was compared to the LPC analysis (with a prediction order of 10) using the Burg algorithm with pre-emphasis. The LPC comparison was used in less than 5% of the token measurements.

In addition to the static formant measurements, two dynamic measurements were calculated. Following Kurowski and Blumstein (1987), the final two glottal pulses of the nasal murmur and the first two pulses of the post-nasal vowel were identified. A rectangular window with a length of two pulses was centered over both regions and rms energy was measured in two critical bands, Bark 5–7 (corresponding to 395–770 Hz) and Bark 11–14 (1265–2310 Hz). The proportion change in rms energy between the nasal murmur and vowel in both bands was computed for analysis. Kurowski and Blumstein (1987) identified changes in Bark 5–7 and Bark 11–14 bands as characterizing labial and alveolar nasals, respectively. Whether these regions were optimal for velar nasal classification was unknown.

3.2. Results

Approximately, 26–32 tokens per nasal place per speaker were included in the analysis. Large variances were found in the temporal measures. This wide spread in murmur and vowel durations may be attributed to across-speaker differences in speech rate. As the frequencies of F2 and F3 (and their relationship) at the CV juncture are known to cue place of articulation (Delattre et al., 1955; Stevens, 1989; Stevens & Keyser, 1989), these measurements were an *a priori* focus of interest in the analysis. The descriptive statistics of the measurements pooled across speakers are given in Table 1.

Measurements were submitted to a linear mixed model analysis,[4] which showed significant effects of Place ([ma], [na], [ŋa]) on murmur duration [$F(2, 261) = 5.96$, $p < 0.005$], F1 at the NV juncture [$F(2, 261) = 29.63$,

---

[4]The linear mixed model controls for the likely correlation of repeated measures on the same speaker (or listener) via the inclusion of random subject effects. This model is used in the analysis of repeated acoustic measures from speakers in Experiment 1 and repeated perceptual measures in the subsequent experiments.

Table 1
Descriptive statistics for acoustic measurements pooled across three speakers

Variable	Place	Mean	SD	SE	Min	Max
Murmur duration (ms)	[ma]	123.20	31.00	3.30	53.41	206.75
	[na]	120.63	32.93	3.47	46.37	213.33
	[ŋa]	134.93	44.19	4.71	38.17	238.84
Vowel duration (ms)	[ma]	305.08	91.86	9.79	130.28	447.95
	[na]	312.18	91.90	9.68	115.53	465.79
	[ŋa]	315.59	91.79	9.78	121.90	470.47
F1 (Hz)	[ma]	430.20	85.58	9.12	273.48	644.65
	[na]	365.83	93.03	9.80	196.58	730.11
	[ŋa]	354.81	94.95	10.12	200.45	686.64
F2 (Hz)	[ma]	1269.31	110.12	11.73	966.39	1527.59
	[na]	1800.48	147.86	15.58	1510.26	2284.48
	[ŋa]	1833.22	170.54	18.18	1517.15	2248.51
F3 (Hz)	[ma]	2778.43	232.62	24.79	2339.71	3261.25
	[na]	3013.45	211.38	22.28	2549.63	3456.81
	[ŋa]	2655.12	168.70	17.98	2291.08	3150.57

$p < 0.001$], F2 at the NV juncture [$F(2, 261) = 479.14$, $p < 0.001$], and F3 at the NV juncture [$F(2, 261) = 137.41$, $p < 0.001$]. These results were further explored in a series of multiple comparisons, the results of which are given in Table 2.

There is no clear evidence in the literature that speakers use murmur duration as an acoustic cue to nasal place, though the current data suggest as much. Indeed, the effect size of Place on murmur duration is small (0.34 for the [na]-[ŋa] difference, and 0.31 for the [ma]-[ŋa] difference).[5]

Descriptions of individual data are given in Appendix A. These data were analyzed in separate mixed models. The mean F1 at the NV juncture was significant for speakers 1 and 2 [$F(2, 76) = 14.31$ and $F(2, 88) = 26.51$, respectively, both $p < 0.001$], with [ma] having a higher F1 than [na] and [ŋa] according to Bonferroni-adjusted pairwise comparisons. The mean F2 at the NV juncture was significant for all three speakers [$F(2, 76) = 173.74$, $F(2, 88) = 240.31$, $F(2, 93) = 181.91$ respectively, $p < 0.001$ for all speakers], with [na] and [ŋa] having a higher F1 than [ma]. Mean F3 at the NV juncture was significant for all three speakers [$F(2, 76) = 31.86$, $F(2, 88) = 84.49$, $F(2, 93) = 58.88$, $p < 0.001$ for all speakers]. For speaker 1, F3 in [na] tokens was significantly higher than in either [ma] ($p < 0.001$) or [ŋa] ($p < 0.001$) tokens. F3 in [ma] tokens was higher than in [ŋa] tokens ($p < 0.05$). For speaker 2, F3 for all three places of articulation were different from one another (all $p < 0.001$). For speaker 3, F3 in [na] tokens were significantly higher than in both [ma] and [ŋa] ($p < 0.001$ for both). Murmur duration was significant only for speaker 3 [$F(2, 93) = 11.36$, $p < 0.001$] whose [ŋa] tokens had longer murmurs than in either [ma] or [ŋa] (both $p < 0.005$).

Individual formant values likely play less of a role in the perception of place than do combination metrics, such as the relation between F3 and F2 (Stevens, 1989). The scatterplot in Fig. 2 captures this relation. These data were analyzed using a discriminant function analysis (DFA). DFA essentially performs a regression for a categorical outcome. It builds a predictive model of group membership based on observed characteristics of each case. The analysis generates a set of discriminant functions based on linear combinations of the predictor variables that allow for the best separation between groups. The discriminant functions are determined from a sample of cases for which group membership is known. These functions are then applied to new cases with measurements for the predictor variables but unknown group membership. DFA of the F2 × F3 relationship resulted in correct classification of 89.5% of the tokens.

---

[5]The standardized effect size associated with mixed models reflects the magnitude of difference between the two means divided by the square root of the sum of the variance components (one due to error and the other due to the effect of different speakers).

Table 2
Results of Bonferroni-adjusted pairwise comparisons

Variable	Token	Comparison	Sig.
Murmur duration	[ma]	[na]	1.0
	[ma]	[ŋa]	**0.023**
	[na]	[ŋa]	**0.004**
F1	[ma]	[na]	**0.001**
	[ma]	[ŋa]	**0.001**
	[na]	[ŋa]	0.759
F2	[ma]	[na]	**0.001**
	[ma]	[ŋa]	**0.001**
	[na]	[ŋa]	0.311
F3	[ma]	[na]	**0.001**
	[ma]	[ŋa]	**0.001**
	[na]	[ŋa]	**0.001**

Significant differences are given in bold.
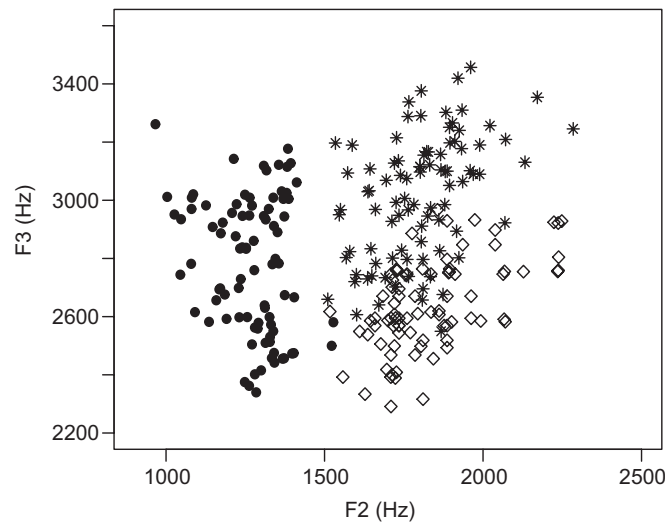
Fig. 2. F2 × F3 scatterplot for three nasal places of articulation measured at the NV juncture. Closed circles represent labial tokens; stars represent alveolars, and diamonds velars. Data represent tokens from three native speakers of Filipino.

The proportion in rms energy change from the murmur to vowel in Bark 5–7 and 11–14 is plotted in Fig. 3. The rms change measurements resulted in an overall correct classification of 78.1%.

When both static (F2 and F3 at NV juncture) and dynamic measurements (change in rms energy) were submitted to the DFA, 91.6% of the tokens were correctly classified. The confusion matrices for the static, dynamic and combined (static and dynamic) analyses are given in Table 3.

The static measurements provide near-perfect classification of [ma] tokens, with most of the errors occurring in classifying [na] tokens as [ŋa]. Interestingly, the DFA predicts [ŋa] better than [na]. This may be due the lower number of [na] outliers along the F3 dimension. Similarly, Kurowski and Blumstein's (1987) dynamic measurement was optimal for the [ma]-[na] distinction, but again showed considerable confusion between alveolars and velars. The combined static and dynamic measurements resulted in the best overall classification scores and very good classification of [ma] and [ŋa] tokens. As with both the static and dynamic measurements, the most errors occurred in [na] being classified as [ŋa].
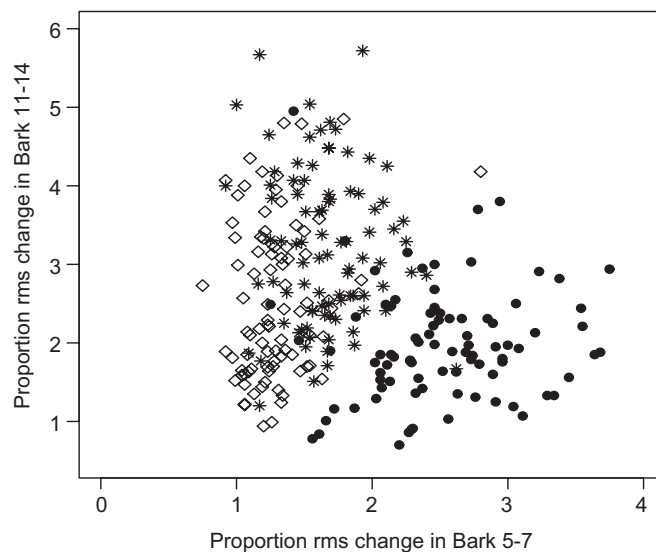
Fig. 3. Scatterplot showing the relationship between proportion rms energy change from nasal murmur to post-nasal vowel in Bark 5–7 and 11–14 for three places of articulation. Closed circles represent labial tokens; stars represent alveolars, and diamonds velars. Data represent tokens from three native speakers of Filipino.

