THE PHONETICS OF NASAL PHONOLOGY: THEOREMS AND DATA

JOHN J. OHALA*
MANJARI OHALA†

*Department of Linguistics
University of California
Berkeley, California 94720

†Department of Linguistics and Language Development
San Jose State University
San Jose, California 95192

1. INTRODUCTION

Physiologically a nasal speech sound is quite simple: it just involves lowering of the soft palate to a degree sufficient to couple the oral and nasal cavities acoustically. With a concomitant oral closure, a nasal consonant is produced; without it, a nasal vowel. Nevertheless, this simple gesture has major and complex phonological consequences due to the interconnectedness of all parts of the vocal tract. In this article we propose to give an account of some of the complex phonological behavior involving nasal segments, summarizing briefly material we have published previously.¹

Another purpose of this article is to explore the extent to which sound patterns in language can be derived (almost) like theorems from first principles, the latter being facts that are empirically verifiable and pertinent as well to domains other than speech. This exercise is thus guided by Lindblom’s charge to “derive language from non-language” (Lindblom, 1984). Such a program should be contrasted with that of mainstream phonology, which rather attempts to account for
sound patterns in language by positing structures, processes, and constraints that are domain-specific. Moreover, these structures and processes rest on a greatly simplified phonetics that neglects the interactions between the articulatory, aerodynamic, and the acoustic–auditory links in the speech chain. The sound patterns of nasal segments constitute an ideal arena where any approach to phonology can be given a trial.

2. AERODYNAMICS

2.1. Introduction

The function of the speech mechanism may be considered the conversion of static or slowly varying air pressure (so-called “DC” air pressure) into the rapid air pressure variations we call sound (“AC” air pressure). This task is accom-

![Diagram of the vocal tract](image)

Figure 1. The vocal tract schematically represented as an aerodynamic mechanism: a series of chambers each capable of having its volume and thus air pressure varied by means of pistonlike structures and all interconnected by means of valves which regulate the flow of air into and out of the chambers. Reproduced from J. Ohala (1983b) with permission.

Phonetics of Nasal Phonology

plished by what is in effect an interconnected series of chambers whose volumes, and thus also air pressures, may be varied by pistonlike mechanisms and whose input and output of air flow are regulated by various valves. This is represented schematically in Figure 1.

Certain fundamental aerodynamic principles govern the behavior of this mechanism (as well as other aerodynamic devices, e.g., bicycle pumps, automobile engines, vacuum cleaners). Assuming no net change in temperature (i.e., adiabatic conditions), pressure times volume is a constant. Thus for a fixed mass of air, a decrease or increase in volume causes respectively an increase or decrease in pressure. For a fixed volume, an increase or decrease of air mass in a chamber causes respectively an increase or decrease in pressure. The volume of air moving through a valve is a function of the difference in pressure on the opposite sides of the valve and of its aperture. Some other principles will be reviewed below. (See also Catford, 1977; J. Ohala, 1983b, 1990c; Scully, 1987, 1990).

The nasal cavity has one opening at the nostrils, which can be narrowed slightly but not closed completely by any of its intrinsic musculature, and a valve, the velum, at the other end, which can open to varying degrees and thus give access to the oral cavity at a point about midway between the lip opening and the glottis. This topological arrangement of the nasal cavity, its opening and valves with respect to the oral cavity in conjunction with the physical principles presented above, gives rise to a number of sound patterns encountered repeatedly in various languages, as reviewed in the following sections.

2.2. Theorems on Phonetic and Phonological Universals Derived from Aerodynamic Primitives

2.2.1. Buccal Obstruents Require Velic Closure

Theorem A. The velic valve must be closed (i.e., the soft palate must be elevated) for an obstruent articulated further forward than the point where the velic valve joins the nasal cavity and the oral cavity.

(For ease of reference we use the term buccal for any place of articulation that is forward of the point where the velic valve joins the oral and nasal cavities.) The purpose of the buccal constriction during the production of a buccal obstruent is to build up the air pressure behind it which, when released, creates audible turbulence. Failure to close the velic valve would allow the air to escape through the nose, thus reducing or eliminating the required pressure drop across the oral constriction. It is the inability to make such a tight valvular closure that greatly reduces the intelligibility of cleft palate speech.

Schadeberg (1982) discusses a possible counterexample to Theorem A. He claims that UMBundu (spoken in Angola) possesses a nasalized voiced fricative [:] but it remains to be established instrumentally whether this is indeed an obstruent, i.e., that it possesses some of the acoustic cues associated with air pressure build-up. Many so-called nonsibilant "voiced fricatives" such as [v, ð, β, γ] do
not have appreciable friction and are, rather, frictionless continuants (J. Ohala, 1983b: 202–203; Pickett, 1980). Ternes (1973) asserts that the Applecross dialect of Scots Gaelic has several nasalized voiceless fricatives, among them [ʕ, ç, ś], in words such as /sśhι/ [sśhι] ‘tame’. If so, this would undercut Theorem A. Before abandoning it, however, it would be best if the position of the velum during these sounds could be determined instrumentally. One need not take the presence of nasalized vowels next to these sounds as unambiguous evidence of the position of the velum during the consonants themselves. In a personal communication (21 August 1991), Ternes indicates that his claim of the existence of nasalized voiceless buccal fricatives was based on kinesthetic sensations during the imitation of these sounds. He concurs with the need for an instrumental study of their phonetic nature. (See Cohn, this volume, for additional references on reported nasalized buccal fricatives.)

