The Acoustic and Visual Phonetic Basis of Place of Articulation in Excrescent Nasals.

Keith Johnson & Christian DiCanio
UC Berkeley

Laurel MacKenzie
University of Pennsylvania

Abstract. One common historical development in languages with distinctively nasalized vowels is the excrescence of coda velar nasals in place of nasalized vowels. For example, the dialect of French spoken in the southwestern part of France (Midi French) is characterized by words ending in the velar nasal [ŋ] where Parisian French has nasalized vowels and no final nasal consonant ([savɔ̃]~[savɔŋ] "soap"). More generally, there is a cross-linguistic tendency for the unmarked place of articulation for coda nasals, and perhaps also for stops, to be velar. In four experiments, we explored why the cross-linguistically unmarked place for the excrescent nasal is velar. The experiments test Ohala's (1975) acoustic explanation: that velar nasals, having no oral antiformants, are acoustically more similar to nasalized vowels than are bilabial or alveolar nasals. The experiments also tested an explanation based on the visual phonetics of nasalized vowels and velar nasals: velar nasals having no visible consonant articulation are visually more similar to nasalized vowels than are bilabial or alveolar nasals. American English listeners gave place of articulation judgments for audio-only and audio-visual tokens ending in nasal consonants or nasalized vowels. In the first and second experiments, we embedded recorded tokens of CVN (N = /m/, /n/, or /ŋ/) words in masking noise and presented them in audio-only and audio-visual trials. We also synthesized "placeless" nasals by repeating pitch periods from the nasalized vowel to replace the final consonant in CVm with nasalized vowel. These stimuli provide a direct test of Ohala's acoustic explanation of coda velarity in nasals. The third and fourth experiments extended these results with tokens in which the last portion of CVN (N = /m/, /n/, or /ŋ/) and Cṽ syllables were obscured with masking noise. These experiments were designed to force listeners to assume the existence of a final consonant and to rely primarily on visual cues in a more direct test of the visual similarity of nasalized vowels and velar nasals. Taken together, the results of these four experiments suggest that excrescent coda nasals tend to be velar because nasalized vowels are both acoustically and visually similar to velar nasals.

Introduction.

In this paper we will consider how acoustic- and visual-phonetic aspects of speech may impact on certain place of articulation phenomena in phonological patterning. The general aim of this research is to contribute to the laboratory study of historical sound changes and the synchronic sound patterns that arise in language history (Ohala, 1974, 1981). We present the results of four experiments that explore the perceptual basis of excrescent nasal velarity.

Ohala (1981) proposed that the listener is the source of sound change, because listeners parse incomplete or ambiguous acoustic signals to determine the linguistic intentions of the speaker, and this parsing process is sometimes incorrect. In this view, sound change occurs when a listener misperceives the speaker and hears a different intended pronunciation than the one that the speaker meant, and subsequently chooses to model his/her speech on the misperception. Similarly, mispronunciation due to gestural misalignment or overlap, or any other aspect of speech production can not result in a sound change unless a hearer interprets the mispronunciation as an intended utterance and chooses to model his/her own speech on the mispronunciation. In either case, the listener-turned-speaker is the crucial step in sound change. Because perceptual mistakes so often mirror sound change, it is possible to test specific listener-based hypotheses in the laboratory by creating ambiguities that we presume listeners also face in nature.

In this study, we were interested in whether this listener-based model of sound change, supplemented with attention to visual speech perception, could help explain a dialectal variation in French. In some words, where standard French has a nasalized vowel, the regional variety of French spoken around Toulouse (Midi French) sometimes has a vowel followed by a velar nasal [ŋ][1]. For example, Durand (1988) cites an alternation in Midi French between [n] and [ŋ] in [savɔne] "to soap up" and [savɔ̃ŋ] "soap". In standard French, "soap" is pronounced [savɔ̃], so the final [ŋ] in Midi French is called an excrescent [ŋ]. Posner (1997) also notes excrescent velar nasals in "southern" French, "where some nasal occlusion tends to be kept", citing pronunciations [pɛɛ̃ŋ] *pain* "bread", and [paãntø] *pente* "slope" in Aix-en-Provence.

1 MacKenzie (2006) found that the number of words that actually show the excrescent nasal in conversational speech is not large.

The puzzle which we wish to address is why the excrescent nasal is velar rather than alveolar or labial. One new hypothesis tested in this paper is that excrescent nasal velarity is partly due to visual phonetics. Seeing the face of the speaker influences perception of place of articulation, consequently if nasalized vowels are perceived as vowel-nasal sequences they will tend to be perceived with excrescent [ŋ] because, like vowels and unlike [m] or [n], velar nasals do not have a visible consonant closure. We also wanted to test Ohala's (1975) hypothesis that listeners will be influenced by acoustic similarity between nasalized vowels and velar nasals. These two effects taken together - acoustic similarity between nasalized vowels and velar nasals, and visual similarity between nasalized vowels and velar nasals - may lead listeners to hear a velar nasal at the offset of a nasalized vowel, thus leading to excrescent [ŋ] in Midi French.

Though having an account of excrescent [ŋ] is interesting, the hypothesis that we are testing has wider implications for sound changes and synchronic patterns in other languages because many languages exhibit sound changes and patterns of alternation that indicate a strong affinity between nasalized vowels and velar nasals. For example, Morais-Barbosa (1962, p. 692) described excrescent [ŋ] in variety of Portuguese that is very similar to Midi French. Standard Portuguese [lã] "wool" is sometimes pronounced [lãŋ] in this dialect. Paradis & Prunet (2000) note, that when Fula borrows French nasalized vowels an excrescent velar nasal is pronounced (1). Wiese (1996) notes a similar tendency when German borrows French words, e.g. [ʀɛstoʀã] > [ʀɛstoʀaŋ].

(1) French borrowings in Fula.

	French	Fula	
(a)	sɛrʒã	sarsaŋ	"sergeant"
(b)	ljøtnã	lijetinaŋ	"lieutenant"
(c)	kɔ̃ferãs	kɔŋfɛras	"conference"

These examples and others like them lead us to posit that there is a cross-linguistic phonological pattern of excrescent nasal velarity (2).

(2) Excrescent nasal velarity

ã > ãŋ

In additon to cases of excrescent velar nasals, a tendency for final nasals to become velar has been noted. For example, Trigo (1988) noted that final alveolar nasals are borrowed in Puerto Rican Spanish as velar, e.g. "train" [treŋ]. Andalusian Spanish has velar nasals where standard Spanish has dental nasals, e.g. "bread" [pan] > [paŋ], "gluttonous" [gloton] > [glotoŋ]. This tendency for final velarity may involve excrescent nasal velarity in the following way. There is a tendency for place of articulation information to be perceptually (Fujimura) and articualtorily (??) weak in final position. If a final nasal segment becomes placeless to some degree, with a nasalized vowel and a small or missing nasal consonant segment, then the same situation observed in excrescent nasal velarity is present. The process of final velarity then involves two steps, the first is nasal place weakening and the second is velar nasal excrescence.

(4) Final velarity

an > ãn > ã > ãŋ

am > ãm > ã > ãŋ

One could object that excrescent velar nasals in French did not arise through misperception of vowel nasality, but rather that they existed historically. Such an objection assumes that languages undergoing the sound changes in (4) pass directly from [ãn] to [ãŋ], without going through a stage of distinctive vowel nasality. If this objection is valid, then the question of perceptual motivation of the change from [ã] to [ãŋ] is moot.

In his work on the evolution of nasal vowels in Romance languages, Sampson (1999: 145) examines this particular hypothesis, first raised by Bec (1968:40). Bec's analysis assumes that in Gascon French changes like the following occurred:

LŪNA	>	'lũ.na	>	'lũŋ.a	>	'lũ-a	(Old Gascon)
PĀNEM	>	'pã.ne	>	'pãŋ.e	>	'pãŋ	

However, Sampson notes that there is both no evidence for [ŋ] having ever existed intervocalically in these forms, nor is there evidence for the syllabic reanalysis which Bec's proposal assumes. After

considering data from within Gallo-Romance, Galician-Portuguese, and Italo-Romance, Sampson concludes:

> "Place of articulation changes have normally *not occurred* amongst conditioning nasals prior to deletion. Confusing the interpretation of the available data from Romance has been the phenomenon of restoration whereby an unconditioned nasal vowel has been restructured into a sequence of nasal vowel + conditioning nasal consonant, Ṽ > ṼN, the typical value of the restored nasal consonant being velar [ŋ]." (1999:342) (emphasis ours)

Sampson is careful to point out that there are certainly sound changes that have led to changes in nasal place of articulation, yet these changes are always motivated by the following consonant's place of articulation and sometimes by the preceding vowel's quality. Importantly, sound changes that lead to excrescent velar nasals (restoration) always seem to reflect the set of sound changes in (4), where a nasalized vowel precedes a coda nasal which is subsequently lost. The third stage in (4) is a necessary precursor to excrescent velar nasals. Thus, we may confidently reject the objection that nasal consonant loss may never have occurred.

There are two elements in the sound change ã > ãŋ. The first is a reanalysis of a nasalized vowel as a sequence of a vowel + nasal consonant. Paradis and Prunet (1999) called this process "unpacking", noting that it is commonly observed in loanwords from languages with nasal vowels. The authors claimed that all nasal vowels are underlyingly biphonemic, which accounts for both their unpacking in loanword phonologies and alternations like [savɔ̃] 'soap, n.' and [savɔne] 'to soap, v.' in French. In this view, since nasalized vowels are bisegmental, they may create VN sequences when unpacked.[2]

The second element in the sound change ã > ãŋ is the determination of the place of articulation of the nasal consonant. Both Rice (1996) and Howe (2004) argued that the unmarked status of the feature [dorsal] allows velar nasals to surface as codas. Following her Default Variability Hypothesis, Rice stated that the default fill-in of an unmarked feature results in a coronal place specification while the failure of filling in an unmarked phonological feature results in the phonetic assignment of velar place of articulation to the consonant. Velar nasals are said to be the result of phonetic interpretation because "the

2 Although any language which has both a vowel length and nasalization contrast while also unpacking long nasal vowels would be problematic for the authors. One would have to assume that long nasal vowels are trimoraic.

velar is articulated with the tongue in a neutral position and involves less movement than the coronal" (1996:525). While Rice (1996:503) suggested that coronal and velar consonants share a common underlying representation in Midi French, a coronal place of articulation is "licensed in syllable-initial but not syllable-final position." She used this hypothesis to explain the alternation between coronal and velar nasals in Midi French [savɔŋ] 'soap, n.' and [savone] 'to soap, v.'

