

UCLA

Working Papers in Phonetics

Title

WPP, No. 14

Permalink

<https://escholarship.org/uc/item/4qt1q81h>

Publication Date

1970-03-01

Peter Ladefoged

working papers in phonetics
 no. 14
 university of california, los angeles

working papers	working papers
in phonetics	in phonetics
no. 14	no. 14
university of california	university of california
los angeles	los angeles

working papers in phonetics
 no. 14
 university of california, los angeles

working papers in phonetics
 no. 14
 university of california, los angeles

march 1970 march 1970

working papers in phonetics no. 14
 university of california, los angeles

March 1970

Abstracts of publications and papers presented at meetings

Shizuo Hiki and Richard Harshman	Speech synthesis by rules with physiological parameters	3
Shizuo Hiki	Control of the tongue movement for dynamic analog speech synthesis	4
George Allen	Temporal structure in speech perception	5
John Ohala and Peter Ladefoged	Subglottal pressure variations and glottal frequency	6
Peter Ladefoged	The measurement of phonetic similarity	7
Victoria Fromkin and D. Lloyd Rice	An interactive phonological rule testing system	8
R.N. Srivastava	Review of A.R. Kelkar's Studies in Hindi and Urdu	9
Peter Ladefoged and Raymond O. Silverstein	The interruptibility of speech	10
George Allen	Acoustic intensity and vocal effort as cues for the intensity of speech	11

Papers on their way to regular publication

John Ohala and Peter Ladefoged	Further investigation of pitch regulation in speech	12
Peter Ladefoged	The phonetic framework of generative phonology	25
Victoria Fromkin	A reply to the previous paper	34
Victoria Fromkin	Tips of the slung -- or -- to err is human	40
George Allen	The location of rhythmic stress beats in English: An experimental study	80
Kerstin Hadding, Minoru Hirano, and Timothy Smith	Electromyographic study of lip activity in Swedish CV:C and CVC: syllables	133

The UCLA Phonetics Laboratory Group

Staff

Peter Ladefoged, Ph.D.	Professor of Phonetics
Victoria A. Fromkin, Ph.D.	Associate Professor of Linguistics
George D. Allen, Ph.D.	Assistant Professor of Linguistics
Shizuo Hiki, Ph.D.	Assistant Research Communications Scientist [on leave from Tohoku University, Sendai, Japan]
R.N. Srivastava, Ph.D.	Post-Doctoral Fellow from Delhi University, India

Research Assistants

John Ohala	[till May 1969; currently NSF post-doctoral fellow, Research Institute of Logopedics and Phoniatrics, Tokyo: from April 1970, Assistant Professor of Linguistics, Berkeley]
William Riley	
Ray Silverstein	
Tim Smith	[till September 1969; currently Assistant Professor of Linguistics Univ. of Calif. San Diego]
Harry Whitaker	[till September 1969; currently Assistant Professor of Linguistics, Rochester]

Teaching Assistants

Mona Lindau
Dale Terbeek

Technical

Willie Martin	Senior Electronic Technician
Larry Grant	Senior Electronic Technician
Richard Harshman	Computer Programmer
D. Lloyd Rice	Computer Programmer
Julie Haaker	Secretary
Renee Wellin	Secretary
Jeanne Yamane	Secretary
Michi Yamane	Secretary

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 1426
and the UCLA Department of Linguistics

Speech Synthesis by Rules with
Physiological Parameters

Shizuo Hiki* and Richard Harshman

[Paper given at the Philadelphia meeting of the Acoustical
Society of America, April 1969]

A speech-synthesis system that consists of a dynamic analog vocal-tract circuit controlled by programs written for the LINC-8 digital computer is being developed. The computer program is designed to be able to control the synthesizer with acoustical parameters for the sound sources plus parameters specifying either vocal-tract configuration or the neuromuscular commands underlying such configuration. The time functions of these parameters or the instantaneous vocal tract shape can be drawn in with a special X-Y curve input device, and can be displayed on the CRO in the course of synthesis. The possibility of effecting simplicity and economy of input for the synthesis by specifying neuromuscular parameter (rather than vocal tract configuration) is discussed.

* Also of Research Institute of Electrical Communication, Tohoku University, Sendai, Japan.

Control Rule of the Tongue Movement
for Dynamic Analog Speech Synthesis

Shizuo Hiki*

[Paper given at the San Diego meeting of the Acoustical
Society of America, November 1969]

As part of a study of the control rules of speech synthesis with the physiological parameters, the control rule of the tongue movement is discussed. Effect of the movement of each of the several muscles involved in the tongue on the change in the three-dimensional shape of the tongue is analyzed based on the anatomical data, and the neuromuscular commands of these muscles are estimated utilizing the EMG data. The result is combined with the control rules of the jaw and lip movements to calculate area function of the vocal tract with a digital computer and to drive a dynamic analog vocal tract circuit.

* Also from Research Institute of Electrical Communication, Tohoku University, Sendai, Japan.

Temporal Structure in Speech Production

George Allen

[Paper given at the San Diego meeting of the Acoustical
Society of America, November 1969]

Temporal sequential dependencies in speech can be realized as negative correlations between the variations in duration of successive articulations; for example, if the consonant in a consonant-vowel syllable is articulated overlong, the following vowel will often be overshoot, making the syllable less variable in duration than the sum of the variabilities of its parts. Temporal dependencies in an articulatory interval may be measured by the interval's *relative variance* (RV), i.e., the variance of the interval's duration divided by its length. Under an hypothesis of no sequential dependency between articulations, RV is approximately constant for all intervals. As sequential dependencies increase, RV decreases; units of articulation then emerge as speech segments with low RV. RV also separates two kinds of dependencies. Since RV, unlike relative error, is sensitive to changes of scale, tempo variations in repetitions of an utterance will artificially inflate the RV of tempo-related intervals (e.g., rhythmic phrases), leaving unchanged the RV of segments whose durations derive more from peripheral constraints (e.g., aspiration). The RVs of some longer intervals drop dramatically when tempo effects are removed, showing clearly the presence of rhythmic constraints in speech.

Subglottal Pressure Variations and Glottal Frequency

John Ohala and Peter Ladefoged

[Paper given at the San Diego meeting of the Acoustical
Society of America, November 1969]

Most investigators agree that variations in subglottal pressure (ΔP_s) and variations in vocal-cord frequency (Δf_o) are related, such that $\Delta f_o / \Delta P_s$ is usually less than 10 Hz/cm aq. in normal speech. Theoretical work by Flanagan indicates that this ratio may vary according to vowel quality. Lieberman's theory, however, predicts a much greater range for the ratio (although his data for normal speech do not). This paper reports data from two more subjects. P_s was recorded directly via a tracheal puncture or by a small transducer inserted through the glottis into the trachea. Transglottal pressure changes were induced by pressing the subject's chest or by obstructing outgoing air at the lips. Electromyographic activity of the laryngeal muscles was monitored in one subject, ensuring the absence of muscular changes that might affect f_o . Variations in vowel quality could not be shown to affect $\Delta f_o / P_s$ as suggested by Flanagan's predictions; however, the ratio remained below 10 Hz/cm aq., refuting Lieberman's claims.

The Measurement of Phonetic Similarity

Peter Ladefoged

[Paper given at the International Conference on Computational Linguistics, Stockholm Sweden, September 1969. A slightly revised version will appear in *Statistical Methods in Linguistics* 6]

Texts in phonetic transcription, can be compared by considering each segment in terms of a set of phonetic features and counting the number of features in common between comparable segments. For the measurement of perceptual phonetic similarity it is seldom appropriate to use simply a fully specified set of binary features. In our computerised work on the topic we have built in options so that any feature may be considered to be relevant or irrelevant (which is, in part, a way of considering it to be marked or unmarked). Our options also allow us to distinguish between two kinds of non-binary features: multi-valued scalar features (in which, e.g., /e/ is regarded as between /i/ and /a/); and multi-valued independent features (in which, e.g., /p, t, c, k/ are each regarded as being an equal distance from each other). All these types of feature specification can be freely intermingled; and each feature can be given a predetermined weight in accord with the amount it is considered to contribute to the degree of perceptual phonetic similarity. Different kinds of feature specifications have been assessed by (1) using them to predict the phonetic confusion matrices that have been published by various investigators (2) using them for comparing short texts in a narrow phonetic transcription in 20 closely related Ugandan Bantu languages for which lexico-statistical comparisons and (for five of them) results of operational tests of mutual intelligibility are also available. In addition to showing that the phonetic comparisons were highly correlated with the lexico-statistical and mutual intelligibility measures, the Bantu phonetic comparisons also revealed interesting facts about which features of segments are most likely to change from language to language.

[In addition to the sources of support for the UCLA Phonetics Laboratory noted on page 2, the work was partially supported by the Survey of Language Use and Language Teaching in Eastern Africa, a Ford Foundation project]

An Interactive Phonological Rule Testing System

Victoria A. Fromkin and D. Lloyd Rice

[Paper given at the International Conference on
Computational Linguistics, September 1969]

According to the theory of generative phonology, the phonological component consists of a set of partially ordered rules which operate on the surface structures of sentences in accordance with the principle of the transformational to correct an abstract representation of a sentence into a phonetic representation. Certain pairs of these rules are disjunctively ordered. The rules are expressions containing a finite set of primitive elements. The theory provides a formalism for the representation and interpretation of the phonological rules.

A means of testing the phonological component of a grammar is required. This can be done by applying the rules to different surface structures, deriving the final phonetic output, and checking the output with what we know to be the required phonetic representation. The complexities of the set of rules needed for any language makes it extremely difficult, if not impossible, for a linguist to perform the necessary derivations without the aid of a computer.

The use of a computer for testing phonological rule systems involves the choice of an input language suitable for rule description. Easy learnability and descriptive generality are conflicting requirements in the choice of such a language. A solution is presented in the form of a rule testing system having two major characteristics. First, the system is designed around a very general internal structure consisting of a binary matrix which contains a feature representation of the form under test. Secondly, a table driven compiler operates on the rule description language and produces an internal coding of the set of rules being evaluated. The use of a table driven compiler to process the rule input language allows the language to be modified so that it pertains specifically to the selected formal rule structures. This allows the rule writer to use the system without learning a more general meta-language. The input format of the system presented is very close to the formal structure described by Chomsky and Halle in the *Sound Pattern of English* and is especially applicable to keyboard entry on a CRT graphic terminal such as the IBM 2260. The program will permit one to test any set of phonological rules. It is also planned for possible use in an on-line classroom system for teaching the properties and operation of the phonological component according to current phonological theory.

A. R. Kelkar's *Studies in Hindi and Urdu, I*

R. N. Srivastava

[Abstract of a review published in *Language*, 45.4, 913-27
(1969)]

Three major questions are discussed in this review of Kelkar's study of word phonology in Hindi and Urdu: the nature of the framework used by the author; the method of phonological analysis; and the authenticity of the phonetic description.

Kelkar does not treat Hirdu as a linguistically composite unit, but considers Hindi (H) and Urdu (U) and their intersection (HU) as distinct systems. This framework fails to reveal the intrinsic nature of the similarities and differences of the systems. The vernacular, i.e., Hirdu (Kelkar's HU) is the primary system, which fact is only revealed in a description which posits identical systematic phonemic (i.e., morphophonemic) underlying-forms. The distinction between *anumāśika* (nasalized vowels) and *anusvāra* (post-vocalic homorganic nasals) is used to justify this position. It is shown that where Old Indo Aryan (OIA) had VNC, certain dialects of New Indo Aryan (NIA) have both VNC and VC. Where both forms may be used for the same 'lexical' item, their use is strictly constrained. One may then ask: (1) What will be the systematic phonemic representation of, for example, [dənt] and [dāt]'tooth?' (2) Is *anusvāra* or *anumāśika* basic to Hirdu? It is shown that *anumāśika* is basic and both forms should be represented on the systematic phonemic level as /dəNta/ using the feature [Native] to distinguish the two phonetic outputs. A number of Morpheme Structure Conditions, and P-Rules are provided in the review to handle such questions. It is further shown, that the distinction between dialects and speech styles can be formally defined: two different dialects demand distinct vernaculars (i.e. different underlying forms and/or rules); different styles always have a common vernacular (i.e. same underlying forms).

The second section of the review reveals specific failures of the taxonomic approach. Kelkar's treatment of nasal consonants for example, forces him to distinguish /ŋ/ and /ɳ/ as separate phonemes. Such an approach obscures important facts about the neutralization of nasals and the sequential constraints which operate. It is shown that phonological minimal pairs, represent the real basic units in the grammar. To corroborate this conclusion, an experiment was conducted at UCLA in which six Hirdu speakers perceived, the final [ŋ] in English words such as *thing*, *sing* etc. as [n] in 28 out of 30 instances.

Finally, regarding the phonetic descriptions, Kelkar's descriptions of phones are obscured at times by undefined and ambiguous terms such as 'neutral quality' 'dark quality' etc. Furthermore, some questions can be raised regarding Kelkar's phonetic findings. For example he describes [h] as unvoiced in V--# position, which is not in accord with the reviewer's experimental data.

The Interruptibility of Speech

Peter Ladefoged and Raymond O. Silverstein

[Paper to be presented at the Atlantic City meeting of
the Acoustical Society of America, April 1970]

We have conducted a preliminary set of experiments on the ability of a speaker to interrupt himself when given a cue to do so at unexpected moments. Subjects were asked to repeat short utterances, each containing only one primary stress, at a fixed rhythm determined by a metronome. Contrary to our expectations, we did not find any particular point in these utterances when a speaker finds it harder to stop saying what he originally intended to say and start saying something else instead. Subjects were able to respond about 350 msec after the presentation of the cue (a 1 KHz tone), irrespective of where in the utterance the cue was presented. But when the cue was presented not within but just before the utterance, subjects took far longer (up to 750 msec) to respond. It is hypothesized that while a speaker is planning the articulations for a short utterance (in this case a single stress group), there is a period during which he cannot readily plan another utterance.

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

George Allen

[Paper to be presented at the Atlantic City meeting of
the Acoustical Society of America, April 1970]

Lane maintains that speech loudness is most directly related to the acoustic intensity (SPL) of the speech signal, according to the sone scale. Ladefoged and McKinney argue, to the contrary, that listeners rate the loudness of speech more in terms of "vocal effort," i.e., work done in producing the sound, which they showed was approximately proportional to the square of the speaker's subglottal pressure (SGP). This paper presents results showing that SPL and SGP are both about equally relevant correlates of speech loudness. The Ladefoged and McKinney experiment was replicated with eleven subjects, whose responses were analyzed for partial correlation between loudness and both SPL and SGP. The partial correlations for SPL were about the same size as those for SGP, showing that among syllables of equal SPL the ones produced with greater vocal effort were rated louder, whereas when SGP is the same the acoustically more intense syllables are louder. This shows that speech loudness is not simply a function of acoustic intensity but also depends upon cues that relate ultimately to the physiological force with which the speech was uttered.

FURTHER INVESTIGATION OF PITCH REGULATION IN SPEECH

John Ohala and Peter Ladefoged

This paper* reports two experiments on the relation between subglottal pressure and glottal frequency, and reviews some previously reported results. The experiments and the review would hardly have been necessary if it had not been for a paper by Lieberman et al. in the June 1969 issue of the *Journal of the Acoustical Society*. Their paper shows that there are still some misconceptions around.

