RELEVANCE OF PHONETICS IN PHONOLOGY

KEIICHI TAJIMA

Hosei University*


**Keywords:** Phonetics-phonology interface, phonetic scale, Optimality Theory, markedness

1. Introduction

Phonetics and phonology both study the sounds of language, so it is not surprising that one is relevant to the other. However, in reality, there seems to be a deep chasm that separates the two fields. For one thing, phonetics has sometimes been treated as “extra-grammatical” and outside of linguistics proper, while phonology is considered to be part of the grammar (Anderson (1981)). There are likely many reasons for this division, one of which is the mutual incompatibility between the continuous, gradient nature of phonetic phenomena and the discrete, categorical nature of phonological patterns. Another reason might pertain to the goals of each subdiscipline; the goals of phonetics are to understand how speech is produced and perceived, what its physical characteristics are, and how it relates to linguistic structure, while the goals of phonology are to make generalizations about language-general and language-specific characteristics of sound patterns, and how they relate to the rest of the grammar. Yet another reason might have to do with dif-

---

* I wish to express my gratitude to the editorial board for providing me an opportunity to write a review article for English Linguistics, and to two anonymous reviewers for their helpful comments and suggestions. All remaining errors and shortcomings are mine.


© 2007 by the English Linguistic Society of Japan
ferences in research methodology; doing phonetics often involves experimentation and analysis of scalar measurements of physical events, but phonology typically involves analysis of discrete linguistic entities based on careful use of inductive and deductive logic. Many attempts have been made in the past to bridge the gap between phonetics and phonology in different ways. These attempts include: natural phonology (Stampe (1973)), the laboratory phonology series (e.g. Kingston and Beckman (1990), Docherty and Ladd (1992), Keating (1994), Connell and Arvaniti (1995), Broe and Pierrehumbert (2000)), articulatory phonology (e.g. Browman and Goldstein (1990)), grounded phonology (Archangeli and Pulleyblank (1994)), functional phonology (Boersma (1998)), and attempts to incorporate speech perception into phonology (Hume and Johnson (2001)).

The present book, entitled Phonetically Based Phonology (hereafter PBP), can be seen as another attempt to overcome the phonetics-phonology divide. The volume is a collection of papers that propose explanations for various phonological phenomena based on constraints that are functionally motivated by mechanisms of speech production, speech perception, and lexical access. The linguistic analyses of most of the papers in PBP are couched within the framework of Optimality Theory (hereafter OT), a model of grammar first proposed in 1993 by Prince and Smolensky (1993), which has virtually replaced traditional rule-based derivational theories (Chomsky and Halle (1968)) and has quickly become a mainstream theory in generative phonology. The majority of the contributors to PBP in fact received their PhD’s in the 1990’s, and thus can be regarded as the new generation of scholars in this research area. Moreover, a handful of the contributors received their doctorate at UCLA, where the first and third editors of the volume hold teaching positions (the third author has moved elsewhere now). These contributors therefore to a certain extent share a common research agenda: to identify the phonetic basis for various phonological patterns. While the chapters treat diverse phenomena and have different perspectives, the commonality of research agenda is clearly evident, and the contributors generally do a good job convincing the reader that phonetics is highly relevant in phonology.

In the present paper, I will argue that PBP rightfully emphasizes the relevance of phonetics in phonological patterning, but I will also attempt to point out several shortcomings; in particular, PBP views the phonetics-phonology interface in a highly narrow-minded fashion, and it more-
over falls short of providing a predictive theory of which aspects of phonetics affect phonology. The present paper is organized as follows. Sections 2 through 6 will review and critique the eleven chapters in PBP, not in their order of appearance in the volume but according to the phenomena they cover. Section 7 will attempt to augment PBP's coverage of topics by presenting some experimental data, some of which are from my own research, on syllable structure as produced and perceived by second-language (L2) learners. Sections 8 and 9 will offer some general comments on the volume, followed by a brief conclusion.

2. The Phonetic Bases of Phonological Markedness

Hayes and Steriade's chapter (chapter 1, "Introduction: The phonetic bases of phonological markedness") presents the central issues addressed in the volume and provides an overview of the remaining chapters. The chapter starts by assuming that markedness laws that characterize the typology of sound systems are part of the linguistic knowledge of individual speakers, which takes the form of ranked and violable constraints in Optimality Theory. The authors then pose the question of where these markedness laws originate and how speakers come to gain knowledge of markedness constraints. The answer entertained by Hayes and Steriade, along with many of the contributors, is the hypothesis that "phonological constraints can be rooted in phonetic knowledge ... the speaker's partial understanding of the physical conditions under which speech is produced and perceived" (p. 1). That is, the search for universal markedness constraints governing sound systems can proceed deductively, by looking for clues in the knowledge that speakers have about how speech production, speech perception, and lexical access work. This is contrasted with an inductive approach to markedness, which proceeds by accumulating factual observations about languages and making generalizations about statistical or categorical tendencies.

As a specific example, the chapter discusses how the tendency for languages to avoid voiced obstruent geminates can be rooted in phonetic facts about the aerodynamics of voicing (instead of assuming an innate constraint that arbitrarily bans these sounds). Specifically, starting with the observation that voicing is more difficult to sustain in long obstruents than in short obstruents, and that voicing is more difficult to sustain in velar than alveolar obstruents, and more so in alveolar than bila-
bial obstruents,⁠¹ a phonetic scale, as in (1), can be constructed that ranks obstruents in order of voicing difficulty (from more to less difficult), and a constraint family with a fixed ranking, as in (2), can be generated based on (1):²

(1) *[@voice]: g: < d: < b: < g < d < b
(2) *g: » *d: » *b: » *g » *d » *b

The scale in (1), replicated from (5) in Hayes and Steriade’s chapter (p. 9), means, for example, that voicing is more difficult to realize in [g:] than in [d:]. The constraint ranking in (2) is assumed to be fixed, based on the phonetic difficulty scale in (1). The ranking indicates, for example, that it is more marked for a language to have [g:] than [d:].

