TR－H－103

Syllables, Internal Structure and Role in Prosodic Organization

Osamu Fujimura

1994. 10. 18

(1994. 9. 16受付)

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2　☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-7749-5-1011
Facsimile: +81-7749-5-1008

ATR/HIP Report
September 1994 (corrected 17.9.94)

To be revised and submitted for
Proceedings of LP'94: Item Order in (Natural) Languages, Prague, August 16-18, 1994,
edited by B. Palek, Institute of Linguistic and Finno-Ugric Studies, Charles University
2, Jan Palach Sq., 116 38 Prague 1
Czech Republic

Syllables, Internal Structure and Role in Prosodic Organization

by Osamu Fujimura
Department of Speech & Hearing Science, The Ohio-State University
Columbus, OH 43210-1002, U. S. A.

1. *Introduction*

This paper discusses a new view of speech organization in relation to the Converter-Distributor (C/D) model of phonetic implementation [Fujimura *et al.* 1991; Fujimura 1992, 1994, in press]. Traditionally, speech signals were interpreted basically as a concatenated string of phonemic segments, each of the segments being represented as a simultaneous bundle of distinctive features [Jakobson, Halle & Fant 1951/52/63]. In this influential classical work, each segment is assumed to be completely specified with all phonetic characteristics as an integral and independent phonetic form, each component feature being associated with its inherent, though inevitably abstract, phonetic manifestation. On this linear string of segmental phonetic events, suprasegmental effects were assumed to be superimposed to form the observable speech signals. Some smoothing process, generally called coarticultion [Lindblom 1963], would be applied to a set of step functions representing phonetic dimensions (such as formants or articulators' positions) formed out of discretely concatenated target values and assigned durations for individual (phonemic) segments, to generate continuously changing and physically realizable time functions for phonetic variables (see Fujimura [1967, 1972]; Vaissiére [1988]). The suprasegmental phenomena have been discussed referring to separately observable speech characteristics, in particular, the voice fundamental frequency (pitch) contour (i.e. time functions) and spectrographically defined segmental durations [Lehiste 1970]. Thus, concatenated string of phonemes, roughly corresponding to alphabetic text, is the primary (abstract) representation of the speech material, modulated by suprasegmental control superimposed on it in actual utterances.

Our new view to be discussed here deviates radically from this interpretation of speech phenomena. It assumes what we will call *prosodic*[1] *organization* of an utterance or its phrasal components (as a phonetic unit) as the basic structure of speech phenomena. This structure is associated with a linear string of syllables as the concatenative "segmental" units. The flow of vocalic gestures characterizing the sequence of syllable nuclei forms the *base function* of the articulatory event that fits in the prosodic structure of the utterance. On this base function, consonantal gestures are superimposed, basically in the way Öhman [1967] depicted in his consonantal perturbation model (see also Carré, R. & Chennoukh, S. [submitted]). The base function is inherently multi-dimensional in the sense that different articulators such as the jaw opening, the tongue body advancing or retraction, and the lip rounding and protrusion, behave more or less independently from each other, and some of these dimensions, in particular, presumably, the mandible abduction/adduction, more directly reflect the prosodic structure. Onto this base function is superimposed consonantal gestures reflecting inherent characteristics of phonological features representing each syllable margin item, or more specifically, in the C/D model terminology, features of onset, coda, or syllable affix(es). The prosodic characteristics of an utterance are specified phonologically by a metrical tree [Liberman & Prince 1977] (or some other symbolic representation, see, *e.g.*, Idsardi [1993]). In the phonetic specification of the speech utterance, which we will need as the input for the C/D system, the metrical tree must be augmented by numeric annotations. The intricate relation between the symbolic, discrete phonological representation and the corresponding continuously variable phonetic characteristics of utterances of the given linguistic message in a given situation is thus explained by an explicitly described generative model of phonetic realization.

To the extent that speech organization is described in terms of articulatory gestures, this theory is similar to the articulatory phonology proposed by Browman and Goldstein [1986, 1988, 1990a, 1990b, 1992a, 1992b]. There are many phonetic observations, particularly allophonic variation of phonemes in the traditional segmental description that are naturally explained by either theory, as the consequence of describing speech phenomena as the result of assembling articulatory gestures. Such variation is typically sensitive to the style of utterance, among other factors. The two theories are also basically different from each other. While articulatory phonology assumes gestures to be the basic ingredients in the lexical phonological representations, apparently eliminating the division of labor between phonology and phonetics, the C/D model respect the distinction between phonology and phonetics according to the linguistic tradition. In terms of the generative phonological theory, one can state that the C/D model receives the output of phonology in the form of metrical tree (or equivalent prosodic representation of the syntagmatic organization of the speech material to be used for an utterance) and phonological feature specifications of the paradigmatic choices of the linguistic materials. Thus, the phonetic structure reflects linguistic control. At the same time, as a phonetic theory, it computes the phonetic signals considering explicitly other (numerical) specifications characteristic of the particular utterance, according to the situation of speaking. In articulatory phonology, in contrast, linguistic control is specified in the form of gesture scores in the lexicon, and the temporal organization is governed basically by biological principles. More specifically, the C/D model assumes syllables to be the basic ingredients of phonological materials that are concatenated into a temporal linear string. The temporal organization of syllables intervened with boundaries for phonetic phrasal organization is represented by a series of magnitude and timing modulated pulse train. The timing of each syllable is derived from the assignment of the magnitude (abstract phonetic strength) of the syllable in the particular context of utterance, and is a phonetically controlled entity. The articulatory phonology, in contrast, assumes an organization of articulatory gestures according to biologically governed principles (*i.e.* task dynamics, see Kelso *et al.* [1986], Saltzman [1985], but also Kröger [1993]), the transfer from one gesture to another being determined by the property of the time functions of the gestures themselves.

This C/D model attempts to offer a bridge between the mathematically well-formulated grammatical description of phonological patterns on the one hand, and continually varying situation-sensitive functional characteristics of phonetic forms, which have escaped any comprehensive and quantitative description in the past, except in the area of intonation theory. The model uses articulatory processes as the descriptive framework, and it is important to emphasize that the signal generator component determines critical characteristics of directly observable physical measures such as articulatory movement patterns and, based thereon, acoustic or spectrographic variables. Therefore, the prescription of articulatory gestures and acoustic and perceptual properties of speech signals, as represented in the articulatory control functions that are composed by the set of actuators, the third component of the model, 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, we claim, by evaluating physically observed signal characteristics, such as x-ray microbeam data [Fujimura *et al.* 1973, Kiritani *et al.* 1975, Nadler *et al.* 1987] and acoustic signals, through the use of powerful computational techniques handling a large mass of data. The effects of nontrivial and nonlinear mapping from the phonetic control functions to observable physical variables can be represented either by direct computational simulation of the articulatory system, or by simplified and systematically schematized computation (including table lookup) based on the information derived from the simulation. An application of the layered abduction by John Josephson, my colleague at OSU, with Kevin Lenzo and some other students, is being explored for this evaluation process, in a novel form of automatic speech recognition.

In this paper, the C/D model is briefly reviewed, even though details are omitted, and its implications are discussed focusing on (1) the general intrasyllabic structure that is assumed in this theory, and (2) the role of syllables in the prosodic organization of speech as depicted in this model of phonetic realization. For supplemental information, the reader is referred to previous publications [Fujimura 1992, 1993, 1994].

2. Some properties of the C/D model

The C/D model describes the phonetic implementation process in three sequentially ordered components, converter, distributor, and a parallel set of actuators, as shown in Fig. 1. The process is inherently multidimensional and superpositionally linear until the set of control time functions are derived, as the output of the actuators, and are delivered to the signal generator. The signal generator, which is a computaional simulation of the physiological and physical system of the human speech production mechanism, constitutes the fourth component of the model. This last component is a complex, highly nonlinear and inherently three-dimensional dynamic system.

The C/D model treats syllables as basic "segmental" units of an utterance, while virtually all theories of phonology with few exceptions, the syllable is a "prosodic" unit comprising segmental (root) elements sequentially ordered. Syllables are the minimal utterable phonetic units, which can be concatenated into a temporal string, interrupted by boundaries. The specification of the utterance begins with the phonological phrasal representation of the sentence-like material to be uttered, and this input (linguistic) representation is augmented by numerical specifications of utterance conditions to complete the phonetic specifications. Thus syllables serve as the link between the segmental specifications and the syntagmatic (prosodic in our sense) elemental units for phrasal formation (the terminal nodes of the tree). The segmental specifications are provided as underspecified privative features. These phonological specifications are passed by the converter to the distributor as corresponding phonetic gesture specifications, with necessary supplemental specifications of redundant information. The phonetic gesture specifications are interpreted by the distributor in terms of elemental articulatory gestures. The converter's role is to evaluate the prosodic pattern as specified in the input in the form of augmented metrical tree, to compute the phonetic strength of each syllable and accordingly assign the magnitude value to each (one dimensional) impulse representing the syllable. The converter also creates a boundary pulse by evaluating the tree configuration of the syntagmatic specification (metrical tree), and assigns the magnitude value of each boundary pulse representing each boundary. In computing the pulse magnitudes, the converter takes into account the numerical phonetic specifications of the utterance conditions and tree augmentation indicating the extrinsic prominence status of any syllable (through numeral marks attached to any dominating tree nodes). Tentatively, it is assumed that this conversion of symbolic representation into numerical specifications of syllable occurrences is performed by the following procedure: Compute the metrical grid based on the metrical tree representation [Liberman & Prince 1977]. Set the magnitude values of the corresponding syllables to be proportional to the grid height (the proportionality coefficient is sensitive to utterance conditions). Identify the terminal syllable nodes that are dominated by the tree node marked by a numerical prominence mark. Multiply the syllable magnitudes by the value of the prominence mark of their dominating node. Repeat the process until all prominence marks are exhausted.

A syllable pulse (thick vertical bars in Fig. 1) associated with the computed magnitude (μ1, μ2, *etc.* shown as bar height) and feature specifications (in Greek letters for manner and associated small Roman capitals for place in converter output) is thus produced for each syllable. A time value (ti) of each syllable pulse is computed based on the magnitude pattern by using a linear "shadow' function algorithm (Fujimura [1992], section 3). A series of such pulses forms a train. In the pulse train, another type of pulse representing boundaries of different types are inserted (thin bars), also with time and magnitude values, computed according to the phonological representation (metrical tree). Note that this syllable-boundary pulse train controlling the prosodic characteristics is one-dimensional, and carries the "prosodic" information of the utterance unit under question.

