# Perception grammars and sound change*

Patrice Speeter Beddor
University of Michigan

The acoustic consequences of gestural overlap afford listeners multiple, time-varying cues for a given linguistic percept. Findings from "offline" perceptual tasks and "online" real-time processing converge in demonstrating that listeners attend to the dynamic cues, tracking the coarticulatory information over time. These findings also converge in showing that listeners systematically differ in their perceptual weighting of the information contributed by the coarticulatory source and its effects; that is, listener attention is selective. One factor contributing to these listener differences in perception grammars may be listener-specific experiences with particular coarticulatory patterns. However, another factor is the quasi-systematic nature of coarticulatory variation, which provides listeners with covarying cues and therefore multiple *possible* weightings that are fully consistent with the input. Of particular interest for sound change are "innovative" listeners, for whom the coarticulatory cues are heavily weighted. These listeners' perception grammars have the potential to contribute to changes in which the coarticulatory effect is requisite and its source may be lost – but only insofar as those grammars are publicly manifested. Such manifestation is likely to occur in conversational interactions either through innovative listeners' expectations about coarticulated speech or through those listeners' own productions.

## 1. Introduction

This chapter assesses the complex nature of perception grammars, their relation to variation in the input auditory signal, and their possible contributions to sound change. I focus on perception grammars for coarticulated speech, that is, on how

* This research was supported by NSF Grant BCS-0118684. Portions of this work were conducted in collaboration with Julie Boland, Anthony Brasher, Andries Coetzee, Kevin McGowan, Chandan Narayan, and Chutamanee Onsuwan. I thank these colleagues, as well as Susan Lin for creative assistance with data presentation. I also acknowledge the helpful comments of three anonymous referees and the valuable input from numerous audiences, including participants in the 2010 Workshop on Sound Change at the Institut d'Estudis Catalans in Barcelona.

listeners systematically organize and respond to the gestural overlap resulting from speakers' coordination of articulatory movements within and across linguistic units. Overlapping articulatory events have the potential to be perceptually informative or disruptive. They are informative in that the resulting "parallel transmission" (Mattingly 1981) of information structures the signal in ways that provide dynamic cues about what the speaker is saying (Whalen 1984; Strange 1989; Hawkins 2003; Fowler & Galantucci 2005). They may be disruptive in that overlap has the potential to blur or mask information, making some gestures difficult to recover (Lindblom 1990; Kochetov 2006). Both types of perceptual consequences of coarticulation are expected to contribute to sound change. The emphasis here is on the former consequence, that is, on coarticulated signals in which information about a target gesture, such as the velum lowering gesture for a nasal or tongue dorsum retraction for a lateral, is unambiguously present in the input to the listener. Drawing from, and expanding on, work emerging out of our lab in recent years, I summarize results from "offline" and "online" perception tests showing that listeners closely track the coarticulatory time course of the target gesture. Of particular significance to theories of sound change is that, for some – but by no means all – listeners, the coarticulatory cues are dominant and sufficient cues for making their perceptual decisions. I argue that this variation in perception across listeners is the expected consequence of the many-to-many relation between acoustics and linguistic units due to parallel transmission, and offer a scenario in which listeners for whom the coarticulatory cues are heavily weighted are especially likely contributors to sound change.

## 2. The nature of the input signal

Speakers adjust the spatiotemporal organization of articulatory gestures so that linguistic goals can be achieved under a variety of contextual influences. These adjustments can result in substantial, yet in many respects systematic, variation in the coarticulated signal that serves as input to the listener. Coarticulatory vowel nasalization offers illustrative examples of these adjustments, providing evidence of influences of syllable structure, stress, consonantal context, vowel quality, speech rate and more on the temporal and spatial extent of an anticipatory velum lowering gesture. (See the contributions to the volume by Huffman & Krakow 1993, for examples of these influences). As Ohala (1981, 1993), Lindblom et al. (1995), Harrington et al. (2008) and others have argued, such coarticulated variants serve as the raw material for sound changes in which a property that was originally due to gestural overlap – for example, vowel nasalization in a nasal consonant context, front vowel backing in a coda lateral context, back vowel fronting in an alveolar



context – becomes an inherent characteristic of the signal. That is, the coarticulatory effect is now requisite and the coarticulatory source may be (but is not necessarily) lost.

A critical issue for sound change theorists is to determine the conditions under which these shifts are especially likely to occur. This determination, in turn, requires understanding the nature of coarticulatory variation. As a step in this direction, aimed at understanding the variation that may contribute to the historical change VN > Ṽ, Beddor (2009) conducted a small-scale acoustic study to determine the detailed characteristics of some of the coarticulatorily nasalized vowel variants to which American English listeners are exposed. The durations of acoustic vowel nasalization and of the nasal murmur were measured for a highly restricted set of words: /C(C)ɛnC/ words in which the coda C was one of /t d s z/ (e.g., *bent, bend, dense, dens*). Figure 1 gives the duration measures for the productions of six speakers (approximately 50 tokens per speaker). (See Section 3.1 for explanation of the shaded portion). The three regression lines correspond to $R^2$ from three linear mixed models of vowel nasalization on nasal consonant duration,

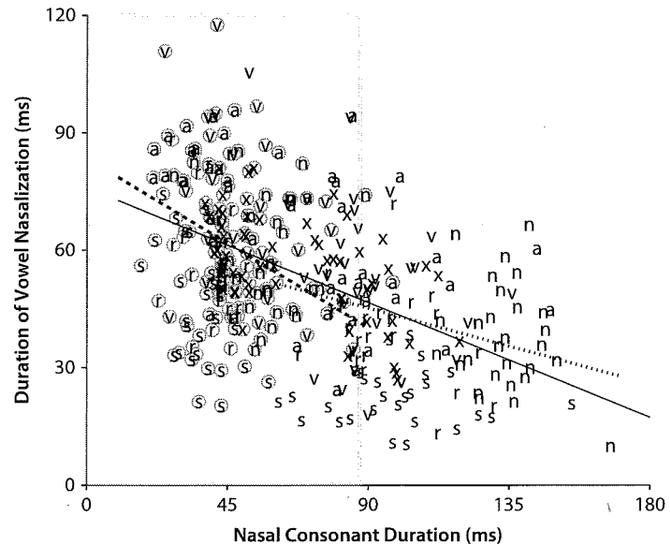


**Figure 1.** Scatter plot of nasal consonant duration by vowel nasalization duration for words containing $VNC_{voiceless}$ (circled letters) and $VNC_{voiced}$ (plain letters) produced by six speakers. Letter type designates speaker. Regression lines correspond to $R^2$ from three linear mixed models (see text): all tokens (solid line) $R^2 = .33$ ($p < .0001$); voiceless (dashed line) $R^2 = .13$ ($p < .0001$); voiced (dotted) $R^2 = .06$ ($p < .005$). Adapted from Beddor (2009)