Proposing a phonetic scale such as (1) and deriving a fixed hierarchy of phonetically motivated markedness constraints as in (2) are crucial strategies in phonetically based phonology, and they are adopted in several other chapters in PBP, e.g. chapters by Jun, Kirchner, Crosswhite, Zhang, and Flemming. The idea of a fixed ranking of constraints seems to be at odds with the idea that OT constraints are by nature permutable (i.e., their ranking can be rearranged) and that the various permutations of the constraints account for typological variation across languages (a notion referred to as factorial typology). However, it appears that there are justifications for stipulating fixed ranking for certain types of constraints (Prince and Smolensky (1993), McCarthy (2002)).

Cross-language typological variations in voiced obstruent inventories are captured by modifying the relative ranking of the markedness constraints in (2) with the faithfulness constraint IDENT(voice), which states that voicing should not be altered between the input and output. For example, Sudanese Nubian permits the obstruents [b d g b:] but not [d: g:]; this can be captured by ranking IDENT(voice) between *d: and *b: in the hierarchy in (2).

As Hayes and Steriade point out, a voiced obstruent inventory such as [b d g b:] would not be expected if phonologies always opted for systems that are formally simple and symmetric. However, given languages such as Sudanese Nubian that have asymmetric inventories, Hayes and Steriade argue that the mapping from voicing difficulty to

---

¹ This is explained by the fact that voicing can continue for a longer time for an obstruent with a larger oral cavity behind the constriction, e.g. a labial obstruent.
² Phonetic symbols are used here as shorthand for feature combinations.
markedness scales is quite direct, and that the grammar reflects in greater detail the complexity of phonetic difficulty, at the cost of formal simplicity.

The approach taken by Hayes and Steriade is conducive to closing the gap between phonetics and phonology, in the sense that it promotes the idea of paying serious attention to phonetic details when accounting for phonological regularities. A similar stance is shared by Zhang and Gordon (section 5). However, the authors are not explicit about which phonetic details are honored by phonologies and which ones are not, thus falling short of a predictive theory of the phonetic grounding of phonology. For example, degree of difficulty of voicing can be influenced not only by length and place of articulation, as incorporated in (1), but also by factors such as whether the obstruent is at the beginning, middle, or end of the word, whether a nasal consonant precedes the obstruent, and whether another obstruent precedes or follows the obstruent (Hayes (1999)). These factors are known to affect voicing in some languages (Ohala (1983)), but not in others. If the mapping between phonetic difficulty and markedness were direct, then some combination of these factors should determine markedness. However, languages tend to opt for simplicity, showing sensitivity to only a subset of factors that affect difficulty. This suggests that phonetic details can be the basis for markedness, but only some of the details can. In Hayes and Steriade’s account, the choice of the relevant phonetic details is left unexplained.

3. Segmental Processes

A number of chapters in PBP focus on phonological processes involving consonants (Jun and Kirchner) and vowels (Crosswhite and Kaun).

3.1. Consonants

Jun’s chapter (chapter 3, “Place assimilation”) offers a phonetically motivated account of patterns of consonant place assimilation in languages. The chapter first reviews what is known about the perceptibility of consonants, focusing on the perceptibility in VC1C2V sequences of the first consonant (C1) which is typically the target of place assimilation (that is, the place of articulation of C1 (the target) assimilates to that of C2 (the trigger)). Several observations are made; for example, cues for place of articulation of the target (C1) differ in perceptibility
depending on its manner of articulation, such that nasals have weaker place cues than stops, which in turn have weaker place cues than continuants such as glides. These observations are shown to be reflected in typological comparisons of languages. For example, there are languages in which place assimilation is observed for nasals only and not for stops or continuants, and other languages in which place assimilation is observed for nasals and stops, but not for continuants. The principle underlying these patterns is that place assimilation, which is assumed to occur as a way to minimize articulatory effort, is more likely to occur for consonants that are less perceptible.

Jun then proposes a formal account of place assimilation within an OT framework, using two phonetically motivated sets of constraints that refer to speech production and perception, respectively: a markedness constraint (WEAKENING) that penalizes extra articulatory gestures (which in effect promotes assimilation) and a set of faithfulness constraints (PRESERVE) that preserve perceptual cues to features (which disfavors assimilation). By hypothesis, the PRESERVE constraints are assumed to have a fixed universal ranking based on perceptibility, such that a constraint preserving perceptually more salient segments are ranked above those preserving perceptually less salient segments (e.g. PRESERVE(place of continuants) » PRESERVE(place of stops) » PRESERVE(place of nasals)). By ranking WEAKENING in an appropriate position within the set of PRESERVE constraints, Jun’s analysis accounts for attested patterns of assimilation in various languages, and successfully rules out unattested patterns.

Furthermore, Jun attempts to account not only for categorical assimilation, which occurs obligatorily regardless of speech rate and style, but also for gradient assimilation, a rate- and style-sensitive process in which residual gestures corresponding to C1 are observable. This is done by proposing that WEAKENING and PRESERVE take on real-valued coefficients that divide the phonetic scales into small steps along a continuum, and by assuming that the relative ranking of the decomposed WEAKENING and PRESERVE constraints varies depending on rate or style.

Kirchner (chapter 10, “Consonant lenition”) argues that various kinds of consonant lenition processes, such as flapping, degemination, spirantization, debuccalization, and voicing are not unrelated phenomena, but constitute a unified phenomenon driven by an attempt to minimize articulatory effort. This is expressed in an OT framework by positing that all candidates generated by GEN are evaluated with respect to “effort
cost" (an estimate of the biomechanical energy required for articulation of the candidate) and by proposing a constraint family called LAZY which penalizes candidates that exceed a certain amount of effort cost. The ranking of LAZY constraints is also fixed such that more effortful articulations incur more severe violations than less effortful articulations. By modifying the relative ranking of LAZY constraints and faithfulness constraints that preserve perceptual distinctions, the theory is capable of explaining typological generalizations in lenition patterns as well as variation in lenition patterns as a function of speech rate and register. The assessment of "amount of articulatory effort" is based on several assumptions about the types of articulatory movements that require energy; for example, it is assumed that it takes more effort to move an articulator for a greater distance, to move it more rapidly, and to sustain an articulatory posture for a longer time. These factors are said to be reflected in the general tendency across languages for consonants to lenite more often when flanked by low vowels than when flanked by non-low vowels (low vowels necessitate articulators to travel further to achieve a constriction), and the tendency for consonants to lenite in fast/casual speech (fast/casual speech requires more rapid articulator movement).

