Beddor, P., A. Brasher & Narayan, C. (2007). Applying perceptual methods to phonetic variation and sound change. In M.J. Solé et al. (Eds.), Experimental Approaches to Phonology. OUP: Oxford. (p.127-143).

9

Applying Perceptual Methods to the Study of Phonetic Variation and Sound Change

Patrice Speeter Beddor, Anthony Brasher, and Chandan Narayan

Linguists have long recognized phonetic variation as a key factor in sound change (e.g. Paul 1880; Baudouin de Courtenay 1972*a*), and researchers' ability to study phonetic variation continues to evolve with methodological advances. For example, the use of conversational corpora to study variation in connected speech has opened the door to large-scale investigation of the ways in which variation in speech production mirrors patterns of sound change. At the same time, a comparable large-scale database is not available, nor possible, for speech perception. For natural communicative settings, even if researchers had access to real-time processing information, we could not determine—short of miscommunications—how listeners categorized naturally occurring variants independent of their meaningful, real-world, context nor would we know how these variants are discriminated or confused. Indeed, the closest that researchers come to a large-scale naturalistic perceptual database are sound changes themselves—that is, the very patterns we seek to explain. Consequently, systematic investigation of the ways in which perception of phonetic variation mirrors patterns of sound change remains largely within controlled laboratory settings, often using refinements of methods originally designed to answer core questions in speech perception.

It is not surprising that experimental methods developed to address theoretical issues in speech perception can be directly applied to questions about sound change. A main challenge to perceptual theorists over the past 40 or 50 years has been to understand how seemingly highly variable acoustic properties, sometimes spread across long portions of the signal, give rise to relatively constant percepts. Such

The research reported here was supported in part by NSF Grant No. BCS-0118684 to the first author. We thank Kathy Welch for assistance in statistical analysis and Sam Epstein, Sally Thomason, members of the University of Michigan phonetics–phonology research group, and participants in the graduate seminar on phonetic paths to sound change for helpful discussions. We also gratefully acknowledge the helpful comments of John Ohala, Manjari Ohala, Maria-Josep Solé, and an anonymous reviewer on an earlier draft of this paper.

investigations closely parallel the goals of the phonologist or phonetician who seeks to understand the conditions under which listeners are more or less sensitive to phonetic variants that have the potential to lead to sound change.

In this chapter, the phonetic variation under investigation is the structured variability introduced by phonetic context. In laboratory and natural communicative settings, listeners use contextual variation to determine what speakers are saying. Such variation, however lawful, means that not all speaker–listeners in a given speech community arrive at the same phonological grammar for their speech variety, and we are especially interested in the perceptual mechanisms by which such differences might arise. The chapter begins with a highly selective overview of certain basic patterns of experimental findings from speech perception, and their relation to patterns of sound change. The major focus of the paper is to apply a well-established perceptual paradigm to new questions about the role of perception in contextually conditioned sound changes, and to interpret findings emerging from this method in terms of a theoretical account of the phonetic motivation for these changes.

9.1 PERCEPTION OF CONTEXTUAL VARIATION

An issue of central interest to speech perception researchers is to understand how listeners achieve perceptual constancy—for example, how they perceive the same phonological category—across the variations introduced by overlapping articulations for adjacent or nearby sounds. The basic perceptual paradigm for investigating coarticulatory variation is to embed target speech sounds, usually varying along a single acoustic dimension, in different phonetic contexts to determine the effect, if any, of context on target identification or discriminability. Across the wide variety of contexts that have been investigated, listeners' responses under these test conditions are consistent with the interpretation that listeners perceptually reduce or factor out the acoustic effects of a coarticulatory context on the target sound, apparently attributing these effects to their contextual source (e.g. Mann and Repp 1980; Martin and Bunnell 1981; Whalen 1989; Fowler *et al.* 1990; Manuel 1995).

J. Ohala (1981*b*, 1993*a*, 2003) recognized the implications for sound change of perceptual compensation for contextual variation. He argued that, while listeners normally adjust for systematic variation, not adjusting could lead listeners to perceive variation as intrinsic to the target. In that situation, listeners might correctly perceive "how" the target was produced (i.e. the phonetic details), but their linguistic interpretation of the relevant sequence of sounds could differ from what the speaker intended. Consider, for example, what happens when English-speaking listeners judge nasalized vowels (which, predictably, although not exclusively, occur in nasal consonant contexts in English), but the expected flanking nasal consonant is either attenuated or entirely absent. Kawasaki (1986) found that nasal vowels in a nasal consonant context ([NṼN]) sounded quite oral to listeners but, as she lowered the intensity—and hence detectability—of the nasal consonants, listeners were increasingly likely to identify the vowel as nasal. Along similar lines, Beddor and Krakow (1999) found that listeners were accurate in rating acoustically identical nasal vowels as "equally nasal" when both were in non-nasal contexts (e.g. [CṼC]–[Ṽ]), but were much less accurate in rating the same vowels when one, but not the other, was in a nasal context ([NṼN]–[Ṽ]).

Yet the experimental picture is not simply that listeners attribute a phonetic property to its coarticulatory source when the source is detected, and otherwise perceive that property as intrinsic to the target. The more complex outcome that has emerged in recent years is that listeners' perceptual adjustments for contextual variation are partial rather than complete. Here again results for vowel nasalization are illustrative. For example, a second experiment by Beddor and Krakow (1999) tested listeners' judgments of nasal and oral vowels using a discrimination task designed to be a more sensitive perceptual measure than the metalinguistic rating task described above. Listeners' responses were again context-dependent in that judgments were least accurate when one vowel was in a nasal and the other in a non-nasal context; at the same time, discrimination of vowels in such pairings as [NṼN]–[V] was consistently above chance, suggesting that listeners attributed some but not all of the context-dependent variation to a (clearly audible) coarticulatory source. Fowler and Brown (2000) investigated similar types of pairings (e.g. [CṼNə]–[CVCə]) and likewise found that listeners' accuracy and reaction times in a vowel discrimination task indicated that compensation for coarticulatory influences was partial.

