UCLA
Working Papers in Phonetics

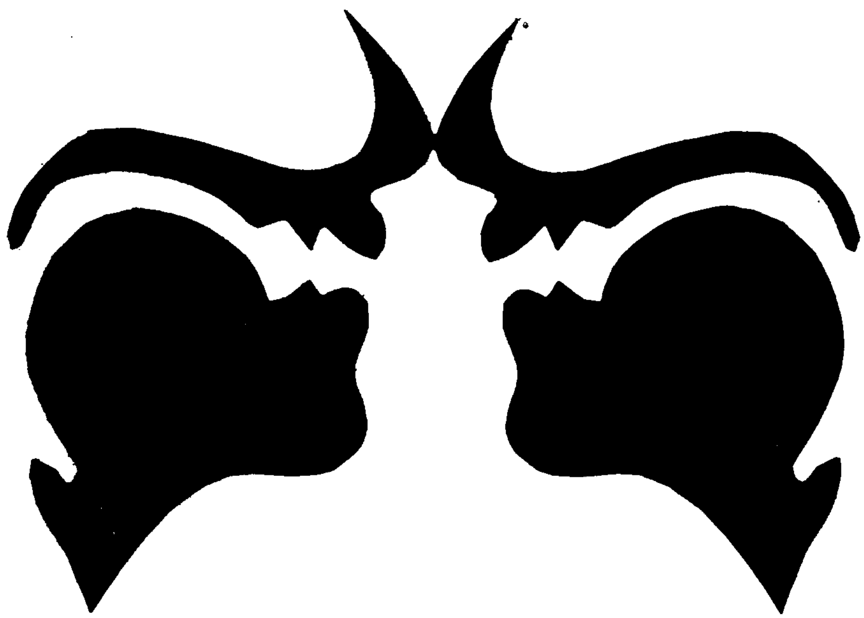
Title
WPP, No. 17

Permalink
https://escholarship.org/uc/item/7s65266q

Publication Date
1971

eScholarship.org Powered by the California Digital Library
University of California

Working Papers in Phonetics Number 17



U.C.L.A. January 1971

UCLA Working Papers in Phonetics 17

January, 1971

Abstract of paper presented at a meeting

Papers on their way to regular publication

The UCLA Phonetics Laboratory Group

Research

George D. Allen [till September 1970; now at University of North Carolina]

Victoria A. Fromkin
Richard Harshman
Leon Jacobson
Steven Krashen
Peter Ladefoged
Mona Lindau
George Papçun
Raymond Silverstein [till August 1970; now at University of Southern Illinois]

Dale Terbeek

Technical and Secretarial

Larry Grant
Julie Haaker
Willie Martin
Renee Wellin
Jeanne Yamane

As on previous occasions, the material which is presented here is simply a record for our own use, a report as required by the funding agencies, and a preliminary account of work in progress.

Funds for the UCLA Phonetics Laboratory are provided through:

USPHS grant NB 04595
ONR contract NR 049-226
NSF grant GS 2741
NSF grant GS 2859
and the UCLA Department of Linguistics

Correspondence concerning this series should be addressed to:

Professor Peter Ladefoged
Phonetics Laboratory
UCLA
Los Angeles, California 90024

There's Many a Slip Twixt the Brain and the Lip

Victoria A. Fromkin

[Paper given at the Houston meeting of the Acoustical Society of America, November 1970]

The difficulties inherent in modeling a speech production system stem from its great complexities and from the fact that the only direct information available to us concerns the output characteristics of the system. Thus, the input units and the properties of the mechanism itself must be inferred primarily from output data. An analysis of over 700 speech errors reveals that these data are an invaluable source of information regarding both the units processed and the sequencing of events in the production of an utterance. The kinds of errors which occur show that whereas the articulators move in an analog fashion producing a semi-continuous acoustic signal, discrete units similar to the abstract units posited by linguists, underlie the speech production process. Distinctive features, segments, syllables, lexical and grammatical morphemes, and 'labelled' syntactic phrases must all be posited as performance units to account for the errors which occur. Furthermore, there are language specific rules, analogous to grammatical rules, which constrain the production process. These rules must be strictly sequenced in the model to insure valid predictions. Evidence is presented in this paper for the reality of certain highly abstract representations of formatives which are actualized phonetically only in speech errors, and for specific morphophonemic rules which must be ordered before phonological-phonetic rules. A sequential ordering of events in the generation of an utterance is proposed.

When is a non-occurring 'possible' verb impossible?

Victoria A. Fromkin

The Main Stress Rule (MSR) proposed by Chomsky and Halle in SPE (1968) would assign primary stress in verbs to a final strong syllable or to the penultimate if the final syllable is weak. This accounts for the stress differences in *exit/exist* etc. To show the 'correctness' of this rule twenty-eight graduate students at the beginning of a class in English phonology were given a number of sentences which included non-occurring 'possible' English words. They were instructed to write the underlined words in phonetic transcription with primary and secondary stress marked. (These were the same sentences used by Ladefoged and Fromkin (1968)). The vowel rules and stress assignment rules worked fine for most of the sentences. Thus in 'Then he will *sempit* the result' all twenty-eight transcribed *sempit* as [sɛ́mpət] or [sɛ́mpɩt] and in 'Everyone will *submist* it' the phonetic form given was [səbmɩ́st]. Unfortunately, my little experiment fell flat on its face with 'Unless he has to *sempist* the lot', where twenty-five students gave the transcription as [sɛ́mpəst], and in 'Then he will decide to *tenist* the process' where twenty-five of the transcriptions given were [tɛ́nəst].

Ross (forthcoming) discusses a number of problems with the MSR and the ASR (Alternating Stress Rule) as given in SPE. But his proposals will not account for the strange stress assignment by the students in all the examples given above. The penultimate stress in *tenist* can possibly be handled by his rules, which also assign a 1-0 stress in *challenge*, *scavenge, govern* etc. but will not account for [sɛ́mpəst]. In fact he states that 'any verb ending in (an *st* cluster) must receive final stress by the MSR.' (p. 41)

How then should we account for the difference in stress between *molest, accost, arrest* and the stress assigned to these 'possible but non-occurring verbs'?

It seems that, *sempist* and *tenist* are *not* possible verbs in English. A quick glance at a reverse word list (Dolby and Resnikoff, 1964) reveals that all verbs ending in *ist* are monosyllabic stems preceded by a prefix. (e.g. *enlist, attrist, subsist, desist, resist, persist, exist* and a few others). Furthermore, there do not seem to be any di-syllabic noun stems ending in *ist*. This particular phonological string is restricted to monosyllabic morphemes -- lexical, as in *cyst, grist* etc. or derivational, as in the suffix *-ist*.

It is no wonder then that the 'verbs' *tenist* and *sempist* were stressed 'incorrectly'. They can't be verbs in English.

The sequential constraints on morphemes in English seem to be more restrictive than we have previously considered.

References

Chomsky, N. and Halle, M. (1968) *The Sound Pattern of English.* Harper and Row.

Dolby, J. L. and Resnikoff, H. L. (1964) *The English Word Speculum Vol. III. The Reverse Word List.* Lockheed Missiles and Space Co.

Ladefoged, P. and Fromkin, V. (1968) "Experiments on Competence and Performance" *IEEE Transactions Vol. AU-16, No. 1.* March

Ross, J. (forthcoming) "A Renanlysis of English Word Stress" (Mimeod paper) MIT

Binary Suprasegmental Features

Ralph Vanderslice[1] and Peter Ladefoged

SUMMARY

The following binary features are proposed for the suprasegmentals of English: (1) Strong, (2) Accent, (3) Intonation, (4) Cadence, (5) Endglide, (6) Emphasis. It is shown that this binary suprasegmental feature system has many advantages. It accounts for the grammar-expounding prosodic distinctions of English in an intuitively satisfying way; it factors out accentual and intonational phenomena conflated in "stress"; and it incorporates linguistically significant generalizations about the syntactic conditioning of prosodic patterns. In addition, it can be used to explain rhythmic and stylistic variations disregarded by previous stress rules; and it enables revision of the stress cycle in ways which enhance its economy and explanatory power while retaining its insights and concinnities.

1. Distinctive features need not be binary features; for some phonetic parameters -- like vowel height -- multivalued ones are more explanatory (Ladefoged 1970). Ironically, in the one place where *The Sound Pattern of English* (Chomsky and Halle 1968; henceforth SPE) uses a multivalued feature -- in the system of multiple stress levels -- this approach is inferior in explanatory power to the binary suprasegmental feature system we are proposing here. Substituting our features in place of stress levels in the cycle enables its revision in a way which preserves all of its insights and concinnities -- our version still derives exactly the correct prosodic difference between cóndēnsátion and cómpĕnsátion, for example -- while enhancing its degree of linguistically significant generality.

The features we propose are summarized in (1). The first distinction is between strong and weak syllables. If a syllable is strong it will have its full vowel quality; and if it is weak it will be completely unstressed and it will often have a reduced, centralized, vowel. The phonetic mechanisms underlying this opposition are not very well understood,

---

1 Ralph Vanderslice is at Hunter College of the City University of New York, New York 10021.

(1)

Feature Name	Phonetic Correlates
±strong	Full articulations vs. reduced timing
±accent	Presence vs. absence of increased respiratory energy, and laryngeal adjustments,causing a pitch obtrusion.
±intonation	(a plus value of one or both of the following features)
±cadence	Presence vs. absence of a falling pitch
±endglide	Presence vs. absence of a terminal pitch rise
±emphasis	Extra pitch obtrusion

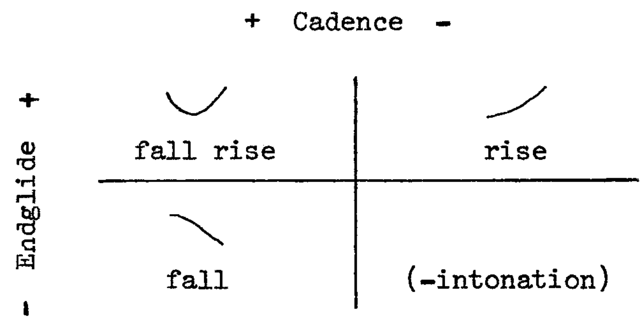
but they presumably involve the systems used in timing the articulations within a syllable.

The next distinction is between accented and unaccented syllables. The phonetic correlates of an accented syllable always include an increase in respiratory energy (Ladefoged 1967), and a concomitant pitch obtrusion (Bolinger 1965). If a syllable is accented it is always [+strong]. Unaccented syllables may be [+strong] or [-strong].

The third feature in our system is called Intonation. This feature differs from the two we have discussed so far, in that it will never appear in the systematic phonetic transcription which is the output of the phonological component, because it will always have been replaced by other features with more specific phonetic correlates which will be discussed below. We do not believe in Chomsky and Halle's dictum "The phonetic features can be characterized as physical scales describing independently controllable aspects of the speech event" (SPE: 297). It has been shown elsewhere (Ladefoged, in press) that there must be features such as Intonation and Consonantal which are needed in phonological rules, but which are merely cover terms for certain combinations of other features, and which consequently cannot themselves be defined in terms of physical scales.

The two features which are used to replace the feature Intonation at later stages in the rules are Cadence and Endglide. Cadence refers to a falling pitch on the nuclear syllable (the syllable which has been marked [+intonation]), and Endglide refers to the possibility of a rising pitch after that point. These two features define three intonation curves as shown in (2). In this system a non-nuclear syllable (one which

(2)



is marked [-intonation]) is not the starting point of any of the three intonation curves.

The final feature listed in (1) is Emphasis. The phonetic correlate of this feature is the presence of an extra pitch obtrusion on a particular syllable. It is thus roughly equivalent to pitch level 4 in the Trager-Smith (1957) system.

At this point we may consider the relation between the first three of these features (Strong, Accent, and Intonation) and the systems which have been used for marking stress in phonetic dictionaries. In his *English Pronouncing Dictionary* Jones (1956) uses only a primary stress mark on the majority of polysyllabic words. In words such as those in (3) he uses a secondary stress mark as well. These words are those in

(3) ˌrelaxˈation , exˌaminˈation , ˌsopoˈrific

which, in our terms, there are two syllables which have to be marked [+accent], the second being also (at least in citation forms) [+intonation].

In *A Pronouncing Dictionary of American English* Kenyon and Knott (1953) are slightly more prolific in their use of secondary stress marks. In addition to using them in words such as those given in (3), they also use them in words such as those in (4). Kenyon and Knott's usage can

(4) ˈhurriˌcane , ˈanecˌdote , ˈhandˌmaid

therefore be interpreted in our terms by saying that, in dictionary citation forms, their primary stress equals our [+intonation, +accent]; and if their secondary stress occurs before the primary stress mark within a word then it equals our [+accent, +strong], but if it occurs after the primary stress mark, then it equals our [-accent, +strong]. In a few cases, such as those in (5), Kenyon and Knott mark two primary stresses.

(5) ˈhandˈmade , ˈarchˈbishop

We will consider the interpretation of these examples in section 4 of this paper.

As a further exposition of the system we are proposing, consider the prosodic aspects of the phrase the telegraph, shown in (6a) in an iconic notation. There is what Bolinger (1965) would call an upward obtruded pitch accent on tel-, followed at once by the pitch drop associated with the intonation pattern. There are two weak syllables, the and -le-, with reduced vowels; the is at neutral pitch, while -le-, being within the domain of cadence, is low-pitched; -graph with its unreduced vowel is a strong syllable, but it is not accented. If it were, the intonation drop would be there instead (télegráph).

(6)(a)

(b)

	the	tel	e	graph
strong	−	+	−	+
accent	−	+	−	−
intonation	−	+	−	−

(c) the télegrāph

the	tél	e	grāph
−	+a	−	+s
	+cadence		

These facts are neatly accounted for with our features as shown in (6b), where the and -le- are [-strong], tel- and -graph are [+strong], and tel- is also [+accent] -- and in this case, being the nuclear syllable, is [+intonation]. Taking account of redundancies and using only dashes for the weak syllables (as in SPE), and assuming the intonation to be specified as falling or [+cadence], the same utterance can be unambiguously represented as in (6c) -- where "+a" means a strong accented syllable and "+s" a strong unaccented one. The diacritics above the syllables merely reduplicate in a more perspicuous form the features shown below (the macron stands for [+strong, -accent]).

2. There are substantial economies to be gained by a proper factoring out of intonation from what Bolinger (1965: 95) justly called the "prosodic souse" of stress. A notable case, ripe for Occam's Razor, is the primariness (in relation to their secondaries) of Trager and Smith's primary stresses, which in our system are merely [+accent, +intonation]; that is, nuclear accents. In the case of (7a), with so-called secondary-primary stress pattern, all that distinguishes the stress on na- of nation from that on e- of either is that the accented syllable of nation serves as the starting point for the falling intonation or cadence. But the grammar must independently account for this, because it contrasts with other intonations such as (7b) where a wholly different pitch pattern (corresponding to the feature [+endglide]) distinguishes the

(7)(a) $\overset{2}{\text{e}}$ither n$\overset{1}{\text{a}}$tion.
+a – +a –
+cadence

(b) $\overset{2}{\text{e}}$ither n$\overset{1}{\text{a}}$tion?
+a – +a –
+endglide

so-called stresses "1" from "2". So it is obvious that, on both phonetic and systemic grounds, any perceptual impression of greater prominence on nation as compared with either should be abstracted away from the accentual system and attributed to intonation. Once this is done, the accentual units ascribed to either and nation may be treated as systemic "sames", and this is what our features do.

Our system makes no provision for postnuclear accents. This amounts to a denial of the Trager-Smith-ical contrast which is assumed to be as in (8). From our point of view the contrastive noun phrase and the compound noun are homophonous; in both sentences bird is just a strong

(8)(a) **He saw a bl$\overset{1}{\text{a}}$ck b$\overset{2}{\text{i}}$rd, but not a green one (NP, constrastive)**

(b) **He saw a bl$\overset{1}{\text{a}}$ckb$\overset{3}{\text{i}}$rd, but not a crow (N, compound)**

unaccented syllable. Whether they can be disambiguated by facultative means such as Bolinger and Gerstman's (1957) disjuncture is not in the least germane to the issue of stress levels. We know of no empirical evidence whatever which would contravene our more parsimonious assumption that all postnuclear strong syllables are systemically equal. To summarize our view, we hold that Trager and Smith's (and Chomsky and Halle's) secondary and primary stresses are the same thing (viz [+accent]) except in postnuclear cases like (8a) where their secondary stress is the same thing as their tertiary (viz [+strong, -accent]). This point is discussed in extenso elsewhere (Vanderslice 1968a: 66-72, 108-111; and 1968b: 28-30, 40-52).

The much-maligned taxonomists held stress distinctions dear as phonetic cues for their IC cuts. Yet it is SPE which shows not two but THREE ostensibly distinct levels of stress among unreduced vowels in postnuclear position, as exemplified in (9):

(9)(a) **$\overset{1}{\text{e}}$lev$\overset{4}{\text{a}}$tor b$\overset{3}{\text{o}}$y**

ch$\overset{1}{\text{e}}$mistry t$\overset{3}{\text{e}}$acher

(a') **élevātor bōy**
+a – +s – +s

chémistry tēacher
+a – – +s –

(b) $\overset{1}{\textbf{e}}\textbf{lev}\overset{4}{\textbf{a}}\textbf{tor}$ $\overset{2}{\textbf{o}}\textbf{per}\overset{4}{\textbf{a}}\textbf{tor}$

$\textbf{ch}\overset{1}{\textbf{e}}\textbf{mistry}$ $\textbf{l}\overset{2}{\textbf{a}}\textbf{borat}\overset{4}{\textbf{o}}\textbf{ry}$

(b') **élevātor ōperātor**
+s – +s – +s – +s –

chémistry lāboratōry
+s – – +s – +s –

We feel that our analysis, transcribed at the right, quite aside from its evident economy and elegance is closer to the empirical facts. In neither of our dialects (roughly, Mid-Western American and British RP) is there any distinction among postnuclear secondary, tertiary, and quaternary stresses in these examples; and it is noteworthy that Chomsky and Halle themselves admit that distinctions of this kind have a "dubious factual status" (SPE: 109). But perhaps this is an appropriate moment to point out that we, too, are uncertain about the status of some postnuclear unstressed syllables. It may be the case that in some dialects a late rule marks all tense final vowels [+strong]. Alternatively there may be oppositions which are manifested by a syllable with a final tense vowel being either strong or weak. The classic examples (éffĭgy̆ ≠ réfŭgēe, wíndŏw ≠ élbōw) like many in the Trager-Smith tradition, scarcely comport with Wang's "important observation about phonemic distinctions", viz "that native speakers respond to them with effortless consistency" (1962: 69).

3. At the level of what Bierwisch (1968) calls Phrase-accent rules -- and especially the level of the sentence and the discourse, as Gunter (1966) and others have shown -- there are accentual perturbations that cannot be accounted for by any conceivable means within the SPE framework. As Bierwisch says: "the assignment of accent patterns is not a matter of the phonological component alone, given the syntactic surface structure as an input. Rather it depends directly on very early syntactic processes in certain respects. It may very well be that the strict separation of syntactic and phonological rules, which has thus far been held to be necessary, is in fact impossible" (1968: 178).

Chomsky and Halle cite contrasts like <u>electrical</u> <u>rather</u> <u>than</u> <u>civil engineering</u> and <u>encourage</u> vs <u>DIScourage</u> as among the problems not accounted for (SPE: 24). Actually the interesting cases involve de-accentuation not by formal identity of components but by semantic niceties of coreference. Thus in (10), an example given by Ohala (1970), <u>to incarcerate</u> is a synonym of <u>to jail</u>. As a result it is obligatorily de-accented; and thereby it becomes postnuclear and part of the cadence --

(10) **Instead of jailing the thief, it's the mayor we should incarcerate**

(11)**(a) Hámlet's hát is bīgger than Horátio's hélmet**

(b) *Hámlet's hát is bīgger than Horátio's héadgear

(12)**(a) Pát$_1$ blūrged$_x$ Chrís$_2$ and then she$_1$ skráffed$_y$ him$_2$**

(b) Déll$_1$ flōobed$_x$ Bért$_2$ and then shé$_2$ clōafed$_x$ hím$_1$

i.e. low pitched. Similarly, though one can say (11a), because *hat* and *helmet* name only partially intersecting semantic categories, one cannot say (11b), because *headgear* is a hypernym of *hat* and must be de-accented (Vanderslice 1968b: 7ff). Then there is the closely related phenomenon whereby in (12a) the accentuation tells us that Pat must be female, Chris male, and *scraffed* different from *blurged* as *kicked* from *hit*, for example; whereas in (12b) Bert is female, Dell male, and *cloafed* and *floobed* synonymous, like *hit* and *struck*.

Finally, consider (13) -- in each version of which the semantic category "house" is assumed to be marked [+mentioned] in the way Heidolph (1966) and others have suggested. The WORD *house* is therefore obligatorily at least de-accented throughout, and in final position there is a progression through optional pronominalization (b) and optional ellipsis (c). We assume that surface structures generated by the syntactic

(13)**(a) This is the brówn hōuse, not the gréy hōuse**

(b) This is the brówn hōuse, not the gréy ōne

(c) This is the brówn hōuse, not the gréy

component encode these obligatory accentual effects by means which take phonological effect after the word-accent rules have applied, but before the late rules which assign rhythmic and stylistic accentual variations of the sort we consider next.

4. It has been a commonplace for upward of two centuries (and Chomsky and Halle are well aware -- e.g. SPE; 117) that of three close-spaced strong syllables, all potentially accentable, the middle one is liable

to be de-accented, especially at conversational rate. Daniel Jones (1964) has given a particularly lucid treatment of this (§9.32, pp. 253-4). A phrase which in slow, emphatic speech might be accented like (14a) will be rendered in rapid speech as (14b):

(14)**(a)** **bíg bád wólf**
+a +a +a

(b) **bíg bād wólf**
+a +s +a

This works also with polysyllables, causing them to exhibit what is traditionally called "accent recession". Examples are given in (15). The last example, for which Bolinger (1965: 161n) gives credit to D. W. Reed,

(15)	A clárinet sólo	He pláys the clarinét
	Ténnessee Wílliams	The státe of Tennessée
	A télegraphic stýle	His stýle was telegráphic

suggests that in the SPE derivation *telegraph-telegraphy-telegraphic*, the 3-1 stress pattern assigned to the adjectival form oversimplifies the facts. It is adequate for the context Chomsky and Halle consider, which is shown in (16a); but in another context, such as (16b), a different arrangement of stress numbers would be needed to fit the facts.

(16)**(a)** **We estáblished tēlegráphic commūnicátion**

(b) **We estáblished it by télegrāphic cóntāct**

We think the lexical accent rules should accent both strong syllables of such words, leaving the determination of which accents are ultimately to be realized phonetically to the care of a late rhythm rule applying after syntactic accent perturbations.

Note in passing that the stress values SPE assigns to *telegraphic* in (16a) are not actually 3-1-, but rather 5-3-; and adding more words to the sentence could modify the stress digits still further, without the slightest phonetic motivation. One of the most patent virtues of our feature system is that it eliminates these spurious complications from the prosodic description, while accounting (as SPE does not) for real accentual variation.

5.0 The question now arises whether our binary features can be assimilated into the stress cycle without mucking it all up. The answer is yes. Revisions of the principal rules are shown here as emendations to the originals in the form given by SPE in chapter five, the Summary of Rules. The cyclical rules are starred (here as there), and we adopt the convention (17a) which states that whenever accent is assigned by

(17) **General Conventions for Accent Rules:**

> **(a) Each application of [+accent] by a cyclical rule (starred) converts other vowels within the word to [−accent].**
>
> **(b) Application of [+accent] to a vowel by any rule implies simultaneous application of [+strong].**

a cyclical rule, any previously accented syllable in the same word is de-accented (but remains strong). This convention is very similar to the SPE convention whereby each assignment of primary stress causes all previously assigned stresses to be lowered by one degree. But note that our convention applies only to the starred rules. It is perfectly possible for the application of non-starred rules to result in there being several syllables in a phrase which are all equally [+accent]; and in most circumstances only the last of these syllables will become more prominent by being subsequently marked [+intonation].

5.1 Rule (18) is the one that precedes the Main Stress Rule in the cycle just in order to get tense suffixes through it correctly, as for

*(18) **FUDGING—I** $\begin{bmatrix}+\text{tense}\\ \text{V}\end{bmatrix} \rightarrow \overset{\textbf{+accent}}{[1\ \text{stress}]}\ /\ + \underline{\quad} C_0\#$

example in <u>molluscoid</u>, <u>recondite</u>, and <u>sychophantize</u> -- all of which apparently are household words in Cambridge. We think it inconvenient to have stress rules without names, and accordingly we propose that this and a later rule be called "Fudging Rules I and II." In view of what the rules do, these names seem natural mnemonics. Our revision of the First Fudging Rule consists merely in providing that tense vowels are assigned accent, rather than primary stress, in the environment shown.

5.2 In the Main Stress Rule (19), the substitution of our features has the salutary effect of restoring it essentially to its pristine, archetypal form. The accretion of all those Greek-letter variables and the conditions on them occurred in response to various problems which turn out to be stress-system dependent. With our features, no such need arises.

*(19) MAIN STRESS

$$V \rightarrow \underset{\textbf{+accent}}{[1\ \text{stress}]} \Big/ \Big[X \text{---} C_0\Big(\underset{\textbf{-accent}}{\begin{bmatrix}-\text{tense}\\ \gamma\text{stress}\\ V\end{bmatrix}} C_0^1\Big(\begin{bmatrix}\alpha\text{voc}\\ \alpha\text{cons}\\ -\text{ant}\end{bmatrix}\Big)\Big)$$

$$\Big/ \text{---}\Big\langle\Big(\begin{Bmatrix}(fik)At\\ \\ [+D]C_0\end{Bmatrix}\Big)\begin{Bmatrix}\langle_1 +C_0\rangle_1 \overset{\textbf{-strong}}{\begin{bmatrix}-\text{stress}\\ -\text{tense}\\ -\text{cons}\end{bmatrix}}[+\text{cons}]_0 \\ \langle_1\begin{bmatrix}-\text{seg}\\ \langle_2 -\text{FB}\rangle_2\end{bmatrix}\rangle_1 C_0 \underset{\textbf{+accent}}{[\beta\text{stress}]} C_0\langle_2 V_0 C_0\rangle_2\end{Bmatrix}\Big\rangle\Big]_{\langle NSP\langle_1 VA\rangle_1\rangle}$$

Conditions: $\cancel{\beta \neq \begin{Bmatrix}2\\1\end{Bmatrix}}$ $\cancel{\gamma \leq 2}$ ≥ 2

where X contains no internal #

Incidentally, one of the remarkably rare SPE typos occurs here in the conditions: gamma is supposed to be "2 or weaker" (SPE: 109), and of course in a perverse system where less is more, this requires that gamma be GREATER-than-or-equal-to 2.

5.3 The revised Alternating Stress Rule (20) shifts the accent from the final syllable (which however remains strong) to the antepenultimate syllable in a multitude of words like matador, formaldehyde, baritone, etc.

*(20) ALTERNATING STRESS

$$V \rightarrow \underset{\textbf{+accent}}{[1\ \text{stress}]} \ / \text{---} C_0(=)C_0VC_0\underset{\textbf{+accent}}{[1\ \text{stress}]}C_0]_{NAV}$$

5.4 Concerning rule (21), with our system there is no need for "Stress Adjustment", otherwise known as "Nuclear Stress", so only the Compound

*(21) COMPOUND, ~~(NUCLEAR STRESS, STRESS ADJUSTMENT)~~

$$V \rightarrow \overset{+\textbf{accent}}{[1\ \text{stress}]} \;/\; [\#\#X \underset{+\textbf{accent}}{\begin{bmatrix} \underline{\qquad} \\ 1\ \text{stress} \end{bmatrix}} Y \langle \#\#Z \rangle \#\#]_{\langle NAV \rangle}$$

where $\cancel{(Y \neq \ldots [1\ \text{stress}] \ldots)}$

$Z \neq \ldots \#\# \ldots$

Rule is kept (by deleting the angle brackets). Also the condition on Y is irrelevant for the Compound Rule (SPE: 92n) and can be deleted. The effect of this rule is that all accents after the first are dropped in compounds such as elevator boy, etc. (cf. (9) above).

Furthermore, the A (adjective) probably should be deleted too, so that only constructs labeled N(oun) or V(erb) would come under the influence of the Compound Rule. It is true that adjectival constructs are usually forestressed in attributive position before nominatives; but they have to be tested in predicative position, preferably sentence final, to determine whether this is their canonical pattern or the effect of rhythmic recession. It appears from (22) that compound adjectives are the exception, and "adjective phrases" -- i.e. (by analogy with noun phrases) those having double accent -- are the rule (cf. SPE: 91, Bolinger 1958, and Webster's NID3: 31a §6).

(22)

Compound Adjectives (forestress)	Adjective Phrases (double accent)
rose colored	hard boiled
mealy mouthed	plain spoken
fatuous sounding	straight forward
bitter sweet	over seas
smoke filled	up standing
	state-of-the-art
	down stream

5.5 For comparison, we now show our own Nuclear Accent Rule (23), which assigns our interim feature "plus intonation" to a vowel if it is accented and there are no later accented syllables in the sentence. Afterward a particular intonation pattern is assigned to this nuclear syllable. The

(23) **NUCLEAR ACCENT**

$$\mathbf{V} \rightarrow [+\textbf{intonation}] \;/\; \left[\begin{bmatrix} \underline{\qquad} \\ +\textbf{accent} \end{bmatrix} \mathbf{Y} \right] \textbf{sentence boundary}$$

$$\textbf{where } \mathbf{Y} \neq \ldots [+\textbf{accent}] \ldots$$

(24) **FALLING INTONATION**

$\begin{bmatrix}+\text{intonation}\end{bmatrix} \rightarrow \begin{bmatrix}+\text{cadence}\end{bmatrix}$ **in certain contexts (notably statements, WH- questions, and citation utterances)**

most important case is covered by our Falling Intonation Rule (24), which converts the cover feature "intonation" to [+cadence] in the informally stated contexts.[1] The order of these rules in the total grammar is as yet unknown.