one model run across voicing contexts and the others within each voicing context. The significant negative correlations indicate that the temporal extent of vowel nasalization covaries with [n] duration both within and across contexts. However, the main generalization that emerges is that vowel nasalization is more extensive, and [n] is shorter, before voiceless than before voiced consonants (see also Malécot 1960; Raphael et al. 1975; Cohn 1990). That is, the data point toward a velum gesture that overlaps more with the oral configuration for the vowel in voiceless contexts and with the oral configuration for the consonant in voiced contexts.

Despite its narrow scope, even this limited study shows that American English-speaking listeners are exposed to considerable variation in anticipatory vowel nasalization. Some of that variation is regular, context-induced variation that could lead listeners to expect an earlier velum gesture in $VNC_{voiceless}$ than in $VNC_{voiced}$ sequences. (See Ohala & Ohala 1991 and Solé 2007 for discussion of the aerodynamic and auditory factors underlying the context effect). Moreover, listeners may be especially likely to expect an early velum gesture in specific lexical items of this structure given that, across speakers, productions of certain words tended toward generally shorter (*spent, sent, sense*) or longer (e.g., *dense, bent*) [n] durations.

In the following sections I consider the consequences of this quasi-systematic variation for perception grammars of coarticulated speech. Because my goals in this work are not restricted to coarticulatory nasalization in English, but are broadly concerned with aligning patterns of production and perception, it is noteworthy that an early velum gesture in contexts with especially short nasal consonants is not unique to voicing contexts nor to English (see, for example, Busà 2007 for Italian; Hattori et al. 1958 for Japanese; and Onsuwan 2005 for Thai). Moreover, preliminary evidence also indicates that variation of the type in Figure 1 is not unique to velum lowering for nasals, but may also hold for tongue dorsum retraction for laterals. Lin, Beddor, & Coetzee (2011) recently conducted an ultrasound study of factors, including voicing, that influence the spatiotemporal characteristics of the tongue tip and dorsum gestures for coda laterals in American English. Like nasal codas, /l/ is shorter when the following consonant is voiceless than when it is voiced. Although our initial analyses have focused on the tongue tip gesture, preliminary analyses of the dorsum gesture for a subset of speakers suggest that retraction for /l/ begins earlier in voiceless (e.g., *help, pelt*) than in voiced (*helm, held*) contexts.

It may be, then, that coda consonants that exhibit particularly extensive variation in coarticulatory overlap with preceding vowels are consonants requiring two supralaryngeal gestures. These gestures are often asynchronous in coda position, with the more open constriction (e.g., velum for nasals, tongue dorsum for laterals) occurring first (Sproat & Fujimura 1993; Browman & Goldstein 1995; Krakow



1999; Byrd et al. 2009). Our data indicate that particularly early onset of the more open constriction often coincides with shorter coda consonants.

Of primary importance for perception grammars is that, for input of the type considered here, listeners have multiple sources of information regarding a coda consonant, and therefore multiple possible weightings of the relevant properties. For a coda nasal, for example, listeners must detect cues corresponding to a lowered velum. But listeners may differ in whether the information for nasality must overlap with the consonantal constriction, the vowel, either configuration, or both. Weightings dependent on voicing context would also be fully consistent with the input data.

## 3. Perception grammars for coarticulated speech

A listener's perception grammar for coarticulated speech includes that listener's weighting of the multiple, dynamic cues for a given linguistic percept (e.g., the percept of *sent* rather than *send* or *set* or perhaps even *scent*).[1] Listeners are expected to attend to the rich information in the input signal afforded by gestural overlap. Yet listeners' attention can nonetheless be selective. For the past several years, my colleagues and I have conducted a series of studies investigating listeners' perceptual weights for cues in coarticulated speech. The downsides to our laboratory approach (as for most laboratory studies) include the absence of interaction with an interlocutor and the absence of non-phonetic cues for deciding what the speaker is saying. Different weights might hold for laboratory speech than for spontaneous interactions in which additional sources of information are available to listeners. Our study of listeners' use of coarticulatory cues has used multiple paradigms – online as well as offline – and multiple sets of stimuli, operating under the assumption that the converging evidence will be reasonably representative of listener behaviors, and of the knowledge underlying those behaviors, in conversational settings.

### 3.1 Listeners' use of coarticulation in real-word categorization tasks

The production data in Section 2 indicate that parallel transmission of vowel and consonant information provides listeners with cues for a coda nasal that are spread

1. The term "perception grammar" has been used by Boersma (1999) and Hamann (2009) to refer to the grammar used by the listener to map from acoustic input (phonetic form) to prelexical phonological form. In both their usage and mine, we are interested in how listener knowledge influences perceptual choices, but my approach does not draw a sharp distinction between perception and word recognition (among other differences).

across the syllable rhyme, although the temporal extent (and likely the spatial magnitude) of the specific cues vary with voicing context. To test listeners' attention to these multiple, context-dependent sources of information, Beddor (2009) created identification and discrimination tests with *bet*, *bent*, *bed*, and *bend* stimuli in which the duration of [n] and duration of coarticulatory vowel nasalization ([ɛ̃]) in naturally produced stimuli were orthogonally varied. The nasal murmur ranged in 10 steps from no nasal (0 ms) to a full nasal murmur (85 ms), and vowel nasalization varied in three steps from oral to 66% nasalized (0 to 124 ms of vowel nasalization). The range of [ɛ̃] and [n] durations, although achieved by cross-splicing in order to control all other aspects of the stimuli, is well-represented by the variation that occurs in natural speech productions, as can be seen by the shaded region of Figure 1. (The lower-left corner of that figure is unpopulated because only /C(C)ɛnC/ productions are represented; that corner would include oral productions such as *bet*, *bed*, etc.).