The imperfect adjustment of listeners for the acoustic variation introduced by overlapping articulations is not surprising and can be viewed as the normal result of perceptual processing of phonetic properties, whose realization depends on many factors. Yet such contextual "residue"—that is, the acoustic effects of coarticulation not attributed to context—has implications for sound change. On the one hand, having contextual residue on the target sound means that change might take place even when the coarticulatory source is detected. Indeed, this is as expected, since not all assimilatory changes involve loss of the conditioning environment. For example, hearing nasal vowels as nasal even when flanked by a detected nasal consonant could create, for the learner, ambiguity in terms of the primary site of nasalization, Ṽ or N. In this regard, we note that, in some languages with contrastive nasal vowels, these vowels are followed by short epenthetic nasal consonants in certain contexts (e.g. M. Ohala and J. Ohala 1991; Shosted 2006*a*); thus phonetic [ṼN] sequences may correspond with phonological /Ṽ/ in some languages and /VN/ in others. On the other hand, perceived contextual residue also means that, when the coarticulatory source is *not* detected, the intended utterance might still be correctly perceived because listeners are presumably accustomed to associating the context-dependent properties (in this case, vowel nasalization) with the sporadically undetected source (the nasal consonant).

In the remainder of this chapter we present an experimental approach to contextual variation that recognizes that listeners retain sensitivity to at least some of the fine-grained phonetic effects of a coarticulatory source on the target sound. That listeners have access to phonetic details is a view shared by exemplar-based models of phonology (e.g. Pierrehumbert 2001*a*; Johnson, in this volume). The view is shared as well by Lindblom *et al.* (1995), although these researchers propose that listeners are only occasionally aware of the unprocessed acoustic form of the input signal, such as when demands on intelligibility are low or when intelligibility is less important than, say, the sociolinguistic information conveyed by phonetic details. In contrast, we see sensitivity to phonetic detail as being systematic, at least for some types of phonetic variation. Our assumption of systematic awareness is based in part on experimental findings such as those just described, and in part on the considerable variation present in the input signal—variation that can render the information contained in such details important to perception. That is, although coarticulatory variation is often viewed as redundant information about segments that are further up or down the speech stream, segments are often deleted in conversational speech. (In one large-scale word corpus of conversational American English, for example, over 20 percent of the words had one or more segments deleted; Johnson 2005*a*.) Residual phonetic cues from these segments thus might be essential, rather than redundant, information in casual speech.

9.2 INVESTIGATING CONTEXTUAL CO-VARIATION

In our current investigations of the ways that phonetic variation may lead to different phonological grammars and to new lexical forms in a speech variety, we are studying variation in segmental durations, especially context-dependent variation that results in extreme segmental shortening and, at times, deletion of the target segment. It is well known that different stress, phrasal, and phonetic contexts trigger different segmental durations (e.g. Klatt 1976). As discussed below, there is evidence from production that shorter durations of a given segment co-occur with particularly extensive coarticulatory overlap of that segment with surrounding sounds. Our first step has been to further investigate this relation in production (Section 3.1). Our next step has been to explore whether such co-variation in production leads to a comparable trade-off in perception, such that the target segment and the coarticulatory influences of that segment are, to some extent, perceptually equivalent (Section 3.2). We hypothesize that articulatory co-variation and perceptual equivalence between segmental duration and coarticulatory details are important phonetic factors in sound changes in which the segment that conditioned a coarticulated variant is lost while that segment's coarticulatory (now distinctive) effects are retained.

To date, we have studied co-variation between nasal consonant duration and the extent of coarticulatory vowel nasalization, that is, variation possibly relevant to the relatively common sound change VN > Ṽ. However, we expect that the general properties under investigation will hold for conditioned sound changes involving other types of coarticulation. In otherwords, there is no reason to expect that either the coarticulatory or the perceptual mechanisms under investigation here are specific to a given articulator or set of acoustic properties.

9.2.1 Co-variation in production

For some articulatory movements, segmental shortening in certain phonetic contexts is offset in part or in entirety by increased temporal overlap of that movement with one or more flanking segments. Velum lowering for a nasal consonant in English provides clear evidence of this phenomenon. Nasal consonants are shorter before voiceless than before voiced consonants in English, especially when the NC sequence is tautosyllabic, and vowels are correspondingly more nasalized when followed by $NC_{voiceless}$ than by NC_{voiced}. The scatterplots in Figure 9.1 illustrate this relation between segmental duration and extent of coarticulation for the productions of two speakers of American English.[1] The speakers were recorded producing a randomized list of words containing /ɛnC/ sequences, where C was either a voiced or voiceless obstruent (e.g. *spend, spent, bend, bent, dens, dense*). The acoustic measures included duration of the nasal consonant and the temporal extent—that is, the acoustic duration—of vowel nasalization. Acoustic duration of vowel nasalization was assessed by inspecting FFT spectra in 10 ms increments throughout the course of the vowel; nasalization onset was identified as the first spectrum with an identifiable low-frequency nasal formant (which increased in amplitude in subsequent spectra) and/or a broadening of F1 bandwidth accompanied by lowering of F1 amplitude (e.g. Maeda 1993, Stevens 1998). These changes in the FFT spectra were verified against the corresponding wideband spectrogram and the waveform display, with the latter typically showing a decrease in overall vowel amplitude at the onset of vowel nasalization.[2]

As shown in Figure 9.1, Speaker 2's vowels are overall more nasalized than are those of Speaker 1, but both speakers' productions exhibit an inverse relation between the acoustic duration of vowel nasalization and the duration of a following nasal consonant. This inverse relation holds not only across the voiced and voiceless contexts, but also (albeit to a lesser extent) within the voiceless context, where $R^2 \approx 0.2$ for each speaker's $VNC_{voiceless}$ tokens (indicated by squares in Fig. 9.1). Similar patterns of temporal co-variation between a nasal consonant and anticipatory vowel nasalization in English also hold for Malécot's (1960) acoustic measures and Cohn's (1990) aerodynamic data.

Importantly, this pattern of co-variation is not unique to English. In some Italian dialects, for example, nasal consonants are shorter and nasalization extends through

[1] The production and perception data reported here belong to a larger set of experiments conducted across languages. See Beddor (forthcoming) for full-scale presentation of the methods and results.