5.6 Returning to the SPE stress rules, those remaining are non-cyclical word-level rules. The Second Fudging Rule (25) exists solely to precondition nouns derived from oxytonic verbs -- <u>relaxation</u>, <u>deportation</u>, <u>attestation</u>, and the like -- in such a way that the Secondary Auxiliary Reduction Rule (27), which is actually a strengthening rule, will apply to them in the correct way, as it does to <u>rhododendron</u> and the rest (SPE: 114-6). When reformulated with our features, rule (25) becomes vacuous, because by our convention (17) there can never be two accents

(25) **FUDGING—II** $\underset{\text{[2 stress]}}{+\text{accent}} \rightarrow \underset{\text{[3 stress]}}{-\text{accent}} \;/\; \_\_\, C_0 \underset{\text{[1 stress]}}{+\text{accent}}$

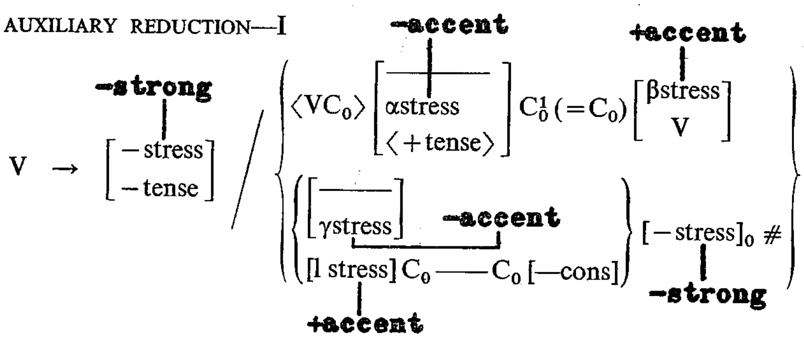
within a word at this stage. But it will turn out that rule (27) expressed in our feature system handles the relevant cases automatically and without any fudging. Our final revision of rule (25), therefore, is to expunge it from the grammar.

5.7 The First Auxiliary Reduction Rule -- the one that really reduces -- is shown with our emendations in (26). Here again the new features capture regularities which render the stress variables and their conditions pointless and unnecessary:

---

1 We are well aware that WH- questions often occur with an intonation in which the WH- word is high-pitched and the rest of the sentence has a low-rising contour. This pattern -- very common in the English dialect of Hawaii (Vanderslice and Pierson 1967: 161ff) -- we regard as indexically marked in General American English, occurring especially with accosting questions (cf. Bolinger 1948).

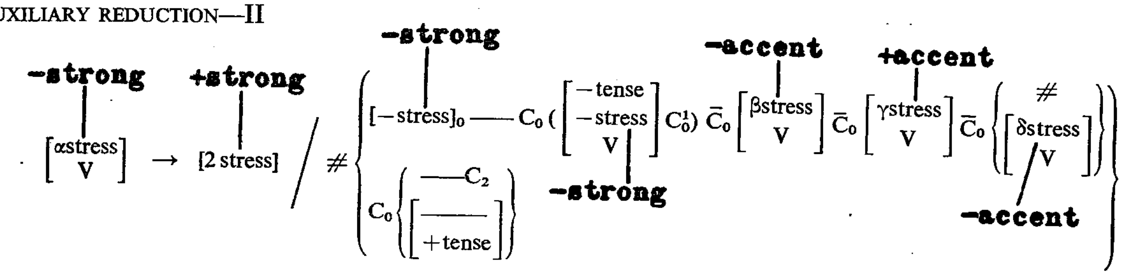
(26) AUXILIARY REDUCTION—I



~~Conditions: β=1, 2, 3~~
~~α is weaker than β,~~
~~γ is weaker than 2~~

5.8 There are two ways to rewrite the Second Auxiliary Reduction Rule in terms of our system. The conservative way is (27):

(27) AUXILIARY REDUCTION—II



where $\bar{C}$ is a consonant or a boundary

~~$\alpha \neq 1$~~
~~β is weaker than 2~~
~~δ is weaker than γ~~

This way assigns our closest equivalent of SPE's 3-1- stress pattern to the words exemplified in (28). That is, the last strong syllable (typically) being accented, this rule changes the correct syllable earlier in the word from weak to strong.

(28) **rhōdodéndron, Ōklahóma, Monōngahéla, Cōnestóga, Mūlligatáwny**

+s – +a – +s – +a – – +s – +a – +s – +a – +s – – +a –

The elegance of this solution, however, is illusory. By far the majority of words having this prosodic pattern exhibit, in the right surroundings, the accent recession phenomenon discussed above in connection with <u>telegraphic</u>, etc. (15, 16). Consider (29) and (30) -- the

(29) **Múlligatāwny stéw, Kálamazōo Cóllege, Wínnipesāukee wáterwāy, Monóngahēla Médical Cēnter, Rhódodēndron Róad**

(30) **to óverthrōw the góvernm(ē)nt, to únderstānd Énglish, to cómprehēnd the cálculus . . . rélāxātion óscillātor; áttēstátion**

last item of which is double-accented even in citation form for many speakers. Of course we agree with Chomsky and Halle that "it is ... unreasonable to expect cross dialectal identity on a minute point such as this" (SPE: 116n), but, all things considered, we think it preferable to assign double accentuation to this class of items at the word level. That is what the more radical revision of the Second Auxiliary Reduction Rule does. We separate rule (27) into an Auxiliary Accentuation Rule (31), which double-accents the cases just discussed, and an Auxiliary Strengthening Rule (32) to assign [+strong] to the first syllable in such words as <u>gestation</u>, <u>cantankerous</u>, <u>pontificate</u>, and <u>vocation</u> (cf. SPE: 118–21). It will be evident that each of these

(31) **AUXILIARY ACCENTUATION**

$$\begin{bmatrix}\text{-accent}\\ \text{V}\end{bmatrix} \rightarrow [\text{+accent}] \,/\, \#\, [\text{-strong}]_{0} \;\_\_\_\; C_0 \left(\begin{bmatrix}\text{-tense}\\ \text{-strong}\\ \text{V}\end{bmatrix} C_0^1 \right) \bar{C}_0 \begin{bmatrix}\text{-accent}\\ \text{V}\end{bmatrix} \bar{C}_0 \begin{bmatrix}\text{+accent}\\ \text{V}\end{bmatrix} \bar{C}_0 \left\{ \begin{matrix} \# \\ \begin{bmatrix}\text{-accent}\\ \text{V}\end{bmatrix} \end{matrix} \right\}$$

(32) **AUXILIARY STRENGTHENING**

$$\text{V} \rightarrow [\text{+strong}] \;/\; \#\; C_0 \left\{ \begin{matrix} \_\_\_\; C_2 \\ \begin{bmatrix}\_\_\_\\ \text{+tense}\end{bmatrix} \end{matrix} \right\}$$

new rules exploits one of the two main paths through the curly-bracketed environment in rule (27). It might be objected that there is a loss of generality in breaking the larger schema down into parts. But on examining into the matter we discover instead a serendipity.

It will be recalled that Chomsky and Halle say of the upper path in (27) that it is "strikingly similar to the rules of primary stress placement, particularly to condition (c) of the Main Stress Rule ... the central difference being that secondary stress is assigned rather than primary stress" (SPE: 114). Further, they say, this "is not a merit of this grammar but rather indicates a defect either in the analysis or in the underlying theory" (SPE: 115). In our feature system, this primary-secondary distinction has been eradicated, and it follows that rule (31) can be incorporated as another case of the Main Stress Rule (19) -- with a corresponding quantum leap in generality. We do not attempt to formulate this here, partly because our sanguine view of the prospects for collapsing (31) and (19) rests heavily on Chomsky and Halle's authority (cf. quotations above). As they appear in the Summary of Rules (SPE chapter five) the environments of the Main Stress Rule and the Second Auxiliary Reduction Rule do not really smite one with their similarity. Nonetheless we think the increase in similarity between these two rules as expressed in our terms furnishes powerful confirmation of the correctness of our approach.

As a first summarizing example of the way in which our rules work, consider the derivation of the stress pattern of aristocracy which is shown in (33). In the first cycle the stem [-kræt] is assigned the

(33) $[_N[_N\text{æristo}[_S\text{kræt}]_S]_N\text{y}]_N$

			+a	Main Stress (19), SPE case (eii).
	+a		+s	Main Stress (19), SPE case (ci).
	+s	+a	+s	Main Stress (19), SPE case (ai).
	+s	+a	-	Auxiliary Reduction I (26), SPE case (c).
+a	+s	+a	-	Auxiliary Accentuation (31).

feature [+accent] by the Main Stress Rule, and [+strong] by the convention given in (17b). Then in the second cycle the sequence [æristokræt] is considered; and by the equivalent of case (ci) of the Main Stress Rule as given in SPE, the last syllable is taken to be part of the context because it is stressed, the syllable before this is also taken to be part of the context because it contains a lax vowel followed by a weak cluster, and the preceding syllable is assigned [+accent]. By the convention in (17a) this reduces the [+accent, +strong] on the last syllable to

[-accent, +strong]. In the third cycle the sequence under consideration is the whole word. In this case we must apply the equivalent of case (ai) of the Main Stress Rule as in SPE. The last syllable is a lax suffix, so it is part of the context; the syllable before it has a lax vowel and a single final consonant, so it too belongs in the context; and the third syllable from the end becomes accented, and the syllable before it becomes deaccented. The First Auxiliary Reduction Rule (26) then further weakens the unaccented strong syllable which occurs before the weak (final) syllable. The Auxiliary Accentuation Rule (31) also applies, since the last accented syllable in the word is preceded by a strong unaccented syllable; accordingly the previous part of the word is accentuated in a way similar to that specified by the Main Stress Rule.

As may be seen in (34), the derivation of the stress pattern in <u>articulatory</u> is similar. But in this case the final line involves the

(34) $[_A[_V\text{artikulAt}]_V\text{Or+y}]_A$

		+a			Main Stress (19).
	+a	+s			Alternating Stress (20).
	+s	+s	+a		Main Stress (19).
	+a	+s	+s		Main Stress (19). Disregard accented syllable followed by no vowels; disregard preceding <u>At</u>; disregard preceding lax vowel (which becomes tense later) since it is followed by a single consonant. Accent preceding vowel.
	+a	-	+s		Auxiliary Reduction I (26), case (d).
+s	+a	-	+s		Auxiliary Strengthening (32).

Auxiliary Strengthening Rule (32), which strengthens the first vowel in a word if it is before a strong cluster. In both cases the correct intonations would be assigned by later rules.

Finally we may consider (35), which compares our derivation (on the left) with that of SPE (on the right) for the word <u>relaxation</u>: ours takes three steps, to SPE's five, because we have no need or use for the Second Fudging Rule nor for the Stress Adjustment Rule.

(35) $[_N[_V$ rElæks $]_V$ At+iVn $]_N$

+a		MSR (19)
+s	+a	MSR (19)
+a+s	+a	AUX ACCENT (31)

$[_N[_V$ rElæks $]_V$ At+iVn $]_N$

	1		MSR (19)
	2	1	MSR (19)
	3	1	FUDGING-II (25)
2	3	1	AUX RED-II (27)
3	4	1	STRESS ADJ (21)

6. We must emphasize that we do not wish to be considered as being in favor of the SPE stress rules as opposed to some revision of these rules such as that proposed by Ross (forthcoming). We merely wish to show that our features can be used with profit in the formulation of such rules. The rules we have discussed above will produce, for example, exactly the same prosodic difference between cóndēnsátion and cómpensátion as that of SPE. The treatment of this pair is often held up as exemplary of the stress cycle's "explanatory power", but actually it falls short of even observational adequacy. Granted that "there is well-known dialectal divergence ... with respect to phonetic minutiae of this sort" (SPE: 39n), no dialect of English known to us distinguishes condensation "process" from condensation "product" as the stress cycle predicts (see also SPE: 116n). In this regard Chomsky and Halle's description does not correctly relate to ANY reality -- perceptual or acoustic. We have not attempted to remedy this defect here, being content to illustrate the way in which their cycle can be advantageously reformulated with our features. We would claim simply that we have presented a binary suprasegmental feature system which accounts for English accentual and intonational distinctions in an intuitively satisfying way. We have shown that our features fit naturally and harmoniously into the stress cycle of *The Sound Pattern of English*; and we have sketched a means of closing a significant lacuna in the grammar of English with explanatory Nuclear Accent and Intonation rules.

References

Bierwisch, Manfred. 1968. Two critical problems in accent rules. *Journal of Linguistics* 4.173-78.

Bolinger, Dwight L. 1948. The intonation of accosting questions. *English Studies* 29.109-14.

-----. 1958. Stress and information. *American Speech* 33.5-20. (Reprinted in Bolinger 1965, 67-83.)

-----. 1965. *Forms of English: Accent, morpheme, order*; edited by Isamu Abe and Tetuya Kanekiyo. Cambridge: Harvard Univ. Press.

-----, and Louis J. Gerstman. 1957. Disjuncture as a cue to constructs. *Word* 13.246-55. (Reprinted in Bolinger 1965, 85-93.)

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English.* New York: Harper and Row.

Gunter, Richard. 1966. On the placement of accent in dialogue: A feature of context grammar. *Journal of Linguistics* 2.159-79.

Heidolph, Karl-Erich. 1966. Kontextbeziehungen zwischen Sätzen in einer generativen Grammatik. *Kybernetika* 2.274-81.

Jones, Daniel. 1956. *An English pronouncing dictionary*, 11th edition. London: Dent and Sons.

-----. 1964. *An outline of English phonetics*, 9th edition. Cambridge: Heffer. (First published 1918.)

Kenyon, J. S. and T. A. Knott. 1953. *A pronouncing dictionary of American English.* Springfield, Mass.: G. and C. Merriam.

Ladefoged, Peter. 1970. An alternative set of vowel shift rules. (*Working Papers in Phonetics*, this issue.)

-----. In press. *Preliminaries to linguistic phonetics.* Univ. of Chicago Press.

Ohala, John J. 1970. Aspects of the control and production of speech. (*Working Papers in Phonetics 15.*) Los Angeles: University of California [Phonetics Laboratory].

Ross, J. Forthcoming. A reanalysis of English word stress. Mimeo paper.

Trager, George L. and Henry Lee Smith, Jr. 1957. An outline of English structure, 7th printing. (*Studies in Linguistics: Occasional Papers 3.*) Washington: American Council of Learned Societies. (First published 1951.)

Vanderslice, Ralph. 1968a. Synthetic elocution: Considerations in automatic orthographic-to-phonetic conversion of English with special reference to prosodic features. (*Working Papers in Phonetics 8.*) Los Angeles: University of California [Phonetics Laboratory]. Also published, Ann Arbor: University Microfilms (No. 68-11887; cf. Dissertation Abstracts 28.588A-9A).

-----. 1968b. The prosodic component: Lacuna in transformational theory. (Report P-3874.) Santa Monica: The RAND Corporation.

Wang, William S-Y. 1962. Stress in English. *Language Learning* 12.69-77.

Webster's third new international dictionary. 1961. Springfield: G. and C. Merriam.

An Alternative Set of Vowel Shift Rules

Peter Ladefoged

A linguistically significant generalization is missed by not considering English vowels in terms of a single height feature with four possible values. The relevant underlying tense vowels are:

	ī	ē	ǣ	ɑ̄	ɔ̄	ō	ū
height	1	2	3	4	3	2	1
back	–	–	–	+	+	+	+

The status of oe and ɨ will be considered later. Note that there are three front vowels and four back vowels. (Incidentally, the argument that my analysis is unsound because languages do not have four vowel heights seems to me to be ridiculous. English *does* have four vowel heights, as my rules show. The same is true for Danish, and probably for several other languages.)

My first equivalent of the rule given by Chomsky and Halle (1968) in *The Sound Pattern of English* (henceforth SPE) IV (34), p. 187, is:

$$(1) \qquad \begin{bmatrix} +\ \text{tense} \\ \alpha\ \text{height} \\ \text{V} \end{bmatrix} \rightarrow [\alpha - 1]\ \text{height}$$

This has the effect of raising all the appropriate vowels by one notch. It is obviously either wrong or nonsensical if applied to the vowels /i, u/ which have height 1. But these are the vowels which undergo a backness shift after the vowel shift in SPE. If the rule for this is put before the vowel shift, so that there are no vowels of height 1 to which my (1) above would apply, there is no problem. Accordingly my equivalent of SPE IV 39, p. 189, is:

$$(2) \qquad \begin{bmatrix} 1\ \text{height} \\ \alpha\ \text{back} \\ \text{V} \end{bmatrix} \rightarrow \begin{bmatrix} 4\ \text{height} \\ -\alpha\ \text{back} \end{bmatrix} \ / \ \_\_\_ \begin{bmatrix} -\ \text{voc} \\ -\ \text{cons} \end{bmatrix}$$

This results in /iy/ becoming /ɑy/ and /uw/ becoming /aw/. Note that in my (IPA) transcription /a/ is a front vowel, which does not occur in underlying forms. (Actually I doubt the phonetic reality of the backness shift. It might be better to get the diphthongs in both *high* and *how* to start with nearer the same, central, low vowel. It seems that for many people there is a neutralisation of the backness feature for a low vowel before a glide. But this is irrelevant to the present comparison between my vowel shift rules and those in SPE.)

If we assume that it is appropriate to combine the backness shift with the vowel shift, then we can reformulate (1) and (2) as a pair of disjunctively ordered rules as in (3):

$$(3) \quad \begin{bmatrix} +\text{ tense} \\ \text{V} \end{bmatrix} \rightarrow \left\{ \begin{matrix} \begin{bmatrix} 4 \text{ height} \\ -\alpha \text{ back} \end{bmatrix} & / & \begin{bmatrix} \overline{1 \text{ height}} \\ \alpha \text{ back} \end{bmatrix} \\ [\beta\text{-}1 \text{ height}] & / & [\beta\ \overline{\text{height}}] \end{matrix} \right\}$$

The two rules (1) and (2), or the combined version (3), account for everything in SPE (34) p. 187 and (39) p. 189 in a phonologically explanatory way, and with much greater simplicity. These rules also let us put vowel shifts such as ɑ to ɔ (as in *laud, brawl*) in the regular framework, without going through any of the curious rounding adjustments which have to occur in SPE. As in SPE there will have to be a rule (= V 6, p. 239 in SPE) which will prevent the /ɑ/ in *Father* from changing. But in my case it will be

$$(4) \quad \bar{\text{ɑ}} \quad \rightarrow \quad -\text{ rule (3) in polysyllables}$$

where rule (3) is the vowel shift rule, rather than rounding adjustment. Again, this is a more explanatory statement of what is going on.

An alternative formulation would be to leave the backness adjustment in a separate rule, but to collapse the other parts of (1) and (2) as in (5). The environments in which this rule would apply may be assumed to be as given in SPE V (33), p. 243. This SPE rule (without the environments) is reproduced here as (6) for comparison.

$$(5) \quad \begin{bmatrix} \alpha \text{ height} \\ \text{V} \end{bmatrix} \quad \rightarrow \quad [\alpha - 1 \text{ height}]$$

condition 0 height = 4 height
(or convention?)

$$(6)\quad \begin{bmatrix}\gamma\text{ back}\\ \gamma\text{ round}\\ \text{V}\end{bmatrix} \rightarrow \left\{\begin{matrix}[-\alpha\text{ high}] & / & \begin{bmatrix}\alpha\ \overline{\text{high}}\\ -\text{ low}\end{bmatrix}\\ \\ [-\beta\text{ low}] & / & \begin{bmatrix}\beta\ \overline{\text{low}}\\ -\text{ high}\end{bmatrix}\end{matrix}\right\}$$

Note that my (5) is not restricted to vowels which are alike in their backness and roundness, and consequently some way will have to be found to prevent it from applying to some cases of /u/, and all cases of whatever is the underlying vowel in *boy*. Recall that Chomsky and Halle deal with the cases of /u/ by first unrounding them to /ɨ/, just so that the vowel shift does not apply, and then rerounding them later. It seems to me to be far more explanatory (less ad hoc) to mark them (probably at the time of y-insertion) - rule (5). I would also rather put the vowel in *boy* directly in the lexicon as /ɔ̄y/, and have a rule, which would be combinable with my rule (4) above, exempting /ɔ̄/ from the vowel shift when it occurs before /y/. I see no virture in trying to claim that English has no diphthongs in underlying forms. It seems ridiculous to force the vowel in *boy* to go through the vowel shift by elaborate Procrustean maneuvers, and far preferable to show that it is an exception within English phonology by having rules specifically for it.

It is difficult to evaluate my rule (3) in comparison with the SPE vowel shift rule, since my rule does more work (it accounts for the backness in addition to the height changes). In addition my rule (5) is not exactly comparable in that it does less work (it does not exclude vowels which differ in their backness and roundness). But note that if, for the sake of making a valid comparison, we add [γ back, γ round] to the left hand side of my (5), then my rule seems to be considerably more elegant and explanatory than the SPE rule. It would then name (counting the condition, but excluding the contexts which would be as in SPE) 6 values of 3 different features, in comparison with SPE's rule which (excluding the contexts) names 8 values of 4 different features. The condition, which implies that vowel heights are members of a cyclically ordered set, is also more reasonable and explanatory than some of the cumbersome conditions in SPE, such as those in V (20) (I), p. 240.

Similar vowel shift rules have been proposed by Contreras (1969), who also cites an earlier formulation of mine (Ladefoged 1967). In a discussion of the Contreras paper, Harris (1970) argues that rules involving relations such as higher than might have to be regarded as more costly, because they "are presumably rare, if they exist at all." This seems a curious argument considering the fact that this relationship occurs in nearly one third of the phonological rules in SPE (all those involving numerical stress levels). Anyone who accepts the notion that different levels of stress can be specified in phonological rules surely should have no difficulty in accepting the notion of vowel height as a multivalued feature. There are an enormous number of experiments and observations over the centuries showing that a speaker knows, as one of the phonetic facts of his language, that there is a continuum of vowel height.

All in all, there seems to be a clear case for dealing with English vowels in one of the ways suggested here. The contortions which Chomsky and Halle go through because they have binary features, only three possible vowel heights, and such curious views on rounding, seem to me to be quite extraordinary.

References

Chomsky, N. and M. Halle (1968), *The Sound Pattern of English*, Harper and Row, New York.

Contreras, H. (1969), "Simplicity, Descriptive Adequacy, and Binary Features," *Language* 45, 1-8.

Harris, J. W. (1970), "Sequences of Vowels in Spanish," *Linguistic Inquiry* 1, 128-34.

Ladefoged, P. (1967), "Linguistic Phonetics," *UCLA Working Papers in Phonetics* 6.

Acoustic Level and Vocal Effort as Cues for the Loudness of Speech

George D. Allen*

Abstract

A review of previous studies of speech loudness shows great variability in the derived psychophysical functions relating loudness to speech power level (SPL) and other physical and psychological measures, such as subglottal pressure (SGP) and vocal effort. This paper argues that, because of the complex nature of speech production and perception, traditional scaling procedures that yield exponential relations between loudness and some other measure should be replaced by multidimensional techniques. Two experiments on speech loudness scaling using partial correlation analysis demonstrate that loudness judgments depend upon both acoustic cues as measured by peak SPL, and vocal effort cues, as measured by peak SGP in one experiment and subjective effort in the other. The relative dependencies upon these cues are different for different listeners and at least somewhat resistant to distortion via signal attenuation and/or masking with white noise. These differences between listeners might be exploited in the search for relationships between speaking and listening.

INTRODUCTION

A number of investigators have studied the relationship between listeners' judgments of speech loudness and various physical and physiological measures of the speech. The results of these studies have varied widely, however, as the following brief review shows. The basic methodology for these studies was established by Stevens's (1955) publication of the sone scale, wherein the loudness of a pure tone is related linearly to its sound pressure level (SPL) to the 0.6

---

* Now at Dental Research Center, University of North Carolina, Chapel Hill, North Carolina, 27514.

power. That is, when listeners are asked to judge the loudness of a pure tone (e.g., 1000 Hz) presented to them at a variety of levels, the function relating the logarithms of their responses to the logarithms of the levels will be well approximated by a straight line with a slope of 0.6.

Pollack (1954) presented speech to listeners from a tape recording, whose playback level was varied by the listeners in order to halve and double the apparent loudness of the speech. The resulting loudness function was decidedly nonlinear in form, being steeper at low levels than at moderate or high levels, and the straight line that best approximated the majority of the data represented an exponent of 0.4. Pollack attributed the difference between his and Stevens's results to speech-constancy effects, analogous to size-constancy effects in vision.

Warren, *et al.* (1958) had listeners judge both the loudness and the apparent distance away of a 1000 Hz tone and of constant peak amplitude tape recorded speech and found that loudness was inversely proportional to apparent distance and directly proportional to SPL to approximately the 0.75 power. They concluded that loudness judgments are based on perceived distance and that the reason why the derived slope was not 1.0, as the distance theory would predict, was that reverberation cues in the listening situation conflicted with similar cues about the recording situation carried on the stimulus tape. In a later study, Warren (1962) had speakers produce sustained /ɑ/ vowels, /ʃ/ fricatives, and pitch-pipe tones first at a moderate autophonic level (AL) and then at a level perceived to be half as loud. The exponents relating AL to SPL for these sounds were approximately 0.85 for /ɑ/ and /ʃ/ and approximately 0.75 for the pitch-pipe tone. Warren concluded from the similarity of these exponents for AL to those for loudness that AL too is judged in terms of distance.

Lane, Catania and Stevens (1961) had speakers rate the AL of sustained /ɑ/ vowels and found AL to be related to SPL with an exponent of 1.1. When listeners rated the loudness of these very same vowels, however, the exponent relating loudness to SPL was 0.7. They related this difference in exponents to a different mode of loudness judgment between speaking and listening, and suggested that, although SPL is the only cue used by listeners in judging loudness, it plays a negligible role in autophonic judgments by speakers, who instead use proprioceptions of vocal effort. Lane (1962) later replicated the results for loudness in a more detailed study involving sustained /ɑ/ vowel stimuli that differed in spectral structure as well as AL and SPL. Again he found an exponent of 0.7 relating loudness to SPL, and again he found SPL to account for most of the variance of his listeners' judgments, although spectral structure also had some effect.

Ladefoged and McKinney (1963) disputed these results, claiming that sustained vowels are not necessarily speech, and showed that their listeners rated the loudness of recorded *words* to the peak SPL of these words with an exponent of 1.2. They further related the physical work of vocalization to the square of the speaker's subglottal pressure (SGP) and showed that their listeners' loudness judgments of a word were related to the word's peak SGP to the 1.9 power, close to the predicted 2.0. They therefore suggested that listeners do not use SPL in their judgments at all, but rather extract cues from the speech that tell them how much work, or effort, went into the production of the speech.

More recently, Moll and Peterson (1969) attempted to remove fundamental frequency cues from loudness judgments by having one speaker produce sustained /ɑ/ vowels over a wide range of levels but at just four fixed fundamental frequencies. The speaker's judgments of AL had exponents ranging from 0.62 to 0.84 for the four different fundamental frequency conditions, whereas listeners rated the loudness of these same utterances with exponents ranging from 0.52 to 0.72. Brandt, *et al.* (1969) had listeners rate both the loudness and effort with which speech had been produced for stimuli that were manipulated electronically to hold either SPL or AL constant while the other varied. The exponents of the derived scales ranged widely for the three conditions of presentation, but the exponent for loudness judgments of the normal speech was 1.12, while that for stimuli electronically varied in level was 0.92. Finally, Mendel, *et al.* (1969) had listeners rate the loudness of a variety of stimuli, both speech and non-speech, with the speech again both autophonically and electronically varied in level, and found the exponents of the derived scales to range from 0.86 to 0.96 for normal speech and from 0.74 to 0.84 for electronically varied speech.

The experiments on speech loudness cited above have contributed to our knowledge of speech production and perception, but they have failed to pinpoint the information that listeners use in their loudness judgments. In this paper I shall argue that the exponent of the derived psychophysical scale is an inappropriate measure for obtaining that information and that multivariate techniques should be used. Then I shall describe two small experiments on speech loudness that illustrate the simplicity of application of one kind of multivariate analysis, partial correlation analysis.

The Exponent Must Go

The relationship between physically measurable attributes of a variety of stimuli and perceptual evaluations of those stimuli by human and other animal observers have been reduced to an elegant algebra by S. S. Stevens and his students over the last three or four decades, an algebra expressing these relationships in terms of exponential functions $\Psi = \Phi^n$, where Ψ is the perceptual magnitude in psychological units, Φ is the physical magnitude of the stimulus, and n, the slope of the function when plotted in logarithmic coordinates, varies according to the particular perceptual mode being investigated. In recent years, however, this calculus has come under attack, and Poulton (1968), for example, has shown that the exponent, n, can vary substantially depending upon experimental conditions. An even more damning criticism of these exponents as measures, however, is that they do not explain the underlying perceptual processes. A science should relate to processes outside itself, thereby validating and finding validation in those external processes; such validations have been sought for the exponents of psychophysical scales of the sort we have been discussing (e.g., Stevens and Davis, 1936), but they have not been found. Lacking this external growth, the scaling technique has turned inward upon itself, and the validity of a scale is typically demonstrated by the internal consistency of cross-modality matching (Stevens, 1959).

Investigators of speech loudness have, to be sure, sought validation of their exponents in a variety of ways, and have often tried to use variations in the size of these exponents from one experimental condition to another as indicators of varying perceptual processes. Unfortunately, the measured variations have not obeyed reasonable laws, perhaps because uncontrolled differences in experimental conditions have produced unwanted variations of the sort Poulton (*op. cit.*) referred to. The exponent has proved too insensitive to measure perceptual variations.

But there is reason to believe that the exponents of speech loudness functions for speech could *never* elucidate the perceptual processes involved, the reason being that speech is multidimensional, whereas the exponent relates loudness to but a single physical or physiological dimension. There probably are many interrelated physical dimensions such as SPL, fundamental frequency, duration, and spectral quality to name the major candidates, that contribute to the listener's judgment of the loudness of a speech stimulus; these dimensions presumably contribute to the loudness of other kinds of stimuli as well. But if these dimensions are highly correlated, as is the case for speech, then no unidimensional analysis can be valid, since it will not allow for contributions to perception from the different dimensions.

