LANGUAGE AND SPEECH, 1993, 36(2, 3), 155-170

COARTICULATION AND PHONOLOGY*

JOHN J. OHALA
University of Alberta
and
University of California, Berkeley

Many sound patterns in languages are cases of fossilized coarticulation, that is, synchronic or phonetic contextual variation became diachronic or phonological variation via sound change. An examination of languages' phonologies can therefore yield insights into the mechanisms of coarticulation. In this paper I discuss (a) the need to differentiate between phonological processes that are and are not due to coarticulation, (b) the need to differentiate between 'on-line' synchronic variation and comparable fossilized diachronic variation, (c) how to determine some of the constraints on coarticulation – especially the higher priority of maintaining acoustic-auditory, rather than articulatory, norms for the shape of speech elements, and (d) how coarticulation presents a "parsing" problem to the listener and, of course, to systems for automatic speech recognition.

Key words: coarticulation, phonology, sound change, nasalization, speech normalization

INTRODUCTION

The attempt to understand coarticulation in speech has taken on renewed interest currently due to its theoretical import (Fowler, 1980, 1986; Strange, Verbrugge, Shankweiler, and Edman, 1976) and its relevance to practical applications in speech technology (Klatt, 1986). Phonology is a source of information on coarticulation in speech for two reasons: (a) the principal focus of phonology is sound change (and its repercussions on present-day sound patterns) and (b) much of sound change is fossilized coarticulation. Over the past two centuries phonologists have accumulated a valuable store of data on sound change that could usefully be tapped by those interested in coarticulation – if they approach the data judiciously. As with any discipline in a stage of development that is pre-experimental – or perhaps one can say nascently experimental (Prideaux, Derwing, and Baker, 1980; Ohala and Jaeger, 1986; Ohala, 1986; Kingston and Beckman, 1990; Docherty and Ladd, 1992; Diehl, 1991) – the data are often overlaid with, interpreted, and labelled in terms of untested concepts. Until we have a unified and empirically supported theory of speech communication, it will be impossible to avoid this. But this should not be an insuperable barrier to the use of phonological data. In this paper I propose to demonstrate how phonological data – especially on sound change – can be used to glean some hopefully valid generalizations about the nature of coarticulation.

* Send correspondence to John J. Ohala, Department of Linguistics, University of California, Berkeley, CA 94720.

TERMINOLOGY

Menzerath and Lacerda popularized the term 'coarticulation' (or *Koartikulation*) in their 1933 monograph. It was coined to denote instances where two successive sounds were articulated together.[1] The term is not as common in the phonological literature as the near-equivalent terms 'assimilation' or, more currently, 'feature spreading', both of which, however, may signify a somewhat wider range of phenomena. All such terms attempt to describe or explain similar phenomena, i.e., the variation in the phonetic manifestation of a given sound due to its taking on some of the features of nearby sounds. Here I will use 'coarticulation' and 'assimilation' as synonyms.

Unfortunately there are instances where coarticulation is, in my opinion, incorrectly invoked to explain a particular case of speech sound variation. Before trying to read the literature on sound change for its relevance to coarticulation, it is necessary to consider what is, and what is not, true coarticulation.

WHAT IS, WHAT ISN'T, COARTICULATION

What seems to be coarticulation but isn't

"Velar softening". 'Velar softening' is the term applied to cases where the dorsal stops [k] or [g] change to the apical obstruents [t͡ʃ t͡s ʃ s] or [d͡ʒ d͡z ʒ z], respectively, typically next to palatal (or labial palatal) vowels or glides such as [i, ɪ, e, y, j]. It is also referred to as 'palatalization'. For examples, see (1).

(1) (a) Latin > Tuscan Italian (Grandgent, 1927, p. 103)

placeo "I please" > *piaccio* (= [plakĕo] > [pjatːʃo])
urceus "pitcher" > *orcio* (= [urkĕus] > [ɔrt͡ʃo])

Latin *radicina* /radikina/ > radit͡sin > French *racine* /ʁasin/ "root".

(b) Old English > Middle and Modern English

OE *ecg* > ME *egge* > Mod.E *edge* [ɛd͡ʒ] (cf. cognate 'to egg on' ("to incite"))

OE *cinn* [kin] > Mod.E *chin* [t͡ʃɪn]

(c) Tai (Li, 1977, p. 221)

Lungchow	Po-ai	
kjaa	t͡ʃaa	"rice seedlings"
kjoo	t͡ʃoo	"drum"

[1] 'Coarticulation' is also used to describe cases where there is more than one simultaneous articulation for a single sound. Both articulations may be equal as in labial velar stops [k͡p g͡b] or one may be secondary as in labialized consonants [t^w g^w].

To begin with, the emergence of frication in these sound changes can be accounted for by aerodynamic factors (Stevens, 1971; Ohala, 1983b): The narrow channel through which the airflow passes enhances turbulence. The important problem is to account for the change in the place of articulation, i.e., from dorsal to apical.

Typical of the phonetic accounts given for this process is that by Grammont (1933, p. 214; see also Bloomfield, 1911; Hock, 1986, pp. 73–76):

> ... les voyelles antérieures *i, é,* œ fermé, *ü, a* antérieur, attirent à elles un *k* qui les précède, dont le point d'articulation passe alors en avant du sommet de la voûte palatine. Dans cette position l'explosion est facilement suivie d'un élément fricatif du genre *y* [= IPA j]; si cet élément ne se développe le résultat est un *k* mouillé; s'il se developpe le résultat est un *ky*, et, pour peu que l'articulation avance encore, l'explosion n'a plus lieu dans les domaine du *k*, mais dans celui du *t*, d'où *ty*, avec un *t* articulé la pointe appuyée contre les alvéoles des incisives inférieures. Si l'articulation est molle ... *ty* subsiste: Vionnaz *tyevra* de *capra;* si l'articulation du *t* est violente ... le *y* s'assourdit. La plupart des langues n'ayant pas de *y* sourd, le remplacent par ce qu'elles ont de plus voisin, *s, s'*, ou *š* ...