Phonological evidence that nasalization is associated with defricativization, that is, that an open velic valve bleeds buccal obstruction and its concomitant turbulence, is given in (1).

(1) Nasalization blocks buccal obstruction:

a. In English, /h/, which in initial position can be considered a voiceless version of whatever sound follows (Lehiste, 1964, Chap. 5), has the allophone [ç] before /ʃ/, e.g., hue [çju], but [h] if there is simultaneous nasalization, e.g., “inhuman” [änhju:mən], not */änçjümən/.

b. In Fante, the word with the underlying form /ḥ/ ‘border’ is realized phonetically as [çɪ], but the word /hɪ/ ‘where’ is phonetically [hɪ], not */çɪ/ (Schachter & Fromkin, 1968).

c. In Yuchi, voiceless spirants appear (apparently epenthetically) between vowels and following lingual stops but not if the vowel is nasalized (Wagner, 1934).

d. In Jivaro, voiceless oral vowels appear when unstressed, but voiceless nasal vowels were not found (Beasley & Pike, 1957). (We assume, following the arguments in J. Ohala, 1983b, that voiceless vowels, which are typically high close vowels, arise in part due to the back pressure created from the narrow buccal constriction; an open velic valve would bleed this back pressure.)

Several corollaries of (A) may be formulated:

Corollary A.1: If, in the transition between a nasal consonant and a buccal stop, the velic closure becomes desynchronized with respect to the oral obstruent and is made during the nasal, a stop necessarily homorganic with the nasal will appear.

This phenomenon, which was noted and explained in phonetic terms as early as 1838 by Bindseil and in 1856 by Weymouth (see also Grandgent, 1896; Millardet, 1910: 94), underlies the sound changes and variants listed in (2).

(2) a. Epenthetic stops in English:

<table>
<thead>
<tr>
<th>Orthography</th>
<th>Phonetic</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>warmth</td>
<td>wɔrmθ</td>
<td>warm + [θ]</td>
</tr>
<tr>
<td>something</td>
<td>sʌmpθɪŋ</td>
<td>some + thing</td>
</tr>
<tr>
<td>Thompson</td>
<td>ˈθʌmpən</td>
<td>Thom + son</td>
</tr>
<tr>
<td>glimpse</td>
<td>ˈɡlɪmps</td>
<td>gleam + s</td>
</tr>
<tr>
<td>teamster</td>
<td>ˈtɪmplstə</td>
<td>team + ster</td>
</tr>
<tr>
<td>youngster</td>
<td>ˈʃʌŋkstə</td>
<td>young + ster</td>
</tr>
<tr>
<td>length</td>
<td>lɛŋkθ</td>
<td>long + [θ]</td>
</tr>
</tbody>
</table>

b. Pre-Latin > Latin:

*em-tos > ēmpthus ‘a purchase’

(These examples show that the voicelessness of the following obstruent was also anticipated during the latter portion of the nasal.) It should be noted that although most such cases seem to involve a sequence of a nasal followed by an oral obstruent, suggesting that anticipatory denasalization is the most common direction of this assimilation, Varma (1961: 123) provides apparent examples in certain Indo-Aryan languages where this order is reversed: /kʂṇat/ > /kʂ[h]nt/ ‘Krishna’. The nonstandard North American English pronunciation of ‘business’ as [bɪ[dnt] may constitute another example, if one assumes the intermediate form */bɪzdnt/.

Since pharyngeal and glottal obstructions block airflow at a point further “upstream” from the velic opening, the action of the velic valve neither breezes nor feeds their obstruction. Although we are unable to cite systematic evidence—evidence of what does not occur is always somewhat difficult to find—it seems clear that although English, for example, shows epenthetic stops in words such as teamster, somepthing, one would never find them when the obstruent following the nasal is an [h], as in *somepthing. Likewise, a hypothetical Arabic word *ʔimfat/ with a nasal + voiceless pharyngeal fricative would not be expected to show an epenthetic stop: *ʔimfat/. It is a challenge to feature geometry to represent in a natural way such asymmetrical behavior of buccal and nonbuccal obstructions.

There are cases quite parallel to those in (2) which involve not buccal obstructions adjacent to nasals but sonorants, including [l] and vowels; see (3). (See also Schourup, 1973.)

