1. Introduction

Formalist and functionalist approaches to language have long been seen as inherently antagonistic. However, in recent years this has begun to change, as formalists recognize the influence of language use on linguistic structure, and functionalists provide more explicit models for how this influence operates. This paper represents an example of this growing trend, attempting to formalize a functionalist approach to tonal phonology. The model of tone proposed here is formal in the sense of aiming to provide a completely explicit account of phonological knowledge, and yet the model is also functionalist, since its elements take the form they do because of the way phonological knowledge is actually used when speaking, hearing, and acquiring a language. More specifically, the model assumes that such grammar-external influences on phonological grammar occur in large part during the lifetimes of individual speakers, not merely through eons of biological natural selection on innate linguistic processing mechanisms (see e.g. Newmeyer 1998 for discussion of these two ways by which functionalism can arise in human language). Hence we strongly reject the claims of strong anti-functionalists such as Hale and Reiss (2000a,b) that phonological theory can only serve its purported purpose of describing "phonological computation" if it purges itself all phonetic substance.

Grounding these philosophical issues in a specific empirical domain, the focus in this paper is tone. For our purposes tonal phenomena have the advantages of being both phonologically quite intricate (e.g. spreading or dissimilating, interacting with or ignoring segmental features) and yet phonetically relatively straightforward (i.e. involving primarily a single perceptual dimension, although laryngeal physiology is rather complex). An additional advantage is that one of us has already presented a functionalist model of tone, in Tsay (1994). Ironically, although its goals were functionalist in the sense we mean here, the model presented in Tsay (1994) unwittingly followed the Hale and Reiss (2000a,b) dictum to separate "phonological computation" from all extragrammatical factors (e.g. perception, physiology, learnability). The result was that the most explicit part of the model was intentionally designed to do virtually no work by itself; all of the most interesting tonal phenomena were relegated to extragrammatical factors, which, given the framework in which Tsay (1994) worked, were necessarily less explicitly formalized.

However, there has since developed a theoretical framework that makes it far easier to give functional models formal rigor: Optimality Theory (Prince and Smolensky 1993; henceforth OT). We think that Hale and Reiss (2000a,b) are correct to claim that the OT literature contains a vast amount of functionalist phonology. While most of it is not explicitly labeled as such, some recent works wear the functionalist label with pride, in particular the theory of Functional Phonology presented in Boersma (1998) (see also Kirchner 1997, 1998, Hayes 1999, and the works discussed by Gussenhoven and Kager 2001, among many others). Functionalist OT models are formally explicit in providing strict algorithms for the construction of constraints and for the evaluation of candidate outputs through constraint ranking. They are functionalist in that the constraints (and often their
ranking as well) implement extra-grammatical factors such as speech physiology, auditory perception, and memory limitations. From a purely formal perspective, functional OT models have a great advantage over approaches that posit a strict separation between "grammar proper" and truly extra-grammatical factors, both anti-functionalist (Hale and Reiss 2000a,b) and functionalist (Tsay 1994): in the OT models every theoretical element essential for capturing the linguistic facts is formally explicit rather than merely hinted at.

This paper, then, builds on the functionalist insights of Tsay (1994) and the OT formalism of Boersma (1998) to argue for a new model that attempts to describe how cross-linguistic tonal generalizations arise from extragrammatical causes. We begin in section 2 by reviewing the methodological and formal underpinnings of the model. Next we apply the model in analyses of some fundamental aspects of tonal phenomena: section 3 concerns the nature of tone features, while section 4 examines the autosegmental behavior of tone. Finally, section 5 summarizes the conclusions that follow from the philosophical, empirical, and analytical discussions of the paper.

2. Functional phonology

Given the strong interest that GLOW attendees have in formal theories, we begin in this section with further justifications and clarifications of the notion of a formal functional model of phonology. First, in 2.1 we justify the logic behind formal functionalism, contrasting it with approaches that posit a strict separation between grammar and non-grammar. Then in 2.2 we introduce a specific model of functionalist OT, the Functional Phonology model of Boersma (1998).

2.1 Formal functionalism

In this rather philosophical section we argue that the best approach to phonology is one that is both formal and functionalist, in particular one that formalizes as much as possible of the extragrammatical factors involved in phonological patterns. In the course of this discussion, we show that the anti-functionalist criticisms of Hale and Reiss (2000a,b) are invalid and that the alternative approach to phonology that they advocate does not provide a feasible research program. We conclude that functional OT formalism provides the most promising approach currently available.

2.1.1 Phonology as an interface

Phonologists are not the only modelers who must be concerned with the tension between formalism and functionalism; without the first, a model has no predictive power, and without the second, a model may have no relevance in the "real world". However, the importance of both elements is particularly apparent in phonology, which is concerned with the parallel dichotomy between mental representations and physical events. For example, why does phonological assimilation show so many similarities with phonetic coarticulation (e.g. locality restrictions, systematic interaction with some features and independence from others, implementation of a single articulatory instruction spread across segments)? A purely functionalist answer might be that phonological assimilation is in fact a form of coarticulation, but this misses the fact that assimilation, unlike coarticulation, involves the manipulation of categorical symbols (e.g. toneme A is replaced by toneme B when adjacent to toneme B). A purely formalist answer might be that the similarities between assimilation and coarticulation are coincidental, or irrelevant, or at most the result of natural selection in our distant biological past for genes that allowed for a more convenient interfacing between the
fundamentally distinct worlds of phonological symbols and articulatory gestures. While this formalist move is perfectly respectable (and in fact something like it has long been standard in generative phonology), it only serves to avoid recognizing what makes phonology phonology. After all, if signed languages have phonology, phonological theory can only be defined as the study of that aspect of linguistic competence directly interfacing with physical gestures and signals. A complete model of phonology should study this interface, not pretend that it is irrelevant; otherwise it becomes formalism for formalism's sake, not an empirical science. For this reason, some understanding of how physical gestures and signals interface with phonological competence, i.e. some form of functionalism (biological or cognitive), is necessary.

2.1.2 Problems with an overly narrow formal approach to phonology

In spite of the necessity of doing it somehow or other, it is no trivial matter to combine formalism and functionalism together. By its very nature, functionalism makes reference to elements outside the formal model; one can't formalize everything, or else the model would cease to be functionalist (similar philosophical problems arise with formal models of semantics). Perhaps because of this, theorists who recognize the role of the physical aspects of speech in phonology vary widely in where they draw the lines around the truly formal core of their model. At one extreme are approaches like that of Hale and Reiss (2000a,b) and Tsay (1994), who reduce "pure" phonological knowledge to an extremely restricted formal minimum, supplemented by independent (and often vague) extragrammatical factors. In such models, the formal core is designed to capture all "computationally possible human grammars" (Hale and Reiss 2000b:167, italics in original), not just the ones that are possible to learn or capable of developing through diachronic processes like the phonologization of automatic phonetic processes. While Hale and Reiss (2000a,b) somewhat dogmatically claim that this formal core should be the true subject matter of phonological theory, Tsay (1994) more insightfully describes this approach as modular: the formal core overgenerates, and the extragrammatical factors trim back the set of logically possible grammars to those that are actually found (see also Anderson 1981).

One major trouble with this approach is that the formal core is too general to have much explanatory power; it essentially exists solely to give the extragrammatical factors something to work on. For example, central to the formalism of Tsay (1994) was the multivalued tone feature [αPitch], which formally was free to vary from [1Pitch] (the lowest tone level) to [nPitch], where n represents any positive integer. This proposal, by itself, only makes two empirical claims: that tone interfaces primarily with perception, not articulation, and that pitches can be kept categorically distinct in a language. The more complex tonal phenomena of interest to practicing phonologists, such as cross-linguistic limits on the number of tone levels, or natural classes of tones, were the responsibility of extragrammatical factors (in particular, perceptual, articulatory, and memory-based constraints on learnability). Similarly, in the only example that Hale and Reiss (2000b) give to illustrate the sort of formalism that they believe that phonologists should practice (in place of the "substance abusing" approaches standard in a large portion of phonological research today), formal conditions are posited to recognize when two consonants are or are not identical in every feature; it is held to be an important empirical discovery that there are no conditions that recognize when two consonants are different in every feature, or identical in only one feature. The extremely general nature of these conditions is meant to preclude any functionalist explanation of them (since vowels delete between consonants both when the consonants are identical and when they are not), but their generality also have the effect of relegating them to a quite marginal role in the analysis of phonological patterns in actual languages. Much
more supplementary work must be done to answer the sorts of questions practicing phonologists are wont to ask (e.g. why do the conditions refer specifically to consonants, not vowels, and do they apply with vowel insertion as well, and if not, why not?). The more phonologically specific the questions, the more likely the answers require reference to factors sullied by physics, which in Hale and Reiss’s view do not belong within the theory of phonological competence. In short, because such models give so much work to the vaguely described extragrammatical modules (such as "the nature of diachronic sound change" required at one point by Hale and Reiss 2000b:179), it is not clear whether they can be called truly "generative" in the sense defined by Chomsky (1965:4). A generative model should describe competence so explicitly that the reader is not required to fill in any gaps in the argument.

A second problem with approaches that define phonological formalism overly narrowly is that, given the interface role of phonology, the choices about what belongs within the formalism and what belongs outside it become quite arbitrary. For example, how does Tsay (1994) know that tone’s behaving like "phonological pitch" is a basic primitive of pure computation, not derived from the extragrammatical factors that play important roles elsewhere in her model? In fact, in this paper we will show that whether or not phonological tone "acts like pitch" depends on the functional motivation for the particular pattern (articulatory vs. perceptual). Likewise, it is not impossible that the similarity conditions of Hale and Reiss (2000b) are derivable from more general aspects of cognitive processing. For example, a functionalist might argue that like all complex events, segments can be processed in two different ways, either as wholes or as collections of features. Given the short duration of consonants and their non-steady-state nature (e.g. being identifiable primarily through formant transitions), consonants may be particularly likely to be processed more as wholes than as feature bundles. Hence consonants will tend to be perceived as entirely identical or not, and phonological processes may make reference to this distinction. To complete the hypothesis, it would have to be made far more explicit, but in doing so it would likely generate a variety of specific predictions that would never emerge from the original, overly formal analysis (e.g. obstruents should be more likely to obey the holistic identity conditions than sonorants, given that they are less steady-state). The point here is that without considering the functionalist alternatives, it is invalid to claim that one has truly reached the true formal core of phonology at last. The dilemma predates Hale and Reiss by over twenty years. Foley (1977) represents an earlier attempt to purge phonology of all phonetics (which for him, but apparently not for Hale and Reiss, included distinctive features), but ironically his proposed formalism relied heavily on notions like "fortition" and "lenition", now taken to be functional principles (e.g. Kirchner 1998). Hale and Reiss (2000b:182) commit a further logical fallacy when they claim that "the focus of phonological theory should be on the cognitive architecture of the computational system", i.e. pure phonology purged of all physics (Tsay 1994 adopts the more widely accepted view of phonology as being concerned with all phonological patterns, regardless of their cause). It does not make sense to define the empirical study of some domain X as being concerned solely with "pure X", since the very existence of "pure X" can only be determined by empirical research (and we feel there are very good reasons for doubting its existence in the case of phonology).

2.1.3 Advantages of formal functionalism

Given the fatal problems with the alternatives, we conclude that the best approach to the problem of formalism and functionalism in phonology is to stretch the formal boundaries in the opposite direction, as widely as possible. If the psychophysics of articulation or
perception are relevant for explaining a phonological pattern, then phonologists should make an effort to incorporate psychophysics into the formal model. The same holds for memory (e.g. processes of lexical access) and for factors only playing a role during language acquisition. This approach faces problems of its own, but they are far more tractable than those faced by the alternatives. The first is the logical problem alluded to above: it is impossible for a model to be fully formal and fully functional. This, however, is an inevitable problem for any scientific model; there will always be elements that seem necessary but which as yet are unformalizable. This is actually a benefit of this approach, since now the boundaries of the formal model are not defined arbitrarily, but rather by the limits of our present understanding.

A second, more interesting, problem is that if the model claims to be a model of competence, it thus claims to model the psychology of an individual speaker-hearer. Yet a complete explanation of synchronic phonological patterns must make reference to physics and to events that occurred long before present-day speaker-hearers were born (e.g. the Great English Vowel Shift). Again, however, this is a problem faced by any theory of phonology; the approaches described above merely ignore its difficult implications (e.g. by dismissing the obvious similarities between autosegmental spreading and coarticulation, or claiming that all descriptively valid phonological patterns must be psychologically real in spite of much psycholinguistic evidence to the contrary; see e.g. the review in Myers 1993 of evidence for the relative lack of productivity of lexical rules in English). For us the interface between the mental and the physical is the core of phonology, and understanding it is the primary task of phonological research; we don't intend to do all of this work in this paper. Even using the approach we advocate, then, there must come a point where something mysterious happens: the physical is mapped onto the mental. Our approach merely puts this off as long as possible, by putting as much as possible into the formal model of phonological knowledge, even those aspects that seem rather "physical" in nature.

This move is even easier to justify than it would have been at the time generative phonology was founded, because we have learned since then that much of what had been dismissed as "pure physics" in speech is in fact under systematic (and learned) psychological control. A famous example is the variability of vowel lengthening effects across languages; in spite of its ultimate causes apparently being physiological (rather than determined by abstract UG conditions), English speakers and French speakers learn to lengthen vowels to different degrees (Chen 1970). Among the numerous other examples are the systematic variations in the articulation of /l/ in English (Sproat and Fujimura 1993) and the interactions between tone and vowel height in Fuzhou, Mandarin, and English that will be discussed in section 3 (Tsay and Sawusch 1994). At the very least, such discoveries demonstrate that speakers' knowledge of the sound system of their language (what is usually called "phonology") goes far beyond the narrow, abstract confines claimed by Hale and Reiss (2000a,b). If the job of phonological theory is to formalize phonological knowledge, even phonetically detailed phenomena must be formalized.

