PROSODIC ORGANIZATION OF SPEECH BASED ON SYLLABLES: THE C/D MODEL

Osamu Fujimura
Department of Speech & Hearing Science, The Ohio State University
Columbus, OH 43210-1002, USA

ABSTRACT

A syllable-based theory of phonetic implementation called the C/D model is reviewed, with remarks on its phonetic implications regarding prosodic control. The phonological feature specification assumed at the input is discussed, in connection with the underspecification scheme. A recent revision of the timing computation scheme accounts for some prosodic effects on temporal behavior of articulatory gestures for English /l/.

C/D MODEL

This paper discusses a new view of speech organization: the Converter-Distributor (C/D) model of phonetic implementation [Fujimura *et al.* 1991; Fujimura 1992, 1994a,b, in press]. It considers the prosodic organization of an utterance as the basic framework for describing the speech production process. A prosodic structure, represented by a metrical tree (see Liberman & Prince [1977]), is assumed with a phonetic augmentation for specifying utterance conditions. The prosodic structure is interpreted as a linear string of syllables and boundaries with varied magnitude values.

The flow of vocalic gestures characterizing the sequence of syllable nuclei forms the *base function* of the articulatory events that fit in the prosodic structure of the utterance. On this base function, consonantal gestures are superimposed, basically in the way Öhman [1966] depicted in his consonantal perturbation model. The base function is inherently multi-dimensional in the sense that different articulatory variables such as jaw opening, tongue body advancing or retraction, lip rounding and protrusion, and pulmonary and laryngeal conditions, behave more or less independently from each other. Prosodic effects are implemented mainly by mandibular, laryngeal, and pulmonary variables. Vocalic gestures are implemented in tongue body position and lip rounding dimensions, which physically interact with mandibular position, and represent a continuous flow of inherent articulatory gestures for unreduced syllable nuclei, constituting one aspect of the base function. The implementation process of vocalic and intonational aspects of the base function may be somewhat similar to existing acoustic models of F0 contours such as Pierrehumbert's [1980] or Fujisaki's [1988]).

To the extent that speech organization is described in terms of articulatory gestures, the C/D model is similar to the articulatory phonology proposed by Browman and Goldstein [1992]. There are many phonetic observations, particularly allophonic variations of phonemes in the traditional segmental description, usually expressed as context-sensitive rewrite rules in generative phonology, that are naturally explained by either theory as the consequence of using an assembly of autosegmental articulatory gestures in variable timing relations. Such variation is typically sensitive to the style of utterance, among other factors.

These two theories, however, basically differ from each other. While articulatory phonology assumes gestures to be the basic units in the lexical phonological representations, integrating everything together from lexical phonology to phonetic signal generation, the C/D model strictly respects the traditional distinction between phonology and phonetics. The phonetics, however, is strongly sensitive to the particular language or dialect, and it also handles abstract features until gestures are concretized at the output of the actuators. What was called the base of articulation in the traditional British literature, for example, is incorporated into the system parameters that prescribe the signal generator design. The phonological structure as the input to the model reflects the lexical specifications and the syntagmatic organization of phonological phrases. At the same time, the numerical specifications attached to any node of the metrical tree produces prominence of the pertinent part of the tree structure, and additional numerical specifications of utterance characteristics including the speaker's habit, determine system parameters for the entire utterance, according to the situation of speaking.

The C/D model describes the phonetic implementation process, apart from the signal generator, in three sequentially ordered system components: converter, distributor, and a parallel set of actuators. The process is inherently multidimensional and superpositionally linear until the set of control time functions are derived. The signal generator, which takes these control functions as its input, is a complex, highly nonlinear and inherently three-dimensional dynamic system [Wilhelms-Tricarico, in press].

The C/D model uses syllables as the basic units of segmental materials that are concatenated into a temporal linear string, intervened by phonetic phrase boundaries. The latter can be empirically observed in articulatory movement patterns, as discussed in Fujimura [1990]. The prosodic structure of an utterance is represented completely, at one level of the phonetic representation, by a series of magnitude-specified pulses. The timing pattern of the series of abstract events for syllables is then derived from the magnitudes (abstract phonetic strengths) of syllables and boundaries.