3.2. Vagueness of Phonetics Scales

Jun’s chapter and Kirchner’s chapter both deal with some kind of reduction of consonantal gestures, and their approaches reveal common problems, so it is instructive to treat them together in this section. The consonant reduction processes discussed by Jun and Kirchner can be seen to be functionally motivated by a tendency to reduce articulatory effort. However, notions such as "articulatory effort" or "ease of articulation," as intuitive as they are, have never been defined in a satisfactory, uncontroversial manner (Ohala (1990)). Thus, the very concept that drives the formulation of the WEAKENING constraint in Jun’s chapter and the ranked LAZY constraints that constitute the core of Kirchner’s proposal rests on unsubstantiated grounds. Kirchner himself admits this: "The analysis must be regarded as tentative, in that it relies upon certain assumptions concerning the relative effort cost of particular articulatory gestures; ultimate confirmation or justification thus awaits a programme of articulatory experimentation and modelling to establish a more objective quantification of effort" (p. 316).

Because phonetic scales of effort are not well-defined, several of the
assumptions seem to be questionable or inconsistent. Jun assumes (pp. 70–72) that a two-gesture sequence such as /tk/ requires greater effort than a single gesture such as /kk/, even though /k/ is lengthened to compensate for the deleted /t/. This is based on a simplified computation of effort, in which all closure gestures, regardless of place, manner, voicing, or length, are associated with a fixed effort, while zero effort is associated with a deleted gesture (p. 70). However, this approach essentially ignores the factors that are mentioned by Kirchner to influence effort; in particular, Jun’s computation of effort is at odds with Kirchner’s assumption that a longer closure requires greater effort than a shorter closure (p. 325). If Kirchner is right, then there is some cost in lengthening the short /k/ to a long /kk/ in the process of place assimilation, even though there is some gain in deleting the /t/. If so, then the assumption that a two-gesture sequence such as /tk/ is always more costly than a one-gesture sequence such as /kk/ may be too simplistic.

In addition, Kirchner assumes that articulatory effort can be equated with biomechanical effort, which is essentially a function of acceleration (rate of change of velocity) and mass. Greater velocity (required by a faster speaking rate or greater displacement in a fixed amount of time) as well as greater mass (required for moving a massier articulator) should both require greater effort. While the former is incorporated into Kirchner’s model, the latter does not appear to be. If we assume that different articulators, such as the tip and body of the tongue, have different masses (Barry (1992)), then a place of articulation that involves an articulator with a greater mass (tongue body) should be associated with greater effort than a place of articulation that involves a smaller-mass articulator (tongue tip). However, effort differences by place are not systematically reflected in Kirchner’s account.

3.3. Vowels

Crosswhite (chapter 7, “Vowel reduction”) surveys a variety of languages that exhibit vowel reduction, a process by which vowels undergo quality changes in unstressed positions. Crosswhite proposes that there are two basic types of vowel reduction, contrast-enhancing reduction and prominence reduction. In contrast-enhancing reduction, certain perceptually challenging vowel qualities are limited to appear only in stressed positions, so that in unstressed positions, vowels change to the corner vowels /i a u/, which are perceptually distinctive. For example, in Belarusian, the vowels /e o/ are lowered to /a/ in unstressed posi-
tions. This is achieved within an OT framework using licensing constraints that allow non-corner vowels only in stressed positions. On the other hand, in prominence reduction, certain high-sonority vowels are not permitted to occur in unstressed positions, and are replaced by lower-sonority vowels. For example, in Bulgarian, the vowels /e o a/ are reduced to /i u ø/, respectively, in unstressed positions. This is formally achieved by using prominence alignment constraints, a constraint family that disfavors high-sonority vowels in unstressed positions but instead favors them in prominent positions. Crosswhite argues that recognizing the existence of two reduction mechanisms is a key to resolving "reduction paradoxes," which are cases where one and the same vowel (e.g. /a/) is both highly marked (i.e. tends to be subject to reduction cross-linguistically, as in Bulgarian) and highly unmarked (i.e. often serves as a reduction vowel, replacing other vowel qualities that are subject to reduction, as in Belarusian).

Consonant lenition (Kirchner's chapter) and vowel reduction (Crosswhite's chapter) are similar in that both involve reduction of segments. The two chapters are also similar in that there are categorical and gradient aspects to both phenomena. Kirchner notes that there are "stable" lenition processes that obligatorily apply regardless of speech rate or style, and "variable" lenition whose application varies depending on rate and style. Crosswhite also notes that there are "phonemic" vowel reduction processes which apply in all contexts, and "phonetic" vowel reduction (also known as "vowel undershoot") that applies to varying degrees depending on conditions such as speech rate. No one would probably disagree that the "stable" and "phonemic" processes are domains that phonological theory should account for, but opinions vary concerning whether the "variable" and "phonetic" processes should also be treated within phonology. In the spirit of bridging the gap between phonetics and phonology, these processes should in fact be accounted for in phonological theory, e.g. by proposing mechanisms that handle gradient phenomena (cf. Boersma (1998), Flemming (2001)).