[2] The vowel nasalization measure identifies the point at which acoustically detectable nasalization appears after the initial oral consonant. Comparison of acoustic measures and airflow data (for another group of speakers) is under way.

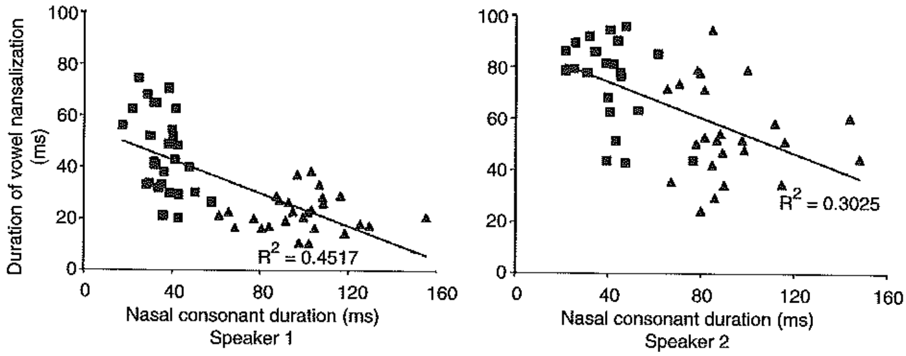

FIGURE 9.1. Scatterplots showing the inverse relation (indicated by the regression line in each plot) between nasal consonant duration and the temporal extent of vowel nasalization for /C(C)εNC/ productions of two American-English speakers in which the coda C was either voiceless (squares) or voiced (triangles). (See text for an explanation of the measure of acoustic vowel nasalization.)

more of the preceding vowel in fricative (e.g. /Vns/) than in stop (/Vnt/, /Vnd/) contexts (Busà 2003 and in this volume); Japanese sequences exhibit a similar relation (Hattori *et al.* 1958). In Thai, which has contrastive vowel length, nasal codas are relatively short after long vowels, but these long vowels are produced with relatively extensive nasalization (i.e. vowel nasalization extends through more of the vowel in V:N than in VN: sequences; Onsuwan 2005). Thus an inverse temporal relation between a nasal coda and its coarticulatory effects holds across various types of coda shortening (or, conversely, lengthening) processes in various languages.

9.2.2 Testing perception of co-variation

Consider, then, the task of a listener conversing with Speakers 1 and 2 (Fig. 9.1), that is, with speakers whose nasal consonants range from quite short (even under laboratory recording conditions) to long and for whom extent of vowel nasalization is likewise highly variable. Presumably the listener–learner must make decisions concerning whether, for these sequences, the speaker intended /εnC/, /εC/, or /ε̃C/; the listener must decide as well if the coda C is voiced or voiceless. English-speaking adult listeners hearing Speaker 1's and 2's productions of these same sequences might be expected to be making decisions between /εnC/ and /εC/ (again, voiced or voiceless), although the extent of vowel nasalization also influences the choices of these more mature listeners, quite possibly in speaker-specific ways.

We are interested in how listeners go about making these decisions. We hypothesize that, in arriving at phonological representations that encompass the wide range of phonetic variants under investigation here, listeners formulate equivalence categories in which the two sites of a lowered velum, N and Ṽ, are perceptually equivalent. In this case, although English-speaking listeners are expected to hear vowel nasalization even in the presence of a nasal consonant, and to use this information in making decisions

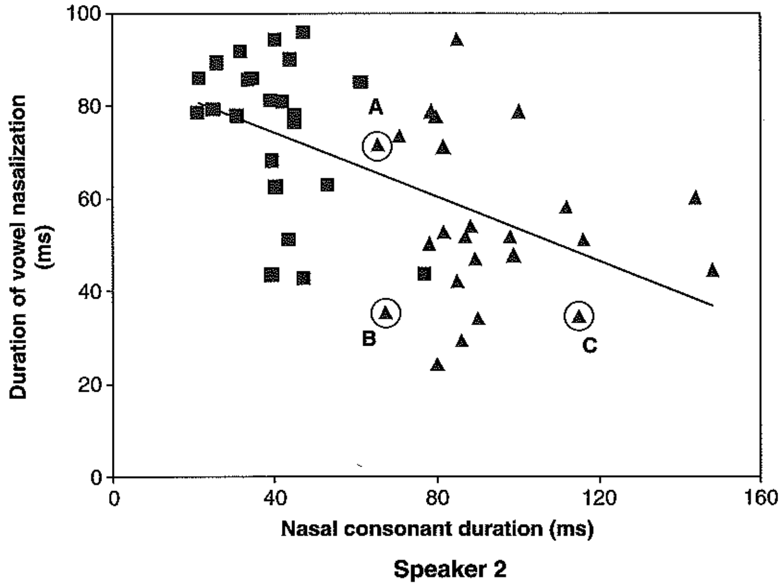

FIGURE 9.2. Scatterplot of productions of Speaker 2 from Fig. 9.1 highlighting *bend* data points with roughly comparable durations of vowel nasalization (B, C), nasal consonant (A, B), and total nasalization across the syllable rhyme (A, C).

about /VNC/, /VC/, and, at least for learners, /ṼC/, they should be relatively insensitive to whether the nasality is primarily on the consonant or on the flanking vowel, possibly formulating the category "nasal" rather than Ṽ or N. Consider, for example, the three data points from Speaker 2 that are circled in Figure 9.2. Under laboratory listening conditions, listeners might be expected to discriminate *bend* tokens B and C, whose nasal consonants differ by 45 ms, as different; the same is expected for *bend* tokens A and B, which differ in more than 35 ms of vowel nasalization. However, if nasalization on the vowel and consonant are heard as perceptually equivalent, then tokens A and C might be heard as similar; despite their relatively large acoustic differences, the total nasalization across the syllable is roughly comparable in these tokens. We hypothesize as well, based on the production data, that the range of variants of Ṽ and N that listeners treat as perceptually equivalent will differ depending on the voicing of the coda consonant. As argued below, such perceptual equivalence might be an important step in the historical loss of a conditioning environment in sound changes such as VN > Ṽ.