Table 3
Confusion matrices produced by discriminant function analysis for classifying [ma], [na], and [ŋa] according to static (F2 and F3 at the NV juncture), dynamic (change in rms energy from murmur to vowel in Bark 5–7 and 11–14), and a combined (dynamic and static measurements) metric

Token	Predicted group		
	[ma] (%)	[na] (%)	[ŋa] (%)
Static			
[ma]	97.7	0	2.3
[na]	1.1	80.0	18.9
[ŋa]	1.1	8.0	90.9
Dynamic			
[ma]	86.6	4.9	8.5
[na]	4.4	70	25.6
[ŋa]	1.1	20.5	78.4
Combined			
[ma]	96.3	1.2	2.4
[na]	1.2	83.5	15.3
[ŋa]	0	4.9	95.1

3.3. Summary and discussion

Experiment 1 examined three syllable-onset nasal places of articulation from Filipino along multiple temporal and spectral dimensions. When static spectral characteristics were considered, information unique to the three places of articulation emerged. For all speakers, the [ma]-[na]-[ŋa] distinction is differentiated by F2 and F3 at the NV juncture. To the extent that these place categories are not fully distinct, the relationship between F2 and F3 as displayed in an F2 × F3 space (Fig. 2) and between the murmur and vowel in Bark 5–7 and 11–14 (Fig. 3) shows that [na] tokens overlap more with [ŋa] tokens than with [ma] tokens. Moreover, the DFA of both static and change-over-time measures shows that the most misclassifications occur between [na] and [ŋa] tokens. These results are consistent with analyses of *oral* obstruent-vowel sequences from American English. Kewley-Port (1982) found a similar grouping of [da] and [ga] tokens in an F2 × F3 space as measured

at the onset of the vowel transition. Consistent with Harrington (1994), the best place classification for the nasal syllables was achieved using combined static and dynamic measurements.

The acoustic distance between [ma], [na], and [ŋa] tokens may well have a perceptual correlate. Given that the F2 × F3 acoustic space is relevant for place perception (Larkey et al., 1978; Liberman et al., 1954; Stevens, 1989), we might predict that the perception of phonetic categories with more overlap in F2 × F3 space would be measurably different from categories which show more separation in this acoustic space. This prediction is tested in Experiment 2, which is the first of two experiments designed to assess the perception of /ma/, /na/, and /ŋa/ by adult listeners whose phonology reflects the three-way contrast (Filipino) and listeners for whom the /na/-/ŋa/ contrast is non-native (English).

4. Experiment 2: cross-language perception of nasal place

Experiment 1 showed that acoustic measures differentiate [ma], [na], and [ŋa], but in an F2 × F3 space as well as the proportion rms energy change between the murmur and vowel, [na] and [ŋa] tokens overlap resulting in imperfect separation of the two place categories in discriminant analyses. This pattern, characterized by robust differentiation between [ma] and [na], and less robust differentiation between [na] and [ŋa], parallels the typological distribution described in the introduction, namely, when languages of the world contrast two nasal places of articulation in syllable-onset position, they most often do so at the bilabial and alveolar/dental places. The asymmetrical nature of the distribution of syllable-onset nasals in Filipino, coupled with their asymmetric acoustic distribution, suggests that a parallel asymmetry will be evident in perception as well.

Experiment 2 asks first whether English-speaking adults, whose phonology does not contrast syllable-onset /n/-/ŋ/, show poor perception of the non-native contrast—a "non-native effect" reflecting heightened perception of only those contrasts which are native (in this case /m/-/n/). Second, this experiment aims to establish whether Filipino speakers will discriminate nasal place consistently with the acoustic distinctions that were described in Experiment 1, thus showing poorer perception of /n/-/ŋ/ than /m/-/n/ on the basis of the relative distance between tokens of the phones in F2 x F3 space. Both of these questions are here assessed in a single cross-language perception experiment in which adult native speakers of either English or Filipino are presented with the contrasts [ma]-[na] and [na]-[ŋa] in a 2AFC (AX) discrimination task.

4.1. Predictions

The predictions are that (1) both English and Filipino listeners will perform well on the native [ma]-[na] contrast, (2) English listeners will show an effect of language experience with poor performance on the non-native [na]-[ŋa] contrast (or the "non-native" effect consistent with Larkey et al. (1978) and Narayan (2004)), and (3) Filipino listeners will show a perceptual consequence of the acoustic distance of [na] and [ŋa] with somewhat depressed performance on the [na]-[ŋa] contrast relative to their performance on [ma]-[na] (consistent with Krull, 1990).

4.2. Acoustic stimuli

Real-speech stimuli were selected from tokens recorded by a single native Filipino speaker described in Experiment 1. Four tokens of each nasal place ([ma], [na], and [ŋa]) were selected based on their showing minimal within-category variation. The resulting 12 [Na] tokens had similar durations of nasal murmur and vowel as well as similar values for F1 at the onset of the post-nasal vowel transition. Variation in the vocalic portion (as measured at the vowel midpoint) of the stimuli was avoided to the extent possible in order to preclude discrimination based on vowel quality. In addition, tokens were selected that differed minimally in the f_0 of the nasal murmur and vowel as well as the overall contour. The acoustic characteristics of the 12 stimuli are given in Appendix B. Formant measures were taken at vowel onset and midpoint as indicated.

An ANOVA revealed a significant effect of Place on F2 [$F(2, 11) = 70.45$, $p < 0.001$] and F3 [$F(2, 11) = 5.56$, $p < 0.05$] at the NV juncture. F2 frequencies were significantly higher for [na] (mean = 1787.83 Hz) than for [ma] (mean = 1335.04 Hz) stimuli ($p < 0.001$). The difference in F2 between [na] and [ŋa] (Mean = 1718.00 Hz)

was not significant, while the difference in F3 between [na] (Mean = 2680.46 Hz) and [ŋa] (Mean = 2507.06 Hz) approached significance ($p = 0.08$). No other acoustic measurements that were taken (i.e., temporal measures, or spectral measures taken at vowel midpoint) were significantly different between [na] and [ma] or [na] and [ŋa]. These results are very similar to the more general pattern of /Na/ acoustics of the productions of the three speakers of Filipino given in Experiment 1. The stimuli were then impressionistically assessed by three phonetically trained listeners to ensure that the tokens sounded similar in terms of overall pitch contour, duration and extra-linguistic factors such as breathiness or creakiness.

4.3. Methods

4.3.1. Trials

AX trials consisted of two types of *non-identical* pairs, different pairs and within-category pairs, for both place-of-articulation contrasts [ma]-[na] and [na]-[ŋa]. For the different (across-category AB) pairs, each of the four tokens from each place category was paired with each of the four tokens from a contrasting category (except [ma]-[ŋa] pairs were excluded). For example: $na_1\eta a_1$, $na_2\eta a_1$, $na_3\eta a_1$, $na_4\eta a_1$; $na_1\eta a_2$, $na_2\eta a_2$..., and so on, for a total of 192 different pairs. Orders were counterbalanced in both different and within-category pairs. For within-category pairs, each token was paired with an acoustically different, but phonemically *same* variant ($A_x A_y$), for a total of 108 within-category pairs. Because within-category pairs did not consist of acoustically identical pairings, listeners would have to generalize across small acoustic differences in order to make categorical decisions.

The experiment was broken up into three blocks. Each block had 100 trials, which consisted of one presentation of each possible pairing. The first eight trials of only the first block were "familiarization" trials meant to acquaint the listener to the sounds and the task. Responses to these trials were not analyzed. The inter-stimulus interval (ISI) was set at 1500 ms to facilitate a "phonemic level" of discrimination of non-native contrasts (Pisoni, 1973; Werker & Tees, 1984).

4.3.2. Participants

Fifteen native Canadian-English-speaking listeners and 13 native Filipino-speaking listeners were recruited from the Department of Psychology subject pool at the University of British Columbia. Listeners received either course credit or $10 CDN for their participation.

Each participant was administered a language-background questionnaire before beginning the experiment. Canadian-English speakers were chosen to participate only if they had not had any formal training in languages from East or Southeast Asia. Many South East Asian languages and some dialects of Cantonese exhibit the velar nasal in syllable-onset position. In order to avoid any listener bias, Canadian-English speakers who were familiar with other East and South East Asian languages were not chosen to participate.

Similarly, only Filipino-speaking listeners who had spent their early years in the Philippines and continued to regularly speak Filipino with friends and family were chosen to participate. All the Filipino participants were immigrants to Canada and some reported fluency in other Austronesian languages of the Philippines (such as Ilocano, Bikolano, Cebuano, Kapampangan, etc.) as well as Spanish and French.[6] The mean length of residence in Canada for the Filipino speakers was 4.92 (SD = 3.42) years. None of the participants had received linguistics or phonetics training prior to the experiment.

4.3.3. Setup

The perception experiment was conducted using the DMDX (Forster & Forster, 2003) presentation software running on a Windows XP laptop computer in a quiet room at the Infant Studies Centre (UBC). Stimuli were presented through a Sony STR-DE197 amplifier over two free-field Bose (Model 101) speakers placed at $30°$ to the right and left of the centerline of the participant at 70 ± 2 dB

---

[6]That the Filipino-speaking participants in this and subsequent studies were often bilingual with other Philippine Austronesian languages was not problematic for the purposes of these experiments, as the syllable-onset velar nasal occurs in all of these languages (Himmelmann, 2005).