Numerically specified utterance conditions such as speed of utterance, formality of utterance, speaker idiosyncrasy (which eventually may be given in terms of continuous measures in many dimensions), as indicated at the upper left corner of Fig. 1, affect the configuration of the pulse train.2 This pulse train functions as the prosodic control of the utterance and determines nonuniform temporal overlapping of gestures. But, at this abstract level of phonetic representation, none of the prosodic conditions including the numerically specified prominence directly affects elements of articulatory movement patterns, apart from the amplitude setting of the IRFs (henceforth IRFs). The basic assumption is that all prosodic information is absorbed into this pulse train representation (possibly with some limited scheme of amendment for a higher order approximation, which is not considered now). It should be noted, however, that the syllable type *via* the feature specifications of each syllable is considered together with the magnitude of the syllable pulse, in computing the time intervals between contiguous syllable/boundary pulses. As discussed in Section 3.5, this consideration of the syllable type is implement ed *via* defining onset-coda-affix pulses subordinate to each syllable pulse, in a recently revised version of the C/D model (unpublished).

The distributor interprets the feature specifications and generates elemental gesture specifications for the next component, which is a parallel set of actuators. The actuators generate time functions for individual articulatory dimensions, superposing stored IRFs evoked by the time series of syllable pulses, to form a single-dimensioned control time function for the utterance unit in each articulatory dimension (see Fujimura 1992). The set of output time functions as multidimensional control signals are fed into a computational simulation model of the physical articulatory system, *i.e.* the signal generator, to produce articulatory movement patterns and the resultant acoustic signal.

In what follows, rather than to give a formal description of the model, which is outlined elsewhere in publications and is being worked on in further details, some predictable consequences of the assumptions made in this theory will be described. Before doing so, however, we need to discuss some salient properties of the phonological feature specification system that is assumed with respect to the input form of the C/D model, since the choice of specifications is impotant for the model to work effectively.

3. Phonological Feature Specification Scheme and Demisyllabic Analysis of Syllable Structure

The C/D theory is based on an idea which was advocated in the current author's demisyllabic analysis of syllables [Fujimura 1976, 1979; Fujimura and Lovins 1978], that consonant clusters (in English) do not require any ordering specification if features are defined with some special consideration of the distributional and phonetic facts, and if syllable affixes (prefixes or suffixes) are separated from the syllable core. In the case of English (and apparently many other Indoeuropean languages), the critical feature is what is named {spirantized}, representing the /s/ part of the consonantal cluster /sp/, /st/, and /sk/ in both initial and final position as an obstruent manner feature, as in 'sky' /skaJ/ and 'task' /tæsk/ (as opposed to 'tax' /tæk.s/[4] which contains a phonological and not morphological suffix (s-fix), see *infra*). The separation of s-fixes, according to a strict principle, results in a fairly limited variety of consonants in the remaining (core) part of the syllable, and the introduction of the feature {spirantized} dispenses with the necessity of order specification in English (syllable-final, as well as initial) consonantal clusters all together. This analysis has been discussed in detail for English [Fujimura 1976, 1979, 1992], but not for other languages (but see Eek&Help 1987; Clements 1989; Maddieson 1992). The basic framework is assumed to be universal for any language, but detailed research is needed for different types of languages to substantiate the claim. English happens to be very interesting and rather revealing about final consonant clusters. Some of the presentations, particularly by Steriade, Leben, and Maddieson, at the informal C/D-model workshop at OSU in the summer 1993, pointed out apparently intriguing counterexamples. Putative C/D solution of some typical types of such counterexamples will be discussed elsewhere. It should be mentioned that even in English, there are many so-called syllabic consonants, in particular sonorants, that must be treated as a separate syllable, even though phonetically there is no vowel. Japanese also has many cases of phonetically nonexistent high vowels, which must be assumed to exist at the abstract level for morphological, if not any other, reasons.

One critical problem in connection with the discussion of possible syllable structures is how to define syllables as abstract phonological units. Like other abstract linguistic units in generative descriptions of language, we assume certain guiding principles and hypothesize a particular syllable-based lexical representations, and see if it results in a consistent and effective lexical and phrasal representations. Before we discuss where syllable boudaries are in polysyllabic forms, we will first be concerned with existence of syllables, identifying syllable nuclei. Some syllable nuclei are not obviously identifiable in orthographic representations. For example, the Czech name 'Petr' may be arguably bisyllabic, both phonetically (considering the voicing pattern, see below, as well as the sonority cycle) and distributionally (simplicity and regularity of the resultant syllable structures assuming that there is a word-medial syllable margin, whatever its syllable affiliation may be), just like English 'Peter' is, but the Czech orthography does not show any vowel beyond the first.

Some guiding principles in identifying phonetically "hidden" syllables are, according to the author's preliminary study:

(1) A syllable must have one and only one continuous stretch of voiced portion in the phonetic signal, unless there is a phonetic reason that voicing has been affected or eliminated due to the unvoiced phonetic environment. If a putative syllable 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 voicing should be observed absolutely and universally.

(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 (underlying) feature specifications, and most probably (as it is the case in English), 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 within an apparent (phonemic) consonant cluster, as in German initial /kn/ (in 'Knabe') and English final /nt/ (in 'tent'), therefore, the two consonantal elements must be both contained within the syllable core.[5] It is likely that one of the consonants as phonemes has a severely limited distribution in terms of place, most likely involving no place specification in the given intrasyllabic environment. In German, for example, /km/, *etc.* is not allowed, and therefore all the specification for the /n/ element in the cluster /kn/ is {nasal} and nothing else, maintaining the number of place specifications in the onset (cluster) to be only one. In English, it can be shown that at most one place specification is allowed in onset or coda, and none is allowed in s-fixes.

(3) In some languages, two phonetic obstruents may be involved in the onset, each of which has to be place-specified because of phonemic oppositions such as /bg/ vs. /gb/. This may appear to refute the concept that temporal ordering of segments (or features) is necessary. However, the available phonetic observation seems to be that the second (closer to the nucleus) obstruent gesture in such an initial cluster is consistently weaker than the first (further away from the nucleus) obstruent gesture, and the choices of place in the same context seem relatively limited for the inside consonant. In such a case, as observed in intial voiced stop sequences of Eggon by Maddieson [personal communication], the C/D interpretation, as a tentative analysis, would be that the outside stop gesture menifests a place-specified obstruent feature, while the inside weak obstruent gesture manifests a glide (it is either labial rounding or dorsal raising for velarization, and there are no glides that form an opposition to the weakened stops in this language).[6]

(4) When more than one p-fix or s-fix is allowed for the language (as in English for s-fixes, *e.g.* in 'sixths'), ordering of feature specifications for the sequence of affixes is

necessary. The inventory of phonemic segments treated as syllable affixes must be small, the specifications for each of them in a string being given by only a very limited number of features (one manner specification only in English). Apart from this paradigmatic parsimony, each affix behaves like phonemes: they form a temporal string with specified sequential ordering.

(5) The mapping from the set of phonological features to the set of phonetic gestures are many to many. For example, in English, {strident} evokes a frication gesture much like {apical, fricative} (*i.e.* /s/), resulting, in part, in a [s]-event as a phonetic segment. Note, however, that {strident} is always implimented by the tongue tip, and it also implies a stop closure, whether implying a concomitant (*i.e.* in the same onset or coda} {stop} (as in English 'spell' or 'ask') or {nasal} (as in English 'smell').[7]

In the C/D model, unlike earlier demisyllable analysis [Fujimura 1976, 1979], vowels are treated separately from consonants all through 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. To avoid confusion, in this document after separating out syllable affixes (*i.e.* within the syllable core), initial (demisyllabic) consonant complexes will be called onsets, and final demisyllabic consonant complexes including the gliding or vowel elongation element in diphthongs will be called codas (see Fig. 2). References will be made to demisyllables when signal properties, particularly acoustic characteristics, are discussed, since these signals inevitably reflect vowel articulations together with consonantal gestures.

A minimal underspecification by means of privative (unary) features is used for the (phonological) input representation in the C/D model. In English, place is specified only in conjunction with an obstruent feature, and at most one place can be specified for each onset or coda. Table 1 shows feature specifications and associated gestures for single consonants and a few clusters (from the phonemic point of view) in English. As an example of more complex clusters, the syllable/skrAmp/ as in 'scrumptuous' may be considered. The onset is specified as {dorsal, stop, spirantized, rhoticized}, and the coda as {labial, nasal, stop} (no meaningful ordering of features intended). The voicelessness for onset and coda is not specified because {spirantized} and {stop}, respectively, without {voiced}, and therefore phonetically there is a voice cessation (vocal fold abduction) in both margins. The feature {spirantized} evokes an oral closure, just as {affricate}, {stop} and {nasal} do. Obstruent features and {nasal} call for a place specification. (In terms of feature geometry, an "obstruent" node dominates "place" and "manner".) The manner feature {spirantized} is always implemented with a voiceless apical frication. When it is combined with {nasal}, it is also implemented with an oral closure temporally following the frication, but since the frication is unvoiced and the nasal murmur is inherently voiced, this manner feature combination can not occur in coda (if it did, then there would be more than one contiguous stretch of voicing within the syllable, see above). Note the /sm/ in 'smoke' is specified by {labial, spirantized, nasal}, reflecting the lack of 'fmoke', but /sl/ in 'slash' is specified by {apical, fricative, lateral} (the specification of apical articulator is for /s/ and not for /l/), reflecting the contrast with 'flash'. Likewise, {affricate} is implemented with (either voiced or unvoiced) frication, and in English (but not in German), the place is palatal (the word-final /t.s/ contains a s-fix, and /ts/ does not occur in onset in English except in borrowed forms). Phonetically, in American English, the fricative /v/ in forms like 'prove' is often pronounced with a [bv]. Lateral and rhoticized consonants involve no "place" specifications.

The English morphemic voiceless interdental theta is peculiar in that it behaves exceptionally freely resulting in a large number of s-fixes in a word (as in 'sixths'; *cf.* the morphemic /s/ which avoids a similar situation by forming a separate syllable as in 'sixes'). It also creates a small set of apparently exceptional words like 'width' and 'warmth', which used to be treated as exceptions in previous C/D papers but are now treated within the core, using {interdental} as a manner feature.