Making use of Halle, et al.'s (2000) Revised Articulator Theory, Howe (2004) proposed that all vowels are "dorsum articulated." The feature [dorsal] may therefore spread from any vowel to any adjacent segment, where codas are the favored targets. Howe (2004) disagrees with Rice's default phonetic interpretation rule for velars because [dorsal] is a feature in Halle, et al.'s feature geometry, instead of a non-terminal node as assumed by Rice. Despite this difference, both Howe (2004) and Rice (1996) concluded that coda consonant underspecification is involved in the explanation of the presence of [ŋ] as the excrescent nasal in languages like Midi French. Moreover, in Howe (2004) and Paradis & Prunet (1999) excrescent velar nasals are predicted to surface because of the preceding *vowel's* phonological specification (either with 2 X-slots or with a [dorsal] feature) while for Rice coda nasal velarity is a consequence of phonetic realization and not phonological specification.

These explanations of final velarity are unsatisfying because they are incomplete. We agree that velar nasals tend to alternate with nasalized vowels because velar nasals and nasalized vowels are similar to each other. However, "share a feature" (Howe, 2004) or "default interpretation" (Rice, 1996) are not explanations but merely ways of saying that the alternating elements are similar to each other. What is needed is a description of the sequence of events, the historical and cognitive mechanisms by which "is similar to" leads to a language sound pattern. Ohala's listener-based model of sound change is appealing precisely because it does describe such a mechanism. In the remainder of this paper we will present the results of four experiments which explore the similarities between nasalized vowels and velar nasals. With this research we sought to test the perceptual basis of excrescent nasal velarity, and to measure, in a preliminary way, the relative strengths of the influences of acoustic and visual similarity on the perception of velarity in nasalized vowels.

Ohala (1975) noted that nasalized vowels are more acoustically similar to velar nasals than they are to labial or coronal nasals. The velar nasal [ŋ] has no oral acoustic antiformants, while the mouth cavity in

[m] and [n] functions as a side branch in the vocal tract and contributes a significant low-frequency acoustic zero or antiformant. Because the mouth cavity does not form a significant oral side branch in the mouth in [ŋ] this acoustic zero is not present, and thus the acoustic spectrum of [ŋ] is more vowel-like than other nasals. Consequently, if a nasalized vowel is misperceived as a nasal segment the place of the segment will be velar because of the acoustic similarity of [ŋ] and nasalized vowels. This acoustic similarity hypothesis has not been adequately tested. We do not know whether or to what extent listeners are likely to think that they are hearing [ŋ] when presented with a nasalized vowel. The experiments here seek to test Ohala's acoustic similarity hypothesis.

The perceptual consequences of visual similarity between velar nasals and nasalized vowels has also not been formally tested. We can speculate that the visual display of a nasalized vowel is more similar to the look of velar nasals than to labial or coronal nasal because labial and coronal stops and nasals tend to be distinct from velars: in [ŋ] there is no visible tongue closure while lip closure in [m] and even to an extent tongue tip closure in [n] is visible (Grant & Walden, 1996). The visual lack of lip or tongue tip closure is thus a property of the visual displays of both [ŋ] and vowels, so we would naturally expect that if a nasalized vowel is misperceived as including a nasal segment, the place of the segment will be velar because of the visual similarity of [ŋ] and vowels. This idea, however, has not been tested in previous research.

We would note in passing here that visual phonetic properties of speech have in general not been the focus of much linguistic/phonetic research, compared with the attention given to acoustic and auditory properties. Evidence from language acquisition suggests that this is a research imbalance that should be addressed. For example, Mills (1987) studied visual influences on language acquisition and found that blind children learn [labial] later than sighted children. Blind children confused [b] with both [d] and [g] in their early word production and babbling, while the main confusion for hearing children was between [d] and [g]. This suggests that the "markedness" of [labial] for these children is partially determined by the visibility of the lips.

The importance of visual phonetics in linguistic/phonetic accounts of sound change is also highlighted by the common finding that visual phonetic similarity is substantially different from auditory phonetic similarity. For example, Grant & Walden (1996) found that the main dimensions of auditory perceptual

similarity among consonants are voicing, manner, and nasality, while in visual speech perception place of articulation is the most salient dimension of contrast (i.e. auditory and visual cues are complementary).

From these general perspectives dealing with a listener-based explanation of sound change and the audio-visual nature of speech communication, we now return to the specific question of excrescent nasal velarity and a set of experiments that were designed to test for possible audio/visual perceptual underpinnings of coda velarity. Experiments 1 and 2 test Ohala's hypothesized acoustic similarity of nasalized vowels and velar nasals, while also providing evidence regarding the strength of visual phonetic cues relative to acoustic cues. Experiments 3 and 4 test the hypothesis that nasalized vowels and velar nasals are visually confusable.

Experiment 1: Nasal place identification in audio-visual stimuli.

Experiments 1 and 2 tested the acoustic similarity hypothesis by presenting stimuli embedded in masking noise. These experiments used both audio-only and audio-visual stimuli and thus also add some new information regarding the strength of visual cues in perception.

In a typical audio-only trial in experiment 1, a listener heard a noise-masked token of the word "seen" (for example) and was asked to identify the token as either "seem", "seen", or "sing"[3]. To test the hypothesis that nasalized vowels are auditorily similar to velar nasals, we also constructed "placeless" stimuli from words ending in [m]. Final consonant place cues were largely removed from these stimuli, retaining only vowel nasalization. If Ohala's hypothesis is correct, the placeless stimuli should be identified as ending in [ŋ] despite the fact that they were derived from tokens that had actually ended in [m].

Methods.

Subjects. Fifteen subjects (8 women, 7 men) participated in experiment 1. They were undergraduate students at UC Berkeley ranging in age from 19 to 28 years. All of the participants were native speakers of English who had no history of speech or hearing disorders. None of these subjects participated in any of the other experiments described in this paper

3 The reader will notice that the velar nasal is redundantly cued in this word set by the unique vowel quality in "sing" in American English. This is discussed in the results section.

Materials. The stimuli used in this experiment were made from a digital video recording of one of the authors (KJ) producing each of the words in table 1.[4] For each of four vowel environments - [i ə ɔ eɪ] we selected three sets of words ending in the final nasals [m], [n], and [ŋ]. This resulted in recordings of thirty-six words. The word list was repeated twice.

Table 1. The thirty-six words used as stimuli in experiments 1 and 2.

[i]	beam, bean, bing
	seem, seen, sing
	ream, reen, ring
[ɔ]	calm, con, kong
	pom, pawn, pong
	rom, ron, wrong
[eɪ]	dame, dane, dang
	fame, feign, fang
	same, sane, sang
[ə]	dumb, done, dung
	rum, run, rung
	sum, sun, sung

Two movie clips of each word were produced by excising frames from the digital video recording. The starting frame for the clip was taken as the speaker began to close his mouth to form the initial consonant and ended after the speaker opened his mouth for the release of the final consonant (so we have closure and release motions for both the initial and final consonants). Then the first and last frames of the movie clip were held motionless for 0.5 second. This gives the listener time to orient to the face before the articulators begin moving. Thus, each clip had a still-frame for 0.5 seconds, then the closure and opening movements for C1, the vowel, and C2, and then a still-frame for 0.5 seconds.

The audio sound track of these movies was recorded with the digital video signal using an inexpensive

4 Though the items in table 1 span a range of lexical parameters such as word frequency, word familiarity, and neighborhood density, in this speech perception experiment (in which listeners identified the final consonant) we neither expected, nor found lexical factors to play a role in the pattern of results.

wireless lavalier microphone and separately to a digital audio recorder using a high fidelity condenser microphone. We used the hi-fi recording to construct the audio stimuli. Tokens were excised at the onset and offset of acoustic energy associated with the word, and the excised tokens were amplitude normalized so that all of the stimuli had the same peak RMS amplitude.

"Placeless" stimuli were constructed from tokens ending in [m] by deleting the acoustic signal from a point two pitch periods prior to the onset of the vowel-consonant formant transitions through to the end of the token. This effectively removes all nasal consonant place of articulation information in the final vowel formant transitions and nasal murmur. We then replicated the last glottal pulse of the truncated token so that the token's duration matched the original to within 5 ms and applied an amplitude envelope which also matched the original [m] token. The resulting stimuli thus had an extended nasalized vowel that trailed off in amplitude as if there was a nasal consonant, but the original place of articulation information was obliterated. As illustrated in Figure 1, the resulting "placeless" stimulus retained significant similarities to the [m] token from which it was constructed - this is particularly obvious in the frequency of F3 during the nasal murmur. However, the key similarity between nasalized vowels and [ŋ] that Ohala (1975) emphasized is also apparent - the spectrograms of "wrong" and "ro(m)" both have relatively prominent F1, F2 and F3 in the "nasal murmur" portion of the spectrogram. The experiment was designed to test whether this similarity is enough to cause listeners to hear [ŋ] in place of a nasalized vowel.