That the point of articulation of the stops in sequences like /ki/ /gi/ is quite fronted is well established.[2] The main problem with a purely articulatory account of the sound changes in (1) is that it is not at all obvious how a shift from dorsals to apicals could occur (Ohala, 1986, 1992a). Accordingly, this is the weakest and most speculative step in the scenario given by Grammont: 'pour peu que l'articulation avance encore ...'. What is the motivation for the place of articulation to inch forward? There are other specific problems with such explanations:

First, the place of articulation of the apical obstruent, the end product of velar softening, is not the same as that of the palatal vowels; a [t] is further forward than an [i] or [e]. When one sound coarticulates in place with another, it is unusual, if not unprecedented, for it to overshoot.

Second, even though it is the tongue that is involved in both the 'before' and the 'after' states of velar softening, they are functionally different parts of the tongue: /k, g/ and /i, j/ are articulated with the tongue dorsum; /t͡s, t͡ʃ/, etc., are articulated at least in part with the tongue blade and apex. For both of these reasons, it is difficult to see how velar softening can be attributed to coarticulation; /t͡ʃi/ is not a combination or a "compromise" articulation between /k/ and /i/.

Third, another difficulty with a purely articulatory account of such changes, pointed out by Blumstein (1986), is that it leaves one unable to explain why it is asymmetrical in its direction, i.e., dorsal stops become apicals near front vowels, but apicals never seem to become dorsals near back vowels.

---

[2] However, dorsal consonants seem always to be coarticulated with adjacent segments. No one seems to have found any case of a non-fronted /k/ in a syllable like /ki/ or a non-backed /k/ in /ka/. So if coarticulation is conceived of as a *combination* of articulatory features from two segments, it is not clear what the "canonical" place of articulation of the /k/ is.

A fourth weakness in Grammont's (and others'[3]) account is the argument:

> La plupart des langues n'ayant pas de *y* sourd, le remplacent par ce qu'elles ont de plus voisin, *s*, *s'*, ou *š*.

Arguments of the sort: 'possible sound *x* changed to *y* in language *A* because that language did not have (did not permit) an *x*', though commonly encountered in the phonological literature, are circular. Often the only evidence one has that language *A* does not permit an *x* is the absence of *x*, which is the fact one is trying to explain. Such arguments therefore reduce to the tautology: lack of *x* because of lack of *x*. Moreover, sound changes do often introduce sounds to languages, i.e., sounds they had not had before, e.g., nasal vowels in French; front rounded vowels in German due to umlaut.

The above difficulties are largely overcome by recognizing that the change of /ki/ to /tʃi/ is motivated by their perceptual similarity. Winitz, Scheib, and Reeds (1972) studied listeners' confusion of CV syllables (C = p t k; V = i a u) under various conditions. Under the conditions where the maximal information was transmitted, the largest number of confusions occurred when /ki/ was misidentified as /ti/. This is understandable given the acoustic similarity of their transitions (Fant, 1973, p. 112). Furthermore, as Ohala (1985) argues, the asymmetry in the direction of change may also be explained by considering these sounds' acoustic shapes: The spectra of their bursts are rather similar except for a sharp peak near 3 kHz in the case of the velar stop (due to the fact that the /ki/ has more of a downstream resonating cavity). It is more probable that listeners would miss this detail in the burst spectra than that they would spuriously imagine it in the spectrum of the /ti/. Thus the most probable direction of confusion is from /ki/ to /ti/ rather than the reverse.

It is safe to conclude that changes of the type /ki/ > /tʃi/ owe more to acoustic similarity than to coarticulation. It is thus similar in origin to sound changes which are more obviously due to acoustic factors since they involve discontinuous shifts in place of articulation, e.g., p^j > t, k^w > p (Ohala and Lorentz, 1977; Ohala, 1978).

Assimilation of place of articulation in clusters. Another example of a phonological process commonly attributed to coarticulatory factors is the case of heterorganic consonant clusters in which there is assimilation of place of articulation of the first consonant to the second, as in (2).

(2)	Late Latin	*scriptu*	>	Italian	*scritto*	"written"
		nocte	>		*notte*	"night"
	Sanskrit	*bʰaktum*	>	Pali	*bʰattum*	"devotee"
		praptum	>		*pattum*	"obtain"

The vast majority of such assimilations of place in medial C_1C_2 clusters have the place of C_1 assimilating to that of C_2 rather than the reverse.[4]

---

[3] Bloomfield (1911) attributes the shift from palatal to dental/alveolar to the following cause:

> In some languages, which tend to articulate either with the back or with the tip of the tongue and not with intermediate points, the palatal character of *t'š'*, *d'ž'* may be nearly or wholly given up, the result resembling *tš*, *dž* ...

[4] Place assimilations in the reverse direction do occur, but some cases of this may have a different explanation; see Murray (1982) and Ohala (1990).

Passy (1890, pp. 183–184) and Hock (1986, p. 65) suggest that such assimilations are especially common when C_1 is a nasal. Thus, in Latin, NC clusters regularly show the N assimilating in place to the C, e.g., *k̂mtóm > *centum* "hundred", *quinctus* > *quintus* "fifth" (where the original 'n' = [ŋ] > [n] / _t), whereas many *stop + C* clusters remained unchanged, e.g., *octō* "eight", *dictus* "statement", *lāpsus* "falling" (Kent, 1932, pp. 112–138).

