Ann. Bull. RILP(1988) No. 22, 59-66

SOME REMARKS ON CONSONANT CLUSTERS

Osamu Fujimura

ABSTRACT

Implications of the demisyllabic theory of syllable structure are discussed in relation to recent studies of consonant clusters. Some basic characterizations of the phonetic implementation system, assuming an unordered set of articulator-based features as the constituents of a demisyllable, are mentioned. It is argued that the temporal organization of syllables must follow from the general framework of a universal rule system that allows for parameter settings of different types. Such a theory would systematically account for various effects of utterance conditions, as well as structural specifications of speech, on allophonic and temporal alterations.

0. Introduction

Consonant clusters in English have been of concern for some time. One of the motivations for my involvement, while I was at the Research Institute of Logopedics and Phoniatrics, was a crosslinguistic interest in the phenomena. I was interested in the difference between English and Japanese, and how this difference, in particular with respect to their syllable organizations, was reflected in the pronunciation/hearing performance of Japanese speakers learning English. With Haruhisa Ishida, I devised a computer-controlled system for systematically and adaptively testing students' oral shadowing performance of English words [Fujimura 1973, Ishida & Fujimura 1972], and my colleagues performed analyses of testing results [Smith *et al.* 1970, Harada 1971]. In designing this method, I used a combination of consonantal constituents with the syllable nucleus for what I called later a demisyllable. The point was that there must be some interesting patterns about the seemingly complex and *ad hoc* consonantal clusters in English, which Japanese students were having difficulties with, and we should be able to find some regular structural patterns empirically, if there is any effective framework for representing such clusters.

We did not go very far, but I did gain some insight into the structure of English syllables. I first proposed a practically useful way of describing English syllables using phonetically stable segmental subunits, *i. e.* demisyllables and syllable affixes [Fujimura 1976]. In this analysis, I noted, along other linguists, that the traditional concept of sonority is a difficult concept from a purely phonetic point of view, and used the term vowel affinity to indicate its functionality and language dependence. It was necessary to devise some deviations from the accepted phonemic theory, to account for regular exceptions, particularly the clusters /sp/, /st/, and /sk/ in English, as

AT&T Bell Laboratories

well as to separate out certain limited classes of phonemic segments at the end of the syllable as syllable affixes (as attached to a syllable core), which were usually but not always morphemic suffixes (see for related discussions including previous work by many scholars, Fujimura & Lovins [1978]).

This put me in a position to state that within a demisyllable, we do not need to specify temporal ordering of phonemic segments. The observed ordering follows automatically by assuming vowel affinity values for individual segments, and applying a vowel affinity principle as a phonotactic constraint Fujimura [1975]. Later [Fujimura 1979] (see also Fujimura [1988]), I abandoned the use of phonemic segments as the constituents of the demisyllable, and described a demisyllable as an unordered set of (demisyllabic) features. Phonetic implementation rules would respect inherent properties of features to determine temporal parameters of articulatory events that are inherently specified for the constituent features. To implement this idea was too big a job for me, and also I have not had any opportunity to seriously study many languages, some of which exhibit apparent difficulties for this analysis. The idea remains, therefore, only a hypothetical descriptive scheme. But the validity of demisyllables as phonetically stable segmental units of English has been demonstrated in speech synthesis [Fujimura *et al.* 1977; Macchi 1980; Browman 1980] and in automatic recognition [Rosenberg *et al.* 1983]. Also, I see some encouragement lately regarding its value for phonological theory, particularly in a recent discussion of the sonority cycle by Clements [in press], based on demisyllables as abstract units in lexical representations, and in the historical account of Estonian rhythm shift by Eek and Help [1987].

In this article, I will try to discuss some of the theoretical predictions of a version of demisyllabi c theory that may be interesting to look into, particularly given that the new x-ray microbeam system [Fujimura *et al.* 1973] is operating now in Madison, Wisconsin, providing us with opportunities to study articulatory phenomena systematically, extensively, and more easily than before [Nadler *et al.* 1987]. Many of the insights, of course, came from the preliminary data I have collected at the University of Tokyo using the earlier implementation of the microbeam system [Kiritani *et al.* 1975], and also from some preliminary studies using the fiberscope [Sawashima & Hirose 1968] in cooperation with Masayuki Sawashima and his colleagues while I was at the University of Tokyo. Also, needless to say, there have been many related contributions emerging both from the RILP and other research groups in these areas more recently.