It is emphasized that the signal generator component, as the last and physical stage of the model, determines critical characteristics of directly observable physical phenomena such as articulatory movement patterns, and based thereon, acoustic or spectrographic patterns, including durations of acoustically defined speech segments. The input to the signal generator may be interpreted to represent basically motor control time functions given to the physiological apparatus for speech production.

The prescription of articulatory gestures in the form of control functions is generated by the set of actuators, the third component of the model. These control variables as time functions can be significantly different from directly observable physical signals, whether articulatory or acoustic. Nevertheless, we claim that the model's general validity can be tested and its parameter values can be inferred, by evaluating physical signal characteristics, if powerful computational techniques are used to handle a large mass of data for inference of the underlying variables.

The basic assumption is that, however complex (with feedback loops, *etc.*) the signal generating system may be, it has a fixed physical design, containing only parameters that are sensitive to the speaker's conditions. In contrast, the process up to the output from the actuators including the table of impulse response functions for consonantal gestures, are parametrically sensitive to the language or dialect spoken.

The classical theory of generative phonology (see Chomsky & Halle [1968]), assuming a level of systematic phonetic representation, ascribes the switching from discrete specifications in phonology to continuous and numerical variable specifications in phonetics to additional suprasegmental variables like segmental duration and tonal inflection, while assuming a large number of phonetic segments (allophones) resulting from detailed but discrete alterations of articulatory states. This can not account for the intricate interaction between articulatory or acoustic gestures of individual consonants or vowels and prosodically conditioned suprasegmental parameters, including variable strengths of phonetic boundaries (see Fujimura [1970]). The continuous nature of phonetic phenomena stems not from the superimposed properties of individual segments, but from the inherently multidimensional nature of the articulatory organization interacting with prosodic conditions. Therefore, the "segmental" characteristics themselves continuously vary.

The C/D model seems to have the potential to account for much of the observed allophonic variation, whether coarticulatory or not, within the phonetic implementation process, according to a general phonetic principle combined with language-specific system parameters. The feature specifications are passed by the converter to the distributor for phonetic gesture specifications. Many apparently supplemental specifications of redundant information are automatically provided by the speech production process itself. For example, unspecified vocalic gestures for reduced syllables in English, can be left unspecified throughout the phonetic process, and computed by the signal generator as continuous time functions, according to the base function control.

Likewise, the place specification for the nasal segment in English coda when combined with a tense obstruent (*e.g.* in 'tent', 'tense', 'camp', 'honk') is not phonologically copied from the stop segment specification, but is implemented as a single articulatory oral closure gesture spanning over the nasal (lowered velum) and oral (raised velum) portions of the coda. In contrast, when an obstruent is voiced and follows a nasal consonant, as in 'lens', 'tend', 'sums', 'songs', *etc.*, the syllable-final voiced obstruent is always apical (alveolar or dental), and the place is specified for the nasal consonant. The final obstruent in such a situation (along with the final voiceless apical obstruent in an obstruent sequence such as 'act' and 'opt') are separated out from the syllable core as a syllable suffix (s-fix), based on the general rule of English syllables that a syllable-final apical obstruent that agree in voicing with the tautosyllabic obstruent in the coda is separated as a s-fix (Fujimura [1979]).

As the first component of the model, the converter's role is to evaluate the prosodic pattern as specified in the augmented metrical tree, to compute the phonetic strength of each syllable, and accordingly, to assign a magnitude value to each impulse that represents the syllable. The converter also creates a boundary pulse by evaluating the tree configuration, and assigns the magnitude value to each boundary pulse. Based on the series of magnitude-specified syllable-boundary pulses, the converter computes time intervals between contiguous pulses by an algorithm which is called a shadow computation (see below).

At the input level for the converter, utterance conditions such as speed of utterance, formality of utterance, and speaker idiosyncrasy (in multidimensional measures) are numerically specified. These affect the pulse train *via* adjustment of shadow slopes. This pulse train functions as the total prosodic control of the utterance (to the extent that the current approximation is effective) and determines the non-uniform temporal overlapping of gestures in each articulator. It should be noted, however, that the syllable type (heavy *vs.* light syllables, *etc.*), as a phonological property of the syllable represented by the feature specifications, controls the shadow coefficients, which affect the time intervals between contiguous syllable-boundary pulses (see Fujimura [1994a]). The numerical augmentation of a tree node for prominence, as an utterance specification, does not affect the shadow slopes.