Kaun (chapter 4, "The typology of rounding harmony") surveys various languages exhibiting lip rounding harmony of vowels, and arrives at a handful of conditions under which harmony tends to occur. They are: (1) the rounding harmony trigger (the segment that causes rounding harmony) is non-high, (2) the trigger is front, (3) the rounding harmony target (the segment to which rounding spreads) is high, (4) the target is back, and (5) the trigger and target agree in height (see (2), p. 92).
These tendencies are then given phonetic grounding by pointing out evidence that the acoustic and perceptual consequence of adding lip rounding to back and high vowels is greater than that achieved by adding lip rounding to front and non-high vowels. Rounding harmony is proposed to be a perceptually driven process, whereby a vowel whose rounding is hard to identify (front vowels and non-high vowels) extends its rounding feature to a neighboring vowel which, when rounded, is easily identified as a rounded vowel (back vowels and high vowels). Furthermore, harmony between vowels of different heights is proposed to be avoided because the degree of lip rounding differs between high and non-high vowels, and there is a preference for an articulatory gesture to be uniformly executed during its span of association. Finally, Kaun proposes a set of OT constraints along with different constraint rankings for different types of rounding harmony systems across various languages.

Kaun’s perceptually motivated account of rounding harmony elegantly explains the asymmetric tendency for non-high and front vowels to become triggers and high and back vowels to become targets. It cites data from articulatory studies (Boyce (1990)) showing that the spreading of the [round] feature to multiple vowels is phonetically realized as a single lip-rounding gesture across the VCV sequence rather than being realized as two separate gestures. The single-gesture idea is critical in explaining the preference for same-height harmony over cross-height harmony (vowels of different heights are associated with different degrees of lip rounding gestures). If such gestural uniformity is crucial in determining likelihood of rounding harmony, however, then one might naïvely wonder if rounding might be blocked or dispreferred when the intervening consonant involves a lip-closing gesture, e.g. labial consonants. This question is not directly addressed by Kaun, but the sample words given in the chapter appear to indicate that labials do not block rounding harmony. It is therefore unclear how a one-gesture account of rounding harmony explains the resistance of harmony in different-height vowels and at the same time captures the apparent non-blocking behavior of intervening lip-closing gestures.

4. Segment Sequences

Two chapters in PBP deal with phonological phenomena involving sequences of segments. In Wright’s chapter (chapter 2, “A review of perceptual cues and cue robustness”), it is proposed that the Sonority
Sequencing Constraint, whose traditional formulations are known to be plagued with exceptions, can be reformulated as perceptually motivated, such that "an optimal ordering of segments is one that maximizes robustness of encoding of perceptual cues to the segmental makeup of the utterance" (p. 35). Wright gives a comprehensive survey of perceptual cues for place, manner, and voicing of consonants, and vowel quality, and describes factors that influence the robustness of encoding of phonological contrasts. For example, the auditory system is more sensitive to the onset of acoustic events than their offset, so that a boost in neural activity is associated with prevocalic consonants in CV sequences while no such boost is associated with postvocalic consonants in VC sequences. Other factors that influence cue robustness are: resistance of certain speech sounds to environmental masking, cue redundancy achieved through gestural overlap in speech production, and the amount of modulation (e.g. change in frequency or amplitude) in the speech stream (the greater the modulation is, the more salient the cue is). Wright then relates these general principles to segmental organization in languages, and argues that segmental patterns that are well-attested in languages are those in which the cues are least vulnerable to degradation and loss. For example, many past formulations of the Sonority Sequencing Constraint have treated initial /s/+C clusters as an exception to the general rule that segments before the vowel nucleus rise in sonority, but according to a perceptually motivated account, sibilant fricatives such as /s/ can occur in such a position because it has sufficiently robust internal cues (noise frequency) so that it need not be adjacent to vocalic segments in order to be identifiable to the listeners, unlike stops which are optimally signaled next to vocalic segments.

Blevins and Garrett (chapter 5, "The evolution of metathesis") deal with metathesis, the inversion of segments within words, and argue for a non-functionalist account of phonology that diverges from positions held in many other chapters. The authors identify four types of metathesis sound changes and provide phonetic explanations for each type. They are: (1) perceptual metathesis, which involves "migration" of a segment that has elongated phonetic cues spanning several segments, (2) compensatory metathesis, which involves a stress-induced inversion of a consonant and a following unstressed vowel, (3) coarticulatory metathesis, in which intervocalic CC stop clusters undergo metathesis due to extreme gestural overlap, and (4) auditory metathesis, in which stop+sibilant and sibilant+stop clusters undergo metathesis due
to auditory stream decoupling of the sibilant (which contains high-energy frication noise) and the stop. In coarticulatory metathesis, Blevins and Garrett report asymmetries in metathesis patterns in CC clusters depending on consonant place, such that labial+velar clusters are often reanalyzed as velar+labial clusters but not the other way around, and coronal+non-coronal clusters are reanalyzed as non-coronal+coronal clusters, respectively, but not the other way around. Blevins and Garrett describe two competing theories concerning phonetic influences in phonology, a “phonetic optimization approach” that maintains that phonology is teleological in that it is functionally driven by constraints that try to optimize phonetic attributes such as articulatory ease or perceptual salience, and an “evolutionary phonology approach” that claims that sound patterns are not necessarily phonetically optimizing, and are instead the result of historical sound changes that stem from intrinsic properties of speech production and perception, such as listeners’ reinterpretation of sound patterns. The authors maintain that the latter approach offers a better explanation of existing metathesis patterns.

In both Wright’s account of segmental organization and Blevins and Garrett’s account of metathesis, perceptual factors play a key role, albeit in quite different ways. It is rather surprising that neither of the chapters considers in any detail speech production constraints that may also influence segmental organization. For example, articulatory factors have been proposed to influence metathesis in some languages (e.g. Hume (2001)) and in child language acquisition (e.g. Kirk and Demuth (2005)). Also, in the organization of segments into syllables, the organization of articulatory gestures seems to show a preference for CV syllables over VC syllables, especially under demanding speaking conditions, e.g. when speaking at very fast speech rates (e.g. Tuller and Kelso (1991), de Jong (2003)). Thus, perceptual considerations may not be the only factors that are responsible for the preference of onsets over codas, and the tendency for some segment sequences to undergo metathesis.

