Phonological Processes and Phonetic Rules*

Patricia Donegan
University of Hawai‘i at Mānoa, Honolulu, Hawai‘i
donegan@hawaii.edu

copyright © 2002 Patricia Jane Donegan

1. Relating phonological representations to phonetic output
In both generative and natural phonology, phonological representations and alternations have been described in terms of categorical feature values, as in Jakobson, Fant, & Halle’s (1963) original conception. This categorical representation contrasts with instrumental phonetic data, which present the speech signal as temporally, qualitatively, and quantitatively non-categorical and continuous. The question that will be addressed here is how phonetic representation (‘surface’ phonological representation) and speech are related.

Generativists and naturalists have taken two quite different views on this. The generativist view, and that of most recent writers on phonetics, has been that phonetic representation and speech are related by language-specific phonetic rules that associate binary phonological values with gradient phonetic values. The naturalist position has been that the relationship is universally determined in the act of speaking. The reason that the generative and natural views are so distinct is related to their rather distinct views of what surface phonological representation is like.

1.1 The generativist view of phonetic implementation
In the generativist view, the surface phonological representations are realized only after the application of phonetic interpretation rules, which differ from language to language (Chomsky & Halle 1968, 295-298). Both the particular gestures and timings that realize a phonological feature and the degree of coarticulation of particular features across segments are specified by the individual language, in ‘numeric’ terms. Arguing that phonetic adjustments are not entirely automatic, Keating has suggested that we consider ‘all phonetic processes, even the most low level, to be phonologized (or grammaticized) in the sense that they are cognitively represented, under explicit control by the speaker, and once-removed from (that is, not automatic consequences of) the physical speaking

* Thanks very much to David Stampe for inspiration and for almost more suggestions than I could cope with. Thanks also to Katarzyna Dziubalska-Kołaczyk for helping me to think about the questions addressed here. All errors are my own.

1 By ‘generative’ phonology, I mean not only the standard generative phonology typified by Chomsky and Halle 1968, but also its autosegmental and lexical descendants. Non-derivational models like Optimality Theory seem to view the division between phonetics and phonology in much the same way as these earlier theories (e.g. Hayes 1996, Myers 1997).
machine’ (1985:128). Kingston & Diehl (1994) concur, pointing out that ‘there actually is substantial variability between contexts, speakers, and languages in whether one articulation covaries with another in its size, and even in some cases its direction.’ This yields ‘variation of a kind different from that predicted by a model in which all the variation is a product of contingencies among articulations’ (423).²

1.2 The naturalist view of phonetic implementation

In the naturalist view, the feature and prosodic specifications of the phonological representation are interpreted through the action of speaking, in an automatic but highly contextually-determined way. The phonological representations specify combinations of features in relative time, rather like a musical score, and the vocal organs ‘interpolate’ as they move from one target or gesture to the next. This includes not only simultaneous but also sequenced combinations of feature specifications, as well as the rhythmic groupings they are part of. For example, other things being equal, labial vowels are more vs less labial if they are tense vs lax, high vs non-high, non-low vs low, plain vs palatal (e.g. [u] vs [y]), accented vs unaccented, long vs short, and they are more labial adjacent to or in the same prosodic domain as other labial sounds. Thus the number of realizations implied by the phonological representation of even a single feature value is enormous, and it is multiplied by ambient tempo, articulatory energy, and many other mental and physical variables.

1.3 A different phonology-phonetics interface

As if these differences were not enough, generativists and naturalists differ sharply on the ‘locus’ of the interface between categorical and noncategorical representation. In the generativist view, phonetic interpretation occurs only after the categorical phonological substitutions, which are also viewed as governed by language-specific phonological rules. In the naturalist view, some substitutions may indeed be governed by language-specific rules, but such rules are part of lexicon or grammar rather than of phonology, and are characterized above all by the fact that their inputs are exactly as pronounceable as their outputs.

An example is the rule governing the Latinate negative prefix in inelegant, illegal, irrelevant. While *in-legal and *in-relevant are wrong, they are no less pronounceable for a mature English speaker than illegal and irrelevant. This is a grammatical rule that applies only to certain constructions (*ul-lawful,*ur-repentant,*suddel-ly). Contrast the

---

² Either view may include the possibility that some features remain unspecified, with their values are determined by interpolation. Keating (1990) has proposed a model of coarticulation in which certain segments remain unspecified for some features, and the phonetic values are determined by the values of adjacent segments. For example, she treats vowels in English as unspecified for nasality, and says that therefore nasalization spreads part way through them from an adjacent nasal, and compares this to French, where vowels are specified [+] nasal or [−] nasal, and nasalized vowels are nasalized through almost their entire duration, and nasality spreads into non-nasalized vowels little, if at all.
phonological process that assimilates the point of articulation of a following obstruent in the phrase, e.g. \( u[m] \) pleasant, \( t[e][m] \) pennies, \( a[b]n \) mire, \( A[u, mp] \) Mary, \( s[houl, g\tilde{n}k] \) go, etc., where the unassimilated pronunciations are difficult except in quite deliberate styles. Examples like \textit{impossible} and \textit{i[n]} credible are often described as due to this assimilation. But \textit{impossible} is due to a rule, with grammatical rather than purely phonological conditions (thus \textit{im-fuckin’-possible}, \textit{il-fuckin’-legal}), while \textit{i[n]} credible is due to the process (\textit{*i[n]}-fuckin’-credible). Rules apply in terms of the phonemes of the language—[\( \eta \)] was not a phoneme of Latin, nor a phoneme of English when it adopted the Latin rule, so it was not a possible rule output—while processes apply in terms of feature values. They may govern phonemic alternations, as in the neutralization of /\( n / with /\( m / in \textit{sane/same policy}, or allophonic alternations, as in the variation of /\( n / with a labiodental in \textit{te[\( n \]} fingers or the dental in \textit{te[\( n \]} things.

1.4 Rules and processes
The phonology that authors like Keating and Kingston & Diehl refer to is generative phonology, which has focused mainly not on natural phonological processes but on rules, and in fact mostly on morphological rather than syntactic phonological rules (internal vs external sandhi). These rules typically apply only to derivative morphological or grammatical structures, and they govern only phonemic, not allophonic, alternations. They often have lexical or grammatical exceptions, and they are indifferent to real-time rhythm or tempo. Some of these same phonological rules also are said to apply ‘post-lexically’ (Kiparsky 1982, 1985, Mohanan 1986). Since postlexical rule applications are thought to share none of the above characteristics, ‘postlexical rules’ have been compared to natural phonological processes (e.g. in Mohanan 1986). Natural processes, however, are not the same as ‘postlexical rules’: natural processes are universal, and they are innate or emergent, whereas rules, however ‘phonetically plausible’ are learned or abducted from other speakers. And natural processes, unlike postlexical rules, do not govern only surface representation, but lexical representation as well, constraining the phoneme inventory, the form of lexical items, and even the possible rules of the grammar. For example, a lexical item like /\( l\ae np / is impossible in English because the assimilation process discussed above is obligatory within syllables, while an item like /\( b\ae mp / is impossible because a context-free process eliminates /\( m / in favor of /\( n / or /\( m /, so that even if [\( m / is invariant in a word like Banff, it will be perceived, mentally represented, and produced as an /\( n / or /\( m /. The same processes constrain the abduction of rules like

\footnote{The application of a natural process changes only a single feature, so it is not surprising that the syllable-final stops in such examples may be doubly-articulated, e.g. that the assimilation does not change the [+coronal] of the nasal stop in e.g. \textit{in bed} but simply adds a [+labial] articulation. The original stop is usually perceived as unassimilated until this complex articulation is simplified by a separate process that optionally eliminates the syllable-final coronality (cf. Jun 1996). The overlapping articulation can be described physiologically in terms of gestural overlap, but phonologically it is an assimilation, and though it is merely allophonic, it feeds a neutralizing process in examples like sane/same policy.}
the one governing impossible, so that the rule can govern an alternation of /n/ and /m/, but not /n/ and [m], since the latter is not a phoneme of English. (See Stampe 1973, 1987, Donegan & Stampe 1979, Donegan 1985, 1995.)