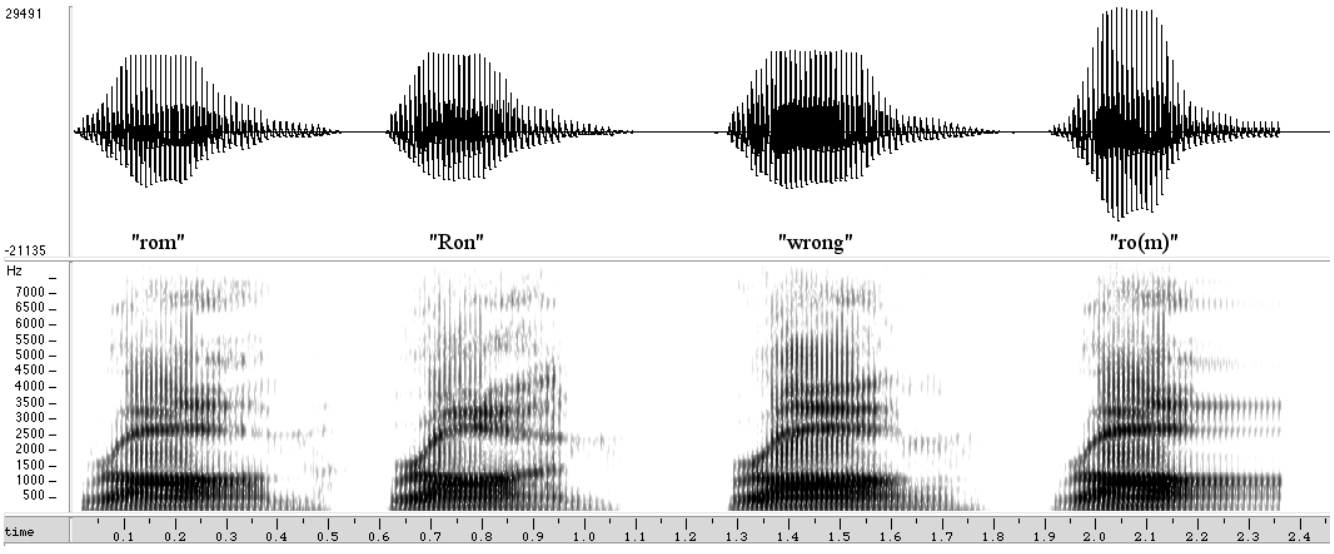

Figure 1. Spectrograms of example stimuli (before noise masking) used in experiments 1 and 2. Note that the token labeled "ro(m)" is the "placeless"

nasal token in this set.

In an error of stimulus creation, the peak RMS amplitude of the truncated stimuli was higher than that of the unmodified stimuli. This can be seen in the example stimuli in figure 1. Because the masking noise was added digitally using an algorithm that measured peak RMS amplitude and then synthesized noise to match a determined signal to noise ratio (SNR), which was constant over all of the tokens, and thus the placeless stimuli did not have an SNR advantage. However, they were louder overall than the other stimuli and this may have been noticeable to listeners. As might be expected given the relatively small amplitude mismatch, none of the listeners remarked on any loudness variation of the stimuli, and we do not see any effect in the data that could be attributed to this small inconsistency in the stimuli.

The excised and edited audio tokens were then embedded in digitally synthesized Gaussian noise. The signal to noise ratio, calculated separately for each token, was 0 dB. Meaning that the RMS amplitude of the noise was equal to the peak RMS of the token. Previous research has shown that noise at 0 dB SNR significantly degrades audio-only speech perception performance so that in a forced choice task such as the one used here listeners are at just above chance performance. The synthetic noise was longer than the speech token so that about 0.5 seconds of noise preceded the onset of the word and 0.5 seconds extended past the offset of the word.

The audio tokens thus created were added to video clips replacing the old sound tracks. They were aligned to the video clip on the basis of the existing audio sound track which had been recorded into the movie using the lavalier microphone. In this way the audio sound tracks were nearly perfectly time aligned with the video frames. Each "placeless" token was used as a sound track for three movies - one in which the talker said (visually) [m], one where he was saying [n], and one where he was saying [ŋ]. These were aligned so that the peak of vowel energy matched in the noise embedded token and in the original recording (see van Wassenhove, et al., 2007, regarding tolerable temporal misalignments between the audio and visual channels).

Procedure. The experiment was run in two blocks with the same order of blocks for all listeners. The first block was composed of presentations of the 96 audio-only stimuli (2 instances of 36 words, plus two instances of 12 placeless nasal tokens), and the second block was composed of trials with the 144 audio-

visual stimuli (2 instances of 36 words, plus 2 instances of 36 tokens constructed from placeless nasal tokens). Order of presentation within blocks was randomized separately for each listener. We chose to present the audio-only block first because we wanted a within-subjects manipulation of the "modality" factor, but we did not want the listeners to have experience with the more informative AV displays before they gave responses to the audio-only stimuli. The listeners were given a three-alternative forced-choice task in which they were to identify the final consonant in each token (pressing one of three buttons on a response box) as "m", "n" or "ng".

Results.

The overall results of experiment 1 are given in table 2, which shows the percentage of times that each response label ("m", "n", or "ng") was given, as a function of whether the stimulus was in the audio-only or audio/visual block, and as a function of whether the stimulus ended in [m], [n], or [ŋ] or was a "placeless" nasal (we are using [x̃] to symbolize these tokens. The row for placeless nasal in the audio-visual part of the table reports the response pattern when the video token was [ŋ]. Table 3 shows the overall data for the placeless nasal tokens as a function of different visual stimuli.

Table 2. Results of experiment 1 and 2. Percent of "m", "n", and "ng" responses (columns) to audio and AV stimuli (rows) in experiment 1, and percent "ng" responses in experiment 2.

		Exp 1			Exp 2
		"m"	"n"	"ng"	"ng"
audio-only	[m]	34	35	31	42
	[n]	19	56	23	28
	[ŋ]	17	28	54	62
	[x̃]	26	33	41	46
audio-video	[m]	97	1	2	3
	[n]	5	79	16	21
	[ŋ]	3	18	79	89
	[x̃]	6	40	54	60

Table 3. Percent "m", "n", and "ng" responses (columns) to audio [x̃] tokens as a function of the video display of the token (rows).

		Exp 1			Exp 2
		"m"	"n"	"ng"	"ng"
video	/m/	93	3	4	1
	/n/	7	50	43	52
	/ŋ/	6	40	54	60

A number of patterns are apparent in tables 2 and 3, prior to any statistical analysis. First, it is evident in the audio-only data that listeners tended to identify the placeless nasal tokens (identified as [x̃] in the table) in a way that is more similar to how they identified tokens that end in [ŋ] as opposed to tokens that end in [m]. This supports Ohala's acoustic similarity hypothesis. Another striking effect is that in the audio-visual condition, visual [ŋ] information had a much stronger impact for tokens ending in [ŋ] than for the placeless nasal tokens. Overall, when the [ŋ] tokens were paired with [ŋ] videos "ng" responses went up to 79% (from 54% in the audio-only condition). The placeless nasals on the other hand did show an increase of "ng" responses when paired with [ŋ] videos but the effect was much weaker (41% audio-only to 54% "ng" responses in the audio-visual condition). As we will see below, the identity of the vowel is a large factor in this result. Table 3 shows that the perception of the placeless nasal stimuli was sensitive to visual information, however the difference between the effect of the /n/ movie and the /ŋ/ movie was relatively small.

The percent "ng" data are of particular interest in this experiment because we are interested in whether listeners associate vowel nasalization with the velar nasal consonant. We see in table 2 that this is the most frequent association, so it makes sense to focus our quantitative analysis on factors that relate to listeners' tendency to use "ng" in response to the various stimuli of this experiment. We used logistic regression to test the effects of the Final consonant, the Vowel, and the stimulus Modality on the number of "ng" responses. This analysis found a main effect for Final consonant [$G^2(3) = 578$, $p < 0.01$]. Averaged over vowel contexts and modality, the percent "ng" responses was higher for tokens ending in [ŋ] or [x̃] (51% and 48% respectively) than for tokens ending in [m] or [n] (12% and 25% respectively). There was also a strong interaction between the Final consonant and Modality [$G^2(3) = 198.5$, $p < 0.01$], as can be seen in the "ng" column for experiment 1 in table 5 where we see that the pattern of percent "ng" responses is very different for audio-only versus audio-visual presentation.

The Vowel main effect was also significant [$G^2(3) = 172.7$, $p < 0.01$] with more "ng" responses to the words in the [eɪ] and [ɔ] sets (50% and 49%, respectively) and relatively fewer "ng" responses in the [i] and [ə] sets (25% and 27%, respectively). Additionally, the interactions between Vowel and Final consonant [$G^2(9) = 106.6$] and Vowel and Modality $G^2(3) = 69.5$, $p < 0.01$] were significant. These interactions are subsumed in the three way interaction between Vowel, Modality, and Final consonant [$G^2(9) = 19.5$, $p < 0.01$] which is shown in figure 2.

Analysis of the effects of vowel identity and visual place on percent "ng" responses to the placeless stimuli (Table 3) found that there were significant main effects of vowel [$G^2(3) = 131$, $p < 0.01$] and visual place [$G^2(2) = 303$, $p < 0.01$], and that the interaction of these was also significant [$G^2(6) = 20$, $p < 0.04$]. This indicates that the pattern seen in Table 3 (for the experiment 1 "ng" column) differed as a function of the vowel identity just as did the relationship between Final consonant and Modality in the main analysis.