The above argument can be made in more formal terms. Suppose we wish to relate loudness judgments (L) to SPL for a set of stimuli in which both SPL and effort (E) have been measured. Then $r_{L,SPL}$, the Pearson correlation between loudness and SPL, is a weighted sum of $r_{SPL,SPL}$, the correlation of SPL with itself ($r_{SPL,SPL} = 1$, by definition), and $r_{SPL,E}$, the correlation of SPL with effort; according to the following equation*:

$$r_{L,SPL} = w_{SPL} \times r_{SPL,SPL} + w_E \times r_{SPL,E}$$

$$= w_{SPL} + w_E \times r_{SPL,E} \; .$$

The weights, w_{SPL} and w_E, represent the degrees to which the listener uses SPL and effort, respectively, as cues for loudness and are formally the partial correlation coefficients between loudness and SPL, holding effort constant, and between loudness and effort, holding SPL constant. That is, if the listener uses only SPL for his loudness judgments and ignores effort cues, as Lane, *et al.* (1961) suggest, then $w_{SPL} = 1$, $w_E = 0$, hence $r_{L,SPL} = 1$; if Ladefoged and McKinney (1963) are correct, however, and the listener uses only vocal effort cues, then $w_{SPL} = 0$, $w_E = 1$, hence $r_{L,SPL} = r_{E,SPL}$.

These two cases appear clearcut enough but may in fact be difficult to distinguish in practice, for the following reasons. First, the actual correlations between loudness responses and SPL or effort (or any other measure) will always be less than unity because of the errors inherent in *any* measurement process, physical or mental. The greater the error variability is, the more will these correlations diminish (in absolute value) from their ideal maximum, and so we should not expect $r_{L,SPL}$ to be equal to 1, even if the listener were paying attention only to SPL for his judgments. Second, however, the correlation between SPL and effort ($r_{SPL,E}$) is high in many stimulus sets used in speech loudness experiments, representing as it does (roughly) the acoustic response of a fixed vocal tract to changes in SGP (Ladefoged and McKinney, 1963). For example, $r_{SPL,E} = 0.92$ for the Ladefoged and McKinney stimuli shown in Figure 1. We are therefore unlikely to distinguish between the case where $w_{SPL} = 1$ and that where $w_E = 1$ simply by looking at $r_{L,SPL}$, since that correlation will be slightly less than unity in both cases. In other words, so long as effort and intensity are highly correlated, one cannot determine from simple regression analysis which dimension is

---

* A more usual notation for this relationship would be $r_{L,SPL} = b_{L,SPL \cdot E} + b_{L,E \cdot SPL} \times r_{SPL,E}$ (Hays, 1963).

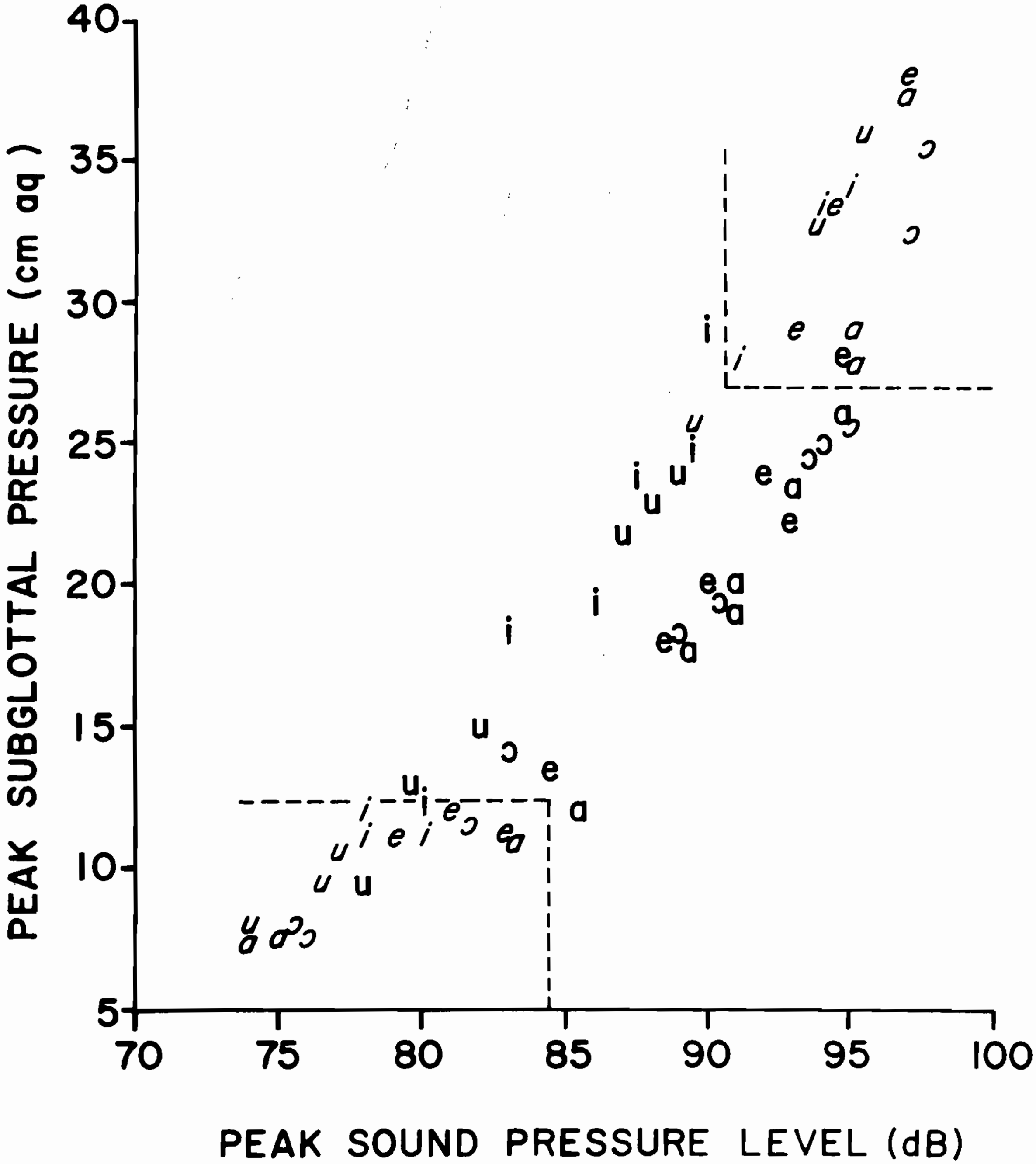


Figure 1. Peak subglottal pressure (cm aq) vs peak sound pressure level (dB) for the 60 stimuli of experiment 1. The dotted lines demarcate roughly the stimuli whose responses were not analyzed; these stimuli are noted in light fact italic. The 30 central stimuli, responses to which were analyzed, are noted in boldface roman.

"responsible" for the loudness ratings. Multidimensional methods must be used.*

To recapitulate the argument against the use of exponents to "explain" the loudness of speech: these exponents are derived from least squares regression analyses of loudness responses; such responses may in fact depend more upon variables not considered than upon the variables in the regression analysis; if the variables considered in the analysis are highly correlated with equally important but not considered variables, then the regression analysis will yield a high correlation with the measured variables, to which the loudness judgments will be erroneously related; only multivariate analysis will reveal the relative contributions of different dimensions to the loudness judgments.**

THE EXPERIMENTS

Two experiments were carried out in order to examine the relative importance of acoustic intensity and vocal effort as cues for speech loudness. These two variables were chosen because of their relative positions in the competing theoretical statements on speech loudness by

---

* Hursch, *et al.* (1964) have presented a theoretical model of perception which presupposes a multiplicity of interrelated cues such as I have described for speech loudness. Their model, based on Brunswik's (1943) notion that the probabilistic structure of the environment plays a crucial role in an organism's perception of that environment, expresses subjective estimates of a physical variable in terms of two multiple regression equations, one for the relationship between the physical variable and all the cues that can be used to estimate it, the other representing the relative importance attached to each of those cues by the organism in making its estimates. The reader is referred to Hursch, *et al.* (*op. cit.*) for an illuminating discussion of how the correlation between perception and reality can vary depending upon how effectively the perceiver uses information available to him.

** It should be recalled that Lane (1962) did in fact structure his stimuli variously along the different dimensions of AL, SPL, spectral quality, fundamental frequency, and duration; he presents only one multivariate analysis, however, in which the loudness of stimuli comprising four levels each (fully crossed design) of AL, SPL and spectral quality depended primarily, but not entirely, on SPL. Likewise, both Brandt, *et al.* (1969) and Mendel, *et al.* (1969) presented stimuli that varied in the relationship between SPL and AL, but neither study used multivariate analysis.

Lane (Lane, *et al.*, 1961; Lane, 1962) and by Ladefoged (Ladefoged and McKinney, 1963; Ladefoged, 1967).

Experiment 1

The first experiment was essentially a rerun of the Ladefoged and McKinney (1963) experiment, using the same stimuli, but with different subjects and a different analysis. A brief description of the materials and method will be given here, and the reader is referred to the original article for further details. "... A number of repetitions of the words 'bee', 'bay', 'bar', 'bore', and 'boo' were spoken by a British speaker in an anechoic chamber. A possible phonetic transcription of these utterances would be [bi:, be:, bɑ:, bɔ:, bu:]. The words were spoken in a natural way at a number of loudness levels with no constraints on pitch" (*op. cit.*, p. 454). From these utterances were selected twelve tokens of each of the five words, chosen to span a wide range of effort levels. These 60 words, plus five more that served as introductory material, were spliced into a tape recording of 65 repetitions of the constant frame utterance, "Compare the words: 'bar' and _____," with the stimulus word in the blank approximately 200 msec following the word "and." Subjects for the present experiment were the 14 members of a speech science class, all undergraduate juniors and seniors at U.C.L.A. They heard the tape played on a tape recorder (Ferrograph Series 5) and loudspeaker (Sony SSA 777) in a normal, i.e., not specially sound treated, classroom. They performed a standard magnitude estimation task, rating the loudness of each stimulus word relative to the constant "bar," to which they were to assign the standard value of 10. Subjects were not paid for their services.

Measurements of peak SPL (dB) and peak SGP (cm aq) had been made prior to this experiment, as described in Ladefoged and McKinney (1963). SPL was measured in true rms volts with a smoothing time constant of 0.23 sec; SGP values were obtained using an esophageal balloon. SPL values ranged from 74.0 to 97.5 dB ($\pm$0.5); SGP ranged from 7.5 to 38.0 cm aq ($\pm$0.5). These measures are shown together for the 60 stimulus words in Figure 1.

The logarithms of the subjects' responses were examined in order to note departures from assumptions underlying linear multivariate analysis, and some data were discarded on two different grounds. First, three of the 14 subjects used five or fewer response values for the 60 stimuli, and because of the imprecision of these subjects, their data were excluded *in toto*. The remaining 11 subjects' logged responses were subjected to polynomial regression analysis to check for linearity with respect to the two independent variables, and substantial quadratic and cubic components as compared to the residual error were found to be present in the data, as noted in Table I.

	Loudness vs SPL				Loudness vs SGP			
Subject	Linear	Quadratic	Cubic	Error	Linear	Quadratic	Cubic	Error
1	2625.	7.0	71.0	4.4	3650.	155.1	60.7	14.0
2	2548.	1.3	84.9	5.7	3701.	139.3	52.0	13.5
3	2507.	8.7	104.7	5.9	3505.	35.2	163.6	16.8
4	2496.	7.6	42.4	7.2	3666.	43.3	66.6	15.6
5	2594.	14.2	41.0	5.4	3891.	287.5	11.3	8.2
6	2463.	1.7	229.0	4.6	3466.	129.5	241.7	14.5
7	2534.	8.1	86.7	5.8	3816.	54.6	72.1	12.6
8	2351.	11.5	108.6	8.6	3380.	31.0	116.6	20.0
9	2236.	9.3	123.5	10.4	3335.	30.6	139.6	20.4
10	2628.	8.2	42.6	4.9	4149.	94.5	26.7	6.7
11	2293.	118.4	60.0	8.6	3379.	583.0	4.5	12.1

Table I. Variance components (squared log loudness units) derived from cubic polynomial regression analyses of 11 subjects' judgments of loudness against SPL and SGP, for all 60 stimuli in experiment 1.

Rather than apply a separate polynomial transformation to each subject's responses, a procedure that would guarantee linearity at the expense of obscuring some intersubject differences and of possibly introducing hetero-scedasticity (non-homogeneity of error variance from one end of the scale to the other), just the responses to the centrally located 30 of the 60 stimuli were analyzed for each subject; i.e., responses to the three most extreme stimuli were excluded for each word type at both ends of the scale, leaving only the responses to the central six stimuli for each of the five words. Extreme values were found on the basis of both SPL and SGP, which agreed as to which stimuli to exclude in all but three cases: one loud instance of "bay" was excluded whose vowel had a peak SPL 2 dB less than one that was retained, the peak SGP being 1 cm aq greater for the one that was excluded; one soft instance of "boo" showed disagreement between SPL and SGP by 1 dB and 1 cm aq, respectively, in the same direction as the "bay" example; and one loud instance of "bee" showed disagreement between SPL and SGP by 1 dB and 1 cm aq, but in the opposite direction from the "bay" example. The remaining 30 stimuli had SPL values ranging from 78.0 to 95.0 dB and SGP values ranging from 9.5 to 29.0 cm aq and are noted in heavy roman type in Figure 1.

Polynomial regression analysis showed the 11 subjects' responses to these 30 stimuli to be reasonably linear (i.e., with comparatively small quadratic and cubic terms), as shown in Table II, justifying partial correlation analysis of each subject's data. The results of these analyses are given in Table III.*

Table III shows that the subjects' responses were positively correlated with *both* SPL and SGP, indicating that both intensity and effort were used as relevant cues for the loudness judgments of these subjects. That is, for two stimuli that were equal in SPL the one with greater SGP would be heard as louder, and for two stimuli equal in SGP the one with greater SPL would be heard as louder.

It is difficult to conclude from the absolute sizes of these correlations whether SPL or SGP was the more important cue, on the average, since we may recall from earlier discussion that the absolute value of a correlation is reduced by measurement error, both physical and perceptual, and careful calibration of these errors was not undertaken in this experiment. Although the data do not allow us to make this kind of absolute judgment of the relative importance of SPL and SGP as cues for loudness, they do however indicate that listeners differ in their relative dependence on these cues. One would expect that some subjects

---

* Typical 95% confidence intervals under the assumption of normally distributed errors are .08 to .60 for an observed correlation of .30 and .16 to .73 for a correlation of .50.)

Subject	Loudness vs SPL Linear	Quadratic	Cubic	Error	Loudness vs SGP Linear	Quadratic	Cubic	Error
1	475.	5.2	16.2	5.4	534.	4.9	63.2	4.9
2	438.	14.3	11.7	6.6	482.	2.4	10.1	9.1
3	441.	38.1	0.1	6.0	459.	18.7	1.5	9.7
4	416.	13.1	3.9	7.8	540.	16.0	22.9	5.8
5	446.	0.0	7.3	7.0	562.	0.0	32.4	5.2
6	476.	22.7	18.0	4.6	476.	20.0	15.5	8.4
7	504.	15.5	2.3	4.4	590.	9.3	16.8	4.4
8	415.	19.1	10.9	7.3	399.	12.5	38.9	10.8
9	344.	31.0	1.3	10.0	386.	56.2	0.7	11.1
10	450.	14.0	0.8	6.6	636.	8.0	1.9	3.3
11	350.	29.6	0.3	9.8	415.	89.0	2.1	8.6

Table II. Variance components (squared log loudness units) derived from cubic polynomial regression analyses of 11 subjects' judgments of loudness against SPL and SGP for central 30 stimuli in experiment 1.

Subject	1	2	3	4	5	6	7	8	9	10	11
Loudness vs SPL	.55	.49	.53	.36	.43	.60	.62	.52	.36	.38	.33
Loudness vs SGP	.51	.41	.35	.59	.61	.35	.65	.23	.32	.79	.38

Table III. Partial correlations of 11 subjects' loudness judgments with SPL and SGP, for central 30 stimuli of experiment 1.

are better raters of loudness than others; i.e., they have less response error variance, and indeed the residual error for linear regression of loudness with respect to SPL (Table II above) correlates moderately across the 11 subjects with the corresponding residual error for SGP (Spearman r = 0.52). One would as a result expect that the accurate raters' partial correlations would in general be higher than the inaccurate raters', but in fact the rank order correlation between the partial correlations with respect to SPL and those with respect to SGP across the 11 subjects is approximately zero. That is, although the listeners do appear to differ in rating ability, the relative correspondence of loudness with SPL and SGP does not follow this rating ability. This lack of correlation, where one might have been expected, indicates that SPL and SGP have different relevance to different listeners as cues for loudness, so that one subject used SPL more heavily in his ratings of loudness, while another used SGP more heavily. The implication that different people have different perceptual "strategies" for decoding the loudness of speech, and the attempt to give greater substance to this notion of the strategy for speech perception, motivated the second experiment.

It is useful to look at the pattern of loudness judgments in experiment 1 to determine the source of the positive correlations with SPL and SGP. Each correlation is the average of a listener's dependence on SPL or SGP across all levels of loudness, whereas this dependence might well be different for soft and loud speech. Figures 2 and 3 show the average loudness of each stimulus plotted against its peak SPL and peak SGP, respectively. Figure 2 shows that for fixed SPL the close vowels, /i/ and /u/ , which were produced with greater SGP (cf. Figure 1), were judged louder than the open vowels, /ɑ/ and /ɔ/ ; Figure 3 shows that for fixed SGP, the open vowels, which have greater SPL, were judged louder than the close vowels. These relationships hold quite uniformly across all levels of loudness and are the sources of the positive partial correlations noted in Table III.

A slight difference between Figures 2 and 3, whose statistical reliability cannot be determined from the present data, is the pattern of variability of response in the loudness range from 1.00 to 1.25 (i.e., slightly louder than the standard). In Figure 2 the variability of response increases in this range with respect to the rest of the scale, whereas in Figure 3 the variability decreases in this range, suggesting that listeners may consider effort cues as the more important information at normal to slightly loud levels and SPL as more important at soft and very loud levels.

Experiment 2

The second experiment was an attempt to manipulate the listener's dependence on one or another cues in the speech signal for his judgments

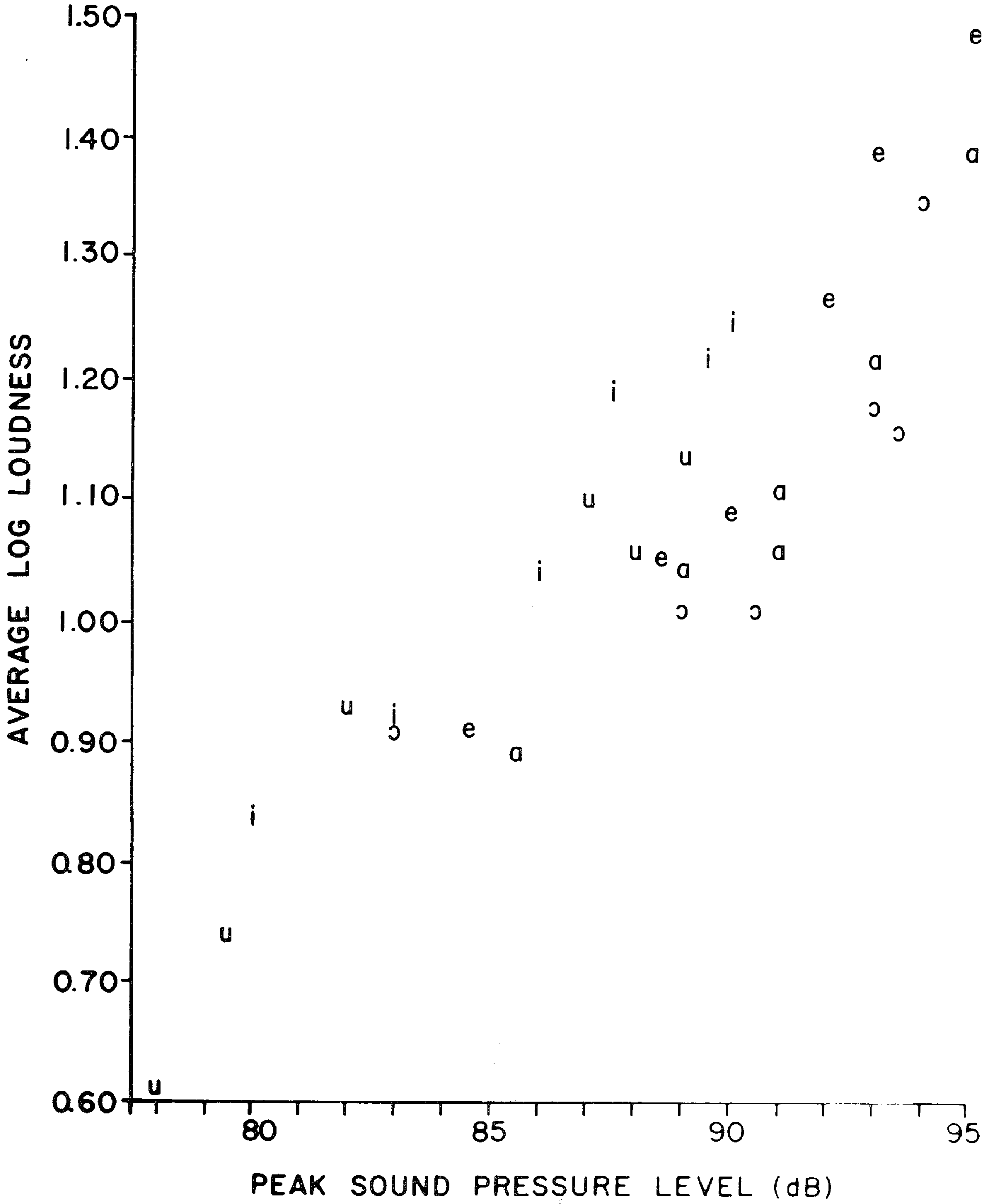


Figure 2. Average log loudness vs peak sound pressure level (dB) for the 11 listeners' responses to the 30 central stimuli in experiment 1.

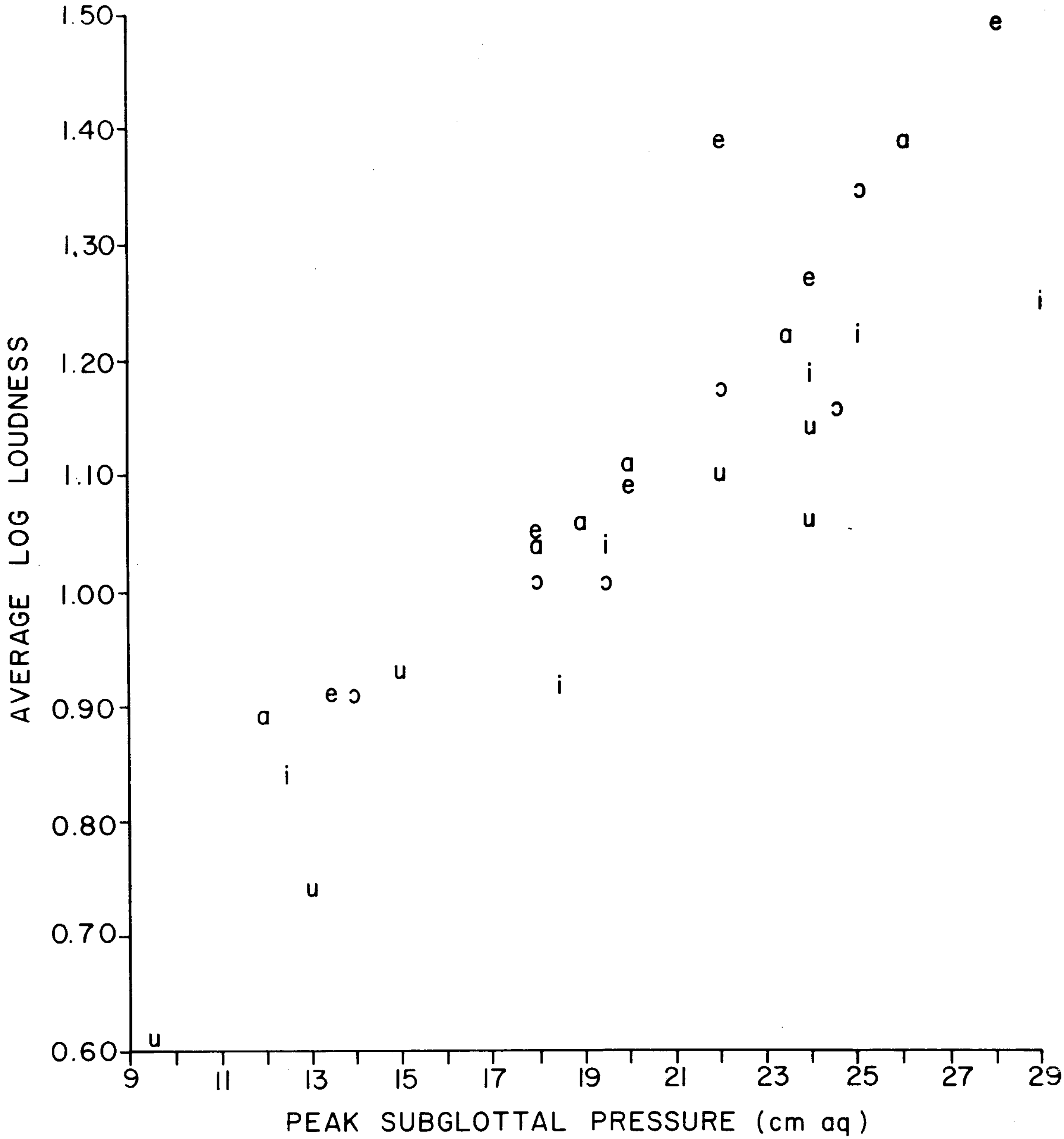


Figure 3. Average log loudness vs peak subglottal pressure (cm aq) for the 11 listeners' responses to the 30 central stimuli in experiment 1.

of loudness. If the listener uses both intensity and effort cues, and if his strategy is changeable, then masking or otherwise disguising one of the cues might increase his dependence on the others. Speech stimuli were presented under a variety of conditions, therefore, to see whether in fact the partial correlations of listeners' loudness judgments with either SPL or effort waxed and waned in a systematic fashion.

A tape recording was constructed of the words "see," "say," "so," and "sue," uttered by one speaker (the experimenter) at 13 different levels of vocal effort, and spliced onto the ends of 52 copies of the constant frame sentence, "Ready? How loud is this word? _________," uttered with average effort. SGP was not measured, and so effort was controlled subjectively at each of the 13 levels in the following manner: the speaker uttered the four words seven times each in a quasi-random order, about one word per second, inhaling anew for each word. The 28 words were uttered with subjectively constant effort, and the list was preceded and followed by three extra instances of "see" to negate end effects of list intonation. During this reading, the speaker heard approximately 110 dB white noise (100 to 20,000 Hz) in both ears to mask sidetone feedback, which had been found to cause smaller differences in SPL among the four stimulus words than is predicted by Peterson and Lehiste (1959).* From the resulting seven instances of each word, the one of median peak SPL was selected for use in the stimulus tape. The four words chosen from a given list were assumed to have the same level of effort, which was arbitrarily assigned a numerical value equal to the mean peak SPL of the four words. This procedure insured that SPL and effort would be linearly related, an assumption underlying the intended correlation analysis. The peak SPLs, as measured on a true rms voltmeter (Ballantine , smoothing time constant = 0.02 sec), and the derived effort levels of the fifty-two stimuli are given in Figure 4. Five additional stimuli, spanning a wide range of effort levels, were included as an introduction to the stimulus tape; responses to these five stimuli were not included in the analysis. Six normal hearing subjects served in the experiment without pay; they comprised four naive listeners and two trained phoneticians acquainted with the experiment and its goals. They heard the stimulus tape under 10 different conditions of presentation, to be described below, and rated the loudness of each stimulus word relative to the frame sentence according to a nine point category scale.** Subjects heard the frame sentence, then the stimulus word approximately 1/4 second later, and then had approximately four seconds in which to respond by pressing one of nine buttons labeled

---

* This feedback phenomenon will be discussed further below.

** I am indebted to Dr. Allen Parducci, Department of Psychology, U.C.L.A., for much help in deciding matters of scaling procedure.

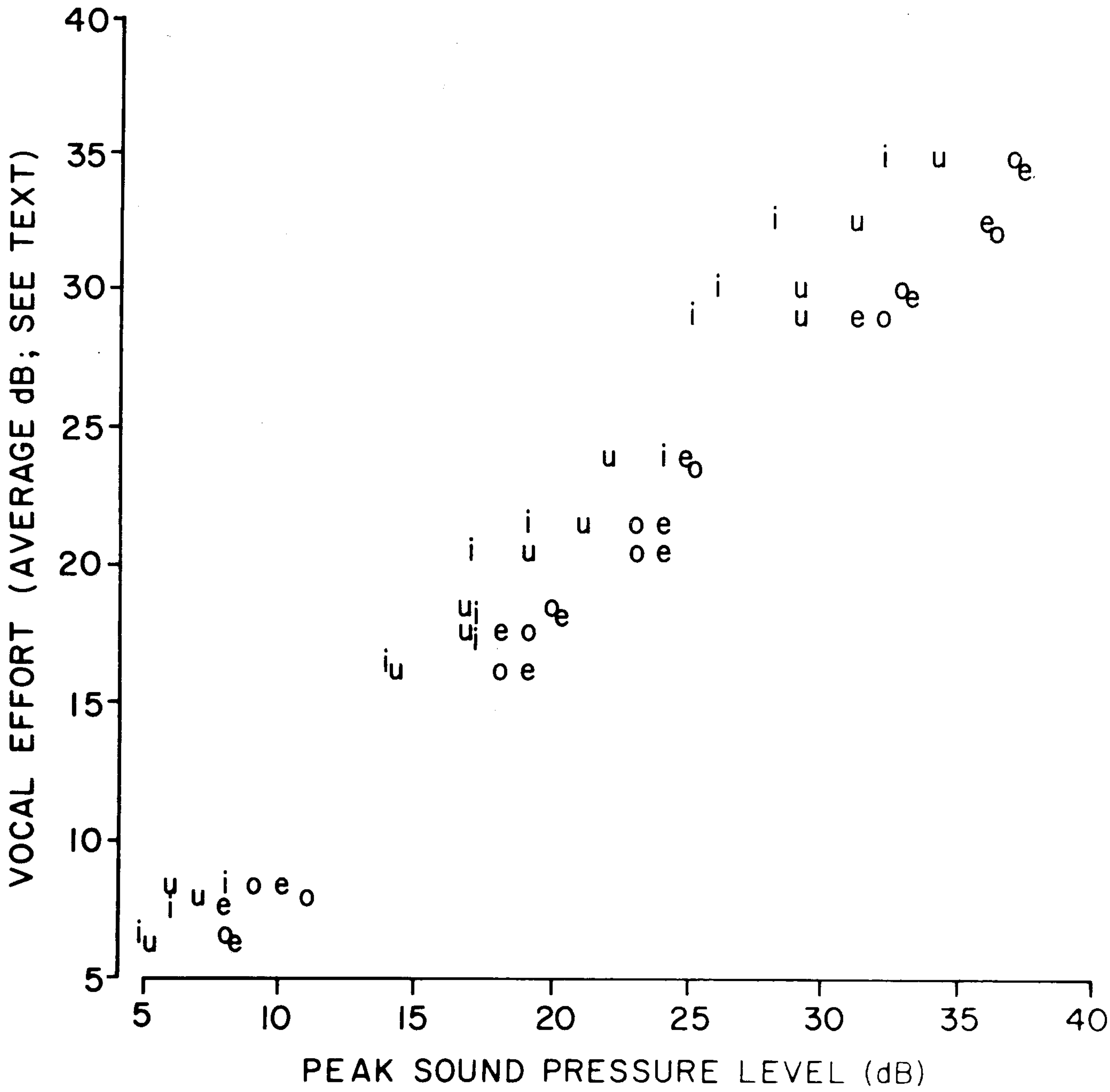


Figure 4. Vocal effort (in average dB; see text for explanation) vs peak sound pressure level (dB) for the 52 stimuli of experiment 2.