1.5 ‘Postlexical rules’ as ‘phonetic rules’
Increasingly, however, generativists have proposed to shift alternations previously viewed as part of the postlexical phonology to the ‘phonetic rules’. For example, Cohn (1993) has argued that vowel nasalization in English belongs to the phonetic component, Zsiga (1995) has claimed that English postlexical consonant palatalization is phonetic, Meyers (1999) regards the morphology-insensitive place assimilation in phrases like /m/ Baltimore as a phonetic matter of gestural overlap. Other phonologists and phoneticians also appear to accept this view. So, for the generativist, surface phonological representation, the output of the phonological rules, would be very close to lexical representation.

1.6 Are phonetic rules natural processes?
In contrast, in natural phonology, lexical representations are affected by a variety of natural phonological processes, and the surface phonological representation is the output of these processes. The question then arises whether the ‘phonetic rules’ proposed in generative phonology can be identified with natural phonological processes. If the ‘phonetic rules’ are like natural processes, then either the phonetic rules are part of the phonology (as natural processes), or perhaps what we have called the natural phonology of languages is really part of their phonetics. The name of the domain is not important; what matters is the nature of the processes (or ‘phonetic rules’).

2. Phonetic rules and natural phonological processes: Similarities
Kingston & Diehl (1994) propose a model that ‘considers phonetic implementation to be governed by constraints that determine what a speaker (or listener) can do, but not what they must do; that is, the constraints limit phonetic behavior rather than predicting it’ (423). Similarly, Keating (1988) says that phonetic rules may derive from a basic set of preferences she calls ‘default options’, but she notes that ‘a language need not include rules that reflect a particular default pattern; non-default options are often chosen instead’ (288). Keating’s ‘default options’—substitutions which are phonetically motivated but not (universally) beyond the speaker’s control and thus not merely mechanical—sound rather like natural phonological processes, and Kingston & Diehl’s constraints sound like the phonetic constraints or ‘difficulties’ that motivate processes. Let us make a comparison.

2.1 Natural processes
Natural phonological processes are phonetically motivated mental substitutions; they apply to enhance some phonetic property of an individual speech sound or to make sequences of segments easier to pronounce. Because they resolve phonetic (articulatory or perceptual) difficulties, they are universal, insofar as human speakers share similar
vocal and perceptual capabilities. But though universally motivated, a process (e.g.,
obstruent devoicing) may apply or not apply in a given language—or it may apply in
various forms or with different degrees of completeness in different languages. It can
be limited or suppressed in language-particular—or for that matter speaker-particular—
patterns, and these limitations follow universal implicational conditions regarding which
configurations are most susceptible to a given substitution. For example, other things
being equal, stops with velar closure are more susceptible to devoicing than stops with
coronal or labial closure, and coronals are more susceptible to devoicing than labials.
Additionally, stops in final position are more susceptible to devoicing than stops in initial
or medial position.

Natural phonological processes are based entirely on phonetic features. They
apply asegmentally, assimilating (spreading) or dissimilating (polarizing) features across
stretches of speech that have specific features, within specifiable prosodic domains, or at
specifiable prosodic boundaries. Processes are not ‘structure preserving’; they may
introduce or alter nondistinctive features, and their application may depend on
nondistinctive features. But neither are they always allophonic; they may cause changes
in the phonemic interpretation of the sequences they affect (synchronously as well as
diachronically); they may merge distinctions—or not. Processes are not morphologically
conditioned. Apparent cases of morphological conditioning in natural processes may be
interpretable as prosodic conditioning, where, for example, syllable structure is
influenced by morphological structure and this in turn influences the application of
processes. Unlike morphonological rules, which are always obligatory, processes may
apply variably, increasing or decreasing their application, depending on rate, style,
dialect, ‘text’ frequency or familiarity, etc.

---

4 Compare the violable constraints of Optimality Theory (Prince & Smolensky 1993). Domination of a
‘markedness’ constraint by a ‘faithfulness’ constraint is the equivalent of process suppression. Implicational
conditions would correspond to cases where a more general form of a markedness
constraint is universally ranked below a more specific form of the constraint (e.g. *[+son +voi]#
universally dominates *[+son +voi]). If a relevant faithfulness constraint (e.g. IDENT I/O [VOICE])
is ranked above the more general but below the more specific markedness constraint (e.g. *[+son +voi]#
>> IDENT I/O [VOICE] >> *[+son +voi]), this would be equivalent to limitation of an
obstruent-devoicing process that has an implicational condition: 1/ __#  (especially in final
position).

5 For example, a process like [+son] → [+nasal] / [+nasal] spreads nasality across any sequence
of sonorant material within a syllable, or within a measure. (The domain may vary with speech
style or other variables.) See Section 5.2.2.2, however.

6 For example, holy vs slowly, where diphthongization in the latter (or monophthongization in the
former) reflects not only the morpheme boundary in slowly, but also a the syllable division—
compare gail.y (where the morpheme boundary is weak/nonexistent) vs the neologism gray-ly, as
in 'she described it so gray-ly'.

5
2.2 Phonetic rules

The phonetic rules of the generative model apply, like natural processes, to sound configurations describable in terms of phonetic properties and prosodic environments. Like processes, phonetic rules are not dependent on morphological classes or boundaries, and they appear to lack lexical exceptions. Their application, like that of natural processes (though unlike that of morphonological rules), is affected by speech rate and prosody and intonation, and it may be sensitive to word frequency. Phonetic rules introduce or alter non-distinctive speech properties—in fact, in the Kingston & Diehl model, contrastiveness is the dividing line between phonological rules and phonetic ones: ‘Contrastiveness rather than language-specificity should be the criterion for assigning a phonic attribute of an utterance to the phonology’ (1994, 425).7 Although implicational conditions on phonetic rules are not, to my knowledge, mentioned in the phonetic literature, the physical, physiological, and perceptual facts that underlie the implicational conditions on processes also underlie phonetic rules, and one might presume that these facts create the same sorts of implicational conditions on phonetic rules.

It would also appear that, like natural processes, phonetic rules apply without reference to segments.8 In the ‘target and interpolation’ model sketched in Keating 1996, segments unspecified for a feature, like nasalized vowels in English, acquire phonetic properties associated with that feature by interpolation between segments that have been assigned specified targets: e.g. in a word of the form CVN, like bean, C is [-nasal], V is unspecified, and N is [+nasal], so the vowel begins as non-nasalized and ends as nasalized. Presumably a CVRN word like barn or film would be interpreted in similar fashion; since neither the vowel nor the /r/ or /l/ would be specified for nasality, the nasalization would increase gradually through the entire VR sequence.

Phonetic rules may also be compared to natural processes in that they govern allophony and certain kinds of coarticulation, and they have effects in both production and perception. For example, the unaspirated stops of Korean are voiced between vowels within words, but they are not perceived as different from initial unaspirated, phonetically voiceless stops. Keating (1996) suggests that the voicing is a purely mechanical effect: the glottal-opening gesture is shorter within words than initially, and is therefore simply obscured by the voicing of the segments on either side. According to Keating this requires a phonetic rule to specify the duration of the gesture in different positions. An

---

7 This seems to ignore the argument, most widely cited from Halle 1959 (19ff), that some phonological rules change features that are contrastive in some combinations but redundant in others.

8 Despite attempts to state rules autosegmentally, the representations of generative phonology remain segmental because they are always seen with reference to a root node. The rules of generative phonology can therefore be constrained by structure preservation. Structure preservation itself refers indirectly to underlying phonemes, although the model does not make it clear where these come from. Phonetic rules are thus distinct from morphonological rules, which, while they may be described with features, actually always manipulate segments.
alternative understanding of this, in a ‘controlled phonetics’ (as in Kingston & Diehl 1994), would be that speakers simply give up the attempt at the very short glottal-opening gesture in question and allow voicing to continue through the stop.

2.3 Effects in perception
There is considerable evidence that speakers perceive speech in terms of the phonological processes or phonetic rules of their language. They discount some of the phonetic properties that can be attributed to these rules or processes, and they depend on others. For example, English speakers perceive the vowel of *bend* [bɛnd] as the same as the vowel of *bed* [bɛd], discounting the context-sensitive nasalization of the vowel. But on the other hand, on hearing a pronunciation like [bɛt] they perceive *bent* /bɛnt/, using the nasalization of the [ɛ] to restore, in perception, a deleted nasal consonant—i.e. the /n/ whose nasality remains on the vowel but whose alveolar closure has become simultaneous with the onset of voicelessness.