2.1.4 Functional Optimality Theory

As Hale and Reiss (2000a,b) themselves note, Optimality Theory makes it quite straightforward to formalize extragrammatical factors. There are two primary reasons why OT is so well suited to functionalist analyses (though there may be other sociological factors at work, e.g. OT's historical relation with connectionism, an essentially empiricist approach to cognition). First is the device of constraint ranking, which implements the functionalist notion of structure emerging through compromises between equally valid grammar-external forces (e.g. ease of articulation vs. perceptibility). Second, there is the close relationship
between OT constraints and the notion of markedness. For example, NoCoda is preferred over NoOnset as a hypothetical constraint because languages tend to disfavor codas but not onsets. This means both that the greater "naturalness" of NoCoda can be directly observed (an empiricist claim), and that NoCoda is assumed to be only a tendency (on the surface at least), not a strict universal (opening the possibility that it is derived through learners' experience with language data, rather than being innate). This contrasts with the rationalist and nativist logic familiar from other generative linguistic theories. For example, abstract Case is presumed to exist in all languages, not because most languages have overt case marking, but because abstract Case accounts both for overt case marking and for a variety of other phenomena not obviously related to overt case, even in languages without case marking.

A key notion in any functionalist OT model is that the constraints are indeed motivated by extragrammatical factors, but are not themselves extragrammatical factors; they are elements in a grammar, after all. Thus a functionalist phonology OT model is not itself phonetics, but what Hayes (1999) calls "phonetically driven phonology". Hayes gives two reasons for maintaining a separation between phonology and physics. The first is the familiar observation noted above in reference to assimilation vs. coarticulation: phonology is categorical, but phonetics is not. The second is a somewhat more subtle observation: phonology is "symmetrical". By this Hayes seems to mean that phonological patterns form natural classes (i.e. divide up the "phonological space") in ways that are best expressed with categorical symbols, even if the natural classes don't exactly match their apparent psychophysical motivations. For example, he cites an aerodynamic model that assigns prohibitively high effort measures to post-obstruent voiced stops, to [d, g] in initial position, and to [g] after oral sonorants, yet languages virtually never treat precisely this set as a phonological class. Instead, languages choose related but simpler descriptions (e.g. disallowing voiced obstruct clusters as in Latin, or disallowing [g] across the board as in Dutch), although aggregating across many languages reveals the underlying asymmetrical but more phonetically natural pattern. These observations suggest that, unsurprisingly, physical forces are filtered through cognition before resulting in phonological patterns. Based on Hayes's observations, the main properties of this filtering are that it values categoricity and simplicity over strict psychophysical "accuracy".

Yet there is no reason to conclude that these properties represent Hale and Reiss's pure "phonological computation", since categoricity and simplicity are themselves expressible with interacting OT constraints acting on representations that are not inherently categorical or simple. Kirchner (1997) and Boersma (1998) independently discovered how categoricity can be made to emerge from competing OT constraints, and pressure to simplify mental representations is hardly unique to phonology. Informally, the key observation is that phonological representations must be easy to articulate, to perceive and to remember, while also preserving lexical contrasts, but since all this can't be done at once, compromises must be struck. For example, the aerodynamic model may describe a class of easily articulated sounds, but they may not be easy to perceive or remember as a coherent class. Languages subject to the same extragrammatical forces may resolve their conflicts in different ways, as can be expressed in OT with extrinsic constraint ranking; this, if anything, is the formal core of a functional OT model.

Note, by the way, that allusions to notions like "ease of articulation" in functional OT are not as vacuous (even fatuous) as assumed by Hale and Reiss (2000a,b), who suggest that a model based on "dysfunctional" principles (e.g. "no pain, no gain") would work just as well, and that functionalists, if honest, should spend research time studying "fundamental features of the human personality ('laziness', 'helpfulness', and so on)" (p. 181). Rather, in spite of its notoriously vague name, a notion like "ease of articulation" is a perfectly ordinary scientific
hypothesis: it is intuited by an observant scientist, formulated into a falsifiable form, then
tested against new data. The logic is the same as that which led phonologists to accept that
place features are unary rather than binary; perhaps the original hypothesis came from
pondering the nature of articulation, but it won acceptance by being formalized and tested.
Likewise, Hayes (1999) doesn’t claim that ease of articulation plays a role in phonology for a
priori reasons (e.g. because humans are naturally lazy), but because his algorithm for
converting the output of the aerodynamic model into OT constraints results in highly accurate
predictions about the aggregate pattern of disfavoring positions for phonological voicing.
Even if Hale and Reiss’s dysfunctional principles could be formalized and applied as
explicitly, it doesn’t seem obvious that they would yield as great an empirical success, any
more than a theory that allowed negative place feature values.

Once these broad issues have been addressed, the remaining challenges for the
functional OT approach are more technical. One particularly important one concerns
modularity. In spite of its other weaknesses, an approach like that of Tsay (1994) does
recognize that extragrammatical forces like physiology, perception, and memory have quite
different behaviors. Is it really appropriate to represent them by constraints ranked together
in a single constraint hierarchy? In the OT literature, questions like this are not new; for
instance, it is still an open question whether syntactic and phonological constraints should be
combined into a single hierarchy, or separated into two independent hierarchies (Prince and
Smolensky 1997). If OT is thought of as an actual model of cognition, a unified hierarchy
would seem to imply that articulation, perception, and memory are in principle directly
subject to each other’s constraints, which does not seem cognitively (or neurologically)
plausible. We therefore follow the suggestion of Boersma (1998) that phonology must be
described by a set of modular subgrammars, in particular distinct grammars for production
and for perception.

One final question that may be asked is whether the processes of production and
perception are best formalized by constraints, and their interactions best formalized with
strict-dominance relations. At this point, however, we must leave philosophy behind and
allow the empirical coverage of the OT approach to speak for itself.

Summarizing, then, a functional OT approach to phonology is promising because it is
more functionalist than earlier generative theories (which ignored obvious parallels between
phonology and physics) and more formalist than approaches like Tsay (1994) or even,
ironically, Hale and Reiss (2000a,b).

2.2 Formalizing functional phonology

Probably the most explicit and complete functional OT model is the Functional
Phonology model of Boersma (1998). This model is still undergoing development (see e.g.
Boersma 1999, Boersma and Hayes 2001) and has a large number of components, not all of
which will be employed in this paper. For the purposes of formal explicitness, however, we
will adopt its claims, analyses, and terminology wherever they relate to our own empirical
issue, phonological tone. This section provides an introduction to these elements of the
theory, in particular the relation between the separate modules for production and perception,
the nature of production and perception constraints, the emergence of categoricality in
perception and production, and functionally motivated constraint rankings. We end with
discussion of a key weakness of the theory, namely its inability to account for lexicalized
patterns; these we pass over to another OT module for the lexicon.

The central claim in the theory of Functional Phonology is that phonological systems
involve separate grammars for production, perception, and recognition. These grammars
divide the familiar speech cycle into three mappings, respectively: between underlying
forms and articulatory forms (instructions to the articulators), between acoustic forms (generated by the articulatory forms via the laws of physics) and perceptual forms, and between perceptual forms and the underlying forms. The division between perception and recognition is made because Boersma (1998) assumes that semantics plays no role in perception, but does in recognition. While a complete theory rightly acknowledges such issues, in this paper we will conflate the perception and recognition grammars, using "perception grammar" to label the conflation. Our perception grammar, therefore, maps between acoustic inputs and underlying forms, which are thus perceptual targets; these same perceptual targets are also used as the inputs to the production grammar (the consequences of this assumption will be discussed further below). The production and perception grammars are OT grammars, defined solely by the ranking of structure constraints that evaluate outputs (articulatory or perceptual) and faithfulness constraints that compare the outputs with their associated inputs (underlying forms in the case of the production grammar, acoustic forms in the case of the perception grammar). The resulting simplified model appears as follows (for more detailed diagrams see Boersma 1998:143, 1999:1; one notational difference is that we use "//" to mark the underlying form, whereas Boersma uses this only for the perceptual input, deleted in our simplification).

(1) The relation between the production and perception grammars

```
/underlying form/
```

perception grammar                               production grammar

[acoustic input]                                 [articulatory output]

physics

Our simplification also leaves out several other elements of Boersma's model. In the complete model, for example, it is claimed that the perceptual forms derived from the articulatory forms generated by the speaker's own production grammar are evaluated for their faithfulness to the underlying forms. Especially important coordination occurs during acquisition, where an algorithm for adjusting constraint rankings is triggered by mismatches between the perceptual forms derived from a learner's own utterances and the perceptual forms derived from the utterances of an adult model (see Boersma and Hayes 2001). Even without these mechanisms, however, it should be clear from (1) that the perception and production grammars will necessarily be highly coordinated, since they share a representation at one end, and at the other end the representations have a deterministic physical relationship.

The production and perception grammars each have their own structure and faithfulness constraints, all of which are couched in detailed psychophysical terms. Prototypical examples of these constraints are given below, categorized as structure constraints (operating solely on outputs) or faithfulness constraints (comparing inputs and outputs).

(2) Production grammar constraints (after Boersma 1998:152, 176; Boersma 1999:7)

a. Structure

*GESTURE(articulator: gesture / distance, duration, precision, velocity):
Do not let a certain articulator perform a certain gesture, along a certain distance, for a certain duration, and with a certain precision and velocity.
b. Faithfulness

*REPLACE(feature: value₁, value₂ / condition / left-env _ right-env):
Do not replace, for a certain feature, a specified value (value₁) with a different value (value₂), under a certain condition and in the environment between left-env and right-env.

(3) Perception grammar constraints (after Boersma 1998:163-5)

a. Structure

*CATEG(feature: value):
The given perceptual feature cannot be recognized as the given value.

b. Faithfulness

*WARP(feature: distortion):
The perceived value of a feature cannot be different from the acoustic value by the given amount of distortion.

We have changed the description of production grammar constraint *REPLACE in (2b) somewhat in order to accommodate it to the simplified theory presented here. In the full theory, *REPLACE evaluates the faithfulness of the speaker's perception of her own production to the underlying form; the features that it mentions are perceptual features. Here we will usually be able to get away with making the simplifying assumption that the features are instead the acoustic features automatically generated by articulation. In either version of the theory, the articulatory output is, strictly speaking, not directly subject to any faithfulness constraints, which while making for unorthodox OT seems quite reasonable from a processing perspective. Moreover, note that the necessity of faithfulness to hold between underlying forms and perceptual/acoustic forms means that the underlying forms must be represented in perceptual terms, not gestural terms, even though they also serve as input to the production grammar. The model thus makes a quite radical break from the standard view of underlying forms in production. Rather than serving as an abstract version of the articulatory program (an assumption made by theories as disparate as Chomsky and Halle 1968, autosegmental phonology, and the articulatory phonology model of Browman and Goldstein 1992), the theory of Functional Phonology assumes that the underlying form and the articulatory form are entirely distinct sorts of entities that are actively coordinated via perception, rather than literally being derived one from the other. This interesting hypothesis is really only expressible in a non-derivational, constraint-based model like OT. However, in the rather naive form we are assuming for most of this paper, it cannot be quite right, since it ignores the well-known problem of the lack of acoustic invariants for determining articulations. To take one familiar example, the presence of a plosive gesture in a form like spill is indicated acoustically both by the formant transitions for [p] and by the duration of silence. Yet these two features operate in perception by trading off, rather than working independently or additively, so with a longer silent duration less accurate formant transitions are sufficient for recognizing the presence of a stop (Liberman 1982). Boersma (1998) seems to assume a mostly psychoacoustic approach to speech perception, but in principle Functional Phonology is not forced to take a stand in the debate over the motor theory of speech perception (e.g. Liberman 1982 vs. MacNeilage 1991). Hence we will occasionally assume that articulatory gestures are directly perceivable, though we keep such
assumptions to an absolute minimum (and future work may eliminate their necessity entirely).

We now illustrate a simple application of these constraints for the fundamental issue of categoricality. The representations and constraints of Functional Phonology are phonetically specific, so categoricity is held to be an emergent property, not inherent. The empirical advantages of this approach are reviewed elsewhere (e.g. Kirchner 1997, Myers 1997, Steriade 2000) and further arguments will be discussed below, when we examine tone features in more depth. Here we merely exemplify the proposal.