"very very soft," "very soft," "soft," "a little soft," "medium," "a little loud," "loud," "very loud," and "very very loud." The use of categories implied that responses could be gathered automatically by computer; the particular choice of category labels gave the additional benefit of a nearly linear relation between responses and both effort and intensity, so that all the data could be analyzed without transformation. The price for these benefits was an increase in response variance over the standard magnitude estimation procedure (Garner, 1962).

The ten conditions of presentation were divided conceptually into condition zero (C0) and conditions one through nine (C1-C9). C0 involved no distortion of the stimulus tape and was given first to each subject as an introduction to the loudness rating task and for comparison with earlier studies, in which there was also no distortion. C1 through C9 involved combinations of three degrees of attenuation with three levels of masking noise, in an effort to mask either intensity or effort cues. It was hoped that attenuation might reduce listeners' dependence on SPL by making all the signals less intense and therefore harder to discriminate (Licklider, 1951) while leaving the relative levels within stimuli, a possible cue for effort (Brandt, *et al.*, 1969), unchanged. Masking with white noise, however, might destroy information about the relative levels of frequency components within the stimulus while preserving overall intensity relationships, thus obscuring effort cues and making SPL cues more relevant.

The tape was played from a tape recorder (Ferrograph Series 5), the output of which was mixed resistively with the output of a white noise generator (General Radio Type 1390-B, 100 to 20,000 Hz). The fixed level white noise was resistively attenuated by any of three amounts (0. -7, or -15 dB) before mixing with the signal, and the resulting signal-plus-noise was attenuated by any of three amounts (0, -6, or -12 dB) before being sent to the headset (Suprex Model ST-M) worn by the subject. It is hard to specify meaningfully the actual levels and signal-to-noise ratios, since the recorded stimuli varied over a 35 dB range, but the peak signal-plus-noise level of the stressed vowels in the frame utterance was about 80 dB above threshold in the condition of -7 dB noise and 0 dB attenuation, and the weakest stimuli were just perceivable in the condition of maximum noise (0 dB) and maximum attenuation (-12 dB). The nine conditions were completely intermixed over nine playings of the tape to a subject, according to quasi-random orders prestored in the computer.

A given stimulus trial was as follows: at the word "Ready" of the frame sentence the experimenter signaled the computer to set the appropriate relays to attenuate the masking noise and signal-plus-noise according to the condition of presentation of the present stimulus; the

Noise Condition (dB)		Subject 1 Gain (dB) 0	-6	-12	2 Gain 0	-6	-12	3 Gain 0	-6	-12	4 Gain 0	-6	-12	5 Gain 0	-6	-12	6 Gain 0	-6	-12
-15	L vs SPL	.06	.27	.36	.33	.24	.30	.05	.21	.04	.37	.31	.06	.14	.33	.24	.36	.17	.27
	L vs E	.51	.31	.27	.32	.38	.14	.60	.37	.55	.19	.26	.41	.41	.27	.35	.36	.43	.20
- 7	L vs SPL	.26	.46	.35	.30	.09	.14	.12	.38	.26	.25	.16	.36	.19	.12	.17	.42	.32	.36
	L vs E	.47	.08	.35	.26	.43	.29	.41	.20	.43	.30	.62	.19	.43	.43	.35	.23	.36	.14
0	L vs SPL	.44	.27	.40	.20	.26	.32	.05	.10	.20	.26	.09	.26	.14	.28	.06	.43	.35	.39
	L vs E	.31	.44	.13	.37	.25	.26	.53	.41	.28	.30	.59	.20	.52	.34	.48	.34	.35	.13
Condition 0	L vs SPL		.43			.17			.13			.33			.34			.37	
	L vs E		.20			.42			.43			.34			.40			.33	

(a)

	Subject 1	2	3	4	5	6
L vs SPL	.33	.15	.21	.20	.18	.30
L vs E	.29	.39	.28	.31	.36	.24

(b)

Noise Condition (dB)		Gain (dB) 0	-6	-12
-15	L vs SPL	.23	.21	.22
	L vs E	.23	.30	.30
- 7	L vs SPL	.18	.22	.18
	L vs E	.28	.30	.35
0	L vs SPL	.19	.22	.25
	L vs E	.35	.35	.35
Condition 0	L vs SPL		.24	
	L vs E		.30	

(c)

Table IV. Partial correlations of loudness judgments (L) with SPL and effort (E)
(a) for six subjects and all presentation conditions of experiment 2;
(b) for six subjects grouped over conditions; and
(c) for all conditions grouped over subjects.

rest of the frame sentence and the stimulus word were then played to the subject under that condition, after which the experimenter signaled the computer to reset the relays to a neutral condition (no signal, low noise) while the subject responded. Succeeding stimulus trials were therefore under different conditions, and responses were stored automatically.

The six subjects' loudness ratings of the 52 stimuli, after being checked for linearity, were subjected to partial correlation analysis, with results as shown in Table IVa. The subjects' loudness judgments correlate moderately with both intensity and effort, in different amounts for different subjects (Table IVb), but there is no clear trend toward increase or decrease of correlation with respect to either of these variables with change in condition of presentation (Table IVc). The correlations are generally smaller than in the first experiment, presumably because of the greater response error variance in the category rating procedure, and we may conclude that, within the precision of this experiment, subjects appear to use both intensity and effort cues in a way that is not obviously manipulable.

The average loudness responses, across subjects and conditions of stimulus presentation, are plotted against SPL and vocal effort in Figures 5 and 6. In agreement with Figure 2, Figure 5 shows that the close vowels are rated louder, on the average, than open vowels of equal SPL; Figure 6 likewise agrees with Figure 3, showing that open vowels are rated louder than close vowels produced with equal effort. The pattern of response variability across all loudness levels appears rather stable with respect to both SPL and effort, except perhaps that soft sounds show slightly reduced variability with respect to effort; this pattern disagrees with that of experiment 1, indicating that we should not place much weight on the interpretation of variability until the different measurement errors have been calibrated in detail.

DISCUSSION

Methodological Considerations

Two methodological points that need further clarification are (1) the kinds and degrees of errors in these and other studies of speech loudness and their effects on the conclusions of the studies, and (2) the kinds of multivariate procedures available for speech loudness research and their relative merits.

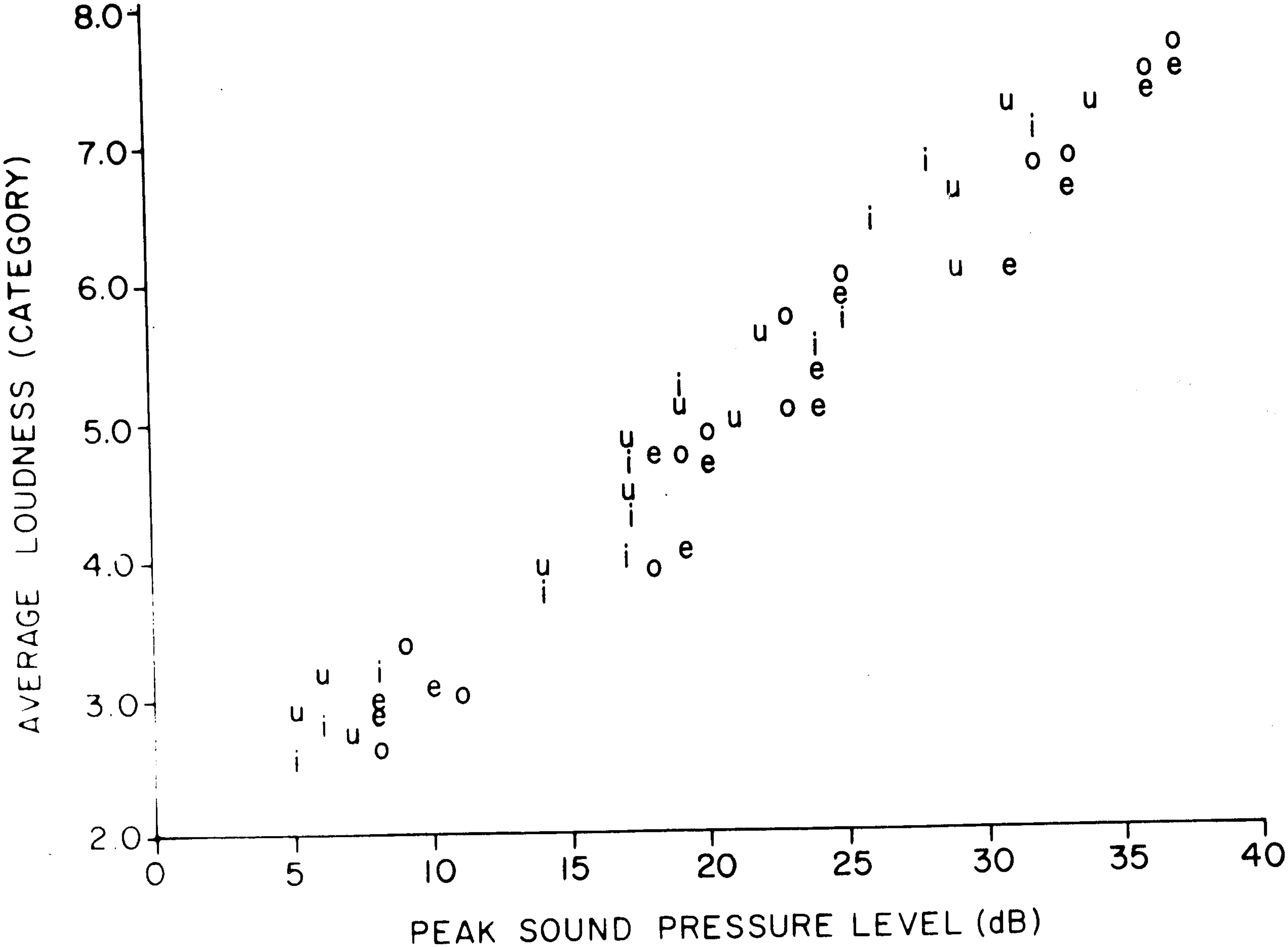


Figure 5. Average loudness (category) vs peak sound pressure level (dB) for six listeners' responses to the 52 stimuli of experiment 2.

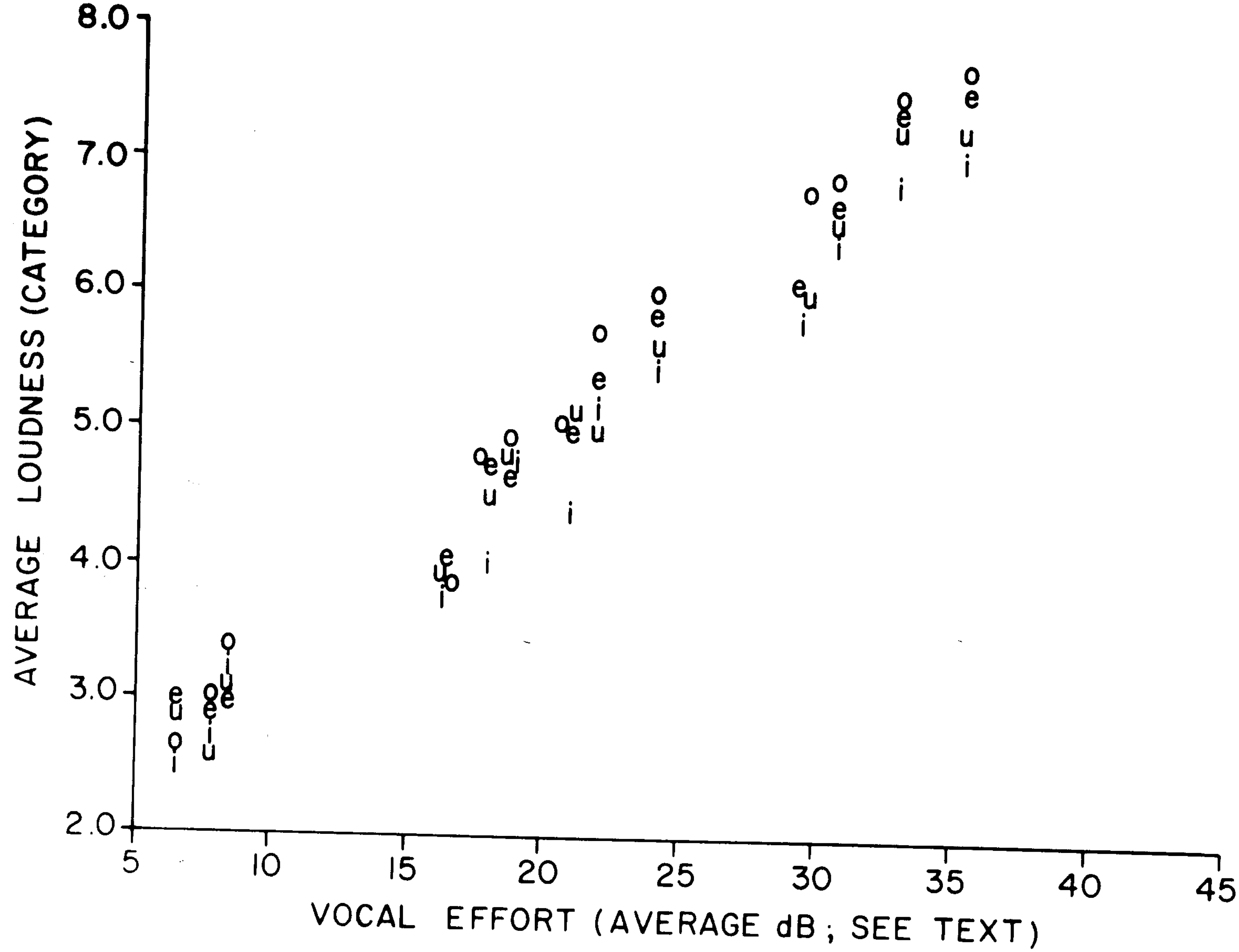


Figure 6. Average loudness (category) vs vocal effort (average dB; see text for explanation) for six listeners' responses to the 52 stimuli of experiment 2.

		Subject										
		1	2	3	4	5	6	7	8	9	10	11
Linear regression	N = 60	0.86	1.39	0.82	0.85	1.30	1.22	1.16	0.66	0.94	0.72	0.77
	N = 30	0.88	1.62	0.94	1.01	1.33	1.16	1.37	0.80	1.16	0.84	1.06
Cubic regression	N = 60	0.65	1.01	0.58	0.66	1.02	0.81	0.86	0.48	0.65	0.60	0.58
	N = 30	0.73	1.25	0.89	0.91	1.15	0.92	1.27	0.69	1.21	0.93	0.73

Table V. Exponents, n, of loudness functions $L = SPL^n$ derived from linear and cubic regression of 11 subjects' responses to all (N = 60) and centrally located (N = 30) stimuli in experiment 1.

Correlation coefficients may be viewed as ratios of variances*, with total experimental variance in the denominator and the variance accounted for or "explained" by the independent variable in the numerator (Hays, 1963). No independent variable of interest will ever be able to account for error variance, and so the correlations we observe in perception experiments are always less than 1. For example, human listeners make errors in their judgments of stimuli, and although they may be trained to reduce the variability of their judgments, small fluctuations will still exist from one identical situation to the next, and even larger variations will obtain between different observers. All physical measurements also contain errors, and although some measurements contain less error than others, an experimenter must always search for the most reliable of the valid techniques available. Since in a partial correlation analysis the perceptual errors are confounded with measurement errors, and since the size of the correlation coefficient will depend on the magnitude of this error variance, measurement error must be calibrated in order to make complete sense of the correlations.** In other words, even if a listener could make perfect estimates of the loudness of a set of speech samples, if SPL were measured by unreliable instruments the correlation between loudness and SPL might turn out to be quite low.

In experiment 1, the errors associated with SPL and SGP were both given as accurate to within ± 0.5 units, suggesting that the error variance was the same for both measures and uniform within each measure. However, the situation is not quite that simple. The measurement of SPL involved at least two arbitrary decisions: namely, what voltage measure and what integration time constant to use. Peterson and McKinney (1961) have shown that there is a rather close agreement, after integration, between full wave rectification and true rms voltage measurement of speech power. There are systematic variations between the two measures, however, depending on the phase relationships of the spectral components within each vowel; that is, rectification is subject to phase distortion, whereas true rms is not, and so an /u/ vowel, whose principal spectral components are closer together in phase, will have a rectified power 0.6 dB less than true rms, while an /i/ vowel, with spectral components far apart in phase, will have rectified power 1.0 dB less than true rms. The range of these differences between rectified and true rms voltage measures is of the order of ± 0.2 dB, which might not seem to be great, but is in fact substantial when viewed in terms of partial correlation analysis.

---

* More precisely, they are ratios of standard deviations.

** The situation is different in an analysis of variance with repeated measures, for example, wherein measurement error variance and perceptual error variance may be estimated separately.

In simple linear regression, the ±0.2 dB would be about 1 percent of the (usually) 35 to 40 dB stimulus range; in the partial correlation analysis used in experiment 1, however, the other variable of interest (SGP) was held constant, and so the range of variation of SPL was greatly reduced, to approximately 5 dB. Thus the ±0.2 dB variation was 8 percent of the range. Even worse, from the point of view of validity, the variations are not independent of other variables of interest but rather are highly related to spectral quality.

Peterson and McKinney (1961) also suggest an integration time constant of 0.02 sec or approximately two glottal cycles, as being valid for speech power measurements. Ladefoged and McKinney (1963) found, however, that such a time constant led to multiple peaks in the speech power while their SGP measures had but a single peak. Rather than have to decide which was the "true" peak, they therefore chose a constant of 0.23 sec, more than ten times as great. In measuring the SPL of the stimuli of experiment 2, however, the rate of increase to peak SPL for an /o/ vowel in the word "so" was often found to be great enough that a 0.23 sec time constant significantly lowered the measured SPL; that is, the jaw opened and began to close so quickly that the slow integration time "smoothed" the peak SPL. The higher vowels in "see" and "sue," however, showed no such SPL distortion for a time constant as long as 0.23 sec. Hence, since the measured difference in peak SPL for /o/ vowels with short and long time constants was a huge relative error, of the order of 2 dB, and this difference is again correlated with spectral quality, the smaller time constant of 0.02 sec was chosen for the SPL measurements of experiment 2.

Partial correlation analysis, though severe in its demands for highly calibrated measurement error, also suggests a method for deciding which of various physical measures is the most valid. If listeners actually judge loudness in terms of SPL, for example, then the best measure of SPL (i.e., the best choice of voltage measure and smoothing time constant) is the one that leaves the least residue of error variance beyond perceptual error in the judgment. All other measures will have greater residual errors and hence lower correlations between loudness and SPL. One then need only correlate loudness judgments with various measures of SPL to see whether one measure consistently outranks the others in terms of size of correlation.* Such an

---

* Included in the set of such measures to be tested should be "speech spurt averages" (Bricker, 1965), "equivalent peak level" (Brady, 1968), and any other automated speech level measures in order to determine how much information relevant to perception is lost by using such measures. Although Nature abhors simple solutions, an automatable speech level measure might turn out to be perceptually as valid as any other.

experiment should probably use multiple ratings by trained observers and a continuous response scale in order to reduce perceptual error to a minimum and thereby highlight the measurement error. (The data of the present experiments were deemed too crude for such an exact analysis.)

The SGP measures of experiment 1 illustrate yet another problem in estimating error variance, namely non-homogeneity. A partial correlation analysis in effect averages the errors for each variable over its entire range of values, and so the errors should be of comparable size across the range for this average to have any meaning.* The values of SGP given in Ladefoged and McKinney (1963) are presumed to be accurate to ± 0.5 cm aq, yet the reliability of the recording system may have been greater in the middle of the range of pressures than at the extremes, especially the lower end (Ladefoged, personal communication). Therefore a partial correlation analysis of all of the data of experiment 1 probably would not have been justified, and even though the centrally located 30 stimuli may have had homogeneous errors in SGP measurement, the necessary error data were unavailable, and so no test for homogeneity was carried out.

Partial correlation analysis is but one of many multivariate techniques available for speech loudness data analysis and was chosen because SPL and SGP, the two independent variables, were highly correlated in the original stimulus set (experiment 1). Analysis of variance is another widely used multivariate technique, and I have noted that Lane (1962) constructed a stimulus set appropriate for this type of analysis. More exciting than either partial correlation analysis or analysis of variance, however, are the more "creative" techniques developed by Kruskal (cf. Pols, *et al.*, 1969) for non-metric data and by Carroll and Chang (1970) and Harshman (unpublished manuscript) for metric data. These techniques presuppose far less about the structure of the data than the more traditional methods and are therefore capable of yielding new information about speech perception. Harshman's technique, for example, can show an exact representation in acoustic space of the three physiological dimensions many phoneticians "know" to be important for vowel description, namely high/low, front/back, and round/spread. Investigators of speech loudness, and of speech perception in general, should avail themselves of these techniques.

---

* This averaging of error variance is analogous to the deriving of within cells (error) variance in analysis of variance. A test for homogeneity of within cells variance is required prior to an analysis of variance, and so should a similar test be required for partial correlation analysis.

Theoretical Considerations

This paper is primarily methodological in intent, but there are some theoretical results worthy of comment. The most important of these relate to the perceptual strategies used to judge loudness by the listeners of these experiments. Proponents of the so-called motor theory of speech perception (Liberman, *et al.*, 1967) have argued that the listener decodes speech by noting in some way the actions he would have had to go through to produce a sound like that which he is hearing. Such a totally internalized process might be quite insensitive to external distortion of the signal, and the results of experiment 2, in which listeners apparently did not come to rely to a different extent upon SPL or effort under different listening conditions, argue for such an internalized process.

We must be cautious in our conclusions about perceptual strategy, however, for three reasons. The first is that negative results (i.e., no change under different conditions) are never strong evidence for anything, and we might be able to produce changes in perceptual strategy under different conditions where, for example, the relevance of the various cues remained steady at different levels for long periods of time rather than jumping around from trial to trial. Second, even if we were able to demonstrate stability of perceptual strategy over a wide variety of conditions, we would still not be able to conclude that this strategy was special for speech, since all of the stimuli in the present experiments were speech. We would have to show the pattern of responses to *the same stimuli* to be different depending upon whether the stimuli are part of a set of speech stimuli or whether they are part of some broader set of noises. Third, the pattern of individual differences in dependence upon cues in these experiments shows that each listener has his own strategy; if this strategy relates to speech production, as the motor theory suggests, then listeners should show analogous differences in their speech. That is, speakers continuously monitor their output, presumably via both auditory feedback and kinesthetic proprioception, and if speaking and listening are similar processes then the listeners who depend more on SPL for loudness judgments should be predicted to use auditory information to a greater extent in monitoring their own speech. In an experiment along these lines, the degree of disruption of speech under conditions of delayed auditory feedback could perhaps be used to measure dependence on auditory information, whereas degree of disruption under sensory nerve block might measure kinesthetic dependency.

Another theoretically important matter that has perplexed the present writer is the nearly instantaneous effect of side-tone feedback on the relative levels of different vowels produced with "the same" vocal effort. I produced a set of repetitions of the words "see," "say," "so," and "sue" for use in experiment 2 *without* white noise in my ears to mask side-tone feedback and found, upon measuring the peak SPL of the vowels, no difference in the median levels of the different vowels. A somewhat similar effect may be noted in the data of Lehiste and Peterson (1958), in which the speaker uttered sustained vowels both with and without 130 dB masking noise in his ears. Although the /o/ vowel remained greater in level than /i/ , /e/ , and /u/ in both conditions, the range of average levels for the latter three vowels was 1.0 dB without noise and 3.2 dB with noise.

It is not unusual to discover that a speaker will alter his vocal level on the basis of information derived from side-tone feedback; but it is hard to imagine a perceptual process that can act so fast that the peak SPL of the medial vowel of a stressed monosyllable is modified by information fed back from initial segments of the same vowel. The solution of this puzzle is left as an exercise for the reader.

References

Brady, Paul T. (1968) "Equivalent Peak Level: A Threshold-Independent Speech-Level Measure," J. Acoust. Soc. Am. 44, 695-699.

Brandt, John F., Ruder, Kenneth F., and Shipp, Thomas Jr. (1969) "Vocal Loudness and Effort in Continuous Speech," J. Acoust. Soc. Am. 46, 1543-1548.

Bricker, Peter D. (1965) "Technique for Objective Measurement of Speech Levels," J. Acoust. Soc. Am. 38, 361-362.

Brunswik, E. (1943) "Organismic Achievement and Enviromental Probability," Psychol. Rev. 50, 255-272.

Carroll, J. Douglas and Chang, Jih-Jie (1970) "Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of 'Eckart-Young Decomposition'," Psychometrika 35, 283-319.

Garner, Wendell R. (1962) Uncertainty and Structure as Psychological Concepts (John Wiley and Sons, Inc., New York).

Harshman, Richard A. (1970) "Models and Conditions for Unique 'Explanatory' Solutions with a Type of Three-Mode Factor Analysis," unpublished manuscript, Department of Linguistics, University of California at Los Angeles.

Hays, William L. (1963) Statistics for Psychologists (Holt, Rinehart and Winston, New York).

Hursch, Carolyn J., Hammond, Kenneth R., and Hursch, Jack L. (1964) "Some Methodological Considerations in Multiple-Cue Probability Studies," Psychol. Rev. 71, 42-60.

Ladefoged, Peter (1967) Three Areas of Experimental Phonetics (Oxford University Press, London).

Ladefoged, Peter and McKinney, Norris P. (1963) "Loudness, Sound Pressure, and Subglottal Pressure in Speech," J. Acoust. Soc. 35, 454-460.

Lane, Harlan (1962) "Psychophysical Parameters of Vowel Preception," Psychol. Monogr. 76 No. 44, 1-25 (Whole No. 563).

Lane, H. L., Catania, A. C., and Stevens, S. S. (1961) "Voice Level: Autophonic Scale, Perceived Loudness, and Effects of Sidetone," J. Acoust. Soc. Am. 33, 160-167.

Liberman, A. M., Cooper, F. S., Harris, Katherine S., MacNeilage, P. F., and Studdert-Kennedy, M. (1967) "Some Observations on a Model for Speech Perception," in Models for the Perception of Speech and Visual Form, Weiant Wathen-Dunn, ed. (The M.I.T. Press, Cambridge, Massachusetts).

Licklider, J. C. R. (1951) "Basic Correlates of the Auditory Stimulus," in Handbook of Experimental Psychology, S. S. Stevens, ed. (John Wiley and Sons, Inc., New York).

Mendel, Maurice I., Sussman, Harvey M., Mersen, Richard M., Naeser, Margaret Ann, and Minifie, Fred D. (1969) "Loudness Judgements of Speech and Non-speech Stimuli," J. Acoust. Soc. Am. 46, 1556-1561.

Moll, Kenneth L. and Peterson, Gordon, E. (1969) "Speaker and Listener Judgements of Vowel Levels," Phonetica 19, 104-117.

Peterson, Gordon E. and Lehiste, Ilse (1959) "Vowel Amplitude and Phonemic Stress in American English," J. Acoust. Soc. Am. 31, 428-435.

Peterson, Gordon E. and McKinney, Norris P. (1961) "The Measurement of Speech Power," Phonetica 7, 65-84.

Pollack, I. (1952) "On the Measurement of the Loudness of Speech," J. Acoust. Soc. Am. 24, 323-324.

Pols, L. C. W., van Der Kamp, L. J. Th., and Plomp, R. (1969) "Perceptual and Physical Space of Vowel Sounds," J. Acoust. Soc. Am. 46, 458-467.

Poulton, E. C. (1968) "The New Psychophysics: Six Models for Magnitude Estimation," Psychol. Bull. 69, 1-19.

Stevens, S. S. (1955) "The Measurement of Loudness," J. Acoust. Soc. Am. 27, 815-829.

Stevens, S. S. (1959) "Cross-modality Validation of Subjective Scales for Loudness, Vibration, and Electric Shock," J. Exp. Psychol. 57, 201-209.

Stevens, S. S. and Davis, H. (1936) "Psychophysiological Acoustics: Pitch and Loudness," J. Acoust. Soc. Am. 8, 1-13.

Warren, Richard M. (1962) "Are 'Autophonic Judgements' Based on Loudness?" Am. J. Psychol. 75, 452-456.

Warren, Richard M., Sersen, Eugene A., and Pores, Edwin B. (1958) "A Basis for Loudness Judgements," Am. J. Psychol. 71, 700-709.