The distributor interprets the feature specifications to distribute corresponding elemental gestures to pertinent articulatory dimensions to be implemented by specific articulatory organs, generating elemental gesture specifications for the next component, a parallel set of actuators. The parallel set of actuators generate time functions by exciting pertinent IRFs by the syllable pulse, which determines timing and amplitude of each IRF. Different IRFs are then superimposed in each dimension to form the time function of the articulatory control for phrasal units.

FEATURE SPECIFICATION

The syllable structure analysis in the C/D model adopts the principle of demisyllabic analysis [Fujimura 1976, 1979; Fujimura and Lovins 1978], that consonant clusters (in English) do not require any ordering specification within the syllable core, after separating out syllable affixes. This principle recognizes CVC as the canonical syllable structure of English syllables, where the initial C can be zero, but the final C is mandatory unless the syllable is reduced as a supplement to the head (with stress) of a foot (there may be more than one such subordinate syllables). Tense vowels and diphthongs in English are treated as a combination of a vowel (V) and a glide (C). The syllable affix to the left of the core is called a p-fix (not applicable to English) and that to the right is called a s-fix.

The C in onset and coda, optimally an obstruent, marks the edges of the syllable core, to which a s-fix (or order-specified string of s-fixes) can be attached, when certain strong constraints are met for each consonant to qualify for the status of a s-fix. The p-fixes are similar in a mirror-image situation. The basic assumption is that within the core, in either onset or coda, no sequential ordering of features is given. Therefore, feature specifications, including sonorant features, for either onset or coda, are given as a set (not sequence) of several privative feature specifications, which may be divided into concomitant feature types (such as place and manner). The temporal organization of tautosyllabic articulatory gestures automatically emerges as the inherent properties of the evoked IRFs.

For this principle to work in English, it is critical to assume an abstract feature called {spirantized}, representing the combination of apical frication and oral closure in the phonemic consonantal sequences /sp/, /st/, and /sk/ in both initial and final position. This feature is an obstruent feature, as a member of the manner feature paradigm opposing it to {stop}, {fricative}, {interdental} and {nasal}. The features {spirantized}, {stop}, and {nasal} all require a place specification and are implemented with an oral stop closure (the place-specified closure is delayed for {spirantized} relative to the frication production according to the pertinent IRF properties). This feature also corresponds to the same phonemic sequences in the coda (*e.g.* 'task' /tæsk/ as opposed to 'tax' /tæk.s/ which contains a s-fix outside the core, as indicated by a dot in the phonemoidal transcription).

It should be mentioned here that in English, there are many syllabic sonorants (as in 'button', 'bottle'), that must be treated as separate syllables, even though, phonetically, there is no vowel. These are not s-fixes, since they do not satisfy the requirement for s-fixes that the voicing status must agree with that in the coda. Japanese also has many cases of phonetically nonexistent (or devoiced) high vowels. These syllables contain vocalic specifications which cause a minimal distinction between /i/ and /u/ in the devocalized environment. In addition, these hidden vowels always show up when the intonation pattern requires a raised pitch, as observed toward the end of a question sentence.

One critical problem in connection with the discussion of possible syllable structures is how to define syllables as abstract phonological units. Before we discuss where syllable boundaries are in polysyllabic forms, we will first be concerned with the existence of syllables, identifying syllable nuclei which may not be phonetically apparent. Some guiding principles in identifying phonetically hidden syllables may be formulated as follows.

(1) A syllable must have at most one continuous stretch of voiced portion in the phonetic signal. If a word manifests itself with an unvoiced portion surrounded by voiced portions on both sides, there must be assumed more than one syllable. Thus the sonority principle (see Clements [1989] and Fujimura [1989]) with respect to phonetic voicing should be observed with the strongest priority (at the top of the constraint hierarchy in the sense of optimality theory, see Prince & Smolensky [in press]), and probably universally (as an absolute requirement).

(2) Consonant clusters at the left and right edges of a phonological word often contain syllable affixes, which are often but not always morphological affixes. The separable affixes (p-fixes and s-fixes) must be strongly limited in phonological feature specifications, and the phonetic voicing status continuously spreads from the onset (backward) or coda (forward) toward the word edge,

thus requiring no feature specification for voicing in affixes. If there is a change in voicing at a syllable edge, as in German initial /kn/ (in 'Knabe') and English final /nt/ (in 'tent'), the two consonantal elements must be both contained within the syllable core. There is strong phonetic evidence that a phonemic minimal pair like /tɛnt/ and /tɛn.d/ must be treated differently (see Fujimura & Lovins [1978]).