Passy, like most writers on the subject, attributed this and all assimilations to 'une tendance à l'économie: on supprime les distinctions superflues ...' (p. 168). Paul (1880, p. 53) likewise declared such complete assimilations as Italian *otto* and *cattivo* to be more convenient (*bequemer*) to pronounce than the original *octo* and *captivus*. (See also Whitney, 1867, p. 69.)

But there are aspects of this process that undercut an explanation for it grounded solely in articulation and notions like ease of articulation and economy (see further arguments in Oertel, 1901, p. 204). The energy expended in producing speech presumably accumulates from the beginning to the end of an utterance. Thus there would be greater motivation to change to less effortful articulations later rather than earlier in an utterance. Yet it is usually C_1, the earlier segment, that assimilates to the later C_2, not *vice-versa*. An why should this change in place affect NC clusters more than those consisting of *stop + C*? In those cases where NC clusters show assimilation, why is it so often the case that just place assimilates and not manner? If speakers were trying to make their speech easier, why do they stop short of complete assimilation?

Again, there is evidence that such place assimilations owe more to acoustic-auditory factors than to articulatory (but see Nolan, 1992). Several studies have shown that C_1–C_2 clusters whose component syllables have been cross-spliced with closure maintained at a duration appropriate to a single C are heard by listeners as either singleton C_2 or in the case of heterorganic NC clusters as homorganic, with the N having the same place of articulation as the following C (Malécot, 1956; Dorman, Raphael, Liberman, and Repp, 1975; Repp, 1978; Fujimura, Macchi, and Streeter, 1978; Ohala, 1990). Apparently, as argued by Malécot, place cues are generally less salient in VC than in CV transitions and are even less salient in VN transitions (see Ohala, 1990; Ohala and Ohala, in press). The CV place cues therefore dominate and overshadow those from the VC transition.[5] Moreover, the classic study of Miller and Nicely (1955) suggests

[5] In addition to the contribution to this effect ('CV dominance') of the relative acoustic salience, the task remains of establishing the contribution of (a) purely auditory factors, e.g., the dominance of the most recent place cues, i.e., those from the CV transition, and (b) linguistic experience. That linguistic experience plays some role is supported, first, by the Fujimura *et al.* study which showed that English and Japanese listeners' reactions to the cross-spliced VC_1–C_2V stimuli are slightly different, presumably because of the languages' difference phonotactics, and, second, by the result reported by Ohala (1990) that the medial consonant duration that induces a singleton C_2 percept is shorter for component voiced than voiceless consonants, presumably mirroring the typical difference of durations of intervocalic consonants as a function of voicing. Such linguistic experience, of course, may develop out of adaptation to acoustic and auditory effects; that is, listeners' biases in how they interpret the speech signal is based on what they have learned are optimal and efficient cues.

that manner cues are more robust than place cues (though, strictly speaking, they established this just for CV, not VC, sequences). Studies such as Malécot's show that heterorganic clusters can change to singletons or homorganic ones through the agency of the listener. These confusions parallel the asymmetries of place assimilation in the diachronic record.

What may seem not to be coarticulation, but is

Epenthetic stops. Intrusive or epenthetic stops appear in various environments as exemplified in (3) and (4).

(3)	English	*warm + th*	[wɔɹmpθ]		
		young + ster	['jʌŋkstɚ]		
	Latin	*hum(i)lis*	"low"	>	English *humble*
	Latin	*sumsī* > *sumpsī*	"to put on" (past part.)		
	Latin	*cam(ĕ)ra*	"vault"	>	French *chambre*
	Arabic	*al hamra*	"the red"	>	Spanish *Alhambra*
	Ulu Muar Malay	[banu]~[band^{u}]	"doorsill"	(Hendon, 1966)	
(4)	English	*else* [ɛlts]			
		pulse [p^{h}ʌlts]			
	Late Latin	*slavo* "slave" > **stlavo* > *sklavo* > Italian *schiavo*			

These stops seem to come out of nowhere and thus are sometimes treated by phonologists as epenthetic (Piggott and Singh, 1985). In essence, however, they are due to coarticulation – and, in fact, this has long been recognized by phonologists familiar with the underlying phonetic details (Bindseil, 1838; Weymouth, 1856; Grandgent, 1896; Passy, 1890, pp. 215–216; Oertel, 1901, pp. 205–206; Phelps, 1937; Wetzels, 1985; see also Ohala, 1974; Ohala, 1992b; Ohala and Ohala, in press). In the case of [mθ] > [mpθ], etc., the [p] is created out of the latter portion of the nasal due to anticipatory assimilation of the velic position and the glottal configuration of the following voiceless oral segment. In the case of [ls] > [lts] or [sl] > [stl], the stop arises from the fact that [s] and [l] have near-complementary areas of tongue-palate contact, respectively: along the sides of the alveolar ridge but not the midline *vs.* in the midline but not along one or both sides of the alveolar ridge. In the transition (or coarticulation) between these sounds it might happen that contact is made all along the alveolar ridge thus creating a stop. Although such stops may originate as purely predictable products of the transition between sounds, listeners may reinterpret them as purposeful, i.e., phonologize them, thus precipitating a sound change (Ohala, 1992a).

Tonogenesis. The development in Chinese and other tone languages of lexically distinct tones out of prior contrasts in voicing has long been noticed (Edkins, 1864; Maspero, 1912; Haudricourt, 1961; Svantesson, 1983). One example is given in (5).

(5)	Southern Kammu	Northern Kammu (from Svantesson, 1983)	
	klaaŋ	kláaŋ	"eagle"
	glaaŋ	klàaŋ	"stone"

Here, insofar as preservation of the voicing contrast and lack of tone is concerned, Southern Kammu is presumed to be similar to the parent form of Northern Kammu. Thus Northern Kammu's high tone is derived from an earlier preceding voiceless stop and the low tone from an earlier preceding voiced stop.