There are a number of factors which affect the frequency of vibration of the vocal cords. The most important of them are summarized in Figure 1. The parameters and the symbols listed on the left are those suggested by Flanagan and Cherry (1969).

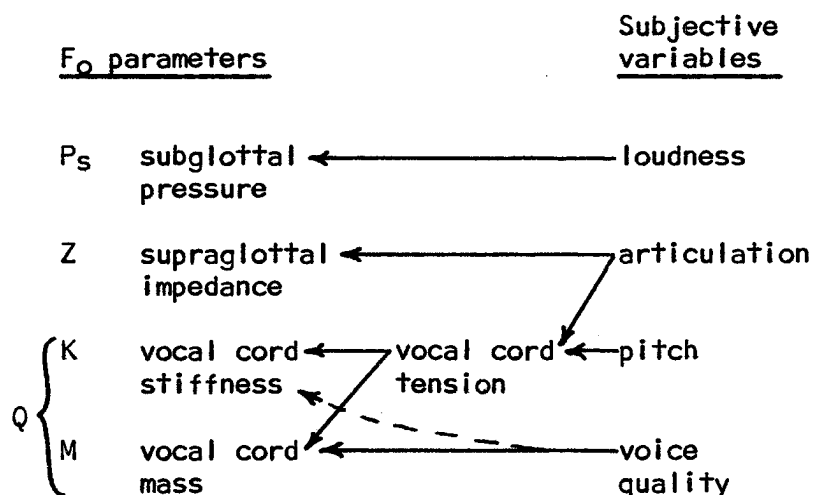


Figure 1. The parameters affecting glottal frequency, and their relation to the parameters which a subject can control.

* The text as here is a slightly revised version of a paper presented at the November 1969 meeting of the Acoustical Society of America.

The first is the subglottal pressure, with which we will be mainly concerned in this paper. Other parameters are the stiffness and mass of the vocal cords, and the supra-glottal impedance. Of course the normal speaker cannot vary each of these parameters independently of each of the others. The parameters he can vary are shown on the right of the figure. He can control the intensity of utterances. In doing this he is mainly varying subglottal pressure, but he is also varying the position of the cords to some extent. He can also aim at producing different pitches. In most circumstances he does this by varying the tension of the vocal cords. Variations in tension affect both the mass and the stiffness of the vocal cords, and, in Flanagan's model, are considered to be a single parameter which he denotes by Q . But, we must note that these parameters are not determined simply by the speaker's pitch control mechanisms. This is to some extent apparent when we consider how variations in the supraglottal impedance can occur. When a subject alters his articulation he alters not only the supra-glottal impedance, but also, in some cases, the tension of the vocal cords. In saying vowels such as /i/ and /u/, and consonants such as /k/ and /g/ the tongue is raised. This results in noticeable changes in the configuration of the laryngeal tissues which we must assume have a definite effect on the tension of the vocal cords*. Lehiste and Peterson (1961) found that, other things being equal, the vowel /i/ might be 6 Hz higher than the vowel /a/; and in work on West African Languages Ladefoged (1964) found that each of the three tones in a tone language was about 4-5 Hz higher in syllables containing /i/ and /u/ than in those containing /a/, presumably because of this mechanical linkage. Similar findings abound in the literature.

In addition a speaker may vary the mass and tension of the vocal cords when he alters his voice quality.** This is the main reason why it is better not to combine mass and tension into a

* For example, van den Berg (1955) and Shimizu (1960) report that for a given pitch the laryngeal ventricle is more enlarged for high vowels such as /i/ and /u/ than it is for low vowels such as /a/ or /ɔ/. Shimizu also noted that the ventricle is more enlarged for high pitches than for low pitches. Tomograms of the vocal cords published by Arnold (1961) also show this.

** In the paper as presented at the Acoustical Society, this section was illustrated by the photographs of the vocal cords described by Ohala and Vanderslice (1965).

single variable. Such a combination is perfectly valid as a first approximation, since we can assume that a change in pitch without a change in voice quality may consist of a simple increase in vocal cord tension. But it is quite possible to make sounds in which this relationship between mass and elasticity does not hold. In a breathy voice the vocal cords are vibrating anteriorly, but they are apart posteriorly, between the arytenoids. Presumably for this kind of voice quality the relation between mass and elasticity would be different. The same is true for a creaky voice in which the cords are vibrating anteriorly but are tightly together between the arytenoids.

Because of the complex relations between the factors which a speaker can control and the parameters which affect the frequency of vibration of the vocal cords, it is difficult to do experiments which would enable us to quantify the effects of each of the primary parameters: subglottal pressure, vocal cord mass, vocal cord stiffness and supraglottal impedance. The only one that is easiest to vary independently of the others is the subglottal pressure. One method used by a number of experimenters to study a subglottal pressure is to have a subject maintain a steady pitch and then push on his chest or abdomen at unexpected moments. Figure 2 (from Ladefoged 1963) shows the kind of record which results. The top line is the fundamental frequency of the utterance, the next the subglottal pressure, and the bottom a somewhat inaccurate record of the airflow (which we will disregard in this paper). In this utterance complex pushes on the chest caused complex pressure variations which resulted in complex variations in fundamental frequency.



Figure 2. The relation between glottal frequency and subglottal pressure which occurs when a subject is given an unexpected push on the chest (from Ladefoged 1963).

It is important to note that the frequency variations are *exactly* in synchronism with the pressure variations, since Lieberman et al. (1969) have suggested that the frequency variations may be due either to some kind of startle response or to a feedback process. But a startle response would take at least 100 msec; and the fastest feedback system that anybody has yet proposed would require a delay of 50 msec. There is no tracking delay of this kind in this record, and there can be no doubt that, at least on this occasion, the frequency variations are due directly to the subglottal pressure variations.

Several investigators have used this kind of technique to quantify the relation between pressure and frequency variations. The first quantified experiment of this kind was conducted by Isshiki (1959). He recorded the pressure via a tracheal puncture, and monitored the activity of the laryngeal muscles (cricothyroid and vocalis, Isshiki, personal communication) by means of electromyography. He did not report a figure for the $\Delta F_0/\Delta P_s$ ratio, but it is possible to calculate from his Figure 19 that it was, on that one occasion, about 3.5 Hz/cm aq. In 1963 Ladefoged reported a figure of 5 Hz/cm aq. In his experiments the pressure was recorded by means of an esophageal balloon. Öhman and Lindqvist (1966) repeated the experiment but recorded the pressure in the trachea directly through a tracheal puncture. They reported a figure of 2.5 Hz/cm aq.

Various investigators have used excised larynges. Van den Berg and Tan (1959) used an excised male larynx and came up with a figure of 5-13 Hz/cm.aq., except for falsetto for which the figure was 17-20 Hz/cm. aq. Furukawa (1967) found the ratio for an excised larynx to be 0-16 Hz/cm.aq. Anthony (1968), using a female excised larynx, reported a figure of 6-8 Hz/cm. aq. These figures should be regarded with caution. Excised larynges lack the usual muscular tone in the vocalis and other parts of the thyroarytenoid muscle. Since contraction of the vocalis can significantly affect the glottal resistance which in turn affects the rate at which a given level of P_s can affect F_0 , values of $\Delta F_0/\Delta P_s$ obtained in such studies may be too high.

Lieberman, Knudson and Mead (1969) used an esophageal balloon technique and a system for varying the transglottal pressure drop by imposing regular fluctuations in the supraglottal pressure. They published figures which show that when their subject was within his usual pitch range the ratio was 3-10.5 Hz/cm. aq, but when

he spoke at higher pitches or in falsetto the ratio was 9-18 Hz/cm. aq. There is additional data available about this speaker in Lieberman (1967). In the technique used by these authors, the tube and

Fo/Ps (in Hz/cm. aq.)		
Source	Normal Voice	High pitch & Falsetto
Isshiki (1959)	c.3.5	---
Ladefoged (1963)	5	---
Ohman and Lindqvist (1966)	2.5	---
van den Berg and Tan (1959)	5-13	17-20
Furukawa (1967)	0-16	---
Anthony (1968)	6-8	---
Lieberman et al. (1969)	3-10	9-18
Ohala and Ladefoged	2-4 2-8	7-10 ---
Lieberman (1967)	16-22	---

Figure 3. The relation between variations in glottal frequency and subglottal pressure as determined by various investigators.

box system which the subject phonated into is bound to have certain resonance and impedance characteristics of its own which will then be added to those of the vocal tract. This will increase the possibility of coupling between the vocal tract and the vocal cords and could significantly influence the value of $\Delta F_o/\Delta P_s$. It is entirely possible that their values for this expression are not at all representative of the conditions encountered in ordinary speech. These authors seem to admit this damaging point in the final

footnote to their article, in which they say: "The effect of the speaker phonating into a tube in this experiment thus, in all likelihood, was to keep the sensitivity of the larynx to air pressure variations at a high level." If there were a resonance in their system at around 150 Hz it might also account for the fact that they found higher values for $\Delta F_o / \Delta P_s$ at 150 Hz than at 110 or 220 Hz.

We have recently conducted two separate experiments of this kind. In the first experiment the activity of the laryngeal muscles (the cricothyroid and lateral cricoarytenoid) was sampled in order to make sure that their tension did not change during the pushes. Subglottal pressure was obtained from a modified spinal needle inserted into the trachea between the thyroid and cricoid cartilage. The needle was connected via a short plastic tube to a Greer photoelectric air pressure transducer. The frequency response of the system was flat within 3 dB from Dc to 25 Hz. The subject was seated comfortably in a chair with his eyes closed and attempted to maintain a given pitch while phonating the open vowel /a/. At unexpected moments he received a slight push on the chest or abdomen. Various intensities, various pitches and various modes of phonation were used.

In calculating $\Delta F_o / \Delta P_s$, the pressure before a given push was subtracted from the peak pressure to give ΔP_s and the difference between the values for F_o at these two points provided ΔF_o . The initial value of F_o was plotted against the initial value of P_s and the peak F_o against the peak P_s , the slope of the line joining these two points then equaling $\Delta F_o / \Delta P_s$. The determination of any given P_s value (from the oscillomink write-out of the pressure) is estimated to be accurate to within ± 0.5 cm. aq. and the accuracy of the F_o measurements (from narrow band spectrograms) are estimated to be accurate to within 2% of the measured value. These errors are compounded in determining the slopes of the lines and are likely to be non-negligible for small changes.

Monitoring the activity of the two laryngeal muscles was found to be important because changes in their level of activity did sometimes occur during a push, especially, when the person doing the pushing stopped being random in the interval between pushes. Apparently the subject was then able to follow the slight rhythm that was developing in the time of the pushes. When this happened there was a decrease in the activity of the cricothyroid muscle which could have had the effect of lowering the pitch at the moment of the push. Only the pressure peaks and F_o deviations that did not show any accompanying change in the activity

of these two laryngeal muscles were used in the calculations for finding the influence of subglottal pressure variations on the F_0 .

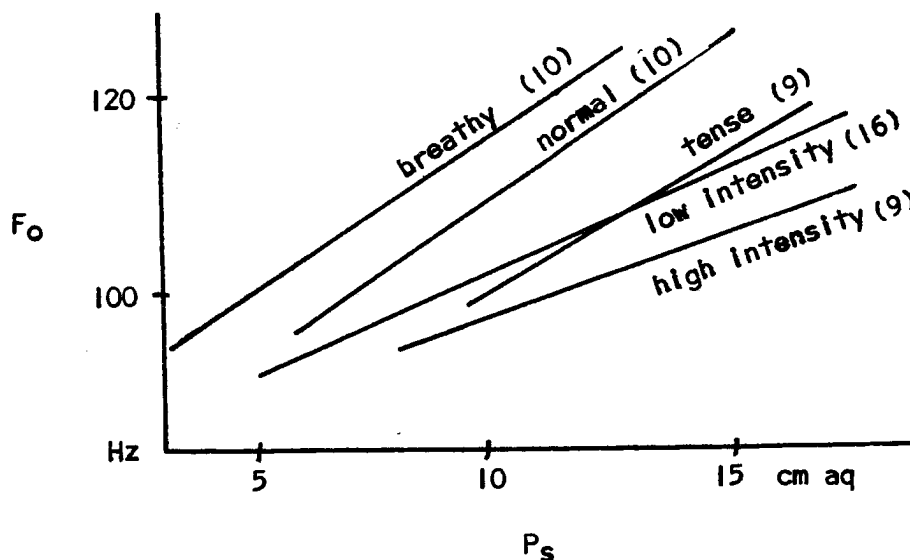


Figure 4. The relation between F_0 and P_s for various phonation types for subject J0 using a low pitched voice in his normal range.

(Estimated means of N observations, where N is shown in parenthesis after each phonation type.)

Figure 4 summarizes some of these data for a low-pitched voice within the subject's normal range. As may be seen there is very little difference in the slope of the lines for any of the different types of phonation. The lines are displaced with relation to the axes, as might be expected. Thus, for example, the line for high intensity indicates a greater mean subglottal pressure than the corresponding line for low intensity. But there are no great differences in the slopes of the lines. The range was 2-4 Hz/cm. aq. for all kinds of voice except falsetto in which it was 7-10 Hz/cm. aq.

In our second experiment, the activity of the laryngeal muscles was not monitored, and the pressure was recorded by means of a small transducer lowered into the trachea in the way described by Koike and Perkins (1968).^{*} This technique is

^{*} We are grateful to Drs. Koike, Takahashi and Ward for their assistance in the conduct of the second set of experiments.

imperfect in that the small wires between the vocal cords undoubtedly affect the voice quality and may well cause the vocal cords to vibrate in a different mode. We may note parenthetically that of the three techniques for recording subglottal pressure -- esophageal balloon, tracheal puncture, and intratracheal transducer -- we find the latter the least satisfactory for most purposes. It is undoubtedly useful when a high-frequency response is required; but in studies of the mean subglottal pressure it is not only inconvenient but also liable to cause a degree of unnaturalness in the speech.

In this experiment the subject said three different vowels, /i, a, u/. The calculations were performed in the same way as in the previous experiment, except that the fundamental frequency was calculated from the interval between successive zero crossings in a low-pass filtered wave, using a computerized technique similar to that used by Lieberman et al. (1969). The results are shown in Figure 5.

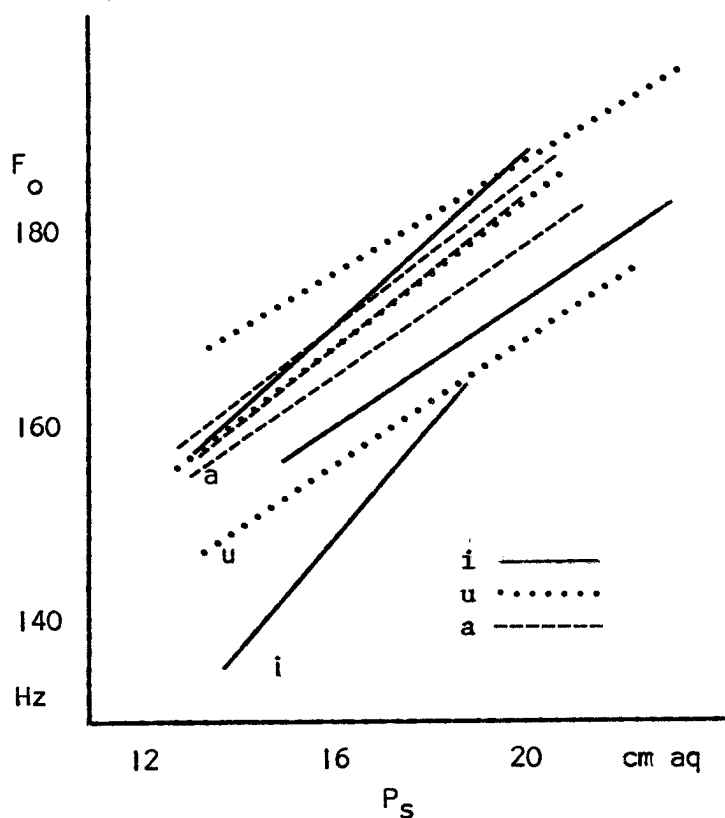


Figure 5. The relation between F_0 and P_s for three vowels as spoken by subject PL. The median and interquartile slopes are shown for each vowel.