SPL.[7] Participants were tested individually and sat in front of the laptop computer, which registered responses from clicks of right and left mouse buttons marked "S" (for same) and "D" (for different). Participants held the mouse on their lap with their right hand (the dominant hand of all the participants) and made responses with their index and middle fingers. Trials advanced automatically when the participant made a button press. There was a one-second delay between the offset of the button press and the onset of the next trial.

Instructions for the English-speaking listeners were given on the computer screen in English. Participants were told that they would be hearing pairs of speech sounds and asked to press the "S" button if the two sounds were from the same category and "D" if they were from different categories. Category membership was explained with "pa," "ta," and "ka" as examples of different categories. Participants were also asked to respond to the trials as quickly as possible. No other instructions were given to the listeners. The English instructions were translated into Filipino and presented online to the Filipino-speaking participants. Although the Filipino speakers were necessarily fluent in Canadian English, experimental instructions were presented in Filipino in order to increase the likelihood that they performed the task according to L1 rather than L2 strategies.

Participants were told that the first 8 trials of the first block were familiarization trials meant to acquaint them with the speech sounds of the experiment and the response apparatus. Listeners did not receive any feedback during familiarization or the experimental phases. There was a short break after the familiarization in case the participant had any questions regarding the administration of the experiment. Listeners were given two short (2–5 min) self-regulated breaks after the first and second blocks. The entire experiment lasted approximately 45 min.

4.4. Results

4.4.1. Proportion correct

Results from all 15 English listeners were included in the analysis. The results of one Filipino listener were discarded due to a misunderstanding of the instructions, leaving data from 12 Filipino listeners in the analysis. Listeners' responses were converted to proportion correct scores for the [ma]-[na] and [na]-[ŋa] contrasts. These results show the predicted pattern of discrimination performance on both the native and non-native contrasts. English listeners accurately discriminated the native [ma]-[na] contrast at an average rate of 98.8% correct. Their performance on the non-native [na]-[ŋa] contrast, however, was slightly below chance at an average of 45.9% correct with considerably larger variances. The Filipino-speaking listeners on average showed the expected native level of performance, with an average of 98.8% correct on the [ma]-[na] contrast and 90.8% correct on [na]-[ŋa].

Data were submitted to a linear mixed model analysis with the fixed factors Language (English and Filipino), Contrast ([ma]-[na] and [na]-[ŋa]) and Block (presentation block 1, 2, and 3). There was a significant main effect of Language [$F(1, 150) = 67.21$, $p < 0.001$], with Filipino listeners performing more accurately, overall, than English listeners. The main effect of Contrast was significant [$F(1, 150) = 122.98$, $p < 0.001$] showing listeners performing better on the [ma]-[na] contrast than the [na]-[ŋa] contrast. The effects of Language and Contrast can only be interpreted in terms of their significant interaction [$F(1, 150) = 67.84$, $p < 0.001$]. The main effect of Block was not significant [$F(2, 150) = 1.475$].

Fig. 4 shows the language × contrast interaction by plotting the average proportion correct on both contrasts for the two language groups. The mixed model showed that Filipino listeners were significantly more accurate than the English listeners on the [na]-[ŋa] contrast [$F(1, 79) = 68.47$, $p < 0.001$]. The language group difference for the [ma]-[na] contrast was not significant [$F(1, 79) = 0.039$]. Thus, the only difference between English and Filipino listeners was in their discrimination of the [na]-[ŋa] contrast.

Within each language group, English listeners performed significantly better on the [ma]-[na] contrast than on the [na]-[ŋa] contrast [$F(1, 88) = 126.12$, $p < 0.001$]. Similarly, within the Filipino-listener group performance was significantly better on [ma]-[na] than on [na]-[ŋa] [$F(1, 70) = 24.0$, $p < 0.001$].

---

[7]Stimuli were presented "free field," rather than with the more common headphone presentation, for consistency with follow-up experiments (not reported here) with infants aged 4–12 months.
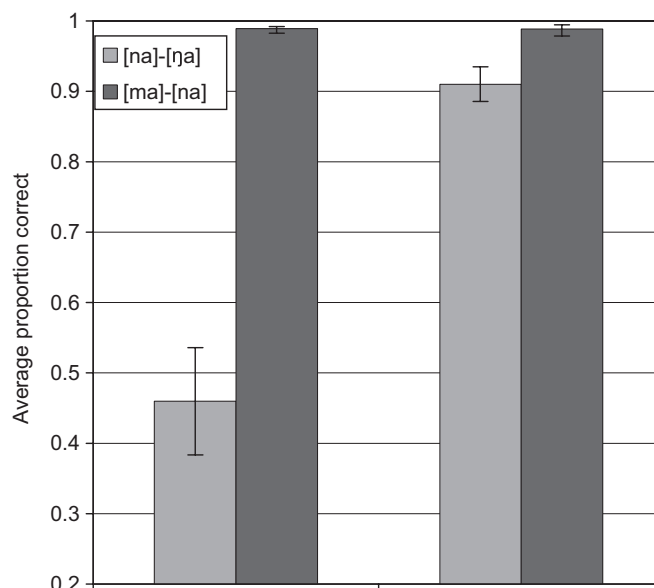
Fig. 4. Proportion correct scores for cross-language AX discrimination test of [ma]-[na] and [na]-[ŋa] contrasts by native English (left two bars) and native Filipino listeners (right two bars).

4.4.2. Sensitivity (d')[8]

In addition to proportion correct scores, listeners' responses were converted to d', a signal detection theoretic statistic that provides a bias-free measure of sensitivity to given contrasts (Macmillan & Creelman, 2005; Swets, 1996). In general, the higher the d' value the more sensitive the listener is to the contrast (i.e., the further apart in perceptual space are the two phonetic categories). The d' value is calculated from *hit* and *false-alarm* rates of the individual for a given contrast. The d' measure is useful in that it provides the researcher with a statistic based on listeners' perception of *both* same (within category) and different (between category) pairings.

As with the proportion correct results, listeners' d' scores were fit to a linear mixed model with dependent variable d' and fixed factors of Language, Contrast and Block. There were significant main effects of Language [$F(1, 147) = 46.87, p < 0.001$] and Contrast [$F(1, 147) = 135.61, p < 0.001$] in the expected directions. There was a significant interaction between Language and Contrast [$F(1, 147) = 36.06, p < 0.001$]. Interestingly the main effect of Block also reached significance [$F(2, 147) = 4.83, p < 0.01$] showing increased sensitivity to the contrasts as the experiment progressed. No interactions with Block reached significance.

Fig. 5 shows the interaction between Language and Contrast. Between language groups, Filipino listeners were more sensitive to the [na]-[ŋa] contrast than were English listeners [$F(1, 79) = 54.98, p < 0.001$]. Within the English group, listeners were more sensitive to the [ma]-[na] contrast than to the [na]-[ŋa] contrast [$F(1, 88) = 175.10, p < 0.001$]. Within the Filipino group, listeners were similarly more sensitive to the [ma]-[na] contrast than to the [na]-[ŋa] contrast [$F(1, 67) = 13.71, p < 0.001$].

The main effect of Block on d' scores is plotted in Fig. 6. The plot shows listeners' increasing sensitivity (pooled across contrast and language background) as the experiment progressed.

4.5. Summary and discussion

Overall, the patterns of results from this real-speech cross-language AX experiment were consistent with results of previous investigation into English listeners' perception of synthetic nasal consonants (Larkey et al.,

---

[8]Standard Signal Detection Theory assumes two types of AX discrimination designs, fixed and roving. In general, the roving design (or differencing model) is assumed when more than two stimulus classes are presented in the experiment. In the present task, listeners heard three classes, [ma], [na], and [ŋa], in any given block, therefore the roving design was used to calculate d' (see Macmillan & Creelman, 2005 for a full discussion of fixed versus roving models of discrimination). Macmillan and Creelman provide a d' table (based on false-alarm and hit rates) assuming the roving model. These d' values were found for each contrast in each block for every participant and used in the analysis.
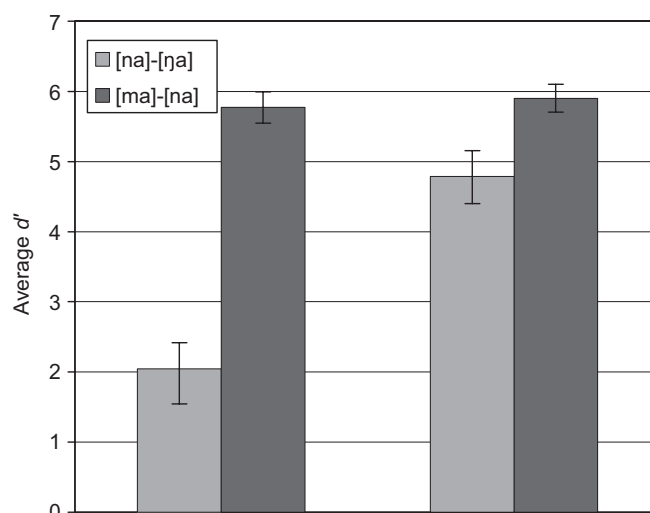
Fig. 5. Sensitivity measures for cross-language AX discrimination test of [ma]-[na] and [na]-[ŋa] contrasts by native English (left) and native Filipino listeners (right). Sensitivity (d′) calculated according to the roving model (see Macmillan & Creelman, 2005, p. 177).
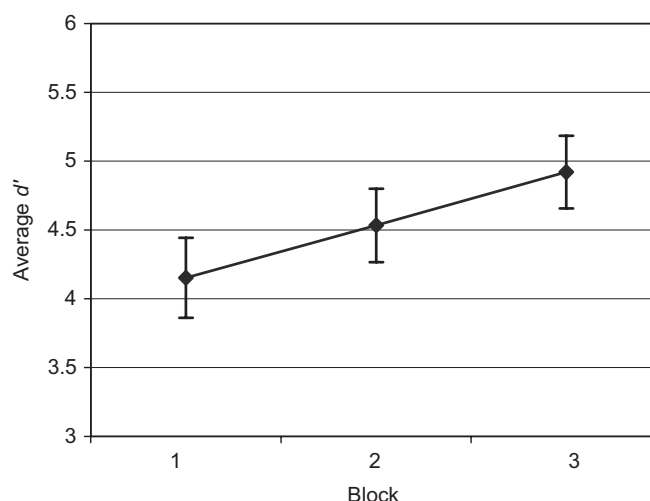

Fig. 6. Block analysis of sensitivity to nasal contrasts overall.