The perception literature shows that each of the multiple acoustic correlates to a given phonetic distinction contributes to perception and that, in combination, the cues trade off against each other (Repp 1982; Pisoni & Luce 1987). Consistent with this literature, [ɛ̃] and [n] were predicted to be in a trading relation: the more extensive the coarticulatory cue, the shorter the [n] required to elicit a *bent* or *bend* (e.g. rather than *bet* or *bed*) percept. That is, the acoustic cues to the single articulatory gesture, a lowered velum, should cohere in perception. However, coherence can emerge through a variety of weightings, and the relative importance of these cues was predicted – and found – to differ across contexts and listeners.

Broadly characterized, the results of both identification and discrimination tests provided evidence that listeners track acoustic information about the lowered velum gesture and use this information in making perceptual decisions about CVNC vs. CVC. Here I present group and individual listener results for the identification tests, re-configured from Beddor (2009) into perceptual "oral-nasal" spaces.

Listeners identified multiple instances of the 60 stimuli (10 [n] durations x three degrees of vowel nasalization x two voicing contexts) as *bet*, *bent*, *bed*, or *bend*. The results, pooled across 30 native American English speakers, are given in Figure 2. The relative darkness of each cell represents the proportion nasal responses such that the darker the cell, the more *bent* or *bend* responses.[2] Two patterns emerge in the group data. First, as expected, listeners traded information for

2. Relatively small nasal murmur step sizes were used towards the oral end of the continuum due to listener sensitivity to short [n] durations in the voiceless context. The use of larger step sizes at longer durations is consistent with Weber's Law. (Relatively large steps in the duration of vowel nasalization helped keep the experiment to a manageable length).



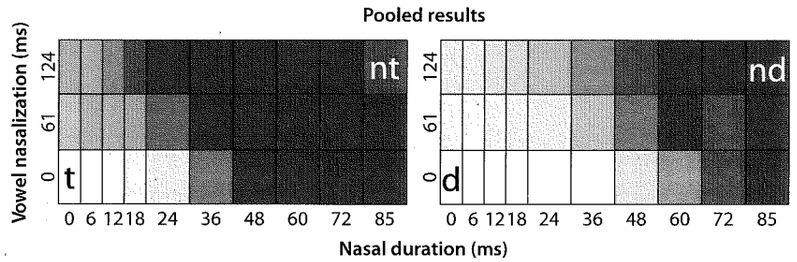


**Figure 2.** Pooled perceptual spaces of 30 listeners based on identification responses to 30 [t]-final *bet-bent* (left) and 30 [d]-final (right) *bed-bend* stimuli. The darker the cell, the higher the proportion nasal (*bent, bend*) responses. (See text for further explanation)

the coarticulatory source (N) and effect (Ṽ): *bent* and *bend* were elicited for increasingly shorter [n] durations as vowel nasalization increased (i.e., cell shading darkens from 0 to 61 to 124 ms of Ṽ). Second, also as expected, voicing context influenced identification. Comparison of the two panels shows that listeners required less nasal consonant duration to perceive *bent* than to perceive *bend*, even in the absence of any coarticulatory cues.

The trading relation between [n] and its coarticulatory effects on the preceding vowel shows that the cues for velum lowering cohere in perception, yielding a unified percept. Moreover, the coherence is exceptionally tight. Typically, only ambiguous stimuli enter into trading relations. However, here [n] durations that elicit unambiguously oral *bet* and *bed* responses when the vowel is oral (e.g., 6 ms of [n] duration in the voiceless context or 36 ms in the voiced) instead elicit predominantly nasal *bent* and *bend* responses when the vowel is heavily nasalized. The context effect – that listeners heard many more stimuli as *bent* than as *bend* for the same range of [n] durations in the two stimulus sets – is in keeping with especially short pre-voiceless [n] in the production data and indicates, in part, that listeners are sensitive to the distributional patterns in the input data. However, nasal murmurs are also more difficult to detect when followed by voicing than when followed by a voiceless closure, which likely further contributes to the perceptual need for longer [n] durations in a voiced context.

The group data demonstrate listeners' sensitivity to the multiple sources of information for a coda nasal as well as their context-sensitive weightings of this information. The individual listener data provide a yet clearer picture of the extent to which the relative perceptual importance of Ṽ and N can vary. Three primary response patterns emerged in the individual data, and these patterns are illustrated by the perceptual spaces of three listeners given in Figure 3. All three listeners show the effect of context (longer [n] required for *bend* than for *bent* percepts) and

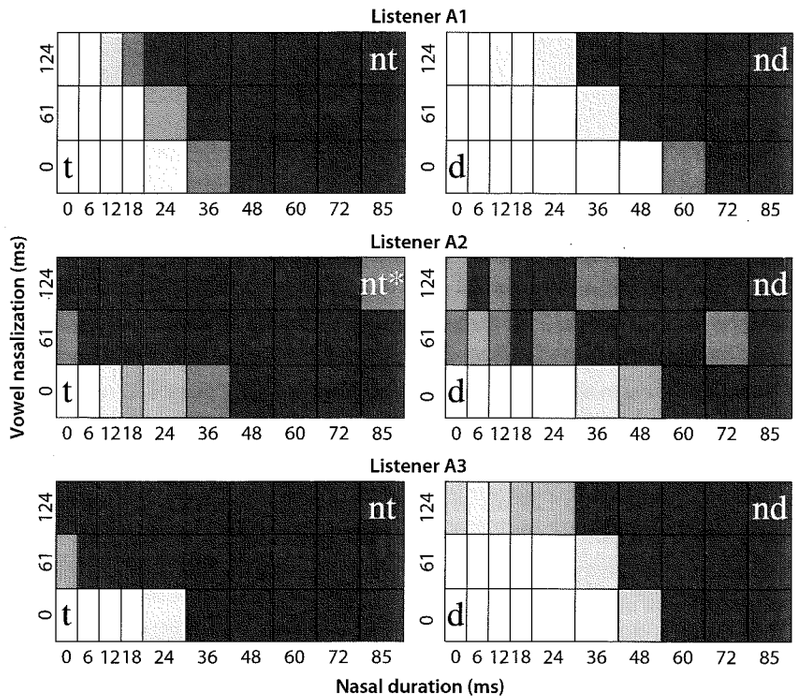


**Figure 3.** Perceptual spaces of three listeners based on identification responses to [t]-final *bet-bent* (left) and [d]-final *bed-bend* (right) stimuli. (See text for explanation of nt*)

the trading relation between Ṽ and [n] duration evident in the group data. Listeners differ substantially, though, in their judgments of many of the stimuli. Listener A1 required some nasal murmur to respond *bent* or *bend*; vowel nasalization alone was not sufficient for this listener to identify the word as containing a nasal consonant. This requirement did not hold for Listener A2, who consistently reported hearing *bent* and *bend* as long as there was some vowel nasalization. Vowel nasalization was also a sufficient cue to elicit CVNC percepts for Listener A3, but only for [t]-final stimuli. For [d]-final stimuli, Listener A3's responses closely mirrored those of Listener A1, with both listeners requiring some nasal murmur to elicit systematic *bend* percepts. The patterns of Listeners A1 and A3 were each representative of slightly over a third of the 30 listeners; the pattern for Listener A2 was less common but clearly held for five participants.