At first glance, as noted in our abstract, it appeared that the different vowels might cause minor variations in the effect of the subglottal pressure on fundamental frequency, as predicted by Flanagan and Landgraf (1968), but it now appears that these variations are not statistically significant and we have found no support for Flanagan and Landgraf's hypothesis. Our abstract will have to be modified accordingly.* The figure that we found for the ratio was 2-8 Hz/cm. aq.

If we now return to a consideration of the values reported by different investigators (figure 3) we see that all of the reports on subjects speaking within their normal voice range show that the ratio is below 10 Hz/cm. aq. But as noted at the bottom of the figure, Lieberman (1967) claimed a value of 16-22 Hz/cm. aq. He deduced this figure from a set of data points relating the fundamental frequency and the subglottal pressure for the quasi-steady state portions of ordinary English sentences of the kind that he calls normal breath groups for three speakers of English. He "assumed that the tension of the laryngeal muscles was unchanged, so that the fundamental frequency was a function of the subglottal air pressure." (Lieberman 1967, p.95-6)** Consequently, he argued, the observed variations in frequency must have been due to the variations in subglottal pressure. In actual fact, if one makes this assumption for a number of other quasi-steady state portions in Lieberman's published data, the ratio will have to go from more than 50 Hz/cm. aq. on some occasions to minus values on other occasions. Obviously, as many investigators have pointed out, the assumption is invalid and the laryngeal muscles are very active in the pronunciation of normal English sentences with unmarked breath groups. Several years ago Ladefoged (1963) examined a sentence of this kind. He factored out those pitch variations attributable solely to changes in the pressure drop across the glottis (especially those owing to supraglottal pressure increases

* The program for the meeting said "Variations in vowel quality indeed caused small differences in $\Delta F_o / \Delta P_s$ supporting Flanagan's predictions." The version to be published in the *Journal of the Acoustical Society* should read "Variations in vowel quality could not be shown to affect $\Delta F_o / \Delta P_s$ as suggested by Flanagan's predictions."

** It is difficult to be precise about the nature of Lieberman's

during consonantal articulations) and thus derived another curve labelled "vocal cord tension," defined as "the sum of the physiological factors affecting the adjustment of the larynx." By far the greater part of the gross variations in F_0 were attributable to this parameter.

Öhman and Lindqvist, having determined the value of 2.5 Hz/cm. aq. experimentally, similarly factored out those pitch variations that could be attributed to the changes in the pressure drop across the glottis and derived curves representing the action of the laryngeal musculature. Again, most of the observed pitch variations were assigned to the action of the larynx. The pressure rises on stressed syllables were said to be capable of

assumptions, because a little later on he says:

The points in Figure 4.35 [the data in question] have a fair amount of horizontal dispersion, which indicates that the laryngeal tension *is not always constant* throughout the non-terminal portion of each breath-group. [p. 102, our italics]

But a few pages later he says:

Quantitative calculations of the relationship between fundamental frequency and the subglottal air pressure were made ... These calculations showed that the tension of the laryngeal muscles *was relatively constant* during the non-terminal portions of the breath group. [p. 107, our italics]

This kind of self-contradiction abounds in Lieberman's work; and it makes it very difficult for anybody else to contradict him; since he can always claim that he has stated the opposite to what is attributed to him. We agree with the Lieberman writing on page 102; we address our criticism in this paper to the Lieberman writing on page 107. Besides the contradictions between these quotes there is also the rather strange procedure of first assuming X (i.e. the inactivity of the laryngeal muscles) as part of the basis of Y (i.e. the calculated $\Delta F_0 / \Delta P_s$ ratio), and then later taking Y as proof of X without, as far as we can see, any other independent supporting evidence.

causing at most only about 10% of the observed pitch rise. They noted as well that

... the ΔP change [i.e., the change in the pressure drop across the glottis] which is due to stress is always much smaller than that due to stop consonants for instance, and ... the F_0 changes during the stressed syllables do not correlate well with the stress-induced ΔP changes either in phase or in amplitude.

Vanderslice (1967) made a qualitative estimate of the relative contribution of the subglottal air pressure and the extrinsic laryngeal muscles. He recorded the esophageal pressure, larynx height, and F_0 (although not simultaneously in the case of the first two parameters) and found the larynx height (a measure of the activity of extrinsic laryngeal muscles which affect pitch adjustment) to be in better correlation with the F_0 than the pressure. He concluded that "intonational pitch control is primarily vested in the larynx." Similar observations were made by Ohala and Hirano (1967) on the basis of their EMG investigations of the laryngeal muscles.

Despite all the data that has been reported Lieberman et al. (1969) say:

Recent studies of the intonation of speech [footnote reference to Ohala and Hirano (1967) and Vanderslice (1967)] have claimed that laryngeal maneuvers are solely responsible for the fundamental frequency variations that typically occur during speech. These studies claim that changes in subglottal pressure have an insignificant effect on fundamental frequency. The basis for these claims is the assumption that the ratio of change of fundamental frequency with respect to transglottal air pressure never exceeds 3 to 5 Hz/cm. H_2O . Our data indicate that this assumption is incorrect.²

There are several errors in this statement. In the first place none of the authors cited have claimed that laryngeal maneuvers are *solely* responsible for fundamental frequency variations, nor that the subglottal pressure effect is *insignificant* (it was claimed rather to be ancillary: i.e., "auxiliary" or "supplemental"). Secondly the *basis* for the claims that have been made concerning the primacy of the larynx is not only observations of the $\Delta F_0 / \Delta P_s$ ratio, but, more importantly, direct

observations of laryngeal activity. Thirdly none of the authors cited have claimed that the $\Delta F_o/\Delta P_s$ ratio *never* exceeds 3 to 5 cm. aq.; they have all read van den Berg and Tan (1959) and know that in higher pitch ranges it is greater. Finally, as we have been trying to show, all investigators including Lieberman, Knudson, and Mead themselves have shown that the ratio is well below 10 Hz/cm. aq. in the normal speech range. Their data, taken in conjunction with that of Lieberman (1967) for the same subject, indicate that a proper figure for them to claim for their subject would be 5 Hz/cm. aq. during the majority of his utterances. Consequently we continue to claim that Lieberman's theory of intonation is wrong in many ways (cf. Ohala forthcoming, Vanderslice in press, and Kim 1968). In particular, in this paper we have shown that it is wrong in so far as it relies on an incorrect value for the F_o/P_s ratio. We are still waiting for Lieberman to acknowledge the fact that *at least 90% of the linguistically significant pitch patterns in English sentences are effected by means of controlled changes in the laryngeal muscles.*

References

- Anthony, J. (1968), 'Study of the Larynx II' *Work in Progress* [University of Edinburgh] 2, 77-82.
- Arnold, G.E. (1961), 'Physiology and Pathology of the Cricothyroid Muscle' *Laryngoscope* 71, 687-753.
- Berg, J. van den (1955), 'On the role of the laryngeal ventricle in voice production' *Folia Phoniatrica* 7, 57-69.
- Berg, J. van den & Tan, T.S. (1959), 'Results of Experiments with Human Larynxes' *Practica Oto-Rhino-Laryngologica* 21, 425-50.
- Flanagan, J. & Cherry, L. (1969), 'Excitation of vocal-tract synthesizers' *J. Acoustical Soc. Amer.* 45, 764-9.
- Flanagan, J. & Landgraf, L. (1968), 'Self-oscillating source for vocal-tract synthesizers' *Trans. I.E.E.E.* Au-16, 57-64.
- Furukawa, M. (1967), 'A study of the mechanism of phonation using excised larynges' (In Japanese) *Oto-Rhino-Laryngology Clinic* [Kyoto] 60, 145-181.
- Isshiki, N. (1959), 'Regulatory mechanism of the pitch and volume of voice' *Oto-Rhino-Laryngology Clinic* [Kyoto] 52, 1065-94.

- Kim, C-W. (1968), Review of Lieberman 'Intonation, perception, and language' *Language* 44.4, 830-42.
- Koike, Y. & Perkins, W. (1968), 'Application of a miniaturized pressure transducer for experimental speech research' *Folia Phoniatrica* 20, 360-68.
- Ladefoged, P. (1963), 'Some physiological parameters in speech' *Lang. and Speech* 6, 109-19.
- Ladefoged, P. (1964), *A phonetic study of West African languages* Cambridge, Cambridge University Press.
- Lehiste, I. & Peterson, G.E. (1961), 'Some basic considerations on the analysis of intonation' *Journ. of the Acoustical Soc. of Amer.* 33, 419-25.
- Lieberman, P. (1967), *Intonation, Perception and Language*, Cambridge, MIT Press.
- Lieberman, P., Knudson, R. & Mead, J. (1969), 'Determination of rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation' *Journ. Acoust. Soc. Amer.* 45.6, 1537-43.
- Ohala, J. & Vanderslice, R. (1965), 'Photography of states of the glottis' *Working Papers in Phonetics* 2, [UCLA] 58-59.
- Ohala, J. & Hirano, M. (1967), 'Studies of pitch change in speech' *Working Papers in Phonetics* 7, [UCLA] 80-85.
- Ohala, J. (1970), 'Aspects of the control and production of speech' *Working Papers in Phonetics* 14, [UCLA].
- Ohman, S., & Lindqvist, J. (1966) 'Analysis-by-synthesis of prosodic pitch contours' *Quarterly Prog. & Status Report* [Royal Inst. of Tech, Stockholm] 4, 1965, 1-6.
- Shimizu, (1960), 'On the motions of the vocal cords in phonation studied by means of the high voltage radiograph movies' *The Oto-Rhino- and Laryngological Clinic (Zibi-Inko-ka Rinsyo)* 53, 446-461. (in Japanese)
- Vanderslice, R. (1967), 'Larynx versus Lungs: Cricothyrometer data refuting some recent claims concerning intonation and archetypality' *Working Papers in Phonetics* 7, [UCLA] 69-79.
- Vanderslice, R. (in press) 'Intonation, Scientism and Archetypality' *Journ. of Linguistics*, 6.1.

The Phonetic Framework of Generative Phonology

Peter Ladefoged

The *Sound Pattern of English* (Chomsky and Halle 1968; henceforth SPE) is a luv. It is full of insights and explanations for facts that I had known and worried about for years without being able to find the correct underlying generalizations. Nevertheless I disagree not only with a number of comparatively minor points of phonetic fact, but also with some basic principles in the theory.

The first disagreement is over what a phonological description is supposed to be doing. Chomsky and Halle would say that they are trying to account for the competence of a native speaker of English. But I do not think this is what they are in fact doing; nor do I think it is what they should be doing at this moment in the history of linguistics. In practice they are exposing (not really explaining) the sound patterns that occur in a language; and this is all we can expect of a phonology at the moment. These sound patterns are the result of numerous interacting historical forces. They exist in the language (or, to be more precise, in the spoken medium of the language); and only some of them exist in the minds of speakers of the language. There is no doubt that we all have the competence to form plurals of new words. These kinds of phonological rules can definitely be said to exist in our minds. But I have serious doubts whether the vowel shift rules have a similar existence. If they did, it would mean that *proper* and *copper* are stored in our lexicons with different vowels, *proper* having a tense vowel (because we "know" that there are related forms such as *appropriate*) but *copper* having a lax vowel (because there are no related forms with tense vowels, and therefore no need to store it in a more complicated way and make it go through the vowel shift rule). Nor do I think it likely that *knee* has the underlying form *ne* but *know* has the form *kno* (because of *acknowledge*). Nor that there is any "psychological reality" to the relations between *amnesia* and *mnemonic*; *debt* and *debit*; *lepidoptera*, *helicopter* and *pterodactyl*. But these are all part of the sound pattern of the language which should be exposed by an adequate phonology.

It is difficult to give a principled reason for what should and what should not be included in the phonological description of a language if we maintain that we are trying to describe what a speaker knows. This is because a speaker's knowledge may be due to different sources. He may know something because it is genuinely part of the sound patterns which are within his competence; this is probably the case in plural formations. But at another level a speaker may know a sound pattern only in the sense that he knows how sets of words are spelled. I would

guess that only in this sense does he know the rules involved in alternations such as *resign*, *resignation*. My guess is slightly substantiated by the fact that my illiterate children (hopefully pre-literate, but I'm not sure) can generate the correct plural forms of new words, and even produce the stress changes predicted by Chomsky-Halle for coined words such as *sublane* and *sublanation* (but usually without reducing the second vowel). But they never convert an aurally presented form such as [səb'sain] to [səbsig'neiŋ]. This sort of relation can probably be made only by those who can assign a spelling such as *subsign*. Then there is also a third stage in the knowledge of the sound patterns of a language. This is that of the grammarian who knows that there are regular processes underlying the formation of *nasal* from *nose*; no doubt this kind of speaker can generate not only [səbsig'neiŋ] from [səb'sain] but perhaps also *snilth* from *snoul*.

Given these three stages of knowledge (pre-literate, spelling influenced, grammatically sophisticated), I cannot see how to motivate a phonological description in terms of a speaker's knowledge or competence. Which speaker? Competence in which sense? The usual answer is that a grammar should describe an idealized speaker, or, in Chafe's (1968) phrase "a maximally efficient rule user." But I agree entirely with Chafe when he says: "I know of no way to determine precisely how many rules of a synchronic set represent a speaker's tacit knowledge -- how deeply he has internally reconstructed in forming subconscious generalizations about his language." We have neither the theory nor the experimental technique to use an *idealized* speaker's knowledge to motivate decisions concerning what should and what should not be in a phonology.

It is true that if we consider phonology to be, for the moment, simply the description of the sound patterns within a language, it makes the subject less interesting; we are merely cataloging the results of interacting historical processes. (Of course, if we can go on and explain these processes in terms of extra-linguistic processes, this is interesting in another way.) But until we have better theory and more appropriate experimental techniques there is little else that we can do.

A weakness of the view that is being expressed here is that it commits one to putting every rule describing a sound pattern (a regular phonological event) into a grammar. It is impossible to motivate leaving anything out; and on the Chomsky-Halle view, if there are observable regularities it is difficult to see why a speaker does not "know" them. The only possible case I can think of might be the regularities in the combinations of sounds such as the fronting of the [k] in *key* as opposed to *car*. I have elsewhere called these different [k] sounds intrinsic allophones (Ladefoged 1966); but I now prefer the term coarticulated allophones. If these combinatory rules are language specific then I think it necessary for them to be included in the phonology of a language; within the Chomsky-Halle framework one would argue that a speaker "knows" that they are the kind of details in the sound pattern of a language which are necessary for the correct production of a native pronunciation. But in so far as coarticulations are universal and not language specific, then they are necessary only for speech to sound human. They are depen-

dent on constraints imposed by the physiology of the vocal organs, and as such have to be stated as part of the meta-theory of linguistic description rather than part of the phonology of any particular language.