This pattern of diachronic variation has a well-documented parallel in synchronic variation. Many languages, even those that are not tonal, show the same differential fundamental frequency (F_0) following voiced and voiceless consonants (Hombert, Ohala, and Ewan, 1979). The diachronic pattern is a fossilization or phonologization of the synchronic pattern (see below).

Superficially, voicing and tone seem like such different entities that a link between the two would not be classified as coarticulation. Recently, however, Löfqvist, Baer, McGarr, and Story (1989) have found an elevated level of cricothyroid muscle activity during voiceless obstruents and that this higher level of activity may persist into following vowels. The cricothyroid has been shown to be the primary muscle that controls F_0 of voice. It appears then that the different F_0 levels result from coarticulation (or perseveratory assimilation) of the state of the cricothyroid for stops with the configuration for following vowels.

Summary

In Ohala (1989, 1992a) I offer a general precept for discovering the phonetic basis of sound changes that are widely attested in diverse languages: Diachronic variation should have a parallel in synchronic variation. This precept would help us in the cases considered above: Whereas stop epenthesis and the consonant-induced F_0 perturbations have well attested parallels in studies of speech *articulation* (and thus could plausibly be attributed to coarticulation), changes of the sort /ki/ > /t͡ʃi/ do not. The latter, however, do have parallels in speech *perception* studies.

After sorting out the genuine from the spurious cases of sound patterns attributed to coarticulation there still remains the question of how the shift from a purely phonetic to a phonological event took place, i.e., how variation that was predictable from context could become purposeful and non-predictable (and thus a carrier of information). I maintain that traditional phonological accounts fail to explain many aspects of the process, i.e., the role of the speaker and listener, and the "cognitive" element.

How Does Variation Due to Coarticulation Become Phonologized?

There is infinite variation in speech

After about a century of instrumental phonetic study we know that the amount of contextually-determined variation in speech is infinite (Brücke, 1871; Meyer, 1903; Lindblom, 1963). The handful of allophones listed in the typical structuralist account of a language's phonology scarcely gives the whole picture. This being the case it would seem that the question "how does synchronic variation become phonologized?" is less important than "how is it that only a very small fraction of synchronic variation becomes phonologized?" If all the variation exhibited in speech were taken at face value by listeners, we would have sound change progressing as rapidly over a few speaker-listener exchanges as it has over several millennia. As noted above, virtually all languages having voiced *vs.* voiceless consonants show F_0 perturbed on following vowels, yet only a few languages have phonologized this in the form of distinct lexical tones. Somehow, pronunciation norms can be extracted from the extreme variation present in speech.

Listeners normalize the variable speech signal

In fact, there is abundant evidence that listeners do *normalize* the speech signal (Ladefoged and Broadbent, 1957; Pickett and Decker, 1963; Lindblom and Studdert-Kennedy, 1967; Mann and Repp, 1980; Beddor, Krakow, and Goldstein, 1986).[6] Exactly how normalization is done is less clear (but see Elman and McClelland, 1986; Nearey and Hogan, 1986). But for the present purposes all we need to know is that perceptual normalization of speech exists and must be globally similar to perceptual normalizations in other domains, e.g., vision (Rock, 1983).

Thus, if a speaker produces a nasalized vowel before a nasal consonant, the listener can factor out the contextually-predictable nasalization (Kawasaki, 1986; Beddor *et al.*, 1986). We can represent this in phonological rule shorthand[7] as in (6).

(6) (a) Speaker: /e/ > [ẽ] / _n
(b) Listener: [ẽ] > /e/ / _n

A process such as that symbolized in (6a) need not be a rule of grammar, i.e., part of the speaker's knowledge; it might represent alterations in the intended signal that occur in what communication engineers refer to as the 'transmission line' outside the message source, including the encoding machinery. The process in (6b), however, must be part of the language-user's knowledge – grammar, if one prefers – and acquired through experience. I have referred to processes such as (6b) as 'normalization' or 'correction' rules (Ohala, 1989, 1992a).

Phonologization = failure to normalize

Phonologization of phonetic contextual variation would occur quite simply through the *failure* of the listener to apply such corrective rules. This could happen for a variety

6 The claim that listeners can factor out the variation in order to arrive at the 'norm' does not imply that the variation itself is thrown away. The variable 'residue' may still be available to identify the speaker and his/her state of health, the style of speaking, the regional dialect, etc.

7 Such shorthand notations should not be reified; they have no formal theoretical status.

of reasons: inexperience or perhaps failure to detect the conditioning environment. The latter cause must be quite common because in fact many sound changes of the sort that have synchronic parallels do involve apparent simultaneous loss of the conditioning environment along with the conditioned change (and this includes the development of distinctive vowel nasalization and distinctive tone, two of the examples considered above). Some speech perception studies have effectively duplicated this type of sound change by demonstrating that listeners interpret the speech signal in a different (one might say more 'veridical') way when certain portions have been deleted or masked out (Ohala, 1983a; Kawasaki, 1986; Beddor *et al.*, 1986). Thus, if (6) characterizes the situation before the sound change occurred, then the sound change itself can be characterized as in (7).

(7) (a) Speaker: /e/ > [ẽ] / _n
(b) Listener: [ẽ] > /ẽ/

Such a two-stage account (i.e., (6) and (7)) is a considerable improvement over traditional accounts, e.g., as in (8).

(8) /en/ > /ẽ/

(8) suggests that the sound change resulted in the vowel being nasalized where it had not been nasalized before. But this is highly unlikely. As represented in (6) the vowel was nasalized all along but this nasalization was perceptually 'removed' by listeners. (8) also sidesteps the question of the relationship between the conditioned change and the conditioning environment.[8]

Summary

Contextually conditioned variation in speech is omnipresent but is normally "hidden" because listeners are able to factor it out if they are aware of the context. Such variation becomes fossilized or phonologized as sound change if and when listeners fail to recognize the variation as totally predictable from context, incorporate it into their own mental lexicons, and base their own pronunciation on the new norm. This often results not simply in a purposeful copying of variant phonetic effects but in their exaggeration – in much the same way, so the story goes, that the way the Roman stone carvers finished a straight incision became exaggerated into the *serifs* on printed letters.