9.2.2.1 Methodological approach

We tested for perceptual equivalence between Ṽ and N using a variant of the well-established trading relations paradigm. In this paradigm, two acoustic properties that co-vary for a given phonetic distinction are independently manipulated to determine whether a change in the value for one property can be offset by an opposing change in the other, so that the phonetic percept—measured by identification and/or discrimination tests—remains the same (e.g. Pisoni and Luce 1987). That co-varying acoustic properties perceptually trade with each other is taken as evidence of the coherence among parts of the acoustic signal that "belong" together. In the current work, the duration of /n/ was co-varied with the temporal extent of nasalization on the preceding vowel /ɛ/. To test the hypothesis that post-nasal voicing might influence the expected perceptual trade-off, /ɛ(n)/ sequences were embedded in /b_d/ and /b_t/ contexts.

Waveform-editing techniques were applied to naturally produced utterances to create a nasal consonant continuum and an oral-to-nasalized vowel continuum. The original stimuli, produced by a female American English speaker, were tokens of *bet*, *bed*, and *mend*. The nasal consonant continuum, created from [n] of *mend*, ranged from 0 to 85 ms of [n] murmur and consisted of eight 12–13 ms increments (two glottal pulses each), excised from /n/ onset to offset. To create the vowel nasalization continuum, portions of oral [ɛ] from *bed* were replaced with same-sized portions of nasal [ɛ̃] from *mend* (beginning at vowel offset), yielding a three-step vowel series from oral [ɛ] to 33 percent nasalized (first two-thirds of the vowel from *bed* and final third from *mend*) to 66 percent nasalized (first third from *bed* and final two-thirds from *mend*).[3] Vowel and nasal murmur portions were orthogonally varied and spliced into [b_d] and [b_t] carriers, where the initial [b] (in both carriers) was from *bed*, and the final [d] and [t] were from *bed* and *bet*, respectively. Consistent with coda [t] and [d] in natural speech, whose closures are longer in CVC than in CVNC syllables, we incrementally shortened the oral closure as longer [n] durations were spliced in (such that each 12–13 ms increment of [n] replaced 6–7 ms of [t] or [d] closure).

Thus there were 48 stimuli (eight /n/ durations × three degrees of vowel nasalization × two voicing contexts), with endpoints *bed*, *bend*, *bet*, and *bent*. Stimuli were paired for perceptual testing using a variant of the same-different discrimination task first used by Fitch *et al.* (1980; see also Best *et al.* 1981). For the same pairings, each stimulus was paired with itself. The different pairings were of three types. In all three types, pair members differed in /n/ duration by a constant 37 ms (i.e. three steps along the /n/ continuum). In N-only pairs, /n/ duration was the only difference between pair members; vowel nasalization was held constant (similar to tokens B and C in Fig. 9.2). In the other two types of pairing, the /n/-duration difference was accompanied by a difference in vowel nasalization. For cooperating pairs, the stimulus with the shorter /n/ had *less* vowel nasalization than did the stimulus with the longer /n/ (i.e. pairs were of the type $\tilde{V}_S N_S$–$\tilde{V}_L N_L$, where $_S$ = slight V nasalization or

[3] Because vowels in pre-voiceless contexts are shorter than those in pre-voiced, the original [ɛ] from *bed* was shortened to an intermediate value so that vowel duration would be equally appropriate for [t]-final and [d]-final stimuli; this duration held for all vowels. We note also that, in creating the partially nasalized vowels, the vowel from *mend* rather than *bend* was chosen to ensure that nasalization extended throughout the excised portion. The nasal murmur was extracted from the same stimulus to preserve naturalness.

shorter N duration and $_L$ = longer V nasalization or longer N duration). For conflicting pairs, the stimulus with the shorter /n/ had *more* vowel nasalization than did the one with the longer /n/ (i.e. $\tilde{V}_LN_S$–$\tilde{V}_SN_L$; similar to tokens A and C in Fig. 9.2). Table 9.1 summarizes the details of the pairings whose results are reported here.[4] Figure 9.3 gives a spectrographic illustration of one cooperating pair (top panel) and one conflicting pair (bottom) from the series. (Comparable stimuli for the *bent* series are identical except for a (longer) voiceless coda closure and a [t] burst.)

For each of the two voicing contexts (final /t/ and /d/), listeners heard eight randomized repetitions of each different pair, with the order of pair members counterbalanced, and four repetitions of each same pair. Trial presentation was blocked according to final voicing; listeners' task was to determine whether pair members were the same or different.

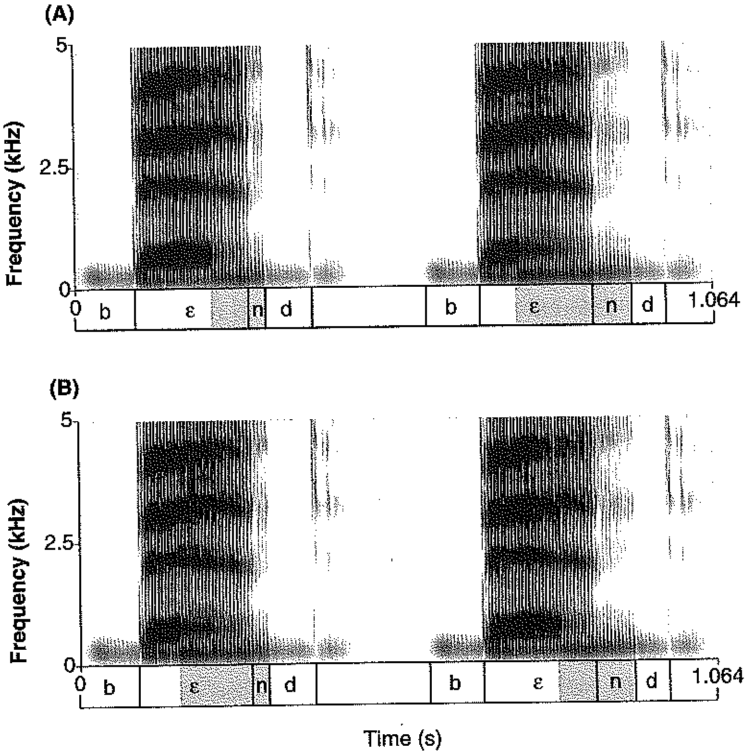

FIGURE 9.3. Illustrative cooperating (A) and conflicting (B) discrimination pairs from the *bend* series. The shaded portion of the transcription grid indicates the nasal ($\tilde{V}$ and N) portion of each pair member.