A New Line Analog Speech Synthesizer for the PDP-12

Lloyd Rice

Introduction

The machine synthesis of speech has found general acceptance as a tool for research into both the production and the perception processes of linguistic communication. One of the techniques of speech synthesis which is particularly suited to the study of the production process is the articulatory synthesis approach, using a controllable transmission line analog of the vocal tract shape. This controllable shape is excited by some form of oscillation, either a direct acoustic signal or a representative form of energy which interacts with the tract shape analogously to the physical air vibrations which would result if the actual acoustic model were used. For example, one possible form for this representation is that of electrical activity. Physical theory shows that electrical vibrations interact with electrical resistances, capacitances and inductances in a manner closely analogous to the way air vibrations interact with the acoustic resistances and cavity dynamics of an acoustic model. A somewhat more abstract but equally valid representation is obtained if no real physical process is used but only the numerical structure of the physical theory is retained. This is the basis of the computer simulation approach to speech synthesis. The numerical structure is more freely manipulable than any real physical process and in principle can be extended to represent any aspect of the simulated process.

The fact is, of course, that no process can be completely and exhaustively described within the bounds of finite processing time. Compromises must be made between those aspects of the physical model which are interesting to the investigator and those which are not. Questions on the implications of the various compromises and approximations used in the present model will arise as the model is discussed in greater detail.

The model to be described consists of a variable area tube which is excited by a controllable glottal pulse input. The tube is represented by a series of equal length adjustable diameter tubular sections placed end to end to obtain a variable tubular resonator (see Figure 1). The diameter of the individual sections is controlled in time by a sequence of commands in the form of area function frames. Each frame is a set of area specifications, one for each of the tubular sections of the resonator. Control of the model ultimately consists of a time sequence

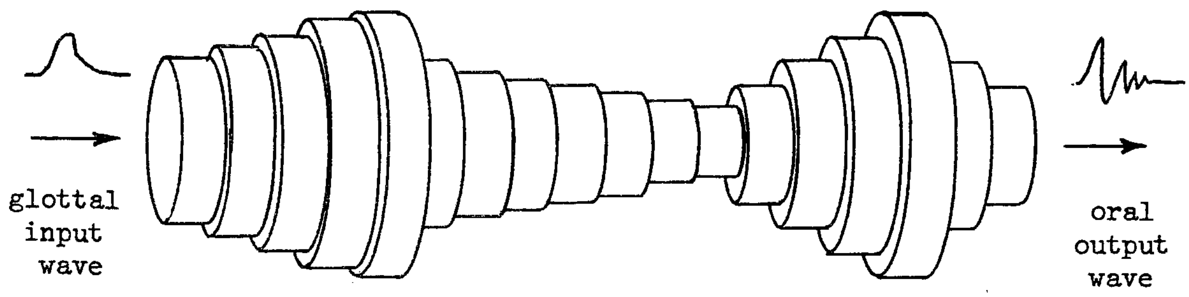


Figure 1.

of frames such that the vocal tract is represented by a time-varying variable area tubular resonator. This resonator is excited by a train of glottal pulses, each a copy of a stored glottal waveform. This glottal pulse train is adjustable both along the horizontal (time) axis to give a variable frequency and along the vertical axis to give a variable amplitude. The stored glottal waveform itself is initially adjustable but in the present system may not be changed during any one sequence of tract excitation.

A view of the model

The basis for the computational model of the acoustic process used in this system is a set of signal values representing the longitudinal volume velocity air flow at various points along the vocal tract. A travelling wave of air movement through such a tube is subject to wave reflection effects when the flow encounters a narrowing or widening of the tube. The effect is such that where the tube narrows in cross-sectional area a wave of the same polarity is reflected and, conversely, where the tube widens a wave of the opposite polarity is reflected. This may be seen in the physical air flow as a pressure wave travelling away in the reverse direction as a result of encountering the increased pressure necessary to force the air current through the smaller opening. Conversely, when the tube becomes less constricted the air current flows more easily and a rarefaction wave travels away in the opposite direction. In the case of the section approximation used in this model, these reflections occur at discrete points along the tract, namely at the boundaries between adjacent sections of different cross-sectional area. The magnitude of the reflected wave is proportional to the sum and difference of the areas of the two sections and may be represented as

$$V_{R_{it}} = V_{I_{it}} \frac{A_{i+1} + A_i}{A_{i+1} - A_i}$$

where $V_{R_{it}}$ is the reflected volume velocity in section i at time t,

$V_{I_{it}}$ is the incident volume velocity in section i at time t,

and A_i is the cross-sectional area of section i.

Where the area function narrows, the advancing air flow is diminished by the same amount as the strength of the positive reflected wave, and for a widening tube, the forward moving air flow is augmented by the amount of the negative reflected wave. This relationship maintains the energy and momentum values at the section boundary.

These forward and reflected waves, computed as described above, are then allowed to move along the tube for a delay period equivalent to the time required to cross a section length at the speed of sound. Assuming a sound velocity of 350 meters/second and a section length of one centimeter, this delay (Δt) is found to be 1/35,000 seconds. In practice, the computer cannot complete the required calculations for all sections in this time interval, requiring that the computed results be stored for later acoustic conversion, as described below. During this time interval the air flow is subject to a resistance due to absorption by the vocal tract surfaces. This absorption is effected in the model by diminishing the volume velocity by a small decay factor for each centimeter of flow through the tract. Thus the advancing wave incident upon the boundary between sections i and i+1 is represented by

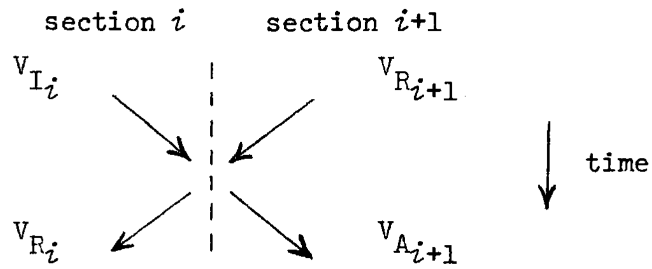
$$\begin{aligned} V_{I_{i+1,t+1}} &= k \quad V_{A_{i+1,t}} \\ &= k \quad (V_{I_{it}} - V_{I_{it}} \alpha_{i,i+1}) \end{aligned}$$

where $V_{A_{i+1,t}}$ is the advancing volume velocity from the boundary $i,i+1$ at time t,

$\alpha_{i,i+1}$ is the area relation $\dfrac{A_{i+1} - A_i}{A_{i+1} + A_i}$,

and k is 1 minus the decay factor.

It is actually necessary to compute the reflections described only at alternate intervals Δt in each section. The reason for this is made clear by a time graph of the signal flow as in Figure 2. If we let the volume velocity values be represented in the form



then we see from Figure 2 that the calculation need be accomplished only at time intervals $2\Delta t$ at each section boundary. Further, the same storage location may conveniently be used to store the reverse wave during alternate time intervals.

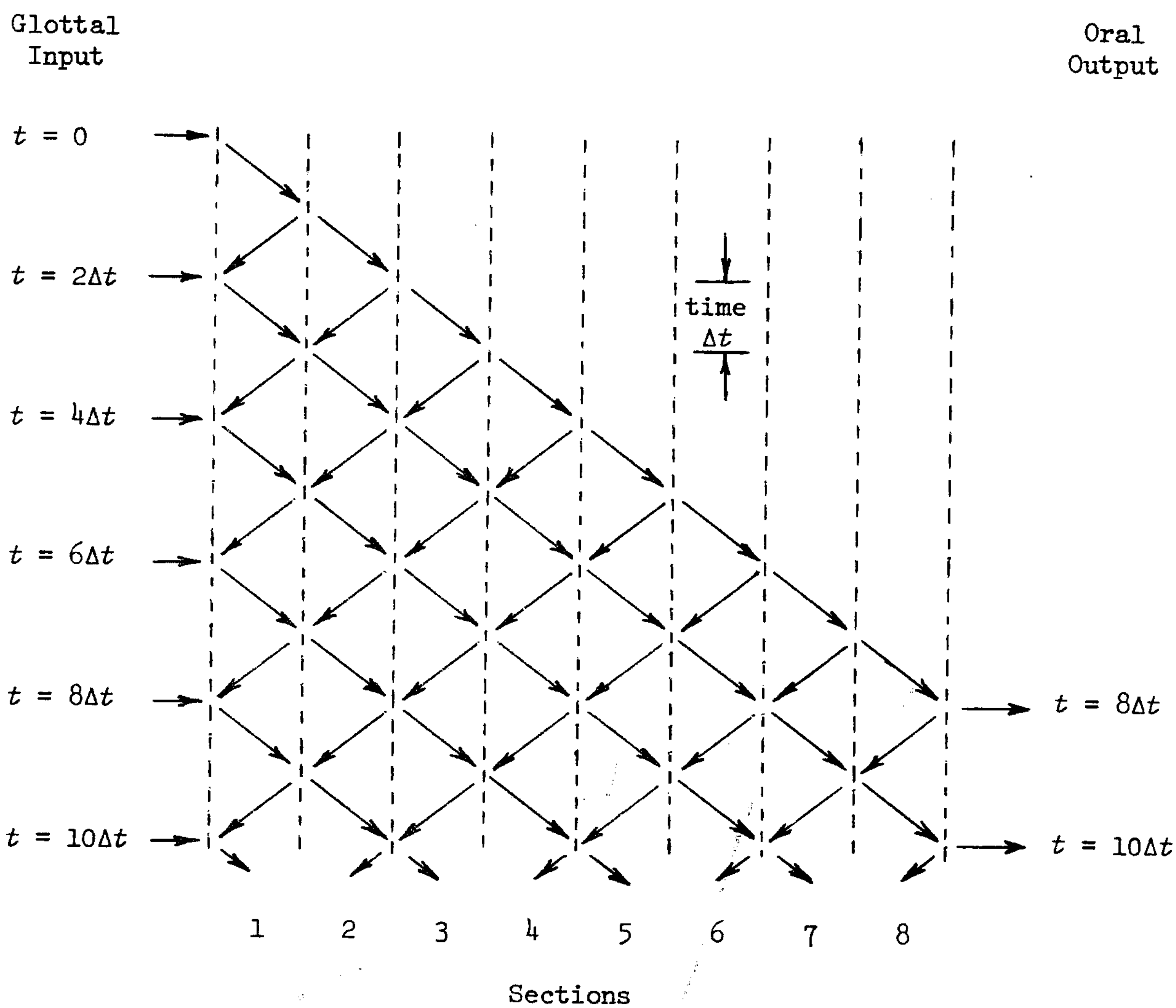


Figure 2.

Consideration of this model reveals that since only longitudinal air flow is considered in the sections the circular shape shown is really not a property of the model itself. In fact, the model makes no statement regarding the cross-sectional shape and implies that any distortion effects due to the geometrical shape of a section will lie outside the frequency range of interest. A comparison shows that the maximum expected section diameter is still small relative to the wavelength in air of the highest frequencies of interest for vowel formant structure. In a later section, the effect of this approximation on the higher frequency structure will be discussed.

One difficulty with the present implementation is experienced in the attempt to simulate small movements of an articulator along the longitudinal axis of the tract shape. It would be very desirable if, for example, twice the number of sections could be used. The present model is based on 18 sections of 1 cm. length each. The difficulty encountered in increasing the number of sections lies in the fact that the amount of computation necessary increases as the square of the number of sections. Thus, doubling the number of sections would multiply the computation time by a factor of 4.

Controlling the model

In designing the software system to control the area function model described above we felt that it would be desirable to specify the tract shapes in terms of the cross-dimension (CD) distance as viewed on a mid-saggital X-ray photograph rather than directly in terms of the cross-sectional area function. This would allow us to transfer measured data from X-rays directly to the model and also would establish a stage within the system at which the vocal tract representation has a clear cut relationship to more or less readily observable physiological structures. The CD (cross-dimension) specification of the tract shape is then converted to an area specification by reference to a prestored table of conversion values. This table gives a cross-sectional area value for each section for each possible CD-input value. Very recently, a new set of vocal tract geometry data has been reported based on an extensive oral casting process (Ladefoged, Riley and Anthony 1970). These new data are presently being set up for use in the synthesis model in the form of a new CD-area conversion table. A forthcoming paper will describe the use of the line analog synthesis model in cross checking these new vocal tract measurements against known formant frequency and CD measurement relationships.

The major software routine at the heart of the CD input process is a program for displaying a mid-saggital (X-ray) view of the vocal tract model on a CRT (cathode ray tube) screen integral to the PDP-12 control console (see Figure 3). The display consists of a fixed upper boundary

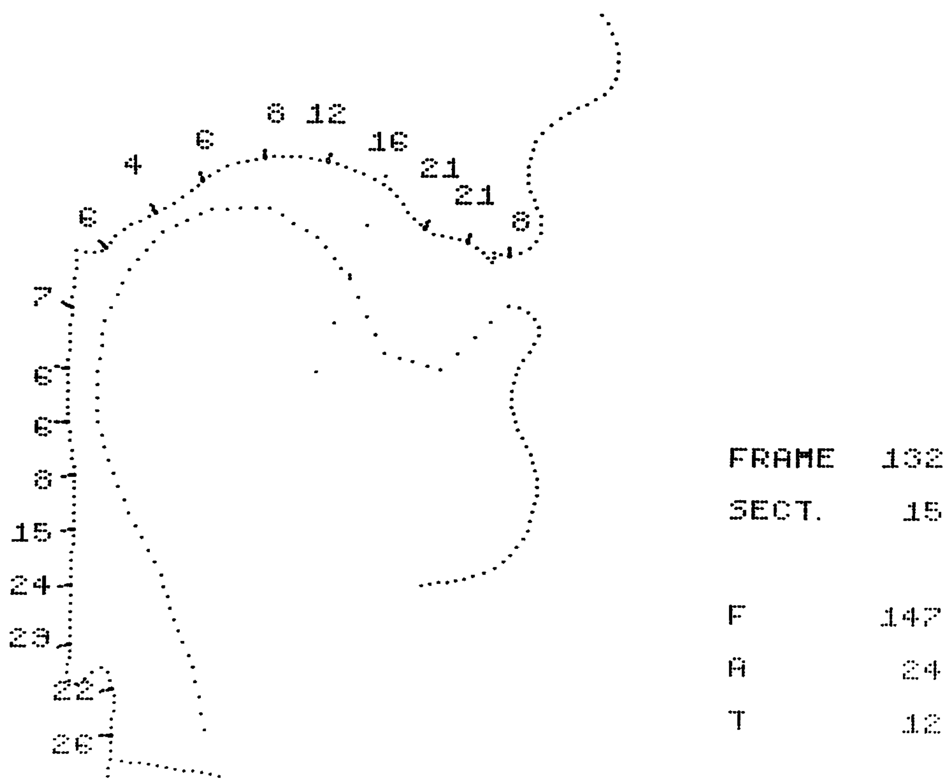


Figure 3.

representing the rear pharyngeal wall, the velum, soft palate, hard palate, teeth and upper lip. A movable lower boundary represents the tongue and lower lip as adjustable articulators. This movable portion is defined by straight lines joining the 18 points along the tract at the midpoints of the 18 sections. The distance between the fixed upper and movable lower tract boundaries at any section represents the CD value for that section. The display on the screen is actually life-size so that tongue and lip positions from X-ray photographs may be traced directly on the screen and the model then adjusted to match those positions.

The display routine allows easy manipulation of the CD value at any point along the tract. A knob at the operator's left hand controls the position of a dotted line cursor which marks the "current" section. Striking the key ">" or "<" then decreases or increases, respectively, the CD setting of the current section. Once the tract shape has been adjusted as desired, that shape may be stored as one frame in a sequence of frames. The tract shape may then be further altered and stored as a subsequent frame of the series until a maximum of 128 frames have been specified. The frames are numbered in octal base beginning with frame 1 and continuing through frame 200_8. The frame number and current section setting are displayed in the lower right-hand corner of the screen along with the frequency, amplitude and time duration settings. The frequency and amplitude parameters will be discussed in the following paragraphs.

Each frame may be specified as to a time interval or duration in milliseconds by typing the desired time followed by a "T". After a frame is stored in the control sequence, the time interval is automatically reset to the standard value of 12 msec. This value was chosen to give sufficiently fine temporal detail so that rapid movements could be made smoothly and still interact reasonably well with the glottal pulse timing constraints as discussed below. If the standard 12 msec. time is used for all 128 frames the total utterance length would be 1.536 seconds.

Glottal pulse excitation is applied to the vocal tract model by generating multiple copies of a single cycle glottal pulse waveform. This waveform is adjustable from the computer console, tracing in a new glottal wave shape by simultaneously moving two panel knobs. When the computation cycle is started to synthesize an utterance from a set of area function frames, this stored glottal pulse cycle is copied repeatedly as needed for the duration of the utterance. During the computation process for each frame this glottal pulse copying process is controlled in two ways by a pair of parameters associated with that frame. These glottal excitation parameters, frequency and amplitude, control respectively the rate of scanning across the stored glottal pulse cycle to produce the copy and a relative amplification factor which is applied to the copied wave before it is injected into the vocal tract model. The frequency and amplitude parameters are specified for each frame via a CRT display as shown in Figure 4.

Each horizontal point on the data curves represents 1 millisecond in time. Since the frame durations are adjustable in milliseconds, a frame set for 10 milliseconds would be displayed as a series of 10 points on these curves. A panel knob moves these two data curves to the left or right under the center screen cursor. As the curves move, the values under the cursor are numerically displayed as is the frame number of the associated frame. By typing either "F" or "A", the operator may make one

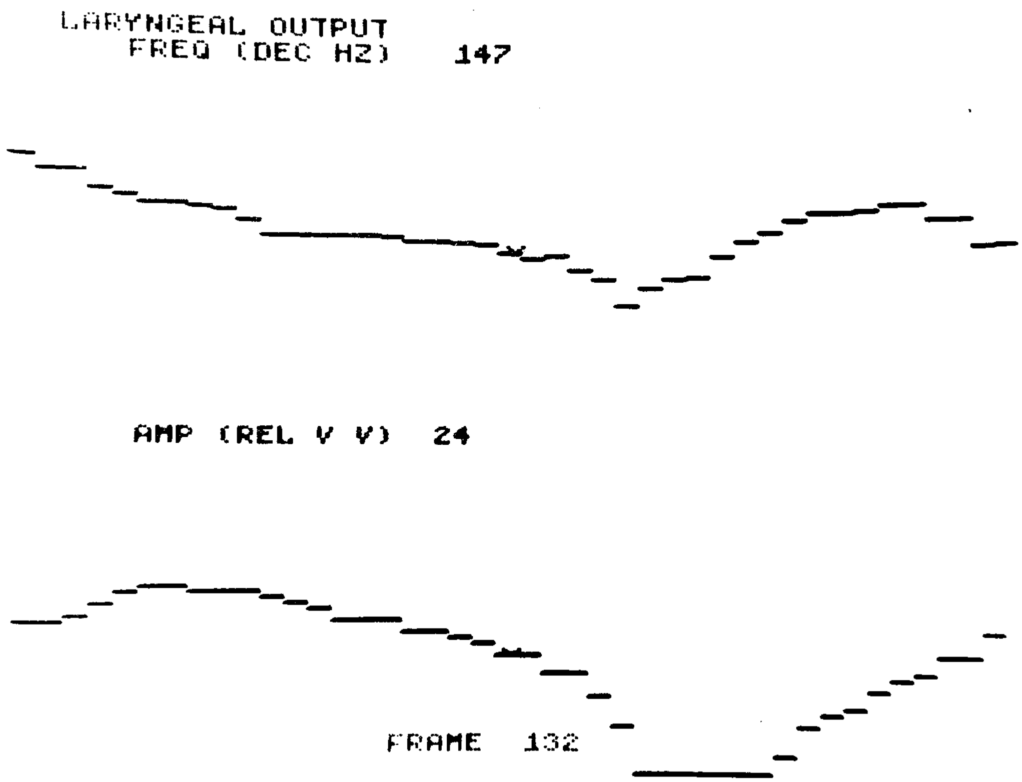


Figure 4.

of the curves adjustable. In this mode, the height of the selected curve at the cursor position is controlled by a second knob while the horizontal position is still under knob control. The horizontal motion may be set for a continuous drift and a new curve may then be traced in as each frame position passes under the cursor. The maximum range of the frequency parameter is 50 to 350 Hz which covers the upper half of the display. The amplitude parameter is a relative value ranging from 0 to 77_8 at full amplification. Because the actual numerical value of this parameter has no significance, it is displayed in octal base form for programming convenience. The amplitude parameter range covers the lower half of the display screen.

Once the computation for a frame has begun, as many cycles of glottal excitation are produced as needed to complete the frame. When the time interval for the frame ends, the current glottal cycle in progress is completed before the area function setting is switched to the next frame. This must be done to prevent unwanted acoustic dynamic effects which result in severe clicks and pops if the area function is changed suddenly while there is appreciable energy in the vocal tract. In order to maintain the correct overall frame timing pattern, the time interval for a particular frame begins when the interval of the previous frame ends although the computation will not actually begin using the area function values of that frame until some time later when the current glottal pulse cycle ends. This offset is noticeable only at very low glottal pulse frequencies where the glottal pulse wavelength may exceed the normal frame timing of 12 milliseconds. The operator should be aware that at low fundamental frequencies some frames may be skipped entirely. For example, if the situation occurs as shown in Figure 5 the frame 24 data would not be used in the computation.

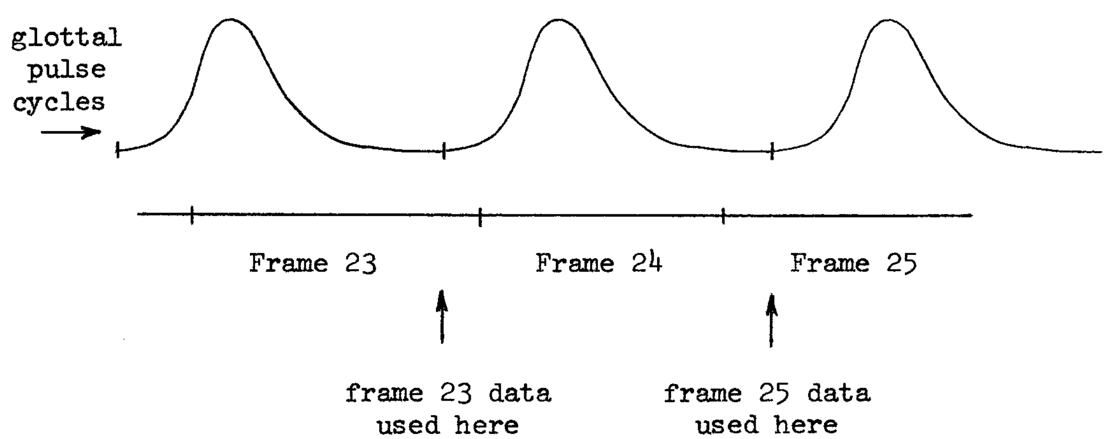


Figure 5.

Using the system

Constructing sequences of frames is facilitated by an editor-like function achieved with the library filing system. In addition, of course, particular sequences of frames may be saved in the library. The library routine allows any number of frames to be saved as a library entry. An index display lists all of the present entries and allows unwanted entries to be deleted from the file. Finally, a library entry may be loaded into the working sequence beginning at any desired frame number. Editing can be done by saving certain frames of the working sequence as a temporary library entry and then reloading the entry beginning at a smaller or larger frame number. Repeated gestures may be formed by successively loading a library entry which contains one repetition of the gesture. Four hundred forty-eight blocks on the LASS system linc-tape* are reserved for library file space. As each block holds up to 16 frames of data, sufficient library space is available, for example, for each student of a class to have several entries of utterances he is currently working on. A library entry includes all values associated with frames of a sequence and therefore includes the fundamental frequency and amplitude curves but does not save the glottal pulse waveform.

During the input stage of setting up a sequence of frames of vocal tract shapes, the experimenter has available a dynamic display mode which produces a rapid sequential display of a desired set of frames. The speed of this display is adjustable from a very slow motion when individual frames may be clearly viewed up to a normal speed display of the sequence giving a life like appearance to the motion. A particular subset of the total sequence of 128 frames may be designated as a temporary working set by specifying a beginning frame number and an ending frame number for the subset. The working subset so designated determines the set of frames displayed in the dynamic display mode as well as the set to be used in the next output computation. This allows the experimenter to single out a short section of an utterance for detailed work without having to go through the computation for the entire set of 128 frames on each computation. These beginning and ending frame limits are initially set to 1 and 200 and may be changed at any time.

When the output computation has been started and sense switch 0 is in the normal (reset) position a display is produced on the CRT screen consisting of simultaneous presentations of the area function setting and the instantaneous volume velocity flow in each section. The area function is displayed symmetrically on a horizontal axis with the glottal section (fixed at 0 cm^2 area) at the left and a large area radiating section outside the lips on the right. As the computation proceeds through the various frames, this area function changes by steps to the setting of each new frame. Superimposed on the area function display at the center of each section is plotted a pair of points representing the forward and reverse air volume velocity in that section. The volume velocity values are plotted on a vertical scale, a pressure wave deflecting the point upward and a rarefaction wave deflecting it downward. The values are plotted as they are computed, independently

---

* A pocket size computer tape containing 512 premarked randomly-accessible blocks of 256 words each.

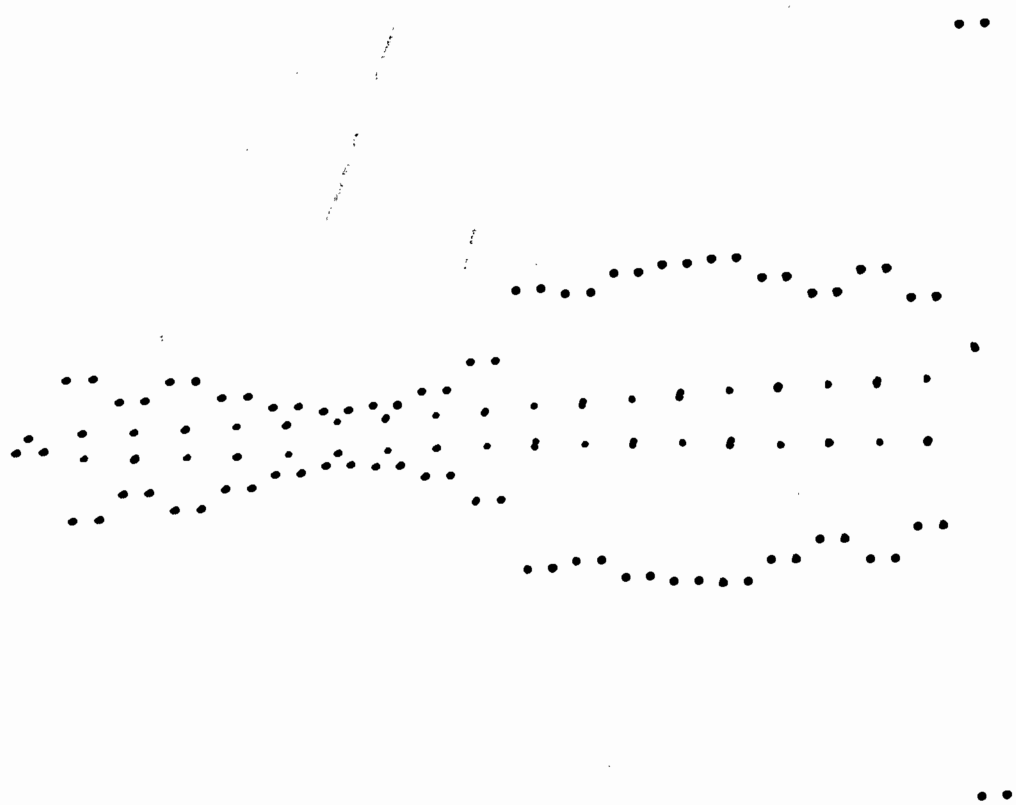


Figure 6.

for the forward and reverse directions as shown in Figure 2. The actual volume velocity in an acoustic model would be the algebraic sum of the forward and reverse flows. This would correspond to the sum of the displacements of the two points plotted at each section.

The two outermost sections, the glottal section on the left and the radiating section on the right, have only one volume velocity point displayed as reverse flow is disregarded in both of these sections.

In the case of the glottal section, the area of the opening is taken to be zero, resulting in 100% reflection of energy back into the upper tract. Under conditions of phonation this is a reasonable approximation, as the area of the vocal fold opening is very small relative to the cross-sectional area of the larynx just above the glottis. Ignoring the downward energy at this point also assumes that acoustic energy transmitted back into the lungs is completely absorbed and has no further effect on the oral dynamics. A different but related effect of acoustic conditions in the upper tract on the glottal waveform will be discussed in greater detail in the section on future improvements.

In the case of the radiating section, the assumption is made that the external acoustic environment has no effect on the operation of the vocal tract model. The radiation effect is simulated by the final large area section which causes energy to be reflected back into the tract with reversed polarity. The model then operates as if this large area section were infinite in length, as no energy is reflected back from the outer end of this section. Thus far, no detailed study has been made of the actual radiation effect achieved with the large area section. Thus far, no detailed study has been made of the actual radiation effect achieved with the large area section. The radiation characteristics may be expected to influence the spectral structure of the generated output wave, primarily in the rate of high frequency rolloff. The present system has a preset value of approximately 400 cm^2 for this radiating section.

If the experimenter desires to watch the volume velocity display while the output computation is in progress, a considerable portion of the computing time is required to produce the display. When the display speed is set to maximum (each set of points computed is displayed once) approximately half the time is required for this display. Setting sense switch 0 causes the display routine to be bypassed thus doubling the rate of computation of output data. Without the display, approximately 100 seconds of processing time are required to compute the output data for 1 second of speech.

When the output computation has been completed for the specified set of frames, the output data tape is rewound and a sampled data playback routine applies the string of data pulses to a system digital-to-analog converter. The output of the D to A converter is used to drive a console mounted loudspeaker so the experimenter may hear the generated output waveform. Audio output channels are available so the signal can be recorded, heard over headphones, etc. A selectable filter is available so the output voltage can be low pass filtered to remove sampling frequency artifacts. The sample frequency of the present system is 8800 Hz. The generated utterance is played repeatedly, rewinding the data tape automatically, until any key on the teletype keyboard is struck. The system then returns to the vocal tract display in CD input mode. At any time in CD input mode, the key "Q" may be struck, causing the last generated output to be "cued" up for playback, again repeated until any key is struck.