(3) a. Latin > Gallo-Roman:

<table>
<thead>
<tr>
<th>Latin</th>
<th>Gallo-Roman</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>camera</td>
<td>tʃəmbre</td>
<td>‘room’</td>
</tr>
<tr>
<td>generum</td>
<td>ʤɛndre</td>
<td>‘son-in-law’</td>
</tr>
<tr>
<td>insimul</td>
<td>ɛnsəmble</td>
<td>‘together’</td>
</tr>
</tbody>
</table>

b. Spanish:

<table>
<thead>
<tr>
<th>Latin</th>
<th>Spanish</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>vent(ire)</td>
<td>vendre ‘sell’</td>
<td>(Pope, 1934: 148)</td>
</tr>
<tr>
<td>al hamra</td>
<td>Alhambra (literally ‘the red’</td>
<td>(Spaulding, 1965)</td>
</tr>
</tbody>
</table>
J. Ohala (1975) suggests that in such cases the sonorant consonants also require closure of the velar valve not for aerodynamic reasons but for acoustic reasons, namely, to reduce the acoustic distortion of these segments that would otherwise happen if they were nasalized. Given that the acoustic effects of nasal coupling are primarily in the lower end of the frequency spectrum, it is segments with low resonances that would show this preference for velar closure. Vowels with low F1 and/or F2 or those that are distinctively oral may also engender intrusive stops in adjacent nasals (assimilation of velar closure), as exemplified in (4).

(4) a. Ulu Muar Malay:
   [ban] ~ [bändi] 'doorsill'  
   (Hendon, 1966)

b. Tenango Otomi:
   /mohi/ [m^o^hi] 'plate'
   /nín/ [nín^e] 'your mouth'  
   (Bliht & Pike, 1976)

c. Korean:
   [mul] ~ [m^b^ul] 'water'  
   (Chen & Clumeck, 1975)

d. Latin > Italian dialects:
   rúnicem > rómaces ~ rómbice 'dock' (botany)
   comméatus > commiatio ~ combiatio 'leave-taking'  
   (Grandgent, 1927:118)

From this we may rephrase Corollary A.1 as A.2:

Corollary A.2: If, in the transition between a nasal consonant and a segment requiring orality, the velar closure becomes desynchronized with respect to this oral segment and is made during the nasal, a stop necessarily homorganic with the nasal will appear.

A further corollary of A is A.3:

Corollary A.3: Assimilating nasalization is blocked by buccal obstruents.

Cases showing the blocking of nasal prosodies by buccal but nonbuccal (pharyngeal or glottal) obstruents are given in (5).

(5) a. Sundanese:
   ṇañiñ 'to wet'
   ṇñphār 'to be rich'
   ṇñřok 'to inform'
   miñasih 'to love'  
   (Robins, 1957)

b. Tereno:
   1st person 3rd person
   pïho m^biho 'i/he went'
   ahja?afó ā^n^2a?afó 'i/he desire(s)'
   iso ī^n^zo 'i/he hoed'
   owoku òwó^g'u 'my/his house'
   ajo ājö 'my/his brother'

In (5a) it is seen that perseveratory nasalization following a nasal consonant spreads throughout an entire word unless blocked by a buccal obstruant such as [k] or [s]; however, it passes through the glottal obstruents [h, ʔ]. In Tereno (5b), nasalization acts like a prosody that is “injected” at the start of a word to mark first person singular. Again, this perseveratory nasalization is blocked by buccal obstruents such as /p, k, s/, but it causes them to become voiced and prenasalized [n^b, n^g, n^z]. The prosody passes through the glottal stop [ʔ] but, curiously, not through [h] or [h], in apparent contradiction to Corollary A.3. However, there is comparative evidence that at least the [h] (and perhaps the [h]) derives from an earlier apical obstruent, /t/, which plausibly passed through an intermediate stage of /s/ before becoming Tereno /h/ in nonnasal environments (see J. Ohala, 1983b).

We would have to conclude that the changes effected by the nasal prosody in Tereno today are morphologically determined; that is, after becoming phonologized they are no longer influenced solely by phonetic factors (this is probably true in the case of Sundanese, too). As a further reminder that the phonetic naturalness of some phonological behavior is to be sought in the phonetic conditions in the given language’s past, not its present, Court (1970) documents cases in Mentu Land Dayak (MLD) where spreading nasal prosodies are inhibited by nasal consonants (among other segments). Further investigation reveals, however, that these nasals derive from earlier prenasalized voiced stops. For example, in MLD [ɲuːa] ‘to rent’, which shows the nasal prosody spreading to the end of the word, the [ɲ] derives from a historical simple nasal (cf. the cognate word in Johore Malay [ɲuːa]); but in MLD [ɲuːa] ‘to sell’, which shows the nasal prosody extending only to the immediately following vowel, the [ɲ] derives from an earlier prenasalized stop (cf. the cognate word in Johore Malay [ɲuːa]).

### 2.2.2. Devoiced, A Nasal Is A Fricative

According to Chomsky & Halle (1968:302), and most phonological theories influenced by them, one can define a critical degree of constriction such that wider constrictions are [+sonorant] and narrower ones are [−sonorant]. However, what was not anticipated in this definition is that some [+sonorant] segments can become [−sonorant] not by changing the degree of their constriction but simply by changing from [+voice] to [−voice].