At this stage there are two possibilities open to us: we can follow Chomsky-Halle's theory; or we can follow their practice. If we follow their theory we will be able to say very little about the sound pattern of English. We have hardly any data resulting from convincing experiments on the identity of sound patterns. Of course we can guess, just as Chomsky and Halle guess, that the /g/ in *resign* has some form reality. But to go on making unverified guesses hardly advances the science very much. Nothing is proved about present day psychological reality by the fact that these unverified guesses all knit together, supporting each other by forming an intricate system of rules. The system as revealed by rules such as those in SPE is simply the sum of a set of historical events, and there is no reason to assume that all these past events are part of a speaker's present knowledge.

Alternatively we can follow the Chomsky-Halle practice, and not worry about data concerning psychological reality. This still leaves us in a number of difficulties. It might be objected, for instance, that this view of phonology cannot be sufficiently constrained, particularly in its choice of primitives, the feature set which it uses. This is a difficult challenge to meet; there is a great deal of appeal to the alternative view (a kind of *deus ex machina*) that the feature set exists in a speaker's mind as an innate capability. And there is, of course, no doubt that the features which are needed for explaining why certain sound changes occur and not others, how a child learns the rules defining the possible words of a language, and what is considered to be a rhyme or a rhythm, the features that are needed for all these and many other purposes must be represented in the brain. But note that these features cannot be thought of as simply a set of independent neural commands to the speech organs, or, alternatively, of neural patterns corresponding to auditory events. The patterns that arise in the sounds of a language are due to intersecting causes. At least two quite different kinds of features are needed to explain them. Some patterns can be explained in terms of acoustic events, others in terms of articulatory events. Thus on the one hand, there is no doubt that /p/ and /k/ go together in the formation of patterns in some languages; this is because of their acoustic similarity (which used to be called gravity), and no amount of guesswork is likely to lead to establishing anything in common in the neural commands to the speech organs which make them. But, on the other hand, patterns of the kind exhibited in the formation of compounds such as /mp, nt, ŋk/ are obviously due to articulatory constraints; and it is difficult to state rules concerning them in terms of meaningful acoustic features.

Putting this another way, the kind of statement I am objecting to is:

The total set of features is identical with the set of phonetic properties that can in principle be controlled in speech: they represent the phonetic capa-

bilities of man... (SPE 294-5)

Of course Chomsky and Halle are using the phrase 'the phonetic capabilities of man' in a new and rather special way, so it is no use arguing against them by saying that many years ago authorities such as Pike (1943) showed that there is a vastly greater range to the phonetic capabilities of man (using this phrase in its more usual sense) than is specified by the Chomsky-Halle framework. Any child can make a vast variety of groaning, sucking, spitting, whistling and sniffing noises, at least some of which cannot be described in terms of their features. But all this is irrelevant if we use the word 'phonetic' in the way which Chomsky and Halle imply, and restrict ourselves to those aspects of sounds which are of linguistic interest.

What is more to the point is that it seems unwise to try to claim for any current feature system any more than that it is a summary of the data we now have available; and there are severe limitations on our current data. We know a considerable amount about the systematic phonetic contrasts in the surface forms of different languages. But for the vast majority of the languages of the world we know almost nothing about the underlying sound patterns or the deeper phonological relationships that are so fully demonstrated for the first time in SPE. In the present state of our ignorance it is presumptuous to make claims about the best way of grouping sounds together so as to form natural classes which are appropriate in the description of all languages.

Furthermore, it is by no means obvious that it is even possible to develop a feature system at the present time. As Chomsky and Halle themselves say:

Certain apparent linguistic universals may be the result merely of historical accident. For example, if only inhabitants of Tasmania survived a future war, it might then be a property of all existing languages that pitch is not used to differentiate linguistic items. Accidental universals of this sort are of no importance for general linguistics, which attempts rather to characterize the range of possible human languages. The significant universals are those which must be assumed to be available to the child learning a language as an a priori, innate endowment. (SPE 4)

Nevertheless nearly all their own arguments for or against a particular system of universal features are based on observation of what goes on in particular languages. Although they claim they would like to use extralinguistic data, in fact nothing crucial is based on evidence concerning a child's language learning ability, nor on our current neurophysiological knowledge concerning his innate endowment. Nor do Chomsky and Halle consider any other kinds of data, such as the various ranges of phonetic events which are potentially linguistically significant, but which (so far) have not been observed. Consequently although their feature system may be observationally adequate as far as it goes, it is unlikely that it will lead to a fully explanatory theory concerning the phonetic capabilities of man.

It also seems to me to be very dangerous to set up a feature system as Chomsky and Halle do, because it may lead to phonological descriptions not showing the sound patterns in the best possible way because they are being constrained by what turns out to be an inappropriate set of features. I must admit that I am somewhat alarmed by the possibility that the premature Jakobsonian system, which was regarded as an orthodoxy for so long, might be blindly replaced by the new feature set proposed by Chomsky and Halle, for many aspects of which the evidence seems rather tentative. I am sure that they agree that their new feature system should be regarded as a suggested framework which might not always lead to the formation of the correct natural classes. But it seems likely that one result of their presentation will be that linguists will be less willing to experiment and develop their own feature systems, which might reveal more about the languages they are describing. Admittedly, when one describes a particular language one has to behave as if the feature set one is using is a sub-set of the universal feature set; and consequently the disagreement between Chomsky and Halle and myself is not very important. It is just that in thinking how best to advance the science of linguistics, it seems tactically foolish, and completely unnecessary in a description of the phonology of English to consider whether any feature set summarizes the phonetic capabilities of man. We are still at the primitive stage of phonological theory, where there should be a continuous process of enrichment of the theory as more data becomes available.

The other basic disagreement I have is over a principle which Chomsky and Halle retain from much earlier work (Jakobson, Fant and Halle, 1951), and which I had also taken over (Ladefoged 1966 and 1967) but which I now think is invalid. The present Chomsky-Halle formulation is:

Each feature is a physical scale defined by two points. (SPE 299)

This is rather like my own principle, if the word feature is substituted for category:

Two different phenomena should be described in terms of the same category if and only if the category has measureable properties and the phenomena can be distinguished simply in terms of numbers specifying the degrees in which they possess these properties.
(*Linguistic Phonetics*, 65)

But it now appears to me that any adequate phonology must contain some features not specifiable in this way. In general these features, which we may call the non-scalar features, will be those required for defining the major natural classes such as vowels, consonants, sonorants, liquids, and glides. These major classes have defied individual phonetic specification for generations. Perhaps from a feature point of view they should be considered as higher relations between those other features which can be defined in terms of scalar quantities. This possibility raises interesting difficulties within the marking conventions and the evaluation procedure; but they should not be impossible to overcome.

In some ways the present Chomsky-Halle treatment of the major class features is very reminiscent of Pike's (1943) definition of two oppositions vocoid and non-vocoid, and syllabic non-syllabic, which enabled him to define vowels as syllabic vocoids, true consonants as non-syllabic non-vocoids, etc. But it certainly seems hard to claim that any of the Chomsky-Halle oppositions concerned with these natural classes are definable in terms of physical scales. For instance, their definition of vocalic is:

Vocalic sounds are produced with an *oral* cavity in which the most radical constriction does not exceed that found in the high vowels [i] and [u] and with vocal cords that are positioned so as to allow spontaneous voicing; in producing nonvocalic sounds *one or both* of these conditions are not satisfied. (SPE 302, my italics)

This is hardly a scale defined in terms of two points; I find it difficult to imagine a metric which could be applied to two sounds in order to determine which of the two was more vocalic; and without such a metric there cannot be a scale. In a footnote they say 'Recent work indicates that in place of "vocalicness" the phonetic framework should contain a feature of "syllabicity." I wish them luck in looking for a physical scale of syllabicity. It is probably even harder.

But this does not mean that there is no such thing as syllabicity. As Halle (1964) has pointed out (when talking about the existence of phonemes) there is no need for us to have operational definitions of all our primitive terms. We need not worry if we cannot specify a physical scale defined by two points for syllabicity. (It is probably not a scalar quantity, but an organizing principle concerned with the coordination and timing of speech movements.) But irrespective of whether syllabicity and other major class features are actually or even potentially definable in physical terms of any known kind, it is quite clear that they are constructs needed within phonological theory. In a way it is odd that Chomsky and Halle, of all people, should be so hung up on finding measureable physical scales for all their features.

At this stage we should perhaps note that we do not gain anything by attempting to define features in perceptual terms. It is impossible to speak of the perception of phonetic events by an unbiased observer, unless we are considering a new born infant. Anybody who speaks a language will perceive sounds in terms of his own particular experience: and an infant's innate endowment for perceiving all possible speech sounds cannot be tested. So attempting perceptual or mentalistic definitions of features gains us nothing except confirmation of our cynicism concerning operational definitions.

A much better approach, once we have abandoned the notion of a simple physical or perceptual scale corresponding to each feature, is to consider the possibility of a hierarchy in which some features are considered to be cover terms denoting groups of other features. In the Chomsky-Halle framework the major class features (which are obviously useful) are most easily described in this way. We have already noted

difficulties in defining features such as vocalic and consonantal. Chomsky and Halle implicitly recognize the difficulties when they find it necessary to present these features in a table listing the complete set of classes of sounds of which they are composed. It would seem likely that the phonological description of almost any language would require the possibility of a hierarchical relation among features, such that some features were available simply as cover terms denoting groups of other features.

It may well be that with the exception of those features which are simply cover terms, we may eventually be able to give physical definitions of all features. We certainly cannot at the moment define the phonetic capabilities of man; and we must be satisfied with regarding some features simply as constructs necessary in phonological descriptions. But we must, of course, go on trying as hard as we can to find physical correlates of all features, so that our phonological statements will have more explanatory power.

It might be useful to compare the position suggested here with those discussed by Kiparsky (1968). I am not advocating the position he attributes to Lamb (1966) and Fudge (1967), in which morphophonemic representation are fully abstract classificatory devices. It is obvious that there must be some constraints of the kind referred to as naturalness by Postal (1968). But this does not mean that the constraints have to be worded in terms of a speaker's competence. Naturalness constraints should be stated in terms of acoustics and physiology, if they are to have any real explanatory power.

The main suggestions being made here are that we stop regarding the sum of past phonological events as being necessarily the same as a speaker's present knowledge; and we stop pretending that we can, at this time, define a phonology as a set of rules in a speaker's head. The rules given in SPE are no doubt partially represented in the brain; and the part that are represented are probably redundantly represented. Perhaps one day we will know how to restrict ourselves to describing just this complex and no doubt inelegant representation. But meanwhile we have neither the experimental techniques nor a sufficiently well defined theory to allow us to do this. It seems preferable at the moment to view phonology as the explanation of the pattern of sounds which occurs in the medium of spoken language.

References

- Chafe, W. (1968), The ordering of phonological rules, *Internat. J. of Amer. Linguistics* 34, 115-36.
- Chomsky, N. and Halle, M. (1968), *The Sound Pattern of English*, New York: Harper and Row.
- Fudge, E. (1967), The nature of phonological primes, *Journal of Linguistics* 3, 1-36.
- Halle, M. (1964), On the bases of phonology, *The Structure of Language* (ed. J.A. Fodor and J.A. Katz), Englewood Cliffs: Prentice Hall.
- Jakobson, R., Fant, G., and Halle, M. (1951), *Preliminaries to Speech Analysis*, Cambridge: M.I.T. Press.
- Kiparsky, P. (1968), *How abstract is phonology?* (unpublished paper).
- Ladefoged, P. (1966), The nature of general phonetic theories, *Georgetown Monograph Series on Languages and Linguistics* 18 (ed. C.W. Kreidler).
- Ladefoged, P. (1967), *Linguistic Phonetics*, UCLA, *Working Papers in Phonetics* 6.
- Lamb, S. (1966), Prolegomena to a theory of phonology, *Language* 42, 536-73.
- Pike, K. (1943), *Phonetics*, Ann Arbor: Univ. of Michigan Press.
- Postal, P. (1968), *Aspects of Phonological Theory*, New York: Harper and Row.

PAGE 33 IS BLANK

A Reply to P. Ladefoged's

The Phonetic Framework of Generative Phonology

Victoria Fromkin

Comments by Peter Ladefoged have been added in the margin

Ladefoged believes that the time is not ripe for any attempt to 'account for the competence of a speaker' in constructing a phonology. He suggests an alternative goal for the present -- 'exposing the sound patterns that occur in a language' which patterns 'exist in the spoken medium of the language.' His position is based on the following arguments:

(1) 'Some phonological rules ... definitely exist in our minds' (e.g. plural formation); others may not (e.g. vowel shift rule). The patterns which do not result from psychologically real rules should still be accounted for (e.g. the relations between *know* and *knowledge*, *amnesia* and *mnemonic* etc.).

(2) Since a speaker's knowledge of the sound patterns may result from actually existing rules, spelling influence, or linguistic (scientific) knowledge of the language there is no way to 'motivate a phonological description in terms of a speaker's ... competence.'

(3) The necessary experimental techniques are lacking whereby we can genuinely test our hypotheses, particularly if we keep talking of an 'idealized' speaker's 'competence.'

The problem Ladefoged poses has no simple answer. On the one hand, there is the difficulty of determining what is or is not psychologically real. On the other hand, without such a criterion we have no constraints on what 'patterns' we include in our grammar, i.e. 'it is impossible to motivate leaving anything out.' If we leave out the 'observable regularities' which a speaker cannot 'know,' we have a motivation but we're back to aiming at a description of competence. One need only look at phonological descriptions which do not aim at accounting for competence to see where the lack of such constraints lead. Such descriptions are observationally adequate in that they account for the data. But 'exposing patterns in the spoken medium' is not the same as exposing linguistically significant patterns.

In Trnka's *A Phonological Analysis of Present-Day Standard English* (1968, University of Alabama Press) Chapter IX is entitled 'Combination Rules of English Phonemes'. (my emphasis). Trnka reveals additional 'patterns' such as the fact that monosyllables which begin with vowels constitute 3 percent of all monosyllables. Ladefoged has no way of excluding such statements from his kind of description of English; he cannot say why they are not rules except by defining the notion rule by reference to something other than the 'spoken medium of language', such as the notion of a speaker's competence. Ladefoged and I certainly agree that a rule of plural formation is a generalization to be included in a grammar of English, and that Trnka's rules should not.

Phonological theory dictates what rules shall be included in a grammar in that it aims at constraining phonological descriptions to just those rules or patterns which any speaker can know, i.e. have in his competence, or can acquire as a child given certain input data. To include in a grammar of English a rule which deletes a pre-final nasal *g* (i.e. $g \rightarrow \emptyset / \text{-- } N\#$) thereby accounting for the relationship between *resign* and *resignation* as well as many other such alternations (*phlegm/phlegmatic*; *malign/malignant*; *paradigm/paradigmatic*) does not imply that the actual stored item (in the performance sense) is so represented. Rather it is a means of showing that these alternations are not idiosyncratic i.e. not particular properties of these lexical items, but are predictable by general rule. In fact when we examine other phenomena we find that

1. $g \rightarrow \emptyset / \text{-- } N\#$ (as in *sign*)
2. $g \rightarrow \emptyset / N \text{-- } \#$ (as in *sing*)
3. $g \rightarrow \emptyset / \# \text{-- } N$ (as in *gnosis*)

Using Bach's and Langacker's 'neighborhood' convention, one rule can capture this generalization:

4. $g \rightarrow \emptyset / N / \#$

(The expansion $[g \rightarrow \emptyset / \#N \text{--}]$ applies vacuously.)