On the issue of how speech perception influences phonology, Blevins and Garrett take a diachronic, non-functionalist view that is in contrast to the synchronic, functionalist view adopted by Wright and other contributors to PBP. Blevins and Garrett maintain that “genuine” metathesis sound changes occur as historical sound changes, and metatheses that appear to be synchronic (e.g. Hume (2001)) are not genuine, reflecting other processes such as analogical changes. They maintain
that sound changes are mainly caused by listener-based reinterpretation of the speaker’s production, an idea that Ohala (1990) refers to as “innocent misapprehension.” While such an account might be reasonable for the case of metathesis, which is often regarded as sporadic and irregular, a strictly non-teleological view of phonology may not be able to account for certain processes that appear to be “phonetically optimizing.” For example, Flemming (2005) mentions the example of languages in which voiced stops have changed into prenasalized voiced stops. Even though prenasalized voiced stops are by themselves rather uncommon typologically, it is claimed that prenasalization is a strategy that speakers have come across as a way to make the voicing contrast, i.e. the distinction between voiced and voiceless stops, more distinct. This type of contrast-enhancing sound changes would not be predicted to occur if sound changes happened only based on listeners’ misperceptions.

5. Prosody

Prosodic phenomena pertaining to the distribution of contour tones and to syllable weight are discussed in Zhang’s and Gordon’s chapters, respectively. Zhang (chapter 6, “The role of contrast-specific and language-specific phonetics in contour tone distribution”) argues that language-specific phonetic details are crucial for understanding the distribution of contour tones. Zhang hypothesizes that the ability of a syllable to bear a contour tone, referred to as \( C_{\text{CONTOUR}} \), is directly related to the duration of the syllable’s sonorous rhyme, and that languages differ in the extent to which various phonological factors affect sonorous rhyme duration (hence the contour-tone-bearing ability of the syllable). A survey of 187 languages indeed indicated that contour tones are more likely to occur in syllables with longer sonorous rhyme duration, i.e. in CVV syllables, stressed syllables, final syllables of words or utterances (which are longer due to final lengthening), and syllables in shorter words (which tend to be longer than syllables in longer words). Acoustic duration analysis of a subset of languages further showed that phonological factors that determine contour tone licensing in a language are also factors that strongly affect sonorous rhyme duration. For example, in Thai, contour tone licensing is determined by the sonorancy of the coda and not by vowel length; that is, CV, CVR, and CVVR syllables (R=sonorant) support contour tones, while CVO and CVVO sylla-
bles (O=obstruent) do not. Acoustic measurements showed that syllables not ending in an obstruent had significantly longer rhyme duration than syllables that do. In contrast, in Navajo, contour tone licensing is determined by vowel length and not by coda sonorancy; that is, CVV, CVVR, and CVVO syllables support contour tones, while CV, CVR, and CVO syllables do not. Acoustic measurements suggested that syllables with long vowels had greater overall sonorous rhyme energy than syllables with short vowels. These results support the view that positional licensing is contrast-specific; that is, since different contrasts rely on different phonetic properties, they are preferentially licensed in different positions rather than being licensed in a restricted set of specified positions. The results also support the view that the licensing patterns are tuned to language-specific phonetic details.

Gordon (chapter 9, “Syllable weight”) examines the role of phonetics in the phonology of syllable weight. Syllable weight, or the distinction between heavy and light syllables, has traditionally been thought to vary as a function of language, but Gordon points out that weight criteria depend more on the phonological process involved. For example, a survey of numerous languages indicated a difference in weight criteria between tone assignment and stress assignment. That is, tone assignment processes often treated CVV and CV[+son] syllables as heavy and CV[−son] and CV syllables as light, whereas stress assignment processes typically treated CVV and CVC syllables as heavy and CV syllables as light. This difference is attributed to phonetic factors: the realization of tone hinges on the presence of sonorous energy (which is salient in vowels and sonorants), but stress can be realized on the energy profile of the entire rhyme, not just the sonorous portion. Gordon then describes a detailed phonetic study of weight-sensitive stress, and tests the hypothesis that languages make weight distinctions according to phonetic effectiveness, i.e. in such a way that the difference between heavy and light syllables along some phonetic dimension (e.g. total energy in the rhyme) is maximized. Results indicated that languages indeed employed weight distinctions that are most phonetically effective, but with the added qualification that the distinction employed had to be phonologically simple, i.e. expressible without disjoint statements. For example, in Khalkha, which treats CVV syllables as heavy, the dichotomy between CVV syllables and non-CVV syllables led to the greatest separation in total energy among the phonologically simple criteria. In comparison, in Finnish, which treats CVV and CVC syllables as heavy,
the dichotomy between CVV and CVC syllables on the one hand and other syllables on the other hand was the most phonetically effective. These results are taken to support the notion that both phonetic effectiveness and phonological simplicity contribute to the shaping of sound systems.

Zhang’s analysis of contour tone distribution and Gordon’s treatment of syllable weight are prime examples of how serious attention to phonetic details can illuminate our understanding of phonological patterning, to such an extent that these proposals require revisions to previous theories. For example, Zhang’s analysis suggests that contour tones can be licensed by factors such as phrase-finality and word length, which are not among factors that have been previously considered relevant to licensing (Beckman (1998)). Similarly, Gordon demonstrated that criteria for syllable weight are not fixed within a language, as has been previously assumed (e.g. Hayes (1995)), but rather vary within a language depending on the phonological process, e.g. tone, stress, metrics, and minimal word requirement. Both chapters clearly demonstrate that cross-linguistic differences in phonological patterns (e.g. which syllables are good tone bearers or which syllables are heavy) are often not just arbitrary choices made by languages, but can find their origins in the fine-grained details of how phonological entities are phonetically expressed.

Gordon’s chapter is reasonably explicit about the definition of the phonetic scale used for explaining weight-sensitive stress (integration of perceived energy over the rhyme portion), and about how the indirect mapping from phonetic scales to phonological patterns is achieved by referring to phonological simplicity. One wonders, however, whether a phonetic scale such as total perceptual energy is influenced not just by syllable structure but also by other factors such as vowel identity, speaking rate, and style, and if so, how particularities of the speech corpora collected (number of speakers, words, and repetitions used, type of words used, etc.) might affect the outcome of the analyses. Moreover, Gordon’s chapter does not provide any explanation, phonetically motivated or otherwise, on why languages vary in weight criteria for tone, although it provides an extensive analysis of weight-sensitive stress.