I cannot do justice here to the perceptual effects of these regularities of phonetic detail, but the topic has received some attention. See, for example, Ohala 1980, Lahiri & Marslen-Wilson 1991, Ohala & Ohala 1998, and the papers in Johnson & Hume 2001, as well as Stampe 1987, and Donegan 1995, 2001. The point to be made here is that phonetic rules and phonological processes are alike in playing an important role in perception.

3. Phonetic rules and natural phonological processes: differences
Despite the similarities, phonetic rules and natural processes differ in a number of ways. The differences may be resolvable, but this remains to be seen. Phonetic rules have two important functions: 1) they govern some context-dependent feature realizations by specifying the degree and nature of different aspects of assimilation or coarticulation, or by assigning context-dependent values for features; and 2) they assign specific, presumably context-free, physical values to phonological features (or to feature combinations). Natural processes share at least part of the first function. They determine which kinds and degrees of coarticulation (and polarization) are allowed, and under what circumstances.

Processes, on the other hand, do not assign gradient physical values of features. The features of natural phonology are categorical, but they are immediately linked to specific gestures and their acoustic effects. Their specific values depend on the values of other features occurring simultaneously and sequentially in speech, and on the prosodic pattern, as interpreted in real time (including speech rate, intonation, etc.). An example of the gradience effect of simultaneous features appears when we find that, other things being equal, a low vowel is less rounded than a corresponding non-low vowel (i.e. [o] is less rounded than [o]). This is true because jaw opening and lip rounding are to some degree incompatible, so increased jaw opening weakens lip rounding. The voice onset time of stops provides another example: VOT may depend on the point of articulation. VOT increases with the backness of the stop (Smith 1978, Sock, 1981). There are many such examples, where a gradient effect in one feature depends on the values of other
simultaneous features. Prosody also creates gradient effects, as when a consonantal constriction is more extreme or has greater duration initially in a broader prosodic domain (e.g. Keating et al. 1999, Fougeron & Keating 1996), or when VOTs for voiceless stops are longer in stressed syllables than in unstressed ones. Gradience also results from adjacent feature values in the sequence of articulatory targets. For example, coronal stop releases may be slower before high vowels than before non-high (leading to affrication before high vowels, as in Japanese or Canadian French).

Natural phonology expects that the phonetic interpretations of features are variable but that they are, in principle, predictable—so that segments with the same features in the same prosodic conditions ought to have the same phonetic quality. Phonetic rules, in contrast, address the possibility of ‘sub-featural’ differences across languages.

Unlike phonetic rules, which seem rather peripheral to the phonology, functioning as last-minute adjustments in production or perhaps as interpreting redundant cues in perception, natural processes pervade the phonology. The application or suppression of natural processes creates the phoneme inventory and determines the distinctiveness or predictability of each phonetic feature for the language, thus affecting the structure of lexical items. For example, the application of a context-free fortitive process that delabializes low vowels ensures that non-labiality is redundant in low vowels, and that /a/ and /o/ are not distinct phonemes; the suppression of this process allows a distinction. Lenitive processes affect the inventory as well. A first-language learner who encounters a phonetic nasalized vowel has to accept this vowel at face value (i.e., as intentionally nasalized) unless the nasality can be attributed to an adjacent nasal. The latter is possible if the language allows assimilative nasalization to apply, but it is not if nasalization is inhibited elsewhere, or if there is no adjacent nasal consonant.

Both phonemic status and allophonic existence are the result of natural processes: allophonic variants exist because of the application of context-sensitive processes. Phonemes exist because of the interaction of processes: 1) some simultaneous feature combinations are ruled out by context-free processes, and 2) some feature combinations can be attributed to the application of context-sensitive processes, but 3) some feature combinations must be perceived as the intentions of the speaker—and thus as the elements of words. (See Stampe 1987, Donegan & Stampe 1979, Donegan 1985, 1995 for further discussion.)

It would take considerable reinterpretation of the concept of phonetic rules to allow them to assume the function of determining language-specific inventories (of phonemes, or of distinctions). Studies of the effects of coarticulation on perception usually take the distinctions made by the language under study at face value—they do not account for the existence of the distinctions. Phonetic rules seem to play no role in the phonology itself, although there is the possibility that the phonology may be ‘reorganized’ diachronically because of the effects of phonetic rules. Kingston & Diehl (1994, 424) give the example of Sindhi, where formerly-geminate voiced stops were pronounced with implosion, and implosion subsequently became the contrastive feature characterizing these stops.
Another important difference between natural processes and phonetic rules is that processes are assumed to have categorical outputs, while phonetic rules ‘assign numeric values’ and have gradient outputs. This is perhaps the principal difference, and it will be treated in Section 5.

4. Universal or language-specific?
Phonetic rules, since they govern phonetic details, are spoken of as language-specific (Kingston & Diehl 1994), while natural processes, since they are responses to phonetic difficulties, are conceived of as universal. But the universality of processes does not mean that they apply in all languages—only that they are motivated in all speakers. In the Sindhi example just noted, the expansion of the oropharyngeal cavity associated with implosion is a response to the aerodynamic difficulty of maintaining voicing in a geminate stop. In natural phonology, the phonetic property of implosion is regarded as a feature, and a phonological process may assign that feature to geminate voiced stops: voiced stops, especially if they are long, become implosive. Speakers could pronounce the voiced stops—presumably, speakers at some stage in the history of the language did so—but if they allow the process to apply, they implode them instead of (or in addition to) voicing them. (If learners of the language assume that this implosion is a contrastive feature, it becomes one.) Alternative responses to the difficulty (i.e. alternative processes) exist, of course. Japanese, for example, has some English borrowings, like /baggu/ ‘bag’ and /beddo/ ‘bed, bedroom’ with geminate voiced stops; these may sometimes be pronounced in Japanese as long voiced stops, but if they are not, they are devoiced rather than imploded (Vance 1987). In another language, such stops could be degeminated. The uniqueness of each language’s natural phonology lies in the selection of the set of processes that it allows to apply—and, correspondingly, the set of difficulties that it requires its speakers to master. Insofar as the phonetic rules of individual languages are also selected from a limited range of phonetically motivated possibilities, phonetic rules are universal in just the same way that phonological processes are. (This ignores for the moment the question of categorical vs gradient outputs.)

Kingston & Diehl (1994) argue for a ‘controlled phonetics’ as opposed to an ‘automatic phonetics’. Consider their treatment of devoicing of initial stops. In their ‘automatic phonetics’, for example, the glottal gesture for voicing in a language like English would be the same for both initial and intervocalic stops—that is, there would be a gesture of glottal ‘closure’ during the stop closure. The devoicing of initial voiced stops would depend entirely on the air pressure required to start the vocal folds vibrating while there is an oral closure. Devoicing would occur in initial position because initiating vocal fold vibration requires higher air flow than keeping the folds vibrating between vowels requires.9 (Voicing of initial stops would be assured only with special adjustments to expand the oral cavity.) In contrast, in Kingston & Diehl’s ‘controlled

9 In English, some initial stops are voiced during part of their closure, but are still phonologically voiced. Perhaps the special adjustments are not always effectively made. Still, the attempt to produce voicing is there.
phonetics’, the speaker allows the glottis to remain open until the stop release, when air pressure conditions are right for voicing without other adjustments. As they put it, ‘The attempt to produce voicing during closure is deliberately abandoned’ (429). This abandonment of the attempt to produce voicing must take place in the central nervous system, where speech is planned and organized—i.e. in the mind. This is why natural phonologists would refer to this event as the application of a natural phonological process—the speaker substitutes an alternative phonetic target, allowing the devoicing to occur.

Learners of some languages may find that it is acceptable to abandon or delay voicing in these circumstances. The process applies regularly and becomes part of their native ‘accent’. Other languages require speakers to make the adjustments that ensure sufficient, or full voicing\(^\text{10}\) during the closure. In the idiom of natural phonology, speakers of the latter sort of language have to overcome or suppress the natural process of devoicing.

The phonologist describes this as \([\text{obstruent}] \rightarrow \text{[-voice]}\); the phonetician says that the onset of voicing is delayed. It hardly matters whether we call this phonetic control or phonological processing. The former designation emphasizes the articulatory and perceptual motivations for process application; the latter emphasizes the psychological status of these adjustments (or substitutions). If phoneticians and phonologists can agree that these adjustments are mental processes that apply for physical reasons, they can go on to the more difficult and more interesting questions.