To summarize salient characteristics of the C/D analysis of the syllable structure, we may list the following speculative points[8] as universal properties of languages, and propose some descriptive convention:

(1) A syllable consists of a core (the central part), p-fixes, temporally preceding the core, and s-fixes, temporally succedding the core. The p-fixes and s-fixes are sequentially ordered. The numbers of permissible p-fixes and s-fixes are designated for the language[7].
(2) The core consists of core components, *viz.* syllable nucleus, onset, and coda, each of which are represented by an unordered set of phonological features. The same feature specification cannot be given more than once within each core component.
(3) The nucleus is specified by a set of privative vocalic features designated for the language/dialect, such as {front, back, high, low, rounded}.
(4) The onset is represented by a set of privative consonantal features named with a superscript o. The set of features allowed for specifying the onset is designated for the language.
(5) The coda is represented by a set of privative consonantal features named with a superscript c. The set of features allowed for specifying the coda is designated for the language.
(6) Each of the syllable affixes, *viz.* p-fixes and s-fixes, are represented by a small number of privative consonantal features, marked with a superscript p or s with a number attached. The number specifies the intrasyllabic position of the affix in the linear ordering counting from inside out (the number is omitted if the language does not allow more than one p-fix or s-fix). No feature is specified more than once within each syllable affix.
(7) Presumably, syllable affixes occur only at the edges of a phonological word. There is also a strong interaction between the permissibility of an affix and the morphological status of the phonological element. English has at most one s-fix, which is always an obstruent implemented by the tongue tip, and the s-fix occurs only at the end of a monomorphemic word; no p-fixes are allowed. Another s-fix (or rarely more) can follow if it represents a specific morphological suffix, as in 'sixths', where the /s/ following the core final voiceless /k/ is a regular s-fix with the feature {fricative} specified, and the {interdental} and {fricative} manner specifications determine the two succeeding morphemic s-fixes, respectively.
(8) Possible obstruent features are designated for the language, and they are grouped into two feature types, *viz.* place features (such as {apical}, {labial}, {dorsal}, {palatal}) and manner features (such as {stop}, {fricative}, {affricate}, {interdental}, {labiodental}[9], {spirantized}). Not more than one place feature is specified in any onset or coda. More than one (different) manner feature can be specified for onset or coda.
(9) A subset of consonantal features comprises a type of phonological features designated for the language as sonorant features (such as nasal, lateral, rhoticized).
(10) A place feature can be specified concomitantly with a sonorant feature within an onset or a coda. Not more than one place feature can be specified within an onset or a coda including obstruents and sonorants (for example, when both the {stop} and the {nasal} features are concomitantly specified, as in the coda of English 'camp', the two consonantal phonemes must be homoorganic; note that if there were more than one permissible place specification within any syllable component, features would have to accompany ordering specfications).
(11) Lexical accent/tone features are also included in the feature representation of each syllable. In the case of tone systems, tone features may be specified in syllable

components, as opposed to the syllable as a whole.

4. Elemental Gestures and IRFs

4.1. English Onset

Feature specifications accompanying syllable pulses are evaluated by the distributor to yield gesture specifications. Each gesture roughly corresponds to a pair of specifications of place and manner for obstruents and nasals. For the manner specification such as {lateral}, {roticized}, {labiovelarized} and {palatalized} (without place features), the corresponding gesture can be quite complex, and the articulators and the types of gestures, *i.e.* the subset of elemental gestures to be evoked, are identified in the feature table that interpretes phonological feature specifications in the particular context, in particular the pertinent syllable component (onset, coda, p-fix, or s-fix). The manner features {interdental} and {spirantized} involve tongue tip articulations and in this sense their implicit place is apical, and they qualify as the specification of a s-fix. A gesture generally involves several elemental gestures, each of which is implemented in a separate articulatory dimension using a single articulator such as tongue body, tongue blade, tongue tip, lip(s), velum, in a dimension-specific manner. In terms of the signal generation process, each elemental gesture designates a muscle group for which the control function (*i.e.* the IRF adjusted for its amplitude according to the syllable pulse magnitude) at the output of the pertinent actuator specifies the muscle contraction pattern. The manner {fricative}, for example, is implemented not as an approximation of the articulator to the upper structure of the vocal tract, with a "critical" degree of constriction as articulatory phonology specifies [Browman & Goldstein 1992a], but as a qualitatively different phonetic gesture with a different manner of articulation, *i.e.* "action", involving, most probably, a different muscle group (of course with overlapping of the use of the same muscles) from that for a homoorganic stop.[10] In English, /p/ and /f/ are both specified as {labial}, but involve different "places" (*i.e.* labiodental as opposed to bilabial) of articulation. Any details of phonetic implementation like this, including more sutle differences such as tip-blade shapes in dental-alveolar obstruents, are treated quantitatively, in terms of the muscular actions (with local proprioceptic feedback adjustments, for example), as part of the definitions of elemental gestures (IRFs) and their selections (*via* feature table). In the C/D model, an articulatory dimension is a combination of manner (action) and place (articulator), if the place distinction is given, and this combination forms a coherently integral unit. In the sense that an elemental gesture can specify any ad hoc details of the articulatory action, according to the phonetics of the given language, phonetics is language specific. Except for the velum which roughly has one degree of freedom of movement, the movement pattern of an elemental gesture is three-dimensionally complex, using a set of muscle units that are controlled by a prescribed motor command. The velum raising elemental gesture, though usually not mentioned in phonetic descriptions, is always involved in implementation of obstruent features as a positive action, and is important in accounting for the observed behavior of the movement in speech (see *e.g.* [Vaissière 1988]).

To give some idea about interrelations among the elemental gestures as time functions for a syllable, Fig. 3 shows for each feature one time function representing a sample of the pertinent elemental gestures. This figure exemplifies the effects of individual features involved in the word 'splash' by using hypothetical IRFs. As discussed in a recent paper [Sproat & Fujimura 1993], English /l/, specified by a single (manner) feature {lateral}, involves (at least) three identified elemental gestures: tongue tip raising, tongue blade narrowing, and tongue body retraction. These elemental gestures are not synchronous with each other, and the timing of the positional peak of each varies systematically depending on the intrasyllabic position (*i.e.* onset or coda) and on the type of the boundary that immediately follows the syllable. In other words, in terms of the C/D model, the temporal characteristics of elemental gestures for the l-gesture varies considerably, depending on whether it occurs in onset or coda, and how large the computed demisyllabic duration (shadow length including adjacent boundary pulse shadow) is. Therefore, Fig. 3 only gives a rough idea about the temporal characteristics of the phonetic manifestations of different features. Each of these time functions is assumed to be stored in a table for the actuator as the IRF of each elemental gesture. Its amplitude is multiplied by the magnitude of the syllable pulse, or equivalently the "pocs" pulse (see Section 6) that evokes this response. The occurrence of the curve along the time axis in reference to the time value of the excitation pulse is determined by the response function itself.

The time scale is fixed, and there is no horizontal compression or expansion of these elemental response functions as the effect of context, such as what other features are specified in the same onset or coda, or even how fast the syllable rate is locally[11]. This implies that all the tautosyllabic elemental gestures will be proportionally augmented or reduced, whether in the onset, core, or s-fix, as well as which articulator is involved. This is one of the strongest and empirically testable first approximations the C/D theory makes, and should be tested against empirical data from different languages. This may well contribute some insight as to the decision how an acoustic phonetic segmental string without involving vowel segments should be divided into syllables. The phonetic effects of reduction and augmentation on the temporal relation between the onset gestures and coda gestures must be discussed after introducing the onset-coda-affix pulses subordinate to the syllable pulse (see Section 6).

A second strong prediction of this model is that the timing relations among different gestures representing different features within the onset or coda are all fixed and prescribed as inherent properties of the IRFs. Of course, the temporal characteristics are differently specified in the table for onset and coda for the same feature, say {lateral}, depending on whether it is in 'lap' or 'pal'. The peak activity of the tongue body retraction, for example, occurs significantly earlier, relative to the tip-raising action, for initial /l/ than for final /l/.

To be more specific, according to this response function approach as given in the current version of the model, the articulatory movement pattern should be kept the same for a given feature specification regardless of concomitant features. Thus, for example, in 'splash' in Fig. 3, the l-gesture occurs in the same way in the time course whether there is the /p/ or not in phonemic terms. In other words, 'splash' and 'slash' should have the same temporal course of the l-events, such as the tongue tip stretching and raising. It should be mentioned that removing the bilabial closure of /p/ in 'splash' resulting in 'slash' in phonemic terms, reveals a signal reflecting the l-gesture which was hidden behind the stop gap in the acoustic signal. For this part of the syllable to be voiced and have audible effects of [l] (rather than as an unvoiced frication), the vocal folds probably must vibrate earlier for 'slash' than for 'splash'. There are a few mechanisms that could cause this difference in voice onset timing. One is the inherent difference between {spirantized} and {fricative}. Since the former implies a concomitant (and physically succeeding) oral closure whereas the latter does not, the IRF for the former may exhibit the peak activity earlier than the latter, relative to the same syllable pulse. Another consideration is, as a property of the signal generator (*i.e.* the physical articulatory system itself), the aerodynamic interaction between the articulatory closure and the vocal fold vibration is a plausible source of explanation. Voice onset is often triggered by the articulatory release due to the latter's aerodynamic reaction to the glottis, provided that the parametric setting of the vocal fold conditions is appropriately given in the temporal vicinity of the release (the precise voice onset time is dependent on these parameters as well as articulatory and pulmonary conditions). In the 'splash' example, vocal fold vibration is prevented during the complete blockage of the airflow due to the stop. If there is no stop, other things in laryngeal and pulmonary control being equal, the vibration will start considerably earlier.