Suppose a language has two distinct level tones. A speaker-hearer thus has two tasks: to perceptually distinguish these two categories, even if presented with pitches not identical to the prototypes, and to produce the two tones distinctively. Suppose that the prototypical acoustic form for the low tone is a pitch located at a point one fifth up from the lowest producable pitch (i.e. 20% of the speaker's pitch range) and the prototypical acoustic form for the high tone is at the 80% point (note that we assume that pitches are represented for the speaker-hearer in a normalized fashion, another simplifying assumption not made in Boersma 1998). In the present framework, the existence of the two tone categories and their prototypical locations in the pitch range is represented by the following two constraints.

\[(4) \quad \star\text{CATEG(\text{pitch}: 20)}\]
\[\star\text{CATEG(\text{pitch}: 80)}\]

These two structure constraints are ranked below a family of faithfulness constraints that disallow the recognition of an acoustic form as a given category if it is too far from the prototype for that category. For example, the following constraint would not allow a pitch at the 70% point to be recognized as a low tone (since \(|20-70|>30\), but it would allow it to be perceived as a high tone (\(|80-70|<30\).

\[(5) \quad \star\text{WARP(\text{pitch}: 30)}\]

These two sets of constraints obey two functionally motivated ranking principles. The first states that "a less distorted recognition is preferred over a more distorted representation" (Boersma 1998:164). This can be formalized as follows.

\[(6) \quad \star\text{WARP(\text{feature}: d_1)} \gg \star\text{WARP(\text{feature}: d_2)} \iff d_1 > d_2\]

The second principle states that only a finite number of \(\star\text{CATEG} \) constraints can be dominated by \(\star\text{WARP} \) constraints. As Boersma (1998:164) notes, this is motivated by learnability considerations: it takes positive evidence for the language learner to demote any given \(\star\text{CATEG} \) constraint. We will exploit this insight in the next section.

The interaction of \(\star\text{CATEG} \) and \(\star\text{WARP} \) constraints in the example of the two-tone language are illustrated in the following tableaux. We arbitrarily assume that both tones in this language tolerate a large amount of distortion, so we rank the two \(\star\text{CATEG} \) constraints below \(\star\text{WARP(pitch: 30)} \). All acoustic inputs are decisively mapped onto the nearest of the two perceptual categories (low tone or high tone), except for the input directly between the two prototypes, which is correctly predicted to be ambiguous.
(7) (Note: [] marks an acoustic input, // marks a perceptual output, X represents an acoustic form with pitch X% from the bottom of the pitch range, "C"="CATEG", "W"="WARP")

a.

<table>
<thead>
<tr>
<th></th>
<th>*C(50)</th>
<th>*C(70)</th>
<th>*W(60)</th>
<th>*W(30)</th>
<th>*C(20)</th>
<th>*C(80)</th>
<th>*W(20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/20/</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/50/</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/70/</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{}/80/$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

b.

<table>
<thead>
<tr>
<th></th>
<th>*C(50)</th>
<th>*C(70)</th>
<th>*W(60)</th>
<th>*W(30)</th>
<th>*C(20)</th>
<th>*C(80)</th>
<th>*W(20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/20/</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/50/</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/70/</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{}/80/$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

c.

<table>
<thead>
<tr>
<th></th>
<th>*C(50)</th>
<th>*C(70)</th>
<th>*W(60)</th>
<th>*W(30)</th>
<th>*C(20)</th>
<th>*C(80)</th>
<th>*W(20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{}/20/$</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/50/</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/70/</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{}/80/$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Categorical perception can thus be handled without assuming that the target perceptual forms are themselves represented with abstract symbols. However, categorical effects are also found in production. For example, the speakers of our two-tone languages probably tend to pronounce the low tone with a pitch value close to [20] and the high tone around [80]. In Functional Phonology, this sort of phenomenon is also ultimately due to categorical perception, this time operating via the *REPLACE constraints (recall that these evaluate the faithfulness of the perceptual/acoustic results of the articulatory output to the underlying form). Like the *WARP family, *REPLACE constraints also conform to a functionally motivated principle, here designed to minimize categorization errors: "a production that gives rise to a less distant categorization error is preferred over one that leads to a more distant error" (Boersma 1998:177). This can be formalized in terms of constraint ranking as follows.

(8) *REPLACE(feature: x, y₁) >> *REPLACE(feature: x, y₂) ⇔ |y₁-x| > |y₂-x|

Speakers will thus tend to produce articulatory outputs that result in acoustic forms as close as possible to the underlying (perceptual) forms, which in the case of our two-tone language, would consist just of /20/ and /80/. However, it is physiologically too difficult to produce articulations that hit these targets precisely all the time. For example, if a low tone is surrounded on both sides by high tones, it is easier to produce the tone in the middle if its articulation is adjusted to be somewhat closer to that of the surrounding high tones. The physiological motivation behind this can be formalized with a *GESTURE constraint of the following form (where "tone" represents the complex articulations involved in producing a pitch, d measures the distance traveled by an articulator, relativized to the articulator's...
(9) *GESTURE (larynx: tone / distance = d, velocity = v)

The principles for the ranking of articulatorily motivated constraints in the production grammar all boil down to the extremely general one stated by Boersma (1998:149): "an articulation which requires more effort is disfavored", represented schematically as follows.

(10) *EFFORT(x) >> *EFFORT(y) ⇔ x > y

This general constraint can be made explicit by grounding it in Newtonian mechanics. For example, changing a gestural articulation from one nonneutral position to another requires the movement of physical bodies, so all other things being equal, shorter articulatory movements are preferred over longer ones, and slower movements are preferred over faster ones. (The "other things" that we abstract away from here include the effort needed to hold muscles in a nonneutral position; see Boersma 1998:149-151 for a more thorough analysis describing further constraints and somewhat more complex notation.)

(11) a. *GESTURE(distance = d₁) >> *GESTURE(distance = d₂) ⇔ d₁ > d₂
    b. *GESTURE(velocity = v₁) >> *GESTURE(velocity = v₂) ⇔ v₁ > v₂

In the case of tone, these constraint rankings mean that the large distance between the articulations necessary to produce /80/ and /20/ should be shortened and/or taken more slowly. To make the analysis concrete, we will use dₓᵧ and vₓᵧ to represent the distance and velocity involved in moving between the articulations that result in the acoustic outputs [X] and [Y]. This allows the following tableau to describe the situation where the underlying form contains a low tone between two high tones, whereas in the actual pronunciation the tone in the middle is raised slightly. As shown earlier, such minor deviations should not cause any problem for the perception grammar.

(12) (Note: // marks an underlying (perceptual) form, [] marks an articulatory/acoustic input, "G"="GESTURE", "R"="REPLACE")

<table>
<thead>
<tr>
<th>/80 20 80/</th>
<th>*R(20,80)</th>
<th>*G(dₓᵧ, vₓᵧ)</th>
<th>*G(dₓᵧ, vₓᵧ)</th>
<th>*R(20,30)</th>
<th>*G(dₓᵧ, vₓᵧ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[80 20 80]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>[80 30 80]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>[80 80 80]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

These Functional-Phonological analyses of the perception and production grammars may appear to be pure (psycho)physics misleadingly couched in grammatical terms, but they are not: they are in fact pure grammar, albeit grammar influenced by physics. Physics only plays an indirect role, in particular in providing the constraint hierarchies. As with any formal model of phonology, the precise way in which physics and phonology come to share these similarities cannot itself be formalized (see the philosophical discussion in 2.1), but to its credit the present model puts much more physically motivated information into the grammar itself than previous models. This means that much less is left to the imagination, and the grammar is therefore more generative, in the sense of Chomsky (1965).

However, one may still object that the model cannot be right, since it claims that all of this physics-like detail is part of phonological competence. We (along with other proponents of functional OT) would respond that this property is precisely what is so
appealing about it, since as mentioned earlier, copious amounts of phonetic and psycholinguistic research conclusively demonstrate that such physical detail is indeed part of phonological competence. For example, the production grammar given above claims that only a certain degree of tonal coarticulation occurs, no more. This degree would be different in another language, if the relevant constraints were ranked differently. This is precisely what we want, since it captures the observation that languages do indeed vary in the degree to which coarticulation occurs, as noted in section 2.1.

There is another problem with the model as described so far: it gives the impression that all of phonology is "natural", which as Anderson (1981) and many others have pointed out, is not in fact true. For example, velar softening in present-day English (e.g. electric-electricity) does not follow automatically from any obvious articulatory or perceptual principles. Boersma (1998:465) acknowledges as much, but other than suggesting that constraints that capture such patterns must be language-specific and intimately related to morphology, his model has nothing else to say about them. In fact, all theories of phonology are challenged by unnatural patterns; velar softening, for example, receives an extremely cumbersome formal representation in Chomsky and Halle (1968) and even Borowsky (1986) makes no attempt to provide an autosegmental analysis of it. From a functionalist perspective, the key observation is that there is a direct correlation between the degree of unnaturalness of a pattern and its degree of lexicalization, as measured independently by the familiar diagnostics (e.g. the presence of lexical exceptions, interaction with morphology, structure preservation, and so forth; e.g. Hargus and Kaisse 1993). This suggests the necessity for another phonological module for the knowledge of lexical patterns, including the lexical representations of words, their morphological relationships, and the phonological generalizations that can sporadically diffuse across the lexicon, often described as analogy (e.g. the voicing of the stem-final consonant in dwarf-dwarves). Myers (in press) presents one model of this lexical module in an OT framework; others include Benua (1997), Burzio (1997, to appear), and Steriade (2000). For the purposes of this paper, all we need to assume is that such a module exists and is separate from the production and perception grammars per se.

The focus of this paper will be the perception and production grammars, as formalized above. In the remainder of this paper we will demonstrate how these grammars operate to account for a variety of fundamental phonological properties of tonal behavior.

3. Tone features

We begin by applying the Functional Phonology models to two basic questions relating to tone features. In 3.1 we argue that the cross-linguistic limits on the number of tone levels should be handled by constraints external to the representational mechanisms of the theory, rather than following directly from the nature of tone features. In 3.2 we argue the same thing for the interactions between tone and other aspects of production, showing that a functional OT approach works better than one employing articulatorily based (i.e. laryngeal) tone features. In general, then, we argue that tone features, as such, should play no role in phonological theory; the behavior of tone follows instead from the representation of pitch and laryngeal articulation, not from abstract phonological features.

3.1 The number of tone levels

It has always been completely standard in generative phonology to expect the formalism to account for cross-linguistic patterns of lexical contrast. For example, the early observation that no language showed more than a three-way contrast in vowel height led to
the [high]/[low] system advocated by Chomsky and Halle (1968), while evidence for four-way and five-way contrasts later led Clements and Hume (1995) to challenge this feature system and replace it with a new one. Thus phonological theory should also be responsible for the fact that no tone language has been found that contrasts more than five tone levels (Maddieson 1978). Yet as we will argue in this section, the theory that handles this fact cannot be one that is purely formal, divorced of all influence by extragrammatical factors. To account for cross-linguistic limits on the number of tone levels, we will provide a functional OT analysis that makes reference to memory limitations on grammar learnability, demonstrating its superiorit to theories that stipulate a representational maximum on the number of tone levels.

The proposed analysis is adapted from Tsay (1994), who began with a more general observation from Maddieson (1978): the more tone levels in a system, the fewer the languages that exemplify that system (or more formally, there are more languages with n contrasting tone levels than n+1 tone levels). The relevant figures from Maddieson (1978) are given below.

<table>
<thead>
<tr>
<th>number of tone levels</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>&gt;5</th>
<th>Total size of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of languages</td>
<td>128</td>
<td>69</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>207</td>
</tr>
</tbody>
</table>

Yip (1980), whose tone system is designed partly to capture cross-linguistic limits on the number of tone levels, does not fare well with the above data. She assumes that the true maximum is four, which is easily derived by combinations of her two binary features; the rare languages with five levels she dismisses as misanalyzed or as requiring an extra, restricted feature.

However, languages making five-way lexical contrasts in tone level are not dramatically more peculiar than another other tone systems, in spite of their relative rarity. Examples of such languages include Heimiao (Kwan 1966), Copala Trique (Hollenbach 1984), and Gaoba Dong (Shi, Shi, and Liao 1987). A minimal set of Heimiao words are shown below (1=lowest pitch, 5=highest pitch, following the IPA).

(14) From Kwan (1966), based on data collected by Fang-Kuei Li.

a. \[la^{11}\] "candle; smooth, polished"
b. \[la^{22}\] "to move away, to even"
c. \[la^{33}\] "cave, den"
d. \[la^{44}\] "a general classifier"
e. \[la^{55}\] "short"

More importantly, in the cross-linguistic pattern there is no sharp cut-off as one might expect if tone levels were fixed by some categorical feature system (e.g. four-level systems common, five-level systems nonexistent or marked). Instead there is a gradual drop off, with four-level languages already rather rare. Given this pattern, the fact that six-tone-level languages are (as yet) unattested should not really come as a surprise, nor would it be surprise if a more thorough search turned one up. One of the facts that Yip's model was devised to explain is thus not in fact explained at all.

Other tone theories take an approach more like that of Hale and Reiss (2000a,b): the maximum number of tone levels predicted by their phonological computation system (i.e. categorical tone feature) bears absolutely no correspondence with the facts. Thus the maximum has been given as three (Halle and Stevens 1971), four (Pulleyblank 1986, Bao 1990), five (Wang 1967, Sampson 1969, Woo 1969, Maddieson 1972), nine (Hyman 1989,
Duanmu 1990), and no maximum (Tsay 1994, who summarizes the others). If they are to account for the gradient drop off shown in the above table, clearly something must be added to their analysis.

Tsay (1994) is the only one to state explicitly what this something might be, though perhaps most of the listed theorists (including Hale and Reiss) would agree with her insight. Namely, the gradient drop-off must be accounted for by a gradient extragrammatical factor, most plausibly involving memory (perception alone won't do, since the mammalian ear allows us to distinguish all the keys on the piano plus hundreds more). In particular, the apparent five-level limit is suspiciously close to the famous "magic number 7±2" limitation on short-term memory discussed by Miller (1956). Memory would play an especially important role when the tone system was being learned. Each tone category would require positive evidence to learn, and learning would also be required to distinguish it from all of its competitors. A more precise learning model might even be able to capture the actual shape of the drop-off curve (necessary to explain why five-level languages have been found but not six-level languages). Nevertheless the most important aspect is already captured by the learnability hypothesis: the curve is monotonic (rather than becoming level in spots or rising again).