1978) and English and Cantonese-listeners' perception of Cantonese nasals (Narayan, 2004). English-speaking listeners' performance on the [na]-[ŋa] contrast was diminished relative to their performance on the native contrasts. Not surprisingly, Filipino-speaking listeners showed a perceptual pattern that reflected Filipino phonology, with highly accurate discrimination of both the [ma]-[na] and [na]-[ŋa] contrasts. These results are consistent with the prediction made in 4.1, that listeners' perception of nasal place contrasts is influenced by their linguistic experience.

The more interesting result, however, was that both English and Filipino listeners were similar in their overall pattern of responses with discrimination of the [na]-[ŋa] contrast significantly poorer than the [ma]-[na] contrast. That Filipino listeners, for whom both contrasts are phonologically distinct, showed a pattern of relative discriminability parallel to that of the English listeners suggests that factors other than the nativeness of the contrast affects perception (cf. Burnham, 1986; Best, McRoberts, & Goodell, 2001). An explanation advanced for this pattern of perception for Filipino listeners is that the [na]-[ŋa] contrast is perceptually less salient than the [ma]-[na] contrast. This reduced perceptual salience is argued to result from the acoustic distances in F2 × F3 space among the three contrasts shown in the results of Experiment 1. The statistical confusion between [na] and [ŋa], as evidenced by the poorer separation of the two categories in a DFA, is

suggestive of the behavioral consequences in the present experiment. Another potential contributor to Filipino listeners' accuracy in discriminating [na]-[ŋa] is the relative infrequency of /ŋ/ in spontaneous Filipino speech.

The sensitivity analysis revealed an interesting pattern not predicted, namely that listeners' perception of both [ma]-[na] and [na]-[ŋa] improved with exposure. This learning, which was not evident in the proportion correct analysis, shows that the native and non-native listener can partially overcome perceptual biases with increased listening experience.

5. Experiment 3: discrimination in noise

The effect of the acoustic similarity between tokens of [na] and [ŋa] in Experiment 2 was small but significant for Filipino speakers. Experiment 3 was designed to assess the validity of the perceptual salience interpretation of Experiment 2 by intensifying the perceptual effect of the acoustic similarity between /n/ and /ŋ/. The slightly less accurate performance on the [na]-[ŋa] contrast by adult Filipino listeners arguably results from the contrast being acoustically less robust than the [ma]-[na] contrast and consequently perceptually less salient. The adult English-listener pattern can be attributed to a combination of language experience and acoustic–perceptual salience. English listeners' perception of the Filipino [na]-[ŋa] contrast is therefore consistent with numerous studies of non-native speech perception, but the effect of acoustic–perceptual salience cannot be disentangled from the effect of experience.

The current experiment tests relative *strength* of the two contrasts by presenting them under degraded listening conditions. As listening conditions worsen, there should be asymmetrical effects on the perception of the two contrasts, with the more robust contrast being less vulnerable to interference. All else being equal, a greater effect of a noisy listening condition on [na]-[ŋa] than [ma]-[na] would support the claim that the [na]-[ŋa] contrast is perceptually less salient than the [ma]-[na] contrast and implicate a role for perceptual salience as an influencing factor in distribution of nasal place in the world's languages and within Filipino.

As most speech is produced and comprehended in less than ideal listening environments, the introduction of noise to the signal has a long history of use in experimental phonetics for mimicking this aspect of naturalistic listening situations. Noise has varying effects on the perception of speech and interacts with phonetic features such as place of articulation and manner (for a review see Benkí, 2003), as well as with phonetic context. In their classic study of consonant confusions in masking white noise, Miller and Nicely (1955) found that correct identification of place of articulation is greatly reduced by broadband noise in a closed-set identification task. Listeners' confusions were most often found in place of articulation: as the signal-to-noise ratio decreased, place-of-articulation errors in identification increased. Similar results were found in other studies assessing the effects of noise on consonant identification (Benkí, 2003; Wang & Bilger, 1973). Alwan, Lo, and Zhu (1999) examined the effects of broadband Gaussian noise on the perception of nasal onsets in English ([m] and [n]). They found that listeners' identification accuracy in /a/ contexts was high (>80%) in very noisy (−10 dB SNR) conditions, but poor in /i/ contexts even in 5 dB SNR conditions. Alwan et al. found that formant transitions into the vowel are more important for identification when the post-nasal vowel is /a/ compared to other vocalic contexts.

As in Experiment 2, the present experiment assesses the salience of *contrast*, rather than the inherent perceptual salience of a particular token as in the identification studies of Miller and Nicely (1955), Wang and Bilger (1973), Alwan et al. (1999), or Benkí (2003). The purpose here is to explore whether or not certain contrasts are in fact more resistant to adverse listening conditions than others, and in doing so, provide a better understanding of the ubiquity of those perceptually robust phonetic oppositions.

5.1. Methods

5.1.1. Stimuli
The speech stimuli were the same 12 tokens (four tokens for each place) used in Experiment 2. The experiment consisted of three listening conditions: one "clean" or unaltered condition and two "noisy" conditions differing in signal-to-noise ratio. Signal-dependent uncorrelated[9] noise was added offline to the clean stimuli using a specially written script for the Matlab environment following Schoeder (1968). A product

---

[9]Signal-dependent uncorrelated noise has been shown to have perceptual effects similar to broadband noise (Benkí, 2003).

of the signal (*s*), the SNR scaling factor and a randomly generated ±1 were added to every sample of the signal as given in

$$s + \left(\left(\sqrt{\frac{1}{10^{\text{SNR}/10}}} \right) s \right) (\pm 1).$$

Special care was taken in determining the appropriate signal-to-noise ratios, which would potentially affect the perception of the two contrasts. To that end, pilot studies with native Filipino listeners showed that performance on [na]-[ŋa] (as measured in proportion correct) slightly decreased at 0 dB SNR (approximately 85% correct), yet did not decrease further at −2 dB SNR. Increasing the noise level in the signal to −5 dB SNR had the most dramatic effect in the pilot studies and so that level was chosen as the noisiest listening condition used in the experiment.

The three listening conditions were assembled into blocks: clean (with no additive noise), 0 and −5 dB SNR. Four tokens from each category were paired with four tokens from each different category as well as non-identical tokens from the same category. This assembly yielded 100 pairs (of pair types AB, BA, AA, and BB). As "same" pairs never consisted of identical tokens (always A_xA_y, never A_xA_x) this approach yielded a stimulus set that was unbalanced with more "different" pairs (64) than "same" pairs (36). In order to balance the set, an additional 28 "same" pairs were added to the list of stimuli, yielding 64 "same" pairs and 64 "different" pairs, or a total of 128 trials. Each trial occurred twice for a total of 256 trials per block (blocked by listening condition: clean, −0, −5 dB SNR). Consistent with Experiment 2, the ISI was set at 1500 ms.

5.1.2. Participants

Eleven native speakers of Filipino (different from those in Experiment 2) participated in Experiment 3. Participants were recruited through email solicitation to University of Michigan Filipino student associations. All participants completed a language background questionnaire in order to assess the languages they spoke. Five participants reported extensive knowledge of one other Austronesian language of the Philippines such as Ilocano or Cebuano. Only respondents who were fluent in Filipino and reported that they used it on a regular basis were included in the study. All of the participants had spent the majority of their life in the Philippines, having lived in the US for an average of 5.2 years. One participant had been in the US for 15 years, but reported visiting the Philippines for at least 6 months out of each year for the past 5 years. All of the participants were fluent in English and reported that English was the medium of instruction beginning in late elementary school. None of the participants knew the purpose of the experiment and all were phonetically naive. None of the participants reported any speech or hearing problems.

5.1.3. Procedure

The experiment was conducted in a sound-attenuated room in the Phonetics Laboratory at the University of Michigan. Participants were seated in front of an Apple *iBook* laptop computer from which stimuli were presented and responses recorded using a specially designed presentation instrument for the Praat (v.4.3.31) (Boersma & Weenink, 2005) software package. Participants listened to stimuli over high-quality Sennheiser HD 625 headphones at a comfortable volume level. All participants were tested individually.