WHAT CAN COARTICULATION IN PHONOLOGY TELL US ABOUT SPEECH?

By examining sound changes which arose out of coarticulation we can get an idea of what can and what cannot coarticulate and in this way perhaps get an idea of the goal the speaker is trying to achieve.

---

[8] Actually, many accounts posit a two-stage change:
en > ẽn
ẽn > ẽ.
But the evidence for these two stages is often quite forced.

Constraints on coarticulation of nasalization

One of the most striking forms of assimilation is that of spreading nasalization or nasal prosody in languages such as Sundanese and Tereno. Examples are given in (9).

(9) (a) Sundanese (Robins, 1957)

[ɲãĩãn]	"to wet"		
[bɤŋh̃ãr]	"to be rich"		
[ɲãh̃õkɤn]	"to inform"		
[mĩʔ̃ãsih]	"to love"		

(b) Tereno (Bendor-Samuel, 1960, 1966)

[piho]	"he went"	[ᵐbiho]	"I went"
[iso]	"he hoed"	[ĩⁿzo]	"I hoed"
[owoku]	"his house"	[õw̃õᵑgu]	"my house"
[emoʔu]	"his word"	[ẽmõʔ̃ũ]	"my word"

As shown in (9), a nasal segment in the word induces perseveratory nasalization which spreads throughout the word unless blocked by a buccal obstruent such as [k] or [s].[9] In Tereno, nazalization, which marks 1st person, is "injected" at the start of the word and spreads towards the end. A buccal obstruent that blocks the spread of the nasalization becomes voiced and prenasalized. Different segment types vary in their propensity to block such spreading nasalization: High vowels are more likely to block it than lower vowels, distinctively oral vowels naturally block nasalization, non-nasal sonorants such as glides, laterals, and "r-sounds" can yield to nasalization but in some languages block it. Buccal obstruents almost always block nasalization; non-buccal obstruents (glottals and pharyngeals) do not. (See Schourup, 1973; Ohala, 1983b; Ohala and Ohala, in press.)

The segments which block spreading nasalization are also those which, when adjacent to a nasal consonant, can give rise to the epenthetic stops discussed earlier (see (3)).

Both aerodynamic and acoustic constraints are at play here. An open velopharyngeal valve (lowered soft palate) would bleed the pressure build-up which, when released, would produce the audible turbulence or burst that is one of the main cues differentiating obstruents from sonorants. Sonorants, of course, do not have an aerodynamic constraint against nasalization but they may block it if the lowered soft palate would distort their distinctive spectrum (Ohala, 1975; Ohala and Ohala, in press).

Epenthetic nasals

A revealing case of spreading nasalization, having some points of similarity with Tereno, above (9b), is what Ohala and Ohala (1991, in press) call 'epenthetic nasals' in the history of Hindi. Examples of crucial data are given in (10).

---

9 Although these patterns of assimilation of nasalization undoubtedly had a phonetic origin, they are now fossilized and owe more to morphological than phonetic factors (Ohala and Ohala, in press).

(10) Old Hindi Modern Hindi

dã:ta	>	dã:t	"tooth"
tʃã:da	>	tʃã:nd	"moon"

When an original ṼC sequence involved a voiceless stop, no change to these elements occurred, but when the C was a voiced stop, a nasal consonant arose in between them. Ohala and Ohala found widespread evidence in other languages that voiced but not voiceless stops tolerate a nasal onset. This happened even in cases where the source of the spreading nasalization and the voiced stop abutted across word boundaries, e.g., a French speaker pronouncing the phrase "saint bel" /sɛ̃ bɛl/ manifested this phonetically as [sɛ̃ᵐbɛl]. The rationale for this pattern, they argued, was as follows:

> ... among the auditory cues for a voiced stop there must be a spectral and amplitude discontinuity with respect to neighboring sonorants (if any), low amplitude voicing during its closure, and termination in a burst; these requirements are still met even with velic leakage during the first part of the stop as long as the velic valve is closed just before the release and pressure is allowed to build up behind the closure. However, voiceless stops have less tolerance for such leakage because any nasal sound – voiced or voiceless – would undercut either their stop or their voiceless character. (1991, p. 213)

Summary

The meta-principle governing these patterns of assimilation of nasalization would seem most simply to be: preserve the acoustic-auditory norm. As far as articulation is concerned, we might expect spreading nasalization to pass through a /d/ about as well as it could through the lateral /l/. From an articulatory point of view a late velic closure during a /t/ should be similar to that for a /d/. We might expect a word-initial /b/ to be articulated in the same way, no matter whether it followed an oral or nasal vowel in the preceding word. But our expectations would be wrong except insofar as we take into account the aerodynamic and thus acoustic-auditory outcomes. Attempts to account for the sound patterns in language, including those derived from sound changes, by reducing speech behavior exclusively or predominantly to articulatory facts are missing a major part of the picture.

Those systems for automatic speech recognition which include models or 'grammars' of segments could exploit some of the above principles. A typical grammar of a voiced stop might include: "abrupt decrease in amplitude simultaneous with an abrupt change in spectrum followed by voicing during the low amplitude portion, with duration not less than 30 msec, followed by an abrupt increase in amplitude (including an optional burst) plus an abrupt change in spectrum, etc." The weight attached to the abrupt decrease in amplitude could be reduced following a nasal element, or the weight attached to the increase in amplitude at release could be given more weight in all cases.