In a modular approach like Tsay (1994), this drop-off, strictly speaking, is not part of the grammar. The grammar merely supplies the feature \([\alpha]\text{Pitch}\), whose numerical value for \(\alpha\) is formally unrestricted; a language with a phonemic tone level for each key of the piano is thus computationally possible according to her proposal, just not learnable. This also absolves her grammar of responsibility for explaining why no more than five-level languages have ever been found. However, we just observed that explaining cross-linguistic patterns of phonemic contrast has always been a standard duty of generative phonological theory. Moreover, since the explanation for the drop-off (and hence the five-level limit) is claimed to be extragrammatical, it is not clear who precisely is responsible for explaining it; certainly few nonphonologists would be interested enough in the problem to do the hard work of creating an explicit formal model for us. In short, as discussed earlier, phonological models that define their formalism too narrowly are not really justified in calling themselves generative.

The model of Functional Phonology, however, can provide an account for both aspects of the pattern directly within its formalism. The first aspect, the gradual drop-off, follows directly from the claim that only a finite number of \(*\text{CATEG}\) constraints can be dominated by \(*\text{WARP}\) constraints. As noted earlier, this claim itself is derived from learnability considerations: positive evidence about the necessity of a new phonemic category, and its prototypical location, is necessary in order to demote a \(*\text{CATEG}\) constraint from its original undominated position. Since each new demotion adds to the overall cost in acquiring the system, a three-level language will be harder to learn, and consequently rarer, than a two-level language. This observation is that same as made in Tsay (1994), but now the formalism allows us to encode it directly.

The second aspect, the steepness of the drop-off (which indirectly causes the apparent five-level limit), also emerges within the formal parameters of the current theory. In Functional Phonology, the creation of a new category is not an all-or-none affair, since it also involves determining the size of the category. In terms of the model, a \(*\text{CATEG}\) constraint must not only be demoted, but demoted below a particular \(*\text{WARP}\) constraint. Given that there are a continuously infinite number of \(*\text{WARP}\) constraints (i.e. in our notation there is a \(*\text{WARP}(x)\) constraint for every real \(x\) such that \(0 \leq x \leq 100\)), we can't assume that acquisition involves demotion by equal-sized steps. Instead, demotion must initially involve large steps that are then gradually reduced in size as the target grammar is honed in on (precisely this approach is assumed by Boersma 1998, Boersma and Hayes 2001 in the Gradual Learning
Algorithm). This will add nonlinearity to the acquisition algorithm, no matter what the
details are, because the amount of precision required to hone in on a workable tone system
will depend nonlinearly on the number of tone categories.

To make this concrete, suppose the learning algorithm involves halving the distance of
constraint reranking at each step (e.g. halfway down, then a quarter way up or a quarter way
down, and so on). If there are \( n \) tone categories, the target for *CATEG demotion has a size
of \( 100/(2n) \) (i.e. any ranking of the two *CATEG constraints below *WARP(100-100/(2n))
will allow for completely nonoverlapping category ranges). Thus languages with different
numbers of tone levels will have target sizes as shown in the second row of the table below.
The number of steps needed to demote a single *CATEG constraint is shown in the third row.
For example, to reach the target range for a two-level language, two steps are needed
(demotion below *WARP(50), then demotion below *WARP(75)). This pattern already
represents a nonlinear function. The fourth row multiplies these numbers by the number of
categories in each system, on the assumption that these demotion steps are entirely
independent. This pattern, a steep nonlinear function, represents the degree of difficulty in
learning each of these tone systems, which is then inversely correlated with the number of
languages representing these systems. A more plausible constraint-demotion algorithm
based on Functional Phonology would give essentially the same results, and perhaps better
ones.

(15) Size of target range in *WARP family for demotion of *CATEG

<table>
<thead>
<tr>
<th>number of tone levels</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>size of target range</td>
<td>25</td>
<td>16.7</td>
<td>12.5</td>
<td>10</td>
</tr>
<tr>
<td>number of demotions to learn each category</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>total number of demotions</td>
<td>4</td>
<td>9</td>
<td>12</td>
<td>20</td>
</tr>
</tbody>
</table>

By employing Functional Phonology, then, the insights of Tsay (1994) can be captured
without having to pass off the crucial explanations to a nonformalized part of the model.
Instead, the functional OT framework makes it possible to describe the memory and
acquisition forces that cause the steep drop-off in the number of tone levels, all using devices
intrinsic to the formal model itself.

3.2 The modality of tone features

Traditionally, features in generative theories tend to reflect articulation rather than
perception (see especially Halle 1983). In the case of tone, several attempts have been made
to describe tone entirely with articulatory features, such as Halle and Stevens (1971),
Ladefoged (1973), Bao (1990), and Duanmu (1990). As with limits on the number of tone
levels, however, we hold that the relation between tone and the larynx should emerge through
the interaction of phonetically detailed constraints, rather than reflecting the inherent nature
of abstract tone features. In this section we provide support for this position by
demonstrating that tone can interact with nonlaryngeal features as well, as in the tone-vowel
1994, 1996, Jiang-King 1999). Such interactions are predicted because they are
physiologically motivated, but since the motivation is not strong, they are correctly predicted
to be rarer than interactions between tone and laryngeal features. After describing the
Fuzhou pattern and its physiological motivation, we show how Functional Phonology can
capture it and also account for why such patterns are rarer than tone-laryngeal interactions.

Our approach is thus in sharp contrast with theories that posit laryngeal features for tone.
In Bao’s (1990) proposed feature geometry for tone, for example, the laryngeal node
dominates (directly or indirectly) nodes for the vocal cords, the glottis, the cricothyroid cartilage, and the vocalis muscle, with the leaves of the tree being the four articulatory features of Halle and Stevens (1971). The model is supported by evidence that tone interacts with laryngeal features in the appropriate way (e.g. there is a correlation between high tones and voiceless consonants, and between low tones and voiced consonants). Functional approaches to phonology would predict the same interactions, given that tone is indeed articulated by the larynx. However, by claiming that the tone features are inherently laryngeal, models like Bao’s are incapable of accounting for any other type of interaction. In fact, while interactions between tone and vowel height are relatively rare, they are not entirely unheard of; in addition to Fuzhou, they have also been reported in Cantonese (Hashimoto 1972), Hausa (Pilszczikowa-chodak 1972, Newman 1975), Lahu (Matisoff 1973), and Mandarin (Chao 1948, Tsay and Sawusch 1994). The Fuzhou pattern is particularly interesting, however, because there are phonetic data (from Tsay and Sawusch 1994 and Tsay 1996) demonstrating that the vowel height changes are categorical, meaning that it cannot be dismissed as a purely phonetic phenomenon.

As in many Sinitic languages, tone sandhi in Fuzhou is quite complex. Of interest here are the vowel height changes triggered by the tone change: higher tones are associated with higher vowels. Sample data are shown below in (16), while (17) summarizes all tone-triggered vowel alternations observed in Fuzhou (including some found in data not shown in (16)).

(16) Data from Wang (1969), Chan (1985). (Note: for clarity tones have been abstracted to just three pitch heights: H, M, L)

<table>
<thead>
<tr>
<th>Isolation form</th>
<th>Context form</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tei[LML]</td>
<td>ti[H]-uon[H]</td>
</tr>
<tr>
<td></td>
<td>&quot;ground&quot;</td>
</tr>
<tr>
<td>b. hou[LML]</td>
<td>hu[HL]-ny[M]</td>
</tr>
<tr>
<td></td>
<td>&quot;woman&quot;</td>
</tr>
<tr>
<td>c. tey[LM]</td>
<td>ty[HL]-3ouk[LH]</td>
</tr>
<tr>
<td></td>
<td>&quot;write&quot;</td>
</tr>
<tr>
<td>d. kaiŋ[LML]</td>
<td>kein[HL]-nyou[M]</td>
</tr>
<tr>
<td></td>
<td>&quot;county&quot;</td>
</tr>
<tr>
<td>e. tau?[LH]</td>
<td>tou?[M]-mu?[H]-tseu[M]</td>
</tr>
<tr>
<td></td>
<td>&quot;peck&quot;</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>tones</th>
<th>vowel alternations</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM, LML</td>
<td>æ   ai  au  ay  ei  ou  ëy  ieu  uoi</td>
</tr>
<tr>
<td>H, HL, M</td>
<td>e   ei  ou  oy  i  u  y  iu  ui</td>
</tr>
</tbody>
</table>

The key generalization is that when low tones (i.e. beginning with L) are replaced with higher tones (i.e. beginning with H or M), the vowels are raised one level of vowel height: low vowels ([æ] and [a]) become mid vowels ([ë], [e], or [o], assimilating the frontness and/or roundness of the offglide if present), while mid vowels ([ë], [o], and [œ]) become high...
vowels ([i], [u], and [y], respectively). Analyses of this pattern that deny a direct tie between tone and vowel height, such as Wright (1983), Chan (1985), and Jiang-King (1999), focus instead on the apparent changes in prosody (e.g. diphthongs becoming monophthongs, or triphthongs becoming diphthongs). However, these alternative analyses neglect the facts that not all of the alternations change prosodic structure (e.g. [æ]-[ɛ], [ai]-[ei]), and that even when the number of segments does seem to change, this can always be understood as involving the merger of two high vowels (i.e. the raised mid vowels always produce a high vowel identical to the original preglide or offglide).

How can this interaction between tone and vowel height be accounted for? Clearly theories that claim that tone is handled by laryngeal features are insufficient, but it is equally unhelpful to propose that tone can have links in the feature geometry to vowel height as well. Adding such extra links would not only be difficult to represent geometrically, but it would miss the point that patterns such as Fuzhou's are intriguing precisely because they are relatively uncommon. Nevertheless, as with tone interactions with laryngeal properties like voicing, the explanation for the Fuzhou pattern also appears to lie in articulation. The articulation of the larynx is such that the raising of pitch height can be enhanced by moving the entire larynx upward with extrinsic muscles. This rotates the thyroid cartilage relative to the cricothyroid, thereby lengthening the vocal folds and raising pitch. This lifting of the larynx in turn raises the hyoid bone somewhat, to which the larynx is anchored at the top. The tongue root is also partly anchored to the hyoid bone, so raising this bone makes it easier to raise the tongue body as well. Thus raising pitch can reduce the energy required to produce a rise in vowel height. However, since the route from pitch raising to vowel raising is rather indirect, the physiological effect is rather small (Honda 1983, Honda, Hirai, and Dang 1994). These articulatory facts explain both why the Fuzhou pattern is able to exist, and yet why patterns like it are relatively uncommon (see Tsay 1994 for further discussion). Moreover, these physiological considerations may also explain why the reverse interaction, where vowel height affects tone, is not attested in phonological systems, given that merely lifting the larynx through tongue body raising is not guaranteed to stiffen the vocal cords in the required way (see also Hombert 1977, who offers perceptual reasons for the asymmetry in tone-vowel height interactions).

Yet in spite of its physiological motivation, the Fuzhou pattern is truly phonological, not merely phonetic, since the vowel height alternations involved are categorical. Tsay and Sawusch (1994; see also Tsay 1996) demonstrated through an acoustic phonetic experiment that the first formants are identical in basic high vowels and in high vowels that participate in the tone-induced vowel height alternations. The ultimate motivation thus appears to be physiological, but the pattern itself is psychological, and so should be represented in the grammar. Moreover, there are reasons to suppose that perception played a crucial role in the phonologization process in Fuzhou. Tsay and Sawusch (1994) examined the effect of F0 on F1 in Mandarin and English, and found that, like Fuzhou, Mandarin showed vowel raising in syllables with lexical high tones. Interestingly, while this raising was not categorical (i.e. raised mid vowels were still lower than lexical high vowels), it appeared to be larger than what is predicted by the purely physiological model of Honda, Hirai, and Dang (1994). That is, Mandarin speakers exaggerate the physiological effect; since they expend extra energy to do this, the motivation for the exaggeration must be perceptual, not articulatory. By contrast, Tsay and Sawusch (1994) were unable to find any effect of pitch production on the height of English vowels. They speculate that the relevant difference between Mandarin and English is that the former is a tone language, which means that speaker-hearers have to be able to clearly distinguish tone categories. Exaggerating the vowel-raising effect provides an extra cue to tone category that listeners can pick up on. In English there are no lexical categories defined by pitch, so there is no reason to provide extra cues to the raising of
pitch. The fact that no effect was found in English either means that a small effect was present but was missed, or more intriguingly, that the physiological effect is actively suppressed in English.

In order to keep the discussion within the scope of this paper, here we will examine schematic versions of the two types of patterns of interest: the cross-linguistically common raising of tone after voiceless consonants, and the cross-linguistically attested but rarer raising of vowel height with high tones. Functional Phonology is responsible both for accounting for these pattern's emergent categoricality and for encoding their physical motivation (including their difference in relative strength). The first is handled by the perception grammar, and works essentially as in the example described in an earlier section. The second is handled by the production grammar, and this is what we will focus on here.

In our analysis, the articulation of voicelessness, pitch raising, and tongue body raising are restricted by the families of *GESTURE constraints represented in (18a-c), respectively. Notice that we do not assume that tone and voicing articulation involve precisely the same articulators or gestures. Following Halle and Stevens (1971) and Bao (1990), we assume that the articulator for voicelessness on consonants is the cricothyroid cartilage, whose relevant gesture is vocal cord stiffening, but the articulation of high tone on vowels is performed by the larynx as a whole, which can accomplish this by a variety of gestures, including rotating the cartilages that compose the larynx, stiffening the vocal cord muscles internally, raising the larynx with extrinsic muscles, or changing subglottal air pressure through other means (see e.g. Ohala 1973, 1978, Schneiderman 1984). Note also the quantitative parameter that will be used in the analysis, namely "distance" (degree of activation of the gesture).