4.2. English Coda

Another example about final clusters may help illustrate the implications of this theory and potential points of empirical test of the theory. Consider a set of words ‘Ted’, ‘ten’, ‘tet’, ‘tend’, and ‘tent’. From a phonemic point of view, there is a parallel relation between the series ‘ted’, ‘ten’, ‘tend’ and the series ‘tet’, ‘ten’, ‘tent’. The final /d/ in the first series is replaced by /t/ in the second series. The structure is, from the traditional segmental point of view, the same for ‘pend’ and ‘pent’, which are opposed to each other minimally with respect to the binary opposition tense-lax (or voiceless-voiced). In contrast, according to the onset-coda-affix analysis, the two forms are structurally different: ‘pent’ has no s-fix and the final stop feature belongs to the core, whereas ‘pend’ contains /d/ as the s-fix, and the core is identical with ‘ten’.[12] The final /t/ in ‘pent’ is voiceless without a co-ocurring voiceless obstruent (cf. ‘act’), and there is only one place of articulation specified in the coda (cf. ‘camp’ and ‘tank’), indicating its status as a coda element. The phonetic experience is that the nasal murmur is short or virtually nonexistent in American English, if the coda contains a voiceless obstruent, whereas it is long and perceptually distinct if it is either not accompanied by any obstruent or is followed by a voiced obstruent (s-fix). In synthesis, a perfectly natural ‘tend’, ‘tens’, *etc.* would be heard if the word ‘ten’ is clipped out as a segment of the acoustic signal and spliced together as a sequence with the [d] or [z] part of, say, ‘hemmed’ or ‘kings’.[13] But ‘tent’ and ‘tense’ are very different. They are phonetically more like ‘Tet’ and ‘Tess’, respectively, both durationally and articulatorily, except for a heavy nasalization of the vowel [Fujimura 1981]. In the case of ‘tent’ or ‘pint’, depending on the dialect and other factors, the final apical stop may be accompanied or replaced by a glottal stop, with no articulatory release heard, and the acoustic signal may be just like ‘ten’ or ‘pine’ except that the signal is sharply truncated before the nasal murmur ends. Malécot [1960] found similar facts and asserted that English has a phonemic distinction between oral and nasal vowels. There is also a more recent and extensive study by Victor Zue [1975] with similar findings.

The point is that phonetically speaking, within a demisyllable (which contains the vowel gestures), the temporal sequencing of consonantal events is not consistently there, whereas phonologically, the order specifications for onset or coda are redundant. But a syllable affix, whether it reflects a morphological affix or not, seems rather independent from the syllable core, and is phonetically (even acoustically) stable. Its associated signal characteristics are relatively free from contextual influences, just as we expect for the classical concept of phonemes. The affixes are very restricted in kinds, and thus require very few feature specifications in comparison with what is expected for phonemes. The only feature specification for an English s-fix is manner: {stop}, {fricative}, {affricate} or {spirantized}.

A feature {nasal} must be accompanied by a place feature; if there is a concomitant obstruent feature, the two features share the place specification, and the nasal murmur (if any) is produced with the articulator which is always the same as that for the obstruent gesture. If the obstruent feature is {fricative}, as in ‘tense’, {nasal} in English evokes an additional articulatory dimension for oral closure with an inherently small magnitude, resulting in a short acoustic segment of nasal murmur next to the nasalized vowel or its gliding (as in the case of ‘ounce’). If the concomitant obstruent feature is {stop}, the elemental gesture for oral closure starts somewhat before the velum is raised to close up the nasal channel, and the voicing stops naturally (in part) due to the cessation of the airflow that causes the Bernoulli effect for maintaining vocal fold vibration. In the case of the voiced stop /d/ (note that there is no /b/ or /g/ in this environment) following a nasal (which can take any of the three possible place specifications), the additional stop is a s-fix, which has a distinctly independent closure gesture, and its (at least partial) voicing is assured by a positive laryngeal adjustment to facilitate the maintenance of the vocal fold vibration, even when a (voiceless) heterosyllabic consonant (or boundary) follows. Interestingly, with this “marked” voicing, the s-fix is never followed by another s-fix.

The so-called t-epenthesis producing *e.g.* ‘dance’ [dænts] with a short t-closure and a weak release[3] which is often treated by a rewrite rule creating a new t-segment in the phonological literature, is an instance where the velum raising, one of the elemental gestures for obstruent features, accompanied by the vocal fold abduction as another elemental gesture unless voicing is specified for the coda, takes place earlier than the release of the complete tongue tip closure (assigned for {nasal} at the lack of the concomitant {stop}, leaving a short segment of oral closure with the (prematurely) raised velum. This is purely a matter of relative timing among elemental gestures, and, therefore, a continuum of unspecified stop closure duration is predicted, in contrast to what the phonological rewrite rule would suggest as an allophonic binary opposition. Note that, due to the nonlinearity of the signal generating process, the duration of the oral closure created in this manner (as in any other case) depends critically on how large the intrinsic amplitude of the IRF of this apical closure elemental gesture is, as prescribed in the response function table for {nasal} in the particular language, and also on the syllable pulse magnitude reflecting the prosodic and utterance conditions of the syllable. The velum lowering almost completely covers the acoustic vowel segment, *i.e.* the left- over period of the core that is not affected by the obstruent consonantal feature. In fact, the peak activity of velum lowering for final nasal elements seems to occur always in the middle of the acoustic vowel portion of the syllable, rather than during the nasal murmur which by definition corresponds to the oral closure [Fujimura 1981; Krakow 1989].

Another case is opposite to the stop-epenthesis, *i.e.* a stop is reduced to be observed as a fricative in casual speech. This phonetic process should be distinguished from cases involving a phonologized change from stop to fricative. Phonologization provides a lexical representation with {fricative}. This may or may not be the case for the intervocalic /b/ in some dialects of Spanish, which seems never produced as a stop even when the syllable magnitude is very large. Note that in some dialects, as in Peruvian Spanish, this allophonic variation involves an alteration between bilabial and labiodental fricatives. Reducing a closure gesture for an English apical stop in casual speech, for example, can produce frication in place of complete closure, as the amplitude of the articulatory excursion is reduced for the same IRF. This reduced gesture, however, is presumably more difficult to produce consistently in different environment, in comparison with the true fricative, which is produced by a stable three-dimensional articulatory gesture (see Fujimura & Kakita [1979] for a related discussion of the vowel [i]). While the articulatory trajectory of a midsagittal flesh point, as revealed by cinefluorography or microbeam, as well as the gross acoustic characteristics of the turbulent noise, may be similar to the production of an intended fricative (*i.e.* using the proper IRF for {fricative}), the identification of the reduced stop as a fricative apart from the apparent acoustic effect may well be only an artifact of the midsagittal observation. The continuous phonetic gradation of the phenomena, as observed by electropalatography [Hardcastle 1976] ranging from a complete closure of varying articulatory force to a weak frication or semivowel, can not be described effectively by allophonic rewrite rules in any case (see Sproat and Fujimura [1993] for a related discussion concerning English /l/; also [Browman and Goldstein 1992a,b; Kohler 1990, 1991, 1992]).

Another example of the nontrivial relation between features and gestures in relation to reduction or lenition is the nasalization of the voiced palatovelar stop in word-medial intervocalic position in Japanese. In Tokyo dialect, it is said that the /g/ in forms like /ari'gatoH/ (thank you) is typically pronounced (by speakers of older generations, particularly trained broadcasting announcers) with a velar nasal. While this appears to be an ad hoc addition of the feature {nasal}, it does have some phonetic explanation: when the dorsum is raised using the palatoglossus muscle which is involved in raising the tongue body also, there is a tendency for the velum to be pulled down, unless there is a strong enough counterbalance due to the palatal levator which an obstruent feature implies. Therefore, probably this nasalization has a phonetic motivation. Slight nasalization is often induced phonetically without {nasal} being specified in most languages, particularly for low vowels.

5. Organization of Articulatory Gestures

Phenomena discussed above are inherently multidimensional, and can be described effectively only by referring to articulatory conditions, particularly when various types of contextual and prosodic influences are quantitatively considered. Similar discussions are given by Browman and Goldstein [1992a,b] and other investigators (*e.g.* [DeJong, Beckman & Edwards in press; Beckman, Edwards & Fletcher 1992]) using the descriptive framework of articulatory phonology. Some insightful discussion, though qualitative, have been offered earlier by Borowsky [1986], and acoustic details have been discussed extensively for German, French and English by Kohler [1990], among others.

Most of the arguments for the gesture score approach apply equally well to the C/D model in accounting for many types of phonetic variation. In articulatory phonology, phenomena called articulatory hiding and blending are explained in terms of the temporal shifting of one gesture relative to another when the two hold no rigid timing (phasing in their scheme) relation. In the C/D model, there are two distinct cases: (1) heterosyllabically, the temporal overlapping of gestures varies depending on the interval between the two syllable pulses, which is affected directly by utterance conditions as well as prosodic conditions, (2) tautosyllabically, overlapping does not change articulatorily, but acoustic effects can significantly change due to the complexity of the factors that determine the onset and offset of vocal fold vibration as mentioned above, articulatory saturation due to the contact of the articulator to its counterpart, *etc.* For example, increasing the stop gesture magnitude may appear as an increase in the articulatory retention exhibiting a longer plateau of the time function of vertical position of the tongue tip, lower lip, *etc.*, rather than increasing the peak height, temporally shifting the ascending and/or descending movement, or, depending on the characteristics of the IRFs and the details of the three dimensional structure, considerable changes in the ramp shape and peak velocity. As a result, a uniform amplitude reduction of elemental gestures within a syllable can be responsible for an apparent deletion of a segment, feature changes, *etc.*, but this may not be the case if we examine the underlying articulatory gestures. Also, as seen in Section 6, the revised C/D model accounts for changes in terms of temporal relations between onset and coda gestures due to syllable pulse magnitude changes.

This does not mean that all allophonic variation can be described and explained equally well by articulatory phonology or the C/D model. For example, as seen in Sproat and Fujimura [1993], the distinction between the so-called light and dark /l/ must assume inherently different physical properties of this consonantal gesture depending on whether it is syllable-initial or syllable-final among other factors. (The distinction between syllable and word in this context is often difficult and only limited evidence is available. Fujimura [1981] gives some evidence in the case of Japanese nasals.) The C/D model critically distinguishes onset phenomena from coda, but this distinction is not built into articulatory phonology, nor other theories including the classical coarticulation theory [Lindblom 1990].[14]

Also, as Kohler [1990, 1991, 1992] discusses in detail, there may be phenomena which he calls "reorganization", which may escape any explanation by articulatory modeling, and, in terms of the C/D model, obviously phonologized changes that must be incorporated within the lexicon set aside, have to evoke some symbolic changes of the converter input, depending on the style of utterance. For example, in Japanese, /sayonara/ is often pronounced as /saJnara/ producing a diphthong /saJ/ from /sayo/ in casual speech, and this is transcribed in kana as sa.i.na.ra in novels, *etc.* The C/D model is based on a simple principle governing the temporal and magnitude patterning of the string of syllables when larger phonological phrasal units are formed by syntactic processes.