Future additions to the model

A number of modifications have been considered which would enhance the articulatory capability of the model and improve the realism of some of the limiting assumptions made in the design of the model. Most of these modifications would, unfortunately, add to the amount of computation needed, so increasing the required processing time.

Probably the most desirable modification to the model would be the addition of a nasal tract. The basic addition to the program for the nasal tract computation would be a routine to compute the acoustic interactions in a 3-way junction of acoustic tubes. My calculations for this interaction are based on the assumption suggested by Dr. Denis Klatt (personal communication) that the second area which figures in the reflection coefficient for each tube in the Y junction is the sum of the areas of the other two incoming tubes. Where one of the branch paths has zero area, the reflection calculation is seen to reduce to the standard two-path calculation described above. Letting A_1, A_2 and A_3 represent the 3 path areas as shown in Figure 7, the reflection coefficient affecting path 1 would be the ratio

$$\alpha_1 = \frac{(A_2 + A_3) - A_1}{(A_2 + A_3) + A_1} .$$

The reflected wave R_1 travels back into section 1 in the same manner as in the 2-path model. The forward transmitted volume velocity plus the forward wave (the inverse of R_1) is shown as one signal, F_1, which is divided between the paths 2 and 3 proportionately to their areas.

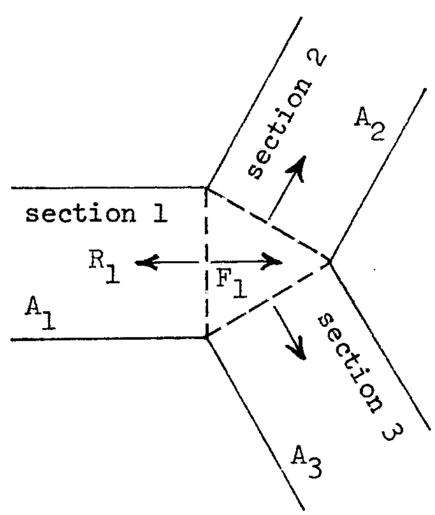


Figure 7.

The equation for the reflected signal is then

$$R_1 = -\alpha_1 V_{I_1} \quad ,$$

while the forward volume velocity is given by

$$F_1 = (1 + \alpha_1) V_{I_1} \quad ,$$

where

α_1 is the reflection coefficient, and

V_{I_1} is the incident volume velocity in section 1.

The total volume velocity leaving the 3-way junction in each tube is then the sum of the wave reflected in that tube plus the proportion to that tube of the forward transmitted wave from each of the other two tubes.

The nasal tract itself would consist of a fixed number of 1 cm. sections of which all but the first 1 or 2 would have fixed areas. As the acoustic energy in the nasal tract would be zero much of the time, it would not be necessary to compute the reflections in this tract continuously but rather a check could be made which would activate the nasal tract routine when the velar section opened to a non-zero area. This routine would then remain active as long as the velar section was open and until the energy in the nasal portion had decayed to a zero or near zero level. Two other effects of the nasal tract would remain to be investigated; first, whether the subsequent branching of the nasal cavity in the nose need be modeled to achieve the desired spectral characteristics, and second, what nature of mixing of the oral and nasal signals would adequately represent the radiation characteristics of the dual (or triple) source.

A problem is encountered with the present model in the simulation of acoustic events when a complete closure occurs anywhere in the tract. The most immediate effect of a closure is that the path of acoustic flow is completely cut off resulting in zero energy output during this interval. This is unrealistic, especially for the voiced consonants, in that the well known "voice-bar" is absent. The energy which contributes to the voice-bar is radiated from the throat, jaw and cheeks of a speaker, particularly when a closure causes acoustic energy to build up in the closed off cavity. A possible solution to this problem would be to add into the output signal a small contribution from each of the tract sections, a contribution approaching the amount by which the signal decays in traversing each section. This added signal would have the effect of smearing the frequency structure of the output; however, spectrograms made of the output of the model show very clearly marked formant bands, in fact these bands are unrealistically separated. Thus, it is felt that such smearing would be an improvement in the realism of the output.

A further effect of a complete closure is the acoustic loading of the glottal vibratory system by the energy buildup in the closed tract. In fact, Flanagan (1969) has suggested that this acoustic loading begins to affect the glottal waveform for some of the relatively constricted vowel configurations. The only means of simulating this acoustic loading in the present model is by reducing the glottal pulse amplitude curve just prior to and in frames in which a closure occurs. The amount of energy built up in the tract could be used as a control to automatically reduce the glottal amplitude factor as a means of freeing the operator from this responsibility. For example, if the previously described signal were computed which consists of a fraction of the energy in each section for adding into the output, it would appear to be a good measure of the energy as needed for the present purpose. Actually, the realistic solution would involve computing the cumulative pressure just above the glottis, a rather more complex process. This solution still would not, of course, affect the actual shape of the glottal waveform, which, as Flanagan shows, does change somewhat with increasing loading. Adjusting the waveform in response to the acoustic loading would require a more complex model of the larynx.

The above discussion brings us to another major improvement which could be made at a considerable cost of computing speed. The realism and dynamic effects of vocal cord vibration could be much more closely achieved with a discrete time vibratory simulation of the vocal folds than is obtained with the present simplistic waveform model. A number of such vocal vibration models have been described in the literature (Flanagan 1969; Flanagan, et al. 1968; Dudgeon 1970), most of which could easily be adapted to work with the present vocal tract model. A vibratory model of the larynx would have a number of advantages over the present system.

1. Glottal waveform and amplitude would correctly respond to the pressure build up that accompanies full closure.

2. Glottal waveform and amplitude would correctly respond to acoustic loading due to a restricted tract configuration.

3. Glottal waveform would vary realistically as a function of subglottal pressure and volume flow from the lungs.

4. Fundamental frequency would be a function of subglottal pressure and vocal cord tension, allowing experimentation with physiologically realistic parameters.

5. Glottal amplitude would be controlled by lung volume flow and the glottal pressure drop, again allowing control with physiologically realistic parameters.

It is difficult to estimate the computation increase that would be required in adding such a larynx model to the system, as the models reported often include more complicated computations to produce evaluations of the performance of the model and have been done on a variety of computer systems. Not all these models have been evaluated as to computer time needed. Nevertheless, some estimates can be made. The vocal tract model itself requires 3 multiplications and 3 additions at each section boundary for each double time interval $2\Delta t$. Assuming that a comparable time interval could be used in the larynx model, say $2\Delta t$, the larynx computation would have to be performed only once for the complete set of section reflection calculations, 57 multiplications and 57 additions. It would appear that such a larynx model would add 20 to 30% to the total computation time.

A problem quite different than any of those discussed above arises in connection with a fairly large class of speech sounds, those that depend on airstream friction or turbulence. These sounds are accompanied by the hissing noise of the air turbulence which is filtered by the vocal tract shape depending on that shape and on the location of the noise source. A cavity resonant at a particular frequency in the noise spectrum will absorb energy at that frequency and harmonically related frequencies if it is located behind the noise source, while it will pass that frequency if it is in front of the sound source (Fant 1960). The nature of the model suggests that the spectral modifications of a hiss source would be properly computed if the hiss were injected at the proper point along the tract. The hiss source itself could easily be simulated by a random number generator. A relatively simple randomizing process could be used as the statistical distribution would probably not be critical. The remaining question, then, is how to control the amplitude of the hiss source. At this point, we must consider in more detail the acoustic representation used in the model.

The volume velocity signal referred to throughout the above discussion is the instantaneous wave velocity of the acoustic energy and does not necessarily correspond to the actual air currents. As an example of this difference, consider the case of a closed off oral cavity which is momentarily excited by a pulse train representing an influx of air. If the excitation is now suddenly cut off, the volume velocity signal will decay to zero within a few milliseconds as would the vibrations in a physical model. In the real case, however, there now exists stored energy in the form of air pressure which is not represented in the volume velocity model. If a small opening is made to release this air pressure, air turbulence will create a hiss source at the opening. In the volume velocity model, all sections have a stored value of zero, and there is no record of the true air currents.

Thus, we see that a true model of the acoustic situation would involve a further computation at each section, a cumulative record of volume velocity inflow and outflow at each section, providing a measure of the air pressure at each point. This would also provide the oral pressure measure mentioned earlier as needed to counter sub-glottal pressure and shut down the vocal cord vibration when an upper tract closure occurs. As a first approximation, the volume velocity signal used in this model could be employed as the basis for computing air friction noise. The glottal input system could be generalized to allow injecting a continuous (non-pulsing) positive volume velocity value, which would allow reasonable simulation of fricative and sibilant sounds. This approximation would, however, be inadequate whenever air compressibility effects become important, as they almost certainly would for affricate sounds and aspirated stop openings. To some extent, this inadequacy could be compensated for by the addition of an artificial burst of glottal input just at the moment of opening. In addition, this solution would rejuvenate the problem of the acoustic effects of vocal tract shape changes with non-zero volume velocities.

One further improvement to the model will be considered; that is the capability of changing the lengths of individual sections or of the entire tract. In at least two situations, modeling vertical movements of the larynx and more accurately representing lip section changes, it would be highly desirable if individual section lengths could be varied as well as section areas. Unfortunately, the algorithm described here rests firmly on the assumption that all sections are of equal length. An obvious solution would be to increase the number of sections to the point where a greater or lesser number of sections could be taken to reach the desired length. The computational cost of this solution has already been noted. Dr. Shizuo Hiki has suggested (personal communication) that within the limitations imposed by the discrete structure of the model, small changes in section length may be adequately modeled by appropriate changes in the areas of adjacent sections. This has been borne out to some extent by experimentation with the existing system. Vowel rounding can be adequately achieved by adjusting the lip section somewhat closer than would otherwise be expected and leaving a relatively large opening just behind the lips.

Changes in the overall length of the modeled tract are relatively easy to simulate by changing uniformly the modeled length of all sections. In effect, this amounts to changing only the pulse rate of the auditory playback system and adjusting a constant in the glottal pulse frequency determination routine. Thus, the basic model is readily useable to simulate large or small physical tract sizes.

Acknowledgements

I am indebted to Dr. Denis Klatt and Dr. Shizuo Hiki for including me in an exchange of ideas on this model and to Johan Lilliancrants for a brief discussion of the basic form of the model. Many relevant observations were made by Dr. Ladefoged and an early version of the section computation routine was written by Giles Peterson. The system has been rendered much more useful and convenient to use by incorporation of a number of suggestions by various people during the development phase.

I especially want to thank Richard Harshman, Mona Lindau and Dale Terbeek. Larry Grant and Margaret Press assisted with the preparation of some of the figures.

References

Dudgeon, D.D. (1970), "Two mass model of the vocal cords," paper presented at the 79th Meeting of the Acoustical Society of America, April, Atlantic City.

Fant, G. (1960), *Acoustic Theory of Speech Production*, Mouton: The Hague.

Flanagan, J.R. (1969), "Use of an interactive laboratory computer to study an acoustic-oscillator model of the vocal cords," *IEEE Transactions on Audio and Electroacoustics AU-17, No. 1.*

Flanagan, J.R. and Landgraf, Lorinda L. (1968), "Self-oscillating source for vocal tract synthesizers," *IEEE Transactions on Audio and Electroacoustics AU-16, No. 1.*

Ladefoged, P., Anthony, J., and Riley, C., *Journal of the Acoustical Society of America*, abstract forthcoming.

Simplicity is a Complicated Question

Victoria A. Fromkin

"Seek simplicity, and distrust it".
A. N. Whitehead

With the publication of the famous "little blue book" (Chomsky, 1957) linguists were freed from attempting the impossible task of finding a discovery procedure for grammars. It will be remembered that in *Syntactic Structures* three alternative sets of requirements for a linguistic theory were suggested: (1) that it provide us with a *discovery procedure*; (2) that it provide a *decision procedure*, i.e, " a practical and mechanical method for determining whether or not a grammar proposed for a given corpus is... the *best* grammar of the language ..." and (3) that it provide an *evaluation procedure* for grammars, i.e. "the theory must tell us which is the *better* grammar of the language [given two proposed grammars]" (p. 51)

For a time, it was the third alternative which was considered the only reasonable one, and linguists sighed with relief at the "lowering (of) our sights to the more modest goal of developing an evaluation procedure for grammars". It all seemed so "simple" (in the non-technical sense of "simplicity"). Three main tasks faced us: first, the necessity "to state precisely (if possible, with operational, behavioral tests) the external criteria of adequacy for grammars."; second, the characterization of "the form of grammars in a general and explicit way";and third, the analysis and definition of "the notion of simplicity that we intend to use in choosing among grammars all of which are of the proper form." (Chomsky, 1957, pp 53-54)

As we proceeded we discovered that the "weakest of the three positions" seemed a lot harder than the "discovery procedures" outlined by our predecessors. Operational and behavioral tests were difficult to design, the correct form of grammars was tied to both the external criteria and the simplicity metric, and the technical definition of simplicity became equated with "a certain property that we may call the 'value' of the grammar" (Chomsky and Halle, 1958, p. 296). As this notion of "simplicity" developed, we found ourselves aiming at a theory which implicitly, if not explicitly, accepted the second of the three alternatives, i.e. a theory which would provide a "decision procedure" for grammars, since by 1968, we were told: "There is some *correct* answer to the question of how lexical items

---

I am grateful to Richard Ogle for raising some of the questions regarding the 'evaluation metric' considered here.

are represented and what the phonological rules are. A particular notion of "values" or "simplicity" will lead to an assumption about lexical items and phonological rules which is *either right or wrong* .... A specific proposal as to the definition of 'value' will make certain assumptions as to what constitutes a linguistically significant generalization, as to what constitutes a 'regularity' of the sort that a child will use as a way of organizing the data he is confronted with in the course of a language acquisition." (Chomsky and Halle, 1958, p. 296) If the grammar selected by our "evaluation procedure" was to be the same "grammar that a child constructs in learning his native tongue (which) will ... be the one that ranks highest in terms of this evaluation measure", (p 251) it is obvious that we now seek "a practical...method for determining... the *best* grammar", not the better of two.

The difficulties facing us as linguists have therefore been multiplied, both in our attempt to constrain our theory to permit such a selection, and to apply the theory as it now stands to the writing (or selecting) of grammars, or parts of grammars.

The three tasks are intricately connected. Since the selection of formal devices and the evaluation metric specified by the theory are "empirical" questions the external criteria become primary. In other words it is not enough for a grammar to account for the facts (i. e, the raw data) in an elegant fashion, the grammar must be a model of the innate grammar constructed by the child, must be 'psychologically real'. Without those behavioral and operational tests, however, we are often faced with solutions which are 'selected' by our evaluation procedure but which seem to fail the 'reality' test, or with non-resolvable alternative solutions.

An attempt to construct a set of morpheme structure conditions (Stanley 1967) for English exemplifies our dilemma. This should be a relatively easy task. Speakers of English 'know' the segmental and sequential redundancies, i.e. the particular set of phonemes and the constraints on their sequential occurrence. Every student in an introductory linguistics course points to the fact that even a child knows that *brick* is a word in English, *blick*, a non occurring but possible word, and *bnick* an inadmissable sequence. Behavioral tests can easily be devised to test this knowledge. Speakers of English also know that "the first segment of an initial consonant cluster must be [s] if the second segment is a true consonant." (Chomsky and Halle, p. 171) Because this a 'true' generalization, a Morpheme Structure Condition(MSC) can be included in our grammar of English:

$$(1)\qquad \begin{array}{lccc} \text{If:} & + & [+\text{consonantal}] & \begin{bmatrix} +\text{consonantal} \\ -\text{vocalic} \end{bmatrix} \\ & & \Downarrow & \\ \text{Then:} & & \begin{bmatrix} -\text{vocalic} \\ +\text{anterior} \\ +\text{coronal} \\ +\text{strident} \\ +\text{continuant} \\ -\text{voice} \end{bmatrix} & \end{array}$$

Our theory tells us that "each redundancy condition (must) result in a saving of features in the lexical representations that is larger than the number of features required to state the condition itself." (Chomsky and Halle, 1968, p. 389)

Given the number of morphemes in English which start with [sp], [sk], [st] by including such a MSC we 'save' a vast number of feature specifications in the lexicon, and reveal the non-existence of initial [pk], [ft] etc., thus explicating some of the speaker's knowledge. The evaluation measure has served its true purpose.

But a speaker has other kinds of knowledge, which is revealed by P-rules, rather than MSC's. Because of the alternations between such forms as for examples *resign /resignation, malign/ malignant, phlegm/ phlegmatic, repugn/ repugnant, paradigm/ paradigmatic* and in keeping with the constraint imposed by our theory that only the idiosyncratic properties of a formative should be expressed in the lexicon, we conclude that speakers of English 'know' that there is a /g/ in the lexical representation of the stems in the above words, and further 'know' that this /g/ is deleted just in case it occurs before a word final /n/. We therefore cannot permit a MSC which would exclude the sequence of morpheme final /gn/ just as we cannot permit MSC's to exclude morpheme final /mb/ because of the pairs *iamb/iambic, bomb/bombastic, crumb/crumble, thumb/ thumbelina, womb/wombic* (compare, *plumb/plum#er*). In addition, morpheme final /mn/ must be permitted by our MSCs to account for *hymn/hymnal, solemn/ selmnity, condemn/condemnation, damn/damnation, autumn/autumnal*, and morpheme final /Cr/ and /Cl/ (where C is a true consonant) must not be excluded because of *hinder/hindrance* as opposed to *potter/pottery* (with a final /Vf/) and *single/singly*.

It seems fairly obvious that a speaker's knowledge concerning the permissibility of these sequences of segments within a morpheme is different than his knowledge of the *brick/blick/bnick* variety, i.e. of the initial cluster [s] rule. This rule represents his knowledge of the *phonetic* redundancy rules of English, differing from Morpheme Structure Conditions, which "represent redundancies at a single level, the systematic phonemic level." (Stanley, 1967, p. 395) And, as Stanley has convincingly argued, "the complexity and depth of ordering which in general exists in the P rules shows that it may be difficult to construct any natural set of phonetic redundancy rules, even in principle." (p. 405) We cannot equate a speaker's knowledge of the phonetic redundancy rules with his knowledge of phonemic or lexical redundancy because the two sets of constraints are often contradictory. *Phonetically* a final [gn] is inadmissible, as is [mn] etc. Phonemically they may be permissable sequences. But 'behavioral tests *of the sort* which reveals the *blick* knowledge would fail in the case of admissible *underlying* sequences. The verification of the lexical representations and the MSC's is therefore totally dependent on the evaluation metric and the formal constraints provided by the theory. The 'psychological reality' of a final /gn/ rests on the existence of [n]/[..gn..] alternating pairs, the overall simplicity of the grammar, and the fact that the outcome of the grammar produces phonetic strings which are acceptable to the native speaker. There seem to be no 'external criteria' by which we can decide on the 'reality' of the underlying strings.

This being the case, perhaps the MSC which permits initial /br/ and /bl/ but not /bn/, and would specify the first of two true consonants in an initial cluster as [s] should not be a constraint on morphemes but rather viewed as a phonetic redundancy rule.

In fact, we do find in English alternating pairs in which one member has an initial [n] and the other a [..gn..] as in *gnaw/agnatous, gnosis/agnostic/agnosia, gnash/agnathic, gnomen/agnomen.* The MSC (1) can be revised to permit initial /gn/:

$$\text{(2)}\qquad \begin{array}{ll} \text{If:} & + \begin{bmatrix} +\text{consonantal} \\ \alpha\text{voiced} \end{bmatrix} \quad \begin{bmatrix} +\text{consonantal} \\ -\text{vocalic} \end{bmatrix} \\ & \qquad \Downarrow \\ \text{Then:} & \begin{bmatrix} -\ \alpha\text{continuant} \\ -\ \alpha\text{strident} \\ -\ \alpha\text{anterior} \\ -\ \alpha\text{coronal} \\ -\ \ \text{vocalic} \end{bmatrix} \end{array}$$

It is immediately apparent that if the evaluation metric contained only the convention regarding redundancy rules mentioned above, this changed MSC would be disallowed, since it would force us to specify the value of the voice feature for every initial /s/ in a cluster, and since there are more #sC formatives than #gn formatives, MSC (1) would save more features, and would therefore result in a more highly valued grammar.

But "Notice that simplicity is a *symstematic* measure; the only ultimate criterion in evaluation is the simplicity of the whole system. ....by simplifying one part of the grammar we may complicate other parts." (Chomsky, 1957, 56)

Would the inclusion of MSC (2) lead to simplifications in other parts of the grammar? Chomsky and Halle (1968) derive the 'g-less' *resign* by first positing a rule which spirantizes the velar:

3. $g \rightarrow [+\text{cont}] \ / \ \text{---}[+\text{nasal}]\#$ (p. 234)

and then, delete the continuant by a rule already posited to derive [rayt] and [rəyčəs] from the stem /rixt/:

$$4.\ \begin{bmatrix} -\text{cor} \\ -\text{ant} \\ +\text{cont} \\ -\text{voc} \end{bmatrix} \rightarrow \emptyset \ / \begin{bmatrix} \text{C---} \\ \text{---C} \end{bmatrix}$$

These same rules, with a slight modification of rule 3. will delete the initial /g/ when followed by /n/.

5. g → [+cont] / ---[+nasal]/ #

By using Bach's neighborhood convention (1968) and Langacker's mirror image rule (1969) a pre-nasal /g/ will become a velar continuant before or after a word boundary, which will then be deleted by rule 4.

Furthermore, to account for the deletion of a word final /g/ after a nasal in *sing* and *sing#er*(mentioned in Chomsky and Halle, 1968, p. 85, but not formalized) we can further extend rule 5, which will permit Rule 4 to apply to a post-nasal final /g/ as well as pre-nasal /g/:

$$6.\ \begin{bmatrix} \text{C} \\ -\text{cont} \\ -\text{coronal} \\ -\text{anterior} \\ +\text{voiced} \end{bmatrix} \rightarrow [+\text{cont}] \ / \ [+\text{nasal}] \ / \ \#$$

which expands to:

6.a.	g → [+cont]	/ -- [+nasal]#	e.g. *sign* as opposed to *signature*
b.		/ [+nasal]--#	e.g. *sing, sing#er,* but not *finger*
c.		/ #--[+nasal]	e.g. *gnosis* as opposed to *agnostic*
d.		/ # [+nasal]---	(vacuous)

The g → ɣ rule, and the final g deletion rule, proposed in(Chomsky and Halle, 1968) are ad-hoc rules; they are now no longer needed. Other rules are generalized and the grammar is thereby simplified. A connection is also revealed between all the cases where an underlying /g/ is deleted when in the immediate vicinity of an initial or final cluster.

But we still seem to be missing a generalization, since not only is /g/ deleted in these environments, but /b/, the other [-coronal] voiced stop is also deleted, whereas the [+coronal] /d/ is retained as are the voiceless stops (*lamb/lamp, limb/limp, sing/sink,* and *band, bend, bent* etc.) If we attempt to simplify the g-spirantization rule to include

/b/ by deleting the [-anterior] specification, and similarly delete the [-anterior] specification from the continuant deletion rule we would end up deleting the /f/ from forms such as *sift*.

What has happened is that in our attempt to exclude two ad hoc rules, and to 'simplify' the grammar by generalizing already existing rules, we have overlooked a real generalization, i.e. that there is a *phonetic redundancy rule* which disallows final or initial nasal plus voiced-stop clusters except where they are both [+coronal]. We can then posit a simple rule to handle this:

7. $$\begin{bmatrix} \text{C} \\ \text{-coronal} \\ \text{+voiced} \\ \text{-cont} \end{bmatrix} \rightarrow \emptyset \; / \; [\text{+nasal}] \; / \; \#$$

This will delete the /g/ in */sing/ /sign/ /gnosis/* and the /b/ in */iamb/* etc.

The rule, as it stands, however, will not account for the /n/ deletion in *condemn, solemn,* etc. and will have to be altered as in 8.

8. $$\begin{bmatrix} \text{C} \\ \left\{ \begin{matrix} \begin{bmatrix} \text{+nasal} \\ \text{+coronal} \end{bmatrix} \\ \text{-coronal} \end{matrix} \right\} \\ \text{+voiced} \\ \text{-cont} \end{bmatrix} \rightarrow \emptyset \; / \; [\text{+nasal}] \; / \; \#$$

Since the first rule expansion will delete the /n/ in *condemn,* the structural description for the second part of the rule will not be met, and the /m/ will remain in final position.

This rule seems to capture our 'intuition' about phonetic redundancy rules and at the same time allows us to remove two ad-hoc rules from the grammar and, in addition, accounts for more of the data than do the ad-hoc rules. This does, of course, weaken the argument for the /x/ in *right* postulated by Chomsky and Halle since the velar continuant deletion rule is now not as general as suggested. The argument that /s/, /f/, and /s/ appear in post-vocalic position, and therefore /x/ appearing in this position fills a "phonological gap" (Chomsky and Halle, 1968, p. 234) assumes that there *is* an /x/ in this position, that speakers of English 'know' this by virtue of their knowledge of other rules, although they have never 'heard' it in this position or elsewhere, for that matter. In addition, to suggest that the "*apparently* (my italics) irregular occurrence of [ŋ] in word-medial position, as in *dinghy, hangar, gingham, Birmingham*" can be explained by assuming a medial /nx/ is not necessarily a better solution than, for example, putting a word boundary after the /g/, (i.e. ding#y). It is interesting to note that in an informal survey of seven native speakers, all of whom spoke a dialect in which *singer* had no [g], four of them pronounced *hangar* as [hæŋgr̩]. Webster's Third also gives this as an alternative pronounciation to [haeŋr]. *Birmingham* also has the alternative pronounciation of [bərmiŋhæm], as well as [bərmiŋəm] and, in fact, six of the seven gave it the first pronunication. There obviously is an underlying /x/ or /h/ which for many speakers is realized phonetically. For this dialect, the /x/ deletion rule would not work. Furthermore, as the rule now stands, it would delete the /x/ in *adhesion* because of the convention which stipulates that a rule which applies within a morpheme also applies across a morpheme boundary.

The difficulty is that our theory with the present evaluation metric does not really help us to decide between the solutions offered. A morpheme structure condition which specified all post-vocalic continuants as [+anterior], thereby disallowing the occurrence of /x/ in *right* certainly results in "a saving of features in the lexical representations that is larger than the number of features required to state the condition itself." (Chomsky and Halle, 1968, p. 389) The decision to specify *right* with an /x/ or to allow initial /gn/ obviously cannot depend on this alone. Chomsky and Halle state that they "have very little to say...about the interesting question of how complexity of the lexicon should be measured against complication of the phonology in evaluating a grammar." (Chomsky and Halle, 1968, p. 216)

Given this huge lacuna in the theory, our dilemma is compounded. Since the cluster simplification rule given above does account for the deletion of /g/ in both a final and initial nasal cluster, do we permit the initial /g/? If not, do we enter *gnosis* and *agnostic* as separate items, thereby denying the regularity of the relationship and the /g/ deletion? Or do we enter them with one initial /g/ stem, and mark them minus the MSC specifying the first of two true consonants to be an [s]?

Without further constraints the theory appears to be too powerful, we can play lots of games, and put forth intricate arguments for underlying /x/'s, for a boundary in confetti (/con=fetti/) which is necessary for the rules to work, for a /ɛ/ glide etc. We can, if we like, account for the relationship and alternations between such pairs as *know/knowledge*, *mnemonic/amnesia* and *pterygoid/dipterous* * if we mark these items as minus the MSC, and extend the P-rules still further.

I think there is general agreement that as now stated the "evaluation metric" is no metric at all. It doesn't tell us how to count or what to count. Does, for example, a symbol stated in a rule condition cost the same as, double, or half as much as a symbol in a rule? Is the cost

---

* Mona Lindau suggests, facetiously, that while we're at it why not also include *bush/ambush*.

of violating the 'invariance condition' simply the number of symbols in the rules which convert it to its phonetic representation, or do we count the number of rules which must apply? If the rules are independently justified for other parts of the grammar should the cost be equal to that of say one "simple" rule which applies only to one segment (e.g. converts an underlying "unmarked" /t/ to a dental click)? Is the cost of readjustment rules more or less, or equal to regular P-rules? Do diacritic features cost more than distinctive feature specifications?

Even where we know how to apply the "metric" it doesn't work. We are told that "the number of symbols in a rule is inversely related to the degree of linguistically significant generalization achieved in the rule." (Chomsky and Halle, 1968, p. 335) Yet, given two rules:

(a) $C \rightarrow [+\text{voiced}] \ / \ V \text{ -- } V$

(b) $C \rightarrow [+\text{voiced}] \ / \ V \ (\#) \text{ -- } V$ *

we know that (b) is the more general and yet it is less highly valued.

Like other sciences, we seek the simplest possible theory. But "The logically simpler is not always the mathematically simpler." (Rosenthal-Schneider, 1959, p. 137) According to Popper (1959) "Simple statements, if knowledge is our object, are to be prized more highly than less simple ones because they tell us more: *because their empirical content is greater, and because they are better testable.*" (p. 23) But we do not have to count symbols to determine which statement tells us more. We are told that "To say that the rules may be given in a simpler form implies that they *must* be given in that form." (Chomsky and Halle, 1968, p. 71) but again it is not because of a counting device which constrains us to do this, but because a simpler rule "tells us more".

What I am suggesting is that the notions of "'simplicity' or 'economy' in the very general (and still very poorly understood) sense in which these terms usually appear in writings on the philosophy of science" (Chomsky and Halle, 1968, p. 335) seem to be better understood than our "technical" notion of simplicity which too often provides us with no answers, or incorrect ones.

We do indeed need constraints in our theory of grammar, constraints which place "more restrictive conditions on the choice of grammars, limiting the kinds of rules that can appear in them and the ways in which these rules can operate" (Chomsky, 1970, p. 7) and these constraints must be based on substantive criteria. An automatic counting machine will not tell us whether or not we have constructed a valid model of a speaker's grammar, and in fact may prevent us from achieving the explanatory theory we seek. We need to go back to our first task -- "The external criteria of adequacy for grammars" -- and devise the "operational behavioral tests" which can provide the information we need.