(18) a. voicelessness:  *GESTURE (cricothyroid: vocal cord stiffening / distance)
   b. high tone:  *GESTURE (larynx: pitch raising / distance)
   c. tongue height:  *GESTURE (tongue body: raising / distance)

The interaction of the gestures of distinct articulators is not examined by Boersma (1998). However, the insight to be captured is exactly parallel to the analysis of ease of articulation when only one articulator is involved. Just as it is easier to articulate a slightly raised low tone between two high tones than to articulate a very low tone in that context, it is likewise easier to articulate a high vowel when high tone is simultaneously articulated, though this effect is not as strong as the effect that the articulation of voicelessness has on the articulation of high tone. To put this in the negative terms required by articulatory constraints in Functional Phonology, it is harder to produce a gesture with a lesser distance in the presence of another gesture with a greater distance. To express this, we will adapt the articulatory coordination constraint mentioned briefly by Boersma (1998:156) to include mention of the distances employed by the two gestures, as shown below.

(19) *COORD(gesture₁/distance₁, gesture₂/distance₂):
   The two gestures gesture₁, produced with distance₁, and gesture₂, produced with distance₂, are not coordinated.

Boersma (1998) notes that articulatory coordination has to be actively learned, so initially all such constraints are undominated, a finite number being demoted during the course of acquisition. By including the distance measures, we can also propose the following general ranking principle for the constraints themselves. The principle claims that all else being equal, it is preferred that the sum of the distances of two coordinated gestures to be greater. This captures the insight that coordinated gestures tend to piggyback on one
another as part of the general trend to reduce articulatory effort.

(20) \( ^*\text{COORD}(\text{ges}_1/d_1, \text{ges}_2/d_2) >> ^*\text{COORD}(\text{ges}_1/d'_1, \text{ges}_2/d'_2) \iff d_1 + d_2 < d'_1 + d'_2 \)

Of course, the ranking of these coordination constraints vary depending on what the gestures are. Partly this ranking will be language-specific, as shown by the difference between Mandarin and English in their recognition of pitch influences on vowel height. Yet we must also capture universal physiological motivations as well. In our formalism, this is expressed by patterns in the ranking of \(^*\text{COORD} \) relative to \(^*\text{GESTURE} \). For pairs of gestures with stronger physiological connections, \(^*\text{COORD} \) constraints will tend to be ranked higher than \(^*\text{GESTURE} \) constraints, while the reverse is true for pairs of gestures with weaker physiological connections. We can illustrate this by comparing a constraint ranking that results in a large effect of voicelessness on pitch articulation (as in the many languages where tones are raised on a vowel after a voiceless consonant), with a ranking that results in a small effect of pitch on vowel height (as in Fuzhou and a small number of other languages).

For clarity of exposition, we use only three levels of tone (L, M, H) and three levels of vowel height ([æ], [e], [i]), and assume that the lowest forms of the pitch and tongue body gestures require the least energy (more realistically, the centralized forms should be easiest). The principle in (20) then requires the rankings in (21a,b); the rankings in (21c,d) follow from the principle for ranking \(^*\text{GESTURE} \) constraints given in section 2.2.

(21) (Note: "C" = "COORD", "G" = "GESTURE", and the symbols within the "()" are articulations; e.g. (p,L) represents the coordination of a slack articulation sufficient to produce a voiceless plosive and a pitch articulation sufficient to produce a low tone).

a. \(^*\text{C}(p, L) >> ^*\text{C}(p, M) >> ^*\text{C}(p, H) \)
b. \(^*\text{C}(H, æ) >> ^*\text{C}(H, e) >> ^*\text{C}(H, i) \)
c. \(^*\text{G}(H) >> ^*\text{G}(M) >> ^*\text{G}(L) \)
d. \(^*\text{G}(i) >> ^*\text{G}(e) >> ^*\text{G}(æ) \)

Full rankings that achieve the desired difference are illustrated in the following tableaux.

(22)

<table>
<thead>
<tr>
<th></th>
<th>(^*\text{C}(p, L))</th>
<th>(^*\text{C}(p, M))</th>
<th>(^*\text{G}(H))</th>
<th>(^*\text{G}(M))</th>
<th>(^*\text{G}(L))</th>
<th>(^*\text{C}(p, H))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{pa}^L)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{pa}^M)</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{æ}^L)</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(^*\text{G}(i))</th>
<th>(^*\text{C}(H, æ))</th>
<th>(^*\text{C}(H, e))</th>
<th>(^*\text{G}(e))</th>
<th>(^*\text{G}(æ))</th>
<th>(^*\text{C}(H, i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{æ}^H)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{e}^H)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{i}^H)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The crucial difference lies in the position of the highest-ranked \(^*\text{GESTURE} \) constraint relative to a mid-ranked \(^*\text{COORD} \) constraint: in (22a) \(^*\text{C}(p, M) >> ^*\text{G}(H) \), while in (22b) \(^*\text{G}(i) >> ^*\text{C}(H, e) \). To make this ranking difference follow from physiology, we need to invoke the general \(^*\text{EFFORT} \) ranking constraint: regardless of their details, articulatory
constraints are ranked by the degree of effort involved in the articulations that they represent. Based on our reading of the tone production literature, we hypothesize that it requires more effort to produce a mid tone with a voiceless plosive than it does to produce a high tone in general, but more effort to produce a high vowel in general than to produce a mid vowel with a high tone. Detailed phonetic work (involving empirical experimentation and ideally also computer modeling) is necessary to know if this hypothesis is correct. However, if it is, the necessary rankings follow automatically from the *EFFORT ranking principle, so automatically that if the physiological hypothesis is wrong, the current analysis must be rejected as well.

The difference in degree of effect established by such rankings then results in tone-voicing interactions being cross-linguistically more common than tone-vowel height interactions. This is because the perception grammar will tend to recognize acoustic forms as the nearest categorical prototype, as shown in an earlier section, which means that smaller acoustic effects will tend not to find their way into the underlying representations.

We have shown, then, that the way that tone interacts with other phonological properties should not be taken to demonstrate the inherent representational nature of tone. Instead, what is needed is a general characterization of these interactions that is capable of predicting which interactions will be favored, and to what degree. This is what is provided by our Functional Phonology proposal.

4. The autosegmental behavior of tone

We now move beyond tone features themselves to their behavior in phonological processes. Our overall theme is that differences between articulatory and perceptual teleologies have an important effect on the sorts of behaviors exhibited by various tone processes. In 4.1 we show that explicit recognition of tone spreading as having an articulatory motivation helps account for cross-linguistic restrictions on the autosegmental behavior of contour tones. In 4.2 we demonstrate a contrast between tonal assimilation and dissimilation processes, in that the former typically respect the laryngeal articulations that cause them, while the latter reflect their origins in pitch perception. In 4.3 we show that differences in the behavior of tonal domain-span vs. domain-límit rules (in the terms of Nespor and Vogel 1986) also derive respectively from differences between tonal articulation vs. perception.

4.1 Contour tones

One issue important in the history of tone theories is the nature of contour tones. While early theories represented them in a unary fashion (e.g. the [rising] and [falling] features of Wang 1967), today it is standard to treat them as sequences of level tones. This accounts for the well-known behavior of contour tones in African languages, which are readily split into or derived from sequences of level tones through the concatenation of morphemes, deletion or insertion of vowels, or other processes affecting the number of tone-bearing units (see e.g. Odden 1995 for a recent review). Yet it has long been unclear if the hypothesis is necessary for Asian tone languages, in which contour tones tend to preserve their shape from the underlying form, rather than splitting up or being derived (see e.g. Yip 1995 for a recent review). In this section we will not address these differences in tonal typology, but instead focus on properties of contour tones that all tone systems seem to obey. In particular, contour tones do not seem to spread as wholes. This is unsurprising for African languages, in which contours clearly behave like sequences of level tones, but from a purely autosegmental perspective it is unexpected for Asian languages, given the general
tendency of such languages to treat tones as wholes. However, this observation has a straightforward explanation if the articulatory teleology of autosegmental spreading is recognized, as can be formalized in Functional Phonology.

Before getting to the explanation, we must first justify our claimed observation. Bao (1990), Yip (1989, 1995), and Chen (2000) all argue that there are in fact examples in Sinitic languages showing the spreading of whole tones and/or of contour alone (i.e. tonal shape independent of specific tone height); the languages that have been claimed to have such patterns are Zhenjiang, Wenzhou, Changzhi, Danyang, and Zhenhai. However, even after years of debate no consensus has emerged over the validity of the proffered examples. Thus Yip (1989, 1995) challenges Bao's (1990) analyses of contour spread in Zhenjiang and Wenzhou, while Duanmu (1990, 1994) has carefully demolished all cases offered by Bao and/or Yip (including an earlier version of Yip 1995), i.e. Zhenjiang, Wenzhou, Changzhi, and Danyang. Even Chen (2000:76) notes that "[u]ncontroversial cases of contour spread, with or without involving the register [i.e. tone height -- JM&JT] are indeed hard to find," citing with apparent approval Maddieson's (1978) earlier claim that no rules of contour tone assimilation have been found. Controversy is inevitable because all putative examples of contour spreading are complex enough (interacting with stress and the irregular tone alternations typical of Asian tone sandhi) that alternative analyses are always available.

To provide some flavor of the ambiguity of the relevant phenomena, consider the following Zhenhai data from Rose (1990) showing tone sandhi in compounds with a weak-strong stress pattern (as far as we know not yet discussed by Bao, Yip or Duanmu). Chen (2000) argues that they demonstrate the spreading of contour independent of tone height. The digits below the segments represent tone contours, and can be considered quite reliable given the careful acoustic studies of Rose (1990).

(23) (Cited in Chen 2000:68)

   a. fû.ke  "bedroom"
      213.441  base tone
      11.334  sandhi form

   b. têju nîq  "last year"
      213.231  base tone
      11.24  sandhi form

   d. meî khwâ"coal mine"
      231.231  base tone
      11.441  sandhi form

In Chen’s analysis, the contour of the unstressed (first) syllable delinks, leaving it a level tone. Then the contour links to the stressed (second) syllable. However, the actual pattern does not seem sufficiently compelling to force either Yip or Duanmu to revise their respective positions. One obvious problem (common to all analyses of Asian tone patterns, in our opinion) is that it is very difficult to distinguish phonetic detail from what is claimed to be the underlying phonological pattern. For example, crucial to Chen's analysis of (23a) is that the [213] of the first syllable of the base form has precisely the same contour as the [334] of the second syllable of the sandhi form. Phonetically it certainly doesn't. The contour [213] begins with a slight drop not found in [334]; this cannot be dismissed as a purely physiological dip towards a low-tone target, since the sandhi form of the first syllable is transcribed as [11], not [21], showing that the dip in [213] is intentional. Moreover, the
contour [213] involves a steep rise (from a short [1] to [3]), whereas the rise in [334] is more gradual (from a long [33] to [4]). Chen completely ignores these phonetic details in the quest for a general description, even though it is presumably only the phonetic detail that is capable of arguing for or against a phonological analysis.

Another, somewhat less obvious problem is illustrated by (23c), in which [441] appears as the sandhi form of the second syllable, even though this corresponds neither to the precise contour of the base form first syllable [231] nor the tone height of the base form second syllable [231]. Instead, it precisely matches the lexical tone [441], seen for example in the base form for (23a). While most sandhi forms in Zhenhai seem to be nonlexical, the appearance of [441] here indicates that Zhenhai sandhi is at least partly precompiled, to use the term of Hayes (1990). That is, in at least some alternations it involves the substitution of one memorized tone with another, rather than deriving one from the other through general rules or constraints. Some Sinitic tone sandhi systems are entirely of this sort; for example, Tsay and Myers (1996) demonstrate that all tone sandhi alternations in Southern Min are precompiled, arguing further that searching for general patterns in its system is fruitless, since base and sandhi forms are associated in entirely arbitrary ways (though in principle explainable from a historical perspective). The lexicalized nature of tone sandhi also appears relevant in Changzhi and Danyang. Duanmu (1994) argues that the putative whole-tone spreading in Changzhi is more perspicuously analyzed as morphological tone reduplication, given that it affects precisely two morphemes (and they happen to be function morphemes, which often have peculiar tonal behavior in Sinitic languages). Similarly, Yip (1995:483) acknowledges that the "historical origins" of the key morpheme is relevant in the Danyang pattern, which we take to imply that it too is at least partly precompiled. Chen (1996:36) himself notes elsewhere that "apparently random tonal substitutions" are "quite common among Chinese dialects".

We conclude that such irregular phenomena should be handled in the lexicon, rather than in the perception or production grammar per se. As noted in section 2, Myers (in press) provides a formally explicit OT model of how lexicalized patterns behave, arguing that they spread and interact through competition among lexical items, rather than being imposed on the lexicon by forces from without, as happens in the functionally motivated phonology we have looked at up to this point. In any event, it has been far from conclusively demonstrated that contours can spread as wholes in Asian languages, and they certainly do not in other tone languages.