Thus, different listeners systematically accessed a given lexical item through different acoustic information. Stimuli with a nasalized vowel but with no [n] or a very short [n] were unambiguously *bet* for Listener A1 but were equally



unambiguously *bent* for Listener A2. Of the [d]-final stimuli, only 44% are *bend* for Listener A1 compared to 78% for Listener A2. [t]-final stimuli with an especially long nasal murmur were sporadically heard as voiced *bend* by Listener A2 (indicated by nt* in the upper, right-most cell for this listener in Figure 3), despite the voiceless release burst. What "counts" as information for N or NC or voicing differed from listener to listener, and consistently so.

### 3.2 Listeners' use of coarticulation in categorizing nonsense items

I argue in Section 4 that, if we assume that listeners closely attend to the coarticulatory information in the input, perception grammars that diverge along the lines illustrated in Figure 3 are precisely what are expected. That is, listeners' attention to time-varying cues arguably offers an account of the type of across-listener variation that occurs. However, there is much that we do not yet understand about why *particular* listeners select the weights that they do. In an exemplar approach, for example, we would expect listeners' perceptual weights to be driven by their specific experiences, including experiences with these specific lexical items. For the study reported in Section 3.1, very general information about the dialect background of those native English-speaking listeners was collected. Although most participants were from Michigan, and all were living in Michigan at the time of testing, others had grown up in other parts of the U.S. It is possible that different listeners were exposed to systematically different patterns of nasalization for long periods of time. An informal comparison of region of origin and listener group (corresponding to the groups represented by Listeners A1–A3 in Figure 3) was not suggestive of any regional link, but of course our limited knowledge of these listeners' linguistic background gives no indication of specific experiences.

Other perceptual data indicate that, whatever the factors are that determine a particular listener's perception grammar, they apply not only to words that listener has actually heard – and therefore with which she or he may have considerable experience – but to nonsense items as well. Chutamanee Onsuwan and I conducted tests of listeners' perception of the nonsense items *gaba* ([gaba]) and *gamba* ([gãmba]). We again manipulated naturally produced stimuli in which we co-varied nasal consonant duration and the temporal extent of vowel nasalization, creating varying proportions of the signal that were produced with a lowered velum: 0–52% vowel nasalization in four steps and 0–70 ms of [m] murmur in nine steps.[3] The top panel in Figure 4 gives the perceptual space based on identification

3. The identification findings of the *gaba-gamba* experiment reported in this section have not previously been published. However, these same stimuli were used in other experiments, and are described in Beddor (2009).

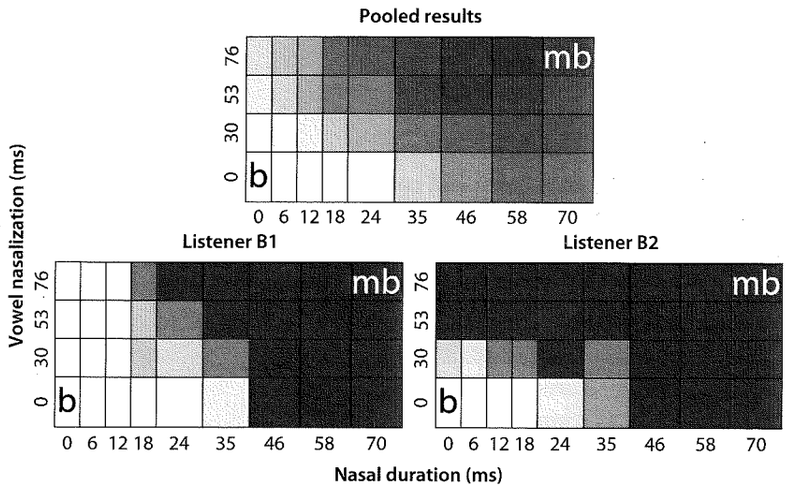


**Figure 4.** Perceptual spaces based on identification responses to *gaba-gamba* stimuli. Top panel: pooled responses of 28 listeners. Lower panels: responses of two individual listeners. The darker the cell, the higher the proportion nasal (*gamba*) responses

responses pooled across a new group of 28 native speakers of American English. The spaces for two individual listeners are given in the bottom two panels. Although I have chosen two particularly distinct respondents, Listeners B1 and B2, to illustrate the different response patterns, each is again representative of a larger group of listeners. (For the *gaba-gamba* continua, about a third of the respondents fell between these two extremes).

The group and individual listener responses to the *gaba-gamba* series show the trade-off between the nasal consonant and its coarticulatory effects that was observed for real words, with increasingly shorter [m] durations eliciting *gamba* responses as the coarticulatory information increases. But the details of the trade-off are distinctly different for Listeners B1 and B2, whose response patterns closely mirror those of Listeners A1 and A2, respectively, for the *bet-bent* and *bed-bend* stimuli. The responses of Listener B2, who required no [m] to hear *gamba* as long as a third or more (i.e., 53 ms or more) of the vowel is nasalized, are particularly interesting in that $VNC_{voiced}V$ is not a context of N – and especially not of [m] – shortening in English, as shown by temporal measures we have taken over the years. That is, native English speakers are not expected to have heard productions of words such as *number*, *jumbo*, *combat*, and *ramble* with no [m] or even with particularly reduced [m]. Nonetheless, the vocalic cues were sufficient to signal *gamba* for Listener B2 (and for several other listeners in that study).