Figure 2 illustrates that there was substantial variation in the response pattern as a function of vowel quality. In the [i] set, the only stimuli to be labeled "ng" with any consistency were the [ŋ] tokens. The placeless nasalized vowels were very rarely identified as ending in "ng". Here it appears that the unique redundant vowel quality for words ending in [iŋ] (or possibly [ɪŋ]) is enough to over ride the vowel nasality and visual cues that we are studying. Indeed, the average F1 at the mid-point of the vowel in the tokens ending in [ŋ] was 412 Hz, while it was 391 and 381 in [m] and [n] final words respectively. Similarly, F2 was different: 1978 Hz before [ŋ] but much higher at 2227 and 2288 Hz for [m] and [n] final words respectively. The words in the [i] set illustrate that listeners use the redundancy and temporal dispersion of acoustic cues to recognize words and segments. However, these tokens were not particularly well suited for testing our specific hypotheses about the acoustic and visual similarity of nasalized vowels and velar nasals.

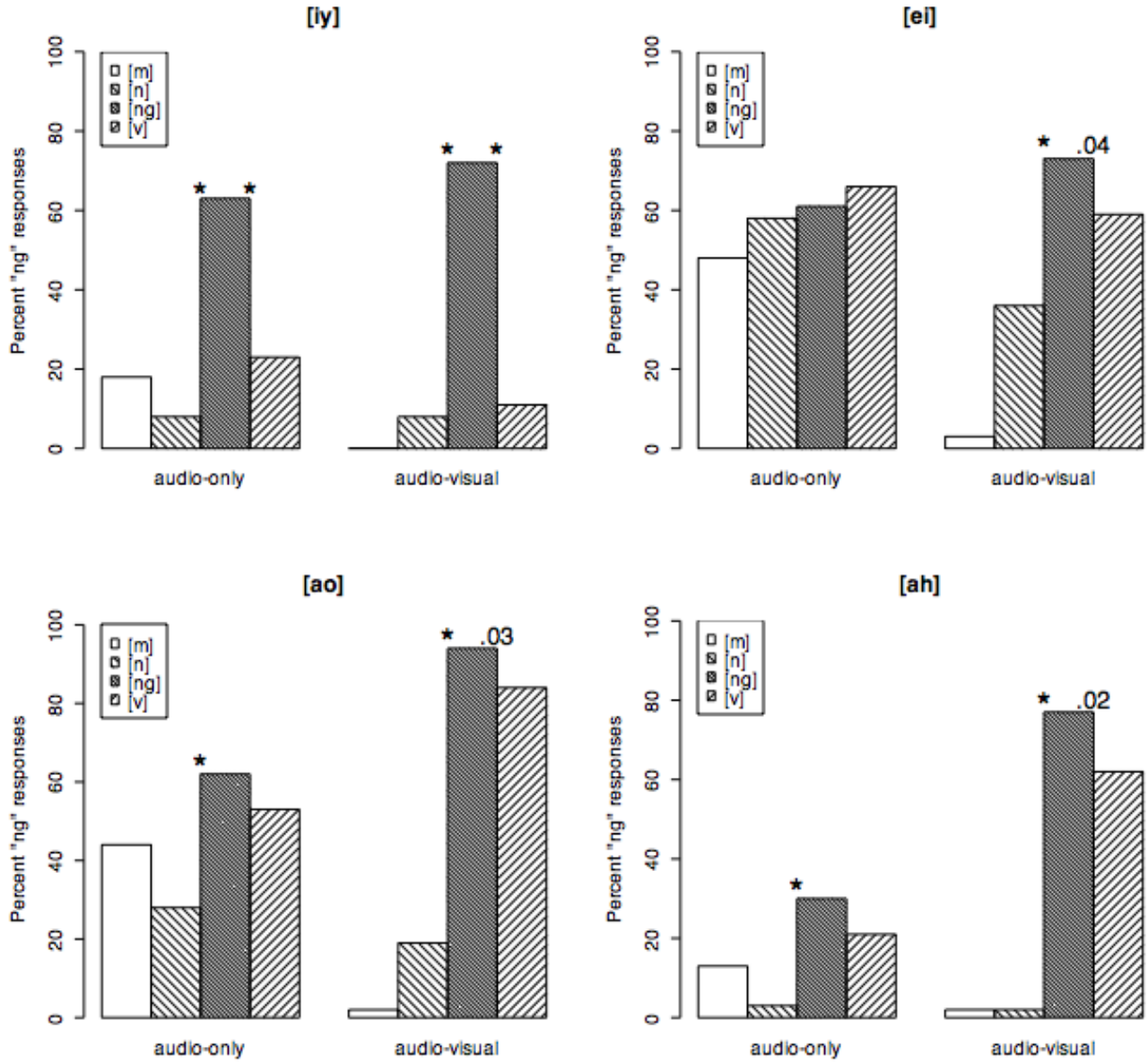

Figure 2. Results of experiment 1 shown separately for each vocalic environment. The height of the bars is proportional to the percentage of responses that were "ng" and the bar shading indicates the audio stimulus type - open bars for tokens ending in [m], right leaning hatch for tokens ending in [n], gray for tokens ending in [ŋ], and left leaning hatch for the placeless nasal tokens. ARPABET is used to identify the vowels - "iy" is [i], "ei" is [eɪ], "ao" is [ɔ], and "ah" is [ə]. Stars at the breaks between bars indicate that the conditions were significantly different in a paired

comparison and *p* values between bars indicate differences that are marginally significant. Only the pairs [n]/[ŋ] and [ŋ]/[x̃] were tested.

With the [eɪ] and especially the [ɔ] and [ə] words (tables 4 and 5) we see strong support for Ohala's hypothesis. "Placeless" nasalized vowel stimuli were much more likely to be identified as "ng" in audio-only stimuli (46% "ng"), and this trend was greatly enhanced in the audio-visual presentation with visual [ŋ] (69% "ng"). When the [i] tokens are removed from the data set and the logistic regression rerun the three-way interaction was not significant but all of the other effects noted above remained significant. Similarly, in the analysis of Visual place and Vowel effects in the perception of the placeless stimuli (table 5), the Vowel by Visual place interaction was not significant [$G^2(4) = 9.9$, $p=0.04$] when the [i] words were removed from the analysis while the main effects for Vowel [$G^2(2) = 15$, $p < 0.01$] and Visual place [$G^2(2)=307$, $p < 0.01$] remained significant.

Table 4. Results of experiment 1 and 2 with responses to [i] tokens removed. Percent of "m", "n", and "ng" responses (columns) to audio-only and audio-visual stimuli in experiment 1 (rows), and percent "ng" responses in experiment 2.

		Exp 1			Exp 2
		"m"	"n"	"ng"	"ng"
audio-only	[m]	33	32	35	48
	[n]	11	58	30	27
	[ŋ]	17	31	51	56
	[x̃]	21	32	46	51
audio-video	[m]	96	2	3	3
	[n]	3	78	19	23
	[ŋ]	2	16	82	89
	[x̃]	3	28	69	72

Table 5. Percent "m", "n", and "ng" responses (columns) to audio [x̃] tokens with [i] tokens removed as a function of the video display of the token (rows).

		Exp 1			Exp 2
		"m"	"n"	"ng"	"ng"
video	/m/	92	4	4	1
	/n/	3	43	53	61
	/ŋ/	3	28	69	72

Discussion.

This experiment found support for Ohala's (1975) explanation of excrescent nasal velarity. "Placeless" nasal stimuli, that were constructed from CVm tokens were consistently identified as more like "ng" than like "n" or "m". This result is striking because the placeless stimuli retained acoustic information that seems closer to their original [m] context. The lack of oral antiformants (and resulting fuller low spectrum) seems to have been enough, as predicted by Ohala, to drive perceptual association of nasalized vowels with velar nasals. These were not naturally produced nasalized vowels though, so this result may not be conclusive. As we saw with tokens constructed with [i], vowel quality variation may be a strong (though arbitrary) cue for final consonant identity. The quality of distinctively nasalized vowels may thus play an important role in the perception of nasal excrescence. We address this concern in experiments 3 and 4.

The visual-phonetic manipulation used in experiment 1 showed that listeners are very sensitive to visual information, and our result pairing the placeless stimuli with visual [m], [n], and [ŋ] also showed that while visual [n] and [ŋ] are similar to each other, listeners were more likely to identify [ŋ] tokens as "ng". However, the design of experiment 1 did not permit us to investigate the visual similarity hypothesis that excrescent nasal velarity is partly due to visual similarity between nasalized vowels and velar nasals. Experiments 3 and 4 fill this gap.

We turn now to an experiment designed to address a different unanswered question left by experiment 1: namely, the impact of the closed set of response alternatives given to listeners.

Experiment 2: [ŋ] detection in audio-visual stimuli.

One possible weakness of experiment 1 is that listeners were forced to choose an answer from the closed set "m", "n" and "ng". There may be some ecological validity of this forced choice in normal linguistic

circumstances - if a listener "hears" a nasal consonant segment then he/she will likely have to choose a consonant place of articulation for that percept. However, as a way to measure the perceptual similarity of [ŋ] and [x̃] the forced choice paradigm of experiment 1 may not have been very sensitive. Experiment 2 is a replication of experiment 1 in which listeners were given a more open response set. We asked a new group of listeners to indicate whether they thought the tokens ended in a "ng". The thinking here is that in a forced-choice situation listeners may feel that the placeless nasals sound more like "ng" than the other options, but they may still be very aware that the stimuli are not really good instances of "ng". On the other hand, with response alternatives "ng" and "not ng" listeners can be more selective about what counts as "ng".

Methods.

Subjects. Fifteen subjects (8 women, 7 men) participated in experiment 2. They were undergraduate students at UC Berkeley ranging in age from 18 to 37 years. All subjects were native speakers of English who had no history of speech or hearing disorder. None of these subjects participated in any of the other experiments described in this paper.

Materials. The audio-only and audio-visual stimuli in this experiment were the same stimuli that had been used in experiment 1.