6. Onset-Coda-Affix Pulses -- an Elaboration of the C/D Model

Finally, a recent elaboration of the C/D model (unpublished) will be outlined. This is particularly relevant to what is called reduction/lenition phenomena, but it also affects significantly the power of the theory in handling syllable types such as heavy vs. light syllables. It is also related to the issue of how suprasyllabic units such as foot can be treated within the C/D framework.

The original version of the C/D model stipulated that the entire temporal structure of an utterance is completely determined, at one level of the model description, *via* syllable pulse train computation, which is based on the input specification including the numerically augmented metrical tree. Articulatory events for each syllable are implemented by evoking IRFs from a stored vocabulary of elemental time functions, using each syllable pulse as the excitation. The magnitude, as well as timing, of each gestural event is modulated according to the magnitude (and derivatively time) values of the syllable pulse. Since the IRFs have a common reference to the same syllable pulse, the temporal relations among elemental gestures within the same syllable (or the same onset, coda, or affix, see infra) must be fixed, while the magnitudes are scaled uniformly according to the magnitude of the excitation pulse. There is a second-order approximation anticipated by this model, that the parameters of the response functions, which are the stored table specifications given a fixed family of functions for elemental gestures, can be made sensitive to some aspects of the C/D model input representation. This effect is specified explicitly below, not by manipulating IRF parameters like peak activity timing, but by setting up subsidiary puleses governed by each syllable pulse, whose time values are sensitive to the result of the shadow computation. It should be noted that this revision does not affect the position that all information concerning prosodic organization of speech utterances is contained in the syllable/boundary pulse train. Further possibilities of adding a second approximation beyond the current C/D model that utterance conditions such as speed and style of utterance can modify certain parameters describing IRFs systematically according to a certain general principle.

It should be mentioned, however, that the first approximation, even without the revision to be outlined below, accounts for a wide range of phonetic phenomena, and this explanatory power may not be obvious in this model using an abstract phonetic representation. According to the C/D model, the vocalic movement patterns are separately implemented as movement from the preceding syllable nucleus to the next, possibly with long distance interaction (*e.g.* a stressed nucleus to the next) which has not been discussed so far. For example, in the case of vowel harmony, a vocalic feature would be specified for a word rather than for each syllable, and its implementation may be reflected throughout the vowel sequence in the specified domain (see Boyce, Krakow & Bell-Berti [1991]) for an articulatory study of Turkish). This should affect all "intervening" consonantal gestures when they appear as articulatory and acoustic signals through the signal generator, since the base function in each articulatory dimension will be different depending on the treatment of the vocalic feature specifications in relation to the syntagmatic effects such as each foot has an inherent effect on the base function at its edges (*via* boundary pulses that are associated with their intrinsic IRFs).

The phonetic consequences are not necessarily obviously predictable, however, because of the interaction of the physical system, the movement of which is governed not only by the paradigmatic (inherent) features but also by syntagmatic factors (configurational features, in Jakobson, Fant & Halle [1951/63]).) For example, mandible movement may reflect primarily stress status of the syllable (see *e.g.* Westbury & Fujimura [1989]; Erickson & Fujimura [1992], *i.e.* syllable pulse patterns rather directly, but it also may reflect syllable margins as such in general as well as effects of interacting articulators and the larynx. The base function controlling the tongue body position changes very slowly, like intonation contours, but they are also perturbed by local vocalic gestures due to consonantal feature specifications, such as palatalization or lip-rounding for diphthongal glides, or vocalic gestures involved in sonorant consonants (see Sproat and Fujimura [1993] for the distinction between vocalic features and vocalic gestures in connection with English /l/) and possibly even palatovelar obstruents.

There is also an interaction between the vocalic base function and consonantal gestures in their effects on articulators' movement patterns. For example, the mandible contour will reflect lip closure gestures *via* physical constraints of the anatomical devices for achieving an articulatory goal of lip closure, which is implemented as the design of the gesture table relating features to elemental gestures of different inherent magnitudes, in such a way that the lip closure is attained under a wide range of prosodic and utterance conditions by involving jaw opening. Presumably, design parameters are neurally optimized during the process of language learning, as well as the development of a language itself. The gesture implementation scheme of the distributor employs a feature table which relates features (Greek letters and Roman small capitals in Fig. 1) to corresponding gestures (Italic Greek), each of which comprise elemental gestures in independent dimensions. Each actuator employs an impulse response table, which relates elemental gestures to their articulatory events in the form of IRFs (specified as parameter values of a family of functions). These tables are designed to reflect the "wisdom" of individual languages considering the communicative functions, just as phonological forms for lexical items are chosen carefully reflecting phonologization of phonetic variants according to their perceptual effects [Ohala 1981], among other factors. The principle of quantal nature [Stevens 1989] also may govern or explain such table designs, and the sonority principle [Hooper 1973/76; Clements 1989] or vowel affinity principle [Fujimura 1975] will represent one of the principles of parameter setting for IRFs. The patterning of speech sounds in different languages [Maddieson 1984] reflects the general design principle that prescribes the inherent dynamic properties of articulatory events.

When there is no vowel specification for a reduced syllable, the base function simply shows a transitional state, except for the possible boundary effects. The observed "vowel segment" in the acoustic signal, which is inherently voiced, would be delimited by effects of consonantal gestures (and may be totally devoiced, or the voice quality may change, due to the "invading" effects of vocal fold abduction and other gestures evoked for the neighboring voiceless consonants and their physical consequences such as blocking of air flow). If the syllable is pronounced in isolation as a monosyllabic word, 'that', for example, the temporal pattern of all the gestures involved will be the same, except for the shared amplitude scaling, regardless of the strength of the pronunciation. When the pronunciation is weak, whether due to stress pattern or intraphrasal position (*e.g.* phrase final weakening due to articulatory declination, see, Fujimura [1990a,b], Browman and Goldstein [1990b], Vayra & Fowler [1992]), the deviation of the vocalic gesture from neutral position will be governed by the adjacent full vowels, producing a colored schwa, and the consonantal gestures will be small in magnitude. Such properties of the C/D model reamain the same in the revised version.

Note that when a consonantal perturbation gesture is reduced in magnitude, the apparent duration of that consonant in the acoustic waveform will also be reduced. In addition to the saturation effect mentioned above, this can be because of the inherent nonlinearity of the source signal generation process. There is a certain threshold value for the constriction aperture for a frication (*i.e.* turbulent noise) to be generated in the vocal tract; if the constriction is not narrow enough, turbulent noise will not be generated at all. This means that, at the left and right margins of a single IRF for a fricative consonant, for example, the skirts will be cut off acoustically, even though in the underlying control function, there are remaining reduced amplitudes representing the same time function as in a stronger utterance. When there is a preceding or following syllable pronounced, then reduction of a syllable results in shorter intervals between consecutive syllable pulses, according to the timing value computation algorithm of the syllable pulse train, and the margins of the syllable in acoustic signals will be obscured by contiguous heterosyllabic gestures even though the consonantal signal of the reduced syllable in question is still there. The result is a significantly shortened syllable duration in comparison with those for more prominent pronunciations of the same syllable. The 'hiding' effect will be stronger if the adjacent syllables have higher magnitudes.

However, the internal timing relation between the initial and final gesture peaks will remain the same regardless of the magnitude of the syllable, according to the original version of the C/D model. In fact, 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 is of course counterfactual, while the shortening of consonantal segments is found true.

A reasonable solution of this problem is 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 value evaluation of the syllable/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 (for p-fix, onset, coda, s-fix) pulses) evokes the IRFs. The pocs pulses inherit the magnitude of the parent syllable pulse. The time value of the leftmost p-fix pulse, however, is computed from the time value of the left edge of the syllable, as evaluated by the shadow computation of the syllable/boundary pulse train, as discussed in the first version of the C/D model [Fujimura 1992]. Each p-fix receives assigned slopes for the shadows to the left and to the right of its pulse, and the edge of the right shadow is made coincident to the evaluated left edge of the syllable as determined by the syllable shadow. In essense, a p-fix adds its own duration to the left of the syllable core duration, which is dictated by the syllable pulse. The preceding syllable or boundary ends its shadow at the point in time where the p-fix (or core, if no p-fix) left shadow starts. This computation scheme is similar to the shadow computation of syllable pulses, which depicts a linear computation equivalent to the elastic model of articulatory timing [Fujimura 1987].[15] Likewise, the time value of the lefttmost s-fix pulse is computed forward from the right edge of the righthand syllable core shadow. The other pocs pulses for more external affixes are recursively computed by a similar algorithm recursively inside out. The p-fix next to the core determines the time value of the excitation pulse for the onset IRFs as well as the pertinent p-fix IRFs. The s-fix next ot the core determines the time value of the excitation pulse for the coda IRFs as well as the pertinent s-fix IRFs. When there is no affix, the shadow computation as previously described applies based on the syllable pulse (which is now interpreted at the same time as the core pulse).

This additional scheme of timing computation makes the temporal relation between initial and final consonantal gestures sensitive to the syllable pulse magnitude. Also, since the slopes of shadows for syllables are sensitive to the internal structure, reflecting the syllable type such as long-short or heavy-light, 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 a short vowel, with or without diphthongal glide, the number of affixes, *etc.* The function of the syllable within its foot, in particular whether it is the head or not, may be another factor that affect the asigned syllable duration, either *via* shadow coefficient manipulation, or entirely by a new version of the hierarchical spring model replacing the shadow computation scheme.

There are many additional details of the theory that have to be worked out. The comparison of prediction and 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 is a critical component of this phonetic implementation process, which reveals the effects of the highly nonlinear nature of the complex physical system for producing speech signals, as discussed above. This component brings the generative description closer to modeling of the speech production process, but some viewpoints, such as motor theory of speech perception [Liberman 1991], may suggest the significance of such a production-oriented theory also as a theory of speech perception.

To summarize some critical points that the C/D model predicts about phonetic organization of speech events, we may note:

(1) Each phonological feature or a combination of specific types of phonological features evokes a phonetic gesture, which comprises one or more than one elemental gesture. For example, {lateral^o} and {lateral^c} evoke different allophonic variants of /l/ in English, both involving tongue body retraction and blade narrowing with accompanying stretching of the blade as characteristic elemental gestures; a related contact of the tongue tip with the alveolar region of the roof is another elemental gesture in the onset but not necessarily in coda (see Sproat & Fujimura [1993]).