That individual listeners differ systematically from each other in their use of coarticulatory cues in nonsense items is not, of course, an argument against the role of specific linguistic experiences in how listeners assign weights to the properties of coarticulated speech. My colleagues and I have argued elsewhere that native-language coarticulatory patterns shape the extent to which listeners compensate for coarticulation (Beddor & Krakow 1999; Beddor et al. 2002), for example. But, as a research community, we seem to have a better understanding of the influences of broad language-specific patterns on perceptual attentiveness (e.g., Strange 1995) than of the influences of certain other types of experiences. At this point, it is clear that listeners selectively attend to properties in the acoustic input (Foulkes & Docherty 2006), yet we remain at the early stages of determining the mechanisms that govern an individual's selections.

## 3.3 Real time processing of coarticulated speech

Before assessing how perception grammars of the type found for Listeners A2, A3, and B2, in comparison to those of Listeners A1 and B1, might contribute to sound change, I provide evidence that different listeners use coarticulatory information to differing degrees in their moment-by-moment processing of the acoustic input. In this work, conducted in collaboration with Kevin McGowan, Julie Boland, Andries Coetzee, and Anthony Brasher, we are investigating listeners' perception of unfolding CVNC and CVC words (Beddor et al. 2010). Listeners in this audio-visual task are fitted with a head-mounted eye tracker. In each trial, they hear a single auditory stimulus and see, on a computer screen, two pictured objects (e.g., black and white line drawings of a chess set and a nose sniffing for the pair *set/scent*). Participants' eye movements are monitored as they hear instructions to look at one of the pictured objects on the screen.[4] Stimuli are CV(N)C words drawn from minimal quadruplets in which the final C is either [t] or [d] (e.g., *bet/bent/bed/bend*, *set/scent/said/send*). Paired visual stimuli differ either in final voicing (*set-said*, *scent-send*) or presence of a nasal consonant (*set-scent*, *said-send*), but never in both properties.

We hypothesized that listeners would use coarticulatory vowel nasalization to anticipate an upcoming nasal consonant. Upon hearing [CṼNC] (e.g., [sɛ̃nd]), participants should – and did – look to the image corresponding to the CVNC word faster when the visual competitor was of a CVC word (image of *said*) than when it was of another CVNC word (image of *scent*). That is, the latency of initial

4. More precisely, for each trial, listeners first hear the instruction "Look at the pictures". After a 3.5 sec pause, they then hear "Fixate cross" (yellow cross in the center of the screen, to direct their gaze away from the images) and "Now look at [target word]".

correct fixations was, as expected, shorter for audio-visual conditions in which coarticulatory nasalization served as a disambiguating cue. Importantly, inspection of the time course of correct fixations in response to auditory [CṼNC] suggests that, for the pooled data, listeners used the coarticulatory cues as soon as they became available to them. This estimate assumes, in keeping with the literature (Dahan et al. 2001), that it takes roughly 200 ms to program an eye movement. In our pooled data, the proportion of fixations of the image of the CVNC item began to increase approximately 200 ms after the onset of vowel nasalization.

If the weights that listeners assign to Ṽ and N in identification and discrimination tasks are indicative of how informative these components of the signal are in conversational interactions, then listener-specific and context-specific patterns should – and, again, did – emerge in this real-time processing task. As an illustration, Figure 5 gives the results of two sets of trials in which the auditory stimulus had a nasalized vowel, with vowel nasalization beginning, on average, 106 ms after stimulus onset. In the trials represented in the left panel, the original [n] was retained (e.g., [sẽnd], [sẽnt]). For these trials, the mean proportion fixations of the CVNC image (relative to all fixations at each point in time) show a nearly identical time course for voiced and voiceless contexts. In the trials represented in the right panel, the auditory stimuli were identical to trials represented in the left panel except that the entire [n] was excised (e.g., auditory [sẽd], [sẽt]). [n] deletion resulted in significantly fewer fixations of the CVNC image overall but with an especially large decrease for the voiced (CV(n)d) contexts.

Although nearly all of the 26 English-speaking participants were more likely to look at the CVNT than the CVND image in the deleted-[n] condition, this general outcome was achieved in different ways by different listeners, as shown by the

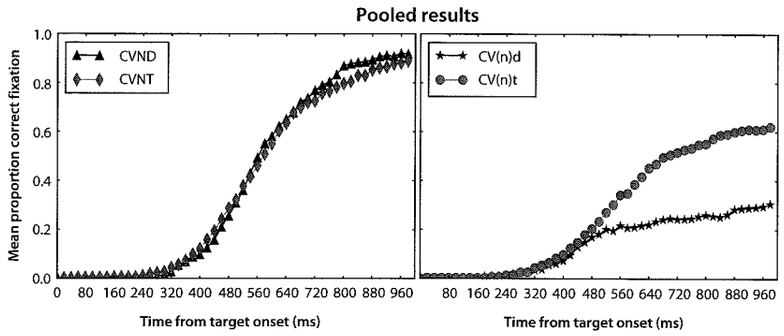


**Figure 5.** Proportion pooled (across 26 listeners) fixations of target CVNC image (e.g., *send*) over time when visual competitor was corresponding CVC image (*said*). Left: auditory stimulus was [CṼND] or [CṼNT]. Right: auditory stimulus [CṼD] or [CṼT], with [n] deleted



individual listener results in Figure 6. The point at which the original [n] was excised was, on average, 206 and 248 ms after target onset for CV(n)t and CV(n)d, respectively. (Recall that these deletions are not expected to influence fixations for another 200 ms or more). Listeners C1, C2 and C3 respond similarly to the CV(n)t (auditory [CṼt]) condition in that they used the coarticulatory cue in the vowel and continued to look at the CVNT image (as opposed to the competitor CVT) after the point at which [n] should have occurred. These listeners differ, though, in the extent to which CV(n)d (auditory [CṼd]) activates CVND lexical items. Listeners C1 and C2 both used the information in Ṽ to activate the CVND item, but Listener C2 then looked away in the absence of [n] (as shown by the decrease in CV(n)d fixations after the peak at 640 ms). Listener C3 did not use the coarticulatory information in Ṽ to activate the CVND item and, unexpectedly, Listener C4 did not use the coarticulatory cues in [CṼC] regardless of whether the final C was voiced (CV(n)d) or voiceless (CV(n)t). (Here again each listener is representative of a larger group, although only one other listener was as extreme as Listener C4 in not using the coarticulatory information in either voicing context).