[4] In the results reported here, all discrimination pairings involved some vowel nasalization (33% or 66%) although the full experimental design included tokens with no vowel nasalization (0%); see Beddor (forthcoming).

TABLE 9.1. Non-identical stimulus pairs for the same-different discrimination task

	N-only $\tilde{V}_S N_S - \tilde{V}_S N_L$	Cooperating $\tilde{V}_S N_S - \tilde{V}_L N_L$	Conflicting $\tilde{V}_L N_S - \tilde{V}_S N_L$
Pair 1	$\tilde{V}_{33\%}N_{0ms}-\tilde{V}_{33\%}N_{37m}$	$\tilde{V}_{33\%}N_{0ms}-\tilde{V}_{66\%}N_{37ms}$	$\tilde{V}_{66\%}N_{0ms}-\tilde{V}_{33\%}N_{37ms}$
Pair 2	$\tilde{V}_{33\%}N_{12ms}-\tilde{V}_{33\%}N_{50m}$	$\tilde{V}_{33\%}N_{12ms}-\tilde{V}_{66\%}N_{50ms}$	$\tilde{V}_{66\%}N_{12ms}-\tilde{V}_{33\%}N_{50ms}$
Pair 3	$\tilde{V}_{33\%}N_{25ms}-\tilde{V}_{33\%}N_{63ms}$	$\tilde{V}_{33\%}N_{25ms}-\tilde{V}_{66\%}N_{63ms}$	$\tilde{V}_{66\%}N_{25ms}-\tilde{V}_{33\%}N_{63ms}$
Pair 4	$\tilde{V}_{33\%}N_{37ms}-\tilde{V}_{33\%}N_{75ms}$	$\tilde{V}_{33\%}N_{37ms}-\tilde{V}_{66\%}N_{75ms}$	$\tilde{V}_{66\%}N_{37ms}-\tilde{V}_{33\%}N_{75ms}$
Pair 5	$\tilde{V}_{33\%}N_{50ms}-\tilde{V}_{33\%}N_{88ms}$	$\tilde{V}_{33\%}N_{50ms}-\tilde{V}_{66\%}N_{88ms}$	$\tilde{V}_{66\%}N_{50ms}-\tilde{V}_{33\%}N_{88ms}$

9.2.2.2 Predictions

As can be determined from Table 9.1, N-only trials have the smallest acoustic difference between pair members of the three trial types, while cooperating and conflicting trials have equally large acoustic differences (33% difference in vowel nasalization and 37–38 ms difference in /n/ duration between pair members). If listeners treat nasalization on the vowel and /n/ as perceptually equivalent, such equivalence should lead to a relatively high proportion of incorrect "same" judgments of pairs whose members are roughly similar in terms of total nasalization across the ṼN sequence. That is, conflicting pairs (illustrated in the bottom panel of Fig. 9.3), despite large acoustic differences between pair members, should be difficult to discriminate—possibly more difficult than the acoustically less distinct N-only pairs. In contrast, cooperating pairs (as in the top panel of Fig. 9.3), whose members have large acoustic differences and large differences in total nasalization, should be correctly judged as different.

The expected influence of coda voicing is that the perceptual judgments of listeners will broadly reflect the distribution of ṼN measures found for the production of VNC_{voiced} and $VNC_{voiceless}$ words, such that vowel nasalization will have a greater influence on judgments in the voiceless (*bent*) than in the voiced (*bend*) context. Specifically, the extreme shortness of N, especially before a voiceless alveolar (shown by our own data, but see also, for example, Kahn 1980), suggests that listeners should be highly sensitive to vowel nasalization in judging *bent*-like stimuli. In this case, it may be that vowel nasalization will override the nasal consonant information, such that discrimination of voiceless conflicting pairs will be similar to that of voiceless cooperating pairs. (In both voiced and voiceless contexts, cooperating pair members should remain relatively easy to judge as different.)

9.2.2.3 Results

Twenty-four native English-speaking listeners participated. The pooled results for each discrimination trial type in which pair members differed, averaged across stimulus pairings (i.e. across the five entries in each column of Table 9.1), are given in Figure 9.4 for final-[d] and final-[t] contexts. Both contexts had a significant main

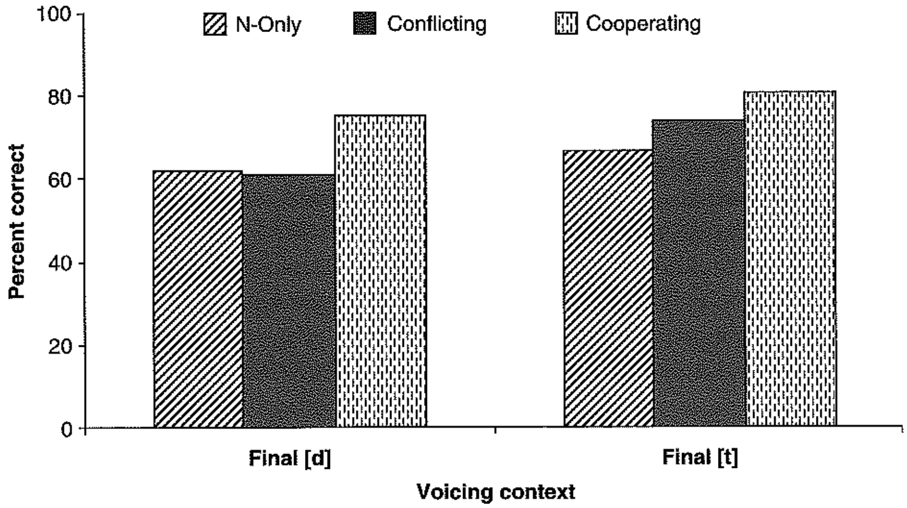

FIGURE 9.4. Pooled responses of 24 listeners to three types of discrimination trial: N-only (extent of vowel nasalization remained constant), conflicting ($\tilde{V}_L N_S - \tilde{V}_S N_L$), and cooperating ($\tilde{V}_S N_S - \tilde{V}_L N_L$).