It has been suggested in the literature that using the listeners' native language for the language of instruction may help them approach the experimental task in a native listening mode (Beddor & Gottfried, 1995; Jenkins, 1979). For this reason, before the experiment began each listener spoke briefly with a native Filipino speaker (the Filipino instructor at the University of Michigan) in Filipino. These conversations lasted approximately 5 min. As in Experiment 2, written instructions were presented to participants in Standard Filipino. The instructions as well as on-line computer display labels of "same" and "different" were given in Filipino in order to maximize the likelihood that listeners were in a "Filipino mode" of listening. These instructions asked the listeners to imagine they were in the Philippines participating in a speech experiment where they would be hearing pairs of speech sounds from Filipino. Examples of "same" and "different" categories were similar to those given in Experiment 2. If the sounds were from the same category in their native language, they were to click the button "same" on the computer screen; if the sounds were from different categories in their native language, they were to click the "different" button. The experiment was self-paced, with trials advancing 1 s
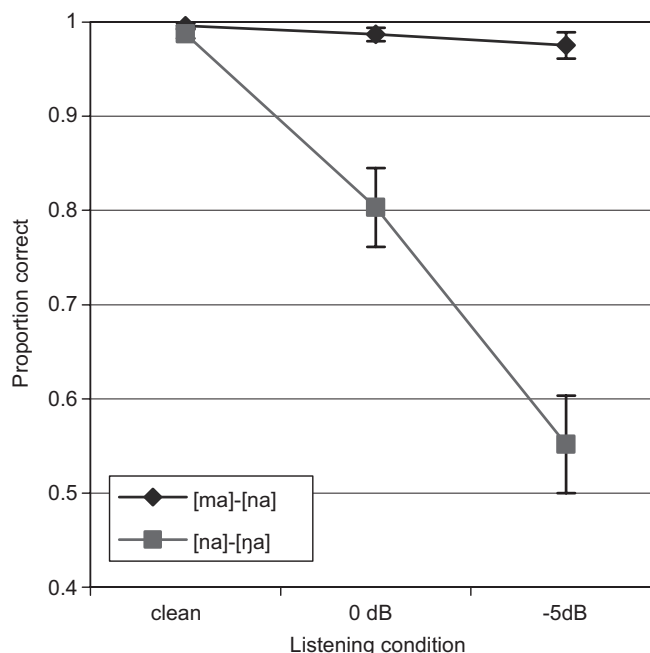
Fig. 7. Proportion correct to [ma]-[na] and [na]-[ŋa] discrimination by native Filipino listeners in three listening conditions.

after a response was made. Listeners were instructed to respond as quickly as possible in order to ensure timely completion of the experiment.

Every listener was tested with the "clean" (non-noisy) stimuli in the first block. The noisy blocks (0 and −5 dB SNR) followed, with their order being counterbalanced across participants. The order of the three listening conditions was not randomized in order to avoid effects of noisy perception on the perception of clean trials. After every 64 trials, there was a break in presentation and participants were allowed to take a self-regulated break. After each block the listener was allowed to take a 5 min break as regulated by the experimenter.

5.2. Results

The results of ten listeners were analyzed. The results of one listener were excluded from analysis due to a misunderstanding of the instructions. Listeners' performance on the AX experiment is presented in Fig. 7 as a function of their proportion correct on the two types of *different* trials, [ma]-[na] and [na]-[ŋa], according to listening conditions.[10] These results were submitted to a linear mixed model analysis with factors: Contrast ([ma]-[na] and [na]-[ŋa]), Listening condition (clean, 0 db, and −5 dB SNR) and Block order (0 dB block first or −5 dB block first).

The analysis revealed significant main effects of Contrast [$F(1, 114) = 164.99$, $p < 0.001$] and Listening condition [$F(2, 114) = 59.57$, $p < 0.001$]. There was no significant effect of Block order. In addition, there was a significant interaction between Contrast and Listening condition [$F(2, 114) = 55.45$, $p < 0.001$]. Within the clean listening condition there was a small, but nonetheless significant difference between the proportion correct on the two contrasts with listeners performing better on the [ma]-[na] contrast than on the [na]-[ŋa] contrast [$\text{Mean}_{[ma]-[na]} = 0.998$, $\text{Mean}_{[na]-[ŋa]} = 0.987$; $F(1,38) = 6.06$, $p = 0.019$].[11] The effect of Contrast was

---

[10]Listeners' responses were also converted to a bias-free measure of sensitivity, d'. The d' results for the current data were similar to the proportion correct results, that is, the same general pattern was observed, with sensitivity to the [na]-[ŋa] contrast dropping dramatically with increasing noise with high sensitivity to [ma]-[na] across all three listening conditions.

[11]The effect size for the difference in the clean condition (computed as the difference in means divided by the square root of the sum of the variance components) was 0.7 suggesting a medium to large effect. Indeed, one-sample t-tests (two-tailed) revealed that discrimination of the [ma]-[na] contrast was no different from 100%, while performance on [na]-[ŋa] was significantly lower than ceiling [$t(19) = −2.9$, $p < 0.01$].
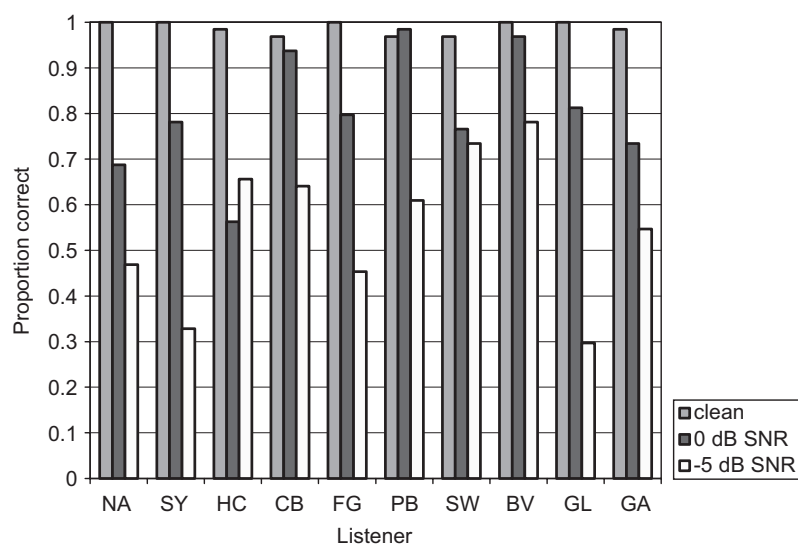
Fig. 8. Individual performance on [ma]-[na] and [na]-[ŋa] discrimination in three listening conditions.

dramatically greater in the 0 dB [Mean$_{[ma]-[na]}$ = 0.995, Mean$_{[na]-[ŋa]}$ = 0.803; $F(1,38)$ = 38.56, $p<0.001$] and −5 dB [Mean$_{[ma]-[na]}$ = 0.990, Mean$_{[na]-[ŋa]}$ = 0.551; $F(1,38)$ = 126.68, $p<0.001$] listening conditions.

Within the [ma]-[na] contrast, there was no significant effect of Listening condition. Listening condition had a significant effect on the [na]-[ŋa] contrast [$F(2, 57)$ = 58.31; $p<0.001$], with proportion correct dropping from 98% in the clean condition to 80% in the 0 dB SNR condition. Accuracy dropped to 55% in the −5 dB SNR condition. This effect was investigated in a series of multiple comparisons. Bonferroni-adjusted comparisons showed significant differences between clean and 0 dB SNR listening conditions ($p<0.001$) and between 0 and −5 dB SNR listening conditions ($p<0.001$).

In general, with the addition of signal-dependent noise, listeners' perception of the [na]-[ŋa] contrast became progressively worse. However, this was not the trend for all of the listeners. Fig. 8 shows individual performance on the [na]-[ŋa] contrast in the three listening conditions.

Note that equal amounts of noise to signal (0 dB SNR) had little effect on [na]-[ŋa] discrimination for listeners CB, PB, and BV. Moreover, for these listeners as well as HC and SW, perception remains above chance (50% correct) level even in the noisiest conditions. Two speakers (SY and GL) performed at substantially below chance levels in the noisiest condition. Below-chance performance is difficult to interpret but suggests that listeners were reasonably confident that different stimuli were actually the same.

5.3. Summary

The crucial finding from Experiment 3 is the interaction between Contrast and Listening condition, with decreases in SNR affecting the perception of the [na]-[ŋa] contrast and *not* the [ma]-[na] contrast. Filipino listeners, for whom both contrasts are native, performed at near ceiling levels on both contrasts in the clean condition. Consistent with the results of Experiment 2, there was a small but significant difference between listeners' perception of the two contrasts even in the clean condition, although mean difference is substantially smaller than in Experiment 2. It is expected that the difference in performance between the two experiments was caused by procedural factors. As the stimuli in Experiment 2 were presented in a free-field fashion, it is likely that the closed headphone presentation in the present study was less susceptible to the effects of ambient noise.

The more important finding is that listeners' perception of [ma]-[na] remained highly accurate (almost perfect) even in the most adverse listening conditions, while their perception of [na]-[ŋa] degraded to chance performance with increasing noise. Why would the presence of noise in the signal disproportionately affect the [na]-[ŋa] contrast? The most obvious reason is found in the acoustics of these two syllables. Alwan et al. (1999)

showed that in noisy conditions listeners rely on transitional information in categorizing /m/ and /n/ syllables. Alwan et al. (1999) Given that [na] and [ŋa] are more similar in F2 × F3 space (as shown by the DFA in Section 3.2), we might attribute the source of the perceptual asymmetry in the present study to the acoustic distance between the tokens. A related and equally likely reason for the disproportionate affect of noise on [na]-[ŋa] contrast is that the critical acoustic cue, namely F3, was masked due its being low in intensity relative to the lower formants.[12] As F3 is generally lower in intensity than F2, the nasal murmur, which undoubtedly nasalizes the initial portion of the vocalic transition, additionally serves to attenuate F3 more than it does F2 (cf. Ohala, 1975). Taken together, the F3 cue to velar place is masked more thoroughly in noisy conditions than the F2 cues discriminating labial and alveolar places.

Another possible factor in adult Filipino listeners' enhanced perception of the [ma]-[na] contrast is that all of the listeners in present study were bilingual (L1 = Filipino, L2 = English). One might argue that listeners in the present study are used to hearing more occurrences of the [ma]-[na] contrast in ambient speech as they all reside in an English-speaking country. This L2 experience might be viewed as being *added* to native language experience, resulting in heightened sensitivity to the more frequent contrast. Although listeners were primed to be attentive to Filipino through the initial conversation with a native Filipino speaker as well as the written experiment instructions, the possible influence of English cannot be discounted.

The most striking potential non-acoustic factor contributing to the Filipino listeners' depressed discrimination of the [na]-[ŋa] in noisy conditions is that of token frequency. When the results of the spontaneous speech corpus analysis in Section 2.1 are considered one could argue that the depressed performance of listeners on the [na]-[ŋa] contrast in noise results from the infrequency of syllable-initial /ŋ/ in Filipino speech.