In this situation, it is likely that one of the consonants as a phoneme has no paradigmatic opposition in place. In English, the final phonemic sequence nasal + voiceless obstruent must be homorganic. In German, for example, /km/, is not allowed, and therefore, the specification for the /n/ element in the cluster /kn/ is {nasal} without any place specification. In English, it can be shown that at most one place specification is allowed for the onset or coda, and none is given for s-fixes. Note again that the feature set {spirantized, labial} for the English words 'spoon' or 'grasp', for example, does not specify the place for /s/.

Likewise, the feature {lateral} in English, does not have any place specified in our analysis, allowing a distinction between 'slight' and 'flight', for example, with only one place specification for the onset (*cf.* 'smell' *vs.* 'snell' for which the place specification is for the nasal element, not for the /s/, reflecting the distributional fact that there is no opposition /s/ *vs.* /f/ in this onset environment). The feature {lateral} automatically evokes the apical gesture for an alveolar contact in onset position as an elemental gesture, by looking up a feature-gesture table. It evokes a similar coronal gesture in coda in some dialects of American English but not necessarily. The most robust inherent gestures seem to be tongue blade narrowing and body retraction (see Sproat & Fujimura [1993]). Therefore, at least in coda, the lateral in English cannot be specified with a place feature from a phonetic point of view.

While the C/D analysis treats onset and coda gestures basically as independently assigned gestures, it does use the same feature name vocabulary in onset (with a superscript *o* as needed) and in coda (with *c*).

(3) When more than one p-fix or s-fix is allowed in the language (as in English for s-fixes, *e.g.* in 'sixths'), ordering of feature specifications for the sequence of affixes is required. The inventory of phonemic segments treated as syllable affixes must be small, and their feature specifications are given parsimoniously (only a manner specification in English). Apart from this paradigmatic parsimony, affixes behave like phonemes: they form a temporal string with specified sequential ordering. They are phonetically very stable, allowing, for example, segmental waveform concatenation.

Syllable affixes (p-fixes or s-fixes) may be assumed to occur only at the edges of phonological words (or morphemes).

In the C/D model, unlike the earlier demisyllable analysis [Fujimura 1976, 1979], vowels are treated separately from consonants throughout the computational process, from the feature specification level (*i.e.* input to the converter) to the control time function level (*i.e.* input to the signal generator). For this reason, the demisyllable approach is adopted in the C/D model only with respect to consonantal features and gestures.

A minimal underspecification scheme by means of privative (unary) features is used for the input representation in the C/D model. For example, in English, the first syllable /skrʌmp / of 'scrumptious' has an onset specified as {dorsal, stop, spirantized, rhotacized}, and a coda {labial, nasal, stop} (no meaningful ordering of features intended). The voicelessness for onset or coda is not specified because obstruent manner features without {voiced} implies unvoiced, implemented as a voice cessation (vocal fold abduction) at the edge of the syllable.

CONSONANTAL TIMING

According to the original first-approximation scheme presented in previous publications, the C/D model specifies that the internal timing relation between the initial and final gesture peaks will remain the same regardless of the magnitude of the syllable. Therefore, when the syllable is reduced, other things being equal, the duration of the acoustic vowel segment probably will increase when the margins are unvoiced, because the glottal abduction gesture will be reduced and the duration of the vocal-fold vibration will be expanded.

This particular difficulty could be resolved by assuming that the default condition for voicing was unvoiced, as in a pause, and each syllable pulse evokes in the laryngeal dimension an adduction gesture (as opposed to the scheme where voicing is the basic gesture unless obstruent features or phrase boundary features evoke voice cessation, *i.e.* laryngeal abduction). The observed voiced duration in the acoustic signal then would depend on the characteristics of the signal generator, balancing the durations of the voiceless consonants and the vowel portion in the resulting acoustic signal as dictated by the nonlinearity of the production mechanism. The vowel elongation due to syllable reduction is, of course, counterfactual, while the shortening of consonantal segments is factual. Which approach is more nearly correct as an approximation theory is an empirical issue. In either case, the total syllable duration, or more exactly, the time interval between contiguous syllables as represented by the syllable pulses, is distinctly shorter (proportional to the syllable pulse magnitude) when the syllable is weak.