Furthermore the above rule, if generalized to include *k*, would also account for *know* and *acknowledge*. However, if we do account for this alternation (and *gnosis/agnostic* etc.) we have to include in the lexicon formatives beginning with *gn* and *kn*. But this

Alternatively, I could define the notion rule as a formalized invariable statement about the medium of spoken language, and omit Trnka's 'rules' because they are only probability statements.

This is fine as long as it implies nothing more than the fact that the patterns occur.

violates a more general constraint on the ways in which words can begin in English.

5. $[+consonantal] \rightarrow s / + \text{ -- } \begin{bmatrix} +consonantal \\ -vocalic \end{bmatrix}$ (SPE 171)

Why maintain (guess) such an unlikely possibility? (My guess is that they do not 'know' even that these pairs form an alternation.)

In fact, speakers of English (I maintain) do 'know' that there is something 'special' or exceptional about the *know/acknowledge* kinds of alternations, and we therefore want to include Rule 5 in a grammar of English, and still show the relationships between those 'forbidden' items. There are a number of alternatives we can pursue. As exceptional formatives they can be marked [-MS 5]. Every segment of the above formatives would have to be so marked, thereby complicating the grammar which is just what we would want, since this extra 'cost' reflects part of the 'idealized speaker's competence'.

You do not account for *why* they are patterns. I would give an answer in historical terms.

I would imagine that Ladefoged would accept the inclusion of such rules but would counter -- 'Fine, but no speaker 'knows' them. They only exist as a pattern in the spoken medium.' But this does not account for why they are patterns. Furthermore Ladefoged's suggestion that 'these are all part of the sound pattern of the language which should be exposed in an adequate phonology' would not motivate differentiating between certain 'patterns' and others. He might counter by saying that since the 'exceptional' items constitute a smaller number of patterns, this less frequent patterning can be revealed in the manner suggested above.

If you can invoke simplicity (of a grammar) so can I (of description of the patterns),

But a resort to frequency will complicate rather than clarify the problem. A quick glance at Webster's 3rd will reveal that there are fewer words beginning with [bɫ] than with [brɫ]. The current theory of phonology 'forces' us to make no distinctions between these initial 'possible' segmental sequences, since any distinction will complicate the grammar, whereas the distinction between /#kn/ and /#sn/ must be made since it will simplify the grammar. The constraints of the theory are supported by the fact that speakers of English will readily accept new 'words' beginning with [bɫ] but will not accept any beginning with [kn]. Thus the theory itself establishes the 'psychological reality' of the rules of the grammar.

My argument is based on acceptance of the goal stated by Chomsky and Halle but with a different view of 'psychological reality' than that which seems to be

emerging in some linguistic circles. It is only if a grammar of competence is an abstract idealization that it can contain all the generalizations which account for the 'patterns' derivable by *grammatical* rule. To ask 'which speaker' is like asking whether a grammar of English should include in its lexicon the word *phonetic* if some speaker never heard the word (and alas there are such speakers). By including such a form, the grammar predicts that some speakers have such a form in their lexicons, and furthermore that any speaker could add it with all its idiosyncratic phonological, syntactic and semantic features specified and, that given his knowledge of English phonology he could also derive *phonetician* pronounced [fəʊnə'tɪʃn] and not *[fəʊnə'tɪkɪjən] or *[fəʊnə'tɛkɪjən].

But such a grammar must include only forms and rules which are 'psychologically real' or it will have to include some of Trnka's rules, or many others equally uninteresting linguistically. Given Ladefoged's short term goal we have no way of deciding between them. A rule is 'psychologically real' if its predictions can be verified, and if it is permitted by our theory of phonology which places constraints on the kinds of rules, and the form of rules which can occur in any language. An abstract representation is 'psychologically real' if a correct phonetic representation can be derived from it. The theory must specify the kind of generalizations which do represent a speaker's linguistic knowledge. Once you eliminate this constraint beautiful elegant patterns can be woven which may reveal nothing about human language, although they may reveal a great deal about the creative ability of the human linguist (and man in general).

It should be clear that I am not arguing that the particular rules in SPE or in any other phonology do represent the phonological competence of the idealized speaker, only that they should. It is perfectly clear that further constraints must be built into the theory, which will provide Ladefoged and all of us with a way to motivate a phonological description in terms of a speaker's competence.

If competence is defined as above, the fact that a speaker's knowledge may be due to different sources should present no problem. Certainly spelling may influence one's knowledge, and literacy, and meeting a stranger in the night. If such knowledge can be revealed by grammatical rule, which rules are permissible by our theory of phonology, fine.

I agree he *might* if he could see (or imagine) the words in written form. But I am sure you agree that we should really do the experiment, and not be like Chomsky/Halle who just make claims.

This can't be done.

Chomsky/Halle have no way of constraining a theory of their kind to permit the addition to the grammar of my English

the fact that I now know (through reading SPE) that the first vowel in the underlying form of *violent* is lax.

There is no question but that we need to develop new experimental techniques to test our hypotheses which constitute the theory of universal grammar. No science, however, can declare a moratorium on theory development until such techniques are developed. Furthermore, theoretical arguments (as well as empirical data) can often decide between opposing hypotheses. Thus, for example, Kiparsky put forth a number of arguments calling for a constraint against 'absolute neutralization'. These arguments have been countered by Hyman, Kisseberth and others both from a theoretical point of view and from data attempting to justify the position that under certain circumstances rules involving absolute neutralization are indeed 'psychologically real.'

Whether the vowel shift rules are 'real' rules, then, will be determined by the constraints on such rules provided by the theory. In other words, at present, there are no agreed on constraints which would disallow the setting up of an underlying /æ/. There are many linguists who believe that this part of the vowel shift rule is wrong, and arguments must be (and are being) put forth to disallow this kind of abstraction. In other words it is argued that the ability to derive the correct phonetic forms from abstract underlying forms does not constrain the grammar sufficiently.

But there is nothing to gain by guessing wildly beyond our present data; (& we have our scientific integrity to lose).

Let me conclude briefly with the second main point in Ladefoged's paper on the proposed feature system. He is concerned because the features are merely 'a summary of the data we now have available; and there are severe limitations on our current data.' I also have serious questions about the new set of features but not because we do not have enough knowledge to form a theory. How much data is enough? The fact is that the hypotheses are falsifiable, and I think some are falsified by the data already available. If we find that in the languages of the world there are natural classes of sounds needed for phonological rules which can not be provided by the current set of features, something is wrong. For example, Lyle Campbell has criticized the current feature set because it fails to show the relationship between rounded vowels and labial consonants, which class is needed in many languages. In addition, one may argue that the current feature set permits specious and false generalizations (for example, the class of [+distributed] sounds).

It is by applying such a feature system to languages that we are able to revise it and develop a more viable theory. Of course the proposed set should not be blindly accepted as a new orthodoxy. No scientific hypothesis should be, and if non-scientific linguists distort the meaning of science that is not a problem for the theory but for the linguists. If linguists are less willing to experiment (should they find the proposed features inadequate) that is also a reflection on them. My suggestion to them would be to use Ladefoged as a model. His concern for theoretical validity, and empirical verification, coupled with his far-flung and meticulous research forced many of the changes which have already been made in the theory of universal phonetics and will undoubtedly create many more.

[There is no doubt that this argument will continue. Editor]

TIPS OF THE SLUNG -- OR -- TO ERR IS HUMAN

Victoria A. Fromkin

In the current linguistic and psychological literature a sizeable number of articles have appeared dealing with 'slips of the tongue' and errors in speech. (See bibliography) The interest in speech errors is not, however, of recent origin. Historically, they have been the source of humour as well as of serious study. Rabelais in the sixteenth Century utilized such errors to display his pungent wit and in the Compleat Gentleman (1622) Henry Peacham refers to a 'melancholy Gentleman' who says 'Sir, I must goe dye a beggar' instead of the intended 'I must goe buy a dagger.'¹ 'Spoonerisms' were uttered before and after the long happy life of The Reverend William A. Spooner, who is credited as the originator of a particular kind of 'lapse'. In fact, if one assumes that the origin of man and the origin of language and speech arose simultaneously, then a further assumption follows -- that the origin of 'spoonerisms' arose with Adam.

Speech error data have been studied as a source of historical linguistic change (Sturtevant 1917, 1947, Jespersen 1922, MacKay forthcoming) as a means for understanding the actual mechanisms in the speech production process (Lashley 1951, Boomer and Laver 1968, MacKay 1969, 1970, Hockett 1967, Fromkin 1968, Nooteboom 1968) and as a means of shedding insight into psychological repressions. (Freud 1901) Speech errors have also been investigated in attempts to show the 'reality' of phonological units and rules, and the relationship between linguistic 'competence' and 'performance.' (Fromkin 1968, Green 1969) Freud, in his 'Psychopathology of Everyday Life' which first appeared in 1901, questioned 'whether the mechanisms of this (speech) disturbance cannot also suggest the probable laws of the formation of speech.' (p. 71) It is to this general question that this paper is directed.

2. Data

Every book and article which refers to speech errors is replete with examples. The most extensive collection, an estimated 8,800 errors, appears in Meringer and Mayer (1895) and Meringer (1908). A rigorous statistical analysis of these errors is contained in a number of articles by MacKay. (1969, 1970) This corpus of German errors, is augmented by errors in spoken Dutch noted by Cohen (1966), by over one hundred tape recorded errors in English by Boomer and Laver (1968), and by the other errors cited in various articles

listed in the bibliography.

The data used in this paper, while considering the extensive published data will primarily be a collection of speech errors collected by myself over the past three years. Over six hundred errors were collected by myself personally or by colleagues and friends who reported in detail errors which they either made or which they heard others produce.²

For each error which I myself noted, the name of the speaker was recorded, the date, and where possible, (particularly in the case of blends) the speaker was questioned as to what he had been thinking of saying. This is scanty information indeed when compared to the data recorded by Meringer for each error in speech which he heard. In true Germanic style, he also included the birthdate of the speaker, the educational background, the time of day, the state of health and tiredness, the rate of speech etc. Sturtevant reports that Meringer became the most unpopular man at the University of Vienna, and since 'no correlations between any of the above factors and the nature of the error were found.' (MacKay, 1970) my own data collecting omitted such information to protect my personal reputation while maintaining the scientific accuracy of the data. It is important to note however that this (my) method of data collecting has a built-in bias, since many errors occur when it is just not feasible to note them, and unquestionably many errors made are not 'heard' at all. The data collection method used by Boomer and Laver (1968) in which they analyzed tapes of conference discussions, psychiatric interviews etc. for the errors which they contained is not similarly biased. Fortunately, there were not sharp discrepancies between the kinds of errors recorded by them and by myself. There are certain kinds of errors included in my corpus which did not seem to occur among the hundred or so errors recorded by them. I only included such errors when other listeners heard and attested to the error, or when the speaker himself caught the error and then corrected it. I felt this precaution necessary to mitigate my own 'desire' to hear certain kinds of errors.

The aim of this paper is therefore not to treat the errors in the corpus as a random sample of all errors made, but to attempt an explanation for the attested errors which were recorded.

3. Discreteness of Performance Units

Sturtevant (1947) defines a 'lapse' or a 'speech error' as 'an unintentional linguistic innovation.' (p. 38)

Boomer and Laver's (1968) definition echoes Sturtevant's: 'A slip of the tongue ...is an involuntary deviation in performance

from the speaker's current phonological, grammatical or lexical intention.' (p. 4) It is because such 'unintentional' or 'involuntary' errors may result in utterances which provoke laughter, that individuals or writers have intentionally used them. Such conscious 'creations' will not be considered here, although one finds that these 'intentional errors' usually follow the same 'rules' as do non-intentional errors.³

Meringer was mainly interested in classifying the kinds of errors which occurred in spontaneous speech, and since Meringer, one finds in the literature different classification schemes and varying terminology.

In Boomer and Laver's classification scheme, we note that speech errors show a 'misordering of units in the string, omission of a unit, or replacement of a unit.' (p. 5) According to them, the units so misordered, omitted or replaced may be segments, morphemes, or words. Nootboom (1968) classifies segmental errors as 'Phonemic speech errors', and 'Non-phonemic errors' including in the latter classification 'meaningless combination of phonemes', morphemes (including affixes and root morphemes) and whole words. Nootboom dismisses the possibility that 'distinctive features' behave 'more or less (like) independent elements just as phonemes do' but Hockett (1967) implies the independence of such features. (915)

Further classification is not the concern of this paper. The interest is rather in how particular errors shed light on the underlying units of linguistic performance, and the production of speech.

What is apparent, in the analysis and conclusions of all the linguists and psychologists dealing with errors in speech is that despite the semi-continuous nature of the speech signal, there are discrete units at some level of performance which can be substituted, omitted, transposed, or added. It should be stated here that were we to find no evidence in actual speech production or perception for such discrete units, this would not be sufficient cause to eliminate discrete units in phonology or syntax. The fact that it is impossible to describe the grammars of languages without such units is evidence itself for their postulation in a theory of grammar. But when one finds it similarly impossible to explain speech production (which must include errors made) without discrete performance units this is further substantiation of the psychological reality of such discrete units. In other words, behavioral data of the kind described here may not be necessary to validate hypotheses about linguistic competence, but certainly is sufficient for such verification.

3.1. The Reality of the Segment or Phone

By far the largest percentage of speech errors of all kinds reveal the substitution, transposition (methathesis), omission, or addition of segments of the size of a phone. These occur both within words and across word boundaries, the greatest percentage in our corpus being the latter. The largest number of these segmental errors are errors of anticipation, which is in keeping with the conclusions reached in the literature. Simple anticipations result in a substitution of one sound in anticipation of a sound which occurs later in the utterance, with no other substitutions occurring. Examples in I. illustrate such errors.

- I. a. John dropped his cup of coffee → --- cuff --- coffee
- b. also share → alsho share [ɔl]o [er]
- c. such observation → sub ...such ---
- d. delayed auditory feedback → --- audif --- auditory ---
- e. week long race → reek long race
- f. M-U values [ɛm juw vɔljuwz] → [ɛm vjuw] values
- g. the third and surviving brother → the sird and ...the bird ... the third ---

Examples I.a. through I.e. illustrate the substitution of one segment for another. In I.f. however, anticipating the future [v] a segment is added where there was no segment in the intended word. And in I.g. the error is compounded: first the s is anticipated, and then in an attempt to correct the error a later b is anticipated.

Perseverance errors are also not uncommon as exemplified in II.

- II. a. I'm not allowing any proliferation of nodes → --- proliferation
- b. John gave the boy → --- gave the goy
- c. Spanish speaking people → --- speaping people
- d. irreplaceable → irrepraceable
- e. Chomsky and Halle → Chomsky and Challe

- IV. a. Kathy can type → tathy Kathy can type
- b. correct class of → collect correct --
- c. shown in the present slide → shown in the pleasant...
I mean present slide
- d. greater pressure → greyfer ... greater pressure
- e. delayed auditory feedback → delayed audiff auditory
feedback

All the above examples reflect the anticipation of consonants. Vowels are also anticipated, metathesized etc. as shown in V.