1. Demisyllabic Interpretation -- Initial and Final Clusters and Syllable Affixes

Phrase initial consonants often differ qualitatively from corresponding final consonants, and this deviation from the linear phonemic principle based on concatenation and smoothing in part constituted our motivation toward the demisyllabic analysis [Fujimura & Lovins 1978]. For example, the difference between English light vs. dark [l]'s in initial and final positions escapes phonetic explanations based on assimilation principles, or, it seems, any other general principle that we know. Our fiberscopic observation [Fujimura & Sawashima 1971] revealed a peculiar use of false vocal fold constriction and associated laryngeal gestures for syllable final /t/ when followed by another /t/ or /d/ across a syllable (in this case always word) boundary. These details may well vary from dialect to dialect and also from speaker to speaker, as well as with miscellaneous utterance conditions like speech rate.

If we are to describe such allophonic variation in different languages by phonetic implementation rules of segmental features, as assumed in Chomsky and Halle [1968], we expect to have a very large collection of *ad hoc* context-sensitive rules, in which the context would have to be specified referring to distant conditions many segments away from the segment affected. We observe such complex rules in speech synthesis rules such as Klatt's [1987]. Note that most, if not all, of such strongly context-sensitive allophonic rules are concerned with the tautosyllabic conditions.

Demisyllabic description takes a basically different approach, where a given feature is in principle implemented in specific ways for individual demisyllables, under general constraints imposed by the principles of phonetic implementation. Such principles must respect the identity of the feature in different demisyllables in such a way that universal rule schemata define phonetic characteristics of each feature, in terms of partially specified target gestures including necessary dynamic specifications, respecting general constraints such as vowel affinity, and allowing parameter specifications. Such a theory, based on a universal vocabulary of features, is expected to reveal different classes of constraints, perhaps hierarchically structured in the form of a parameter setting scheme, reflecting the types of susceptibility of feature implementations to various utterance as well as language dependent conditions. The technical capability to handle such structured constraints efficiently within a complex rule system has become ours only relatively recently, in conjunction with AI-type symbolic manipulation tools, which are strong influencing virtually all subfields of theoretical linguistics [Fujimura to appear-a].

Among the most salient characteristics of English syllables is the peculiar status of the consonantal clusters /sp, st, sk/. These consonant sequences occur in both initial and final positions without changing phonemic ordering, for example 'sky' and 'task', resulting in an apparent violation of the generally (albeit vaguely) accepted sonority principle. Thus we see a minimal contrast such as 'tax' *vs.* 'task' in American English, even within the domain of monomorphemic syllables (see Clements [in press] for detailed discussions). Observing the paradigmatic pattern in English of initial /b, p, sp/, *etc.*, where /sp/ is implemented as though it were really an initial /s/ and an initial /b/ combined together, we can describe the feature bundle for /b, p, sp/ as:
{stop, lax, labial},
{stop, tense, labial},
{stop, spirant, labial},
respectively (I use here privatative features, see Clements, *ibid.*). The treatment of the s-element as the feature spirant (as opposed to apical and fricative for the place specified fricative /s/ (in 'sat' vs. 'fat') makes it possible to claim that, in English, place is specified only once in a demisyllable. The same feature spirant is used in the

initial clusters /sm/ vs. /sn/, as well as for characterizing the unusual occurrences of /sf/ (as in 'sphere').

Note that the final clusters /ps, ts, ks; bz, dz, gz/ as well as /mz, nz, ngz/ are all analyzed as part of a sequence of (final) demisyllable + affix. The affricate 'ch' is analyzed as a palatal stop in English. The unusual occurrence of initial 'ts' is as exceptional as the initial 'ps', and for these we probably should assume an exceptional syllable affix /t/ and /p/ as a prefix. English syllable affixes are generally obstruents that agree with the tautosyllabic place-specified obstruents (of the adjacent demisyllable) in voicing. Thus the /t/ and /d/ in 'act' and 'send', respectively, are affixes, but the final /t/ in 'sent' is not.

Syllable affixes (we called them synonymously phonetic affixes also) have a quasisyllabic status, and phonetically they are quite stable regardless of context [Macchi & Nigro 1977]. They also occur as a rule, apparently with some lexical exceptions, only morpheme finally (sometimes word medially) [Fujimura & Lovins 1978].

2. Articulator-Based Representation

If demisyllables and affixes are phonetically effective concatenative units, we may expect that they also serve as good units for predicting temporal patterns of speech organization based on duration or time interval computation. In order to prescribe the temporal organization of speech correctly, however, we need to consider a higher-level organization in addition to the so-called segmental (in our case demisyllabic) units. Roughly speaking, an utterance consists of a string of words, but it is easy to demonstrate by speech synthesis, that a simple concatenation at the phonetic level of lexical words uttered in isolation does not at all resemble what we expect as an utterance of a sentence. Furthermore, it is known that such prosodically inadequate speech signals are hard to process for human cognitive mechanisms.