Procedure. As in experiment 1, all listeners participated in two blocks of trials; in the first block of 96 trials the audio-only stimuli were presented and in the second block of 144 trials the audio-visual stimuli were presented. Order of presentation within the blocks was randomized separately for each listeners. Unlike experiment 1, the listener's task in this experiment was to determine whether the token ended in [ŋ]. The listener pushed a button labeled "yes" if he/she thought the token ended in [ŋ] and pushed a button labeled "no" if he/she thought the token did not end in [ŋ].

Results and Discussion.

The results of experiment 2 have been given in tables 2, 3, 4, and 5, along side the results of experiment 1, and figure 3 shows the percent "ng" data as a function of the three factors: Final consonant, Modality, and Vowel. Statistical analysis of these data had the same signficant factors that were found in the analyses of the experiment 1 data. As can be seen in the tables and figure, the results of experiment 2

were highly similar to the results of experiment 1. In general, listeners in experiment 2 used the response "ng" more often than did listeners in experiment 1, as might be expected given that chance performance in experiment 1 (with three response alternatives) would have resulted in 33% "ng" responses, while chance performance in experiment 2 (with only two response alternatives) would result in 50% "ng" responses.

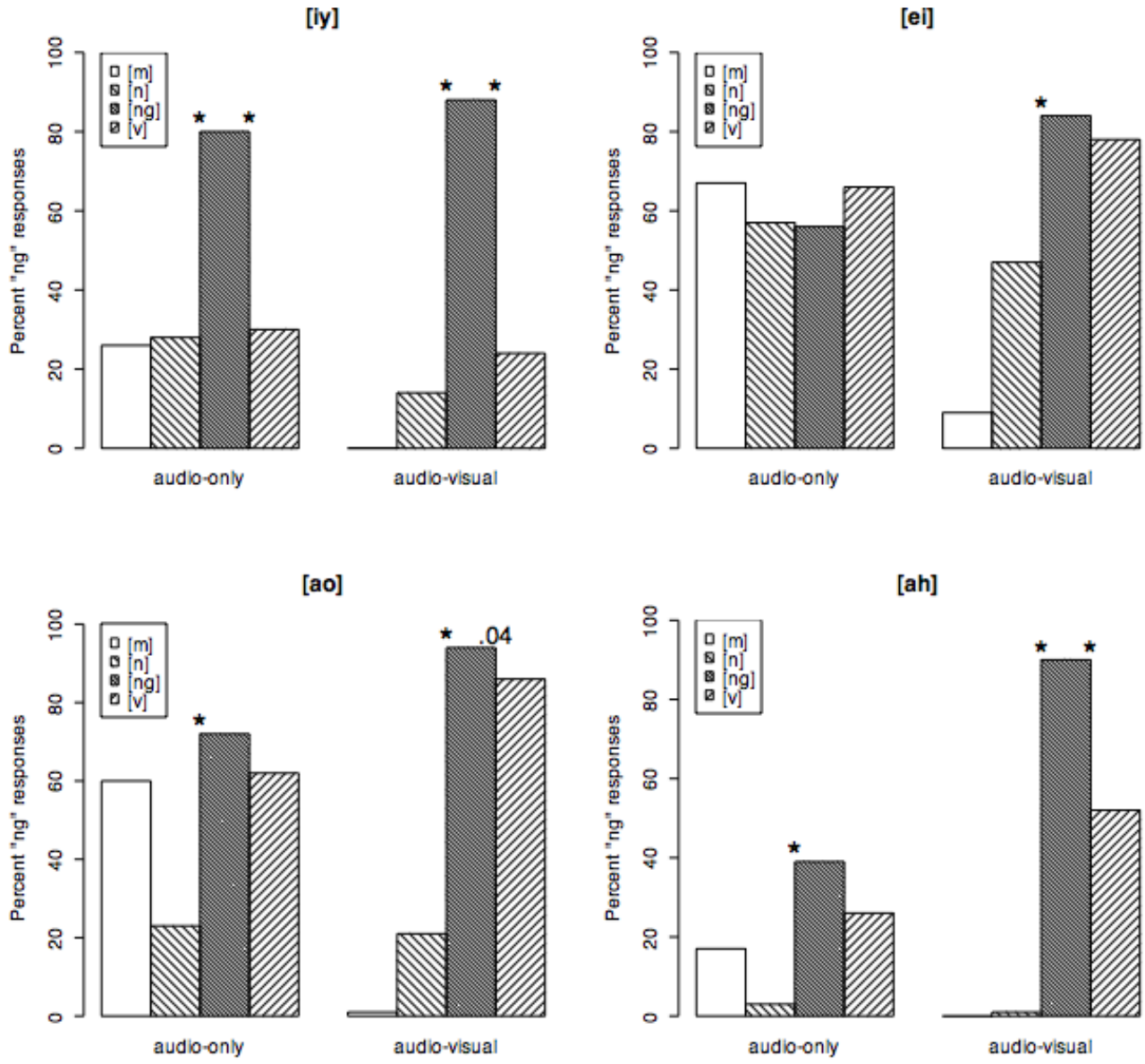

Figure 3. Results of experiment 2 shown separately for each vocalic environment. See the caption for figure 2.

The experiment then serves as a replication of experiment 1 and importantly confirms that listeners thought that the placeless nasal stimuli did actually sound like velars. They didn't simply label them as "ng" in experiment 1 because there was no alternative. We turn now to two experiments that included naturally produced nasalized vowels rather than the artificially constructed "placeless" nasals of experiments 1 and 2.

Experiment 3: Nasal identification in audio-visual stimuli.

Experiment 3, like experiment 1, used a three alternative forced choice paradigm. The difference is that for this experiment we used a new set of stimuli in which the speaker produced nasalized vowels in addition to CVN syllables.

Methods.

Subjects. Nineteen subjects (12 women, 7 men) participated in experiment 3. They were undergraduate students at UC Berkeley ranging in age from 18 to 33 years. All of the participants were native speakers of English who had no history of speech or hearing disorder. None of these subjects participated in any of the other experiments described in this paper.

Materials. The stimuli used in this experiment were made from a digital video recording of one of the authors (LM) producing each of the "words" in table 6. We dropped the [i] tokens used in experiment 1 because the vowel quality in [ŋ] context was different from [m] and [n] contexts. For each of three vowel environments - [ə ɔ eɪ] we selected three sets of words ending in the final nasals [m], [n], and [ŋ]. The speaker, who is a phonetically trained, native speaker of English, and L2 speaker of French, also produced a "word" which consisted of the initial consonant and a nasalized version of the vowel. This resulted in recordings of thirty-six words.

Table 6. The thirty-six "words" used as stimuli in experiments 3 and 4.

[ə]	dumb, done, dung, [də̃]
	rum, run, rung, [rə̃]
	sum, sun, sung, [sə̃]

[ɔ] calm, con, kong, [kɔ̃]
pom, pawn, pong, [pɔ̃]
rom, ron, wrong, [rɔ̃]
[eɪ] fame, feign, fang, [fẽɪ]
dame, dane, dang, [dẽɪ]
same, sane, sang, [sẽɪ]

Movie clips of each word were produced by excising frames from the movie using the same method as in experiment 1. The starting frame for the clip was taken as the speaker began to close her mouth to form the initial consonant and ended after the speaker opened her mouth after saying the final consonant (so we have closure and release motions for the initial and final consonants. Then the first and last frames of the movie clip were held motionless for 0.5 second. This gives the listener time to orient to the face before the face begins moving. Thus, each clip had a still-frame for 0.5 seconds, then the closure and opening movements for C1, the vowel, and C2, and then still-frame for 0.5 seconds.

The audio sound track of these movies was recorded with the digital video signal using an inexpensive wireless lavalier microphone. Unlike experiment 1, we used this lower quality signal for the experimental tokens[5]. The audio sound track was excised into a sound file (44kHz, 16 bit samples) and edited to remove the final nasal and most place of articulation information associated with the nasal. The last half of the vowel and the nasal segment were deleted and a burst of Gaussian white noise of the same duration as the deleted segment was inserted in place of the deleted segment. The resulting audio files (see figure 4) had highly obscured final nasals, though conceivably some place information could remain in the quality of the first part of the vowel. The audio-visual stimuli were always phonetically matched so that audio [m] was used as the sound track for the visual [m] token, etc.

5 The lavalier microphone recording had less low frequency energy and thus a reduced amount of acoustic/phonetic information. This was deemed acceptable because the results of pilot work showed that enough information was captured to permit correct identification of the stimuli in audio-only presentation and because the focus of this study was more on the role of visual information in excrescent nasal velarity.