(2) An elemental gesture may require a combination of feature specifications. For example, a bilabil closure-release gesture is an elemental gesture evoked by a combination of {stop} and {labial} in English.

(3) When a phonetic gesture consists of more than one elemental gesture, elemental gestures have different temporal courses as specified as their IRFs (IRFs). The time functions may not be synchronous in terms of the peak articulatory activities, and also may considerably differ in the time span of their effective manifestations. These IRFs are inherent to individual elelmental gestures, separately specified for onset and coda gestures even if the name of the elemental gesture is the same.

(4) All the alterations of the time course, according to the C/D model, is implemented *via* manipulation of the syllable/boundary pulse train.[16] Thus, for example, a change of utterance speed is implemented parametrically by manipulating (perhaps simply by a multiplication factor) the shadow coefficients in time interval computation between contiguous syllable/boundary pulses, and the phonetic gestures in the abstract form (up to the signal generator) themselves are not altered (IRFs remain the same but their relative timing and therefore the patterns of overlapping change).[17]

(5) Boundaries, in the sense of the C/D model, are specified with features for their types. Intonation patterns are consequences of specifying, phonologically, both lexical features and boundary features (of the metrical tree) as paradigmatic information at the level of phonetic representation, along with the syllable-boundary pulse train (derived from the augmented metrical tree) as syntagmatic information of the utterance material.

(6) Boundary features are also associated with articulatory/phonatory elemental gestures. Boundary tones, for example, are implemented as a set of elemental gestures in the form of IRFs that determine the time functions of control signals governing temporal characteristics of parametric control of the laryngeal (and possibly pulmonary) conditions. An elemental gesture for pitch accent (phonological tones, for both lexical and phrase accents and boundary tones in terms of current intonation theory [Bruce 1979, Pierrehumbert 1980, Poser 1984, Pierrehumbert & Beckman 1988] (also see Ladd [1990], Kubozono [1993]) is likely to affect F_0 as well as voice quality and intensity modulation [Löfqvist *et al.*, in press; Fujimura *at al.* in press], whereas an increase in the syllable magnitude amplifies all tautosyllabic phonetic gestures including (pulmonary as well as laryngeal) elemental gestures related to voice intensity and quality changes, accompanying natural F_0 change apart from the separate tone control (in the same or different muscles).

7. Acknowledgment

It is gratefully acknowledged that this work has been supported in part by ATR Human Information Processing Research Laboratories and ATR International Telecommunications Research Laboratories.

Notes

1. It should be mentioned that in Jakobson-Fant-Halle's original work, the term "prosodic" was used in a specifically narrow sense, which is basically different from the usage here, as well as conventional usage in current phonological and phonetic literature. In this paper, in contrast to most contemporary usage, on the other hand, prosodic characteristics refer to nonlexical syntagmatic characteristics of speech signals in general, thus excluding paradigmatic lexical specifications such as lexical tone/accent, while including supralaryngeal articulation as well as voice source and temporal modulations of speech due to phrasing characteristics and phrasal intonation control. Thus, unfortunately, under the influence of the current usage, the same term "prosodic", which was used in JFH only for paradigmatic but noninherent distinctive features, is used in the sense more like their "configurational features".

2. The utterance conditions as specified at the upper corners of the input representation (Fig. 1) also affect certain parameters of the signal generator, including the anatomical dimensions of the articulatory system, shadow coefficients, and certain parameters involved in the IRFs. Presumably, the extrinsic prominence specifications augmenting tree nodes affect the utterance configuration only through syllable magnitude (and derivatively time) manipulations. Note that the syllable magnitude and the associated timing patterns, in part reflecting the prominence augmentation, affect intensity and quality of voicing as well as articulatory force, whereas the paradigmatic (both lexical and phrasal, such as statement/question) tone/accent control pertains feature specifications which evoke laryngeal gestures independently from the syllable magnitude manipulation.

3. In a "phonemoidal" transcription in this paper, surrounded by slashes, a period separates s-fixes from the core.

4. The English word 'warmth', therefore, must be all in the core. The C/D analysis for the coda of this syllable is {labial, rhoticized, nasal, interdental}. The feature {interdental} is a obstruent manner feature. The lack of voicing for the last part of the coda in the phonetic manifestation is unmarked for obstruents (thus, the coda specification for 'smooth' is {interdental, voiced}. The specification of place {labial} assigns the feature {nasal} to be implemented by the lips, all other coda features being manner specifications (if there is a specification of {stop} in the same coda, for example, the nasal-stop cluster must be homorganic, the articulator being commonly specified by the place feature).

5. Note that {spirantization} is implemented with an IRF which shows a temporally earlier (with respect to the demisyllabic impulse) peak activity than that of {stop}, regardless of whether it appears in onset or coda.

6. In other words, the inside weak stop-release elemental gesture, which we would treat as a (strong) manner feature without a place specification functionally, has its IRF with an intrinsically lower amplitude and most probably slower movements than the outside stop-release gesture, but the amplitude happens to be larger than usual glides in other languages. Perhaps more importantly, the use of the articulatory musculature is different so that, unlike other cases of glides, there is no mechanism that prevents the collapsing vocal tract when the force of articulation is strong due to the prosodic environment (i.e. high syllable magnitude). Our prediction, then, would be that, this type of weaker stops will not be articulated as a stop, when the syllable is in a weak position, and this reduction would be observed more often than that in, for example, in English casual speech.

7. The constraints given here may have to be relaxed as we study various languages. In that case, most likely, a description would be possible in terms of a hierarchy of preference of individual constraints with some interaction among them, as proposed in the optimality theory [McCarthy & Prince, to appear; Prince & Smolensky, to appear]. On the other hand, presumably, these constraints, as presented here speculatively as absolute constraints, could be described in terms of a generative rule system.

8. In English, the phonetic labiodental fricatives are specified as {labial, fricative}.

9. Note that the C/D model's segnal generator is a three-dimensional dynamic system, and is free from the most common limitation of other models stemming from the 2-dimensional consideration in the midsagittal lateral picture of the articulatory state.

10. Unless there are errors in the motor program or command execution, which often occur in casual speech, and apart from the speaker/style dependent parameter adjustments of the response functions. Also, temporal shifting (without compression/expansion or other deformation, of the IRF, due to the context within the same articulator is a different matter, as discussed later.

11. According to the demisyllabic feature analysis [Fujimura 1976, 1979] restated according to the underspecification scheme we adopt here, English s-fixes must be (1) implemented by the tongue tip, and (2) the concomitant voiceing state must follow that of the core-final voicing specification (if {voiced} in code, the s-fix automatically is specified {voiced} by implication).

12. Phrase final effects must be taken care of appropriately by giving appropriate contexts, as usual.

13. Similar examples may be found more abundantly in German (*e.g.* 'Menschen' pronounced with an epenthetic [t]).

14. The obviously separate treatment of initial consonants from final consonants even in an adjacent position in speech errors (see *e.g.* [Fromkin 1973]) is explained in the current psycholinguistic literature by a concept of "frame" in performance rather than the lexical representation itself, see *e.g.* [Shattuck-Hufnagel 1985]. Kubozono [1989] discussed larger segmental units in explaining speech error patterns.

15. Note that this scheme makes it possible to adopt the hierarchical spring model [Fujimura 1987], considering the effects of foot structure in computing syllable durations. Instead of using the shadow algorithm, we can compute the syllable durations as the sum of the syllable core duration and affix durations, according to the hierarchical spring model. Note also that at the time the spring model was discussed, the model had to be assumed for individual articulatory dimensions separately. This clumsy situation is now removed, since the signle linear pulse train governs all dimentions properly according to the C/D model.

16. Apart from the time scale alteration (time warping) toward the end of a phrase as a possible elaboration of the current C/D model. This may be implemented, if it proves empirically correct, as a local and non-uniform expansion of the physical time scale over a certain period of time toward the end of the syllable materials, to absorb part of duration assigned to the succeeding boundary (the remaining part may be implemented as a silent pause, which may be similarly absorbed into the succeeding syllable material).

17. The only exception so far noted to this is the articulatory clash, which seems to affect a parameter of the relevant IRFs themselves due to the context (such as shifting an IRF in time when there is a certain condition of clash with the following elemental gesture is met within the same articulatory organ). This phenomenon has been observed and discussed [Fujimura 1986, 1990], but more data will be needed to accounmt for this correctly as a second-order amendment of the current C/D model. It is also possible that this property is incorporated into the signal generator as a physiological characteristic of the articulatory system.

References

Beckman, M.E., Edwards, J. & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G.J. Docherty and R.T. Ladd (eds.), *Papers in laboratory Phonology II. Gesture, Segment, Prosody,* (pp. 68-86) Cambridge, U.K.: Cambridge University Press.

Borowsky, T. (1986). *Topics in the Lexical Phonology of English.* PhD diss. U. Mass., Amherst.

Boyce, S.E., Krakow, R.A., & Bell-Berti, F. (1991). Phonological underspecification and speech motor organization. *Phonology,* **8**, 219-236.

Browman, C.P. & Goldstein, L.M. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (ed.) *Phonetic Linguistics* (pp. 35-53). New York: Academic.

Browman & Goldstein [1988]. Some notes on syllable structure in articulatory phonology. *Phonetica*, **45**, 140-155.

Browman, C.P. & Goldstein, L.M. (1990a).Gestural specification using dynamically-defined articulatory structures. *J. Phonetics*, **18**, 299-320.

Browman, C.P. & Goldstein, L.M. (1990b). Tiers in articulatory phonology, with some implications for casual speech. In In J. Kingston and M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 341-376). Cambridge, UK: Cambridge University Press.

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

Browman, C.P. & Goldstein, L.M. (1992b). Response to commentaries. *Phonetica* , **49**, 222-234.

Bruce, G. (1983). Word prosody and sentence prosody in Swedish. *Proc. 9th ICPhS, Copenhagen, Vol. 2,* 388-94.

Carré, R. & Chennoukh, S. (submitted). Vowel-consonant-vowel modeling by superposition of consonant closure on vowel-to-vowel gestures.

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.

DeJong, K., Beckman, M.E., & Edwards, J. (In press). The interplay between prosodic structure and coarticulation.]