5.4. Experiment 4: acoustic similarity or token frequency?

Experiment 4 follows from two possible interpretations of the results of Experiments 2 and 3. The argument thus far suggests that typological patterns in nasal place distribution are influenced by the acoustic structure of nasals and their perceptual consequences. In light of the corpus analysis of spontaneous Filipino speech in 2 where tokens of syllable-initial /ŋ/ are shown to occur considerably less than /m/ or /n/, the results of Experiments 2 and 3 can also be interpreted as suggestive of a frequency effect. There is some evidence that token frequency affects adult speech perception. Newman, Sawusch, and Luce (1999) show that lexical neighborhood effects on speech perception are subordinate to the effects of high-frequency phonemes, while lower-frequency phonemes show effects of lexical neighborhoods. That is, listeners were biased towards identifying an ambiguous speech sound according to its phoneme frequency rather than its lexical neighborhood when the phoneme frequency was high. Indeed, the effects of frequency are central to exemplar or instance-based theories in phonology, providing a mathematical model for capturing features of speech perception and phonetic learning (e.g., Johnson, 1996; Pierrehumbert, 2001a, 2001b). Given the potential for frequency effects on perception, the relative infrequency of syllable onset /ŋ/ in spontaneous speech as well as the lexicon perhaps affects perception in the direction of a weakened sensitivity to /ŋ/ in the context of the more frequent /n/. Experiment 4 serves to disambiguate these two interpretations, acoustic salience and token frequency, by introducing the [ma]-[ŋa] contrast for discrimination in noise.

The acoustic distance between /n/ and /ŋ/ is argued to contribute to the weakened salience of the [na]-[ŋa] contrast relative to the [ma]-[na] contrast whose components are acoustically more distant. Given this reasoning, if perception of native nasal contrasts in noise is affected more by acoustic distance than frequency, the prediction for the present experiment is similar discrimination performance on both the [ma]-[na] and [ma]-[ŋa] contrasts, with diminishing performance on [na]-[ŋa] with increasing noise in the signal. However, the prediction offered from an acoustic distance argument cannot be separated from those made by an argument that broadband noise disproportionately affects the relatively low intensity F3 at the NV juncture. The intensity argument would predict that a noise masked F3 essentially renders the [ma]-[ŋa] contrast perceptually equivalent to the [ma]-[na].

---

[12] I thank an anonymous reviewer for pointing out frequency-specific masking effects.

5.4.1. Methods

5.4.1.1. Stimuli. The stimuli were identical to those used in Experiments 2 and 3, with the addition of a set of [ma]-[ŋa]/[ŋa]-[ma] pairings. As in the previous experiments, AX trials consisted of two types of non-identical pairs, different and within-category pairs, for three place contrasts [ma]-[na], [na]-[ŋa], and [ma]-[ŋa]. The assembly was identical to that in Experiment 3, yielding a total of 300 pairs, 192 different and 108 same pairs. In order to balance the set, an additional 25 same pairs from each category was added to the block, yielding a total of 375 pairs, 192 different and 183 same pairs. As in Experiment 3, the experiment was divided into three blocks of 375 trials: clean, 0 and −5 dB SNR.

5.4.1.2. Participants and procedure. Eleven participants (different from those in Experiments 2 and 3) were recruited via solicitation to University of Pennsylvania Philippine cultural associations. As in Experiment 3, all of the participants were native speakers of a Philippine Austronesian language (such as Tagalog, Cebuano, etc.). Participants had lived in the United States for an average of 2.92 years (SD = 2.8 years) and reported speaking Filipino on a daily basis with friends and family. None of the participants reported speech or hearing problems. Participants were paid $15 US for the hour-long experiment.

The experiment was conducted individually or in groups of two in the Speech Perception Laboratory in the Institute for Research in Cognitive Science at the University of Pennsylvania. Each participant sat at a MS Windows machine where the experiment was presented via Praat (v. 4.5.08). Participants listened to stimuli at a comfortable volume over Sennheiser HD 280pro headphones. The experimental procedures were identical to those in Experiment 3, though participants did not initially speak with a native Filipino speaker before beginning the experiment.

5.4.2. Results

The results of 10 participants were used in the analysis. The results of one participant were not collected due to an experimental error. Fig. 9 shows participants' discrimination (percent correct)[13] performance on different pairs in three listening conditions. Results were submitted to a linear mixed model analysis (Block order: 0 dB SNR first, -5 dB SNR first; Contrast: [ma]-[na], [na]-[ŋa], [ma]-[ŋa]; Listening condition: clean, 0 and −5 dB SNR).

The main effects of Contrast and Listening condition were significant [$F(2, 72) = 105.38$, $p < 0.001$; $F(2, 72) = 51.70$, $p < 0.001$, respectively]. There was no effect of Block order. The main effects can only be understood in light of their significant interaction [$F(4, 72) = 44.79$, $p < 0.001$]. Within the clean listening condition there was no significant effect of Contrast [$\text{Mean}_{[ma]-[na]} = 0.99$, $\text{Mean}_{[na]-[ŋa]} = 0.971$, $\text{Mean}_{[ma]-[ŋa]} = 0.983$], though the effect was approaching significance [$F(2, 27) = 2.71$, $p = 0.084$]. The effect of Contrast was dramatically greater in the 0 dB [$\text{Mean}_{[ma]-[na]} = 0.989$, $\text{Mean}_{[na]-[ŋa]} = 0.859$, $\text{Mean}_{[ma]-[ŋa]} = 0.983$; $F(2,27) = 18.78$, $p < 0.001$] and −5 dB [$\text{Mean}_{[ma]-[na]} = 0.989$, $\text{Mean}_{[na]-[ŋa]} = 0.551$, $\text{Mean}_{[ma]-[ŋa]} = 0.975$; $F(2,27) = 102.25$, $p < 0.001$] listening conditions. Bonferroni-adjusted multiple comparisons revealed no significant difference between the [ma]-[na] and [ma]-[ŋa] contrasts any of the three listening conditions, while the [na]-[ŋa] differed significantly from both [ma]-[na] and [ma]-[ŋa] in the 0 and -5 dB SNR conditions ($p < 0.001$ in all comparisons).

5.4.3. Summary

The suggestion that the token frequency of /ŋ/ affects the perception of [na]-[ŋa] is tempered by the results of the present experiment because listeners' accuracy on the [ma]-[ŋa] contrast is not different from their performance on [ma]-[na]. If acoustic distance was not an influencing factor, the perception of [ma]-[ŋa] in noisy conditions would have been more similar to [na]-[ŋa] than [ma]-[na]. Overall, the results of the perception in noise experiments follow from (1) an acoustic distance interpretation, where [na]-[ŋa] is predicted to be poorly discriminated in noisy contexts due their acoustic similarity in F2 × F3 space, and (2) an F3

---

[13]As in Experiment 3, d' scores were computed for the results of Experiment 4. The d' pattern was no different from the percent correct pattern, with sensitivity to the [na]-[ŋa] contrast falling with increasing noise and sensitivity to both [ma]-[na] and [ma]-[ŋa] remaining high in all three listening conditions.
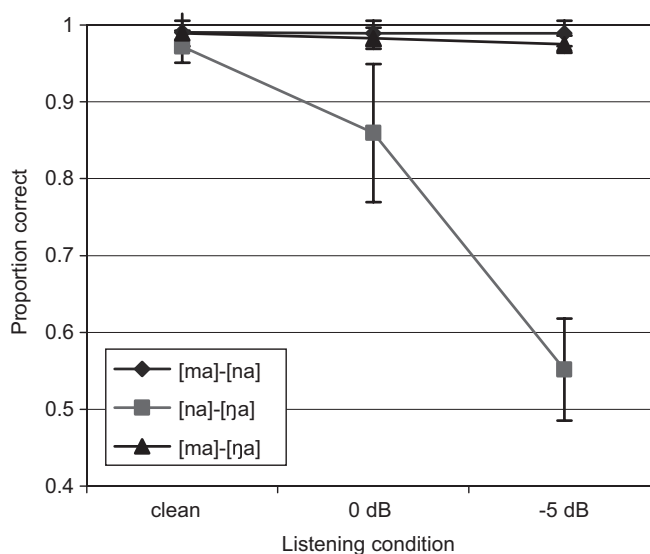
Fig. 9. Proportion correct to [ma]-[na], [na]-[ŋa], and [ma]-[ŋa] discrimination by native Filipino listeners in three listening conditions.

intensity interpretation, where broadband noise disproportionately masks the F3 place cue for velars rendering listeners to rely on the more intense F2 cue.

6. Conclusion and general discussion

When we consider the synchronic typology of phonological systems it becomes apparent that languages exhibit more of certain types of contrasts than others. The present work asks why this might be the case. The literature on phonetic influences in phonology points towards the effects of acoustics and corresponding perceptual patterns as providing an answer. The hypothesis following from this literature is that certain contrasts are better represented in the world's languages in part because they are perceptually robust, while contrasts that are less represented are less perceptually salient or perceptually weaker. In order to test this hypothesis, I examined the acoustics and corresponding perception of syllable-onset nasal contrasts in Filipino, a language that exhibits both the common distinction (/m/-/n/) as well as the less common distinction (/n/-/ŋ/).