The observation established, we now turn to the explanation. Why can't a tone contour spread as a unit? Our answer has two parts. First, autosegmental spreading arises through the phonologization of coarticulation, i.e. the anticipation or perseveration of a gesture. Among other things, this hypothesis explains why the features whose autosegmental spreading is most widely attested (e.g. tone, nasality, vowel height) are associated with gestures that have been shown to be involved in long-distance coarticulatory effects due to the relative independence and/or slowness of their articulators. By contrast, the features that spread least readily, if at all (e.g. consonantality and sonorancy; see McCarthy 1988), are not associated with well-defined gestures of any sort (Kaisse 1992 makes the same observation). The second part of our answer concerns the behavior of a contour tone undergoing coarticulation. If one extends the duration of the gesture(s) associated with a contour tone, the result is either a contour with a longer duration relative to other gestures, or a contour tone that begins or ends with an extended level tone; it is impossible for repetitions of contour tones to arise through coarticulation. Putting the two points together, it is therefore impossible for the apparent autosegmental spreading of contour or whole contour tones to arise. What may easily arise is exactly what is widely attested: the spreading of one end of a contour, realized as a sequence of level tones, or the extension of a contour across a
wider domain (e.g. in Shanghai; Duanmu 1997).

To capture this analysis in terms of Functional Phonology, we merely have to build on
the schematic production grammar presented in our introduction to this theory in section 2.2. The same *GESTURE constraints, used there to reduce the degree of change in tone articulations across syllables, can be used here to account for tone spreading patterns. The relevant constraint families and their universal rankings are repeated below.

\[(24)\]

\[a. \quad \text{*GESTURE(distance = } d_1) \gg \text{*GESTURE(distance = } d_2) \Leftrightarrow d_1 > d_2\]

\[b. \quad \text{*GESTURE(velocity = } v_1) \gg \text{*GESTURE(velocity = } v_2) \Leftrightarrow v_1 > v_2\]

All else being equal, these constraint rankings favor gestures that move no distance at zero velocity, i.e. that are held in position (in Boersma 1998 they are countered by *HOLD constraints, not discussed here). Thus replacing one gesture with another is costly, as is changing the degree of a gesture (we assume that this is true even when reducing the degree of a gesture, since such reductions are typically controlled and thus take energy themselves). Unlike the case with the evaluation of many familiar OT constraints, these constraints do not incur more violations the more segments in a candidate output are produced with a forbidden gesture; the number of segments is not the crucial measure, but the number of gestures.

(Boersma 1998:448 avoids some of this discussion by positing a *MOVE constraint, but we are trying to restrict the proliferation of theoretical mechanisms as much as possible.)

The following tableau illustrates the effect of these structure constraints, free from the faithfulness force of *REPLACE. We assume as before that changing pitch height is the responsibility of the larynx as whole. Thus the contour tone in [pa\textsuperscript{51}] and the sequence of level tones in [pa\textsuperscript{5} tu\textsuperscript{1}] both involve the same distance parameter of a single dynamic gesture, differing only in velocity (faster vs. slower). The *GESTURE constraints may be violated multiply by a single candidate if a gesture changes more than once; for example, the candidate [pa\textsuperscript{51} tu\textsuperscript{51}] in (25h) receives three stars from the relevant constraints, since three changes are involved: from [5] to [1], then from [1] to [5], and finally from [5] to [1] again. The candidate outputs are listed vertically in order from most to least optimal; the focus here is not on picking a single optimal output, but showing how the constraints and their ranking impose a hierarchy on the candidates. Note that the output [pa\textsuperscript{51} tu\textsuperscript{51}], with a sequence of repeated contours, is the least optimal.

\[(25)\]

(Note: "G"="GESTURE", "d," represents the distance between the articulation of the pitch levels [e.g. 5-1=4], "v_y" represents the velocity needed to move between [5] and [1] in y half-syllable time intervals.)

<table>
<thead>
<tr>
<th></th>
<th>*G(d\textsubscript{1})</th>
<th>*G(d\textsubscript{2})</th>
<th>*G(v\textsubscript{1})</th>
<th>*G(v\textsubscript{2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>pa\textsuperscript{5} tu\textsuperscript{1}</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>pa\textsuperscript{5} tu\textsuperscript{1}</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>pa\textsuperscript{5} tu\textsuperscript{1}</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>pa\textsuperscript{5} tu\textsuperscript{1}</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>e</td>
<td>pa\textsuperscript{51} tu\textsuperscript{1}</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>pa\textsuperscript{51} tu\textsuperscript{1}</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>g</td>
<td>pa\textsuperscript{51} tu\textsuperscript{15}</td>
<td>**</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>pa\textsuperscript{51} tu\textsuperscript{15}</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
</tbody>
</table>

Languages may show any of these tone patterns on the surface if forced by *REPLACE constraints; even (25h) may appear if the two syllables each have a falling tone in their
underlying representations. The patterns in (25a-e) may also be chosen as optimal outputs for inputs not identical to their surface forms, in order to obey *REPLACE constraints. For example, given the input /pa⁵¹ tu⁵/, the ranking in (26a) chooses [pa⁵ tu¹] as optimal, while the ranking in (26b) chooses [pa⁵¹ tu¹].

(26) (Note: "R"="REPLACE", *R(aₓ,bᵧ) means "do not replace [a] with a duration of x syllables for [b] with a duration of y syllables").

<table>
<thead>
<tr>
<th>Input</th>
<th>*G(d₁)</th>
<th>*G(d₂)</th>
<th>*G(v₁)</th>
<th>*R(5₁,5₁)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[pa⁵¹ tu¹]</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[pa⁵ tu¹]</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[pa⁵¹ tu¹]</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A contour tone may also be derived, as [tu⁵¹] may be derived from /tu⁵ u¹/ through deletion of the second vowel, but the tonal coalescence must be motivated solely by faithfulness since this change increases articulatory complexity. For example, in the following tableau tonal coalescence is forced by *REPLACE.

(27)

<table>
<thead>
<tr>
<th>Input</th>
<th>*R(5₁,5₁)</th>
<th>*R(5₁₂,1₁)</th>
<th>*G(d₁)</th>
<th>*R(5₁₂,5₁₁)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[tu¹]</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[tu]</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[tu¹]</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

By contrast, the output [pa⁵¹ tu⁵¹] cannot be selected as an optimal output for an underlyingly simpler structure, e.g. /pa⁵¹ tu¹/, because there is no mechanism for evaluating the copying of [5₁] as a reduction in articulatory complexity or as the optimal way to preserve underlying structure.

Notice that the pattern [pa⁵¹ tu¹⁵] in (25g) also cannot be derived from input like /pa⁵¹ tu⁵/ by any ranking of *GESTURE or *REPLACE, since it neither makes the output easier to articulate nor preserves underlying information. The Functional Phonology analysis thus disallows a further subset of patterns that are uncontroversial even among tone theorists that reject the notion of tone contour units. While we have not yet carried out a thorough survey, we expect that any putative examples showing a pattern like (25g) will prove to be misanalyzed. For example, Chen (2000:64-67) claims that it occurs in another area of Zhenhai tone sandhi involving strong-weak stress patterns, but the evidence offered for his analysis is as ambiguous as the data discussed earlier: the base-sandhi alternation of [3₂₃,₄₄₁]-[3₃₄,₅₁] is analyzed as [MH.L]-[MH.HL], and [2₁₃,₂₁₃]-[₁₁₄,₅₁] as [LH.L]-[LH.HL] (the base form second syllable [L] having been filled in by default after deletion of the original tone). This means that [MH] is realized variously as [3₂₃] or [3₃₄], and [LH] as [2₁₃] or [₁₁₄], with no explanation offered for the lack of a match with the phonetic forms or for their variation. It seems likely that, like the other Zhenhai pattern, this pattern is
essentially lexicalized rather than being derived in the production grammar.

Properly speaking, the above analysis is a model of coarticulation, albeit under grammatical control, not of spreading in the classic autosegmental sense. The major difference between the two is that autosegmental spreading produces categories that are also used to make lexical contrasts. In the case of tone, this means that coarticulatory spread of an underlying /5/ tone across two syllables will actually result in an articulation something like [543.321] (where "." marks a syllable boundary), whereas the phenomenon typically analyzed as autosegmental delink and spread would result in an articulation of [5.1]. From a purely production standpoint, [5.1] will always be disfavored relative to [543.321], since the former involves a much higher gesture velocity. To complete the analysis, then, the perception grammar must also play a role, converting [543] into /5/ and [321] into /1/ in essentially the same manner as formalized in section 2.2. This process, first occurring sometime during language acquisition, and then reinforced throughout adulthood, would increase the categoricality of the underlying forms involved in the production grammar. Then, through the faithfulness force of the *REPLACE constraints, the articulatory forms would themselves become much more categorical than "pure" coarticulation.

4.2 Assimilation vs. dissimilation

In this section we look at another prediction of the hypothesis that phonological tone spreading originates in the production grammar, namely that tone spreading respects the articulatory constraints on tone production. We then review arguments that dissimilation is motivated by perception rather than production, and consequently, we predict that tonal dissimilation processes respect constraints on pitch perception rather than production. This correlation between the extragrammatical motivations of assimilation versus dissimilation and their respective grammatical behaviors is a strong argument in favor of a functional approach to phonology.

4.2.1 Assimilation and articulatory gestures

There are two important pieces of phonological evidence that demonstrate the role of articulation in tone spreading (in addition to the arguments given in section 4.1). The first is that tone spreading may be blocked by segments with incompatible laryngeal properties, and the second is that tone spreading respects articulatorily defined natural classes of tones. As an example of the first sort of evidence, Odden (1995) discusses the following pattern in Bade (data from Schuh 1978), which shows that spreading of a high tone does not occur across a voiced obstruent. As we have pointed out, the articulation of voiced obstruents is similar to the articulation of a low tone, but this is not true of sonorant consonants, whose voicing is not produced through an explicit articulatory command but instead occurs spontaneously due to their less constricted vocal tract configuration (Chomsky and Halle 1968). Hence it makes articulatory sense to find blocking of low tone across voiced obstruents, but not across sonorant consonants. An analysis using ranked *COORD constraints, along the lines of those in section 3.2, would formalize this pattern rather straightforwardly.

(28) (Note: Due to font problems, tone diacritics have been replaced with "H", "L".)

   a.  nənH kəH tawH → nənH kəH tawH "I returned"
   b.  nənH laH wawH → nənH laH wawH "I ran"
   c.  nənH gaH fawH → nənH gaH fawH "I caught"
Nevertheless, as Odden (1995:452) notes, "[p]honological tone-consonant interactions are actually rare; there are many languages where the phonatory features of consonants are transparent to tone spreading, and only a few where they are not" (though some of these few, or cases like them, are found in Asia, not an area where tone spreading is common; see Yip 1995:485). Odden's observation would be a serious challenge to our claims if the only way to account for it were to posit abstract tone features whose behavior was completely unrelated to their physical implementation, especially if their spreading behavior did not relate in any way to their articulation. However, this abstract hypothesis is both unnecessary and problematic. It is unnecessary because, as we demonstrated in section 3.2, we do not need to assume that tone and voicing involve precisely the same articulators and gestures in order to describe their interaction in a principled way. In fact, it appears that the articulation of voicing involves a much more restricted set of gestures than does tone articulation, or at least a much more restricted range of articulator movement (the reader can confirm, for example, that articulating a contrast in low vs. high tone produces much greater movement of the larynx than does a contrast between a voiced [z] vs. a voiceless [s]). Acoustically, voicing only creates rather small and short-lived disturbances in the overall pitch contour. Most importantly it appears that a high tone has little trouble spreading coarticularly across a voiced obstruent (e.g. the pitch of the second syllable in [a'lb'å] appears to be slightly higher than in [a'bå]). Since tonal coarticulation can occur across consonants, the seeds of voicing-insensitive tone spreading can indeed be planted in the way that our model requires.

Hypothesizing entirely abstract tone features is also problematic, because this would miss an important generalization: there is a direct correlation between the degree of the articulatory "unnaturalness" of tone-voicing interactions and the degree of their lexicalization, as measured by independent diagnostics. Odden (1995:452) makes a related observation, noting that in many languages when tone-voicing interactions occur they are restricted to "later in the grammar", but this makes it appear as if they are therefore more due to physics than psychology. We have already discussed the fallacy of searching for sharp lines between psychology and physics in phonology, but more importantly, from our experience with phonological patterns "earlier in the grammar" generally means "more lexicalized" and therefore of less, not more, phonological interest (see e.g. Myers 1993 for extensive arguments and illustrations).

This correlation between unnaturalness and lexicality in tone patterns can be illustrated by visiting a few points along the scale. At one extreme is the Bade pattern shown above; fully consistent with articulation, it properly treats voiced obstruents and sonorant consonants as separate classes. It also appears to be a classical postlexical process, since the spreading occurs across words and produces a nonlexical downstep (indicated by ['] in (28a); see Pulleyblank 1986 for discussion of downstep as a postlexical process). A case somewhere further along the scale is found in Nupe, another example cited by Odden (1995) as showing tone-voicing interactions. The data below show that a rising tone only appears when preceded by a low tone and an intervening voiced consonant, whether obstruent or sonorant. As Kenstowicz and Kisseberth (1979:267-8) show, this can be analyzed as autosegmental spread of the first syllable's low tone rightward, but only across voiced consonants. This time voiced obstruents and sonorants behave as a class in spite of articulatory differences in their articulation of voicing. Note also that the alleged spread results in a configuration shown in the previous section to be articulatorily disfavored (i.e. an increase in articulatory complexity from [L.H] to [L.LH]), further suggesting that the pattern is no longer actively applied in the production grammar.
As we predict, this pattern is also more lexicalized than the one in Bade: all of the examples cited by Odden (1995) and Kenstowicz and Kisseberth (1979) involve the same nominal prefix [eL]. Apparently it does not apply across word boundaries and is not triggered by other morphemes. This process thus appears to be of the same quasi-morphological nature as Changzhi tone sandhi. It is not structure-preserving (i.e. the rising tone is apparently only found in these predictable contexts), but it is known that not lexical processes are (see e.g. Borowsky 1993).