In order to explain this fact, some additional aerodynamic principles must be reviewed first (see also Catford, 1977; J. Ohala, 1976, 1983b, 1990c; Scully, 1987, 1990; Shadle, 1990; and Stevens, 1971). Principles of fluid dynamics (which
cover the flow of gases) reveal that under ideal conditions of airflow through a tube one can identify a critical threshold at which smooth or ‘laminar’ flow changes to turbulent flow. The resistance to airflow also increases sharply when this threshold is passed. One might be tempted to associate [+ sonorant] with conditions of airflow which are lower than this threshold and [− sonorant] to those which exceed it except for two factors. First, conditions of airflow in the vocal tract are not “ideal”; there is some turbulence at almost any rate of airflow (J. Ohala, 1990c). Second, among the relevant variables which determine this threshold (or which under nonideal conditions contribute to any increase in turbulence and resistance to airflow) are not only the area of the constriction but also the velocity of the airflow.4 Thus, given a certain vocal tract configuration, say that for a palatal glide [j], there might be no appreciable turbulence when it is voiced, since the vibrating vocal cords offer sufficient resistance to the pulmonic airflow so that the velocity of the air flowing past the palatal constriction is relatively low. However, under voiceless conditions the same configuration may lead to noticeable turbulence, since now the escaping pulmonic air is virtually unchecked and reaches much higher velocity levels. Thus, as mentioned earlier (1a), /h/, which in initial position is phonetically a voiceless version of the following sound, has the fricative allophone [ç] preceding [j], e.g., English /hu/ [çu] ‘hue’.

The same principles apply to nasal consonants when they become voiceless. Distinctive voiceless nasals are rather rare; this rarity itself may stem from the principles just reviewed. First, the principal point at which the turbulence is generated in voiceless nasals is the nostrils and, of course, this will apply to all voiceless nasals no matter what buccal place of articulation they have. There will thus be no difference between [ŋ], [ŋ], [ñ], and so on during the consonantal constriction, though of course they could still be differentiated via their transitions in adjacent vowels. This may be the reason why all distinctive voiceless nasals, such as those in Burmese, are phonetically just half-voiceless—thus, /ŋ/ = [ŋm] (Ladefoged, 1971:11). Second, speakers have relatively little control over the degree of constriction at the nostrils, and even when they are maximally constricted (which occurs automatically only when breathing in and thus not during the expiratory airflow of speech), it is not sufficient to generate very intense noise. Additionally, such voiceless nasals share with bilabial nasals the lack of a downstream resonator to amplify and shape the noise. As a result, the fricative created during voiceless nasals is low in intensity and devoid of much distinctive spectral shaping. Thus, auditorily, they are nonoptimal as speech sounds. Nevertheless, on the rare occasions when they do occur, they may show obstruent-like (i.e., [− sonorant]) behavior, as given in (6).

(6) a. The voiceless nasals in Burmese stem from original /s/ + nasal clusters; e.g., corresponding to spoken Burmese /ŋa/ ‘nose’, one finds orthographic Tibetan sna, which gives evidence of the earlier phonetic form (Graham Thurgood, personal communication).
3.2. Theorems on Phonetic and Phonological Universals Derived from Acoustic Primitives

3.2.1. BACK NASALS ARE LESS CONSONANTAL THAN FRONT NASALS

Based on the above principles we derive Theorem C:

Theorem C. The further back a nasal consonant is articulated, the less “consonantal” it is.

The basis for this is as follows. The further back the oral constriction is, the higher will be the antiresonances contributed by the oral cul-de-sac branching off the pharyngeal–nasal airway. Given this and the fact that in the spectrum of a nasal, as in all voiced sounds, energy decreases with increasing frequency (Fant, 1960), the antiresonance will fall in the high end of the spectrum, which has very little salient acoustic energy or spectral peaks. The listener is therefore more likely to overlook it. Insofar as the auditory effects of the antiresonance may be weak-

ended, the spectrum that remains will be dominated by the resonances of the pharyngeal–nasal airway and will resemble a simple nasalized vowel. A back-articulated nasal consonant will therefore not be too dissimilar from a flanking nasalized vowel.

A second factor also affects back nasals’ consonantality: a “good” consonant, according to Stevens (1980, 1989), is one that creates an abrupt change in amplitude and spectrum with respect to immediately neighboring vocalic segments. This follows from the auditory system’s greater sensitivity to abrupt rather than slow modulations of acoustic parameters (Møller, 1971). Back consonants—nasal or not—by virtue of having the massive tongue dorsal as an articulator have longer, slower transitions than those produced by the lips or tongue tip (Kewley-Port 1982; Lehiste & Peterson 1961). Thus with two of the principal cues for consonantality weakened, back nasals may be expected to be less common than front nasals, to have a more restricted distribution, and to alternate with nasalized vowels or glides. Evidence supporting this is given in (7) and (8).

(7) Maddieson (1984) in a survey of 317 languages found the following incidence of nasals at different places of articulation:

<table>
<thead>
<tr>
<th>Articulation</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dental/Alveolar</td>
<td>316</td>
</tr>
<tr>
<td>Bilabial</td>
<td>299</td>
</tr>
<tr>
<td>Velar</td>
<td>167</td>
</tr>
<tr>
<td>Palatal</td>
<td>107</td>
</tr>
</tbody>
</table>

(8) Data showing association of back nasals and nasalized vowels:

a. In Acatlan Mixtec, a word-final Ñ is optionally followed by a lenis velar closure. (Pike & Wistrand, 1974).

b. In Mbay, final Ñ in CVCV words often has a light velar closure, e.g., korñ ~ tìnc “is [korun].” (Caprile, 1968).

c. In Vietnamese, French loanwords with Ñ are replaced by VN, e.g., Fr. /ätñ/ “antenne” > Vietnamese [ǎntìnì].

d. House (1957) offers phonetic evidence of Ñ being confused with the velar nasal [ŋ].