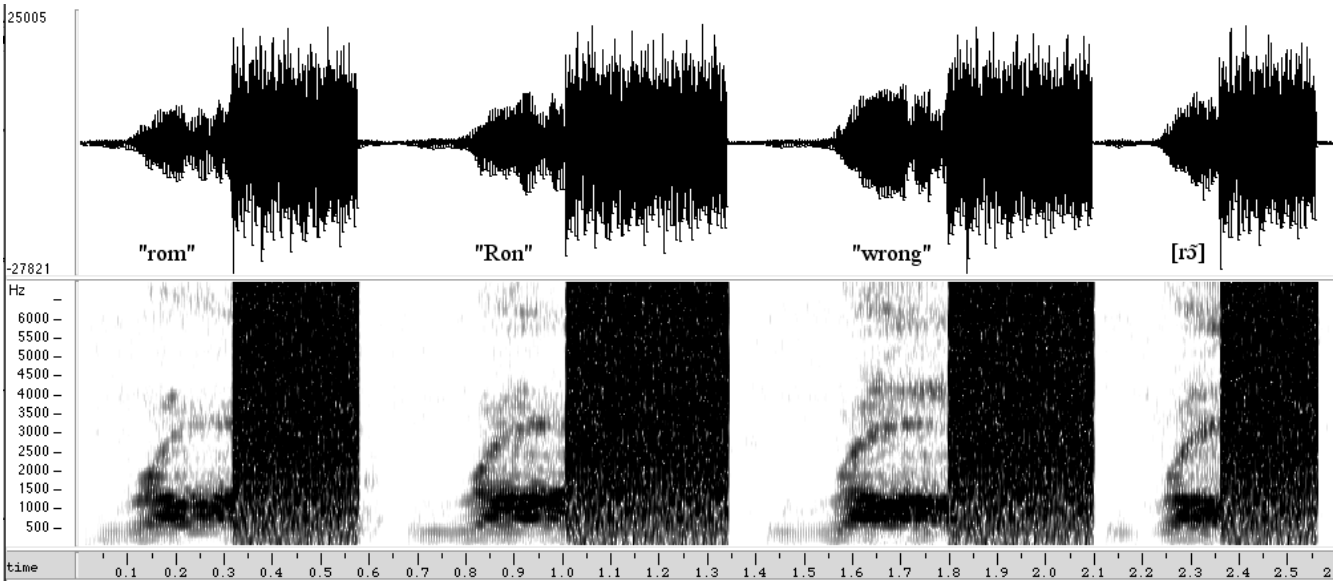

Figure 4. Spectrograms of example stimuli used in experiments 3 and 4. Information for the initial consonant and the first part of the vowel was presented in the clear, but the last half of the word was replaced by noise

Procedure. The listener's task in this experiment was the same as in experiment 1, a forced choice identification of each token as ending either with "m", "n" or "ng". The audio-only stimuli were presented first in a block of 36 trials, and then the 36 audio-visual stimuli were presented. Order of presentation within blocks was randomized separately for each listener.

Results and Discussion.

The overall results of experiment 3 are shown in table 7. That table lists percent "m", "n" and "ng" responses in columns and lists the stimulus set (audio-only [m], audio-only [n], etc.) in rows. Looking first at the audio-only block of trials we see that listeners tended to identify the [m] and [n] audio tokens as ending in "n" (55% and 56% respectively), but the [ŋ] and [ɤ̃] audio tokens were both more likely to be identified as [ŋ] (42% for both sets of stimuli). Comparing the % "n" responses and %"ng" responses for audio-only [ŋ] and [ɤ̃] stimuli it appears that the tendency to identify [ŋ] and [ɤ̃] as "ng" was actually stronger for the nasalized vowel tokens than for the velar nasal.

Looking at the audio-visual data in in the bottom half of table 7 we see that adding visual information

dramatically improved the identification accuracy for [m] and [n] tokens (92% and 75% correct). Identification accuracy for the velar [ŋ] token was also higher than in the audio-only condition (59% versus 42% correct). Also listeners were more likely to identify the final nasal as "ng" in the nasalized vowel [ɛ̃] audio-visual stimuli than they did with the audio-only stimuli (56% versus 42%). The change in "ng" responses to the nasalized vowel (from audio-only to audio-visual presentation) was about the same as the change seen for [ŋ] tokens.

Table 7. Results of Experiments 3 and 4. Percent "m" "n" and "ng" responses (columns) to the audio and AV stimuli ending in different nasal consonants or nasalized vowel in experiment 1 (rows), and percent "ng" responses in Experiment 4.

		Exp 3			Exp 4
		"m"	"n"	"ng"	"ng"
audio-only	[m]	25	55	22	33
	[n]	21	56	23	33
	[ŋ]	17	40	42	47
	[ɛ̃]	28	31	42	39
audio-visual	[m]	92	5	3	4
	[n]	4	75	22	45
	[ŋ]	4	38	59	69
	[ɛ̃]	7	38	56	47

As with the data in experiment 1, we used logistic regression to test the effects of the Final consonant, the Vowel, and the stimulus Modality on the percent "ng" responses. This analysis found a main effect for Final consonant [$G^2(3) = 168.6$, $p < 0.01$]. Averaged over vowel contexts and modality, the percent "ng" responses was higher for tokens ending in [ŋ] or [ɛ̃] (51% and 48% respectively) than for tokens ending in [m] or [n] (12% and 25% respectively). There was also a strong interaction between the Final consonant and Modality [$G^2(3) = 42.6$, $p < 0.01$], as can be seen in the "ng" column for experiment 3 in table 7. Fewer "ng" responses were given to [m]-final tokens in the audio-only condition than in the audio-visual condition, while the pattern was reversed for [ŋ]- and [ɛ̃]-final tokens.

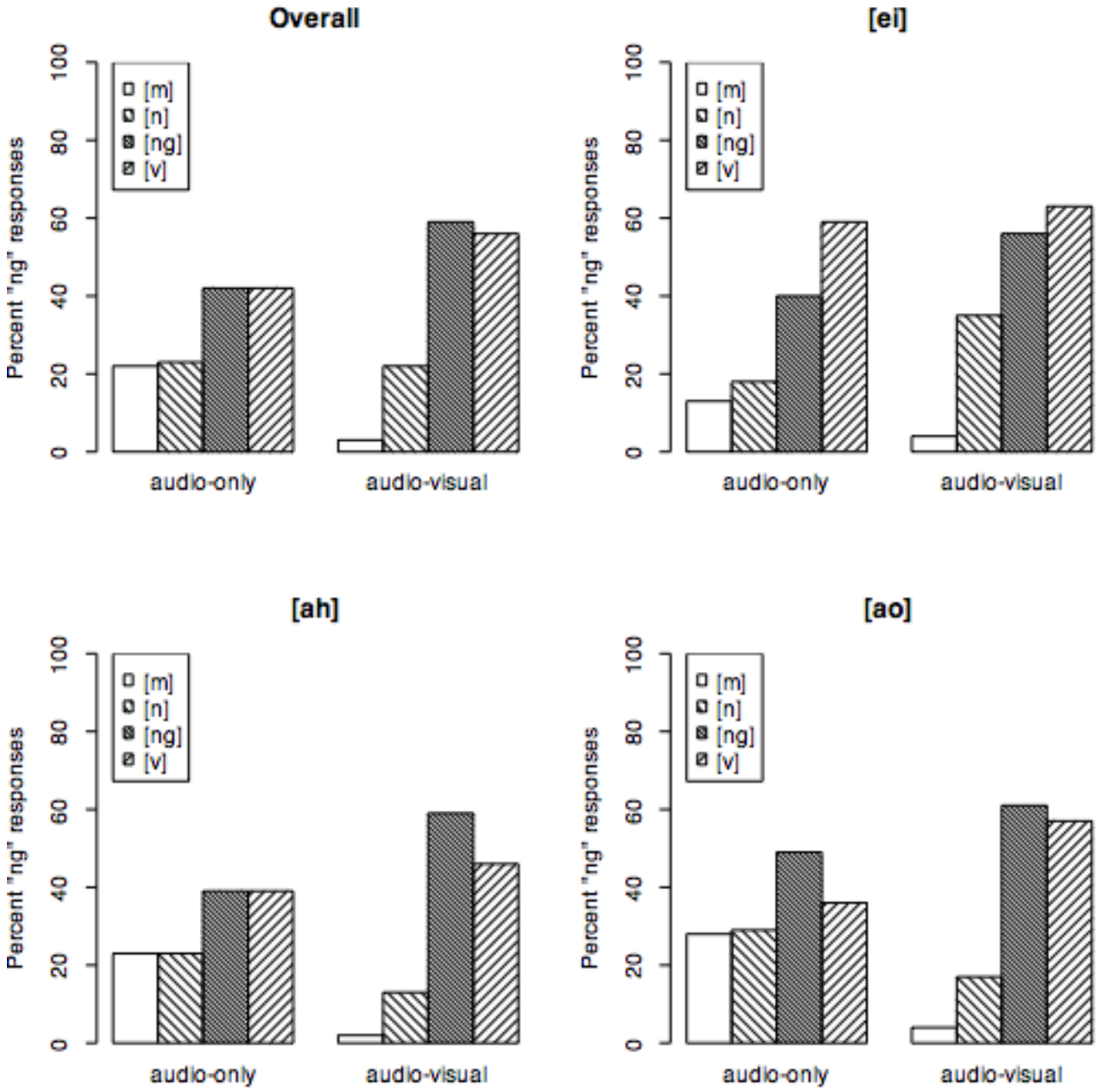

Figure 5. Results of experiment 3. These graphs show the percent "ng" responses for audio-only and audio-visual presentations for the experiment overall and in different vowel environments. See the caption for figure 2.

The Final consonant by Vowel interaction approached significance [$G^2(6) = 13.6$, $p = 0.03$], as did the

three way interaction of vowel, final consonant, and modality [$G^2(6) = 10.7$, $p = 0.1$]. The three way interaction is shown in figure 5 and the differences among the vowels that we see there were tested in a set of separate analyses for tokens with different vowels. These three logistic regressions (one each for [eɪ], [ɔ], and [ə] tokens) all found Final consonant to be a significant factor in "ng" responses, but for [ɔ] and [ə] tokens the final consonant by modality interaction was significant [$G^2(3)=22.3$, $p<0.01$ and $G^2(3)= 23.0$, $p < 0.01$, respectively] and the final consonant effect was not significant in an analysis of audio-only tokens, while for tokens with [eɪ] there was no Final consonant by Modality interaction and the final consonants did significantly differ from each other in the audio-only data [$G^2(3) = 35$, $p < 0.01$]. As can be seen in figure 5, adding visual information did not change the number of "ng" responses as much for tokens with [eɪ] than for tokens with the other vowels. This makes sense because there is a small vowel quality difference between the vowels of "same" and "sane" on the one hand and the vowel of "sang" and [sẽɪ], similar to the difference in the [i] stimulus set used in experiments 1 and 2[6]. The average second formant (F2) frequencies of the tokens used in the experiment are: 2713, 2753, 2544, 2504 Hz for the [eɪm], [eɪn], [eɪŋ], and [ẽɪ] words respectively (with standard deviations of less than 100 Hz). The other two vowel environments did not show any F2 separation of the tokens conditioned by the final consonant.