Another example of a tone pattern that treats voiced obstruents and sonorants as a class (though not one that necessarily requires a spreading analysis) is found in Mandarin, which disfavors a high level tone after these classes, but not after voiceless obstruents. This can be seen by the low number of syllable types in the shaded cells in the following table (statistics from Ho 1976). The table shows that the pattern may also be articulatorily unnatural in another way: the falling tone, which like the high level tone begins with a high pitch, is not disfavored by a voiced onset. Yet the existence of exceptions (including high-frequency words like [ma55ma55] "mother") clearly show that the pattern is lexical, and hence less interesting from the perspective of Functional Phonology. Another possible similar example is Wujiang, mentioned by Yip (1995:486), which like Mandarin treats voiced obstruents and sonorants as a class in a pattern restricted to the making of lexical contrasts.

(30)

<table>
<thead>
<tr>
<th>Onset</th>
<th>Tone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[55]</td>
</tr>
<tr>
<td>p</td>
<td>55</td>
</tr>
<tr>
<td>pʰ</td>
<td>39</td>
</tr>
<tr>
<td>t</td>
<td>53</td>
</tr>
<tr>
<td>tʰ</td>
<td>38</td>
</tr>
<tr>
<td>s</td>
<td>67</td>
</tr>
<tr>
<td>m</td>
<td>8</td>
</tr>
<tr>
<td>n</td>
<td>3</td>
</tr>
<tr>
<td>l</td>
<td>7</td>
</tr>
<tr>
<td>z</td>
<td>1</td>
</tr>
</tbody>
</table>

The other extreme of the correlation between unnaturalness and lexicalization is illustrated by a final example from Odden (1995:452). Kanakuru shows a tone-voicing interaction that is precisely the opposite of the one expected from articulatory considerations: verb stems beginning with voiced obstruents have a tonal melody beginning with a high tone (HL), while those beginning with voiceless obstruents have the tone pattern LH. Moreover, these melodies are laid down on the word from right to left, against the near typological universal of left to right spreading (e.g. Odden 1995:460). This further observation demonstrates both the marked behavior of this tonal pattern as well as suggesting that the processes that gave rise to it involved something other than ordinary coarticulation, since anticipatory (leftward spreading) coarticulation requires a greater degree of planning (i.e. intentional control) than perseverative (rightward spreading) coarticulation. Given the
unnaturalness of the pattern, we expect that the selection of tonal melody must be a lexical process, and indeed it is: in words not beginning with obstruents the choice of HL vs. LH is entirely unpredictable.

We conclude from this survey that in spite of its relative rarity, the interaction of voicing and tone in tone spreading processes provides further evidence that the nature of these processes is best understood when their articulatory origins are explicitly recognized. The same conclusion follows from the observation that phonological tone spreading respects articulatorily defined natural classes of tones, to which we now turn.

Raising pitch above the baseline and lowering pitch below it involve distinct laryngeal-internal articulators and gestures (Ohala 1978). In some languages these different articulations reveal themselves acoustically by the addition of distinct phonation on the lower tones, commonly breathy and/or creaky voice (see Yip 1995:486-7). Such observations have led several tone theorists (in particular, Yip 1980, Bao 1990, and Duanmu 1990) to use the term register to refer both to these laryngeal differences (in vowel phonation and consonant voicing) and to pitch height. The consensus over register spreading independently of the whole tone is precisely the opposite to that of contour spreading: the only major theorist who rejects it is Yip (1989, 1995). The existence of distinct articulators for different registers does imply that register could be involved in coarticulatory effects independent of other aspects of tone production, consistent with our functionalist approach. However, even its proponents admit that register spreading is relatively rare. Duanmu (1990) suggests that this results from the blocking caused by the laryngeal properties of intervening consonants, a suggestion also consistent with our viewpoint, even if we reject his further claim that register and voicing are articulated precisely the same way. Another possible explanation of the relative rarity of register spreading will be given below.

Nevertheless, relatively unambiguous examples of register spread do exist. Odden (1995) cites the example of Ewe (see also Kenstowicz 1994:242-3, using data from Clements 1978). As the data below show, mid tones (i.e. the second level, counting from the bottom) become superhigh tones (i.e. the fourth level) when surrounded on both sides by high tones (i.e. the third level). This superhigh tone may optionally spread to the preceding high, as shown in (31d).

(31) Kenstowicz (1994:243); data from Clements (1978); due to font problems, the four phonetic tone levels are marked 1, 2, 3, 4 rather than with the original diacritics.

```
a. [a'kplo2 o3]    "spears"
b. [a'kplo2 me3gbe3]    "behind a spear"
c. [e'kpe3 o3]    "stones"
d. [e'kpe3 me3gbe3] or [e'kpe4 me4gbe3]    "behind a stone"
```

Note that this process applies across word boundaries, and it also creates a nonlexical output: in Ewe only low, mid, and high tones are contrastive. The pattern is also sensitive to syntactic constituents, which is often a diagnostic of precompiled phrasal phonology (see Hayes 1990), but it may also indicate merely the presence of prosodic structure intervening between the syntax and fully productive patterns. Thus it appears that the production grammar should be responsible for this pattern. Yet raising to superhigh tone overshoots what would seem to be the simplest articulation, namely a steady high tone across a three-syllable span.

An articulatory solution to this apparent problem is given some plausibility by Odden's (1995) purely formal analysis. He suggests that if the mid tone is considered to be of low register, but with a higher pitch (encoded as a separate feature) than low tone, this pattern can
be understood as register spreading from the high tone. This will force the mid tone to skip a level, as the high register is added to the high pitch of the mid tone. It is possible to make sense of this in physiological terms if register contrast and fine pitch discrimination within registers involve distinct articulators, and Bao (1990) has summarized evidence for just this claim. Specifically, he implicates the cricothyroid cartilage as the articulator involved in making register contrasts (i.e. changes in tone above and below the resting articulation), and the vocal-cord-internal vocalis muscle as the articulator responsible for fine-tuning tone. Thus if this physiological hypothesis is correct, the production of the superhigh output preserves the vocalis articulation unchanged while spreading the cricothyroid articulation across it. Of course, given the fact that, as we have emphasized, the larynx has numerous options for producing tone changes, the particular gestures that we assume in order to understand the Ewe pattern are not universal but must be learned by Ewe speakers, as encoded by the *COORD constraints. The fact that not all languages need articulate tones precisely this way may go part of the way towards explaining why register spread is relatively rare: not all languages treat register as phonologically relevant to tone in the production grammar.

Before giving a Functional Phonology analysis of the Ewe pattern, we must face a serious technical problem. Crucial to our articulatory interpretation of Odden's analysis is the assumption that the mid tone in Ewe is normally produced with a "low" gesture of the cricothyroid but a "high" gesture of the vocalis. In principle, since Ewe is a three-level system, the mid tone could have been articulated precisely the opposite way, i.e. "high" cricothyroid plus "low" vocalis. A speaker of Ewe must therefore know precisely how a mid tone is articulated, not just how it sounds. Yet Functional Phonology assumes that underlying representations are perceptual targets, not articulatory targets. There are several possible responses to this problem that do not require giving up this assumption. One response would be to suppose that one way to articulate a mid tone (i.e. the way found in Ewe) is inherently easier than the other way; the production grammar would then settle on it as the preferred way to achieve the perceptual target. If it turns out that languages can vary in how the mid tone is articulated, then Functional Phonology predicts that these mid tones are not in fact perceptually identical. For example, in one three-level language the mid tone may share breathy voice with the low tone while the high tone is clear, and in another language the mid tone may be clear like the high tone. Likewise, vocalis stiffness must produce an acoustic signature that can be recognized by the perception grammar even in quite distinct pitches, or else, more realistically, speaker-hearers associate two distinct acoustic forms with a single perception (through learning, or perhaps innately), as English speaker-hearers do with the silence and formant transition cues for the stop in words like *spill* (see section 2.2).

Regardless of precisely how this issue is dealt with, our analysis requires that in the mature Ewe grammar there must be a direct correspondence between the perceptual representation of mid tone and its articulation as "low" cricothyroid plus "high" vocalis, since the faithfulness constraints in the production grammar are motivated by perception, not articulation. In particular, the *REPLACE constraints must be able to recognize that both the mid tone and the superhigh tone are faithful to an underlying perceptual target associated with the same vocalis articulation. For our purposes here, we assume that there is an abstract perceptual feature associated with the articulation of the vocalis, called VP in the following tableau (for "vocalis perception"). The analysis then operates as illustrated, where [eˈkpe ˈmeːɡbe³] is selected as more optimal than [eˈkpe ˈmeːɡbe³] because the latter changes the value of the VP feature from the underlying /eˈkpe³ meˈɡbe³/ whereas the former does not (these assumptions are made explicit in the leftmost column of the tableau).
(32) (Note: the constraints here only evaluate the center two syllables, i.e. /pe me/; for *GESTURE constraints: L, H = low, high gestures, *G(X:Y) disallows change of X articulator to Y gesture over time; for *REPLACE constraints: H, L = high, low along a perceptual scale, *R(F0:X,Y) disallows replacing the indicated pitches)

<table>
<thead>
<tr>
<th></th>
<th>*R(VP:H,L)</th>
<th>*G(cric:L)</th>
<th>*R(F0:3,4)</th>
<th>*G(voc:H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[e'kpe'me'gbe'] {VP=H}</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[e'kpe'me'gbe'] {VP=L}</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[e'kpe'me'gbe'] {VP=H}</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>[e'kpe'me'gbe'] {VP=H}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this analysis, the trigger for changing the original mid tone of the targeted syllable is articulatory; in the above tableau, it is a *GESTURE constraint that rules out this original form, since it would involve a change from a high to a low cricothyroid articulation across the two middle syllables. The *GESTURE constraints have no complaint with the second candidate, but this is also ruled out by articulatory considerations, though necessarily indirect in the present formalism: the *REPLACE constraint disallows change in the original articulation of the vocalis. The two remaining candidates win because both obey the two highest ranked *REPLACE and *GESTURE constraints; whichever is chosen in any particular utterance depends on the particular ranking fixed on for the two lower ranked floating constraints (see e.g. Anttila 1997, Nagy and Reynolds 1997, Hayes and MacEachern 1998 for other uses of floating constraints in the analysis of phonological variation).

Summarizing the previous two sections, then, the phonological spreading of tone shows three characteristics that reveal its origins in articulation: contours cannot spread as wholes, tone spreading can interact with the laryngeal features of intervening consonants, and tone spreading respects articulatorily defined concepts such as register and vocal cord stiffness, such as making mid and superhigh tone a natural class in spite of their noncontiguity in pitch.

4.2.2 Dissimilation and perceptual features

By contrast with assimilation, it has been argued by several independent researchers that local dissimilation has a perceptual rather than articulatory teleology. Ohala (1986) provides evidence that dissimilation can arise when listeners misperceive an intentional articulatory gesture as mere coarticulation and remove it from the underlying representation. This hypothesis explains why dissimilation tends to involve features that can spread through coarticulation, while the hypothesis that reconstructions of perceptual representations are involved explains why dissimilations are always lexical processes (Kiparsky 1986, 1995). Frisch (1996) provides further evidence that dissimilation involves lexical factors, demonstrating the success of a model premised on these factors at accounting for OCP effects in Arabic. Boersma (1998) also argues that the OCP originates in the perception grammar, pointing out for instance that breaking up sequences of identical segments after morphological concatenation adds to articulatory cost but benefits the listener. The prediction of Functional Phonology, then, is that tonal dissimilation patterns should not respect articulatory factors, but rather perceptual factors.

The model thus predicts that the production grammar should be able to target classes of tones produced with distinct articulatory gestures as long as they are acoustically similar. For example, in the four-level tone system of Nikki Bariba (Welmers 1952), the two mid tones behave as a class in the dissimilation process illustrated in (33) because they are perceptually similar (purely psychoacoustically). This is so in spite of the fact that these two tones do not form an articulatorily coherent class, since according to the articulatory
model discussed in 4.2.1, the tone transcribed as 2 may be articulated with a low cricothyroid gesture and high vocalis gesture, while the tone 3 would have precisely the opposite articulation.

(33) (Note: digits are used in place of the nonstandard diacritics used in Welmers 1952)

a. go'na2 "guinea fowl"
   go'na3 ye'ni4 "that guinea fowl over there"
   bi'i2 "child"
   bi'i3 we'ni4 "that child over there"

b. na3 ta3 su3 duu2 ra2 "I planted yams"
   na4 ra1 ta4 su4 duu1 re1 "I plant yams"
   na3 boo3 wa3 "I saw a goat"
   na3 boo3 wa4 k21 "I saw a goat" (a different usage)

Not only would a tone theory that relies solely on articulatory features be forced to treat these alternations with two separate rules, but neither of them would provide much insight into the underlying processes. The only mechanism for dissimilation available in autosegmental phonology is OCP-triggered deletion/delinking, which is not relevant for this case, since here the dissimilation involves shifting tones; Tsay (1994) argues that such tonal shifts require a new formal device allowing for movement up and down the values of her scalar feature [Pitch]. By contrast, under an analysis that allows for perceptual factors to influence phonology, the pattern can be understood (and formalized with Functional Phonology). The dissimilatory nature of this process is key to the present analysis: unlike the raising of mid to superhigh in Ewe, there is no way to view the tone process in Nikki Bariba as assimilation, since the upward shifting makes the targeted tone more distinct from the following low tone. This means that carrying out the raising process increases articulatory cost, but this cost is offset by the fact that it makes the affected tone more perceptually distinct from its context. In the production grammar, then, the relevant constraints are the faithfulness *REPLACE constraints, not the *GESTURE constraints. The fact that the process is structure-preserving (as dissimilation generally is) is consistent with this hypothesis (note, for example, that unlike the tone-raising in Ewe, no tone is output beyond those already used to make lexical contrasts). Thus the dissimilation results from the fact that the production grammar here serves the lexical representations rather than imposing new structure on the outputs.