3.2.2. LABIOVELAR NASALS TEND TO BE VEURAL, NOT LABIAL

A second acoustically derived theorem is D:

Theorem D. Doubly articulated nasals, e.g., [m̥n] and the nasal assimilating to the following labiovelar stops, will tend to pattern with consonants sharing the rearmost place of articulation rather than the frontmost.

There are two major cues to place of articulation in nasals: the spectrum of the nasal itself, and the transitions leading into or out of it. As regards the former, it is the rearmost constriction which defines the essential dimensions of the resonating cavity, that is, the length of the cavity from the point it branches from the
Figure 2. Schematic representation of the vocal tract during the production of /m n ñ ñ/ glide at the lower right of each shape; nostrils at upper left. The solid arrows mark the nasal-pharyngeal resonating cavity, which is the same for all four sounds, and the dotted arrows mark the portion of the oral resonating cavity branching off the pharyngeal cavity. It is the latter which helps to differentiate one nasal place of articulation from another. This oral cavity is the same for the velar nasal /ŋ/ and the nasalized labial velar /w/.

Pharyngeal-nasal airway to the point in the oral cavity where it terminates. Any additional constriction forward of that point is acoustically irrelevant. This is represented schematically in Figure 2. This is why doubly articulated nasals will, based on this cue, be most similar to a singly articulated nasal having the same place as the rearmost constriction of the double stop. As for the other cue, the transitions, in all likelihood they would be best unique to the doubly-articulated stop—see Gerya (1975) on the acoustic correlates of the labiovelar stop /kp/—or at worst share the same ambiguities for place as transitions do generally (Winitz, Scheib, & Reeds, 1972).

By far the most common doubly articulated consonants, stops and glides as well as nasals, are labiovelars, i.e., /j,m, kp, gb, w/ (J. Ohala & Lorentz, 1977). According to Theorem D, labiovelar nasals will tend to pattern with velars rather than labials. For the same reason, nasals assimilating to the other labiovelar consonants will tend to be velar rather than labial (insofar as they deviate from labiovelar, the default case).

Supporting evidence is given in (9).

(9) a. Tswana: nasal assimilating to [w] is velar:
-roma 'send' + wa 'passive marker' > -rojwa
-fena 'conquer' + wa > -fena
(Cole, 1955)

b. Kpelle: [w] patterns with velars in nasal assimilation:

<table>
<thead>
<tr>
<th>Indefinite</th>
<th>Definite</th>
</tr>
</thead>
<tbody>
<tr>
<td>ṭōk</td>
<td>'mɔ́k</td>
</tr>
<tr>
<td>ōw</td>
<td>'mũ w</td>
</tr>
<tr>
<td>yila</td>
<td>'gilai</td>
</tr>
<tr>
<td>wē</td>
<td>'jwe</td>
</tr>
</tbody>
</table>
(Welmers, 1962)

(10) The Efik stop inventory, in addition to /kp/, has /k/ and /kʰ/ but lacks a /p/.

The /kp/ is realized as [p] occasionally. Thus some analysts have proposed that the /kp/ is underlyingly [+anterior]. Nevertheless, the nasal assimilating to this /kp/ sometimes manifests itself as the [−anterior] nasal [ŋ] (when it is not [ŋ]) (Anderson, 1976; Cook, 1969; Welmers, 1973).

Second, in some cases the same labiovelar consonants pattern both with velars and with labials depending on the particular phonological process involved. The velar patterning dominates when [nasal] is involved, but labial patterning dominates when oral obstruents are involved (the latter for reasons detailed by Ohala & Lorentz). See (11) for some examples. These patterns undercut a simple structuralist account which would predict phonological behavior by reference to the relations these sounds have to the rest of the language's structure, including its phoneme inventory. While not denying the role of language structure, we maintain that this should not eclipse the important role of the physical phonetic content of speech sounds in influencing their behavior.

(11) a. The Yoruba labiovelar glide /w/ (along with the labiovelar stops /kp/ and /gb/) patterns with the labials /b f m/ in causing the merger of following /a/ with /i/. Nevertheless, the nasal assimilating to /w/ is the velar [ŋ]. (Ward, 1952)

b. In Tenango Otomi, /h/ becomes the labial fricative [F] before /w/ but /n/ assimilating to /w/ appears as the velar [ŋ]. (Blight & Pike, 1976)

These generalizations and data also undercut the assumption of generative phonology and its heirs that the phonological behavior of a segment derives from an abstract underlying feature specification—that is, that doubly articulated stops...
like labiovelars have to be either [+anterior] or [−anterior], not both (Anderson, 1976; Chomsky & Halle, 1968—but see Anderson, 1981).