Zhang adopts a more abstract approach than Gordon to compute the phonetic scale used to capture the contour-tone-bearing ability of syllables (C_{CONTOUR}). Instead of integrating energy over the rhyme, Zhang simply measures the duration of the vowel and sonorant in the rhyme,
multiplies the vowel portion by a factor (called \( a \)), and takes the sum of the weighted duration of the vowel and the sonorant. Zhang’s chapter gives no description of how to compute the coefficient \( a \), so the phonetic scale, \( C_{\text{CONTOUR}} \), cannot be calculated precisely. This means that comparisons across different syllable types shown in Figures 6.2–6.4 in the book are only approximations to the actual values of \( C_{\text{CONTOUR}} \). Furthermore, since \( C_{\text{CONTOUR}} \) is a durational measure, it may face a problem similar to Gordon’s total perceptual energy, in that other factors that affect syllable duration may also affect \( C_{\text{CONTOUR}} \), such as speaking rate and style. Zhang attempts to circumvent this problem by assuming that phonetic parameters such as \( C_{\text{CONTOUR}} \) are calculated based on an idealized, “canonical” token produced in a “standard” mode of speech. This reflects an attempt to eliminate variation in factors such as speaking rate and pitch range. However, in real-life communicative contexts, even a “standard” mode of speech potentially contains considerable phonetic variation; according to Lindblom (1990), speakers can dynamically tune their production to meet short-term demands along a continuum of output- or listener-oriented hyperspeech and system- or speaker-oriented hypospeech. As such, the notion of a “canonical” token seems rather difficult to support.

In addition, Zhang assumes a unidirectional causal relationship between sonorous rhyme duration and contour tones, such that a longer duration causes contour tones to become more likely to occur. However, an influence in the opposite direction is theoretically possible, such that the presence of a contour tone causes a syllable to become longer. For example, the need to realize a contour tone on a particular syllable may to some extent cause the syllable to become phonetically longer, or even make the syllable a stressed syllable. Such possibilities are not properly addressed in Zhang’s chapter. The same point could be raised about Gordon’s chapter as well. That is, Gordon does not properly rule out the possibility that the distinction between heavy and light syllables in a language may cause certain phonetic dimensions of syllables, e.g. total rhyme energy, to change.

6. Segment Inventories and the Lexicon

The two remaining chapters in PBP deal with phenomena that are not considered typically in conventional phonological analyses. Flemming (chapter 8, “Contrast and perceptual distinctiveness”) deals with segment
inventories rather than particular phonological processes, and proposes that the markedness of speech sounds is not a property of the individual sounds but rather depends on the sounds that they contrast with. Flemming puts forth a theory of phonological contrasts called Dispersion Theory, which elaborates on the idea that selection of contrasts in a language is subject to three functional goals, which can be formalized as OT constraints: (1) maximize the distinctiveness of contrasts, (2) maximize the number of contrasts, and (3) minimize articulatory effort. Numerous examples are given that show evidence for such constraints on contrasts. For example, many languages contrast vowels in the front-back dimension, realized acoustically as a contrast in second formant frequency (F2). Typologically, fully front and back vowels such as /i/ and /u/ are highly common, but central vowels such as /i/ are not as common. Also, front vowels are typically unrounded, while back vowels are often rounded. In conventional accounts, these patterns have been explained by positing that certain vowels and feature combinations, such as front rounded /y/, back unrounded /u/, and the central vowel /i/, are inherently marked. On the contrary, Flemming argues that these vowels are dispreferred not because of inherent properties, but because they have intermediate F2 values, which make contrasts involving these intermediate vowels, such as /y/-/u/, /i/-/u/, /i/-/i/, and /i/-/u/, to be less distinctive than a contrast involving vowels with extreme F2 values, i.e. /i/-/u/, which is in fact preferred by languages. The relevance of contrasts in the markedness of sounds is demonstrated by the fact that in languages that do not have front-back vowel contrasts, intermediate vowels such as /i/ are quite common. For example, Flemming cites examples of Kabardian and Shapsug, whose short vowels have a “vertical” system of two vowel /i ə/. The common occurrence of such intermediate vowels would not be predicted from conventional treatments of segmental markedness.

In the last chapter of the book (chapter 11, “Language processing and segmental OCP effects”) Frisch deals with long-distance phonotactic constraints that disfavor homorganic consonants to appear multiple times within a word. This has traditionally been proposed within autosegmental phonology to be an instantiation of the Obligatory Contour Principle (OCP), which bans consecutive features to be identical (McCarthy (1986)). However, Frisch shows that the patterns of similarity avoidance in languages such as Arabic vary not in a categorical fashion but in a gradient fashion depending on such factors as segment
similarity, distance between segments, and segment frequency. Also, segmental OCP effects are not considered to be unique to Arabic, but occur in varying degrees in many languages including English. Frisch argues that the richness of these similarity avoidance patterns cannot be captured by the conventional all-or-nothing OCP approach, and proposes a functional explanation, based on evidence from online language processing that indicates that people have difficulty correctly determining the serial order of similar segments. That is, sequences with repeated items are more difficult to process than those that do not contain repetitions, presumably because repetition introduces a potential confusion in linear ordering. The functional account posits that OCP effects exist in languages as a way to avoid sound patterns that are difficult to process due to repetition and similarity.

Flemming’s and Frisch’s chapters demonstrate that fine-grained phonetic details have an impact not just on segmental, phonotactic, and prosodic phenomena, but also on other aspects of linguistic organization such as segment inventories as well as the structure of words in the lexicon. They also suggest that the notion of constraint interaction, whether implemented as OT analyses or connectionist models, is also pervasive, having an effect on inventories of sounds as well as the mental lexicon.