Similarly, the designation ‘universal’ versus ‘language-specific’ is in some ways a matter of emphasis. Phonological processes are universal in motivation and in form, but they do not apply universally, and when a process is allowed to apply, a selected subset of its universal implicational conditions may determine its application, and this subset is not the same in every language. Phonetic rules are also based on universal constraints and may represent ‘default options’. The extent to which they must specify gradient properties that differ in degree from one language to another—and must thus be language-specific in a way that processes are not—is not clear.

5. Gradience and categoriality
The speech signal and the articulations that produce it are undeniably continuous, and the continuity results in gradation into and out of target positions.\(^\text{11}\) The question of gradience is about whether phonological processes or phonetic rules specify features (or targets or gestures) categorically or in some gradient fashion. Many articulatory gestures (such as glottal opening) may occur in varying degrees, and independent articulations, such as glottal ‘closure’ for voicing or velic opening for nasalization or articulator contact

\(^{10}\) Other languages allow speakers to exaggerate voicing: Pre-nasalization or post-nasalization of stops may arise in an attempt to ensure or even enhance voicing during closure, as may implosion or special laryngeal qualities. E.g. final voiced stops are post-nasalized in Mundari (Osada 1992) and Temiar (Benjamin 1976). This is an alternative to devoicing.

\(^{11}\) This is equally true if we speak in terms of gestures rather than target positions.
for stop closure, may occur with varying degrees of simultaneity. There is no reason to suppose that a contrast of aspiration in stops, for example, should be manifested in the same way in all contexts within a language, or that a voicing contrast in stops should always be represented in the phonetic signal by stops that are voiced throughout their closure duration versus stops where voicing ceases precisely at the time of closure and begins again precisely at the time of release. Different laryngeal feature specifications may be introduced and may interact to influence the phonetic output (e.g. in English, the contrast is arguably one of voicing, but aspiration is introduced in some environments and this influences the phonetic result). The question is whether these specifications below the level of contrast may be categorical, or whether they must be gradient to account for the data.

Gradient effects do not always require gradient specifications. For example, although speakers’ categorizations of stops with respect to VOT seem consistent, we need not suppose that an aspiration contrast will be phonetically the same in all languages or in all contexts within a language, *even when the stops have the same laryngeal feature specifications*. Within a language, different prosodic contexts, different points of articulation, and different adjacent segments may affect the onset of voicing (Westbury & Keating 1986, Keating et al. 1983). In different languages, different prosodic patterns (for example, stress in English vs pitch accent in Japanese) may also affect the physical instantiation of laryngeal contrasts. These prosodic and segmental factors may produce gradient effects, like devoicing of more vs less of the duration of the stop closure, short VOT lag vs no VOT lag, differences of burst strength, etc., even when the laryngeal features are the same.

Two kinds of phonetic rules have been proposed. Both are said to require specification of values more specific that the categorial values associated with phonological features. One kind of phonetic rule assigns context-free language-specific phonetic values in the implementation of lexical feature specifications. The other specifies phonetic values in particular segmental or prosodic contexts. Assuming that the first set of rules apply context-freely, this second set would presumably substitute context-sensitive phonetic specifications for the original context-free ones.

### 5.1 Assignment of language-specific phonetic values

Phonetic rules are said to assign language-specific phonetic values to phonological features which, in generative phonology, have rather abstract or general phonetic correlates. The claim is that segments with the same feature specifications are realized differently in different languages, and that phonetic rules must specify articulatory-acoustic values more finely than the level of categorical features. Phonological processes, on the other hand, refer to features that are more directly phonetic, and they are thought to require no such language-specific interpretation rules. We should look at some cases, then, where different phonetic values are said to be the manifestations of the same combination of features.
5.1.1. Vowels with the same features

In an effort to compare universal features in vowels with their language-specific instantiations in a number of Germanic languages, Disner (1983) characterized language-specific differences among vowels that are transcribed with the same phonetic symbol and which presumably have the same phonological feature values. Her claim was, basically, that /i/ in one language (for example) may not be equivalent to /i/ in another. Disner found significant differences in formant values between some pairs of ‘shared vowels’ in pairs of languages, even when the pairs of languages were spoken by the same speaker.\footnote{In an attempt to eliminate differences due to individual vocal tracts, Disner took pains to find balanced bilingual speakers and to have the nativeness of their pronunciation in each of their languages evaluated by native speakers.}

But Disner’s study does not establish sufficiently that language-specific formant values (or ranges) have to be specified by phonetic rules. There are factors that Disner does not seem to have considered. One is the possibility of overly-standardized transcription: /i/ in a given language may not really be appropriately transcribed as /i/. An example of this might be the transcription of Tausug, an Austronesian language that Disner used to exemplify a small vowel system. Tausug is supposed to have the system /i, u, a/, but the formant values Disner gives for this language suggest that the vowel system might really better be described as /i, u, a/ or even /e, o, a/. One suspects that the /i/ and /u/ symbols may have been chosen to represent the Tausug phonemes because /i/ and /u/ are supposedly the ‘unmarked’ high vowels—not because they represent the actual phonetic quality.

A second possibility that must be considered before one can establish the need for specification of ‘numeric’ or ‘sub-featural’ phonetic values is that of additional features. Generative phonology has tended to refer only to a relatively small set of relatively abstract features that are used contrastively in languages, but there are clearly many articulations that speakers can learn to control (and that are associated with perceptual qualities that hearers can learn to perceive) that are part of speech processing. These articulations are what phonetic rules presumably specify; if we call them features, it is only the question of gradient vs categorical specification that distinguishes them from the features of natural phonology. An unanswered question is whether the apparently sub-featural differences Disner found could be specified categorically with an expanded set of features.

For example, Disner found that an important difference between the vowels of English and German is ‘peripherality’, associated with the fact that German labial vowels are more rounded than English labial vowels (and German non-labial vowels are more spread). German /e/ was considerably higher (lower F1) and somewhat more fronted (higher F2) than English /e/ (Fig 4.4, 88), and German /i/ and /u/ were somewhat higher (lower F1) than English /i/ and /u/ (Fig. 3.2, p. 38). German /u/ also showed lower F2 than English /u/ and German /o/ showed lower F2 than English /o/ (see Disner’s Figs. 3.2,
p. 38 and Fig. 4.4, p. 88). The stronger rounding in German might be thought to require ‘numeric’ phonetic specification, but if the phonological/phonetic feature set allows a distinction of ‘tense’ (extreme) versus ‘lax’ (weak) labiality, the different degrees of rounding are not incompatible with categorical specification.

Another difficulty is that Disner’s measurements do not appear to take diphthongization into account.13 The English-German differences noted above might also result from the measurement of diphthongized vowels as if they were monophthongs. English vowels are more susceptible to diphthongization than are those of German. Especially in English /i, e, u/, the first element may be lax, and in /u, o/, the first element may be laxed or delabialized. American English /u/ is rarely a real [u]; it is more often [ʊu] or [iʊ], and /o/ is usually [ɔu] or [aʊ]. We also have [ɪj] and [ɛj] for /i/ and /e/. Since the vowels were measured at the beginning of their steady state, such diphthongization, even if it did not always occur, would affect the collective values of the vowel formants, even for a single speaker. Differences due to diphthongization can be described with categorical features.

Many apparent language-specific vowel differences may perhaps be accounted for by feature differences—either by allowing for additional features or by taking diphthongization into consideration. Although it is possible that Disner has characterized some truly gradient language-specific differences, it is notable that the language-specific descriptions that inspired her work (statements like “[ɛ] is more open in Dutch than in German”, or Swedish [i] is “closer than the vowel in English ‘seen’”—see Disner’s Ch. 1) are not often confirmed by significant differences in her acoustic data. And the specification of language-particular phonetic values is further complicated by the fact that such ‘language-particular’ values are based on averages over groups of speakers, although her study of bilinguals’ vowels attempts to avoid this problem.

There is also the fact that speakers of some languages can have remarkably different vowel realizations and still be accepted as ‘native speakers’. (Consider examples given in Wells 1982, or Labov 1994, for English, where even quite similar dialects show quite transcribable differences.) There is also the question of whether speakers can actually identify other speakers as ‘native’ or ‘foreign’ on the basis of such ‘sub-featural’ vowel quality differences.