In addition, it presents a challenge to those who believe that phonological processes can be represented better if the proper structure of features is specified via simple, transitive, asymmetrical, dependency relations, that is, using “feature geometry.” Most conceptions of feature geometry have [MANNER] and [PLACE] nodes strictly separated on the feature hierarchy. With this configuration, there would be no way for [NASAL] to influence [PLACE] (and to restrict this influence primarily to cases where labiovelar sounds were involved). The challenge, as we see it, is to be able to account for these feature-specific interactions using sufficiently but not excessively powerful principles which have validity both inside and outside phonology (see also J. Ohala, 1990a, 1992; J. Ohala & M. Ohala, 1991).

Finally, we consider three somewhat complex patterns involving nasals that draw on perceptual principles in addition to some of the constraints of speech production and acoustics reviewed above.

4. NASAL SOUND PATTERNS WITH A COMPLEX ETIOLOGY

4.1. Nasal Epenthesis before Voiced Stops

In the history of Hindi, it appears that an epenthetic nasal emerged between a nasalized vowel (V) and a following voiced stop but not between a V and a voiceless stop. The background for this is the following: In the development of the New Indo-Aryan languages from Middle Indo-Aryan, there occurred cluster (or geminate) simplification with compensatory lengthening of the preceding vowel. Thus Sanskrit rātri > MIA ratti > Old Hindi ra:ti6 ‘night’. If the cluster consisted of a nasal plus a stop, the vowel was not only lengthened but also nasalized, as shown in the third column of (12).

(12) Evidence of nasal epenthesis in Hindi:

<table>
<thead>
<tr>
<th>Sanskrit</th>
<th>Middle Indo-Aryan</th>
<th>Old Hindi</th>
<th>Modern Hindi</th>
</tr>
</thead>
<tbody>
<tr>
<td>aṅgana</td>
<td>aṅgana</td>
<td>ā: gana</td>
<td>[āṅgān]</td>
</tr>
<tr>
<td>candra</td>
<td>t[ə]nd</td>
<td>ċa: da</td>
<td>[tfənd]</td>
</tr>
<tr>
<td>danta</td>
<td>danta</td>
<td>dā: ta</td>
<td>[dāt]</td>
</tr>
</tbody>
</table>

However, in Modern Hindi [fourth column of (12)] words with voiced stops (but not those with a voiceless stop) following the nasalized vowel have a homorganic nasal (M. Ohala, 1983): [āṅgān] and [tfənd] but [dāt]. Is it plausible that such intrusive nasals would appear before a voiced but not a voiceless stop? Instrumental evidence suggests that it is (M. Ohala & J. Ohala, 1989, 1991; J. Ohala & M. Ohala, in press).

We recorded nasal air pressure (using a nasal “olive”) and audio from two speakers each (one male and one female) of Hindi and French. The utterances recorded included word-final nasalized vowels followed immediately by word-initial voiced or voiceless stops. When uttered in isolation these words would not show a nasal consonant, so any nasal consonant that would emerge when they abut would be a purely phonetic event, a product of the transition between V and a voiced stop. Thus the French utterances were of the type dit ‘saint’ pour moi ‘say “saint” for me’ /di sɛ̃ puð mwał/, dit ‘saint bel enfant’ say “saint” beautiful baby’ /di sɛ̃ bel ɑ̃fə/, and the Hindi ones: /ap ʃɔhə tɔko/ ‘you glance here’, /ap ʃɔhə dəkə/ ‘you see here’. The results showed that a nasal of up to 70 msec appeared before the voiced stop (i.e., the stop was prenasalized). For example, the sequence saint bel uttered by the French speakers was realized as [sɛ̃bɛl] (see also Cohn, 1990:108). Before voiceless stops, however, such a nasal was either absent or very brief, about 30 msec. This was true for each speaker of both languages and for all places of articulation.

The synchronic epenthetic nasal parallels the historical development of the nasal in Modern Hindi words such as [āṅgān] ‘courtyard’. In Modern Hindi, of course, the nasal in such words is a phonological nasal (i.e., present at the lexical level) and is not epenthetic except from a historical perspective. There is evidence that other languages exhibit similar behavior in that voiced stops, but not voiceless ones, tolerate nasal onsets (Cohn, 1990:108; Aguilar Cuevas, Machuca Ayuso, & Martínez Dauden, 1991; Childs, 1991; Duce, 1991; Paradis, 1988–1989; Roberts & Babcock, 1975; Suen & Beddoes, 1974; Yanagihara & Hyde, 1966).

M. Ohala & J. Ohala (1991:213) offered the following explanation for this pattern:

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 not met even with velar leakage during the first part of the stop as long as the velar 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.

The presence of the phonetic epenthetic nasal sets the stage for a sound change, that is, listeners may reinterpret the phonetically predictable event as a distinctive phonological event (J. Ohala, 1989, 1991).

4.2. Spontaneous Nasalization

Bloch (1920, 1965), Turner (1921) and Grierson (1922), studying Indo-Aryan languages, have called attention to what they call “spontaneous nasalization,” i.e., the development of distinctive nasalization on vowels in words that never had any lexical nasal consonant (the usual source of nasal vowels). One type of segment
that reappears in many of their examples is one characterized by high airflow, including any voiceless fricative, especially [h], aspirated stops, and affricates (M. Ohala, 1975, 1983); see (13).