An alternative general solution of this problem and some others can be provided by specifying a little more detail of the mechanism that evokes IRFs, without affecting the principle that all prosodic structure of speech articulation is computed *via* the time and magnitude evaluation of the syllable and boundary pulses. The current idea is as follows.

A syllable pulse generates separate pulses for the onset, the coda, and each of the affixes. Each of these subsidiary pulses (which may be called *pocs* pulses, standing for p-fix, onset, coda, and s-fix) evokes the IRFs. The pocs pulses inherit the magnitude of the parent syllable pulse. Pocs pulses have shadows which extend only outward from the syllable pulse, the center being the syllable pulse under discussion which excites vocalic and prosodic gestures directly.

The onset pulse is erected at the external (*i.e.* left) end of the left-hand shadow of the syllable pulse, and the coda pulse is erected at the external (right) end of the right-hand shadow of the syllable pulse. The most internal p-fix requires a p1-pulse erected at the external end of the shadow of the onset pulse, and the next external p2-pulse stands at the external edge of the p1-pulse. The s1-pulse, s2-pulse, *etc.* for the s-fixes are similar, forming a mirror image. The most external edge of the shadows to the left or right of the most external component of the syllable determines the temporal limit of the syllable in question as a whole, and the contiguous syllable or boundary pulse is placed to make this limit coincide with its associated external temporal limit, *i.e.*, the edge of its most external pulse shadow.

The pocs pulses generally delimit the time domain in which articulatory gestural activities of the pertinent syllable component (p-fix, onset, coda, or s-fix) are primarily contained. The syllable pulse covers primarily the vocalic activities corresponding to either vocalic or consonantal features, but tense consonantal gestures tend to invade into this vocalic time interval. Note that the IRFs are continuous time functions and never exhibit any sharp boundaries for activities. The segmental discontinuity as observed in the acoustic signals arise due to nonlinearity of the signal generating process. The onset pulse is the pulse that triggers IRFs of onset elemental gestures, the values of the IRFs being subdued and their acoustic effects tending to be invisible beyond the shadow edges particularly if any gesture of the next external component (tautosyllabic or heterosyllabic) manifests predominating effects. The most internal s-fix pulse (S1) marks the nominal end of the coda gesture activities, and the next external s-fix pulse (S2) marks the nominal end of the internal s-fix. The p-fix situation is a mirror image.

Some readers may find it awkward to see a response of the triggering pulse temporally before the latter occurs: the IRFs we consider in this description are not physically realizable. This is a matter of convenience of the description. There is always a considerable delay between the cortical motor control planning and the physical execution, even if we take this model to represent a direct computational simulation of the physiological process of speech production, the actual triggering pulses must occur well ahead of the hypothetical time values of syllable or boundary pulses. Shifting the time values of all pulses by a sufficiently large and universally fixed time interval as a constant delay of responses resolves this seeming contradiction.

This revised scheme of timing computation using pocs pulses makes the intrasyllabic temporal relation between initial and final consonantal gestures more directly sensitive to the syllable pulse magnitude in general. Also, since the slopes of shadows for syllables are sensitive to the internal structure, reflecting the syllable type, the apparent vowel duration may vary not only reflecting the prominence condition and speed of utterance, but also whether the syllable is specified for a long or short vowel. In our analysis, a phonologically long vowel is specified with a "monophthongal" glide *i.e.*, the elongation feature {long^{c}}), and a diphthongal vowel with a more conventionally recognized glide ({palatalized^{c}}, *etc.*).

One distinct advantage of this pocs approach is the differential treatment of vocalic gestures from consonantal gestures. As Sproat & Fujimura [1993] pointed out, lateral and nasal consonants exhibit different intrasyllabic timing behaviors between what may be considered vocalic vs. consonantal gestures, while, phonologically, both {nasal} and {lateral} are consonantal manner features. Vocalic gestures, such as velum lowering and tongue body retraction, seem more closely linked to the center of the syllable, while tongue tip or lip gestures are linked to the margins of the syllable. Assuming that vocalic gestures are evoked by the syllable pulse while consonantal gestures are evoked by the onset or coda pulse, the correlation between the relative timing difference between the consonantal and vocalic gestures to prosodic conditions as observed in the articulatory studies (see also Krakow [1989] may be accounted for by a general phonetic principle as prescribed by the C/D model.