While one cannot rule out the possibility that these sub-featural differences exist, and that phonetic rules may be required to specify them, it is difficult to establish that such rules are necessary to specify ‘context-free’ vowel qualities below the level of a sufficient set of phonological features.

5.1.2 Language-specific manifestations of [voice]

Cross-linguistic differences in voice-onset time are sometimes cited as an example of language-specific differences in the phonetic implementation of the [voice] contrast.

13 Disner does note that most of the languages considered include diphthongs. In some cases, she even notes that the vowels she is measuring are usually somewhat diphthongized, although she measured formant values at only one point for each vowel in the study.
Kingston & Diehl (1994) point out that the voicing contrast in English, Swedish, and German is manifested initially as a contrast between voiceless aspirated stops vs unaspirated (or occasionally voiced) stops; posttonically, the contrast is between voiceless unaspirated and voiced stops. In Icelandic, the contrast is manifested as voiceless aspirated vs unaspirated. In Dutch, the contrast is between voiceless unaspirated vs voiced stops.

Kingston & Diehl argue that these are phonetic manifestations of the same feature, and that ‘[+voice]’ can be applied crosslinguistically to the stop series in which voicing begins earlier and ‘[-voice]’ to the one where it begins later’ (427).\textsuperscript{14} Natural phonologists would argue for more phonetically precise definitions of features, and would claim that the adjustments that aspirate ‘[-voice]’ segments or devoice ‘[+voice]’ segments assign additional features. Kingston & Diehl note that F\textsubscript{0} depression occurs next to ‘[+voice]’ stops, regardless of their phonetic realization as pre-voiced or short-lag, and that it only occurs next to contrastively voiced stops, regardless of actual voicing. This enhancement effect may indeed depend on the lexical specification of the stops, as do other enhancement effects like closure duration and preceding-vowel duration, but the rules that effect these different phonetic manifestations may be manipulating categorical features. In fact, the similarities between English, Swedish, and German in this respect suggest categorical processing. In order to argue for sub-featural differences, one would have to establish consistent (and perceptible) differences between, say, the aspirated stops of English and those of German or Swedish.

One might object that something like ‘lowered F\textsubscript{0}’ should not be treated as a feature, but if we consider the role that lowered F\textsubscript{0} plays in the development of tonal systems, we cannot dismiss this possibility so readily. In tonogenesis, lowered F\textsubscript{0} seems at first to be an enhancement feature, but later it becomes a distinctive one (Hombert et al. 1979).

5.2 Substitution rules with gradient results
There is increasing evidence from phoneticians that, as Whalen says, coarticulation is planned. (See, for example, Whalen 1990.) Speakers do not simply line up a sequence of phonemic targets and allow the articulators to get from one to another as best they can; instead, the activity of articulation is centrally planned, so that features spread (or gestures overlap) in regular ways. This planning differs from language to language, and it requires either the application of language-specific phonetic rules—or phonological processes, which are inhibited in different ways in different languages. The gradience of articulatory movements and of the acoustic signal have led most phoneticians to the conclusion that phonetic rules specify phonetic outputs that refer to differences smaller than feature categories (see above) and to durations that do not correspond to linguistic units like prosodic domains, feature combinations, or segments. An attempt to specify

\textsuperscript{14} It is not clear whether, if this were accepted as a definition of voicing, it would imply that Korean, which contrasts aspirated vs plain vs tense or glottalized stops, the ‘tense’ stops would be considered ‘[+voice]’ because they have the shortest voice onset time.
the linguistic processing of speech in terms of linguistic units, as phonological processes, raises some difficult questions. This section will examine a few examples.

5.2.1 Gradient retraction of English /l/
Sproat and Fujimura (1993), in a study of the articulation of /l/ in English, argued against descriptions of English /l/ as having two allophones, one syllable-initial and ‘light’ and the other syllable final and ‘dark’. They found, instead, that ‘the single phonological entity /l/ is phonetically implemented as a lighter or darker variant depending upon such factors as the /l/’s position within the syllable, and the phonetic duration of the prosodic context containing the /l/’ (291). They did not find a categorical difference between light and dark /l/s, but instead found that some of the syllable-final /l/s were as light as or lighter than the syllable-initial /l/s.¹⁵ Sproat and Fujimura determined that the darkness of English /l/ is a gradient quality that covaries with the strength of the following boundary and the corresponding duration of the rhyme in which the /l/ appears, and they concluded that ‘there is no reason to treat the light and dark allophones as categorically distinct phonological (or phonetic) entities in English’ (291).

Sproat and Fujimura offer an interesting account of why dark [h]’s tend to occur in syllable-final position and light [l]’s in syllable-initial position: /l/ is of course [+lateral], and laterality requires a drawing in of the sides of the tongue. Since the mass of the tongue is constant, this narrowing lengthens the tongue, and this lengthening involves two potential gestures: one is advancement of the tongue tip and the other is retraction and lowering of the dorsum. Sproat and Fujimura found that darker [h]’s are characterized by a greater retraction and lowering of the tongue dorsum and by a relatively earlier timing of the dorsal retraction, relative to the apical advancement. They propose that the advancement gesture is consonantal, since it produces an extreme obstruction in the mid-sagittal plane, and the retraction gesture is vocalic, since it does not produce an extreme mid-sagittal obstruction (304). They claim that since consonantal gestures tend to be stronger (having greater displacement) in syllable-initial position, and vocalic gestures tend to be stronger in syllable-final position, this would account for the stronger apical gesture syllable-initially and the stronger dorsal gesture syllable-finally. Regarding the timing difference, they say that consonantal gestures ‘are attracted to syllable margins’ and vocalic gestures ‘are attracted to syllable nuclei’ (306). Therefore the apical gesture is earlier syllable-initially and later syllable-finally, and the reverse is true for the dorsal gesture. Sproat and Fujimura point out that in other languages where light/dark variation of /l/ occurs, e.g. Dutch and Portuguese, the variation is in the same direction as in English. This speaks in favor of their account of the phonetic motivation of the variation.

But where Sproat and Fujimura’s account falls short is in explaining why not all languages that have /l/ in both syllable-initial and syllable-final position show such

---

¹⁵ All /l/s were elicited between high front vowels. ‘Darkness’ was determined by several measurements: F1 – F2, Dorsum Retraction Extremum, Mid Lowering Extremum, and Tip Delay (a positive value characterized darkness, a negative one, lightness).
variation. They note that German and some dialects of English do not. They suggest that ‘German /l/s involve a dorsal retraction gesture as in English, but that for some reason the articulatory differences between light and dark /l/s … are simply not as extreme as in English’ or that ‘German /l/s, although they are [+lateral], do not involve retraction of the tongue dorsum’ (310). But whichever alternative turns out to be correct, the crucial fact is that German differs from English by not allowing retraction of the tongue dorsum in syllable-final /l/s—a difference between the sound system of English and that of German. Whether we call L-retraction a phonological process or a phonetic rule, it must be under the control of speakers, because English speakers do it and German speakers do not. This is why phonologists are willing to say that English shows two different allophones of /l/, and German does not.

What of the gradience found in the darkness characteristics of /l/? Could a ‘categorical’ phonological process account for the fact that there are not two neatly-divided clusters of phonetic characteristics—that there is no bimodal distribution in the measurements? Sproat and Fujimura point out that the degree of darkness increases with the strength of the following boundary and the corresponding duration of the rhyme in which the /l/ appears. It seems that some /l/s before weak boundaries (in rhymes of short duration) do not darken at all. Other /l/s (in longer rhymes) are darkened, but the degree of darkness depends on the rhyme duration. A categorical process,

(1) L-retraction: [+lateral] \( \rightarrow \) [+retracted] in the rhyme of a syllable,

which applies in English but not in German, might be expected to give less gradient results.

There are two factors that introduce physical gradience into the categorical picture of (mental) speech processing via natural phonological processes. First, a phonological process may be subject to various implicational conditions which determine its applicability, making it obligatory in some circumstances and optional or inapplicable in others, depending on the language or dialect. The process is universal, but the particular conditions that are relevant are ‘selected’ by the language.