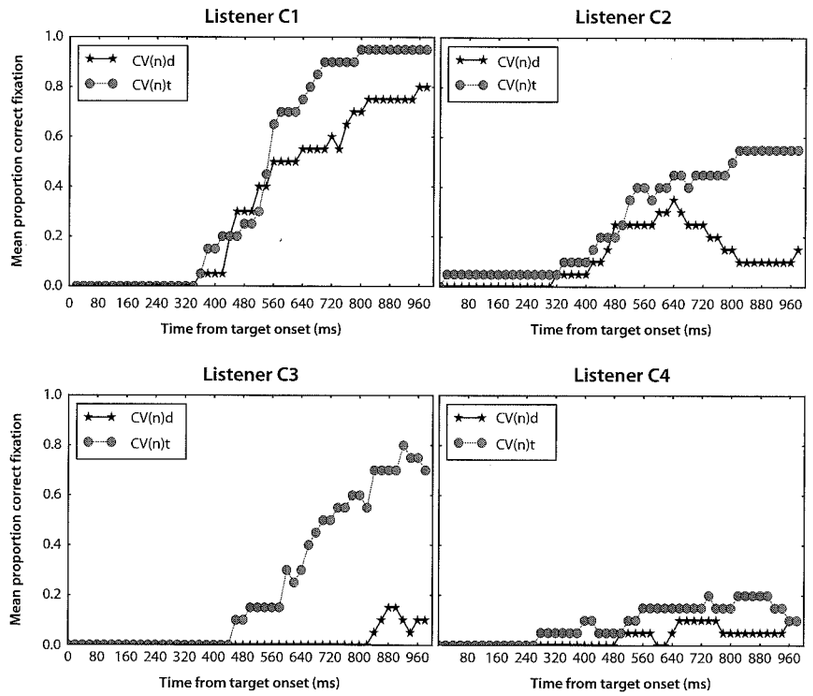


**Figure 6.** For four listeners, proportion of fixations of target CVNC image over time when visual competitor was corresponding CVC image. Auditory stimulus was [CṼD] or [CṼT], with [n] deleted

In summary, listeners' eye movements to target CVNC images, as opposed to competitor CVC images, confirm that, for most listeners, anticipatory nasalization speeds the time course of activation of CVNC words. For a subset of these listeners, the coarticulatory cues need to be reinforced by [n] for continued fixation of the CVNC image, and this reinforcement is especially important in voiced contexts, where N in American English speakers' productions tends to be relatively long and not prone to deletion (Section 2). Thus the time course of listeners' processing of the unfolding coarticulated signal closely parallels the perceptual spaces calculated from listeners' offline categorization of real word and nonsense stimuli.

## 4. Perception grammars of coarticulated speech and sound change

I suggest that two key factors contribute to these robustly distinct perception grammars for different listeners. The first factor is the nature of the overlapping information in the coarticulated signal. Listeners use this time-varying information, but they differ in the perceptual weights they assign to the linguistic units that carry cues for a lowered velum in large part because they *can* differ, that is, because the preponderance of the data, particularly in voiced contexts, contain both sources of information. In voiceless contexts, where the nasal consonant is a less reliable cue, all listeners had weaker requirements for [n] as long as the vowel was nasalized. Moreover, the *co*variation in the input signal, discussed in Section 2, would seem to be especially conducive to across-listener differences in perception grammars for nasals and nasalization in that the vocalic cues are especially strong in some contexts and for some speakers, whereas the consonantal cues are particularly strong under other conditions and for other speakers. That is, although the exposure of a given listener to specific input patterns surely contributes to that listener's selective attention, the covarying nature of the coarticulated signal itself contributes to selectively attending to specific properties of the input. Different weights can be assigned to the relevant acoustic dimensions (e.g., heavy weighting of vowel nasalization even in contexts in which a nasal consonant is regularly produced) and yet yield the same linguistic judgments, under many – although, crucially for sound change, not all – circumstances.

A second factor contributing to different perception grammars is the nature of perception of coarticulated signals. Listeners only have choices between the weights for $\tilde{V}$ and N, for example, if they attend to both properties. And, typically, they do. The results of the eye-tracking study indicate that most listeners are using the coarticulatory information as soon as it becomes available (see also Dahan & Tanenhaus 2004). Sensitivity to the relative timing of gestural events, such as the timing of velum lowering in relation to the oral configuration, facilitates



determining what has been said and may also increase the likelihood that different listeners will assign different perceptual weights to the same set of inputs.

Up to this point, this chapter has begged the question of what constitutes a sound change. The literature tends not to explicitly address this question, perhaps in part because of a tacit assumption that, if two listeners or two groups of listeners differ systematically in perception, then they will exhibit corresponding differences in their productions (and vice versa). I address the issue of isomorphism between perception and production grammars below. However, for the purposes of the present discussion, I assume that, in sound change, *sound* – i.e., production and the resulting acoustic signal – must change. Thus, for an ongoing change, the productions of one group of speakers must differ, in a regular way, from that of other speakers of that linguistic variety. For evolving vowel nasalization, these regularities would presumably include extensive overlap of the lowered velum gesture with the vowel and very little, if any, overlap with a consonantal constriction. As another example, the innovative speakers for evolving /l/ vocalization would be expected to produce laterals with greater tongue dorsum retraction, and a reduced or absent tongue tip gesture (e.g., Recasens & Espinosa 2010).

How, then, do perception grammars contribute to sound (i.e., production) change? For listeners to foster change, their perception grammars cannot simply mirror the statistical patterns of variation present in the input signal. As suggested above, across-listener perceptual differences of the type reported here are not solely the consequence of corresponding across-listener differences in the acoustic input for $\tilde{V}N$ sequences. Exemplar and other speech perception theorists agree that listeners, in determining (and categorizing) what speakers are saying, are not simply storing input signals. Listeners with broadly similar acoustic inputs for the relevant sequences or lexical items effectively transform the input via different weightings. For those who might be labeled the innovative listeners, $\tilde{V}$ is a sufficient and dominant cue. On the most conservative end are listeners who require N and for whom $\tilde{V}$, as shown by their real-time processing, does not appear to activate CVNC words. Intermediate are listeners who perceptually use $\tilde{V}$ but for whom N is also necessary. (The labels "innovative" and "conservative" are relative to the development of distinctive vowel nasalization).