Languages differ in the use of mechanisms for phonetic manifestations of the higher-level organization. Japanese and English sound very different, for example. In both languages, as in others, however, there is a strong indication of phonetic phrasing. One apparently universal characteristic is that the end of a phrase generally is marked by some slow-down process and characteristic pitch change (usually descent). What is meant by a "phrase" is not obvious, however, and there are reasons to believe that at least three levels of hierarchical phrases, or correspondingly three different boundary marks, are necessary to account for some of the prosodic phenomena (see, for example, Beckman and Pierrehumbert [1987]).

There is a further complication of temporal patterns that we have to be concerned with, when we need an exact description of speech signals. It is related to the inherent multidimensionality of the speech production process. Different articulatory organs have separate motor controls, and they bear different functions as physiological correlates of linguistic structural elements. Thus the mandible and the lower lip both contribute to the [p]-closure, but their contributions are shown to be independent bearing different linguistic functions, i. e. suprasegmental or configurational as opposed to segmental or inherent (see Macchi [1985, to appear]). Among the inherent features of segmental units, there is apparently a temporal organization program that functions for each articulator separately, with relatively loose linkage among different articulators. For example, a repulsion in time between contiguous articulatory events may be observed within the same articulator without affecting gestures of other articulators within the same phrase (see Fujimura [1981], [1986]). Some new findings from microbeam data concerning deviations from segmentally predicted timing patterns of gestures observed for different articulators are reported by Browman and Goldstein [to appear].

The articulator-based multidimensional (segmental) organization is particularly interesting in view of the emerging theory of nonlinear phonology which assumes different autosegmental tiers that basically correspond to individual articulators (McCarthy [to appear]).

3. Spring Model

The temporal organization of speech patterns is thus rather complex. In the multidimensional representation of the articulatory program, which Browman and Goldstein [*ibid.*] call the score, phonological specifications must be given sparsely rather than completely segment by segment for every feature. These specifications are given in terms of dynamic movement patterns as crucial events within the articulatory dimension of the proper articulators that are responsible for the phonological feature specifications [Fujimura 1987]. While no comprehensive algorithms nor implementations have been proposed to my knowledge, efforts are being made in this direction (see Fujimura [to appear-b]), and relevant discussions are taking place among some phonologists and phoneticians (see Kingston & Beckman [in press]).

A general framework for constructing articulatory scores may be described in a form of a multidimensional spring model (see Fujimura [1987-b]). Such a model will have to be structured hierarchically, reflecting the effects of phrasing at different levels and representing prosodic controls attached to relevant units such as stressed syllables. It seems to me that it is mandatory to employ such computational models to be able to account for the seeming complexity of emerging observations about speech production processes, and thereby design general enough computational schemes for future speech technology. Linguistic insights and careful observations of pathological and developmental speech phenomena will also serve pivotal roles in such basic research. There is much to be expected from the Research Institute of Logopedics and Phoniatrics, about which Masayuki Sawashima and myself share a lot of nostalgia.

REFERENCES

Browman, C. P. (1980), Rules for Demisyllable Synthesis Using LINGUA, A Language Interpreter, *ICASSP 80 Proceedings, Vol. 2,* Piscataway, NJ, IEEE Service Center, 561-564.

Browman, C. P. and Goldstein, L. (to appear), Some Notes on Syllable Structure in Articulatory Phonology, in *Articulatory Organization -- Phonology to Speech Signals,* O. Fujimura (ed.), Basel, S. Karger.

Clements, G. N. (in press), The Role of the Sonority Cycle in Core Syllabification, in *Papers in Laboratory Phonology I,* J. Kingston and M. Beckaman (eds.), Cambridge, Cambridge Univ. Press.

Eek, A. and Help, T. (1987). The Interrelationship between Phonological and Phonetic Sound Changes: A Great Rhythm Shift of Old Estonian, *Proc. XIth International Congress of Phonetic Sciences, August 1-7, 1987, Tallinn, Estonia, USSR, Vol. 6,*218-233.

Fujimura, O. (1973), Technological Development in Language Learning, in *Applications of Linguistics,* G. E. Perren and J. L. M. Trim (eds.), Cambridge, Cambridge Univ. Press., pp. 41-64.

Fujimura, O. (1975), Syllable as a Unit of Speech Recognition, *IEEE Transactions in Acoustics, Speech and Signal Processing 23,* 82-86.

Fujimura, O. (1979), An Analysis of English Syllables as Cores and Affixes, *Zs. f. Phonetik, Sprachwissenschaft u. Kommunikationsforschung 32,* 471-476.