Experiment 1 revealed that the acoustics of [ma], [na], and [ŋa] as spoken by three native Filipino speakers showed an asymmetry reminiscent of the typological asymmetry for nasal place detailed in 2. The acoustic groupings (in an F2 × F3 space) were skewed in the direction of a greater [ma]-[na] and [ma]-[ŋa] distance than [na]-[ŋa] distance. The perceptual consequence of these relative distances was tested in Experiment 2. It was hypothesized that native listeners would be more accurate at discriminating the acoustically distant categories than they would the more similar categories. As predicted, Filipino listeners perceived both contrasts very well (>90% correct) but were better at discriminating the [ma]-[na] contrast. English listeners performed as well as Filipino listeners on the [ma]-[na] contrast (native to English as well), but near chance levels on the non-native contrast as expected. The small (90% versus 98%) but significant difference in Filipino listeners' perception of the two contrasts was predicted to be affected by a more realistic listening situation. In Experiment 3, Filipino listeners were presented the same two contrasts but in three listening conditions: no noise, equal amounts of signal to noise (0 dB SNR), and slightly more noise than signal (−5 dB SNR). The [ma]-[na] contrast was perceived near ceiling in all three listening conditions. Listeners' perception of [na]-[ŋa] fell precipitously as noise was added to the signal. Experiment 4 followed-up the results of Experiment 3 by introducing a third contrast, [ma]-[ŋa] and found no evidence that results from Experiments 2 and 3 could be explained by the low frequency of onset /ŋ/ in Filipino. Listeners discriminated the acoustically distant [ma]-[ŋa] contrast as well as they did the acoustically distant [ma]-[na], while performance on the acoustically similar [na]-[ŋa] fell in noisy conditions.

Taken together, these results suggest that the acoustic distance between syllable onset /m/ and /n/ contribute to a robust percept of their contrast. The acoustic similarity between /n/ and /ŋ/ renders a weaker perceptual contrast relative to /m/-/n/ and /m/-/ŋ/. These results are consistent with the argument that sound systems of the world's languages tend towards a balance in the features that they contrast, such that the acoustic outcome leads to minimal perceptual confusion (Flemming, 2006)—from a dispersion theory point of view, the acoustic spaces occupied by /n/ and /m/ are maximally distinct, contributing to their ubiquity in sound systems. But why do languages generally contrast onset /m/-/n/ rather than onset /m/-/ŋ/? The results of Experiment 4 would suggest that both contrasts are equally salient. We might speculate that the answer lies in a more general phonological understanding of coronality and markedness (see Paradis and Prunet, 1991). As coronals (in this case the dental/alveolar nasal) pattern across languages as the unmarked place of articulation, the most functionally robust contrast would be with the acoustically distant /m/.

Given the potential consequences of acoustically similar segments in a phonological inventory (such as merger), an important question that follows from this conclusion is how a language like Filipino maintains a phonological contrast between acoustically similar phones. Proto-Austronesian is reconstructed with three nasal places (*n,*m, and *ŋ) in syllable initial position. This pattern remains robust in many of its daughter languages but *n and *ŋ have merged in Thao (Blust, 2003), Malagasy (Dziwirek, 1989), Tetun (Timor) (Morris, 1984), Hawaiian, Tahitian, and the languages of Vanuatu (Blust, pc). The less "destructive" consequence would be to phonologically restrict the distribution of the otherwise confusable contrast to syllable positions that are perceptually enhancing. For example, stop voicing contrasts are often neutralized in word-final position (e.g., German) perhaps because VOT cues are unavailable for perception (Steriade, 1997). The least harmful, and perhaps the most productive solution to the distance problem would be to not encode perceptually enhancing restrictions in the phonology, but rather restrict the occurrence of the confusable contrasts in the lexicon. The Filipino lexicon reveals such a pattern, with the majority of syllable onset velar nasals appearing word medially, leaving relatively few (<100) lexical items with word-initial /ŋ/ (Rubino, 1998). Otherwise, acoustically confusable nasal contrasts in VNV position provides the listener with added cues (in the closure acoustics of the pre-nasal V and the onset acoustics of the post-nasal V) for identification. Further, the restriction on word-initial /ŋ/ appears to operate in spontaneous speech as well. When the corpus presented in 2.1 was analyzed according to position within the word, we find that 30% of the token occurrences of [ŋa] in spontaneous speech are located in word-initial position, while 76% of [na]s and 65% of [ma]s are word-initial. Thus, languages that show phonetic oppositions that are predicted to be "difficult" in terms of their perceptual salience also show a regulated distribution of the sub-optimal contrast that is restricted to those environments that augment its perceptibility.

Given the results of this study, we might make predictions about the perceptual salience relations among other typologically uncommon contrasts in the world's languages. For example, we might expect speakers of a language with a three-way backness contrast in high vowels (/i/-/y/-/u/, 94% of front vowels are unrounded in the UPSID so an /i/-/y/ contrast is rare) to be better at discriminating the acoustically distant (in F2 space) /i/-/u/ contrast in adverse listening conditions than /i/-/y/. If we extend this idea beyond segmental contrasts by including suprasegmentals, we might expect speakers of a densely populated tone language (like Cantonese) (Yip, 2002) to be better at discriminating contour tones of opposite direction (e.g., rising versus falling) than tones that are acoustically more similar and typologically less common in contrast (e.g., low-level versus mid-level).

Another question this work raises is how the first language learner accommodates acoustically similar sounds in the ambient language. Are contrasts that are typologically rare as a result of perceptual salience relations (like those discussed in this paper) perceived as well as acoustically distinct and perceptually robust contrasts? Do infants perceive an onset /n/-/ŋ/ contrast as well as they might a /m/-/n/ contrast? Recent work addressing this line of research suggests that the infant language learner is initially biased towards discriminating acoustically robust nasal contrasts like onset /m/-/n/ and requires native language experience to discriminate an acoustically similar contrast like /n/-/ŋ/ (Narayan, 2006).

Appendix A

Descriptions of individual data are given in Table A1.

Table A1
Acoustic measurements for individual speakers

Speaker	Place	F1 (Hz)	F2 (Hz)	F3 (Hz)	Murmur duration (ms)
1	[ma]	396.65 (94.55)	1240.07 (127.21)	2896.70 (167.19)	98.74 (24.37)
	[na]	309.99 (72.54)	1905.53 (149.89)	3119.72 (147.43)	89.58 (21.86)
	[ɳa]	296.14 (45.18)	1996.18 (195.74)	2778.28 (160.23)	89.61 (24.87)
2	[ma]	402.14 (75.30)	1260.99 (103.98)	2936.39 (154.46)	135.18 (21.14)
	[na]	326.66 (38.99)	1784.75 (117.93)	3149.57 (124.82)	132.29 (29.48)
	[ɳa]	304.02 (42.86)	1756.92 (88.91)	2670.78 (156.49)	139.63 (27.04)
3	[ma]	483.76 (58.09)	1300.87 (95.19)	2534.25 (97.90)	131.83 (32.89)
	[na]	450.90 (85.11)	1727.08 (122.65)	2791.93 (126.33)	135.54 (25.82)
	[ɳa]	450.12 (84.71)	1772.35 (107.72)	2540.37 (98.51)	167.34 (38.32)

Spectral measures were made at the NV juncture. Standard deviations are given in parentheses.

Appendix B

The acoustic characteristics of the 12 stimuli are given in Table B1.

Table B1
Nasal stimuli acoustics

	Token	Total dur	Nas dur	V dur	Total f0	V f0	F1 onset	F2 onset	F3 onset	F1 midpt	F2 midpt	F3 midpt
[ma]	1	505.39	132.98	373.09	215.42	211.88	508.21	1301.13	2327.72	930.64	1473.78	2675.71
	2	505.74	116.60	389.13	213.77	209.77	467.427	1337.13	2572.99	994.59	1469.71	2648.50
	3	486.05	135.73	350.32	209.89	206.83	559.53	1339.20	2564.08	949.72	1461.60	2684.64
	4	484.18	128.42	355.75	210.62	205.90	505.48	1362.71	2439.68	882.30	1405.78	2778.79
	Avg.	495.34	128.43	367.07	212.43	208.60	510.16	**1335.04**	**2476.12**	939.31	1452.72	2696.91
	SD	*11.83*	*8.44*	*17.62*	*2.61*	*2.74*	*37.81*	*25.41*	*116.15*	*46.51*	*31.70*	*56.71*
[na]	1	487.94	134.45	353.49	212.46	207.84	425.16	1721.44	2778.33	741.25	1549.05	2606.51
	2	499.00	112.61	386.39	206.20	203.94	355.68	1749.20	2624.78	979.64	1507.00	2714.03
	3	462.38	118.89	343.49	214.38	210.93	349.28	1868.20	2674.7	1020.63	1539.65	2746.57
	4	500.15	119.28	380.87	209.47	207.84	373.19	1812.48	2644.05	969.56	1534.62	2703.76
	Avg.	487.37	121.31	366.06	210.63	207.64	375.83	**1787.83**	**2680.47**	927.77	1532.58	2692.72
	SD	*17.54*	*9.28*	*20.82*	*3.58*	*2.86*	*34.41*	*65.74*	*68.40*	*126.29*	*18.07*	*60.30*

Table B1 (continued)

	Token	Total dur	Nas dur	V dur	Total f0	V f0	F1 onset	F2 onset	F3 onset	F1 midpt	F2 midpt	F3 midpt
[ŋa]	1	489.73	121.26	377.46	208.23	204.52	461.57	1756.14	2539.93	922.39	1458.28	2557.80
	2	487.84	112.93	374.91	209.87	208.24	422.58	1675.19	2576.39	900.49	1519.81	2761.29
	3	510.90	128.39	382.51	206.90	204.69	417.18	1642.35	2375.92	981.77	1637.98	2969.69
	4	461.85	120.37	341.49	209.60	207.27	261.07	1798.33	2536	834.71	1606.22	2741.86
	Avg.	487.58	120.74	369.09	208.65	206.18	390.60	**1718.00**	**2507.06**	909.84	1555.57	2757.66
	SD	*20.09*	*6.32*	*18.67*	*1.37*	*1.86*	*88.59*	*71.79*	*89.30*	*60.73*	*81.86*	*168.49*

Bold averages represent critical spectral dimensions distinguishing nasal place of articulation. Duration measurements are in ms and spectral measurements in Hz.