This experiment found evidence for the two main hypotheses that we are testing. First, we found that in noise-masked audio-only displays nasalized vowels were identified as "ng" more often than "m" or "n". Second, we found that in audio-visual displays the number of "ng" responses to nasalized vowels increased. These results are compatible with the results of experiments 1 and 2 in suggesting that the effect of auditory similarity between nasalized vowels and velar nasals may explain why excrescent nasals tend to be velar. But where experiments 1 and 2 did not address the question of how natually produced nasalized vowels are perceived, in this experiment we have evidence that nasalized vowels are both acoustically and visually similar to coda velar nasals.

Experiment 4: [ŋ] detection with noise-obliterated stimuli.

In experiment 4, listeners heard (and watched) the same stimuli that were used in experiment 3. Again we were interested in determining if listeners would associate nasalized vowels with final [ŋ] as the

6 The [eɪ] tokens in experiments 1 and 2 were also somewhat different in the [ŋ] context, but that difference seems to have not been large enough to have much impact on the perceptual results. It is possible that the noise replacement method used in this experiment increased listeners' reliance on the vowel quality difference.

acoustic similarity similarity hypothesis predicts and we were also interested in testing whether listeners would have a greater tendency to make the [x̃]-[ŋ] association when presented with audio-visual displays. Experiment 4 differs from experiment 3, just as experiment 2 differed from experiment 1, because we asked listeners to make a 'yes' - 'no' decision about the stimuli, deciding whether the stimulus ends in [ŋ] or not. As with experiment 2, the rationale for this change in procedure was that with a three-alternative forced choice task used in experiment 3 listeners were required to label the stimulus as either "m", "n", or "ng" and if the nasalized vowel stimuli didn't really fit any of these alternatives, listeners still had to choose one of the nasal consonants as their response. Thus, the experiment showed that nasalized vowels sound more like [ŋ] than like [m] or [n], but we still don't know how similar [x̃] and [ŋ] really are. Experiment 4 addresses this question by giving listeners an open-ended response option - they can say that the stimulus doesn't sound like [ŋ] even if they aren't sure what exactly it does sound like. If listeners still associate [x̃] with [ŋ] in this [ŋ] detection task, then we have stronger evidence for the association in audio and/or audio-visual stimuli.

Methods.

Subjects. Fifteen subjects (8 women, 7 men) participated in experiment 4. They were undergraduate students at UC Berkeley ranging in age from 18 to 35 years. All subjects were native speakers of English who had no history of speech or hearing disorder. None of these subjects participated in any of the other experiments described in this paper.

Materials. The audio and AV stimuli in this experiment were the same stimuli that had been used in experiment 3.

Procedure. As in experiment 3, all listeners participated in two blocks of trials; in the first block of 72 trials the audio-only stimuli were each presented twice and in the second block of 72 trials the audio-visual stimuli were each presented twice. The order of presentation within blocks was randomized separately for each listener. The listener's task in this experiment, as in experiment 2, was to determine whether the token ended in [ŋ]. The listener pushed a button labeled "yes" if he/she thought the token ended in [ŋ] and pushed a button labeled "no" if he/she thought the token did not end in [ŋ].

Results and Discussion.

The overall percent "ng" responses in experiment 4 have already been shown in table 7. It is instructive to compare the percent "ng" responses in experiments 3 and 4. Chance performance in experiment 3, if listeners were equally likely on each trial to push any one of the three response buttons, was 33%. Chance performance in experiment 4 on the other hand was 50%.

In experiment 4 listeners could respond "no" indicating that the word did not sound like it ended in [ŋ] without having to identify it as any other nasal segment. As expected this change in task produced greater separation between nasalized vowel stimuli and velar nasal stimuli. Listeners were less likely to label [ɛ̃] stimuli as "ng" in experiment 4, while the percent of "ng" responses increased (compared to experiment 3) for all other stimuli.

Conclusion

To summarize the experimental results that we found in this study, we found that listeners were more likely to identify nasalized vowels (whether artificially constructed or naturally produced) as [ŋ] in experiments in which they were forced to choose among the three alternatives "m", "n", or "ng" and in experiments in which they were given the opportunity to identify the token as "ng" or "not ng". Nasalized vowels were less likely to be perceived as "ng" when compared with words that actually ended in [ŋ], even when vowel quality was not a redundant cue for final nasal velarity, but the trend toward nasalized vowel identification as "ng" was unmistakable.

We also found in experiments 3 and 4 that adding the visual display of a person producing a nasalized vowel increased the likelihood of an "ng" response for nasalized vowel tokens in much the same way that adding visual [ŋ] information increased "ng" responses (again, in both forced choice and "ng" detection experiments).

These results support two hypotheses about excrescent nasal velarity: that the historical change of nasalized vowel to a sequence of vowel followed by velar nasal is grounded in the acoustic and visual similarity of nasalized vowels and coda velar consonants. Given the results of our experiments, we conclude that it makes sense to believe that innocent misperception of nasalized vowels could lead listeners to posit the existence of a coda velar nasal, and thus that a plausible historical phonological explanation of excrescent nasal velarity is that the change originates in misperception of nasalized

vowels.

We would like to close by commenting briefly on two issues which place this work in a larger theoretical context. The first of these is a comment on the limits of phonetic factors in explaining sound change. Although many phonologists seek to find grounding of phonological patterns in phonetic properties of speech (Archangeli & Pulleyblank, 1994; Blevins, 2004), there is clearly tension as to whether such a program is adequate or desirable (e.g. Hyman, 2001). In our studies we sought, and found, evidence tying excrescent nasal velarity to the acoustic and visual phonetic properties of nasalized vowels and velar nasals. But we do not believe, and would not wish to suggest, that a full understanding of historical or synchronic phonological processes is limited to those which can be grounded in phonetic processes. For example, we could imagine a dialect of French that would exhibit excrescent nasal coronality. In this hypothetical dialect "soap" is pronounced [savɔn] where standard French has [savɔ̃]. If this dialect where to be discovered we would not take this to be evidence against the perceptual basis of excrescent nasal velarity, but would instead seek a different, perhaps not strictly phonetic, basis for understanding the excrescence of a coronal nasal. For example, we might suggest that a process of analogical leveling is at work in which the morphological relationships among forms having the same stem are made more transparent (e.g. now [savɔne] "to soap up" and [savɔn] "soap" have the same surface realization of the stem). Perhaps we could also argue that the appearance of the coronal nasal is supported by the spelling of [savɔ̃] as "savon".[7] The point of this small example is simply to suggest that the discovery of a case of excrescent nasal coronality is not an argument against the perceptual basis of excrescent nasal velarity that we think we have found. Instead, such a discovery would support the view that phonetic grounding is not the only way to understand phonological patterning and that existing phonotactic patterns (Pierrehumbert, 2006), patterns of morphological alternations, or perhaps even spelling conventions may function as a different sort of grounding for historical phonological processes (see Hume, 2004, for an example of this more nuanced conception of grounding).

The other more general issue that we would like to address by way of conclusion is the relationship between work in experimental historical phonology (Ohala, 1974) and one's conception of synchronic phonology. In essence, our comment on this topic is that our work in historical phonology indicates that a non-generative conception of synchronic phonology is more plausible than the more common

7 In a related study, MacKenzie (2006) found that Parisian French speakers when required to identify the "final nasal" in nasalized vowels almost always chose "n" rather than "ng" perhaps reflecting such analogy or spelling based biases.

generative conception. To back up slightly, we should explain how we reach this conclusion. The task of historical phonology is to explain why synchronic phonological patterns arise - to understand how language ends up with the phonological patterns that is has, while the task of synchronic phonology is to describe the speaker/hearer's knowledge and the mental representation of these phonological patterns. Obviously, these two research areas are intertwined with each other very closely and will inform each other. One connection between them that we see has to do with the important role of variation in historical phonological explanations. To explain sound change, it is necessary to assume that speakers, when they are presented with variant pronunciations of words, remember them. Some of these variants are no doubt merely speech errors, others are slips of the ear (Bond, 1999), neither of which will make their way into the speaker's vocabulary of produced forms (but see Pierrehumbert, 2001). However, when a variant (say CVŋ) takes on sociolinguistic significance it changes from an error to an intended linguistic form. Thus, misperception, by itself, is not enough to drive sound change, but attention and conscious reanalysis can significantly magnify the effects of misperception (McGuire, 2007). The key point for our argument is that for variants to take on significance and become part of a new way of speaking, the variant forms themselves must be remembered. A hypothesized perceptual process which removes variation and does not store variant pronunciations with some degree of phonetic detail eliminates the possibility of sound change, and must therefore be an incorrect view of perception because we know that sound change occurs[8].

This conception of historical phonological processes is incompatible with the usual understanding of phonology as a generative system, in which surface forms are generated from underlying forms. In the generative conception, a morphophonemic alternation such as [savɔne] "to soap up", [savɔ̃] "soap" is indicative of a rule which changes the underlying form /ɔn/ into a surface form [ɔ̃], or the interaction of a set of constraints including one that prevents the underlying /n/ from being realized without a following vowel in the word, and another that forces the nasality of /n/ to be realized near the location of the missing /n/. Correlated with this conception of generative speech production is a theory of generative perception (Liberman, 1996) in which it is assumed that listeners recover the underlying representation during perception and thus the underlying representation is the only long-term phonetic/phonological record of the form. Such a "unitary representation" conception of the generative phonological system is,

8 Arguably, it is the perceptual systems of children that are important for the view of sound change presented here (Labov, 2007), however, the fact that phonetic accommodation to interlocutors is found in adults as well as children (Pardo, 2006) suggests that the adult system is also implicated to at least some extent, and importantly, with no qualitative change in mode of operation.

as we emphasized earlier, incompatible with the facts of sound change. It is thus, from the perspective of historical phonology, more plausible to conceive of alternations like [savɔne] "to soap up", [savɔ̃] "soap" non-generatively as relationships among forms in memory. That is, the variants are not generated from a single invariant form, but instead both exist in the speaker's memory and the alternation between [ɔn] and [ɔ̃] is a fact about related forms which is recoverable from these remembered variants and which may impact on the perceptual interpretation and production of similar forms through a process of exemplar activation (Johnson, 1997). This non-generative conception of phonology is compatible with patterns of historical sound change, including lexical diffusion of sound change (Wang, 1969) and near merger (Labov, 1994), as well as with emergent properties of synchronic phonology (see e.g. Bybee, 2001).