Fujimura, O. (1986), Relative Invariance of Articulatory Movements: An Iceberg Model, in *Invariance and Variability in Speech Processes,* J. S. Perkell and D. H. Klatt, (eds.), Hillsdale NJ, Lawrence Erlbaum Assoc., pp. 226-242.

Fujimura, O. (1987), Fundamentals and Applications in Speech Production Research, *Proc. XIth International Congress of Phonetic Sciences. August 1-7, 1987. Tallinn, Estonia, USSR, Vol. 6,*10-27.

Fujimura, O. (1987b), A Linear Model of Speech Timing, in *In Hornor of Ilse Lehiste,* R. Channon and L. Shocky (eds.), Dordrecht, Foris Pub., pp. 109-123.

Fujimura, O. (in press), Demisyllables as Sets of Features -- Comments on Clements's Paper, in *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech,* J. Kingston and M. E. Beckman (eds.), Cambridge, Cambridge Univ. Press.

Fujimura, O. (to appear-a), Current Trends in Linguistic Research, in *Proc. Kobe Conference in Japanese Syntax,* M. Shibatani (ed.).

Fujimura, O. (ed.)(to appear-b), *Articulatory Organization -- Phonology to Speech Signals*, Basel, S. Karger.

Fujimura, O. and Sawashima, M. (1971), Consonant Sequences and Laryngeal Control, *Ann. Bul. No. 5, RILP, U. Tokyo*. 1-6.

Fujimura, O., Kiritani, S., and Ishida, H., Computer-Controlled Radiography for Observation of Movements of Articulatory and other Human Organs, *Comput. Biol. Med. 3*, 371-384.

Fujimura, O., Macchi, M. J. and Lovins, J. B. (1977), Demisyllables and Affixes for Speech Synthesis, *Proc. Nineth International Congress on Acoustics, July 4-9, Madrid, Spain, Contributed Paper 1*, p.513.

Fujimura, O. and Lovins, J. B. (1978), Syllables as Concatenative Phonetic Units, in *Syllables and Segments*, A. Bell and J. B. Hooper (eds.), Amsterdam, North Holland Pub., pp. 107-120.

Harada, K. I. (1971), Experiments on the T-T System, *Ann. Bul. No. 5, RILP, U. Tokyo*, 51-68.

Ishida, H. and Fujimura, O. (1972), A Computer-Based Pronunciation-Hearing Test System Using a Hybrid Magnetic Tape Unit, *Form and Processing 71*, 1037-1040.

Kingston, J. and Beckman, M. E. (eds.) (in press), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, Cambridge, Cambridge Univ. Press.

Kiritani, S., Itoh, K. and Fujimura, O. (1975), Tongue-Pellet Tracking by a Computer-Controlled X-ray Microbeam System, *J. Acoust. Soc. Am. 57*, 1516-1520.

Klatt, D. H. (1987), Review of Text-to-Speech Conversion for English, *J. Acoust. Soc. Am. 82*,737-793.

Macchi, M. J. and Nigro, G. (1977), Syllable Affixes in Speech Synthesis, *J. Acoust. Soc. Am. 61*, S6.

Macchi, M. J. (1980), A Phonetic Dictionary for Demisyllabic Speech Synthesis, *ICASSP 80 Proceedings, Vol. 2.*, Piscataway, NJ, IEEE Service Center, 565-567.

Macchi, M. J. (to appear), Labial Articulation Patterns Associated with Segmental Features and Syllable Structure in English, in *Articulatory Organization*, O. Fujimura (ed.).

McCarthy, J. J. (to appear), Feature Geometry and Dependency, in *Articulatory Organization*, O. Fujimura (ed.).

Nadler, R., Abbs, J. H. and Fujimura, O. (1987), Speech Movement Research Using the New X-ray Microbeam System. *Proc. XIth International Congress of Phonetic Sciences, August 1-7, 1987, Tallinn, Estonia, USSR, Vol. 1*,221-224.

Rosenberg, E. A., Rabiner, L. R., Wilpon, G. J. and Kahn, D. (1983), Demisyllable-based Isolated Word Recognition System, *IEEE Transactions in Acoustics, Speech and Signal Processing 31*, 713

Sawashima, M. and Hirose, H. (1968), New Laryngoscopic Technique by Use of Fiberoptics, *J. Acoust. Soc. Am. 43*, 168-169.

Smith, D., Ito, K., Sato, C., Ishida, H., and Fujimura, O., The Pronunciation-Hearing Test Using a Hybrid Magnetic Tape System, *Ann. Bul. No. 4, RILP, U. Tokyo*, 111-113.