However, perception grammars only contribute to sound change, as defined here, if they are publicly manifested. Public manifestation need not entail that the listeners-turned-speakers exactly replicate the perceptual weightings in their productions. As Mark Hale (p.c.) has suggested, perception grammars can be manifested through other interactions with interlocutors, such as confusions about what a speaker has said. The eye-tracking data in Figure 6 are suggestive of the types of confusions that intermediate and conservative listeners might experience: [C$\tilde{V}$d] led to uncertainty for (intermediate) Listener C2; [C$\tilde{V}$d] and [C$\tilde{V}$t] led to CVC

decisions for (conservative) Listener C4. Presumably the additional information available in conversational settings would minimize confusions, but a mismatch between a speaker's productions and the listener's perception grammar can nonetheless be expected to influence interactions (and possibly elicit explicit comments from the listener). Of greater interest are potential confusions of innovative listeners, for whom Ṽ is a dominant cue under conditions of little vowel nasalization in the input; here our eye-tracking data are not informative because all stimuli corresponding to CVNC images in that study had moderate to heavy vowel nasalization.

Of course, a more systematic public manifestation of perception grammars would occur if those grammars were reflected in listener-turned-speaker differences in production. Production surely must align to *some* degree with perception. A conservative listener for whom [bɛ̃t] does not access *bent* would seem unlikely to produce that realization with any regularity. In contrast, productions of innovative listeners might be expected to be characterized less by what they do not produce than by a relatively wide range of variation in what they *do* produce. This expectation is based on the finding that no listener in our experiments failed to use [n] duration in their perceptual judgments; moreover, listeners for whom Ṽ was a sufficient cue did not require especially long [n] durations when Ṽ was unavailable. Figure 3, for example, shows that innovative Listener A2 hears *bend* over a considerably wider range of stimuli than do more conservative Listeners A1 and A3. If Listener A2's productions mirror that listener's perception, then *bend* articulations should be similarly variable. This perspective is consistent with Labov's (2007) account of advancement of change by incrementation, according to which children both reproduce and advance the system of their parents.

I have not yet conducted the requisite comparison of perception and production by the same group of participants to substantiate these expectations. Production and perception have recently been directly compared by Harrington and colleagues, who analyzed /u/ and /ʊ/ as produced and perceived by older and younger speakers of Standard Southern British English in a study of ongoing back vowel fronting (Harrington et al. 2008; Kleber et al. in press). Their data for /u/ (the more advanced change in this process) showed close parallels between production and perception for each of the two age groups, whereas the /ʊ/ results were suggestive of a misalignment in which production lags behind perceptual differences between the groups (see also Harrington, this volume). The present study calls for a comparable investigation of nasalization. However, among the many differences between ṼNC in American English and back vowels in Southern British English is that the former, despite exhibiting great variation, may well be in a stable pattern of covariation between Ṽ and N rather than in a state of change. My aim here has been to delineate how variation in perception grammars could emerge from this type of variation, and how these grammars might, in the future, contribute to change. This



account of change is not as restrictive as it might initially appear because, as noted in Section 2, covariation between coarticulatory source and effect is not unique to English or to voicing contexts, and likely not to nasalization.

In summary, I have speculated that coda consonants for which two supralaryngeal gestures must be coordinated, as in the case of nasals, might be especially likely to yield substantial coarticulatory variation. Listeners are sensitive to the distributional patterns within the variation. Overall, and unsurprisingly, they are more likely to use coarticulatory information (here, Ṽ) in contexts where the source of coarticulation (N) is reduced, and are less likely to do so in contexts where the source is reliably present. However, that there are multiple yet variably realized input cues means that the attentive listener has perceptual choices: different weights of coarticulatory source and effect are compatible with the input. These different weights emerge in the responses of individual listeners. Irrespective of whether listeners are responding to real words or nonsense items, there are more conservative listeners who primarily use the information from the coarticulatory source and more innovative listeners who heavily weight the coarticulatory effects. These weights shape how listeners categorize, discriminate, and access words in real time. The perception grammars of innovative listeners have strong potential to contribute to sound change in that they are likely manifested in conversational interactions either through their expectations about coarticulated speech or through their own productions. These innovations are not unexpected, but are rather the predicted outcome of listeners' close but selective attention to the dynamic coarticulated signal.

## References

Beddor, Patrice Speeter. 2009. "A Coarticulatory Path to Sound Change". *Language* 85.285–821.

Beddor, Patrice Speeter & Rena Arens Krakow. 1999. "Perception of Coarticulatory Nasalization by Speakers of English and Thai: Evidence for partial compensation". *Journal of the Acoustical Society of America* 106.2868–2887.

Beddor, Patrice Speeter, James Harnsberger & Stephanie Lindemann. 2002. "Acoustic and Perceptual Characteristics of Vowel-to-Vowel Coarticulation in Shona and English". *Journal of Phonetics* 30.591–627.

Beddor, Patrice Speeter, Kevin B. McGowan, Julie Boland & Andries Coetzee. 2010. "The Perceptual Time Course of Coarticulation". Poster presented at the 12th Conference on Laboratory Phonology, University of New Mexico, Albuquerque, July 2010.

Boerma, Paul. 1999. "On the Need for a Separate Perception Grammar". Manuscript, Rutgers University.

Browman, Catherine P. & Louis M. Goldstein. 1995. "Gestural Syllable Position Effects in American English". *Producing Speech: Contemporary issues. For Katherine Safford Harris* ed. by Fredericka Bell-Berti & Lawrence J. Raphael, 19–33. New York: AIP.

Busà, M. Grazia. 2007. "Coarticulatory Nasalization and Phonological Developments: Data from Italian and English nasal-fricative sequences". *Experimental Approaches to Phonology* ed. by Maria-Josep Solé, Patrice Speeter Beddor & Manjari Ohala, 155–174. Oxford: Oxford University Press.

Byrd, Dani, Stephen Tobin, Erik Bresch & Shrikanth Narayanan. 2009. "Timing Effects of Syllable Structure and Stress on Nasals: A real-time MRI examination". *Journal of Phonetics* 37.97–110.

Cohn, Abigail C. 1990. "Phonetic and Phonological Rules of Nasalization". *UCLA Working Papers in Phonetics* 76.1–224.

Dahan, Delphine, James S. Magnuson, Michael K. Tanenhaus & Ellen M. Hogan. 2001. "Subcategorical Mismatches and the Time Course of Lexical Access: Evidence for lexical competition". *Language and Cognitive Processes* 16.507–534.