Eek, A. & Help, T. (1987). The interrelationship between phonological and phonetic sound changes: A great rhythm shift of Old Estonian. *Proceedings of the XIth International Congress of Phonetic Sciences, Tallinn, Estonia,* Vol. **6**, pp. 218-233.

Erickson, D., & Fujimura, O. (1992). Acoustic and articulatory correlates of contrastive emphasis in repeated corrections. *Proceedings of The International Conference on Spoken Language Processing,* Banff, Canada.

Fromkin, V. (1973). *Speech errors as linguistic evidence.* The Hague: Mouton.

Fujimura (1967) Speech of Japanese -- from a phonological description of the lin guistic form to the sound wave (in Japanese) In *Collected Papers in Commumoration of the 20th Anniversary of theNHK Publication Office. Tokyo: NHK Publishing Office, pp. 363-404.*

Fujimura (1972). *Onseikagaku* (in Japanese) Tokyo: University of Tokyo Press.

Fujimura, O. (1975). Syllable as a unit of speech recognition. *IEEE ASSP* **23 (1)**, 82-87.

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. *Zeitscrhift für Phonetik, Sprachwissenschaft und Kommunikationsforschung,* **32**, 471-476.

Fujimura, O. (1981). Elementary gestures and temporal organization--What does an articulatory constraint mean? In T. Meyers and J. Anderson (eds) The Cognitive Representation of Speech (pp. 101-110).Holland: North Holland Publishing Co.

Fujimura, O. (1986). Relative invariance of articulatory movements. An iceberg model. In

Perkell, J.S.& Klatt, D.H.(eds.), *Invariance and Variability in Speech Processes.* (pp. 226-242). Hilldsdale: Lawrence Erlbaum.

Fujimura, O. (1987). A linear model of speech timing. In R. Channon and L. Shockey (eds.), *For Ilse Lehiste* (pp. 109-123). Dordrecht, Holland: Foris.

Fujimura, O. (1989). Demisyllables as sets of features: Comments on Clement's 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. (1990b). Articulatory perspectives of speech organization. In W.J. Hardcastle and A. Marchal (eds.) *Speech Production and Speech Modeling* (pp. 323-342). Dordrecht: Kluwer Academic Publishers.

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

Fujimura, O. (in press). C/D model: a computational model of phonetic implementation, in E. Ristad (ed.), *DIMACS Workshop on Human Language.* Am. Math. Soc.

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

Fujimura, O., Kiritani. S. & Ishida, H. (1973). Computer-controlled radiography for observation of movements of articulatory and other human organs. *Computer Biol. Med.* **3**, 371-384.

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, North Holland.

Fujimura, O. & Kakita, Y. (1979). Remarks on quantitative description of the lingual articulation, in S. Öhman & B. Lindblom (eds.), Frontiers of Speech Communication Research. London: Academic Press.

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.

Fujimura, O., Cimino, A. & Sawada, M. (in press). Voice quality control within a sentence: expressive effects of source spectral envelope change. In O. Fujimura & M. Hirano (eds.), *Voice Quality Control: Proc. VIIIth Vocal Fold Physiology Conference, Kurume, April 1994.* San Diego: Singlular Publ.

Hardcastle, W.J. (1976). *Physiology of Speech Production: An introduction for speech scientists.* London: Academic Press.

Hooper, J.B. (1973/76). *Aspects of Natural Generative Phonology.* PhD diss. U. Southern California. *An Introduction to Natural Generative Phonology.* New York: Academic Press. Stevens, K.N. (1989). On the quantal nature of speech. *J. Phonetics,* **17**, 3-45.

Idsardi(1993) PhD dissertation, MIT.

Jakobson, R., Fant, G. & Halle, M. (1951/63). Preliminaries to speech analysis. *Technical Report, Acoustics Laboratory, MIT; Cambridge,* MA: MIT Press.

Kelso, J.A.S., Saltzman, E.L., & Tuller, B. (1986). The dynamical perspective on speech production. *J. Phonetics,* **14**, 29-59.

Kiritani, S., Ito, K. & Fujimura, O. (1975). Tongue-pellet tracking by a computer-controlled x-ray microbeam system. *J. Acoust. Soc. Am.* **57**, 1516-1520.

Kohler, K. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations.. In W.J. Hardcastle and A. Marchal (eds.) *Speech Production and Speech Modeling* (pp. 69-92). Dordrecht: Kluwer Academic Publishers.

Kohler, K.J. (1991).The phonetics/phonology issue in the study of articulatory reduction. *Phonetica*, **48**, 180-192.

Kohler, K. (1992). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica*, **49**, 180-192.

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

Kröger, B. J. (1993). A gestural production model and its application to reduction in German. *Phonetica* **50**, 213-33 .

Kubozono, H. (1989). The mora and syllable structures in Japanese: Evidence from speech errors. *Language and Speech,* **32** (3), 249-278.

Kubozono H. (1993). *The Organization of Japanese Prosody.* Tokyo: Kuroshio Publ.

Ladd, D. R. (1990). Metrical representation of pitch register, in J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: between the Grammar and the Physics of Speech.* Cambridge, Cambridge Univ. Press.

Lehiste, I. (1970). Suprasegmentals. Cambridge, MA: MIT Press.

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

Liberman, A.M. (1991). Afterthoughts on modularity and the motor theory. In I.G. Mattingly and Michael Studdert-Kennedy (eds.) *Modularity and the Motor Theory of Speech Perception.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 35, 1773-81.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H Theory. In W.J. Hardcastle and A. Marchal (eds.) *Speech Production and Speech Modeling* (pp.403-439). Dordrecht: Kluwer Academic Publishers.

Löfqvist, A., McGowan, R. & Koenig, L. L. (in press). Voice source variations in running speech: a study of Mandarin Chinese tones. In O. Fujimura & M. Hirano (eds.), *Voice Quality Control: Proc. VIIIth Vocal Fold Physiology Conference, Kurume, April 1994.* San Diego: Singlular Publ.

Maddieson, I. (1992). The structure of segment sequences. *Proceedings of The International Conference on Spoken Language Processing,* Banff, Canada.

Malécot, A. (1960). Vowel nasality as a distinctive feature in American English. *Language,* **36**, 222-229.

McCarthy, J. J. & Prince, A. (to appear). *Prosodic Morphology I: Constraints Interaction and Satisfaction.*

Nadler, R. D., Abbs, J. H. & Fujimura, O. (1987). Speech movement research using the new x-ray microbeam system. *Proc. XIth Int'l. Cong. Phonetic Scs., Tallin,* Vol.1 pp.221-224.

Ohala, J.J. (1981). The listener as a source of sound change. In C.S. Maesk, R.A. Hendrick, and M.F. Miller (eds.), *Papers from the Parasession on Language and Behavior* (pp. 178-203). Chicago, IL: Chicago Linguistic Society.

Öhman, S.E.G. (1967). Numercal model of coarticulation. *J. Acoust. Soc. Am.*, **41**, 310-320.

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

Pierrehumbert, J. B. & Beckman, M. E. (1988). *Japanese Tone Structure.* Cambridge, MA: MIT Press.

Poser, W. J. (1984) *The Phonetics and Phonology of Tone and Intonation in Japanese.* PhD Dissertation, MIT.

Prince, A. & Smolensky, P. (to appear). *Optimality Theory.*

Saltzman, E. (1985). Task dynamic coordination of the speech articulators: A preliminary model. *Haskins Laboratories Status Report on Speech Research,* **84**, 1-18.

Shattuck-Hufnagel, S. (1985). Context similarity constraints on segmental speech errors: An experimental investigation of the role of word position and lexical stress. In J. Lauter (ed.), *On the Planning and Production of Speech in Normal and Hearing-Impaired Individuals* (pp. 43-49). *ASHA Reports* **15** (ISSN 0569-8553).

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

Stevens, K.N. (1989). On the quantal nature of speech. *J. Phonetics,* **17**, 3-45.

Vayra, M. & Fowler, C.`A. (1992). Declination of supralaryngeal gestures in spoken Italian. *Phonetica*, **49**, 48-60.

Vassière, J. (1988). Prediction of velum movement from phonological specifications. *Phonetica*, **45**, 122-139.

Westbury, J. & Fujimura, O. (1989). An articulatory characterization of contrastive emphasis in correcting answers. *J. Acoust. Soc. Am.,* **85**, Suppl. 1, S98.

Zue, V. W. (1975). Acoustic phonetic data base for the study of selected English consonants, consonant clusters, and vowels. *J. Acoust. Soc. Am.*, **57**, Suppl. 1, S34.