- V. a. ad hoc → odd hack [æd hæk] → [ad hæk]
- b. Wang's bibliography → Wing's babliography
- c. turn the corner → torn the kerner [tɔrn] --- [kɔrnɹ]
- d. feet moving → [fuwt mijving]
- e. fish and tackle → fash and tickle [fæʃ] --- [tɪkl]
- f. the new Sony → the no suny [no sunij]
- g. place the stress → [plɛs] the [strejs]
- h. dissertation topic [dɪsərtɛjŋ təp k] → [dɪsərtɛŋ tɛjpɪk]
- i. available for exploitation → avoilable for ...
- j. prevailing temperature → [prejvijliŋ]
- k. the waterfall [wɔtərfɔl] isn't working → -- isn't wokɪŋ

3.2 Clusters as sequences of discrete phones or segments

The above examples show errors of transposition, substitution, omission and deletion of individual segments, which may be either vowels or consonants. The error may be either anticipatory (i.e. the interfering segment follows the error), or one of perseveration (i.e. the interfering segment precedes the error) or one of transposition (i.e. the order of sound segments is changed). Further justification for assuming that individual segments are units in speech performance is suggested by the fact that many errors in

which the intended utterance included consonant clusters reveal that only one segment of the cluster is involved.

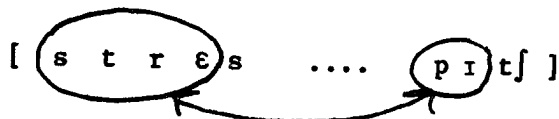
- VI. a. fish grotto → frish gotto
- b. fresh clear water → flesh queer water
- c. split pea soup → plit spea soup
- d. brake fluid → blake fruid
- e. no strings attached → no strings attached
- f. at the Broadway stores the prices are → --- spores
the prices are
- g. in a split second → --- slit second
- h. that's a sticky point → --- spicky point
- i. a car with a stick shift → --- [tɪk sɪft] 4.

As is seen in VI.a., the intended fish grotto has been pronounced frish gotto [frɪʃ gaDo] (the [D] represents a voiced flap), the addition of an [r] in the first word producing an initial cluster instead of the intended single segment. The substitution of the single [g] for the intended cluster [gr] may be explained by postulating that the cluster [gr] is composed of two segments [g] followed by [r] which can be 'broken down' into individual segments. This being so, the individual segments can themselves be transposed. Similarly the error cited as example VI.b., can be explained by an anticipation of the [l] in clear creating the substitution of the intended [fr] in fresh by a [fl]. The substitution of [kʷ] in [kʷɪr] for the intended [klɪr] may again be explained by an anticipation of the [w] in water. It is of course true that example VI.b., may be simply an error in word substitution, since flesh is a word, as is queer. Such an explanation will not, however, explain a number of the other examples given, i.e. [frɪʃ] is not a word, nor is [gaDo], nor is [plɪt], [spiʃ], [blejk], [fruɪd], [ətrɪtɪt], [sprɪkɪj] etc. If we are looking for an explanation of how such errors can occur, it seems highly likely that just as in the examples of single segmental errors, we find here also single segmental errors, the difference being that the segments involved occur in consonant clusters.

- X. a. pussy cat → kassy put
 b. foolish argument → farlish ...
 c. a heap of junk → a hunk of jeep
 d. stress and pitch → piss and stretch
 e. lost and found → faust and [lond]

Example a. shows the monosyllable [kæt] as a sequence of three segments, [k + æ + t] with the first two segments transposed with the first two segments of [p + s + t].

In d. the transposition which occurs can easily be explained as



Another explanation which can be put forth is that the word piss is substituted (the reasons for such a substitution I leave to Freud) for stress and stretch for pitch, or, instead, that the speaker started to say pitch and stress and the error is one of final consonant substitutions. There are, however, numerous examples which show errors involving CV or VC sequences which can not be so explained.

3.3 Affricates

The assumption that clusters on a performance level should be interpreted as sequences of consonants raises the question of affricates. It is interesting to note that while [str], [pl], [kr], [bl], [fr] etc, as well as final clusters reveal the splitting of clusters into segments, not a single example in my own data, or the English examples cited by others shows a splitting of [tʃ] or [dʒ] into sequences of stop plus fricative.

- XI. a. pinch hit → pinch hitch, but not *[pint hiʃ]
 b. pretty chilly → chitty pilly [tʃɪtɪj pɪlɪj]
 c. entire chapter → enchire [ɛntʃəjɹ]

- d. further surgery→ furdjer [fərdʒər] surgery
- e. Ray Jackendoff→ Jay Rackendoff
- f. last cigarette Tim had in June→ ... Jim had in time
- g. in St. Louis John said→ in St. Jouis John said.

We do not find cases of St. [duəs], or st. [ʒuis].

One may assume that the old phonemic controversy as to whether such affricates should be considered as one segment or two, is solved for linguistic performance and that affricates should be considered as single segments in the production of speech.

3.4. Complex Vowels

One finds a similar situation with vowel diphthongs. If [ey] or [uw] or [əw] are interpreted as a succession of V + y, or V + w, one could expect the non-glide section of the diphthong to be subject to substitution without a change of the particular glide. In other words one would anticipate that feet moving might be articulated as [fuɪt mɪwɪŋ]. Examples XII show that where a VG or a [Vr] is involved the error always includes the entire diphthong or the vowel with its 'r quality.'

- XII.
- a. first and goal to go→ first and girl to go
 - b. took part in the first→ took pirt [.] in the farst
 - c. dissertation topic→ [dɪsərtəʃn tɛjprɪk]
 - d. we're going to have to fight very hard → we're going to have to fart very [fayd]
 - e. feet moving→[fuɪt mɪjvɪŋ]
 - f. available for exploitation→ available for ...
 - g. sop is served→ serp is suvd

These examples are of course only English examples and the conclusion regarding affricates and complex vowel nuclei have meaning only for English.

Furthermore, the fact that one never finds an error which results in a 'non-permissible' sequence of, for example, front vowel plus back glide (e.g. [ɪw], [ɛw]) or back vowel plus front glide (e.g. [ʊj]) may have an alternative explanation which is discussed below. One example above (V.i.) (available [əvejləbl] available [əvɔjləbl] could be interpreted as a switch only in the non-glide portion of the vowel nucleus as could all examples of errors which involve all tense front vowels or tense back vowels. The errors involving both front and back diphthongs, and those involving a vowel followed by r cannot be explained in this way, and seem to suggest that the complex vowels are single units or that errors which 'violate' phonological constraints are 'corrected' after the substitution occurs. (See below for discussion on this point).

3.5. The status of [ŋ] in English

Sapir (1925) and Chomsky and Halle (1968) present arguments for deriving [ŋ] from an underlying sequence of /ng/. Their phonological analysis is justified in itself. It is of interest, however, that behavioral data, found in speech errors, point to the fact that at one level of performance, [ŋ] may derive from the sequence of [n + g] or (because of the constraints which change an [n] → [ŋ] before a velar, the sequence of [ŋ + g].

XIII. a. sing for the man [sɪŋ mæn] → [sɪ g mæn]

b. Chuck Young [tʃʌk jʌŋ] → [tʃʌŋk jʌg]

c. shilling [ʃɪ lɪŋ] → shingle [ʃɪŋgl]

A possible explanation for the [g]'s in the actual utterances, is to postulate that prior to the execution of the articulatory commands the following transposition of segments occurred:

a. [sɪŋg....mæn] → [sɪθg....mæn]

b. [tʃʌj jʌŋg] → [tʃʌŋk jʌθg]

c. [ʃɪlɪŋg] → [ʃɪŋgəl]

If this highly speculative hypothesis can be demonstrated by other experimental data the postulated phonological rule for English, $g \rightarrow \emptyset / n\_\_\#$ appears to have reality in that when the nasal is deleted the [g] emerges.

The data however can be given an alternative explanation. Example a. may show a persistence of the velar articulation from [sɪŋ] producing [mæŋ], and a simple loss of the nasality of the final velar in sing.

In b. since in English a vowel is nasalized preceding a nasal consonant, this may be an example of a transposition of oral vowel with nasal vowel, and a concomitant non-nasalization of the final nasal i.e.

[tʃʌk jɪŋ] → [tʃʌkjʌg]

Example c. may be similarly disposed of. The examples are given, however, since they permit speculation as to the reality of the g in [ŋ] utterances.

3.6. The Reality of Phonetic Features

Research on the perception of speech has shown that units smaller than the segment properties or features of speech sounds are 'perceived' and confused. (Miller and Niceley, 1955, Wicklegren 1965, 1966 et al.) Thirty two errors in the corpus used in this study can be explained by postulating that certain properties or features constitute independent elements in the production of speech, as well. The fact that one finds no errors in which consonants and vowels are involved (i.e. vowels do not switch with consonants etc.) may be explained by suggesting that true vowels ([+ vocalic, - consonantal]) constitute one class of segments in a performance model, as opposed to another class composed of true consonants, glides and liquids [+ consonantal] or [- vocalic] but that the segments which are members of these two non-intersecting sets cannot be further analyzed into independent features.

As we shall see below, there are other explanations for why consonants and vowels do not 'interfere' with each other (e.g. are not transposed, anticipated etc.). The data, however, suggest that while a hierarchy probably exists, other features are independently involved in speech errors.

- XIV. a. spell mother → smell [bAðr]
- b. pity the new teacher → mity the due teacher...I
mean...nity the poor teacher...no...pity the new teacher
- c. bang the nail → mang the mail
- d. Did you hear what Malcomb said → did you hear what
balcomb...malcolm said?
- e. Cedars of Lebanon → Cedars of Lemadon

Examples under XIV. show a change in the value of the feature nasality, acting in many cases, independent of other features. In a. and b. the [- nasal] of [p] becomes [+ nasal] (i.e. [p] → [m]. If the [m] of mother remained [+nasal] this example could be dismissed as merely an anticipation of the segment [m]. However, since [m] → [b], or since the value of the nasality feature in the [m] of mother switches from [+ nasal] to [- nasal] all other features remaining the same, a better explanation for the error is that what occurred was a single feature switch. Otherwise, no explanation is provided for the [m] → [b] substitution.

Example b. illustrates the same phenomenon. [p] which is [- nasal] becomes [m] which is [+ nasal], all other features remaining intact, and [n] is changed from [+ nasal] to a [d] which is [- nasal].

Example c. shows a switch of two features. [b] which is [- nasal, + anterior, - coronal] switches to [+ nasal], and [n] switches from [+ coronal] to [- coronal].

Even if one wished to explain the [m] of mail as a perseveration of the [m] which has occurred in [mæn], the substitution of [b] → [m] would be left unexplained. The anticipation of the lowered velum which accompanies the following [n] seems to be some explanation of what occurred.

Example d. represents a simple substitution in the first segment of Malcomb from [+ nasal] to [- nasal].

Examples XV. represent a change of the value for the feature [Voiced].

- XV. a. What does the course consist of → what does the gorse consist of
- b. referendum → reverendum
- c. clear blue sky → glear plue sky
- d. reveal → refeel
- e. define → devine
- f. big and fat → pig and vat

In these example only the value of the feature [voice] is changed, all other features remaining intact.

Other errors which appear to involve properties or features of whole sounds rather than the segments as a whole are given under XVI.

- XVI. a. pedestrian → tebestrian ([p] → [t] and [d] → [b])
- b. scatterbrain [skæDərbrejn] → [spæ Dərgrejn]
- c. spaghetti → skebeti
- d. death valley [dæθ væliʃ] → [fæθ ɔəl]

In example XVI. a. only the value of the feature [coronal] is switched. In other words, only the place of articulation is changed.

It is of course possible to argue that this is rather to be interpreted as segmental transposition, with [p] → [t] in anticipation of the subsequent [t]. This however does not explain why [d] → [b], which can be explained by [p] → [t] being a switch of labial ([- coronal]) to alveolar ([+ coronal]), and the [d] → [b] the result of a change from alveolar ([+ coronal]) to labial ([- coronal]).

Similar errors are seen in examples b. and c.

Again one can suggest a segment transposition, particularly since the voicing feature of the [g] is neutralized after an [s]. But then how does one explain the [g] → [b] switch. If a mere segment transposition was involved we would expect [sgəpɛtiʃ].

A more complex error is seen in example d.

The switch from [d] → [f] seems to be an anticipation of the subsequent [v]. Why the voicing feature is switched is unexplained. The coronality of the [d] seems to influence the switch from [v] to [d] with the [+ voicing] and [+ continuant] features remaining.

It is certainly true that errors which involve a substitution of features are rare compared to errors involving larger units. They nevertheless require some explanation and one can only conclude that some features appear to be independently extractable as performance units. Many of the segmental errors may also be examples of such feature errors, but since they can also be accounted for as errors of larger units we are unable to conclude that individual features are independently involved in all cases. However, the examples in XVII show that feature errors may be obscured in this way.

- XVII. a. extracted → extrapted [k] → [p] or $\begin{bmatrix} +\text{back} \\ -\text{ant} \end{bmatrix} \begin{bmatrix} -\text{back} \\ +\text{ant} \end{bmatrix}$
- b. goofing off → guping off [f] → [p], or [+cont] → [-cont]
- c. call the girl → gall the curl [k] → [g], or [-voice] → [+voice]
- d. documentation → documendation [t] → [d]/d .x.--, or [-voice] → [+voice]

In fact, most of the segmental errors can be so interpreted. In the transposition of brake fluid to blake fruid, one might suggest that what is involved in a transposition of the feature lateral ([anterior]) rather than transposition of the two segments. If segmental errors are analyzed as feature errors we would find that many of the distinctive features other than those cited above do indeed represent a reality, in speech performance.

This suggestion is supported by the findings of Nooteboom (1968): 'In significantly more cases than is to be expected in a random distribution the two elements involved in a substitution error are phonetically similar to one another.' .7). MacKay (1969) found that 'Most pairs of reversed consonants differed in only one distinctive feature (56 percent) and very few (2 percent) differed in all four distinctive features.' (p.12-13) ⁵. This is in contradiction to the conclusion of Boomer and Laver 'that articulatory

similarity is not an important determinant' in speech errors (p.8), although they do note two exceptions: 'sequences of voiceless fricatives seemed to encourage mistakes of place of articulation, and (b) alveolar consonants showed a slight tendency to interact.' (p.8 fn). But they were examining the errors to see if any particular features were involved more often than any others. Their data was not analyzed as to the degree of similarity of the segments which were involved. It is interesting to note that an analysis of jargon aphasia errors also show that most errors involve no more than a confusion of one distinctive feature. (Green, 1969) Whether or not further analysis of substitution errors confirms or contradicts the MacKay-Nooteboom conclusions regarding the 'similarity' of substituted segments, the only conclusion one can draw from the examples of feature switching given above is that at least some of the proposed distinctive features are independent behavioral units.

But an examination of the errors, whether analyzed as errors of whole segments, or of independent features, definitely show a hierarchy and interdependence of certain features. Thus, while there are errors showing just addition or subtraction of nasality, one does not find a 'nasality' switch which results in a voiceless nasal. At least for English, nasality and voicing seem to be interdependent features. This again prevents the occurrence of an 'inadmissible' sound in English.