The application of L-retraction clearly depends in part on the duration of the syllable-rhyme, so the process may be better characterized as:

(2) L-retraction: [+lateral] \( \rightarrow \) [+retracted] in the rhyme of a syllable, esp. when longer.

This additional condition, which we would abbreviate \( !\text{longer} \), is implicational. It means that, other things being equal, if a shorter lateral is retracted, a longer lateral will be retracted as well. It is thus possible that in most English dialects where L-retraction applies, it applies obligatorily to /l/s in longer rhymes, but only optionally to shorter /l/s. To the shortest /l/s, it may not apply at all. So the /l/s in the shortest rhymes, those where the darkness values are like those of syllable-initial /l/s, may not, in fact, be retracted.

Second, the physical interpretation of retraction depends on the rhyme duration. The greater the duration, the greater the articulatory and acoustic effects of retraction.
Duration is gradient because it depends on a wide range of variables, which can be universal (e.g. intrinsic length of segments), language-determined (here illustrated by boundary strength), and even non-linguistic (speech rate). Thus, given specified rhyme durations, the degree of L-retraction could be entirely predictable or ‘mechanical’, although the fact of retraction vs non-retraction is not.

5.2.2. Partial assimilation

More difficult is the question of partial assimilation. A number of studies suggesting that assimilation processes are ‘phonetic’ show that phonetic outputs resulting from assimilative phonological processes or phonetic rules are not the same as the phonetic outputs that result from lexical specifications with the same features.

5.2.2.1. ‘S-palatalization’

For example, a study by Zsiga (1995) shows that palatalization of English /s/ before a high palatal glide has different results when it applies within words, as a morphonological rule, as in *confession*, and when it applies across word boundaries, as in *confess your*. The first application produces a lexical form /kɒnˈfeʃən/. The /ʃ/ of this lexical form, though morphologically related to the /s/ of *confess*, is identical to underlying /ʃ/ as in *fresh*. In natural phonology, this is expected, because /ʃ/ is the speaker’s target. Although the speaker may be aware of the relationship *confess/confession*, the word is lexically /kɒnˈfeʃən/, and there is no sense in which the articulatory target of *confession* includes the sequence [ʃj].

In a sequence like *confess your*, however, the sequence [ʃj] is the speaker’s articulatory target at some point in the processing. If asked to repeat slowly, or even syllable by syllable, speakers will say an [s] in *confess your*, but not in *confession*. Zsiga finds that the [ʃj] target is altered only partially, producing something that might be transcribed [ʃʃ].

It is not particularly surprising that there should be a difference between the output of an obligatory morphonological rule and that of an optional phonetic one. The different palatalizations clearly represent the segmental nature of morphonological rules on the one hand, and the asegmental nature of the phonetic rules or processes, on the other. If we identify the phonetic rule with an asegmental phonological process of palatalization, describing the substitution as

\[
(3) \; [+\text{coronal}] \rightarrow [-\text{anterior}] / \_\_ [+\text{coronal} \; -\text{anterior}],
\]

then, when it applies, we would have to expect the output of the process to be the same as that of the rule. Since it is not, the problem arises how to describe this partial assimilation. If this assimilation affects only part of the /s/, natural phonology cannot...
easily explain the facts without reference to some prosodic boundary which limits the assimilation.\(^{17}\)

Zsiga does not consider that the correct description may involve assimilation of the /\j/ to the /\s/, rather than assimilation of the /\s/ to the /\j/; this would leave the /\s/ intact. Evidence in favor of this analysis comes from the parallel American English palatalization of /\t, \d/, which become affricates [t\t, d\d] before /\j/ in phrases like met your and had your. In such phrases, the /\t/ retains its alveolar closure, so we might conclude it is the /\j/ that becomes sibilant, and the rule might be better stated as:

\[
(4) \{+\text{consonantal}, +\text{palatal}\} \rightarrow \{+\text{sibilant}\} / \{+\text{coronal –sonorant}\}.
\]

Of course, if careful phonetic study shows that only the beginning of the /\j/ is assimilated, the problem of partial assimilation remains.

5.2.2.2. Vowel nasalization

In a similar vein, Cohn’s studies (1990, 1993) of nasalized vowels in English and French\(^ {18}\) suggest that the English assimilation of nasality does not result in the same kind of nasalized vowels as the French contrastive specification of nasality does. Cohn found that the nasalization of vowels in French C\V\C,\(^ {19}\) where vowel nasalization is lexically specified, begins immediately after the release of the initial obstruent and extends through the entire vowel. In English, where vowel nasalization is allophonic and depends on the adjacent consonant, Cohn found a different pattern. The vowel nasalization of English increases gradually during the vowel in CVN sequences, and it decreases gradually during the vowel in NVC sequences. French non-nasalized vowels adjacent to nasals seem different still; they are nasalized only briefly—just adjacent to the nasal consonant.

If the nasalization of vowels\(^ {20}\) that takes place in English is represented in terms of features, and if the phonological process simply states that sonorants are nasalized adjacent to nasals (within the same syllable), then the [+nasal] [\u00e0] of English pond ought to be nasalized in the same way as the [\u00e0] of French vendre.

---

\(^{17}\) One might propose that the /\s/ is ambisyllabic and that the assimilation of palatality is limited to apply within the syllable, so that it affects only the syllable-initial part of the /\s/, which is in the same syllable as the /\j/. But evidence for segments that are ambisyllabic, as opposed to syllable-final, is sparse.

\(^{18}\) Cohn also studied nasalized vowels in Sundanese. Since Sundanese nasalization is a morphophonological (lexical) rule (Cohn 1990:53), the vowels are nasalized in the lexical representation, and their nasalization is categorical, like that of the underlying nasalized vowels of French.

\(^{19}\) \(C\) is a non-nasal obstruent, \(V\) is a vowel, \(\bar{V}\) is a nasalized vowel, and \(N\) is a nasal consonant.

\(^{20}\) In fact, it is sonorants that are nasalized. Nasalization spreads through sonorant consonants in words like barn and palm, as well as into vowels that are immediately adjacent to nasals.
5.2.3. Stop-insertion

Fourakis and Port (1986) studied insertion of a stop between nasal and fricative (/ns/ and /nz/) or between lateral and fricative (/ls/ and /lz/). They found that American speakers inserted stops 100% of the time in /ns/ and /ls/ clusters (i.e. when the fricative was voiceless). They occasionally inserted stops (2%–10% of the time) before the voiced fricatives in /nz/ and /lz/ clusters, and they usually deleted the stop in /ndz/ clusters, but not in /ldz/ clusters. South African speakers neither inserted nor deleted stops in these clusters. The difference between the American and South African speakers makes it clear that stop insertion is not an inevitable effect of producing nasal-fricative or lateral-fricative sequences, but rather a controllable difference in the timing of the required articulations. While they recognize its phonetic origins, Fourakis and Port regard this rule as ‘language-specific’, since it does not apply in all languages. (Fourakis and Port do not seem to consider the possibility that processes may be universal, but might be limited or suppressed in some languages.21)

Since stop epenthesis is not an unavoidable phonetic effect, Fourakis and Port consider its potential status as a phonological rule. They say that a description would have to involve 3 rules: 1) an obligatory epenthesis rule that inserts a voiceless stop between a sonorant and a voiceless fricative, 2) an optional epenthesis rule that inserts a voiced stop between a sonorant and a voiced fricative, and 3) a rule that deletes a voiced stop after a nasal and before a fricative in a syllable coda.22 The main problem Fourakis and Port encounter in regarding stop epenthesis as a phonological rule, however, is that the result of stop epenthesis is not phonetically identical to a lexical sequence of nasal-stop-fricative (or lateral-stop-fricative). Although speakers perceive tense and tents as homophonous, there are phonetic differences between the two words, so that the pair constitutes an instance of what they call ‘incomplete neutralization’.

Fourakis and Port claim that categorical phonological rules would predict that the nasal-stop-fricative that results from epenthesis should be identical to a lexical nasal-stop-fricative sequence. They propose a set of ‘phase rules’ to account for many phenomena that have been described as phonological substitutions—but which in more recent phonetic literature have been viewed as phonetic rules. These ‘phase rules’ govern

---

21 Their mention of Donegan & Stampe 1979 suggests that they think that natural phonology would regard stop insertion as ‘inevitable’ if it results from universal constraints. D & S mention stop insertion among the universal processes, but like all processes, it is suppressible.