effect of trial type, as shown by two multivariate repeated measures ANOVAs, one for each voicing context [$F(2, 22) = 12.83$ for [d] (*bend*) and 13.76 for [t] (*bent*) contexts, $p < .0001$].[5] As expected, for both contexts, discrimination was most accurate for cooperating pairs, whose members differed substantially in total nasalization across the $\tilde{V}N$ sequence ($\tilde{V}_S N_S$–$\tilde{V}_L N_L$). (It is not surprising that no trial type was highly discriminable across stimulus pairings, which included many within-category judgments for English listeners.) Listeners also showed the expected greater sensitivity to vowel nasalization in the [t] than in the [d] context. That is, as shown by pairwise comparisons, responses for the [d] context to the two trial types in which vowel nasalization varied (conflicting and cooperating) were significantly different ($p < .003$), but this same comparison was not significant for the [t] context ($p > .15$), indicating that listeners were more willing to trade nasalization on the vowel with nasalization on the consonant for the conflicting pairs in the [d] context.

An unexpected outcome was that, in the [d] context, discrimination was not less accurate for the conflicting trials than for the N-only trials. If listeners truly treat nasality on $\tilde{V}$ and N as perceptually equivalent in the conflicting trials, poor discrimination is expected since pair members should sound highly similar, leading to incorrect same responses, but this outcome does not strongly emerge in Figure 9.4. However, closer inspection of the data shows that the pooled results in Figure 9.4 are not representative of the responses of individual listeners, many of whom discriminated the conflicting trials at

[5] To control for listener bias, the same analysis was performed with the raw data transformed into d′ scores using the independent-observation model from Macmillan and Creelman (1991: 147). The d′ scores take into account listeners' performance on identical trials, while the percent correct analysis does not. However, due to listeners' high performance on identical trials, both analyses yielded precisely the same results.

chance level. As shown in Figure 9.5 for the [d] context, two distinct listener groups emerged from this inspection (excluding two respondents who performed at chance level on all trial types): listeners who consistently discriminated the conflicting trials more poorly than the acoustically less distinct N-only trials (left panel), and listeners whose overall accuracy on conflicting trials was similar (within 10%) to that on cooperating trials (right panel). Responses to each stimulus pair are given in order to show that the differences between groups hold across the stimulus set. Moreover, of the ten respondents to the final-[d] stimuli whose responses showed clear evidence of perceptual equivalence between Ṽ and N, only six of these showed the same pattern for the final-[t] stimuli; Figure 9.6 gives these results.

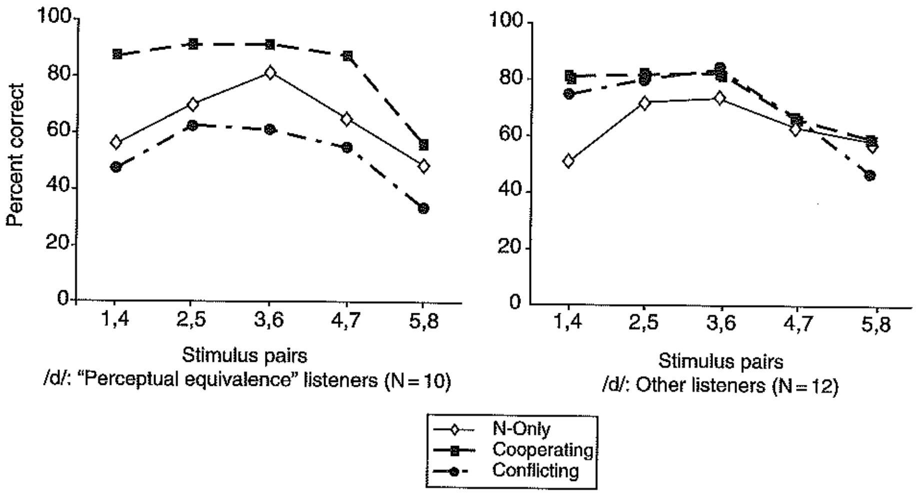

FIGURE 9.5. Responses of two listener groups (see text for explanation) to the /beεd/–/bε̃nd/ stimuli.

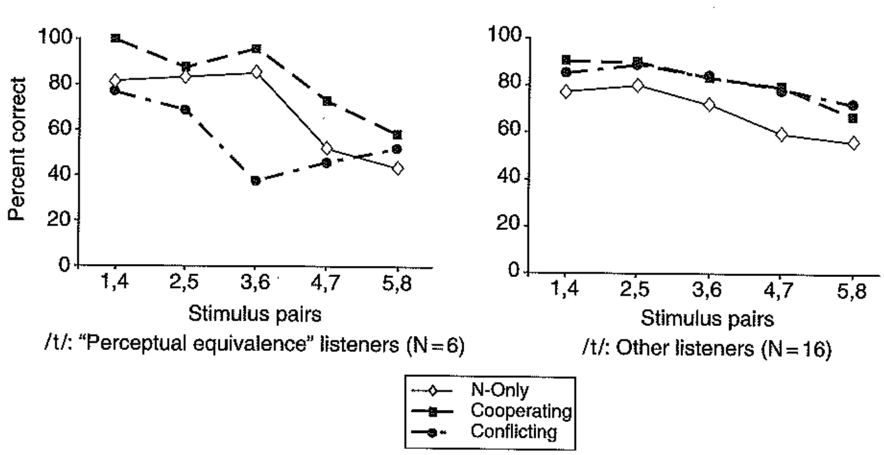

FIGURE 9.6. Responses of two listener groups (see text for explanation) to the /bεt/–/bε̃nt/ stimuli.