The claim that certain features are independent units, and must be postulated as such in a model of performance, seems to contradict the earlier hypothesis that segments (or feature complexes) are 'real' performance units. Actually, there is no real contradiction. Features cannot exist except as properties of larger segments (just as segments, as we shall see exist as parts of larger units). In other words, in the generation of speech, there is a hierarchy of different sized units. A linear ordering of the segments (discussed below) occurs, and this linear ordering may be disrupted. Since the discrete segments are specified by actual physiological properties (or neural commands) some of these properties or features may also get disordered, i.e. 'attached' to other segments. The claim, however, that all distinctive features which are proposed by Chomsky and Halle (1968) are identical with phonetic properties that can in principle be independently controlled in speech is not borne out by the data of speech errors. Unless 'controllable in speech' is defined in some new strange and abstract way it would appear that whatever the needs for certain separate phonological features may be, in actual speech performance only certain of these phonological features have their counterpart

as phonetic features.⁶ Thus, while the two features [consonantal] and [vocalic] very nicely divide segments into four separate classes, needed for phonology, the idea that such features have any independent phonetic reality seems highly improbable. To suggest that a substitution of a [p] for a [k] involves the phonetic substitution of [- anterior] for [+ anterior, [- high] for [+ high], [- back] for [+ back] etc. is saying no more, on an articulatory level than stating that there is a change from a velar articulation to a bilabial articulation. The motor commands to the muscles specifying a bilabial or velar articulation specify the part of the tongue to be raised, where it is to be raised etc. In other words, on a phonetic level a complex of the features [- anterior, + high, + back] is indissolubly a velar place of articulation, and one does not expect to find (as indeed, one doesn't find) a simple switch of the feature [+ coronal] for example, without other phonetic effects. In the example given above, where pedestrian → tebestrian, the error can be specified as a switch in the feature of coronality, but it is obvious that this feature is not 'independently controlled.' What I am suggesting is that segments as feature complexes do exist, and that some of these features or properties can be independently controlled, such as nasality, voicing, place of articulation (if considered as a single multivalued feature) etc, and that other properties are highly dependent on the existence of other properties of the segment. It is thus that [delayed release] does not seem to be independent of affricates, and one can only suppose that on the neuro-physiological level there is some command for a stop closure combined with delayed release, which command cannot be segmented into two segments. What I am claiming here is that the command for the initial and final consonants of church at one level of the generation of speech is a command for just such an affricate. On the other hand in rapid speech when one says did you as [dɪdʒu] the affrication occurs by a different process, i.e. by automatic and mechanical movements of the vocal organs. The results, however, at the level of muscle movements are identical.

3.6 The Reality of the Syllabic Unit

While it seems plausible to assume, as was done above, that units smaller than syllables (segments and features) have independent status as behavioral units, this does not negate the possibility that syllable size units are also units of speech performance. In fact all the evidence from tongue slips support such a view. Nooteboom (1967-68) suggests that since 'the distance between origin and target (or the substituted segments) does not generally exceed seven syllables, (and) since we know that the short memory span of man

may contain about seven units...we might interpret our findings as an argument for the syllable to be a unit in the phonemic programming system.' (p. 5-6). Nooteboom, (1963-68, MacKay (1969-70), Boomer and Laver (1968), and Fromkin (1968) all support the statement that 'Segmental slips obey a structural law with regard to syllable-place; that is, initial segments in the origin syllable replace initial segments in the target syllable, nuclear replace nuclear, and final replace final.' (Boomer and Laver p. 7). Furthermore, Nooteboom (ibid) points out that 'when the second consonantal element of a CVC form is immediately followed by an initial vowel of the next word...final consonantal elements do not tend to become prevocalic.' (p.7) In other words in a string CVC#VC..... CV#CVC one never finds in errors a substitution of the final consonant of the first word for the initial consonant of the final word. My own English data supports the analysis of Nooteboom's Dutch data, and seems to contradict the position taken by Kozhevnikov and Chistovich where they suggest that in the production of Russian utterances a CVC#VC sequence is reorganized into articulatory programs for each CV sequence. This does not seem to be the case in English or Dutch. The evidence for the syllable suggested by Nooteboom can, of course, also be used as evidence for the reality of the unit 'word', which will be discussed below.

MacKay (1970) found that 'the syllabic position of reversed consonants was almost invariably identical.' (p. 16)

The only examples in my data which do not support this finding are two examples of metathesis occurring within words of two sequential segments as in examples XVIII.

XVIII. a. whisper → whipser

b. ask → aks

It has been suggested (Ladefoged, personal communication) that the rarity of such examples, and the fact that all such errors seem to involve the sibilant s should not be overlooked. In a number of perception tests the hiss such as that which occurs with [s] is often 'misplaced', i.e. it is difficult for subjects to judge where the noise occurs in an utterance. This perpetual difficulty seems to be reflected in production errors of this kind.

All other examples of errors occurring within the same word show sequential ordering of segments within syllables as in Examples XIX, (a hyphen represents a syllable division).

- XIX. a. harp-si-chord → karp-si-hord
 b. ma-ga-zine → ma-za-gine
 c. phi-lo-so-phy → phi-so-lo-phy
 d. el-eph-ant → eph-el-ant
 e. an-im-al → am-in-al
 f. spe-ci-fy → spe-fi-cy
 g. ra-be-lais → ra-le-bais
 h. pan-cakes → can-pakes
 i. neural mo-de-ling → neural mo-le-ding

Because of the coarticulation effects of segments within a syllable it is impossible to omit the syllable as a unit of articulation, even if one were to ignore the evidence of the fixed order in the reversal, anticipation or perseveration of segments. (Fromkin, 1968)

There are of course many errors which involve the substitution, omission, replacement, addition, etc. of one or more whole syllables, which further substantiates the claim that syllabic units are real performance units.

- XX. a. Morton and Broadbent point → Morton and broadpoint
 b. revealed the generalization → reeled the generalization
 c. tremendously tremenly
 d. which I analyze as the following → which I analyze as the follow
 e. butterfly and caterpillar → butterpillar and catterfly

XX. f. opacity and specificity → opacity and specifity

g. we want to reveal all the fact → we want to feel all the facts.

Many of the above examples reveal errors which involve several factors at work. Some of the above are what are commonly called 'blends' as are the examples in XXI.

- XXI. a. Did you bring your clarinola (a blend of clarinet + viola)
 b. switch and changed → swunged [swɪndʒd]
 c. importance of [ədʒəʊjsnt] rules (a blend of 'adjacent' + 'adjoining')
 d. my data consists [monliʃ]....[mejstliʃ]... (a blend of mainly + mostly)

The above examples, showing the fixed order of the syllabic segment involved, the errors involved enter syllables, as the co-articulation evidence seems to justify the reality of a syllabic unit in performance.

4. The Reality of Phonological and Morphological Constraints

The speech of jargon aphasics as well as errors made by non-pathological speakers all reveal that 'normal' slips of the tongue and aphasic jargon utterances are constrained by the linguistic system. One does not find 'phonemes' or more correctly 'phones' which are not found in regular utterances. For example, an English speaker does not substitute a rounded front vowel in anticipation of a rounded back vowel, nor a lateral click for a lateral liquid. Furthermore, only permitted segmental sequences occur. Rulon Wells (Wells, 1951) stated this as his 'First Law' of tongue slips 'A slip of the tongue is practically always a phonetically possible noise.' It is obvious that Wells means a 'phonetically possible noise' in a particular language. As stated in an earlier article (Fromkin 1968) 'The segments constituting each syllable must have sequential ordering, so that only initial consonants or clusters, vowels, and final consonants may interchange, if and only if the transpositions are in keeping with the phonological rules of the language.' (64) This 'First Rule' appears to explain a 'Spoonerism' attributed

to Spooner.

*sphinx in moonlight → minx in spoonlight.

What is of interest here is the transformation of the [sf] in sphinx to [sp] when the cluster is transposed with the [m]. While [sf] does occur in words like [sphincter], sphere, and sphinx, such words (and the dozen or other 'technical' words listed in WNI, 3rd Edition) are 'exceptions' to the regular morpheme structure rule in English which permits only voiceless stop obstruents after an initial s. [sfuwnlight] would thus not be a permitted sequence, and consequently the [f] → [p].

All the examples already cited include only permitted English sequences.

Further examples will support the 'reality' of such constraints.

XXII. a. play the victor → flay the pictor

b. tab stops → tap [stabz]

c. plant the seeds [sijdz] → plan the seats [sijts]

c. bloody students [blʌdi stuwðnts] → bloodent studies [blʌðnt stuwðiz]

There are two ways of interpreting the error shown in example a. One might suggest that it is simply the manner of articulation (stop ↔ fricative) which is switched. If such an interpretation is given, one must also add that when the [v] is changed to a stop the place of articulation changes from labio-dental to bilabial. Another possible explanation is that the two segments switch (p ↔ v), and that since [vl] is not a permitted sequence in English the [v] is devoiced. This suggests that these phonological constraints, when learned, become behavioral constraints which occur after the segmental transpositions occur.

A similar example is shown by b. in which the final consonant (or just the voicing feature) of the first word is transposed with the penultimate consonant of the second word (or the final stem consonant, prior to the plural morpheme addition). Again, the intended [ps] is changed not to [bs] but to [bz] in keeping with the phonological (and morphological) constraints of English.

Examples b, c, and d, represent another phenomenon. In the errors that occur the original syntactic structure of the phrases remain intact in that the intended plural nouns occurring as the last words of the phrases, despite the errors which occur, remain as words with plural endings, but the phonetic realization of the plural morpheme changes as well as the segments which occur before them. While [stabz] and [sijts] can be explained simply as due to phonological or phonetic constraints, since [bs] and [tz] never occur as final clusters, the error in example d, is more complex. [js] can occur in English as in Reese [rijs] mice [majs], feast [fijs] face [fejs] etc. But [ij+s] cannot occur when the final sibilant represent the plural morpheme. One can then suggest that the phonetic representation of the plural morpheme is specified prior to the automatic phonetic specifications which serve as the units for articulatory commands. If this was not the case, one can not understand the change of the [s] to [z] in [stuwdijs].

Further examples of the reality of morphophonemic rules are evidenced in errors which include the alternation of the non-specific determiner a/an.

- XXIII. a. a current argument → an arrent curgument
[ə kʌrnt ʌrgjʊmənt] → [ən ʌrnt kʌrgjʊmənt]
- b. an eating marathon → a meeting arathon
- c. a history of an idiology → an istory of a hideology
- d. an ice cream cone → a kise ream cone

The changes from a → an, and from an → a indicates that in the generation of speech the segmental errors, or transpositions must take place prior to the actual neural muscular commands, since they are possible sound sequences as in America is ...[-əɪz] etc.

5. Stress

MacKay (1970) Boomer and Laver (1968) and Nooteboom (1968) all investigated the influence of stress on errors in speech. Boomer and Laver concluded that 'The origin syllable and the target syllable of a slip are metrically similar, in that both are salient (stressed) or both are weak (unstressed), with salient-salient pairings predominating.' (p.7) Nooteboom agrees with this conclusion stating

that 'In significantly more cases than is to be expected in a random distribution the elements involved in a speech error belong to stressed syllables! (p.8) He disagrees, however, with Laver and Boomer's finding that 'Slips involve the tonic (primary stressed) word, either as origin or as target, with tonic origins predominating.' But from Nooteboom's own data the disagreement seems to be the result of a misinterpretation of the difference between primary stress (tonic word) and salient stress. The differences may also be due to differences between English and Dutch. MacKay found that transpositions occurring within words occurred in syllables with different stress while in between-word reversals his findings corroborate those of Laver and Boomer.

What seems to be of greater interest is that when vowels or syllables or parts of syllables or whole words are substituted or transposed there is no change in the stress pattern or contour of the sentence. Laver and Boomer cite an example in which a speaker, instead of saying 'how *bad* things were', said 'how things *bad* were.' As is evident by the placement of the acute accent which represents primary stress, there was no transposition of the stress, despite the transposition of the words. The examples under XXIV show the same phenomenon. (an acute accent ['] represents primary stress, as does '1' above the vowel; a grave accent ['] represents non-primary stress -- secondary or tertiary; a '2' above the vowel represents secondary stress, and a '3' tertiary stress).

- XXIV. a. ³hammer and ¹sickle → ³sickle and ¹hammer
- b. ³peoples ¹park → ³parklz ¹pījp
- c. ²verge of a ³nervous ¹breakdown → ²nerve of a ³vergous ¹breakdown
- d. he's been ²around a ³long ¹time → he's been ²long ³around ¹time
- e. a ²computer in our ³own ¹laboratory → a ²laboratory in our ³own ¹computer
- f. ²examine the ³eyes of the ¹horse → ²examine the ³horse of the ¹eyes
- g. ³broke the ²crystal on my ¹watch → ³broke the ²whistle on my ¹crotch
- h. in the ²theory of ¹phonology → in the ²phonology of ¹theory

Examples e, and h, show that while the word position of primary stress in the phrase is not transposed, the stressed syllable of the word in isolation is the syllable which receives the sentence stress. That is, if the primary stress is to be placed on 'laboratory' it would be placed on the first syllable, and if it is to be placed on 'computer' it is placed on the second syllable.

Thus there seems to be two aspects of stress to be accounted for: first, the word stress moves with the word itself (i.e. the syllable of the word which in isolation receives main stress, receives the primary stress when the word is moved); second, the stress contour of the phrase is fixed by the syntactic structure of the phrase itself, and must be generated independently of the word order in the utterance.

One may then suggest that the word stress is stored as part of the articulatory specifications of the stored unit 'word', but the sentence or phrase stress and overall intonation contour is generated separately as part of what Boomer and Laver call the 'tone-group.' I would therefore agree with them that 'The pivotal role of the tonic word in slips suggests that its phonological, syntactic and semantic prominence is matched by an analogous neuro-physiological prominence, coded in the brain as a part of the articulatory programme' (p.8) and further with their view that 'the tone group is handled in the central nervous system as a unitary behavioural act, and the neural correlates of the separate elements are assembled and partially activated, or 'primed' before the performance of the utterance begins.' (p.9)

But, in the construction of a model of linguistic performance it is necessary to specify the nature, i.e. the syntactic structure, of this tone group, for the 'priming' of the 'tonic' syllable depends on the syntactic structure of the utterance.

The suggestion above that the stress placement on words is fixed in the lexicon does not mean that one can not, or should not, attempt to generalize stress assignment rules in the phonology of English. In fact there may be some evidence from speech errors that not only in a grammar of competence, but also in the actual stored lexicon words (or perhaps formatives) are stored in a more abstract form than by their actual articulatory specifications. There are speech errors which display a movement of stress, and in certain cases a change in the vowel qualities depending on where the stress is placed.

- XXV. a. This can be viewed alternately....alternatively...no,
....alternately
- b. símilarly → simlæ̀rəlii
- c. homogéneous → homodʒəniyəs
- d. əblígatory → áblətəriɹ ábləgətóriɹ
- e. in favor of [həmádʒə....homodʒiɹniəs]
- f. syllabíf....syllábification → [sɹləbíf....sɹlábəfəkeɹjɹn]
- g. opácites and specifícity → opácity and spécifity

One may speculate (probably wildly) that in the generation of speech a word is selected, stress is assigned and then the articulatory program is assembled to produce the sounds, reducing unstressed vowels etc. By this hypothesis the differences between [ər] → [æ̀r] in b., [o] → [a] and [iɹ] → [ə] in c., [ə] to [á] and [í] to [ə] in d., are 'explained' by the suggestion that the words are stored as stems plus endings, and with their unreduced vowel qualities. While such a suggestion can not be entirely ruled out, for all the examples in XXV above, alternative explanations can be provided which, from the performance viewpoint, seems more intuitively satisfying. In b, for example, the speaker might have begun to say 'similarity', or in Carroll's terms have had 'similarly' and 'similarity' in mind at the same time just as he clearly had both 'alternately' and 'alternatively' in mind in a. It should be clear, without belaboring the point, that all the above examples of errors involving stress can be similarly explained. Before one can seriously put forth the hypothesis that the stress of words is generated by phonological rules, and not stored as part of the specification of the word (in one's performance lexicon) a crucial experiment must be found.