22 For independent reasons, English has to have a process like 3), that optionally deletes a voiced stop after a nasal in a syllable coda, especially when the stop is unreleased: phrases like send for flowers, send some flowers, send me flowers become [sen fr flau r z], [sen s m flau r z], and [sen mi flau r z] (not to mention [sen fr flau r z], [sem mi flau r z]) in casual speech, and ten flowers and tend flowers seem as homophonous as tens and tends do. This is actually an assimilation process, where the [d] assimilates to the nasality of the [n] within a syllable-fall (or coda). And the relationship between 1) and 2) is consistent with their being the same process but an implicational condition that reflects stronger motivation for epenthesis when the stop is voiceless.
local articulatory gestures, including their timing. They are controlled in part by phonological context, in terms of features, and ‘they are very sensitive to pragmatic communication needs that change rapidly from moment to moment during speech’ (218), so that speakers constantly adjust their degree of applicability, which could account for fairly continuous control over parameters like voice-onset time, flapping, devoicing, etc. Fourakis and Port say that the inputs to these ‘phase rules’ may be specified in categorical features (but in terms of the necessarily gradient durations of real time speech), and their outputs are gradient. Fourakis and Port come very close to identifying these ‘phase rules’ with phonetic implementation rules, or with post-lexical phonological rules. We might compare phase rules, then, as we are comparing other phonetic rules, to phonological processes.

Like phase rules, phonological processes are described with categorical specifications of their inputs. The outputs of processes, however, are thought to be categorical, with the gradience of the phonetic signal resulting from the physical implementation of their outputs. But Fourakis and Port say that phase rules would specify gradient output values. However, they do not offer any examples of the form phase rules would take. For example, would the phase rule governing final devoicing specify a percentage of the final obstruent duration to be devoiced? or a range of percentages? Or would the rule simply allow devoicing to affect the final stops ‘as much as necessary’, given the timing of other gestures, the rate of speech, the energy being devoted to speaking, etc.? In stop epenthesis, would these rules specify a degree of lengthening of the [–continuant] closure, or simply require that the closure be maintained past the point at which the velar closure is made and the voicing ceases? Such issues must be examined.

Fourakis and Port also specify that phase rules like stop insertion or final devoicing are learned, because some languages have them and others do not. But the learnability of such rules is problematic. If speakers are not able, for example, to perceive the difference between a devoiced obstruent and a lexically-specified voiceless obstruent, how do they learn to devoice to a particular degree?

5.2.4 Incomplete neutralization
We might question whether the outputs of phonetic rules—or of processes—have to be gradiently specified to account for incomplete neutralization. Like the work of Fourakis and Port, a number of studies (reviewed in Dinnsen 1985) have suggested that certain ‘neutralization’ processes result in phonetic forms that are not identical to phonetic forms of their supposed lexical equivalents. For example, in languages with final devoicing, like German, Polish, and Catalan, words with phonologically voiced final obstruents appear to have phonetic forms that are not identical to those with phonologically voiceless finals. Dinnsen’s review of incomplete neutralization ends with the claim that speakers consistently produce differences that they cannot perceive. A corollary is that the production grammar must be separate from the perception grammar. As noted above, it is not clear how speakers could learn a production grammar that requires them to produce differences that are not specified in the perception grammar—i.e., differences
they cannot perceive.\textsuperscript{23} And there is considerable evidence that speakers ‘undo’ the effects of phonological processes and coarticulation in perception, which argues for a single grammar for production and perception.

It remains controversial whether complete neutralization is the correct description of the phenomena in question (see Manaster-Ramer 1996 and Port 1996). Regardless of how this question is resolved, there needs to be an account of the intermediate voicing values that appear in studies that are cited in the literature. One could propose a phonetic rule (or, for Fourakis and Port, a ‘phasing rule’) that results in devoicing of part of the obstruent. This is problematic if the phonetic change is perceptually categorical. Instead, we might propose that a devoicing process (an abandoning of the attempt to voice final obstruents) applies at least some of the time in languages with such devoicing—but that speakers, for orthographic or other reasons, may attempt to suppress the process and preserve the phonological voicing contrast\textsuperscript{24}, and that this results in incomplete devoicing. When speakers of devoicing languages do not abandon the attempt to voice but instead, make the effort to voice, they may be unable to make sufficient articulatory adjustments to produce truly voiced outputs, and end up producing partly-voiced obstruents that are phonetically different from their lexically voiceless ones.

\textbf{5.2.5. Gradient results of process non-application}

In a 1996 paper, Hayes proposed that phonological constraints, including universal, phonetically grounded constraints, must be constructed by the child, because phonological constraints (or the substitutions they are responsible for) apply to classes of sounds that share features, not to classes of sounds defined purely on the basis of phonetic difficulty. According to Hayes, a real phonetic constraint might rule out, for example, any voiced stop that characteristically poses more than a specified degree of difficulty. Since there are a variety of sources of difficulty (e.g., for obstruent voicing: place of articulation, post-obstruent position, post-nasal position, and initial position create difficulty) constraints based entirely on phonetic difficulty might have to rule out complicated sets of segments (e.g. all post-obstruent voiced stops, \([d, g]\) in initial position, and \([g]\) after oral sonorants). In Hayes’ proposal, phonological constraints must be abstracted away from their phonetic motivation; he proposes a constraint-creation

\textsuperscript{23} Dinnsen notes that, in L2 phonology, production of some contrasts may be better than perception. But this is not real evidence for the independence of a production grammar from a perception grammar, since L2 learners usually receive instruction of some sort in making non-native distinctions. Almost any English speaker can produce dental vs retroflex distinction if told where to place the tongue, but learning to perceive and remember the difference is much more difficult.

\textsuperscript{24} In all of these studies, the speakers are speaking carefully, and the target word is pronounced alone, or in a frame sentence designed to place it at the intonational peak.
algorithm that works by comparison of two phonetic categories that share all features but one.  

Hayes claims that phonology is categorical and phonetics is gradient, and he offers an example of the phonology-phonetics difference. He notes that in Ecuadorian Quechua ‘at suffix boundaries, it is phonologically illegal for a voiceless stop to follow a nasal, and voiced stops are substituted for voiceless; thus sača-pi ‘jungle-loc.’ but atam-bi ‘frog-loc.’ He contrasts this with English, where a post-nasal voicing distinction is maintained phonetically (as in simple vs symbol, or camper vs camber), but where from 13% to 60% of the duration of the stop may actually be voiced.

Hayes claims that ‘in English we see post-nasal voicing “in the raw,” as a pure phonetic effect, whereas in Ecuadorian Quechua the phonology treats it as a categorical phenomenon.’ He continues, ‘The Quechua case is what needs additional treatment: it is a kind of leap from simply allowing a phonetic effect to influence the quantitative outcomes (in a variable way) to arranging the phonology so that an entire contrast is wiped out’ (1996, 6–7). But one might claim, instead, that speakers of a language like Quechua, where stops are voiced after nasals, have the more ‘naive’ or ‘natural’ response to the difficulty of nasal-plus-voiceless-stop sequences: they abandon the effort to maintain voicelessness, and simply allow voicing to continue through the stop. It is the English speakers who must learn to make the special adjustments that are needed to keep the stops voiceless—at least voiceless enough to maintain the contrast. The resultant voicing of the stop is indeed partial or gradient articulatorily, but the stop remains voiceless perceptually—a categorical effect.

The question remains whether the partial voicing of these stops in English is unavoidable or ‘mechanical’, or whether there are languages in which a stop following a nasal remains totally voiceless, or whether the degree of voicing differs systematically in different languages.

---

25 Interestingly, Hayes’ algorithm for the child’s construction of constraints depends crucially on comparison of configurations that are the same in all respects except one—the exact kind of comparison that must take place in the application of a natural process.

26 The two voicing effects that Hayes is discussing here are in different grammatical domains, but the comparison is actually one of Quechua morphonology (since the substitution occurs only at suffix boundaries) vs English phonetics. Nevertheless, one might find a language where the Quechua-type substitution is a freely occurring, automatic one—one that could be called a natural process. Post-nasal voicing (a sub-case of voicing assimilation) may have been a natural process at some point in the history of Quechua. The examples cited in the literature all seem to be morphologically conditioned, but the comparison allows me to illustrate the difference between process application and suppression.