COARTICULATION AND THE "PARSING" PROBLEM

Coarticulation figures in another very important asymmetry in languages' phonologies (due to sound change).

For the listener coarticulation creates a potential parsing problem. 'Parsing' here means determining the cause-and-effect relations between events. Given the juxtaposed phonetic events A B C, is B to be parsed with A and/or C (and thus one of their perceptual cues) or is it independent? For example, in the pronunciation of the North American (NA) English "measure" /mɛʒɚ/ there is a sharply rising F_2 and F_3 leading into the [z]. Is this to be parsed with the consonant or the vowel? In most dialects, including my own, this is parsed with the palatal consonant [z], but some listeners must have parsed it with the vowel thus giving rise to the pronunciation [mejʒɚ]. In the case of the epenthetic stops discussed above, listeners must start out by parsing them with the surrounding segments but in words such as *humble* and *Thompson* some listeners must eventually have regarded them as independent events (i.e., to be parsed with themselves). So the type of sound changes mentioned above can be viewed as parsing errors made by the listener.

We might speculate plausibly that the more often speech events occur independently of each other, the easier it is to see them as independent when on occasion they do appear together. Conversely, the more often phonetic events occur together, the less easy it will be to parse them into independent entities. It is a significant fact that most sound changes occur in the domain of the **word** where phonetic events occur in permanent and immediate juxtaposition, i.e., under situations where there is likely to be maximal parsing ambiguity. (Cf. Passy, 1890, p. 168.) Looked at another way: Speech may exhibit as much across-word as within-word coarticulation (recall the case of the across-word coarticulation where French /sɛ̃ bɛl/ becomes [sɛ̃ᵐbɛl]); nevertheless, across-word coarticulation leads to sound change far less often than the within-word variety – conceivably because words are far more permutable and thus isolatable than speech sounds.[10]

DIFFERENTIATING FOSSILIZED FROM REAL-TIME COARTICULATION

Given that there are parallels between real-time coarticulation and the fossilized result of it embedded in the phonology of languages, there can be some difficulty in trying to differentiate the two. A case like the sound change, Late Latin /bon/ > French /bõ/, is not a problem because here the environment conditioning the vowel nasalization has been lost, clearly marking this as a case of fossilized or phonologized coarticulation.

[10] This needs qualification. On the one hand, across-word assimilations are certainly not rare. Some well-known examples include the case of palatalization in English (e.g., where the /-s#š-/ sequence in *gas shortage* > becomes [ʃ]), Sanskrit external word sandhi, and Insular Celtic consonant mutations. Some of these cases undoubtedly came about due to the words involved being in frequent juxtaposition and thus difficult to parse. On the other hand, it is possible for word boundaries to be marked by phonetic means, e.g., the vowel /i/ in "sea made" is less nasalized than that in "seam made" (Ohala, 1975). But, in general, word boundaries in connected speech are indistinct.

But nasalization of vowels before nasal consonants in NA English, e.g., "bean" as [bĩn], is problematic. On the one hand such contextual phonetic nasalization of the vowel is not unexpected, but on the other hand NA English nasalizes vowels far more than other languages do, i.e., well before the VN intersection (Clumeck, 1976). Some of the vowel nasalization may be phonological and not strictly due to coarticulation. There is as yet no general solution to this problem. However, Solé (1992) offers an approach for differentiating the two sources of nasalization on vowels: CVN words spoken at different speeds show little temporal variability in the extent of nasalization in Spanish; in English, however, the duration of vowel nasalization is highly variable, correlated with the duration of the vowel. Thus at least some fraction of this nasalization may be introduced at a high level. (See also Ohala, 1981.)

CONCLUSION

Both theoretical and applied phonetics, e.g., speech technology, should benefit from the extraction of the 'laws and the lore' of coarticulation to be derived from phonological study. However, in order to figure out how coarticulation works from an examination of phonological data:

- There is a need to differentiate sound patterns that are due to coarticulation from those that are not. If we failed to do this we would run the risk of attributing to coarticulation phenomena due to other factors or *vice-versa*.
- There is a need to differentiate between "real-time" coarticulation and the fossilized or phonologized coarticulation due to sound change. The primary difference between the two is that only in the former is there a mechanical link between the variation noted and the contextual (conditioning) environment. Thus in fossilized coarticulation one often encounters variation that is disproportionate with respect to the conditioning environment, i.e., exaggerated either in magnitude or temporal extent. Of course, clear cases of fossilized coarticulation are those where the conditioning environment has been lost through sound change.

From an inspection of selected examples of fossilized coarticulation, I conclude that:

- One of the important constraints on coarticulation is that there is higher priority on maintaining an acoustic-auditory norm for speech elements rather than an articulatory norm.
- The variation in the form of speech caused by coarticulation is usually "undone" or normalized by the listener but that this may be less easy to do when the blended sounds always occur together as happens especially within a word.

REFERENCES

BEDDOR, P.S., KRAKOW, R.A., and GOLDSTEIN, L.M. (1986). Perceptual contraints and phonological change: A study of nasal vowel height. *Phonology Yearbook*, **3**, 197–217.

BENDOR-SAMUEL, J.T. (1960). Some problems of segmentation in the phonological analysis of Tereno. *Word*, **16**, 348–355.

BENDOR-SAMUEL, J.T. (1966). Some prosodic features in Tereno. In C.E. Bazell, J.C. Catford, M.A.K. Halliday, and R.H. Robins (eds.), *In Memory of J.R. Firth* (pp. 30–39). London: Longmans.

BINDSEIL, H.E. (1838). *Abhandlungen zur allgemeinen vergleichenden Sprachlehre*. Hamburg.

BLOOMFIELD, L. (1911). The Indo-European palatals in Sanskrit. *American Journal of Philology*, **32**, 36–57.