(13) Spontaneous nasализation in Indo-Aryan:

<table>
<thead>
<tr>
<th>Sanskrit</th>
<th>Prakrit</th>
<th>Old Hindi</th>
<th>Modern Hindi</th>
<th>Bengali</th>
</tr>
</thead>
<tbody>
<tr>
<td>pakṣa</td>
<td>pakṣa</td>
<td>pākha</td>
<td>[pāṅkha]</td>
<td>‘a side’</td>
</tr>
<tr>
<td>akṣi</td>
<td>akṣi</td>
<td>[āṅkhi]</td>
<td>[āṅkhy]</td>
<td>‘eye’</td>
</tr>
<tr>
<td>uccaka</td>
<td>uccaka</td>
<td>[ūcā]</td>
<td>uṇča</td>
<td>‘high’</td>
</tr>
<tr>
<td>satya</td>
<td>satya</td>
<td>sāccha</td>
<td>sāccha</td>
<td>‘truth’</td>
</tr>
<tr>
<td>sarpa</td>
<td>sarpa</td>
<td>[sāp]</td>
<td>[sāp]</td>
<td>‘snake’</td>
</tr>
</tbody>
</table>

(based on Grierson, 1922)

The same phenomenon exists in other languages; see J. Ohala (1975), Matisoff (1975). The following passage from Jackson (1967) is typical.

It is remarkable that in French loanwords with “cz”, and in others with Fr. ss, there is a certain tendency for the vowel to become nasalized, giving MB [Middle Breton] “ncz” or nts (s), Mod. B. ńś in both cases. E.g., French maçon > Early Mod B mańczonner “mason”, Mod B. mańsoner; French rosse (with plural suffix, -ed) > MB ronceń or ronceet “horses” > Early Mod. B. ronćet, ronceet; French vis [“screw”] > MB vicę > Early Mod. B. vicz, hńcz or hńś > Mod. B. hńś. [p. 174]

As reviewed in J. Ohala and M. Ohala (in press), there is a plausible, if phonetically complex, argument to explain the association between high airflow segments and perceived nasализation:

1. High airflow segments like voiceless fricatives and aspirated stops require for their production a greater than normal glottal opening (vis-à-vis comparable voiceless segments like voiceless unaspirated stops).
2. This greater than normal glottal opening may spread via assimilation to the margins of adjacent vowels, even though these vowels may remain completely voiced.
3. This slightly open glottis creates acoustic effects due to some coupling between the oral and the subglottal cavities that mimic the effects of coupling of the oral and nasal cavities, i.e., lowered amplitude and increased bandwidth of F1.
4. Vowels that sound nasal to listeners, even though they are not physiologically nasal, can be reinterpreted and produced as nasal, thus precipitating a sound change.

Points (1–3) are well documented in the phonetic literature (Beddor, this volume; Fant, 1973; Fujimura & Lindqvist, 1971; Klatt, Stevens, & Mead, 1968; Sawashima, 1969). Ohala and Amador (1981, summarized in J. Ohala 1983a), studying both American English and Mexican Spanish, demonstrated the validity of point (4). They showed that vowel stimuli made by iterating single periods from the portions of vowels immediately adjacent to voiceless fricatives were judged by listeners to be about as nasalized as comparable periods made from vowel margins near nasal consonants. Such pseudonasalization is liable to be misinterpreted by listeners as actual nasализation and reproduced by them as such, thus leading to the sound changes mentioned above. Ohala and Amador’s study was also substantially replicated for Hindi by J. Ohala and M. Ohala (in press).

4.3. Place Assimilation by Nasals

One of the most common forms of place assimilation is found with nasal consonants taking on the place of following nonnasal consonants. Passy (1890: 183–184) remarks, “De toutes les consonnes, les nasales sont les plus sensibles à cette sorte d’assimilation [assimilation de place complète].” Hock (1986: 65) includes it among a list of “associations [that are] quite common in the languages of the world.” Although place assimilation is common enough with nonnasals, too, there are well-known instances where nasal consonants show this more than nonnasals do. An example is given in (14).

(14) Nasals assimilating in place to following nonnasals:

In the development of Latin, -NC- sequences regularly show the N assimilating in place to the following C. (In the following, starred forms represent the Proto-Indo-European parent forms; in other cases alternating forms are given, one of which indicates a historically earlier form.)

*kmēm > centum, cf. Lith. šimūnas
quam ~ quandi
septem ~ septingenti (where the first n = [ɡ])
immortālis < *en-
nomen ~ nomenclātōr (where the second n = [ɡ])
quinctus > quīntus (where the original n = [ɡ] > [n] / _i_

But, although other -CC- sequences also show changes of various sorts, many survived intact or with no change in place of the first C.

pro-petrosν, cf. Greek πτέρυξ (= ptéryξ) ‘wing’
*-hs- > -ps-, e.g., lápsus ~ lábī
*okτό[u > octō ‘eight’
dīxī, dīctus
*estōd > estō, cf. Greek ἔστω (= éstō)

(Kent 1932: 112–138)