27 The English degree of voicing serves to maintain the contrast, but perhaps in a language where the voicing contrast is not supported by a vowel-length difference, speakers would have to take more care to keep the stops voiceless.
5.3 Features or gestures
Note that the identification of phonetic rules with natural phonological processes does not depend on whether the processes or rules—e.g. the assimilations discussed above—are represented in terms of features that assimilate or gestures that overlap. In natural phonology, a feature must be associated with regularities in the processing of speech, but what the feature is is the connection between an independently controllable gesture or gesture type and its acoustic or perceptual effect. Features must be translatable into gestural specifications; these specifications will include information about the intrinsic durations of gestures (i.e., duration associated with the gesture’s own phonetic quality—which is based on physiology, aerodynamics, etc.). If we regard the units of phonological processing as gestures or gestural types (degrees of constriction, etc.), we must also require that they be connected by speakers to perceptual qualities. Best (1995) for example, emphasizes the lawful relationships between gesture and acoustic effect, stating that ‘the range of vocal tract gestures that are (or could be) harnessed by human languages, and their lawful perturbations of the energy transmitted through various media (acoustic, optic, tactile…) can be considered to define the contents of a ‘universal phonetic domain’, akin to Catford’s (1977) anthropophonic space … it is assumed that perceivers recover information about actual speech gestures from the rich flow of lawfully shaped, multidimensional stimulation resulting from those gestures’ (186).

But whether we speak in terms of features or gestures, the question is still whether the extent of an adjustment like assimilatory spread (or gestural overlap) must be specified arbitrarily or whether it can be deduced from a) the full extent of other features (or gestures) specified in the input and b) the prosodic boundaries within the utterance.

6. Problems for the identification of ‘phonetic rules’ with phonological processes
Phonetic rules and phonological processes are intended to account for largely overlapping sets of data. Both appear to be mental operations and to involve selection from a universal set of default options—with the possibility of the language requiring non-defaults. But even if we maintain 1) that a number of interacting factors influence the actual output of the articulators, and 2) that feature and timing changes below the level of contrast can be regarded as phonological, and 3) that a wide variety of articulatory patterns, controlled by the speaker and perceptible to the hearer, are potential phonological features—there are still differences, which cannot at this point be entirely resolved. I will simply make a number of suggestions about the kinds of questions that remain.

---

28 Information on lexically specified duration, and context-sensitive duration and speech rate are also critical to the application of both natural phonological processes and phonetic rules, as noted in Section 5.3.
6.1 Assignment of language-specific phonetic values to features

In natural phonology, features have phonetic values, and therefore, segments with consistently different phonetic characteristics should be described with different features. Yet in many language descriptions, we find cross-linguistic identifications that nonetheless specify differences that would usually be thought to be below the level of phonological features: /i/ in language X is different from /i/ in language Y. Those who assume only the rather limited set of phonological features required by morphonological rules conclude that languages must specify articulatory (or acoustic) differences below the level of phonological features. It is important to consider a wider variety of possible feature descriptions. We must also consider the problem of transcription, as noted in Section 5.1.1-2. The question remains open whether, for example, different qualities of ‘the same’ vowel or different characteristic manifestations of ‘the same’ voiced-stop series in different languages can be characterized with categorical features.

6.2 Substitutions that result in cross-linguistic differences below the feature level

In the interests of phonetic accuracy, differences like those between the phonetic forms of French /N/, English /VN/, and French /VN/ must be described. If the natural phonologist is to claim that the ‘phonetic nasalization’ of vowels adjacent to nasals in English is the result of a phonological process, then there has to be a way to describe both English contextually nasalized vowels and French lexically nasalized vowels in categorical terms.

Similarly, if /s/ and /ʃ/ differ only by a single feature (e.g. [anterior]), a categorical process of assimilation ought to make a sequence of /s -ʃ/ become [ʃʃ] (or [ʃʃ]). But in the /s -ʃ/ sequences Zsiga examined, the result was more like [ʃʃ]. The situation here is similar to the nasalization problem, where a sound sequence that is the result of assimilation is not phonetically the same as a lexically specified sequence that would appear to have the same feature specifications.

6.3 Examination of crosslinguistic differences in articulation

If phonetic rules are ‘controlled,’ and in that respect similar to phonological processes, it is important to sort out the phonetic effects that are, in some sense, under speaker control from those that are ‘uncontrollable’. One might compare the nasalization of vowels before nasals in English, which is clearly something that speakers allow to happen, with the nasalization of vowels before nasals in French, where it is not ‘allowed’ and proceeds only minimally. Are the gradual contextual (assimilative) nasalization of English and the minimal contextual nasalization of French the only two options, or are there languages where contextual nasalization is ‘complete,’ so that the nasalization is comparable to that of French lexically nasalized vowels? And are there languages where it is non-existent, so that not even a minimal part of the pre-nasal vowel is nasalized?

Similar questions need to be explored regarding other assimilations, like /s/-palatalization. Are there languages where the palatalization of /s/ (or /ʃ/) is complete, even though it is optional and occurs across word-boundaries? Certainly there are
dialects of English where such palatalization does not occur, so that it cannot be attributed to a mechanical effect.

6.4. Evidence from cross-linguistic perception
Before we can refer confidently to language-specific phonetic specification below the feature level, it is crucial to determine whether such differences are perceptible, yet there is much that we do not know about the perception of foreign accent in this respect. If ‘sub-featural’ differences of phonetic quality are truly characteristic of different languages, so that languages need phonetic rules that specify these differences ‘numerically’, the question arises whether speakers can actually identify other speakers as ‘native’ or ‘foreign’ on the basis of these sub-featural differences. For example, English speakers can identify pronunciations of voiceless stops as foreign because they expect the aspiration that is typical of English voiceless stops (in appropriate environments), and that aspiration is absent in the L2 English of, say, French or Czech speakers. But it is not equally clear that English speakers will also recognize the Czech or French speaker’s voiced stops as ‘foreign’ because they are more reliably voiced through a greater part of their duration than English stops. Similarly, English speakers may identify a foreign speaker’s L2 English vowels as ‘foreign’ if they do not undergo the diphthongizations that are typical of the English speaker’s dialect, or if they diphthongize in different ways. But it is not equally clear that English speakers whose /i/ is high, palatal (front), tense, unrounded and monophthongal can reliably identify as ‘foreign’ an /i/ that is also high, palatal, tense, unrounded and monophthongal—but ‘somewhat more high and front than English /i/’.

7. Conclusion
Although physical speech output is continuous, both articulatorily and acoustically, and although languages differ considerably in phonetic detail, there is a great deal of evidence that, in the representations of long-term memory and in perception, speakers store and process speech categorically. Therefore, it would seem that the conservative claim is that speakers also process speech categorically in production, and that this categorical processing is sufficient to account for the differences between lexical representation (the output of the morphonology) and phonetic representation, and that the gradience of the speech signal results from multiple interacting factors in the physical production of speech.

In contrast to this, models of phonetic implementation typically assume specification of numerical values with respect to degrees of assimilation or gestural overlap, sub-featural specification of segmental quality, etc. But we have little in the way of illustration of how this kind of numeric specification works in speech processing. (No one writes these phonetic rules.) We also have little evidence of the relevance of numeric specification in perception, and no explanation of how such differences in degrees of features might be learned, if they are not perceptible.

Phonological processes, though categorical, are sensitive to the gradience of their phonetic output, which results from simultaneous and sequential feature combinations,
timing differences, etc. This attention to gradient results is expressed in the implicational conditions that govern the application of phonological processes even at the deepest levels of phonology, and it is a reflection of the speaker’s implicit knowledge of his or her own abilities and the lawful relationships between gesture and acoustic effect.

It might be said that the challenge for natural phonology is to confirm the hypothesis that speech processing is categorical, or phonological, down to the level of the actual phonetic (pronounceable) representation, and that, consequently, if we have phonological processes, we do not need ‘phonetic rules’. And certainly it is a challenge to account for the observed continuity and gradience of the phonetic signal as the physiologically mediated results of the interpolation between targets, the combination of simultaneous features or overlapping gestures, and differences of accent and duration. But in the logic of science, it is for proponents of the distinction between phonological processes and ‘phonetic rules’ to refute the natural hypothesis, and I hope I have been able to show that this has not yet been accomplished.
References