Acknowledgments. This research was supported by NSF grant #9817243. Our analysis and presentation of these results benefited from audiences at "The Phonetic Basis of Distinctive Features" in Paris, 2006. Many thanks to the volunteers in California and France who participated in the study, and also to John Ohala for comments and inspiration.

References.

Archangeli, D. & Pulleyblank, D. (1994) *Grounded Phonology*, MIT Press, Cambridge, MA.

Auer, Jr., E. T. (2002) The influence of the lexicon on speechread word recognition: Contrasting segmental and lexical distinctiveness. *Psychonom. Bull. Rev.* 9(2), 341–347.

Bec, P. (1968) *Les Interférences linguistiques entre gascon et languedocien dans les parlers du Comminge et du Couserans*. Paris: PUF.

Beddor, P.S. & Evans-Romaine, D. (1995) Acoustic-perceptual factors in phonological assimilations: A study of syllable-final nasals. *Rivista di Linguistica* 7, 145-174.

Benot, C., Lallouache, T., Mohamadi, T. & Abry,C. (1992) A set of French visemes for visual speech synthesis, in: G. Bailly, C. Benot (Eds.), *Talking Machines: Theories, Models and Designs*, Elsevier B.V., (pp. 485--501).

Blevins, J. (2004) *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

Bond, Z. S. (1999) *Slips of the Ear: Errors in the Perception of Casual Conversation*. San Diego, CA: Academic Press.

Brancazio, L. (1998). 'Contributions of the lexicon to audiovisual speech perception, unpublished Ph.D., dissertation, University of Connecticut.

Burnham, D. & Dodd, B. (2004) Auditory-visual speech integration by pre-linguistic infants: Perception of an emergent consonant in the McGurk effect .*Developmental Psychobiology* 44 , 209-220.

Bybee, J. (2001) *Phonology and language use*. Cambridge: Cambridge University Press.

Clements, G. N., and Hume, Elizabeth (1995) The internal organization of speech sounds. In J.A. Goldsmith (ed) *The Handbook of Phonological Theory.* Cambridge, MA, and Oxford, UK:

Blackwell (pp. 245-306).
Grant, K.W., Braida, L.D., Renn, R.J., (1994). Auditory supplements to speechreading: Combining amplitude envelope cues from different spectral regions of speech. *J. Acoust. Soc. Am.* 95, 1065-1073.
Grant, K.W., and Walden, B.E. (1996). Evaluating the articulation index for auditory-visual consonant recognition. *J. Acoust. Soc. Am.* 100, 2415-2424.
Grant, K. W., Walden, B. E., and Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *J. Acoust. Soc. Am*. 103, 2677–2690.
Hajek, John (1991) The Hardening of Nasalized Glides in Bolognese, in Bertinetto, P. M. and Kenstowicz, M. and Loporcaro, Michele, Eds. *Certamen Phonologicum II*, pages pp. 259-278. Turin: Rosenberg and Sellier.
Halle, Morris (1995) Feature geometry and feature spreading. *Linguistic Inquiry* 26:1-46.
Halle, Morris, Vaux, Bert, and Wolfe, Andrew. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* **31**, 387-444.
Hampson, M., Guenther, F. and Cohen, M.(1998) Visual influences on the perception of alveolar/velar place discrimination. *J. Acoust. Soc. Am.* 104, 1854.
Hess S. (1990) Universals of nasalization, development of nasal finals in Wenling. *Journal of Chinese Linguistics* **18 (1)**: 44-94.
House, A.S. (1957) Analog studies of nasal consonants. *Journal of Speech and Hearing Disorders* **22 (2)**: 190-204.
House, D. (1982) Identification of nasals: An acoustical and perceptual analysis of nasal consonants. Lund University, Department of Linguistics, *Working Papers* 22, 153-162.
Howe, Darin (2004) *Vocalic Dorsality in Revised Articulator Theory*. Ms, Univ. of Calgary.
Hume, E. 2004. The Indeterminacy/Attestation Model of Metathesis. *Language* **80**(2), 203-237.
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.
Hyman, L. (2001) On the limits of phonetic determinism in phonology: *NC revisted. In Beth Hume & Keith Johnson (eds), *The Role of Perception in Phonology*. New York: Academic Press, 141-185.
Lachs, L. and Pisoni, D.B. (2004) Specification of cross-modal source information in isolated kinematic displays of speech . *J. Acoust. Soc. Am*. 116 , 507
Liberman, A.M. (1996) *Speech: a special code.* Cambridge, MA: MIT Press.
MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception process. *Perc. & Psych*. , 24, 253- 257.
Malécot, A. (1956) Acoustic cues for nasal consonants: An experimental study involving tape-splicing technique. *Language* 32, 274-284.
Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Erlbaum: Hillsdale.
Massaro, D. W. (1998). *Perceiving Talking Faces: From speech perception to a behavioral principle*. MIT, Cambridge, MA.
McGuire, G. (2007) *Phonetic category learning*. Unpublished PhD dissertation, Ohio State University.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264:, 46-748.
Ohala, J. J. 1974. Experimental historical phonology. In: J. M. Anderson & C. Jones (eds.), *Historical linguistics II. Theory and description in phonology.* [Proc. of the 1st Int. Conf. on Historical Linguistics. Edinburgh, 2 - 7 Sept. 1973.] Amsterdam: North Holland. 353 - 389.
Ohala, J. J. 1975. Phonetic explanations for nasal sound patterns. In: C. A. Ferguson, L. M. Hyman, & J. J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*. Stanford:

Language Universals Project. 289 - 316.
Ohala, J. J. 1981. The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Ling. Soc. 178 - 203.
Ohala, J.J. (1990) The phonetics and phonology of aspects of assimilation. In Kingston, J. and Beckman, M.E. (eds) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press (pp. 258-275).
Pardo, J. (2006) On phonetic convergence during conversational interaction . *J. Acoust. Soc. Am.* **119**, 2382-2393
Pierrehumbert, J. (2001) Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (eds.) *Frequency effects and the emergence of lexical structure*. John Benjamins, Amsterdam. 137-157.
Pierrehumbert, J. (2006) The statistical basis of an unnatural alternation, in L. Goldstein, D.H. Whalen, and C. Best (eds), *Laboratory Phonology VIII, Varieties of Phonological Competence*. Mouton de Gruyter, Berlin, 81-107.
Pols, L.C.W. & Schouten, M.E.H. (1978) Identification of deleted consonants. *J. Acoust. Soc. Am.* 64 (5): 1333-1337.
Posner, R. (1997) *Linguistic Change in French*. Oxford, Oxford University Press.
Recasens, D. (1983) Place cues for nasal consonants with special reference to Catalan. J. Acoust. Soc. Am. 73 (4): 1346-1353.
Rice, K. (1996) Default variability: the coronal-velar relationship. *Natural Language and Linguistic Theory* 15. 493-543.
Rosenblum, L. D. (1994) How special is audiovisual speech integration? *Curr. Psychol. Cognition* 13(1), 110–116.
Sagey, Elizabeth (1986) *The representation of features and relations in nonlinear phonology.* MIT: Doctoral dissertation.
Sampson, Rodney (1999) *Nasal Vowel Evolution in Romance*. Oxford University Press, Oxford.
Schwartz, J.-L., Robert-Ribes, J., & Escudier, P. (1998). "Ten years after Summerfield: A taxonomy of models for audio-visual fusion in speech perception," in *Hearing by Eye II*, edited by R. Campbell, B. Dodd, and D. Burnham. Psychology, East Sussex, UK, pp. 85–108.
Sharf, D.J. & Ostreich, H. (1973) Effect of forward and backward coarticulation on identification of speech sounds. Language and Speech16, 196-206.
Shockey, L. & Reddy, R. (1974) Quantitative analysis of speech perception: Results from transcription of connected speech from unfamiliar languages. Speech Communication Seminar, Stockholm.
Shosted, Ryan K. 2005. Vocalic context as a condition for nasal coda emergence: aerodynamic evidence. *UC Berkeley Phonology Lab Annual Report 2005,* 49-62.
Sumby W.H. & Pollack, I. (1954) Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212-215.
Summerfield, Q. (1979) Use of visual information in phonetic perception. *Phonetica*, 36(4--5):314--331.
Summerfield, A. Q. (1981~87?). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip reading* (pp. 3-51). London: Erlbaum.
ven der Torre, Erik Jan. (2003) Dutch Sonorants: the role of place of articulation in phonotactics. Utrecht, the Netherlands: LOT.
Walker, Dale F. (1976) A grammar of the Lampung language: The Pesisir dialect of Way Lima. *NUSA, Linguistic Studies in Indonesian and Languages in Indonesia* **2**,1-49.
Wang, W. S.-Y. (1969). Competing changes as a cause of residue. *Language* **45**, 9-25.

van Wassenhove, V., Grant, K.W. and Poeppel, D. (2007) Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* **45** (3), 598-607

Wiese, R. (1996) *The Phonology of German*. Oxford: OUP.

Zee, E. (1981) Effect of vowel quality on perception of post-vocalic nasal consonants in noise. *Journal of Phonetics* 9 (1): 35-48 1981