7. Syllable Structure

Phonological patterns pertaining to syllable structure are not discussed in detail in PBP, but syllable structure has been an established topic in phonology including OT (e.g. Prince and Smolensky (1993), Blevins (1995)). In OT terms, typological variations in permissible syllable structures would be captured by the relative ranking of faithfulness constraints with markedness constraints such as ONSET (syllables must have onsets), NO-CODA (syllables must not have codas), *COMPLEX-ONSET (no more than one C may be associated with the onset position), and *COMPLEX-CODA (no more than one C may be associated with the coda position), or variants of such constraints. The tendency to favor onsets and to avoid codas is captured by ONSET and NO-CODA, but whether a similar tendency holds for complex onsets and complex codas has been an open question. According to Blevins (1995), some languages allow complex onsets but not complex codas, while other languages go the other way around, thus precluding an implicational relationship.
A number of studies in second-language (L2) speech learning have investigated native Japanese learners’ acquisition of English syllable structure. Since Japanese does not permit tautosyllabic consonant clusters and permits only a few types of syllable-final consonants, Japanese learners of English often avoid such sound sequences by inserting epenthetic vowels between consonants or following word-final consonants, or by deleting consonants, e.g. “blog” \(\rightarrow /\text{burogu}/\) (epenthesized vowels are underlined). Studies that have investigated Japanese learners’ production and perception of English syllable structure have reported asymmetries in difficulty between onsets and codas. When native Japanese listeners are asked to count syllables in spoken English words and English-like nonwords, they are less able to correctly count the number of syllables in items that contain complex onsets, e.g. splep, than in items that contain complex codas, e.g. pemps (Tajima et al. (2002), Tajima (forthcoming)). This suggests that complex onsets are more difficult for Japanese listeners than complex codas when the task is to count syllables. In contrast, data from an experiment in L2 speech production (Hancin-Bhatt and Bhatt (1997)) reveal that native Japanese speakers make fewer production errors in complex onsets than in complex codas, suggesting that complex onsets are easier than complex codas in speech production, an opposite pattern from that observed in speech perception. In addition, the same production data show that errors in complex onsets were predominantly vowel epenthesis errors, while errors in complex codas were predominantly consonant deletion errors.

If we adopt a simplistic view that phonetic difficulty directly maps onto markedness, as was done in (1) and (2) above in Hayes and Steriade’s chapter, then the perception difficulty observed in Tajima et al.’s (2002) study suggests the scale of phonetic difficulty COMPLEX-ONSET \(<\) COMPLEX-CODA and the constraint ranking \(^*\)COMPLEX-ONSET \(>\) \(^*\)COMPLEX-CODA. However, the production difficulty observed in Hancin-Bhatt and Bhatt’s (1997) study suggests exactly the opposite scale and the opposite constraint ranking. The two rankings are obviously mutually incompatible.

These apparently conflicting results can be reconciled if we propose that the onset-coda asymmetry is related to acoustic and perceptual factors that render onsets as more perceptually salient than codas. There is evidence that onsets (or cues in CV transitions) are more acoustically distinct and perceptually salient than codas (or cues in VC transitions)
If onsets are perceptually salient, then learners might be likely to maintain the onsets in their productions, as in Hancin-Bhatt and Bhatt’s (1997) study. The greater perceptual salience of onsets may also explain why learners tend to make vowel epenthesis errors in onsets (which retain the onset consonants at the cost of producing a superfluous vowel), and conversely, the relative non-salience of codas may explain why learners tend to make consonant deletion errors in codas. Also, in Tajima et al.’s (2002) perception study, greater perceptual salience of onsets may have led listeners to perceive some of the onset consonants as “syllabic,” leading them to incorrectly count more syllables than there actually are. In fact, there are reports suggesting that native Japanese listeners hear “illusory” epenthetic vowels within consonant clusters even if a vowel was not produced between the consonants (Dupoux et al. (1999)). If so, then onset clusters, by virtue of being perceptually salient, may be more susceptible to such “perceptual vowel epenthesis” than coda clusters, consequently leading to more syllable-counting errors for items with complex onsets.

Under this view, the apparently conflicting results in production and perception of onset and coda clusters can be resolved by referring to a phonetic scale of relative perceptual salience of onsets and codas. That is, complex onsets are more successfully produced but less perceptually integrated into single syllables because onsets are more perceptually salient than codas. This could be captured within an OT framework by, for example, positing a phonetic scale of perceptual salience that renders onset consonants as more perceptually salient than coda consonants. Based on this phonetic scale, a set of constraints could be proposed that favors retention of consonants in the onset position but deletion of consonants in the coda position. Such constraints could then account for both the more successful production of onset clusters than coda clusters in Hancin-Bhatt and Bhatt’s (1997) study, and the less successful perception of onset clusters than coda clusters in Tajima et al.’s (2002) study. Furthermore, this type of analysis provides a functional explanation for why there are asymmetries between onset and coda positions; for example, onsets are favored and codas are disfavored because onsets are more perceptually salient than codas (cf. Steriade (2001), Wright (2001)). This view is therefore a viable alternative to conventionally proposed analyses based on constraints such as *COMPLEX-ONSET and *COMPLEX-CODA, which do not provide a princi-
pled explanation for onset-coda asymmetries. This view is also in line with PBP’s central claim that many phonological patterns can be motivated on phonetic grounds, and suggests that reference to phonetic scales can be useful not just for synchronic phonological analyses, such as those presented in the majority of the chapters in PBP, but also in other domains such as L2 speech production and perception.

8. General Discussion

8.1. Intended Audience and Book Title

While PBP attempts to address the relationship between phonetics and phonology, it is ultimately a book on phonology, and its intended audience is phonologists, rather than phoneticians. Bluntly put, PBP views the phonetics-phonology interface as one in which phonetics is used as the means to achieve the goals of phonology. In fact, one of the goals of this book, as stated in chapter 1, is to show “how the OT search for the right constraint set can be speeded up on the view that markedness is phonetically based” (p. 2). Furthermore, the book assumes that the reader is already familiar with aspects of OT. As such, for someone relatively new to OT, there is insufficient discussion of and justification for why the phonetically motivated OT analyses proposed in various chapters are significant, and how they are superior to other possible accounts. On the other hand, data from speech production, acoustics, and speech perception that are necessary to understand the phonetic bases of phonological phenomena are discussed extensively in the volume. Thus, the book may serve as a good source to look for phonetic studies that have crucial implications for phonology.