Thus the perceptual picture is that, when listeners discriminate stimuli that differ in the acoustic duration of vowel nasalization and the duration of the nasal consonant, they are more likely to treat these two sites of nasality as equivalent when ṼN is followed by a voiced than by a voiceless stop; in the voiceless context, vowel nasalization is the predominant perceptual cue for most listeners. However, hearing Ṽ as the predominant cue when Ṽ and N co-varied in $\tilde{V}NC_{voiceless}$ does not mean that listeners are insensitive to N duration for [bɛ̃nt]. For N-only stimuli, listeners were overall as accurate in their judgments in the voiceless as in the voiced context, an outcome that may seem surprising given the shortness of pre-voiceless /n/ in production. However, we attribute this outcome to the fact that nasal murmurs are more likely to be detected when followed by silence (the voiceless closure) than when followed by glottal pulsing (the voiced closure).[6]

At the same time, individual listeners clearly differed in their weightings of these properties. As seen in the results given in Figures 9.5 and 9.6, some listeners heard these properties as perceptually equivalent in both voicing contexts, others did so in the voiced but not the voiceless context; yet others weighed vowel nasalization heavily in both contexts. As would be expected, corresponding across-listener differences emerged in the identification tests conducted with these same listeners on the same (in this case, unpaired) stimuli. In the interest of space, full details of the identification results are not presented here. Instead, Figure 9.7 illustrates the range of identification patterns via the responses of three individual listeners, one from each of the three discrimination categories just described.[7] Listener 1 (top panels), who discriminated conflicting pairs poorly in both voicing contexts (i.e. one of the four listeners who heard nasalization on the vowel and consonant as equivalent; see left panels of Figs. 9.5 and 9.6), in identification systematically traded vowel nasalization and nasal consonant duration for both *bed–bend* and *bet–bent*. For this listener, as vowel nasalization (shown by line types) increased, so did *bent/bend* responses; similarly, as [n] duration (abscissa) increased, so did *bent/bend* responses. Listener 2 (middle panels)—one of the four listeners who poorly discriminated conflicting pairs in the voiced but not voiceless context—traded nasality on the vowel and consonant in identifying *bed–bend*, but in the voiceless condition identified any stimulus that had vowel nasalization as *bent*. Listener 3 (bottom) identified nearly all stimuli—voiced or voiceless—as having /n/ if the vowel was nasalized, and was highly sensitive to vowel nasalization in discrimination pairings (right panel of Figs. 9.5 and 9.6).

[6] Malécot (1960) also found evidence of the perceptual importance of vowel nasalization in voiceless contexts but, contrary to our findings, reported that the nasal murmur contributed little to listeners' identifications of words such as *camp* and *can't*. We attribute this difference in the perceptual weight given to pre-voiceless murmurs to stimulus differences in these studies: Malécot's tape-splicing technique included only extremely short "vestigial" murmurs whereas our study included a wider range of N durations (consistent with our production data).

[7] Figure 9.7 shows that ten /n/ murmur durations (stimuli 1–10) were used in the identification experiment, compared to only eight in the same–different discrimination experiment. Identification stimuli 2 and 4, which differed from flanking stimuli by only a single pitch pulse, were omitted from discrimination testing.

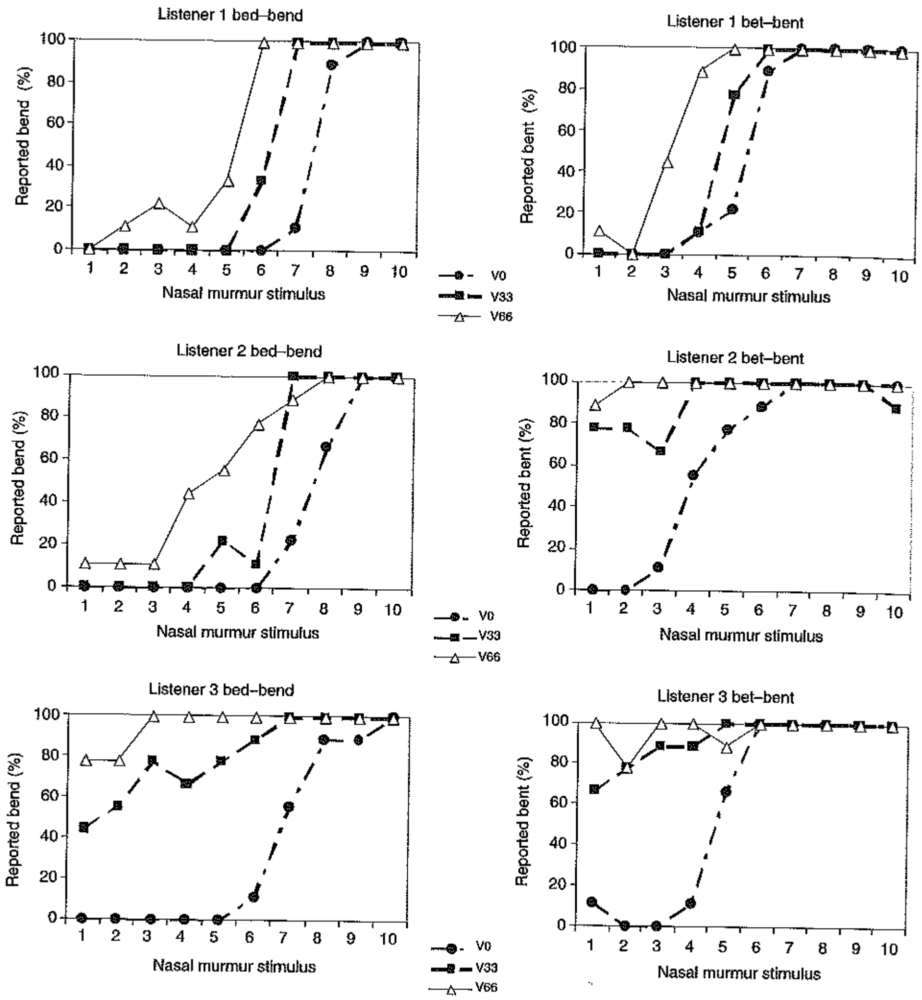

FIGURE 9.7. Identification responses of three listeners to the /bɛd/–/bɛ̃nd/ (left) and /bɛt/–/bɛ̃nt/ (right) stimuli for three degrees of vowel nasalization (oral, 33% nasalized, and 66% nasalized).