BLUMSTEIN, S.E. (1986). On acoustic invariance in speech. In J.S. Perkell and D.H. Klatt (eds.), *Invariance and Variability in Speech Processes* (pp. 178–193). Hillsdale, NJ: Lawrence Erlbaum.

BRÜCKE, E. (1871). *Die physiologischen Grundlagen der neuhochdeutschen Verskunst*. Wien: Carl Gerold's Sohn.

CLUMECK, H. (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics*, **4**, 337–351.

DIEHL, R. (ed.) (1991). Theme issue: On the relation between phonetics and phonology. *Phonetica*, **48**, 77–278.

DOCHERTY, G.J., and LADD, D.R. (eds.) (1992). *Proceedings of the Second Laboratory Phonology Conference: Gesture, Segment, Prosody*. Cambridge, U.K.: Cambridge University Press.

DORMAN, M.F., RAPHAEL, L.J., LIBERMAN, A.M., and REPP, B. (1975). Some maskinglike phenomena in speech perception. *Haskins Laboratories Status Reports on Speech Research*, **SR–42/43**, 265–276.

EDKINS, J. (1864). *A Grammar of the Chinese Colloquial Language Commonly Called the Mandarin Dialect*. Shanghai: Presbyterian Mission Press.

ELMAN, J., and MCCLELLAND, J. (1986). Exploiting lawful variability in the speech wave. In J.S. Perkell and D.H. Klatt (eds.), *Invariance and Variability in Speech Processes* (pp. 360–380). Hillsdale, NJ: Lawrence Erlbaum.

FANT, G. (1973). *Speech Sounds and Features*. Cambridge, MA: M.I.T. Press.

FOWLER, C.A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, **8**, 113–133.

FOWLER, C.A. (1986). An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics*, **14**, 3–28.

FUJIMURA, O., MACCHI, M.J., and STREETER, L.A. (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech*, **21**, 337–346.

GRAMMONT, M. (1933). *Traité de Phonétique*. Paris: Librairie Delagrave.

GRANDGENT, C.H. (1896). Warmpth. *Publications of the Modern Language Association*, **11** (New Series **4**), 63–75.

GRANDGENT, C.H. (1927). *From Latin to Italian*. Cambridge, MA: Harvard University Press.

HAUDRICOURT, A.G. (1961). Bipartition et tripartition des systèmes de tones dans quelques langues d'Extrême Orient. *Bulletin de Société Linguistique de Paris*, **56**, 163–180.

HENDON, R.S. (1966). *The Phonology and Morphology of Ulu Muar Malay*. Yale University Publications in Anthropology, **70**.

HOCK, H.H. (1986). *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.

HOMBERT, J.-M., OHALA, J.J., and EWAN, W.G. (1979). Phonetic explanations for the development of tones. *Language*, **55**, 37–58.

KAWASAKI, H. (1986). Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. In J.J. Ohala and J.J. Jaeger (eds.), *Experimental Phonology* (pp. 81–103). Orlando, FL: Academic Press.

KENT, R.G. (1932). *The Sounds of Latin*. Baltimore: Waverly Press.

KINGSTON, J., and BECKMAN, M.E. (eds.) (1990). *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech.* Cambridge, U.K.: Cambridge University Press.

KLATT, D.H. (1986). The problem of variability in speech recognition and in models of speech perception. In J.S. Perkell and D.H. Klatt (eds.), *Invariance and Variability in Speech Processes* (pp. 300–319). Hillsdale, NJ: Lawrence Erlbaum.

LADEFOGED, P., and BROADBENT, D. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America,* **29,** 98–104.

LI, F.-K. (1977). *A Handbook of Comparative Tai.* Honolulu: University of Hawaii Press.

LINDBLOM, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America,* **35,** 1773–1781.

LINDBLOM, B., and STUDDERT-KENNEDY, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America,* **42,** 830–843.

LÖFQVIST, A., BAER, T., MCGARR, N.S., and STORY, R.S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America,* **85,** 1314–1321.

MALÉCOT, A. (1956). Acoustic cues for nasal consonants: An experimental study involving tape-splicing technique. *Language,* **32,** 274–284.

MANN, V.A., and REPP, B.H. (1980). Influence of vocalic context on perception of the [ʃ]–[s] distinction. *Perception & Psychophysics,* **28,** 213–228.

MASPERO, H. (1912). Études sur la phonétique historique de la langue Annamite: Les initiales. *Bulletin de l'École Française d'Extrême Orient,* **12,** 114–116.

MENZERATH, P., and LACERDA, A. (1933). *Koartikulation, Steuerung und Lautabgrenzung: Eine experimentelle Untersuchung.* Berlin and Bonn: Ferd. Dümmler.

MEYER, E.A. (1903). *Englische Lautdauer. Eine experimentellephonetische Untersuchung.* Uppsala: Akademiska Bokhandeln.

MILLER, G.A., and NICELY, P.E. (1955). Analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America,* **27,** 338–353.

MURRAY, R.W. (1982). Consonant cluster developments in Pali. *Folia Linguistica Historica,* **3,** 163–184.

NEAREY, T.M., and HOGAN, J.T. (1986). Phonological contrast in experimental phonetics: Relating distribution of production data to perceptual categorization curves. In J.J. Ohala and J.J. Jaeger (eds.), *Experimental Phonology* (pp. 141–161). Orlando, FL: Academic Press.

NOLAN, F. (1992). The descriptive role of segments: Evidence from assimilation. In G.J. Docherty and D.R. Ladd (eds.), *Papers from the Second Laboratory Phonology Conference: Gesture, Segment. Prosody (pp. 261–280). Cambridge, U.K.: Cambridge University Press.*

OERTEL, H. (1901). *Lectures on the Study of Language.* New York: Charles Scribner's Sons.