The title “Phonetically Based Phonology” is succinct, but is ambiguous and therefore could be potentially misleading. The title is appropriate if “phonetics” and “phonology” are used in the restrictive sense of “principles governing speech production and perception” and “OT-like analysis of phonological phenomena,” respectively. But if “phonetics” is interpreted to mean “methods of experimental phonetics,” then such a reader would be unsatisfied with the paucity of chapters that actually conducted experiments. If “phonetics” and “phonology” are interpreted to mean “gradient” and “categorical” processes, as is sometimes done (Beckman and Kingston (1990)), then such a reader might also be disappointed with the lack of discussion of the interface between continuous and discrete aspects of sound systems.
8.2. Phonetics-phonology Interface

As mentioned above, the stance that PBP takes on the relationship between phonetics and phonology is that markedness constraints that operate on phonological systems have their origins in phonetic principles. It is important to point out that this is only a limited aspect of how phonetics and phonology relate to each other. Many issues surrounding the phonetics-phonology interface are left out in PBP, as the editors themselves admit, such as the nature of phonological representations and the extent to which they are based on phonetic principles. For example, there have been claims that the basic building blocks of phonological representations are not features but dynamically defined articulatory gestures (e.g. Browman and Goldstein (1990)). There is also increasing evidence, put forth by so-called exemplar-based theories, suggesting that phonological representations contain many more phonetic details than have been previously assumed, and are rich with information pertaining to particular words, utterances, and speakers (Bybee (2001), Pierrehumbert (2002)).

The nature of phonological representations obviously has significant implications for phonological theories including OT, so any theory that simply assumes the presence of features as a priori entities should be approached with a grain of salt. In addition, a related fundamental problem surrounding the phonetics-phonology interface concerns how the discrete categorical entities of phonology relate to the continuous scalar parameters of phonetics. This is a relatively old but persistent problem that continues to be addressed today (e.g. the Laboratory Phonology series starting with Kingston and Beckman (1990), Flemming (2001)), but PBP has nothing substantial to say about this problem.

8.3. Phonetics and OT

Several problems arise in the course of attempting to incorporate phonetics into OT phonology. First, some of the chapters in PBP propose constraints with real-valued coefficients that divide a particular phonetic scale into arbitrary small steps along a continuum, e.g. the WEAKENING and PRESERVE constraints by Jun and the LAZY constraints by Kirchner. Nothing in OT prevents constraints from being associated with real-valued coefficients, but there is something fundamentally unappealing and inelegant about trying to capture continuous, potentially infinite variation using a finite set of discrete constraints. Also, the number of constraints proposed to divide the continuum into discrete steps is often
chosen on an arbitrary basis. A mechanism that incorporates into OT the gradient nature of phonetically motivated constraints might need to be devised (e.g. see Boersma (1998), Flemming (2001)).

Second, phonetic details are important for many phonological processes (e.g. as shown by Zhang and Gordon), but apparently not all phonetic details are. Factors such as speech rate are often irrelevant, so ways to filter out these factors have been proposed, e.g. by evaluating candidates according to a “canonical” mode of speech (Zhang). However, as pointed out by McCarthy (2002: 225), moving away from phonetic details has several problems. First, even though phonetically motivated constraints evaluate properties of physical events, the candidates themselves abstract away from these events if canonical forms are adopted. Also, the choice of which phonetic details are included in the canonical forms and which are excluded seems to be arbitrary. Moreover, using canonical forms as candidates re-introduces the distinction between phonology, which deals with idealized forms, and phonetics, which deals with details included in actual speech events, and runs counter to a functionalist approach to phonology.

Third, many of the OT analyses presented in PBP are based on provisionally proposed constraints. One cannot help but get the impression that some of the constraints proposed are ad hoc patches tailor-made to explain specific phenomena. These include: Kaun’s *RoLo (non-high rounded vowels are avoided), *RoFro (front rounded vowels are avoided), and GestUni (gestural uniformity); Zhang’s *Dur(τ) (rhyme duration cannot be τ or more); Flemming’s Unstressed Vowels Are Short, *Short Low Vowel, *Normal Duration, *Extra-Short, *HighEffort, *Implosive and other constraints banning particular segments; Gordon’s Stress[XX] and One Stress. Unless these constraints are called for in other analyses, or are shown to be special cases of more general constraints, they are unprincipled and poorly motivated.

9. Conclusion

In this review, I have attempted to present a balanced introduction to the contents of the eleven chapters in PBP, to give the reader a taste of the diversity of phenomena addressed and the research interest that this area has received recently. I have also discussed some of the issues raised in the book, particularly concerning the question of how phonetic scales, rich with information but highly variable, are incorporated into
relatively coarse-grained phonological patterns. I have also presented some data bearing on how reference to phonetic scales such as perceptibility helps us understand not just synchronic phonological patterning, but also patterns of L2 speech production and perception. For someone with a stronger background in phonetics than phonology like myself, it was a bit discouraging that the book took for granted substantial prior knowledge of phonology (particularly OT). A better balance in coverage between phonetics and phonology might have been ideal, especially because the book should be read by phoneticians and phonologists alike. It is hoped that the book will stimulate readers to further fill in the gap between phonetics and phonology.

REFERENCES
Browman, Catherine P. and Louis Goldstein (1990) “Tiers in Articulatory Phonology, with Some Implications for Casual Speech,” Papers in Labo-


Stampe, David (1973) *A Dissertation on Natural Phonology*, Doctoral dissertation, University of Chicago. [Distributed by Indiana University Linguistics Club, Bloomington, IN, 1979]


Society of Japan, 103–108.


Department of Psychology
Hosei University
2-17-1 Fujimi, Chiyoda-ku
Tokyo 102–8160
e-mail: tajima@hosei.ac.jp