Nevertheless, the raised tones are of course not literally faithful to the inputs. To force the change, we need to exploit the context-sensitivity of *REPLACE, as indicated in its full form, repeated below.

(34) *REPLACE (feature: value1, value2 / condition / left-env _ right-env):
Do not replace, for a certain feature, a specified value (value1) with a different value (value2), under a certain condition and in the environment between left-env and right-env.

In the case of Nikki Bariba, the dissimilation is triggered by the perception of [2] and [3] tones as higher (closer to /3/ and /4/, respectively) in the context of an adjacent [1] tone than they would otherwise (tone, much more than other phonological categories, relies heavily on relative contrasts). This psychoacoustic factor can be encoded with the following constraint ranking, which works to favor outputs that increase the distance between the perceptual
feature value for a target sound and that of its environment.

(35) \*\text{REPLACE}(x, y_1 \text{ / env}=z_1) >> \*\text{REPLACE}(x, y_2 \text{ / env}=z_2) \iff |y_1-z_1| < |y_2-z_2|

The dissimilation pattern in Nikki Bariba thus receives the analysis indicated by the following tableaux, which show the dissimilatory raising for /2/. An input of /3/ is raised by the same amount through similar ranking of similar constraints.

(36) (Note: \*R(x,y/z) = \*\text{REPLACE}(x,y \text{ / env}=z); \*R(x,y) = \*\text{REPLACE}(x, y \text{ / env} = \text{isolation}).

\begin{tabular}{|l|l|l|l|l|l|l|l|}
\hline
/s2/ & \*R(2,2/1) & \*G(4) & \*R(2,3/1) & \*R(2,4/1) & \*R(2,3) & \*R(2,4) & \*G(3) & \*G(2) \\
\hline
\hline
\hline
2 & * & & & & & & & * \\
3 & & & & & * & * & & \\
4 & & & * & * & & & & \\
\hline
\end{tabular}

\begin{tabular}{|l|l|l|l|l|l|l|l|}
\hline
/s2.1/ & \*R(2,2/1) & \*G(4) & \*R(2,3/1) & \*R(2,4/1) & \*R(2,3) & \*R(2,4) & \*G(3) & \*G(2) \\
\hline
\hline
\hline
2.1 & * & & & & & & & * \\
3.1 & & & & & & * & * & \\
4.1 & & & * & * & & & * & \\
\hline
\end{tabular}

The present analysis simultaneously solves a variety of problems with purely formal autosegmental alternatives. First, no new formal device (e.g. Tsay's 1994 transposition rules) need be added to the repertoire of phonological computations; the independently motivated \*REPLACE constraints are sufficient. Second, dissimilation is explained rather than merely described, since employing context-sensitive \*REPLACE constraints captures both the difference-enhancing effect of context in perception and the association of dissimilation with lexicality (since the perception grammar creates the lexical representations, and \*REPLACE constraints enforce articulatory obedience with lexical representations). Third, and most importantly, the analysis explains why it is possible for dissimilatory patterns like that in Nikki Bariba to completely overlook the fact that tones /2/ and /3/ do not form an articulatory natural class, since here they are being treated as perceptual, pitch-based entities rather than articulatory instructions.

4.3 Tone and prosodic structure

As a final argument for our formal functional model of tone, in this section we briefly examine Nespor and Vogel's (1986) typology of prosodically sensitive rules, arguing that domain-span rules are articulatorily motivated while domain-limit rules are perceptually motivated; this contrast then accounts for observed differences in the behavior of tonal phenomena. Standard theories of tone do not predict any particular relationship between they way tone interacts with prosodic structure and the relevance/irrelevance of laryngeal features, and thus miss important empirical observations.

According to Nespor and Vogel (1986:15), domain span rules and domain limit rules have the structures indicated below.
Typologies are only theoretically relevant if they make empirically testable predictions, and in our version of functional phonology, this distinction is indeed useful. This is because from a functional perspective, it is relevant to the structure of phonology that speakers and listeners use prosodic constituents differently. For speakers, prosodic constituents group articulations, whether phrases or utterances (between which the speaker can inhale) or syllables (which are sometimes viewed as articulatorily optimal units; see e.g. Levelt 1989:452-4). For listeners, by contrast, prosodic constituent boundaries are useful as pointers towards the boundaries of the constituents encoding the meaning of utterances (e.g. syntactic phrases or words; see e.g. Morgan and Demuth 1996). Hence speakers should tend to treat prosodic constituents as wholes, smoothing articulations within them but not across them, whereas listeners should be particularly sensitive to the edges of constituents rather than their middles. If these factors play a role in shaping the form of phonological systems, functionalist phonology predicts that domain span rules should tend to be articulatorily motivated, while domain limit rules should tend to be perceptually motivated.

We have already encountered some evidence for this in one of the tone patterns we have discussed: the assimilation in Ewe is both articulatorily motivated and a domain-span process. The latter point is demonstrated by the fact that it is blocked from applying across the left edge of a syntactic XP (which presumably defines the left edge of a prosodic domain); see Kenstowicz (1994:243, citing Clements 1978).

Perceptually motivated domain-limit rules are also found, such as that illustrated by a set of tone raising processes in Igede (Bergman 1971). While these processes have sometimes been analyzed as involving articulatorily motivated tone features (e.g. Yip 1980, Clements 1983), Tsay (1994) has shown that they do not unambiguously conform to the natural classes expected from the standard articulatory model of tone. Instead, separate rules are required, none of which do anything to decrease articulatory difficulty. The observation that the processes cannot be articulatorily motivated then correlates with the fact that all of them are triggered at a linguistically significant boundary. Hence they appear to be perceptually motivated domain-limit rules.

Igede is a four-level language; the tone raising processes involve them all, either as input or output. Tsay (1994) demonstrates that these processes cannot be reduced to a smaller set of rules because they differ from each other in unpredictable ways. What they have in common is that they raise tone at the right edge of a verb stem, specifically before what Bergman (1971) calls a "hyphen juncture" (before a pronoun). Otherwise they are different: one process raises the lowest tone [1] to [2] in monosyllabic verb stems (see (38a)), another raises tone [3] to [4] in disyllabic verb stems (38b), and a third raises [2] to [4] in both monosyllabic and disyllabic verb stems (38c-d). Moreover, the [2]-[4] alternations should probably be split into two, since according to Bergman (1971), the version found with monosyllabic verb stems has lexical exceptions (38e), while the version with disyllabic verb stems does not.
In a functionalist analysis, these patterns owe their similarities to a single motivating factor: the need to highlight the boundaries of linguistically significant units for the listener. The fact that this is accomplished consistently in Igede by tone raising rather than tone lowering is likely not a coincidence, as domain-limit tone rules often favor high tones, such as with utterance-internal phrase boundaries in English; even the fully lexicalized tone alternations in Taiwanese have a tendency to put the higher-toned allomorph at the edge of a prosodic group (Tsay 1994). Likely this is because higher pitches are perceptually more salient than lower ones (as reflected for example in the perception of music, in which the most prominent voice tends to be higher pitched than the others). This suggests the following ranking of context-sensitive *REPLACE constraints (where "middle" and "edge" refer to positions within a prosodic constituent).

(39) *REPLACE(pitch: x, y / env=middle) >> *REPLACE(pitch: x, y / env=edge) ⇔ y > x

Completing the analysis of the tone raising processes would then involve distinct rankings of further context-sensitive *REPLACE constraints, some referring to monosyllabic stems and some referring to disyllabic stems. In all cases, however, the crucial motivation comes from perception, not articulation. The hypothesis that perception plays a crucial role in the Igede patterns leads to the expectation that they are at least partly lexicalized, and this appears to be true: like tonal dissimilation in Nikki Bariba but unlike tonal assimilation in Ewe, the Igede patterns are structure preserving, and at least one of the alternations, as we have seen, has lexical exceptions.

Full justification of our claimed correlation between domain-span vs. domain-limit and the relevance or irrelevance of laryngeal gestures would require analyzing a further examples like Ewe and Igede, but we have as yet not come across any clear counterexamples to our key hypotheses.
6. Conclusions

As noted in the introduction, this paper presents its case for a formal functionalist model of tone from three perspectives: philosophical, empirical, and analytical. We first discussed philosophical arguments for rejecting criticisms of the general enterprise, arguing that given the interface nature of phonology and the obligation of generative theories to be explicit as possible, the most successful theories of phonology are those that spread their formalism as wide as possible, even into domains previously considered "mere phonetics". Empirically, we provided a wide variety of evidence from tonal patterns that emphasize the importance of extragrammatical factors in the form that phonological grammar takes. In particular, we showed that the number of tone levels is restricted by learnability, not by the form of the tone features; that tone-laryngeal interactions do not prove that tone features are laryngeal features, since these interactions are just part of a wider set that also includes interactions between tone and vowel height; that phonological tone spreading is articulatorily motivated, because contour tones cannot spread, tone spreading may be blocked by laryngeal features, and spread respects articulatory factors like register; that tonal dissimilation is perceptually motivated; and finally, that differences between domain-span and domain-limit rules result from the different roles that prosodic constituents play for speakers vs. listeners.

These disparate observations, many of which had already been made by Tsay (1994), were then given improved analytical rigor through the formal devices of Functional Phonology. In our proposed model, ease of learnability is formalized as ease of constraint demotion, as defined in a strict mathematical way; interactions of tone with other phonological features are handled by *COORD constraints, whose ranking reflects the physiological strength of the coordination; tone spreading is formalized through ranked sets of *GESTURE constraints, favoring short and/or slow changes in articulation, kept in check by *REPLACE constraints to prevent gestures from changing too much beyond their underlying targets; and dissimilation is handled by *REPLACE constraints, perceptually motivated and sensitive to tonal and prosodic contexts.

It must be admitted, however, that the formalism also faces some important challenges. Details of the learnability argument, for example, are still vague, as the search continues for the most accurate formal model of language acquisition within the Functional Phonology approach. The role of lexical factors (e.g. in the irregular tone sandhi patterns of Sinitic languages) is still not entirely clear, which creates the dangerous methodological temptation of treating the lexical module as a dumping ground for all recalcitrant "unnatural" patterns, when these patterns might actually represent decisive falsifications of the functionalist approach. This is of course precisely how the critics of functionalism see such patterns, and if functionalists cannot respond to their critics, they are not really engaged in science; a careful examination of unnatural phonology, as distasteful as it is to the functionalist soul, is therefore essential. The assumption of Functional Phonology that underlying representations are perceptually rather than articulatorily specified, while quite useful in many ways, also raises serious challenges. We saw this particularly in the discussion of register spread in Ewe, where our problem was not the spreading of register itself (perfectly natural given an articulatory view of spread) but rather the maintenance of the vocalis gesture, an articulation that apparently has no consistent acoustic correlate. Our response was to speculate that the vocalis articulation is perceptually recoverable by some other means, but this is not very satisfying without additional support for it; better would be if more careful examination of the Ewe pattern opened up an entirely different analysis. More generally, it may prove to be overly restrictive to claim that the input to the articulatory grammar consists solely of perceptual representations derived through acoustics; surely speakers also receive proprioceptive information about the motion and positions of the articulators themselves,
even within the larynx. If so, underlying representations could include some specific articulatory instructions as well, though these instructions may vary greatly across idiolects given the difficulty of learning them through acoustic cues. Finally, given its wide ambitions, the case for our tone model must remain quite incomplete in a paper as short as this. Much more work, involving the collection and analysis of many more tone patterns, both supportive and apparently problematic for a functionalist approach, remains to be done.

References


Ohala, J. J. (1986) "Phonological evidence for top-down processing in speech production."
Pilszczikowa-chodak, N. (1972) "Tone-vowel height correlation and tone assignment in the

Generative Grammar. Rutgers University and University of Colorado ms.


College-Hill Press.

Linguistics 9:247-283.


Broe and J. B. Pierrehumbert (eds.) Papers in Laboratory Phonology V: Acquisition and
the Lexicon, 313-334. Cambridge University Press.

K. Li, C-R. Huang, and Y-C. Lin (eds.) Proceedings of the International Symposium
on Chinese Languages and Linguistics 2, 389-408. Academia Sinica, Taiwan.


Seventh North American Conference on Chinese Linguistics and the Fifth International
Conference on Chinese Linguistics, 344-361.


Tsay, S. J. and J. R. Sawusch (1994) "The interaction of F0 and vowel quality in tonal

Yat-Sen Cultural Foundation: Taiwan.

Linguistics 33.2: 93-105.


Massachusetts.

dissertation, University of Massachusetts.

Garland in 1990 under the title The Tonal Phonology in Chinese.


Yip, M. (1995) "Tone in East Asian languages." In J. A. Goldsmith (ed.) The Handbook of