OHALA, J.J. (1974). Experimental historical phonology. In J.M. Anderson and C. Jones (eds.), *Historical Linguistics II. Theory and Description in Phonology* (pp. 353–389). Amsterdam: North Holland.

OHALA, J.J. (1975). Phonetic explanations for nasal sound patterns. In C.A. Ferguson, L.M. Hyman, and J.J. Ohala (eds.), *Nasálfest: Papers from a Symposium on Nasals and Nasalization* (pp. 289–316). Stanford: Language Universals Project.

OHALA, J.J. (1978). Southern Bantu vs. the world: The case of palatalization of labials. *Berkeley Linguistic Society, Proceedings, Annual Meeting,* **4,** 370–386.

OHALA, J.J. (1981). Speech timing as tool in phonology. *Phonetica,* **38,** 204–212.

OHALA, J.J. (1983a). The phonological end justifies any means. In S. Hattori and K. Inoue (eds.), *Proceedings of the XIIIth International Congress of Linguists, Tokyo, 29 Aug.–4 Sept. 1982* (pp. 232–243). Tokyo. [Distributed by Sanseido Shoten.]

OHALA, J.J. (1983b). The origin of sound patterns in vocal tract constraints. In P.F. MacNeilage (ed.), *The Production of Speech* (pp. 189–216). New York: Springer-Verlag.

OHALA, J.J. (1985). Linguistics and automatic speech processing. In R. De Mori and C.-Y. Suen (eds.), *New Systems and Architectures for Automatic Speech Recognition and Synthesis* (pp. 447–475). Berlin: Springer-Verlag.

OHALA, J.J. (1986). Discussion. In J.S. Perkell and D.H. Klatt (eds.), *Invariance and Variability in Speech Processes* (pp. 197–198). Hillsdale, NJ: Lawrence Erlbaum.

OHALA, J.J. (1989). Sound change is drawn from a pool of synchronic variation. In L.E. Breivik and E.H. Jahr (eds.), *Language Change: Contributions to the Study of its Causes* (pp. 173–198). Berlin: Mouton de Gruyter.

OHALA, J.J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 258–275). Cambridge, U.K.: Cambridge University Press.

OHALA, J.J. (1992a). What's cognitive, what's not, in sound change. In G. Kellermann and M.D. Morrissey (eds.), *Diachrony within Synchrony: Language History and Cognition* (pp. 309–355). Frankfurt/M: Peter Lang Verlag.

OHALA, J.J. (1992b). What is the input to the speech production mechanism? In Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (eds.), *Speech Perception, Production, and Linguistic Structure* (pp. 297–311). Tokyo: Ohmsha.

OHALA, J.J., and JAEGER, J.J. (eds.) (1986). *Experimental Phonology.* Orlando, FL: Academic Press.

OHALA, J.J., and LORENTZ, J. (1977). The story of [w]: An exercise in the phonetic explanation for sound patterns. *Berkeley Linguistic Society, Proceedings, Annual Meeting,* **3**, 577–599.

OHALA, J.J., and OHALA, M. (in press). The phonetics of nasal phonology: Theorems and data. In M. Huffman and R. Krakow (eds.), *The Feature Nasal: Phonetic Bases and Phonological Implications.* Orlando, FL: Academic Press.

OHALA, M., and OHALA, J.J. (1991). Nasal epenthesis in Hindi. *Phonetica,* **48**, 207–220.

PASSY, P. (1890). *Études sur les Changements Phonétiques.* Paris: Firmin-Didot.

PAUL, H. (1880). *Principien der Sprachgeschichte.* Halle: Max Niemeyer.

PHELPS, J. (1937). Indo-european initial *sl. Language,* **13**, 279–284.

PICKETT, J.M., and DECKER, L.R. (1963). Time factors in perception of a double consonant. *Language and Speech,* **3**, 11–17.

PIGGOT, G.L., and SINGH, R. (1985). The phonology of epenthetic segments. *Canadian Journal of Linguistics,* **30**, 415–451.

PRIDEAUX, G.D., DERWING, B.L., and BAKER, W.J. (eds.) (1980). *Experimental Linguistics. Integration of Theories and Applications.* Ghent: E. Story-Scientia.

REPP, B.H. (1978). Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics,* **24**, 471–485.

ROBINS, R.H. (1957). Vowel nasality in Sundanese. In *Studies in Linguistic Analysis* (pp. 87–103). Oxford: Blackwell's.

ROCK, I. (1983). *The Logic of Perception.* Cambridge, MA: M.I.T. Press.

SCHOURUP, L.C. (1973). A cross-language study of vowel nasalization. *Ohio State University Working Papers in Linguistics,* **15**, 190–221.

SOLÉ, M.-J. (1992). Phonetic and phonological processes: The case of nasalization. *Language and Speech,* **35**, 29–44.

STEVENS, K.N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America,* **50**, 1180–1192.

STRANGE, W., VERBRUGGE, D., SHANKWEILER, D., and EDMAN, T. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America,* **60**, 213–224.

SVANTESSON, J.-O. (1983). *Kammu Phonology and Morphology.* [Travaux de l'institut de linguistique de Lund 18.] Lund: Gleerup.

WETZELS, W.L. (1985). The historical phonology of intrusive stops: A non-linear description. *Canadian Journal of Linguistics,* **30**, 285–333.

WEYMOUTH, R.F. (1856). On the liquids, especially in relation to certain mutes. *Transactions of the Philological Society* [London], pp. 18–32.

WHITNEY, W.D. (1867). *Language and the Study of Language.* New York: Charles Scribner and Co.

WINITZ, H., SCHEIB, M.E., and REEDS, J.A. (1972). Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *Journal of the Acoustical Society of America,* **51**, 1309–1317.