References

Alwan, A., Lo, J., & Zhu, Q. (1999). *Human and machine recognition of nasal consonants in noise*. Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco.

Anderson, G. D. S. (2005). The velar nasal (N). In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world Atlas of language structures*. Oxford: Oxford University Press.

Beddor, P. S., & Gottfried, T. L. (1995). Methodological issues in adult cross-language research. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press.

Benkí, J. R. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica*, *60*, 129–157.

Best, C. T., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775–794.

Bladon, R. A., & Lindblom, B. (1981). Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America*, *69*, 1414–1422.

Blevins, J. (2006). *Phonetically-based sound patterns: Typological tendencies or phonological universals*? Paper presented at the LabPhon10, Paris.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, *66*, 1001–1017.

Blust, R. (2003). *Thao dictionary. Language and linguistics monograph series no. 5*. Taipei: Institute of Linguistics (Preparatory Office), Academica Sinica.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.3.31).

Bubenik, V. (1996). *The Structure and Development of Middle Indo-Aryan Dialects*. Delhi: Motilal Banarsidass.

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, *7*, 207–240.

Carlson, R., Granström, B., & Klatt, D. H. (1979). *Vowel perception: The relative perceptual salience of selected acoustic manipulations* (*No. STL − QPSR* 3 − 4/1979). Stockholm, Sweden: Royal Institute of Technology.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gesterman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, *24*(6), 597–606.

Cotsomrong, P., Sunpetchniyom, T., Kasuriya, S., Thatphithakku, N., & Wutiwiwatchai, C. (2005). LOTUS: Large vocabulary Thai continuous speech recognition corpus. Presented at *NSTDA annual conference S&T in Thailand: Towards the molecular economy* (*NAC2005*). Transcripts available from National Electronic and Computer Center at ⟨http://vaja.nectec.or.th:8083/lotus/download_pre.jsp⟩.

Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language, Vol. 2: Phonology* (pp. 93–152). Stanford: Stanford University Press.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, *27*(4), 769–773.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*(2), 121–144.

Disner, S. F. (1983). *Vowel quality: The relation between universal and language-specific factors*. UCLA Working Papers in Phonetics, 58.

Dziwirek, K. (1989). Malagasy phonology and morphology. *Linguistic Notes from La Jolla 15*, 1–30.

Ferguson, C. A. (1963). Assumptions about nasals: A sample study in phonological universals. In J. H. Greenberg (Ed.), *Universals of language* (pp. 53–60). Cambridge, MA: MIT Press.

Flemming, E. (2001). *Auditory representations in phonology*. New York: Garland.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-based phonology*. Cambridge: Cambridge University Press.

Flemming, E. (2006). *The role of distinctiveness constraints in phonology*. Cambridge, MA, unpublished manuscript.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers*, *35*, 116–124.

Fujimura, O. (1962). Analysis of nasal consonants. *Journal of the Acoustical Society of America*, *34*, 1865–1875.

Guion, S. (1996). *Velar palatalization: Coraticulation, perception and sound change*. Unpublished doctoral dissertation, University of Texas.

Harnsberger, J. D. (2000). A cross-language study of the identification of non-native nasal consonants varying in place of articulation. *Journal of the Acoustical Society of America*, *108*(2), 764–783.

Harrington, J. (1994). The contribution of the murmur and vowel to the place of articulation distinction in nasal consonants. *Journal of the Acoustical Society of America*, *96*(1), 19–32.

Himmelmann, N. (2005). The Austronesian languages of Asia and Madagascar: Typological characteristics. In N. Himmelmann, & A. Adelaar (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 110–181). London: Routledge.

House, A. S. (1957). Analog studies of nasal consonants. *Journal of Speech and Hearing Disorders*, *22*, 190–204.

Hume, E., & Johnson, K. (2001). A model of the interplay of speech perception and phonology. In E. Hume, & K. Johnson (Eds.), *The role of speech perception in phonology*. San Diego: Academic Press.

Hura, S. L., Lindblom, B., & Diehl, R. L. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech*, *35*, 59–72.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.

Jakobson, R. (1963). Implications of language universals for linguistics. In *Selected writings*, *Vol. II. Word and language* (pp. 580–592). The Hague: Mouton.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak, & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, NJ: Erlbaum.

Johnson, K. (1996). Speech perception without speaker normalization. In K. Johnson, & J. Mullennix (Eds.), *Talker variability in speech processing*. San Diego: Academic Press.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, *69*, 505–528.

Johnson, K., Flemming, E., & Wright, R. (2004). Response to Whalen et al. *Language*, *80*(4), 646–648.

Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *Journal of the Acoustical Society of America*, *72*(2), 379–389.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*, 419–454.

Krull, D. (1990). Relating acoustic properties to perceptual responses: A study of Swedish voiced stops. *Journal of the Acoustical Society of America*, *88*(6), 2557–2570.

Kurowski, K. M., & Blumstein, S. E. (1987). Acoustic properties for place of articulation in nasal consonants. *Journal of the Acoustical Society of America*, *81*(6), 1917–1927.

Larkey, L. S., Wald, J., & Strange, W. (1978). Perception of synthetic nasal consonants in initial and final syllable position. *Perception and Psychophysics*, *23*(4), 299–312.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, *68*(379).

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality: The role of perceptual contrast. *Language*, *48*(4), 839–862.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala, & J. J. Jaeger (Eds.), *Experimental phonology*. Orlando: Academic Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Amsterdam: Kluwer.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed). New Jersey: Lawrence Erlbaum.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, I. (1986). The size and structure of phonological inventories: Analysis of UPSID. In J. J. Ohala, & J. J. Jaeger (Eds.), *Experimental phonology*. Orlando: Academic Press, Inc.

Malécot, A. (1956). Acoustic cues for nasal consonant. An experimental study involving a tape-splicing technique. *Language*, *32*(2), 274–284.

Masica, C. (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, *27*, 338–352.

Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., & 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*, 331–340.

Mohr, B., & Wang, W. (1968). Perceptual distance and the specification of phonological features. *Phonetica*, *18*, 31–45.

Morris, C. (1984). *Tetun-English Dictionary. Pacific Linguistics C-83*. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Nakata, K. (1959). Synthesis and perception of nasal consonants. *Journal of the Acoustical Society of America*, *31*(6), 661–666.

Narayan, C. R. (2004). A phonological asymmetry and its perceptual consequence: The case of the velar-nasal onset. *Journal of the Acoustical Society of America*, *116*(4), 2572 (Abstract).

Narayan, C. R. (2006). *Acoustic–perceptual salience and developmental speech perception*. Unpublished doctoral dissertation, University of Michigan.

Newman, R., Sawusch, J., & Luce, P. (1999). Underspecification and phoneme frequency in speech perception. In M. Broe, & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Language acquisition and the lexicon*. Cambridge, UK: Cambridge University Press.

Nord, L., & Sventelius, E. (1979). *Analysis and prediction of difference limen data for formant frequencies (No. STL/QPSR 3 − 4/1979)*. Stockholm, Sweden: Royal Institute of Technology.

Ohala, J. J. (1974). *Phonetic explanation in phonology*. Paper presented at the Parasession on Natural Phonology, Chicago Linguistics Society.

Ohala, J. J. (1975). *Phonetic explanation for nasal sound patterns*. Paper presented at The Nasalfest: Symposium on Nasals and Nasalization, Stanford.

Pandharipande, R. (2003). Marathi. In G. Cardona, & D. Jain (Eds.), *The Indo-Aryan languages*. London: Routledge.

Paradis, C., & Prunet, J.-F. (1991). Introduction: Asymmetry and visibility in consonant articulations. In C. Paradis, & J.-F. Prunet (Eds.), *The special status of coronals: Internal and external evidence*. San Diego: Academic Press.

Passy, P. (1890). *Études sur les changement phonétiques*. Paris: Firmin-Didot.

Pierrehumbert, J. (2001a). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. (2001b). Stochastic phonology. *Glot International*, *5*(6), 195–207.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, *13*(2), 253–260.

Qi, Y., & Fox, R. A. (1992). Analysis of nasal consonants using perceptual linear prediction. *Journal of the Acoustical Society of America*, *91*(3), 1718–1726.

Repp, B. H. (1986). Perception of the [m]-[n] distinction in CV syllables. *Journal of the Acoustical Society of America*, *79*(6), 1987–1999.

Rubino, C. R. G. (1998). *Tagalog-English, English-Tagalog Dictionary*. New York: Hippocrene Books.

Schachter, P., & Otanes, F. T. (1972). *Tagalog reference grammar*. Berkeley: University of California Press.

Schoeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America*, *44*, 1735–1736.

Seitz, P., McCormick, M., Watson, I., & Bladon, R. A. (1990). Relational spectral features for place of articulation in nasal consonants. *Journal of the Acoustical Society of America*, *87*, 351–358.

Smyth, D. (2002). *Thai: An essential grammar*. London: Routledge.

Steriade, D. (2001a). *The phonology of perceptibility effects: The P-map and its consequences for constraint organization*. Cambridge, MA, Unpublished manuscript.

Steriade, D. (2001b). Directional asymmetries in place assimilation: A perceptual account. In E. Hume, & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219–250). San Diego: Academic Press.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, K. N., & Keyser, S. J. (1989). Primary feature and their enhancement in consonants. *Language*, *65*(1), 81–106.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press.

Steriade, D. (1997). *Phonetics in phonology: the case of laryngeal neutralization*. Unpublished manuscript, Los Angeles.

Sussman, H. M., & Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception and Psychophysics*, *58*, 936–946.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America*, *54*, 1248–1266.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*(6), 1866–1878.

Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., & Iskarous, K. (2004). Vowel production and perception: Hyperarticulation without a hyperspace effect. *Language and Speech*, *47*(2), 155–174.

Wheeler, M. (2005). *The phonology of Catalan*. Oxford: Oxford University Press.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Zvelebil, K. (1970). *Comparative Dravidian phonology*. The Hague: Mouton.