There are many additional details of the model that have to be worked out. The comparison of prediction with observation is not easily achieved, but has to be approached step by step in successive approximation comparing data and the updated tentative descriptive framework for interpreting data. The signal generator brings the generative description of this theory closer to direct modeling of the speech production process.

ACKNOWLEDGMENT

The author wishes to express his gratitude for the support he has been given by the ATR Interpreting Telephony Research Laboratories and the ATR Human Information Processing Research Laboratories. It is also acknowledged that this article is based in part on a longer and earlier manuscript for the Proceedings of LP'94 -- Item Order in Natural Languages, edited by B. Palek and P. Janota, to be published from Charles University, Prague, with the editors' permission.

REFERENCES

Browman, C.P. & Goldstein, L.M. (1992). Articulatory phonology: An overview. *Phonetica* , **49**, 155-180.

Chomsky, N. & Halle, M. (1968). *Sound Pattern of English.* Cambridge MA: MIT Press.

Clements, G.N. (1989). The role of sonority cycle in core syllabification. In J. Kingston and M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, (pp. 283-333). Cambridge, UK: Cambridge University Press.

Fujimura, O. (1970). Current issues in experimental phonetics. In Jakobson, R. & Kawamoto, S. (eds.) *Studies in General and Oriental Linguistics* (pp. 109-130). Tokyo: Tec Co.

Fujimura, O.(1976). Syllable as concatenated demisyllables and affixes. *J. Acoust. Soc. Am.*, **59**, Suppl. 1, S55.

Fujimura, O. (1979). An analysis of English syllables as cores and affixes. *Z. f. Phonetik, Sprachwiss. u. Komm.*, **32**, 471-476.

Fujimura, O. (1989). Demisyllables as sets of features: Comments on Claimants' paper. In J. Kingston and M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 334-340). Cambridge, UK: Cambridge University Press.

Fujimura, O. (1990). Methods and Goals of Speech Production Research. *Language and Speech*, **33**, 195-258.

Fujimura, O. (1992). Phonology and phonetics- A syllable-based model of articulatory organization. *J. Acoust. Soc. Japan* (E), **13**, 39-48.

Fujimura, O. (1994a). Syllable timing computation in the C/D model. *Proceedings of the Third International Conference on Spoken Language Processing, Yokohama* .

Fujimura, O. (1994b). C/D model: a computational model of phonetic implementation, in E. Ristad (ed.), *Language Computations* (pp. 1-20). Providence, RI: Am. Math. Soc.

Fujimura, O. (in press). The syllable: its internal structure and role in prosodic organization, Palek, B. & Janota, P. (eds.), *Proceedings of LP'94: Item Order in (Natural) Languages.* Prague: Charles Univ.

Fujimura, O. & Lovins, J. (1978). Syllables as concatenative phonetic units. In A. Bell and J.B. Hopper (eds.), *Syllables and Segments* (pp. 107-120). Amsterdam: N. Holland.

Fujimura, O., Erickson, D. & Wilhelms, R. (1991). Prosodic effects on articulatory gestures-a model of temporal organization. *Proceedings of XIIth International Congress of Phonetic Sciences, Aix en Provence.* Vol. 2, 26-29.

Fujisaki, H. (1988). A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In O. Fujimura (ed.), *Vocal Physiology: Voice Production, Mechanisms and Functions* (pp. 347-356). New York: Raven Press.

Krakow, R.A. (1989). *The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velar Gestures.* PhD diss. Yale University.

Liberman, M. & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, **8** , 249-336 .

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, **39**, 151-168.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation.* PhD Dissertation, MIT (distributed by the Indiana University Linguistics Club).

Prince, A. & Smolensky, P. (to appear). *Optimality Theory,* Cambridge MA, MIT Press.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic variation. *J. Phonetics* **21**, 291-311.

Wilhelms-Tricarico, R. (in press). Physiological modeling of speech production: Methods for modeling soft-tissue articulators. *J. Acoust. Soc. Am.*