Dahan, Delphine & Michael K. Tanenhaus. 2004. "Continuous Mapping from Sound to Meaning in Spoken-Language Comprehension: Immediate effects of verb-based thematic constraints". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30.498–513.

Foulkes, Paul & Gerard Docherty. 2006. "The Social Life of Phonetics and Phonology". *Journal of Phonetics* 34.409–438.

Fowler, Carol A. & Bruno Galantucci. 2005. "The Relation of Speech Perception and Speech Production". *The Handbook of Speech Perception* ed. by David B. Pisoni & Robert E. Remez, 633–652. Oxford: Blackwell.

Hamann, Silke. 2009. "The Learner of a Perception Grammar as a Source of Sound Change". *Phonology in Perception* ed. by Paul Boersma & Silke Hamann, 111–149. Berlin: Mouton de Gruyter.

Harrington, Jonathan. This volume. "The coarticulatory basis of diachronic high back vowel fronting".

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2008. "Compensation for Coarticulation, /u/-Fronting, and Sound Change in Standard Southern British: An acoustic and perceptual study". *Journal of the Acoustical Society of America* 123.2825–2835.

Hattori, Shirô, Kengo Yamamoto & Osamu Fujimura. 1958. "Nasalization of Vowels in Relation to Nasals". *Journal of the Acoustical Society of America* 30.267–274.

Hawkins, Sarah. 2003. "Roles and Representations of Systematic Fine Phonetic Detail in Speech Understanding". *Journal of Phonetics* 31.373–405.

Huffman, Marie K. & Rena A. Krakow, eds. 1993. *Phonetics and Phonology: Nasals, nasalization, and the velum*. New York: Academic Press.

Kleber, Felicitas, Jonathan Harrington & Ulrich Reubold. In press. "The Relationship between the Perception and Production of Coarticulation during a Sound Change in Progress". *Language & Speech*. DOI: 10.1177/0023830911422194

Kochetov, Alexei. 2006. "Syllable Position Effects and Gestural Organization: Articulatory evidence from Russian". *Laboratory Phonology 8: Varieties of phonological competence* ed. by Louis M. Goldstein, D. H. Whalen, & Catherine T. Best, 565–588. Berlin: de Gruyter.

Krakow, Rena A. 1999. "Physiological Organization of Syllables: A review". *Journal of Phonetics* 27.23–54.

Labov, William. 2007. "Transmission and Diffusion". *Language* 83.344–387.

Lin, Susan S., Patrice Speeter Beddor & Andries W. Coetzee. 2011. "Gestural Reduction and Sound Change: An ultasound study". *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, 17–21 August 2011* ed. by Wai-Sum Lee & Eric Zee, 1250–1253. Hong Kong: City University of Hong Kong.



Lindblom, Björn. 1990. "Explaining Phonetic Variation: A sketch of the H&H theory". *Speech Production and Speech Modelling* ed. by William J. Hardcastle and Alain Marchal, 403–439. The Netherlands: Kluwer Academic.

Lindblom, Björn, Susan Guion, Susan Hura, Seung-Jae Moon, & Raquel Willerman. 1995. "Is Sound Change Adaptive?" *Rivista di Linguistica* 7:1.5–36.

Malécot, André. 1960. "Vowel Nasality as a Distinctive Feature in American English". *Language* 36.222–229.

Mattingly, Ignatius G. 1981. "Phonetic Representation and Speech Syntesis by Rule". *The Cognitive Representation of Speech* ed. by Terry Myers, John Laver, & John Anderson, 415–420. Amsterdam: North-Holland Publishing.

Ohala, John J. 1981. "The Listener as a Source of Sound Change". *Papers from the Parasession on Language and Behavior* ed. by Carrie S. Masek, Roberta A. Hendrick, & Mary Frances Miller, 178–203. Chicago: Chicago Linguistic Society.

Ohala, John J. 1993. "Coarticulation and Phonology". *Language and Speech* 36.155–170.

Ohala, Manjari & John J. Ohala. 1991. "Nasal Epenthesis in Hindi". *Phonetica* 48.207–220.

Onsuwan, Chutamanee. 2005. *Temporal Relations between Consonants and Vowels in Thai Syllables*. Ph.D. dissertation, University of Michigan.

Pisoni, David B. & Paul A. Luce. 1987. "Trading Relations, Acoustic Cue Integration, and Context Effects in Speech Perception". *The Psychophysics of Speech Perception* ed. by Marten E. H. Schouten, 155–172. Dordrecht: Martinus Nijhoff.

Raphael, Lawrence J., M. F. Dorman, Frances Freeman & Charles Tobin. 1975. "Vowel and Nasal Duration as Cues to Voicing in Word-Final Sop Consonants: Spectrographic and perceptual studies". *Journal of Speech and Hearing Research* 18.389–400.

Recasens, Daniel & Aina Espinosa. 2010. "A Perceptual Analysis of the Articulatory and Acoustic Factors Triggering Dark /l/ Vocalization". *Experimental Phonetics and Sound Change* ed. by Daniel Recasens, Fernando Sánchez Miret & Kenneth J. Wireback, 71–82. München: Lincom Europa.

Repp, Bruno H. 1982. "Phonetic Trading Relations and Context Effects: New experimental evidence for a speech-mode of perception". *Psychological Bulletin* 92.81–110.

Solé, Maria-Josep. 2007. "Compatibility of Features and Phonetic Context: The case of nasalization". *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 6–10 August 2007* ed. by Jürgen Trouvain & William J. Barry, 261–266. Saarbrucken: Universität des Saarlandes.

Sproat, Richard & Osamu Fujimura. 1993. "Allophonic Variation in English /l/ and Its Implications for Phonetic Implementation". *Journal of Phonetics* 21.291–311.

Strange, Winifred. 1989. "Evolving Theories of Vowel Perception". *Journal of the Acoustical Society of America* 85.2081–2087.

Strange, Winifred. 1995. "Cross-Language Studies of Speech Perception: A historical review". *Speech Perception and Linguistic Experience* ed. by Winifred Strange, 3–45. Timonium, MD: York Press.

Whalen, D. H. 1984. "Subcategorical Phonetic Mismatches Slow Phonetic Judgments". *Perception & Psychophysics* 35.49–64.