Place assimilation and, indeed, most other forms of assimilation, are usually explained by appealing to the notion of “ease of articulation,” the assertion that it is easier to make a sequence of a homorganic nasal + consonant than a heterorganic one. But there is evidence that such assimilations owe a great deal to acoustic–auditory rather than articulatory factors. Although nasal consonants as a class
are highly distinct from other consonants, their place cues are less salient than those for comparable obstruents (Mohr & Wang, 1968; Shockey & Reddy, 1974; Singh & Black, 1966; Wang & Fillmore, 1961). On the other hand, the place cues for stop releases are generally quite distinct. Therefore, in a cluster nasal + stop the place cues for the stop (the burst and rapid transitions) will dominate the percept. This perceptual fact was demonstrated by Malécot (1956) and J. Ohala (1990b) by presenting to listeners heterorganic clusters made by cross-splicing dissimilar closure onsets and releases, including those obtained from nasal + stop sequences as well as singleton intervocalic stop. Listeners generally judged such phonetically (and artificially) heterorganic sequences as exhibiting just one place of articulation: the nasal + stop sequences were judged to be homorganic and the stop + stop sequences to be singleton stops; in both cases the single place of articulation was that cued by the second segment.

It would therefore appear that nasal + stop assimilation can come about due to the listener’s misapprehension, not only from shortcuts taken by the speaker. Examined carefully, the “ease of articulation” scenario for such place assimilations lacks credibility. Articulatory effort, the thing the speaker is supposed to reduce, presumably accumulates from the beginning to the end of words. This being the case, why should it almost invariably be the first of two consonants that changes its place, not the second, even though accumulated effort would have reached a higher level on it? Moreover, if constraints on articulatory effort are important, why should nasals be more subject to place assimilation than other manners of consonants, even though nasals would seem to be less effortful in their articulation than, say, voiceless obstruents, which, unlike nasals, require active velic elevation and glottal abduction?

5. CONCLUSION AND DISCUSSION

Our aim has been to review a variety of sound patterns typical of nasal and nasalized speech sounds and to attempt to derive them as theorems from first principles of speech production and perception—that is, to explain them. A thread running through all these sound patterns is that they owe their existence to the high degree of interconnectedness of the parts and processes of the whole speech production system due to articulatory, aerodynamic, and acoustic–auditory principles. Although the “derivations” presented are just sketches of a formal derivation, we hope that the potential for deriving facts of language from nonlanguage has been made evident.

We regard the following as the most important goals of phonology, though these are wholly neglected by mainstream schools of phonology:

1. Explaining the behavior of speech sounds in terms that are distinct from the phonological data themselves, i.e., are not simply new labels for or recordings of the data.
2. Employing explication that are familiar parts of the universe in which we live, i.e., are not ad hoc entities invented solely for the phonological problems at hand.
3. Formulating the derivations/explanation in a testable way and, if possible, actually doing the tests.

NOTES


2 The tildes above the [h] and [ʔ] in (5a) are justified in part by the kymograph tracings published in Robins (1957) and the nasal pressure recordings of Sundanese by Condas, Howard, Ikranagara, Lin, Crosetti, & Yount (1974). Robins’ instrumental records have been misinterpreted by Anderson (1972) and Vago (1988), who concluded that nasalization must somehow have leapfrogged over segments they assumed to be oral. See Ohala (1990a) and Cohn (this volume).

3 That the nasal prosody in Sundanese is a morphological process no longer directly shaped by phonetic factors is evident from the fact that the infix i- is a ‘plural marker’, does not completely block nasalization although an i- does when it is part of a stem, e.g., niʔiʔi ‘to cool oneself’, nāriʔi (ditto, pl.) vs. [maʔi] ‘to examine’. See also Cohn (this volume).

4 It is necessary to differentiate between volume velocity (how much air flows past a defined point per unit time) and particle velocity (how fast the air is moving). These are generally related, of course, but what matters in the present discussion is particle velocity.

5 The vocal tract can also be branched, in effect, any time there is coupling between a downstream resonator and an upstream resonator with the sound source in the middle. This situation occurs in some oral fricatives where the cavity behind the constriction where the friction is generated interacts with the front cavity. For the most part, such a situation has little relevance for nasal sounds. But the same configuration of resonators may occur if the vibrating vocal cords (the source) are slightly open such that the nasal and the subglottal resonators may interact. This circumstance does have indirect relevance for nasals, as described below.

6 Transcriptions of these and subsequent Indo-Aryan forms consist of conventional transliterations. Modern Hindi forms are given in IPA. Note that Modern Hindi [a] is a long vowel (phonologically paired with the short vowel [ə]) but is not overtly marked for length.

7 We reject the possible scenario that the nasal consonant was never deleted. Given the evidence of the long vowel, we assume compensatory lengthening took place, indicating that there was a nasal and that it was deleted.
The female Hindi speaker was the second author.

The phonological process of liaison in French, e.g., [bɔ̃gœ̃lɔ] ‘good boy’ but [bɔ̃nam] ‘good friend (masc.’), has been interpreted by some as evidence of an underlying word-final nasal consonant in the lexical item /bon/ which is manifested only as [bɔ̃] when uttered in isolation. But what triggers the appearance of the nasal consonant in liaison is a vowel at the start of the next word; since we juxtaposed V to a following word-initial consonant, any nasal that appears could not be due to liaison and a putative underlying nasal.

Quoted by permission of S. Karger, AG, Basel.

REFERENCES


Phonetics of Nasal Phonology