---

* Example suggested by T. Vennemann

References

Chomsky, N. (1957) *Syntactic Structures*, Mouton, 'S-Gravenhage.

Chomsky, N. & M. Halle (1968) *The Sound Pattern of English* Harper & Row. New York.

Chomsky, N. (1970) "Some Empirical Issues in the Theory of Transformational Grammar" (pre publication manuscript.)

Popper, K. (1959) *The Logic of Scientific Discovery*: London.

Rosenthal-Schneider, I. (1959) "Presuppositions and anticipations in Einstein's Physics". in P. A. Schlipp (ed) *Albert Einstein, Philosopher Scientist*, V.I., Harper, New York, pp. 129-146.

Stanley, R. (1967) 'Redundancy rules in phonology' *Language* 43:3. 93-436.

A Comment

One of the advantages of producing this series is that one gets the opportunity to slide in sly notes at the ends of articles such as this. In the last general issue (*Working Papers in Phonetics*, 14) we promised that we would continue the argument over whether the aim of phonology is to account for part of the linguistic competence of a speaker (Fromkin), or whether the aim is to describe the patterns which occur in the medium of spoken language (Ladefoged). I cannot resist expressing surprise that a reputable scientist (such as Fromkin) should have the temerity to imply that complex rules such as those in the foregoing article should be part of the competence acquired by an English speaking child. It seems to me to be folly to propose that speakers, in any sense, "know" rule (8) without having a shred of evidence other than observations of *what occurs in the medium of spoken language*. The fact that the patterns occur is enough reason for me to include them in a phonology. But because she is claiming that a phonology should account for more, Fromkin should have additional evidence before she writes rules of this kind.

Peter Ladefoged

The Interruptibility of Speech

Peter Ladefoged, Raymond Silverstein and George Papçun

This is a preliminary report on a new technique which we hope may reveal something about the way in which a speaker plans his utterances. The initial hypothesis was that there would be some moments in the stream of speech when a speaker would find it more difficult to interrupt himself than at other moments. Thus we thought it likely that a speaker might find it more difficult to interrupt himself in the middle of a syllable than at the end; and perhaps that interruptions might be much easier at the end of a word or phrase rather than in the middle. Similar effects have been reported in the performance of other skilled actions. As long ago as 1947 Craik pointed out that certain hand movements were more difficult to interrupt than others because the actions were, in his words, "triggered off as a whole." And any tennis player or golfer knows how difficult it is to interrupt himself in the middle of a stroke.

In our initial experiments the subject was asked to repeat a word or short phrase after he had heard it on a tape. Then on some occasions the experimenter gave him a signal to stop saying that word, and to say "ba" instead. This kind of task proved to be very complicated to interpret. We did not know whether the difficulties that the subject experienced were due to difficulties he had in perceiving the varying phrases spoken by the experimenter, or in deciding what it was he was going to say next, or in actually saying a particular utterance. Accordingly we revised the procedure so that we eliminated the difficulties the subject had in perceiving what it was he had to say, and in deciding to say it, and kept only that part of his task which entailed his organising himself so that he could say a given phrase.

The experimental set-up we used is shown in Figure 1. On seeing a light which flashed on at about 4 second intervals, the subject, who was sitting in a sound proof room,had to repeat a word or short phrase. The signal from the light went to a variable delay. Since the subject said the phrase about the same length of time after the light on each occasion, the experimenter was able to set the delay so that the stimulus, a 1 Khz tone, was produced in the subject's headphones at variable times throughout the word or phrase.

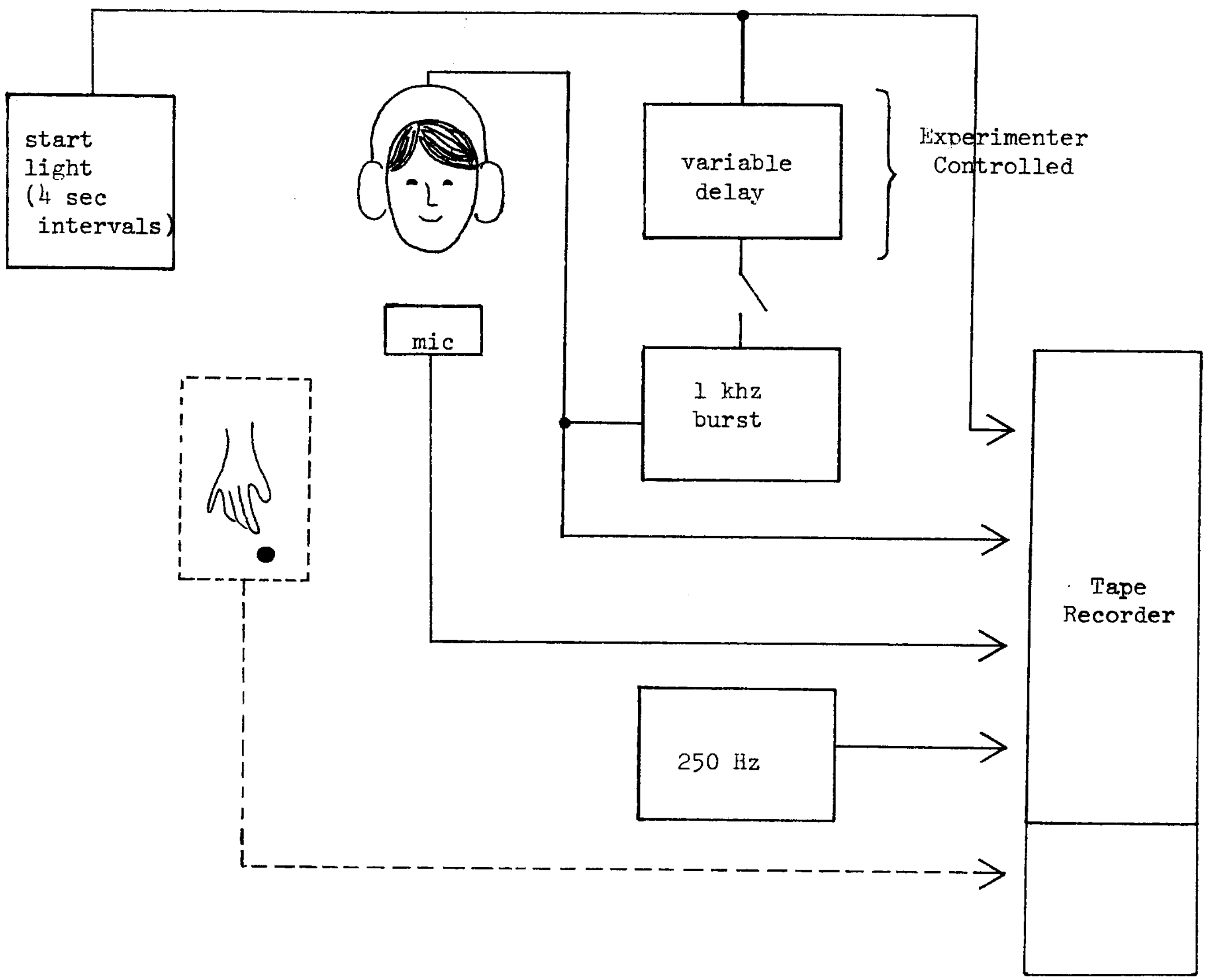


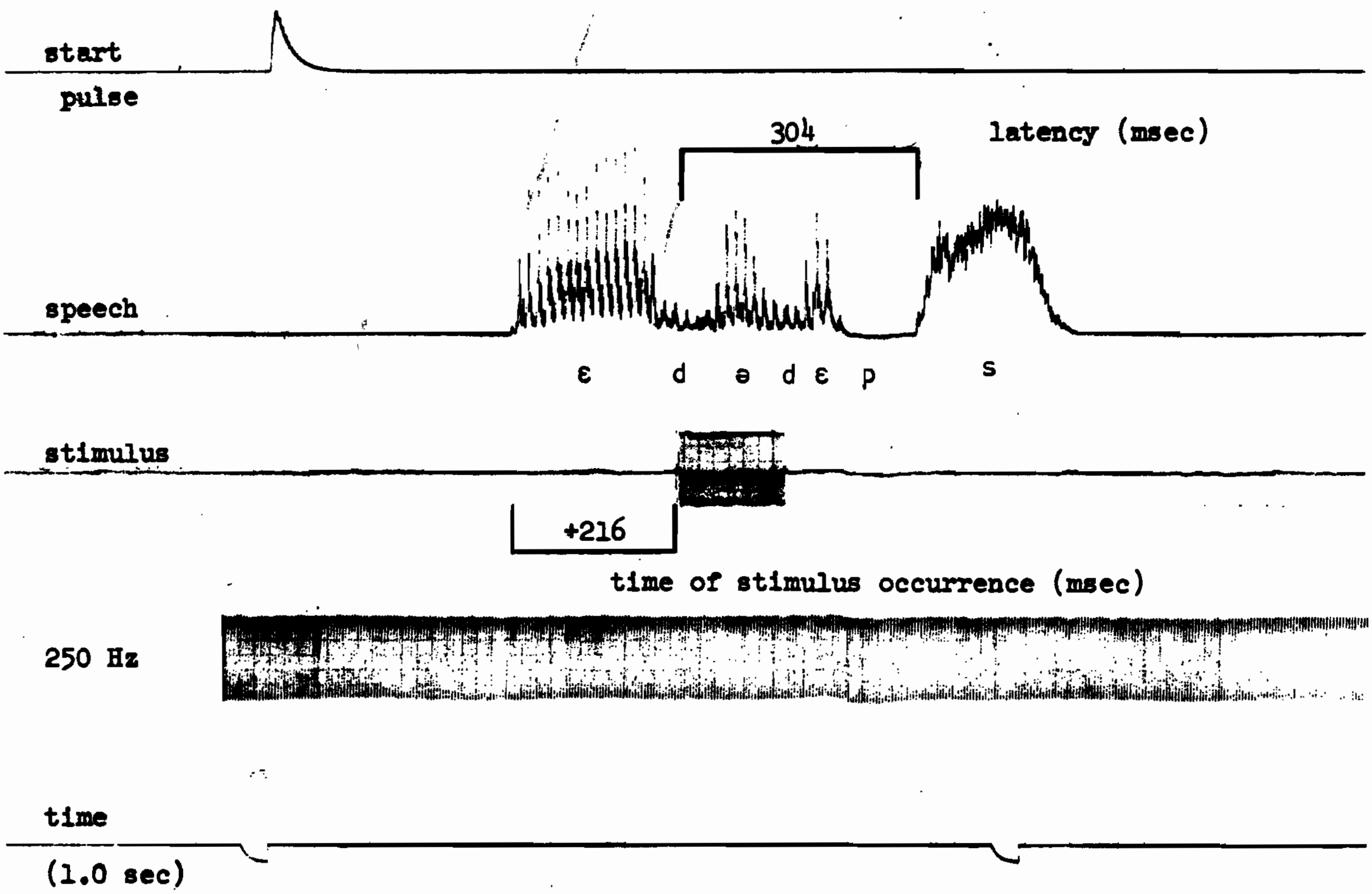
Figure 1. Arrangement of apparatus. The dashed box and line is as used in the second set of experiments.

The subject was instructed on hearing this stimulus to stop saying the original word and to say /ps/. The experimenter arranged the delay in the presentation of the stimulus in a way such that it appeared at all places within the word and sometimes just before the word began. Typically, the subject would repeat the phrase about sixty to seventy times in one session; and on about half of these occasions no stimulus would be presented.

We have run experiments using several different words and phrases, but the results we wish to report here were all obtained while repeating a single phrase "Ed had edited *Id*." This phrase was chosen because it consists basically of the same consonantal gestures with slightly differing vowels in between. It can also be stressed in a number of different ways.

Figure 2 shows a typical record we obtained when we played back the tape-recorded results onto an ink writer. The top channel shows the start pulse which occurred when the light went on. The next channel, which is a rectified integrated audio signal, shows that on this occasion the subject started saying the phrase about 300 msec. later. As may be seen from the next line, the stimulus occurred 216 msec. after he had begun speaking; and as the speech line shows, 304 msec. after this he had started saying the /s/ of the required /ps/ response. The lower two lines are time markers. We recorded a 250 KHz tone simultaneously with the other three channels of information. This signal was used as a time base in computer processing of the data.

So far we have recorded four subjects saying this particular phrase and have substantially the same results for all of them. Contrary to our initial expectation, as long as the stimulus occurred within the phrase, we did not find any particular part of the phrase where the subject found it more difficult to interrupt himself. The data for one subject are shown on Figure 3. It will be seen that throughout the phrase the points are on a fairly straight line. The data are somewhat noisy since subject's attention often varies so that some of the responses are far longer than the average. But it seems that there is no particular part of a word, and no particular part of a stress group which the subject consistently finds more difficult to interrupt than any other part. In all cases he manages to produce the /s/ of the /ps/ about 300 msec. after the stimulus; and he succeeds in interrupting his original utterance a little over 200 msec. after the stimulus. We should emphasize, however, that these are only preliminary results, and improvements in our experimental design and in our statistical procedures may enable us to find some phonological units which are more difficult to interrupt than others.



175

Figure 2. Records from the start pulse and the 4 channels of recorded data.

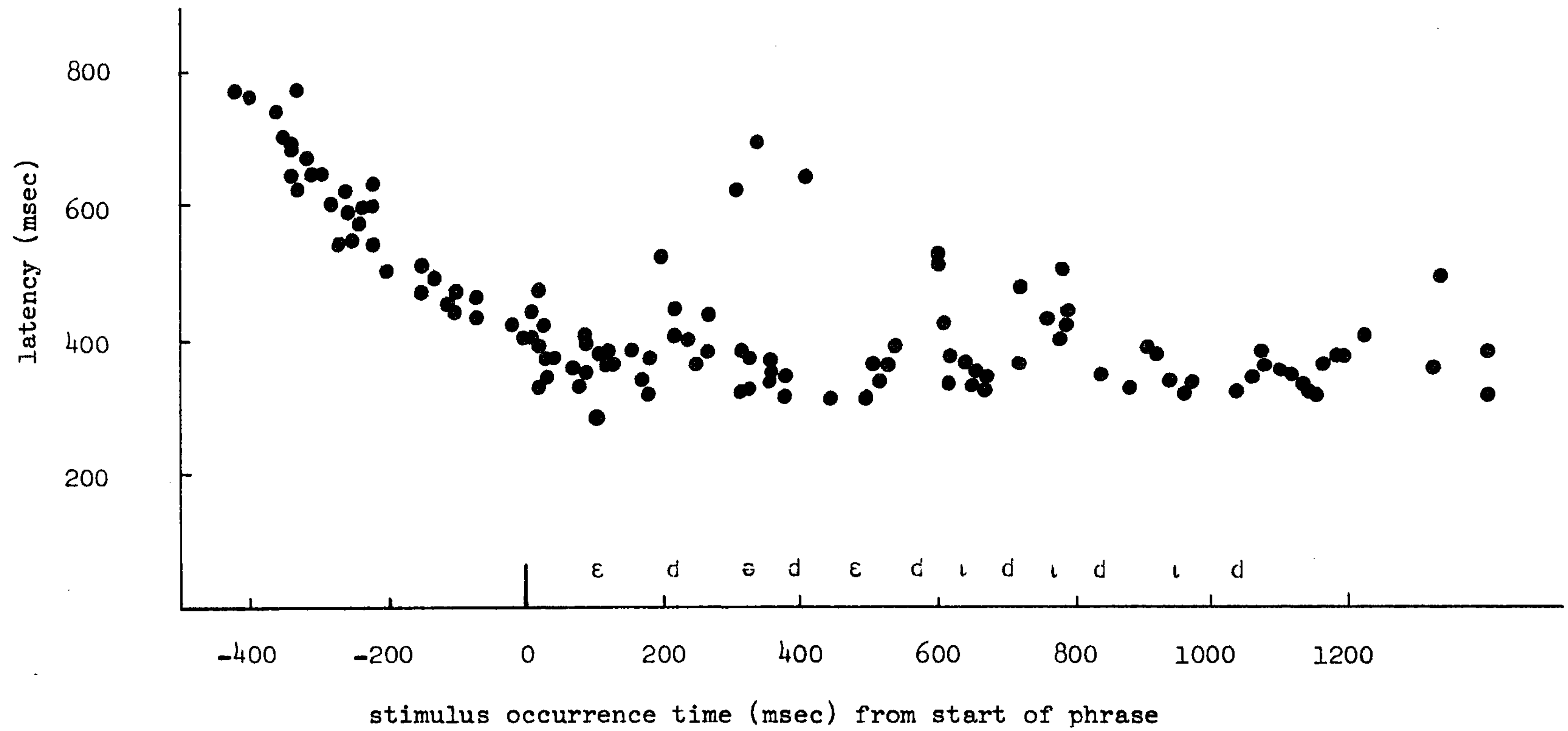


Figure 3. The relation between the latency of the /s/ speech response and the time of occurrence of the stimulus.

But there is a point at which the subject finds it much harder to respond by saying /ps/. This is if the stimulus occurs shortly before he was about to say the original utterance. When the stimulus occurs at this moment there is a striking rise in the latency.

This finding led us to hypothesize that the speaker might find it difficult to plan to say one utterance when he was in the middle of planning to say another utterance. But of course we could not tell whether this was a plausible hypothesis without first knowing whether the increased latency which occurs just before a phrase is due not to any difficulty in *producing* a different utterance but simply to increased difficulty in *perceiving* an incoming stimulus at that time.

In order to check out this latter possibility we conducted an experiment using the additions to the setup shown by the dashed lines in Figure 1. In this experiment the subject's task was not only to stop saying whatever it was he was saying when he heard the stimulus to interrupt himself as previously, but also to tap on a metal plate as soon as he heard this stimulus. The taps were recorded on another channel of the multichannel tape recorder.

The results of an experiment of this kind are shown in Figure 4. The solid points show the latency of the finger taps, and the arrowheads indicate that for the interruption of speech. The arrowhead is pointing upwards when the speech was interrupted after the tap had occurred and downwards when the speech was interrupted before the tap. It may be seen that on this occasion the subject had a slightly increased latency of the finger taps when the stimulus occurred just before the phrase was spoken. But the increase in latency for interrupting the speech was much greater at this time. We may note, incidentally, that interruption by stopping while saying a given phrase appears to be an equivalent task to interruption by stopping while saying a phrase and saying something else instead, since the latencies for interrupting speech in this experiment are very similar to those for interrupting and producing an alternative sound in the previous experiment. It may be seen later in the utterance most of the arrowheads are pointing downwards, indicating that the subject could interrupt his speech more quickly than he could tap with his finger when stimuli occurred at these times. It seems probable that the differences between interrupting speech and tapping with the finger at these moments are of the right order of magnitude to be due to the differences in conduction time between the muscles of the larynx and those of the finger.

When this experiment was repeated a number of times it became apparent that there was a learning phenomenon involved for the tap

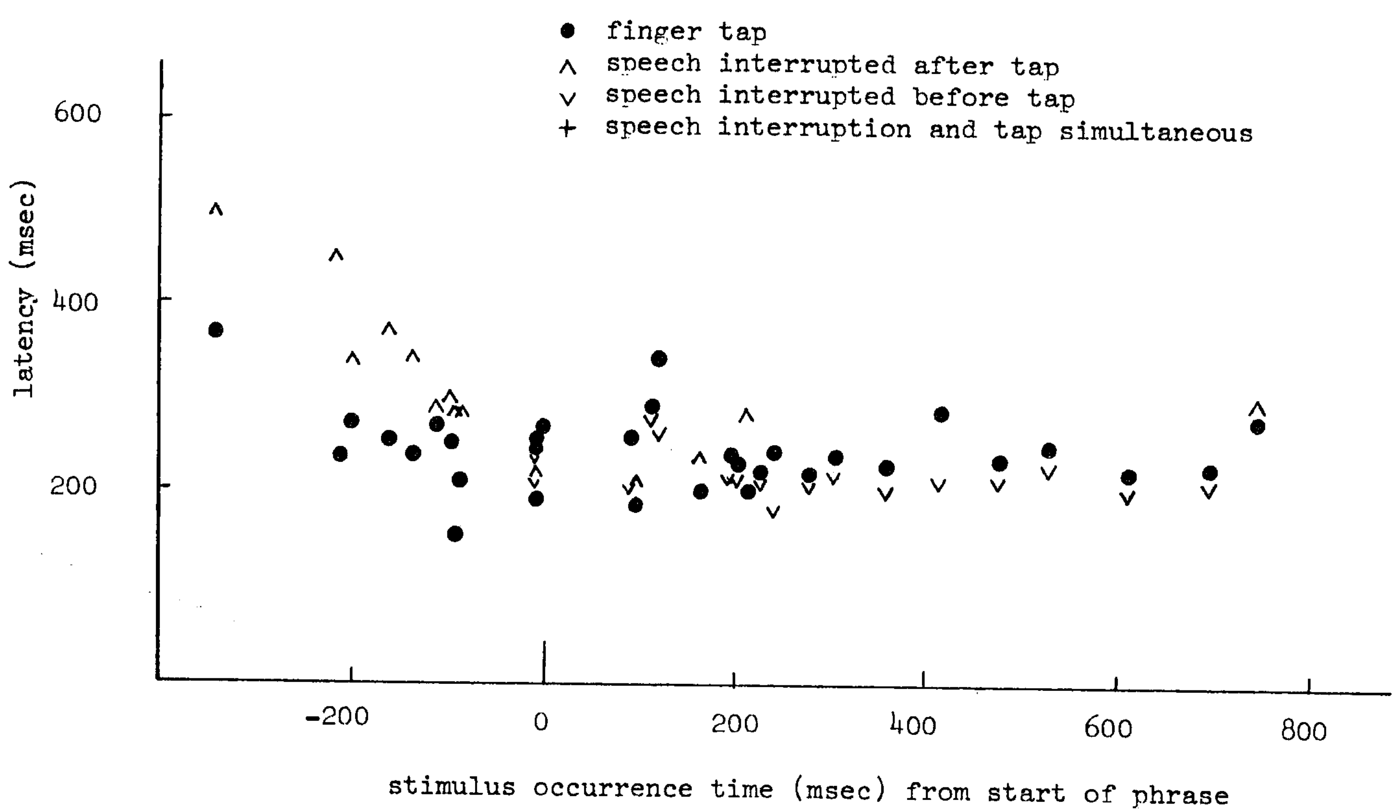


Figure 4. The relation between the speech response, the finger tap, and the time of occurrence of the stimulus. When the subject was not very experienced in this task.

in a way that was not true for the speech. Figure 5 shows a later trial for the same subject. It may be seen that on this occasion there is no increased latency for the tap when the stimulus occurs before the utterance; but the increased latency is still there for interrupting the speech. In addition there is still a crossover in the latencies for the two tasks; towards the end of the utterance the taps are occurring slightly after the speech had been interrupted.

Considering the data in this figure it seems that our hypothesis is valid and that the reason for the increased latency in the interruption of speech is *not* that the subject does not *perceive* the incoming stimulus when it occurs just before the speech; but instead that he finds it difficult to change what he is about to say when the stimulus occurs at this time. We would imagine that what happens is that when a speaker decides to say a phrase, something which we may call the speech organizing mechanism gets to work and stores the information necessary for the motor system in some kind of buffer. It is difficult to make an interruption at the time when this organization and storing is going on. But when the speaker is retrieving information from the buffer and producing a phrase, then his interrupt switch is on and it is an easier task. While one is talking one is, after all, continuously monitoring oneself to see if any changes are necessary. The obvious candidate for the location of this buffer is the cerebellum.

As we said at the beginning this is a very preliminary report; and we have many additional experiments in mind. For instance we intend varying the speech material in systematic ways. In some trials we have already tried slightly longer phrases in the hopes of finding a second point within the phrase where the speaker had to organize the remainder of the phrase. So far we have been unsuccessful when using a phrase of twice the length of "Ed had edited *Id*." It may be that subjects who are merely repeating a phrase are capable of storing in a buffer considerable lengths of speech at one time. Another way in which we intend varying the utterance is to try making it harder for the subject to say the phrase. Thus we might try getting him to say phrases in French. We would predict that under these circumstances the rise in latency just before the phrase would be even greater.

Lastly an intriguing possibility is to compare the speech organizing mechanism with other mechanisms such as that which must be used by a pianist repeating a short musical phrase. It might be that a pianist would find it considerably harder to interrupt his finger movements by making a *tap* when the stimulus for him to do so occurred just before he was about to begin the phrase. But he might have no increased

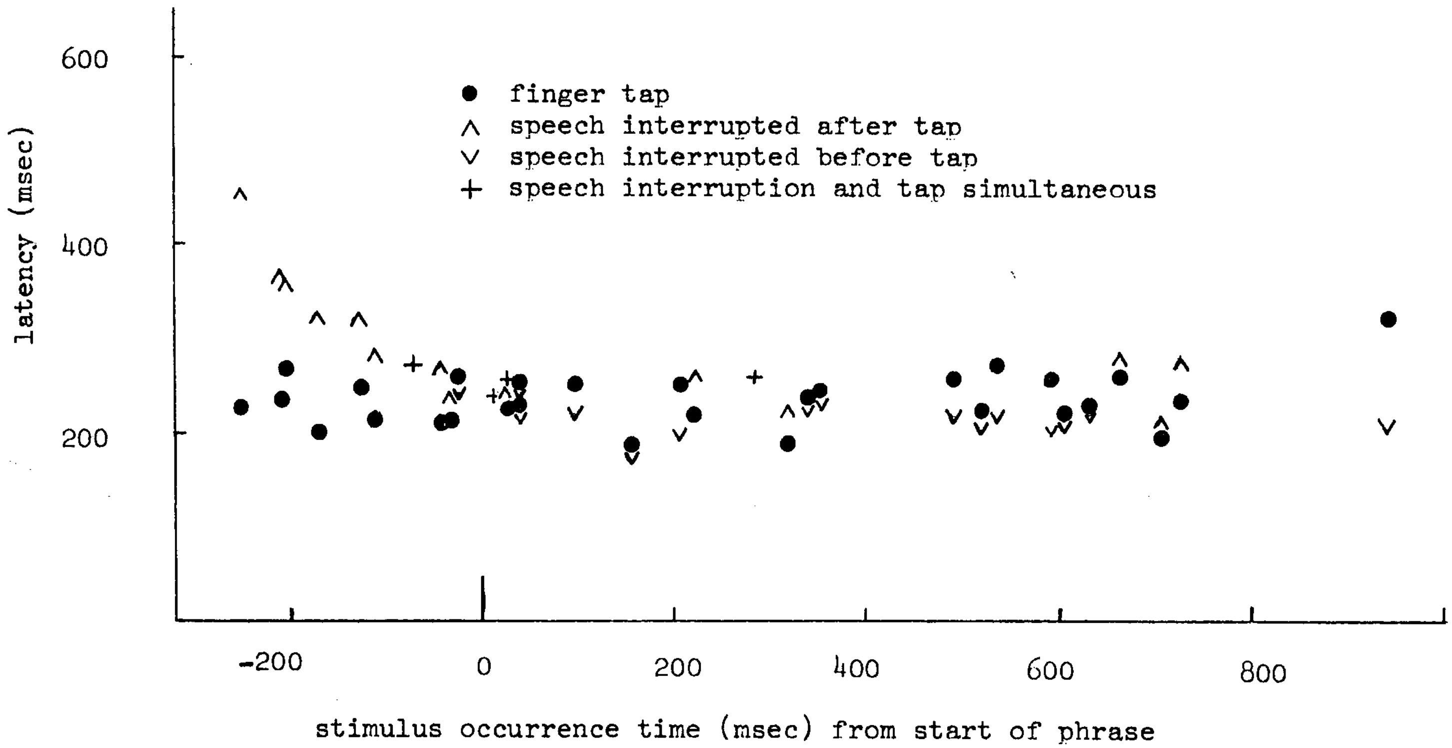


Figure 5. The relation between the speech response, the finger tap, and the time of occurrence of the stimulus. When the subject has become experienced in this task.

difficulty in making a verbal response to a stimulus. Thus if his task were to say "ba" on hearing a stimulus, he might be able to do that very rapidly, even at times when he found it very difficult to alter his intended finger movements by making a tap. This would show that the speech organizing mechanism and the finger movement organizing mechanism were indeed separate entities, each capable of being hampered by having to make an alteration while in their state of greatest activity.

A Basic Linguistic Phonetics Laboratory

Peter Ladefoged

There is no doubt that any adequate program in linguistics must include training in phonetics. By training in phonetics I do not mean just developing the ability to recognize, imitate and write down what an informant says. These are, of course, indispensible skills for any linguist wanting to handle real data; one must be able to transcribe and categorize sounds. But nowadays a properly trained linguist should be able to do a great deal more. He should be able to describe the sounds of a language not just in terms of symbols or sets of features which he cannot understand, but meaningfully, in terms of observable physiological or acoustic facts; and he should know enough about the production and perception of speech sounds to be able to make an intelligent evaluation of some current arguments in phonological theory. He should, for example, be able to assess the implausibility of some of the features proposed by Chomsky and Halle (1968). This kind of training demands at least exposure to the basic facilities available in a phonetics laboratory; and in all probability some of the students in a linguistics program will wish to get their hands dirty with data and specialize in this area.

Presumably no institution would try to set up a phonetics laboratory without having, or intending to have, a phonetician on the faculty. This person would have his own areas of special interest, and would no doubt wish to design the laboratory to meet his special needs. However, it is fair to say that certain basic equipment would be found in any adequate phonetics laboratory. This note is directed primarily to those who are considering the provision of such a laboratory. It discusses the estimated cost, lists some of the basic equipment, and suggests a set of priorities.

The first point to be noted in considering the cost of a phonetics laboratory is that it is not a one-time expenditure. Commitment to a laboratory, like that to a library, must be on-going. It is true that the heaviest expenditure is probably at the beginning. But electronic equipment wears out quickly (five years is a realistic time for writing it off); and new techniques and new instruments are continually coming out. If it is to be up to date, a laboratory must be provided with an annual sum for the purchase of equipment, over and above the sum required for supplies and expenses.