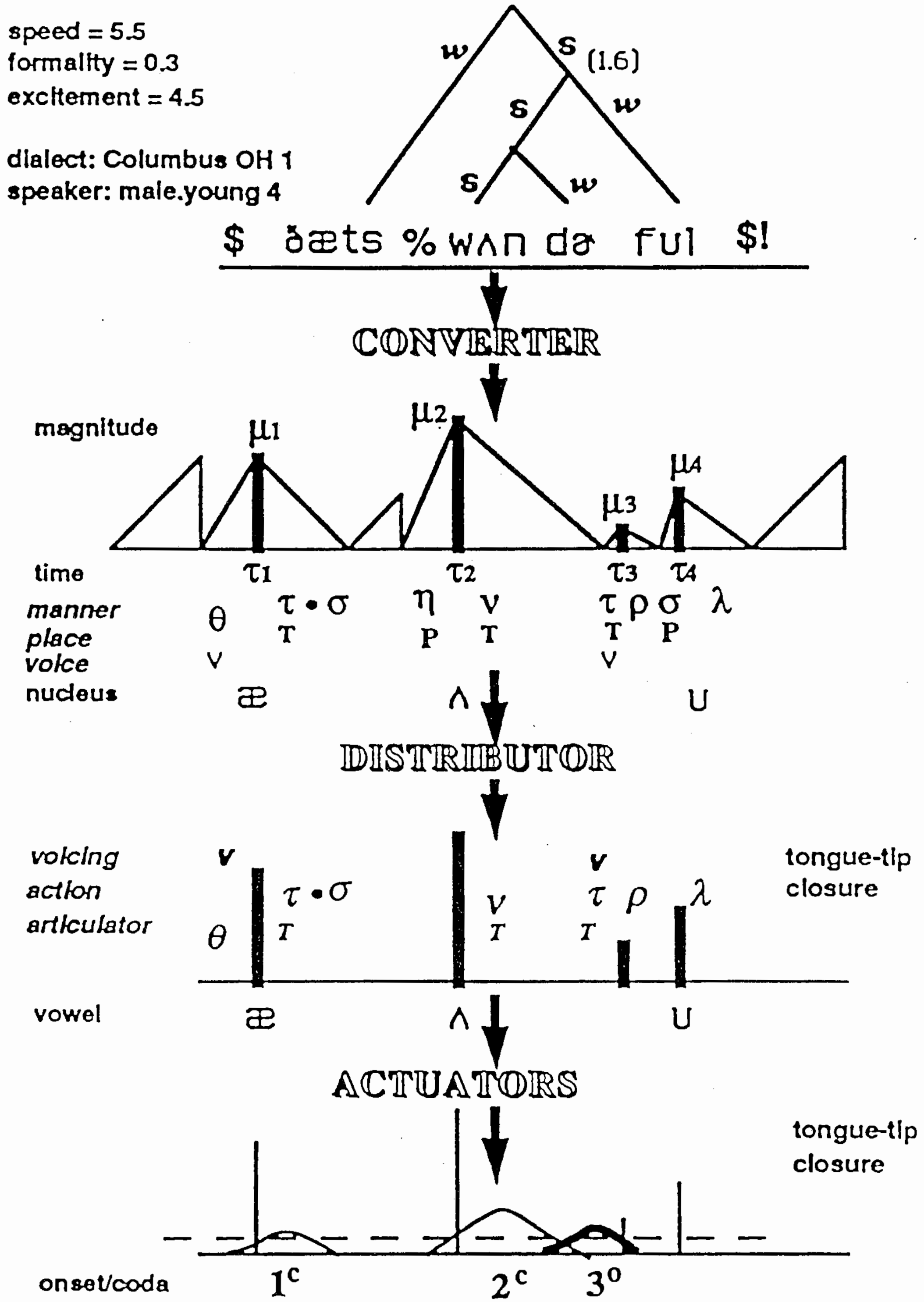

Figure 1: C/D Model, overall organization (signal generator not shown)

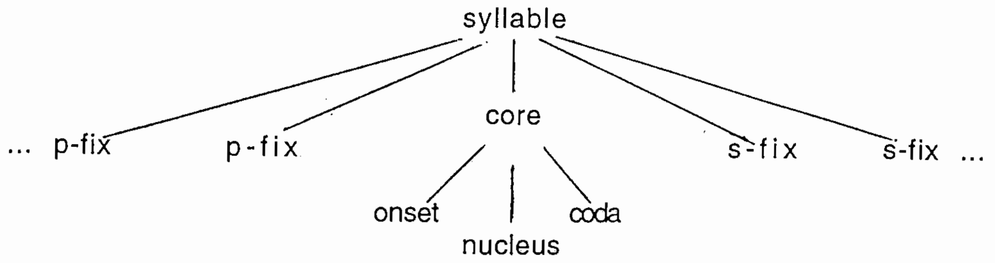

Fig. 2: Syllable structure

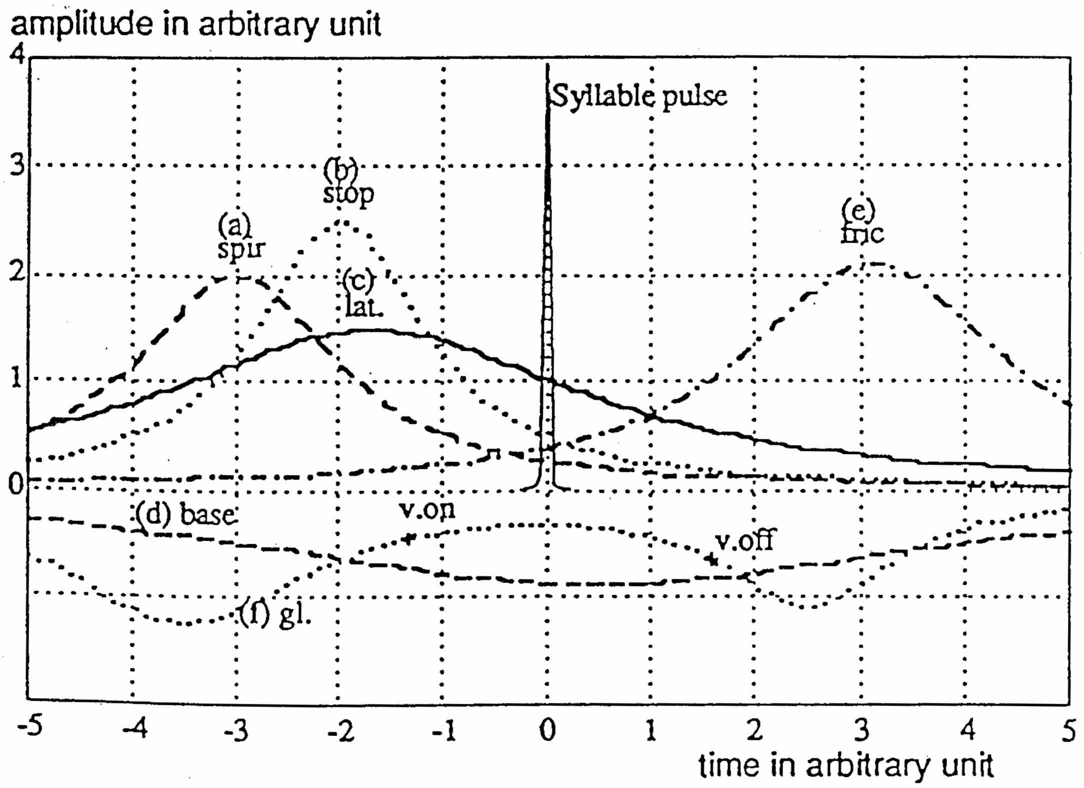

Figure 3: Impulse response functions (IRFs) for 'splash' (sample elemental gestures)

a. Frication constricting & releasing action in the tongue tip for spirantization
b. Closing & opening action in the lips
c. Narrowing/extending & releasing gesture for lateral in the tongue blade
d. Base function in vertical movement of the dorsum
e. Frication constricting & releasing action in the tongue blade for palatoalveolar fricative
f. Abduction & adduction gesture actions in the larynx for voice cessations in onset and coda.

gesture	*art.dim.-1*	*art.dim.-2*	*art.dim.-3*		*features*	
				manner	*place*	*voice*
p	lip-cl.			τ (stop)	P (labial)	-
sp	lip-cl.	spirant.		γ (spirant.)	P	-
b	lip-cl.			τ	P (labial),	V
t	tip-cl.			τ	T (apical)	-
st	tip-cl.	spirant.		γ	T	-
d	tip-cl.			τ	T	V
tʃ	blade-cl-grov.			τ	C	-
dʒ	blade-cl-grov.			τ	C	V
k	body-cl.			τ	K (dorsal)	-
sk	body-cl.	spirant.		γ	K	-
g	body-cl.			τ	K	V
m	lip-cl.		velum l.	ν (nasal)	P	-
sm	lip-cl.	spirant.	velum l.	γ ν	P	-
n	tip-cl.		velum l.	ν	T	-
sn	tip-cl.	spirant.	velum l.	γ ν	T	-
*ŋ	body-cl.		velum l.	ν	K	-
f	lip-retr.			σ (fricative)	P	-
sf	spirant.	labiodent.		γ σ	P	-
v	lip-retr.			σ	P	V
θ	tip-protr.	body-adv.		θ (interdent.)	-	-
ð	tip-protr.	body-adv.		θ	-	V
s	tip-grov.	tip-rais.		σ	T	-
z	tip-grov.	tip-rais.		σ	T	V
ʃ	blade-grov.	blade-rais.		σ	C (palatal),	-
ʒ	blade-grov.	blade-rais.		σ	C	V
l	blade-narr.	tip-cl.	body-retr.	λ (lateral)		-
r	tip-retr.	body-retr.		ρ (rhoticized)		-
w	lips-protr.	body-bunch.		ω (labio-velarized)		-
y	blade-grov.			η (palatalized)		-

All obstruent gestures accompany velum raising.

Table 1: Sample elemental gestures and feature speicifications for sample consonants in English

Here are corrections of errors in my technical report entitled The Syllable: Internal Structure and Role in Prosodic Organization (TR-H-103).

Section 1, p. 2, Paragraph 2, line -6 (minus line numbers means counting back from bottom): Idsardi [1993] should be Idsardi [1992]. The reference (p. 22) should be completed as: Idsardi, W. J. (1992). The Computation of Prosody (PhD Diss., MIT, Linguistics & Philosophy).

Section 2, p. 4, Paragraph 3, line 4: "ti" should be "τi" (Greek tau with subscript i).

p. 5, line 3: "IRFs (henceforth IRFs)" should be "impulse response functions (henceforth IRFs)".

Section 3, Para. 1: text line 9: Note number 4 should be 3.

p. 6, (2) line 9: note number 5 should be 4.

(3) line 3-4: "is necessary" should be "is not necessary".

(3) line -1: note number 6 should be 5.

p. 7, (5), lines 2 and 4: {strident} should be {spirantized}.

line -1: note number 7 should be 6.

Para -2 (starting with "A minimal"), line 11: Attach footnote number 7 to the end of sentence: {nasal} do.

p. 8, (1), line -1: note number 7 should be deleted.

p. 11, end of Para. 2: {affricate} or {spirantized} should be {affricate}, {spirantized}, or {interdental}.

end of Para. 3: Delete the last sentence "Interestingly,... s-fix".

p. 11, line -1: note number 3 should be deleted.

p. 15, line 4 "features to elemental gestures" should be "gestures to elemental gestures".

p. 18, Notes: Note numbers 5 and 6 should be exchanged.

p. 19 (Notes): Note number 7 should be 8. Before this, the following note 7 should be inserted: 7. When, in English, {spirantized, nasal} are both specified concomitantly, there is an oral closure implemented just as in the case of {spirantized, stop}, and the place feature also needs to be specified as in the case of {stop} for this reason.

p. 19: Note number 9 should be 10. Also "segnal" (first line) should be "signal".

Note number 10 should be 11. Also the second to the last line should read: "deformation) of the IRF, due to the context within the same articulator, is a ..."

Note 11 should be deleted.