9.3 CO-VARIATION IN PRODUCTION AND PERCEPTION: IMPLICATIONS FOR THEORIES OF SOUND CHANGE

The production data (reported here and elsewhere) show that, in VNC sequences, the timing of velum lowering relative to oral closure is systematically variable: less overlap of the lowered velum with the oral closure results in concomitantly greater overlap with the preceding vowel (in other words, the shorter the N, the longer the nasalized portion of the V), and this timing pattern is more likely to occur in $VNC_{voiceless}$ than in VNC_{voiced} sequences.[8] Perception parallels these timing differences. Treiman *et al.* (1995) reported that young children's spellings of English and their responses to phoneme counting tasks indicate that 6–7 year olds sometimes interpret VNC sequences as /ṼC/, and that such interpretations are more common for $VNC_{voiceless}$ than VNC_{voiced} sequences. Our behavioral data show that, in differentiating CVC and CVNC words, some mature listeners are insensitive to the precise timing of velum lowering when differentiating such words, but are highly sensitive to differences in total nasalization across the syllable rhyme (left panels of Figs. 9.5 and 9.6); other listeners are particularly sensitive to differences in vowel nasalization (right panels of Figs. 9.5 and 9.6), especially in pre-voiceless contexts.

We speculate that the production–perception scenario that has emerged here is a factor in the historical change VN > Ṽ. Generally speaking, when coarticulation is phonologized in the transmission from a speaker to a listener or possibly, over time, within a listener, a predictable property of the signal becomes a distinctive one. We propose that, for coarticulatory nasalization, phonologization is facilitated by co-variation in production and equivalence in perception between Ṽ and N. Although the speaker may intend /CVNC/, under equivalence listeners attend less to the segmental source of nasalization and abstract away the feature "nasal" rather than /CVNC/—or /CṼC/. Of course, as we have seen, some listeners do attend to Ṽ as the primary source of nasalization. Clearly we cannot determine whether, at an earlier point in time within the relevant linguistic community, perceptual equivalence of Ṽ and N was an even more prevalent pattern among listeners; phonologized [bɛ̃t] and, to a lesser extent, [bɛ̃d] are arguably already well established for some dialects of American English. Of importance here is that listeners arrive at categorizations of [CṼ(N)C] that do not appear to clearly include Ṽ or N, and a given listener may arrive at context-dependent categorizations of [CṼ(N)C] (e.g. abstracting Ṽ in voiceless but not in voiced contexts).

We have argued that listeners are led to formulate equivalence categories in the face of articulatory co-variation; listeners are sensitive to phonetic detail and arrive at representations that encompass the relevant variation. Of course, sensitivity to phonetic detail depends on the psychoacoustic salience of the details. Perceptual studies that we are currently conducting show that, as duration of vowel nasalization increases in ṼN stimuli, listeners are increasingly unable to detect differences in nasal consonant duration. Specifically, when members of stimulus pairs differed only in N duration (N-only), increasing vowel nasalization in pair members from 25 percent to 75 percent of total vowel duration resulted in a 26 percent decrease (to chance performance) in listeners' discrimination accuracy. Psychoacoustic factors account

[8] The timing differences as related to voicing can be understood in terms of a combination of vocal tract aerodynamics and possibly auditory factors. See, for example, J. Ohala and M. Ohala (1993) and Hayes and Stivers (2000) for relevant discussion.

for, or at least contribute to, this outcome, but of interest here is that, even when total nasalization in the syllable rhyme differs, listeners are insensitive to N duration if the preceding vowel is heavily nasalized.

That the production and perception patterns that emerged in this study shed light on the historical change VN > Ṽ is supported by the ways in which the phonetic patterns mirror the historical situation in many languages of the world. Our approach predicts that contexts that give rise to the concomitant processes of nasal coda shortening and heavier vowel nasalization should also be those contexts in which phonological nasal vowels are particularly likely to develop historically. At the risk of oversimplifying complex sound changes, broadly speaking, our prediction holds up. Most relevant to the data presented here is that VN > Ṽ is more likely to develop when N is followed by a voiceless, as opposed to a voiced, obstruent. For example, Hajek's (1997: 141–2) study of nasalization patterns in Northern Italian dialects showed that six dialects had extensive vowel nasalization and systematic N deletion in $\text{VNC}_{\text{voiceless stop}}$ contexts, but only two of these dialects had the same pattern in $\text{VNC}_{\text{voiced stop}}$ contexts. More generally, Hajek noted that his extensive cross-language study of nasalization showed "no counter-evidence to the claim that the development of distinctive nasalization is predictably affected by the voicing contrast in post-nasal consonants" (1997: 53; see also M. Ohala and J. Ohala 1991; Tuttle 1991; Sampson 1999: 256). Additionally, other phonetic contexts that show a clear inverse relation between the extent of vowel nasalization and nasal consonant duration are also paralleled by the historical data. As discussed in Section 9.2.1, fricative contexts trigger short N durations and temporally extensive vowel nasalization, as shown by phonetic data from Italian and Japanese, and historically VN > Ṽ is especially likely to occur in pre-fricative contexts (Foley 1975; Rubach 1977; Tuttle 1991; J. Ohala and Busà 1995; Hajek 1997: 144, Sampson 1999: 182, 253). Similarly, paralleling the phonetic finding from Thai that long vowels are more nasalized than short ones and are followed by short nasal consonants is the historical tendency for long vowels to become distinctively nasalized (Whalen and Beddor 1989; Hombert 1986; Sampson 1999: 340).

To summarize and conclude, our approach to phonetically motivated sound changes takes as its starting point that (a) the coarticulatory variation introduced by phonetic context is lawful and (b) listeners have knowledge of these coarticulatory timing patterns, but their often imperfect adjustments for the consequent acoustic variation mean that listeners remain sensitive to phonetic details. We focused our account on sound changes in which a coarticulatory source is lost over time, but its effects remain on flanking sounds. In such cases, the target gesture is retained, but its coordination relative to other gestures has changed. We hypothesized that we would find evidence of this shifting coordination in production, in terms of increased coarticulatory effects as the duration of the source shortens, and in perception, in terms of overall sensitivity to the acoustic effects of the gesture rather than its precise site. Both hypotheses were upheld, although the responses of some listeners indicated that the predominant cue was the coarticulatory effect rather than the original source.

The convergence of these phonetic patterns with the phonological data leads us to propose that such co-variation in production and perception between coarticulatory source and effect serves as a phonetic path to certain categories of sound changes. We expect that, at a later stage in the evolution of such changes, listener insensitivity to changes in source duration leads to a point at which the original source—which is only variably present in the input—is no longer informative.