There is also another very important recurrent expense which must be considered before deciding to set up a laboratory. Funds must be provided for at least a part-time technician. His major function is to maintain the equipment; there is no point in having expensive specialized equipment if it is not kept in good order so that it is always available. A second function for a technician is the development of new apparatus; much of the equipment is not commercially available, and is most conveniently constructed by the person who will be responsible for keeping it going. Thirdly, although advanced students can operate machines for themselves, there are many beginning students who need help even in such simple things as finding the tapes and operating a tape recorder; and who of course need to be assisted in obtaining more complicated instrumental records to illustrate points in term papers etc. In addition members of the faculty often need help in using the apparatus for their research projects. Adequate technical assistance is vital for the proper functioning of a laboratory.

Given the clear understanding that technical assistance will be provided, it is possible to consider the purchase of equipment for a linguistic phonetics laboratory. The suggestions made below are summarized in Table 1, which also indicates the priority which might be assigned to the various items. Those marked priority 1 need to be bought in the first year of a laboratory's existence. The purchase of items marked priority 2 could be put off for a year or so, if money was not available right away. Those marked priority 3 are of less general importance and might be selected only if the laboratory developed in such a way that the need for them became apparent.

A. <u>Audio recording equipment</u>

1. The first requirement of any phonetics laboratory is that it should be able to make high-quality recordings. Since there are often occasions when a commentary has to be added to an existing recording, preferably without being actually superimposed, a two-channel tape recorder is desirable. This will also provide for the possibility of simultaneously recording two kinds of data (e.g. an utterance and the electromyographic activity of one of the muscles involved in its production, see below).

2. In addition to a good tape recorder, the making of high-quality recordings requires some kind of studio or sound-insulated room. This room should preferably be big enough to conduct experiments in. It should also provide electric shielding for experiments involving very small voltages (emg, see 15 below).

3. Microphones: it is useful to have both an omnidirectional and a cardiod directional type.

4. Being able to make good recordings is of little use unless it is possible to listen to them conveniently and get the maximum information out of them. At least two pairs of high-quality earphones and a good loudspeaker are necessary.

5. In addition a convenient way of repeating the playback of short phrases is very desirable. The criteria for a good listening system include being able to select any section of a recording for repetition over and over again by means of controls that are so simple that one can operate them with one's left hand (because the right hand is holding a pencil) and with one's eyes shut (because one is listening). In my experience a good loop repeater system is far more useful (and far more used) than any other listening device, such as a speech-stretcher or time-expansion device. Hearing a tape recording slowed down, even if the frequencies have been electronically restored to their original values, is not usually very helpful; the sounds are different — they are not as in real speech.

   Tape loop repeater systems of the kind required are not commercially available. But any competent technician can make one from two tape recorders. All that it is necessary to do is to connect them (as shown in Figure 1) so that the master recording is being continually copied, played back, and then erased from a loop on the slave recorder. Then if the erase system of the slave recorder is coupled with the start/stop switch of the master recorder, stopping the master recorder will result in no new material being copied onto the slave loop, the existing material on the loop will not be erased, and the playback function of the slave loop will continue. It is well worth having a system of this kind permanently connected so that it is available for use at any time. If the switches are properly arranged it also makes a convenient tape copying system.

6. One (or, better, two) portable tape recorders for fieldwork, complete with microphone, carrying case, battery charger, etc.

B. <u>Acoustic analysis</u>

7. The sound spectrograph is still the single most important piece of equipment in a phonetics laboratory. There are models produced by two different manufacturers available. They differ in the same way as a Rolls Royce and a VW; both will get you where you want to go, but one will do so with more elegance (and at a greater cost) than the other. For laboratories with little money the Kay Sona-Graph is satisfactory; but the Sound Spectrograph manufactured by Voiceprint Laboratories (P.O. Box 835, Somerville, N.J. 08876) is definitely a better machine.

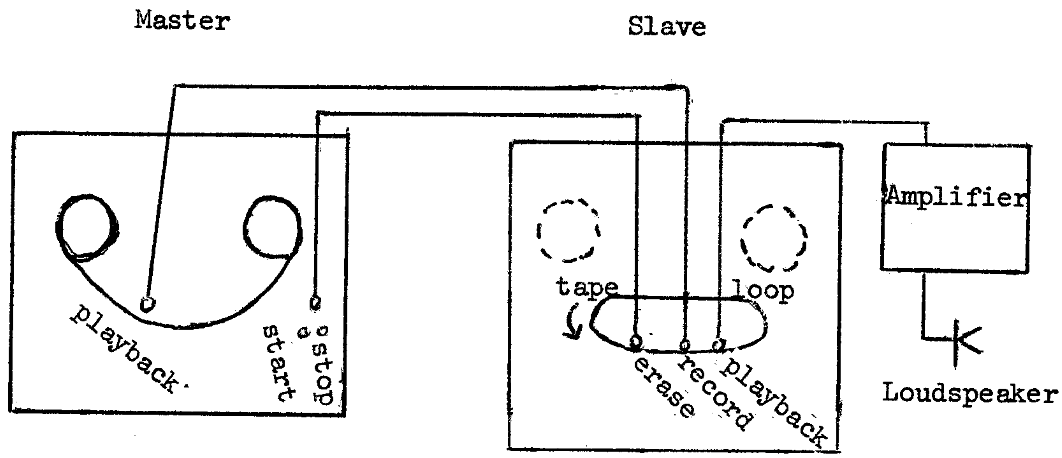


Figure 1. A tape loop repeater system (based on a design originally suggested by J. F. K. Anthony, University of Edinburgh). Note that the Master tape recorder is used only in the playback mode. The loop on the Slave recorder is in continuous motion. The reels on the Slave recorder are shown by dashed lines to indicate that the spooling motors are available to be switched in if the system is required as a tape copying facility.

8. A pitch meter (fundamental frequency extraction system) is obviously very useful in prosodic studies.

9. An intensity meter, particularly one with two channels, one of which can be high pass filtered, is useful in studies involving the segmentation of long sections of speech.

10. An audio frequency filter is needed, both for combination with the intensity meter above, and in many experiments on auditory perception.

11. Frequency and time interval measurements may be made conveniently with an electronic counter.

12. Nowadays it is possible to purchase Terminal Analog Speech Synthesizers which will produce reasonable quality speech when supplied with time-varying voltages corresponding to acoustic parameters. The problem with these devices is that the only adequate way of providing proper control voltages is through a computer (see below). However, even if it is not possible to provide a connection to a computer, it is probably worth having a speech synthesizer for the synthesis of comparatively steady state sounds.

C. <u>Physiological investigations</u>

13. Two pressure transducers for use in recording oral or subglottal air pressures.

14. Palatography apparatus for recording contacts between articulators. The Polaroid dental camera is very convenient for this purpose, requiring only a powder spray in addition.

15. Electromyography preamplifiers for use in recording muscle action potentials. (The power amplifiers used in emg work might be those in the recording equipment; see 19 and 21 below).

16. Laryngoscopy lights and mirrors for viewing the vocal cords.

17. (a) Photo-electric glottography and/or (b) electroglottography; these are two different systems for use in recording (a) the area between the vocal cords in the horizontal plane; (b) the area of contact between the vocal cords in both the horizontal and the vertical plane. The former is the more generally useful.

18. Electroaerometer for recording rate of flow of air in and out of the mouth and nose.

D. <u>Recording instruments</u>

19. An oscilloscope is needed to display the output of many of the instruments listed above. It also makes a convenient high-gain DC or AC amplifier, for use in, e.g., electromyography experiments. A model with a time scale expansion is useful in detailed investigations of audio-frequency waveforms; the waveform of a loop of tape may be expanded as required. In addition a good oscilloscope is required as a basic instrument for the measurement of voltage and frequency, in experimental studies, and in calibration, maintenance, and construction of other equipment.
20. A Polaroid oscilloscope camera is a useful accessory, particularly for photographing waveforms and other events which last only a short time.

21. Permanent records of the outputs of many instruments (e.g. pitch and amplitude displays, pressures, electromyographic data, etc.) are best obtained through an ink writer with a high-frequency response, such as an oscillomink. The usefulness of this device cannot be over-estimated.

E. Computer

22. There is no doubt that any beginning laboratory should think of obtaining its own computer as soon as possible. A small computer which could be used for controlling a speech synthesizer, processing physiological data, etc., and (most importantly) giving students experience in using a computer costs less than $10,000. A more versatile facility including mag-tape units, 'scope display, and 8,000 words of core memory, which is more than adequate for most phonetic research, costs about $35,000.

F. Development, testing and maintenance equipment

23. Voltmeter

24. Ohmeter / voltmeter / ammeter

25. Signal generator

26. Workbench, tools, etc.

Reference

Chomsky, N. and M. Halle (1968), *The Sound Pattern of English,* Harper and Row, New York.

Table 1. Equipment

Priority	Item	Function	Model and Supplier	Estimated Cost
1	1	Two channel studio recorder and microphone preamplifier	Ampex AG350 (Ampex Corp. 410 Broadway Redwood City, Calif.)	$ 2,400
1	2	Sound-insulated room installed, with forced ventilation	IAC Model 402A (Industrial Acoustics Corp. 380 Southern Blvd. Bronx, New York, 10454	3,000
1	3	Directional microphone	MD 421 N (Sennheiser Electronic Corp.	120
2		Omnidirectional microphone	500 Fifth Ave., New York N.Y. 10036)	100
1	4	Earphones	Superex SM600 (Superex Electronics Corp.	30
2		Second pair of earphones	4-6 Redford Place Yonkers, New York)	30
1		Amplifier and loudspeaker		250
1	5	Two additional tape recorders to be coupled together forming a loop repeater system		600
1	6	Portable tape recorder for fieldwork	Uher 4000	425
2		Second portable tape recorder	Nagra (Kudelski 6 ch. de l'Etang Paudex-Lausanne, Switzerland)	1,300
1	7	Sound spectrograph	Sona-Graph 6061A	3,130
1		Scale magnifier/amplitude display	Kay 6076C (Kay Electric Company Maple Avenue Pine Brook, New Jersey,07058)	525
1	8	Pitchmeter	B. Frøkjaer-Jensen (B. Frokjaer-Jensen Vestre Paradisvej, 46 Dk-2840 Holte, Denmark)	590

Priority	Item	Function	Model and Supplier	Estimated Cost
1	9	Two-channel intensity meter	(as 8, but this model is too expensive)	$ 800
2	10	Audio frequency filter	(as 8)	625
2	11	Electronic timer	HP 52223L (Hewlett Packard 1101 Embarcadero Road Palo Alto, Calif. 94303)	1,275
2	12	Terminal Analog Speech Synthesizer	Digital Speech Synthesizer 4516 (Rockland Systems Corp. 131 Erie St. Blauvelt, New York. 10913)	8,000
1	13	Pressure Transducer	Model 420.02, 0.25 psid (Simmonds Precision, Rockingham Rd. Bellows Falls, Vermont 05101)	310
2		Second pressure transducer	Model 420.03,2.0 psid	270
1	14	Palatography apparatus	Polaroid CU-5 Dental Camera #88-1, 88-3, 88-8, 88-33	350
2	15	Three channel electromyography preamplifiers	(as 8)	550
1	16	Laryngoscopy equipment (Light sources, mirrors, etc.)	Mr. J. Anthony, University of Edinburgh, Scotland	150
3	17	Photo-electric glottograph	(as 8)	775
3		Electro - glottograph	(as 8)	695
3	18	Electroaerometer	(as 8)	1,410
1	19	Dual-beam Oscilloscope	Tektronix 502 (Tektronix Inc. S. W. Millikan Way P. O. Box 500, Beaverton,Oregon)	1,150
3	20	Oscilloscope camera	Tektronix C-27-547 (as 19)	450
1	21	6 channel Oscillomink	M290-A1 (Siemens America Inc. 350 Fifth Ave. New York,N.Y.)	6,500
3	22	Computer	8K PDP 12 (Digital Equipment Corp. 146 Main Street Maynard, Mass.,01754)	(35,000)

Priority	Item	Function	Model and Supplier	Estimated Cost
1	23	Multifunction meter (DC, AC, Ohms)	HP 427A (as 11)	$ 225
1	24	General-purpose Ohmeter/ voltmeter/ammeter	Simpson 260	60
1	25	Audiofrequency generator	HP 201C (as 11)	250
1	26	Workbench, tools, etc.		300

Table 2. Summary of annual costs for a basic linguistic phonetics lab (no computer, few luxuries)

	Year 1	2	3	4	5
New equipment	21,000	6,000	2,000	2,000	2,000
Replacement	---	---	1,000	4,000	4,000
Supplies and parts	500	1,000	1,000	1,000	1,000
Technician	5,000	10,000	10,000	10,000	10,000
Total	26,500	17,000	14,000	17,000	17,000

Index of publications by members of the UCLA Phonetics Laboratory,

1963 - 1970

Compiled by Leon Jacobson

George D. Allen (September 1967 - August 1970)

1968 "Experiments on the rhythm of English speech." Working Papers in Phonetics 10:42-46. Abstracted as "Experiments on speech rhythm." *Journal of the Acoustical Society of America* 44:377.

"New kind of laryngeal whistle." *Journal of the Acoustical Society of America* 44:365.

"On testing for certain stress-timing effects." Working Papers in Phonetics 10:47-59.

"The place of rhythm in a theory of language." Working Papers in Phonetics 10:60-84.

1970 "Acoustic intensity and vocal effort as cues for the loudness of speech." *Journal of the Acoustical Society of America* 48:95.

"The location of rhythmic stress beats in English: An experimental study." Working Papers in Phonetics 14:80-132. To be published in *Language and Speech*.

"Temporal structure in speech production." *Journal of the Acoustical Society of America* 47:58.

James Anthony (July 1963 - October 1964)

1964 "Replica of the vocal tract." Working Papers in Phonetics 1:10-14.

1965 "Construction of a replica of the vocal tract." Working Papers in Phonetics 2:5-17.

Kay Atkinson (February 1967 - June 1968)

1968 "Language identification from non-segmental clues." Working Papers in Phonetics 10:85-89. Abstracted in *Journal of the Acoustical Society of America* 44:378.

J. C. Catford

1968 and Peter Ladefoged. "Practical phonetic exercises." Working Papers in Phonetics 11. 22 pp.

Elizabeth Dunstan (July 1964 - December 1964)

1965 "Electromyographic recordings in the mouth." Working Papers in Phonetics 2:55-57.

Victoria A. Fromkin (July 1963 - present)

1964 "Lip positions in American English vowels." *Language and Speech* 7:215-25.

"On system-structure phonology." *Language* 41:601-09.

"Parameters of lip positions." Working Papers in Phonetics 1:15-21. Abstracted in *Journal of the Acoustical Society of America* 36:1037.

1965 "Location of lip muscles by means of electromyography." Working Papers in Phonetics 2:51-54.

"Some phonetic specifications of linguistic units: An electromyographic investigation." Working Papers in Phonetics 3. 184 pp.

(See Peter Ladefoged, "Electromyography in speech research.")

1966 "Neuro-muscular specification of linguistic units." *Language and Speech* 9:170-99.

"Relationship between linguistic units and motor commands." *Journal of the Acoustical Society of America* 39:1219.

Review of *Set theory and syntactic description* by William S. Cooper. *International Journal of American Linguistics* 32:195-98.

Victoria A. Fromkin (continued)

"Some requirements for a model of performance." Working Papers in Phonetics 4:19-39. Revised and published in 1968 as "Speculations on performance models."

and Peter Ladefoged. "Electromyography in speech research." *Phonetica* 15:219-42.

1967 and Walter J. Dowling. "Effects of the location of stress and constituent breaks on the perception of clicks." Working Papers in Phonetics 7:85-90.

1968 "The computer as a research tool in the construction of models of linguistic performance." In *Proceedings of the International Federation for Information Processing Conference* (Edinburgh), pp. 1599-602.

Review of *Elements of general phonetics* by David Abercrombie. *Phonetica* 18:242-44.

"Role of physical phonetics in a theory of universal phonetics." *Journal of the Acoustical Society of America* 44:378.

"Speculations on performance models." *Journal of Linguistics* 4:47-68.

"Universal phonetics as an explanatory theory." *Reports of the Sixth International Congress on Acoustics* (Tokyo) C:37-40. Also published in Working Papers in Phonetics 10:169-71.

"The what and why of intrinsicalness." Working Papers in Phonetics 10:156-68.

and John Ohala. "Laryngeal control and a model of speech production." Working Papers in Phonetics 10:98-110. Published in the Preprints of the Speech Symposium (Kyoto) C3:1-5.

(See Peter Ladefoged, "Experiments on competence and performance.")

(See Paul Schachter, "A generative phonology of Akan: Akuapem, Asante, and Fante.")

Victoria A. Fromkin (continued)

1969 and D. Lloyd Rice. "An interactive phonological rule testing system." Preprint 53 of the International Conference on Computational Linguistics (Stockholm).

1970 "The concept of 'naturalness' in a universal phonetic theory." *Glossa* 4.1 (in press).

"In defense of systematic phonemics." *Journal of Linguistics* (in press).

"A reply to 'The phonetic framework of generative phonology.'" Working Papers in Phonetics 14:34-39.

"Tips of the slung -- or -- to err is human." Working Papers in Phonetics 14:40-79. To be published in *Language* as "The non-anomalous nature of anomalous utterances."

Paul L. Garvin

1963 and Peter Ladefoged. "Speaker identification and message identification in speech recognition." *Phonetica* 9:193-99.

S. Robert Greenberg (September 1964 - September 1969)

1969 "An experimental study of certain intonation contrasts in American English." Working Papers in Phonetics 13. 129 pp.

Kerstin Hadding (at that time Kerstin Hadding-Koch; January 1968 - December 1968)

1970 Minoru Hirano and Tim Smith. "Electromyographic study of lip activity in Swedish CV:C and CVC: syllables." Working Papers (Lund University Phonetics Laboratory) 1:1-9. Also published in Working Papers in Phonetics 14:133-39.

Richard Harshman (May 1966 - present)

1967 and Peter Ladefoged. "The LINC-8 in speech research." Working Papers in Phonetics 7:57-68. Revised and published in *DECUS Proceedings*, Fall 1967, as "The LINC-8 in research on speech."

Richard Harshman (continued)

1968 (See John Ohala, "Photoelectric methods of transducing lip and jaw movements in speech.")

1969 (See Shizuo Hiki, "Speech synthesis by rules with physiological parameters.")

1970 "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis." Working Papers in Phonetics 16. 84 pp.

Frank Heny (October 1966 - September 1968)

1967 "Non-binary phonological features." Working Papers in Phonetics 7:91-120.

"Towards the separation of classificatory and phonetic features." Working Papers in Phonetics 7:121-24.

Shizuo Hiki (April 1968 - April 1970)

1968 Rick Ratcliffe, Stan Hubler, and Pete Metevelis. "Notes on LASS circuitry." Working Papers in Phonetics 10:12-41.

(See John Ohala, "Photoelectric methods of transducing lip and jaw movements in speech.")

1969 and Richard Harshman. "Speech synthesis by rules with physiological parameters." *Journal of the Acoustical Society of America* 46:111.

1970 "Control rule of tongue movement for dynamic analog speech synthesis." *Journal of the Acoustical Society of America* 47:85.

Kenneth Hill (September 1963 - June 1964)

1964 "The musculature of the tongue." Working Papers in Phonetics 1:22-28.

Minoru Hirano (July 1967 - August 1968)

1967 and John Ohala. "Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles." Working Papers in Phonetics 7:35-45. Revised and published as Hirano and Ohala 1969.

Minoru Hirano (continued)

John Ohala and Tim Smith. "Current techniques used in obtaining emg data." Working Papers in Phonetics 7:20-24. Incorporated in Hirano and Ohala 1969.

and Tim Smith. "Electromyographic study of tongue function in speech: A preliminary report." Working Papers in Phonetics 7:46-56.

(See John Ohala, "Control mechanisms for the sequencing of neuromuscular events in speech.")

(See John Ohala, "An experimental investigation of pitch change in speech.")

(See John Ohala, "Studies of pitch change in speech.")

1968 William Vennard and John Ohala. "The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation." Working Papers in Phonetics 10:111-25. Abstracted in *Journal of the Acoustical Society of America* 44:354. Revised and published as Hirano, Ohala, and Vennard 1969.

(See John Ohala, "An electromyographic study of laryngeal activity in speech and singing.")

(See Tim Smith, "Experimental investigations of the muscular control of the tongue in speech.")

1969 and John Ohala. "Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles." *Journal of Speech and Hearing Research* 12:362-73.

John Ohala and William Vennard. "The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation." *Journal of Speech and Hearing Research* 12:616-28.

1970 William Vennard and John Ohala. "Regulation of register, pitch and intensity of voice." *Folia Phoniatrica* 22:1-20.

(See Kerstin Hadding, "Electromyographic study of lip activity in Swedish CV:C and CVC: syllables.")

Reuben Hofshi (March 1968 - June 1969)

1968 (See D. Lloyd Rice, "An interactive phonological rule tester.")

Stanley Hubler (September 1964 - February 1968)

1967 "A high input impedance electromyography preamplifier." Working Papers in Phonetics 7:25-34.

(See Shizuo Hiki, "Notes on LASS circuitry.")

(See John Ohala, "Photoelectric methods of transducing lip and jaw movements in speech.")

1968 "Quantifying muscular activity in the analysis of human speech." *DECUS Proceedings*, Fall 1968:55-59.

Chin-Wu Kim (February 1964 - September 1966)

1965 (See Peter Ladefoged, "Human, replica, and computer generated formants.")

1966 "The linguistic specification of speech." Working Papers in Phonetics 5. 126 pp.

Peter Ladefoged (September 1962 - present)

1963 Review of *English phonetics* by Yao Shen. *Language* 39:581-84.

"Some physiological parameters in speech." *Language and Speech* 6:109-19.

(See Paul L. Garvin, "Speaker identification and message identification in speech recognition.")

1964 Comment on "Evaluation of methods of estimating sub-glottal air pressure." *Journal of Speech and Hearing Research* 7:291-92.

A phonetic study of West African languages. First edition. West African Language Monograph, 1. London: Cambridge University Press.

"Physiological characterization of speech." Working Papers in Phonetics 1:2-9. Revised and published in *Language and Speech* 7:205-14, as "Some possibilities in speech synthesis.

Peter Ladefoged (continued)

Review of *Manual of articulatory phonetics* by William A. Smalley. *International Journal of American Linguistics* 30:424-27.

1965 "The nature of general phonetic theories." In *Linguistics and Language Study*, edited by Charles W. Kreidler, pp. 27-42. Georgetown University Monograph Series on Languages and Linguistics, 18. Washington: Georgetown University Press.

Review of *Acoustical characteristics of selected English consonants* by Ilse Lehiste. *Language* 41:332-38.

Review of *Phonetics: History and interpretation* by Elbert R. Moses, Jr. *International Journal of American Linguistics* 31:182-83.

and Victoria Fromkin. "Electromyography in speech research." Working Papers in Phonetics 2:37-50. Published with slight revision as Fromkin and Ladefoged 1966.

and Chin-Wu Kim. "Human, replica, and computer generated formants." Working Papers in Phonetics 2:18-26.

1966 "New research techniques in experimental phonetics." *Study of Sounds* (Phonetic Society of Japan) 12:84-101.

(See Victoria Fromkin, "Electromyography in speech research.")

(See Norris McKinney, "Terminal analog speech synthesizer.")

1967 "Linguistic phonetics." Working Papers in Phonetics 6. 98 pp.

"Research possibilities in phonetics." Research Institute of Logopedics and Phoniatrics Annual Bulletin 1:31-34.

Three areas of experimental phonetics. London: Oxford University Press.

and Ralph Vanderslice. "The 'voiceprint' mystique." Working Papers in Phonetics 7:126-42. Abstract published in *Journal of the Acoustical Society of America* 42:1164, as Vanderslice and Ladefoged.

Peter Ladefoged (continued)

(See Richard Harshman, "The LINC-8 in speech research.")

1968 "Linguistic aspects of respiratory phenomena." *Annals of the New York Academy of Sciences* 155.1:141-51.

A phonetic study of West African languages. Second edition. London: Cambridge University Press.

and Victoria Fromkin. "Experiments on competence and performance." *IEEE Transactions on Audio and Electroacoustics* AU-16:130-36.

(See J. C. Catford, "Practical phonetic exercises.")

1970 "Clicks." *Encyclopaedia Brittanica* C:911.

"The measurement of phonetic similarity." *Statistical Methods in Linguistics* 6:23-32.

"The phonetic framework of generative phonology." Working Papers in Phonetics 14:25-32. Revised and to be published in *For Eli Fischer-Jorgensen*, as "The limits of phonology." Copenhagen: Akademisk.

and Raymond Silverstein. "Interruptibility of speech." *Journal of the Acoustical Society of America* 48:102-03.

(See John Ohala, "Further investigation of pitch regulation in speech.")

Mona Lindau (September 1967 - present)

1968 "Juncture and the topic-comment distinction, a pilot study." Working Papers in Phonetics 10:90-97.

Norris McKinney (October 1965 - February 1966)

1966 Marcel Tatham and Peter Ladefoged. "Terminal analog speech synthesizer." Working Papers in Phonetics 4:10-18.

Peter Metevelis (August 1967 - July 1969)

1968 (See Shizuo Hiki, "Notes on LASS circuitry.")

John Ohala (February 1965 - May 1969)

1965 and Ralph Vanderslice. "Photography of states of the glottis." Working Papers in Phonetics 2:58-59.

1966 "A new photo-electric glottograph." Working Papers in Phonetics 4:40-53. Abstracted as "Studies of variations in glottal aperture using photoelectric glottography." *Journal of the Acoustical Society of America* 41:1613.

1967 and Minoru Hirano. "Control mechanisms for the sequencing of neuromuscular events in speech." Preprints of the Conference on Speech Communication and Processing (MIT, 6-8 November 1967):164-69. Abstracted in Working Papers in Phonetics 7:10.

and Minoru Hirano. "An experimental investigation of pitch change in speech." *Journal of the Acoustical Society of America* 42:1208-09.

and Minoru Hirano. "Studies of pitch change in speech." Working Papers in Phonetics 7:80-84.

(See Minoru Hirano, "Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles.")

1968 Shizuo Hiki, Stanley Hubler, and Richard Harshman. "Photoelectric methods of transducing lip and jaw movements in speech." Working Papers in Phonetics 10:135-44. Abstracted (1969) as "Transducing jaw and lip movements in speech." *Journal of the Acoustical Society of America* 45:324.

Minoru Hirano and William Vennard. "An electromyographic study of laryngeal activity in speech and singing." Working Papers in Phonetics 10:126-34. Also published in the Preprints of the Sixth International Congress on Acoustics (Tokyo):B-1-2.

(See Victoria Fromkin, "Laryngeal control and a model of speech production.")

(See Minoru Hirano, "The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation.")

1970 "Aspects of the control and production of speech." Working Papers in Phonetics 15. 192 pp.

John Ohala (continued)

and Peter Ladefoged. "Further investigation of pitch regulation in speech." Working Papers in Phonetics 14:12-24. Abstracted as "Subglottal pressure variations and glottal frequency." *Journal of the Acoustical Society of America* 47:104.

Elite Olshtain (September 1963 - June 1964)

1964 "Annotated bibliography on the muscles of the tongue." Working Papers in Phonetics 1:29-39.

Rick Ratcliffe (March 1966 - January 1968)

1968 (See Shizuo Hiki, "Notes on LASS circuitry.")

D. Lloyd Rice (September 1967 - present)

1968 and Reuben Hofshi. "An interactive phonological rule tester." Working Papers in Phonetics 10:206-18.

1969 (See Victoria Fromkin, "An interactive phonological rule testing system.")

William Riley (September 1967 - June 1970)

1968 "Competence or performance? A problem for dialectology." Working Papers in Phonetics 10:200-05.

Paul Schachter

1968 and Victoria Fromkin. "A phonology of Akan: Akuapem, Asante, and Fante." Working Papers in Phonetics 9. 268 pp.

Timothy Shopen (February 1965 - June 1965)

1965 "Index of some of the UCLA linguistic phonetic research 1964-65." Working Papers in Phonetics 2:27-37.

Raymond Silverstein (June 1968 - September 1970)

1970 (See Peter Ladefoged, "Interruptibility of speech.")

Timothy Smith (September 1965 - September 1969)

1966 "Index of some of the UCLA phonetic research in 1965-66." Working Papers in Phonetics 4:54-60.

1967 (See Minoru Hirano, "Current techniques used in obtaining emg data.")

(See Minoru Hirano, "Electromyographic study of tongue function in speech: A preliminary report.")

1968 and Minoru Hirano. "Experimental investigations of the muscular control of the tongue in speech." Working Papers in Phonetics 10:145-55. Abstracted in *Journal of the Acoustical Society of America* 44:354.

1970 (See Kerstin Hadding, "Electromyographic study of lip activity in Swedish CV:C and CVC: syllables.")

R. N. Srivastava (August 1968 - September 1969)

1969 Review of *Studies in Hindi and Urdu I* by A. R. Kelkar. *Language* 45:913-27.

Marcel Tatham (January 1966 - August 1966)

1966 (See Norris McKinney, "Terminal analog speech synthesizer.")

Ralph Vanderslice (November 1964 - August 1968)

1966 (See John Ohala, "Photography of states of the glottis.")

1967 "Intonation: Accent and emphasis, cadence, and endglide." *Journal of the Acoustical Society of America* 42:1181.

"Larynx vs. lungs: Cricothyrometer data refuting some recent claims concerning intonation and 'archetypality'." Working Papers in Phonetics 7:69-79.

Ralph Vanderslice (continued)

and Peter Ladefoged. "The 'voiceprint' mystique." *Journal of the Acoustical Society of America* 42:1164. Abstract of Ladefoged and Vanderslice 1967.

1968 "The prosodic component: Lacuna in transformational theory." Rand Corporation Publication P-3874. 57 pp.

"Synthetic elocution: Considerations in automatic orthographic-to-phonetic conversion of English with special reference to prosodic features." Working Papers in Phonetics 8. 131 pp.

1970 Review of *Intonation, perception, and language* by Philip Lieberman. *Journal of Linguistics* 6:138-44.

Harry A. Whitaker (January 1966 - September 1969)

1968 "Rules versus strategies as a distinction between competence and performance." Working Papers in Phonetics 10:172-90.

"Stylistic tone-changing rules in Thai." Working Papers in Phonetics 10:191-99.

1969 "On the representation of language in the human brain." Working Papers in Phonetics 12. 169 pp.