6. The Reality of Syntactic Word Classes and Syntactic Phrases

Nooteboom found that 'a mistakenly selected word always or nearly always belongs to the same word class as the intended word (indicating) that the grammatical structure of the phrase under construction imposes imperative restrictions on the selection of words.' In my own corpus of errors, a similar conclusion can be

drawn. When words are switched, nouns transpose with nouns, verbs with verbs etc.

- XXVI. a. a computer in our own laboratory → a laboratory in our own computer
- b. that no English manufacturer could name these projects...
products..
- c. naturalness of rules → nationalness of rules
- d. bottom of page five → bottle of page five
- e. I have some additional proposal to hand out → hang out
- f. book of sixes → book of twos
- g. chamber music → chamber maid
- h. a speaker doesn't go through all the worlds...rules he has in his head
- i. While the present ...pressure indicates
- j. How come if you're a Scorpio you don't read...wear oriental spice?

The fact that in many cases the substituted word has some phonetic (or phonological) similarity to the target word was also noted by Nootboom. This suggests that our stored lexicon is ordered in some dictionary like fashion. Any cross-word puzzle addict can attest to this fact. But there must be a complicated addressing system in the computer-like brain mechanism since each listing must be specified under semantic features, phonological features, number of syllables, syntactic features etc. Thus, in example h, the phonetic similarity of [wɜːldz] and [ruːlz] is based on two identical segments which however do not have the same sequenced ordering in the words. This, of course may be a chance error.

The reality of the word as a unit is evidenced by the above. Furthermore, speech errors show that derivationally complex items may be stored as combinations of separate formatives, i.e. stems and affixes. Example c, above, natural+ness → national+ness attests

to this as do the following examples:

- XXVII.a. infinitive clauses → infinity clauses
- b. grouping → groupment
- c. intervening node → intervenient...intervening node
- d. and so in conclusion → and so in conclusement

Example d, suggests again that conclusion may be stored as [conclude + ion] with rules changing the d → [ʃ]. It is however possible and highly probable that we have here a blend of concluding and conclusion.

Hockett's analysis would relegate such affix substitutions to what he calls 'analogy.' Unfortunately, analogy does not explain how such analogy takes place. One possible explanation is there are rules of word formation, plus a vocabulary of stems and a vocabulary of affixes which, as the above examples show, can be manipulated to create neologisms which do not occur in the language, such as 'groupment'.

MacKay's (in preparation) finding that among syllable errors, affixes are involved with a probability greater than chance would support the hypothesis that affixes do form a separate sub-set of the lexicon.

The constancy of the syntactic structure, and the reality of performance units larger than words, morphemes, stems etc, is seen in the following:

- XXVIII. a. I wouldn't buy macadamia nuts for the kids → I wouldn't buy kids for the macadamia nuts
- b. A fall in pitch occurs at the end of the sentence → an end of the sentence occurs at the fall in pitch
(Notice here too the switch from a → an)
- c. He's a far better man than anyone here → he's a farther man than anyone better here

(The displacement of better also results in an adjectival ending added to the adverb 'far', maintaining a correct and intended syntactic structure).

In structures such as NP[Adj[macadamia] N[nuts]] (or, this may be a compound noun) and N[N[fall] pp[in pitch]] and N[N[end pp[prep[of] NP[[det the] N[sentence]] syntactic phrases can interchange, as entire units. Also the similar word classes interchange. Further when (as in c,) an intended adverb + adjective + noun is involved in an error a shift in the adjective to another place in the sentence seems to simultaneously change the remaining adverb 'far' to an adjective, thus maintaining the overall structure. Such facts seem to justify the real existence of syntactic phrases and the syntactic features of words.

7. Semantic Features

Blends occur in which non-existent words are produced as the result of composites of two words with similar semantic features. These are indeed common errors, not only invented by Lewis Carroll, but occurring naturally. In the examples given in XXIX the speaker was questioned as to what he had in mind, or as to what he thought the reason for the blend was. The subject's answers are given in parentheses.

- XXIX. a. My data consists [mownlij]...[mejstlij].... (for mainly/ mostly)
- b. I swindged [swindZd]....(switch and changed)
- c. It's a lot of [ba]....[brAðl] (bother and trouble)
- d. She's a real [swɪp] chick (swinging/hip)
- e. it's a [spajrətɪv] (spirant/fricative)
- f. a tennis [æθlɜr] (player/athlete)

Such errors seem to support Carroll's assumptions. A speaker has in mind some meaning which he wishes to convey. In selecting

words it appears that he is matching semantic features and where there are a number of alternative possibilities rather than make an immediate selection he brings them both into a buffer storage compartment, with their phonological specifications. Either a selection occurs at this point, or the words are blended resulting in the above kind of errors.

The literature and my own data attest to the fact that besides the phonological similarity in substituted words, very often there are semantic features in common, or antonyms substituted, i.e. words with the same features with opposite values.

- XXX. a. I really like to ... hate to get up in the morning
 b. It's at the bottom...I mean...top of the stack of books
 c. This room is too damn hot...cold
 d. the oral...written part of the exam...

Nooteboom presents a number of examples which seem 'to involve a semantic switch from the space to the time dimension' (p. 14) as in the following:

- XXXI. a.*the two contemporary, er...sorry, adjacent buildings
 b.*during the apparatus, er...behind the apparatus
 c.*the singular, sorry, the present time

Evidence from aphasia studies also show that substituted words often fall into the same semantic class, as in cases where patients will read 'tree' for 'flower', 'night' for 'dark', 'spoon' for 'fork', 'liberty' for 'democracy' etc. (Marshall 1966, Luria 1959, Jakobson, 1966 et al).

Such errors provide important evidence as to the storage of vocabulary and the generation of speech.

8. Implications of Speech Errors for a Model of Linguistic Performance

8.1. The Lexicon

When one learns a language, he learns among other things, a

vocabulary. It seems plausible to assume both from the errors of speech and from the ability speakers of a language have to form new 'words' by adding derivational affixes to stems (e.g. 'He's a real computerish type') and by inflecting newly coined words in keeping with the rules of the language (e.g. 'studies' as in [stuwdiɹz]) that the stored lexicon consists of stems and affixes, as well as idioms, compounds, whole words etc. Given the higher than chance probability (MacKay in preparation) that affixes (prefixes and suffixes) are involved in syllable errors, one can further assume that even if words are stored with their affixes, the stem and affix have a separate status. Thus it is not unlikely that grouping is stored as group + ing, which permits a substitution of ment for the affix ing. The fact that one does not find stems substituting for or transposing with affixes further justifies their separate status.

Since phonological or phonetic specifications, semantic features, and syntactic word class features all play a role in the speech errors that occur, it is obvious that the vocabulary items must be stored with such features indicated. But it is not enough to suggest that there is one dictionary-like storage starting with all words beginning with [a] and ending with all words beginning with [z], with other features given. Semantic errors show that words are selected to convey certain meanings as specified by their semantic features. And for literate speakers the listing must also specify the orthography to account for the ability of people to play 'geography' a game in which one must find a country, river, city etc. beginning with the same letter as the previous word ended with. Thus, Passaic ends with the letter 'c' which is pronounced [k] and the next player could offer Charleston, which begins with c, pronounced [tʃ]. The relationship between orthography and sound must be accounted for. Cross word puzzles, double-crostics, and the 'tip of the tongue' phenomenon also attest to this fact. It is for example often the case that in trying to remember someone's name, forgotten at the moment, a speaker will say, 'I know it begins with a 'C'. The name may be 'Cohen', which begins with a 'C' but with the sound [k]. And of course a game like 'Geography' is further evidence for the storage of words in semantic classes.

One may then suggest that the vocabulary is stored in a thesaurus like lattice structure. It is possible to conceive of this network as a listing of all the stems and affixes in some fixed phonological order, each one with all of its feature specifications, and each one with a particular address. The separate semantic section of this lexicon may then be divided into semantic

classes, with semantic features under which are listed, not the particular vocabulary item but the addresses of those items which satisfy the features indicated. One might suggest also that the listings under the semantic headings are grouped under syntactic headings such as [+ noun] [+ verb] etc. to account for the proper grammatical selection in the generation of utterances.

Since the 'tip of the tongue' phenomenon (Brown and McNeil 1966) suggests that the number of syllables is recalled -- the metrical beat of the word -- a further division under the full phonological listing is suggested. In other words, it is not unfeasible to assume that all monosyllables beginning with the same phonological segment constitute one block, followed by disyllables etc.

The error cited in XXX.a. might then occur in the following way: the speaker wishes to say (at least on a conscious level -- we leave the unconscious motivations to be explained by others) 'I really hate to get up in the morning.' At the point in the generation of the utterance prior to the selection of the words, in the 'slot' representing 'hate' the features [+ verb, + desire....] occur, and an address for a word is sought from the semantic class which includes [+ desire]. But either because of unconscious wishes, or due to a random error the address for a verb with the feature [+ desire] rather than one specified as [- desire] is selected, the item at that address called forth with its accompanying phonological features producing [layk] rather than [hejt].

This highly speculative, oversimplified model of the lexicon is suggested as a first approximation to what must be a most complicated storage mechanism. What seems certain, however, is that any model of the lexicon must account for the kinds of errors which do occur, which errors require the specification of various kinds of properties which we have called phonological, syntactic, and semantic features, and no lexicon including a single listing of items can explain what occurs.

8.2. The Generation of Utterances

It seems quite evident from all the examples of speech errors cited above that in the production of speech it is not true that 'anything goes,' that speech performance obeys no rules, and that the errors are totally random and unexplainable. (See discussion of this in Fromkin 1968) While we may not be able to explain, as yet, the exact mechanisms involved in speech errors the errors made not only are highly constrained but provide information about speech

performance which non-deviant speech obscures. In other words, had we no record of errors in which consonant clusters are split into segments, we would not be able to justify the assumption that clusters in performance are strings of individual discrete segments.

Any model of speech performance must therefore account for the kinds of errors which do occur. Such a model must therefore account for the following:

- 1) that features, segments, syllables constitute units in the production of a speech utterance;
- 2) that segments are ordered within a syllable, and that only segments similarly ordered are involved in the same error;
- 3) that 'root morphemes may be interchanged but root morphemes and an affix cannot take each other's places.' (Nootboom, p. 16) or, that words of the same syntactic or morphological class usually exchange for each other;
- 4) that the intonation contour (including the placement of primary stress) remains fixed, and is generated separately from the individual word stresses;
- 5) that morphological constraints and phonetic or phonological constraints must occur at different times in the production of an utterance.
- 6) that the non-permissible phones or phonetic sequences do not occur.
- 7) that errors may be semantic in nature, as in the case of blends or word substitutions involving similar semantic features.
- 8) that the similarity of the phonological form of words appears to play a role in word substitutions.

To account for such phenomena we may suggest the following (oversimplified) order in the actual generation of an utterance:

Stage 1) A 'meaning' to be conveyed is generated.

Stage 2) The 'idea' or 'meaning' is structured syntactically with semantic features associated with parts of the syntactic structure.

For example, if a speaker wishes to convey the fact that 'a ball' rather than 'a bat' was thrown by a boy, the utterance 'A ball was thrown' or alternately 'He threw a ball' (with stress on ball) is structured here. If it is the second structure he uses, part of the features specified for the final nouns must include [+ emphasis] together with the features selected for 'ball,' i.e. [- animate, - human, + count, + round, + used in games etc]

This suggests that the structure itself is put into a buffer storage prior to actual articulation of the utterance to account for the switching of noun for noun, verb for verb etc. when such transpositions do occur.

Stage 3) The output of stage 2 is thus a syntactic structure with semantic and syntactic features specified for the word 'slots'. In order to explain the fact that 'the tone group is handled in the central nervous system as a unitary behavioural act,' (Boomer and Laver, p. 9) one can suggest that the intonation contour, and the placement of primary stress occurs at this stage. Since a transposition of words in the utterance will cause a transfer of primary stress to the main stressed syllable of the word in a given position, one can only posit that at this stage the position of the primary stress is indicated here, and not the particular syllable. That is, the generation of the sentence intonation contour must occur prior to the selection of the words themselves.

Stage 4) We now have in the buffer a syntactic phrase with semantic features indicated and with sentence stress assigned. A lexicon lookup now occurs, going first to the semantic class subsection of the lexicon, with features being matched, obtaining the direction to go to a certain address in the overall-vocabulary. The item in the specified address is then selected, this word being specified as to its phonological segments, which are specified and ordered into syllabic units. At this stage in the process, errors resulting in a choice of a 'wrong' word may occur. Such errors may be errors in the matching of values of semantic features resulting in a wrong address being specified. The correct address may however be specified but a different address substituted which is 'in the vicinity' of the intended address. Thus if the intended word is 'like' and the produced word is 'hate' the error occurs in the selection of the wrong address in the 'semantic component' of the lexicon. But if the intended word is 'pressure' and the produced word is 'present' the correct address is obtained, but the wrong address selected, given that 'pressure' and 'present' have addresses in the same section of the vocabulary. This would be due to the phonological similarity of their first three segments. This

process thus results in a string of phonological segments, each segment specified by certain features or properties, each segment also specified as to a syllabic order, and with the syntactic bracketing remaining intact. But it is at this stage when the string of phonological segments is put into the buffer that a mis-ordering of segments may occur. In other words as the segments are 'sent' into the short-term memory buffer, segment 1 of syllable 1, may be substituted for segment 1 of syllable 4. I am not concerned at this stage with an explanation of why and how this occurs, but with the fact that it can occur at this stage with the syllabic ordering remaining intact. It is also here that whole syllables, or parts of syllables may get transposed or misplaced. These errors must occur before Stage 5, which is where the Morphophonemic Rules or constraints take over.

Stage 5) The morphophonemic constraints of the language at this stage change, if necessary, or perhaps, 'spell out' the phonological shapes of morphemes. The segmental errors must occur before this stage to account for the alternations of a/an and s/z of the plural.

We have now reached the stage where automatic phonetic and phonological 'rules' take over, converting the sequences of segments into actual neuro-motor commands to the muscles in the articulation of the utterance.

The above stages may be diagrammed as shown in Figure 1. Again it must be emphasized that the various 'black-boxes' are highly schematic, and what actually occurs in them is outside of the concern of this paper. Rather the attempt is to show a possible ordering of events in the production of an utterance which can account for non-deviant utterances as well as utterances containing errors in speech.

References

- Bawden, H.H. (1900), 'A study of lapses' *Psychol. Rev. Monograph Suppl.* III. No. 4.
- Boomer, D.S. and Laver, J.D.M., (1968), 'Slips of the tongue' *Brit. Journ. of Disorders of Communication*, Vol. 3, No. 1, 1-12.
- Broadbent, D.E. (1966), 'The well ordered mind' *Amer. Educ. Research Journ.* 3.3, 281-295.

- Brown, R. and McNeil D., (1966), 'The tip of the tongue phenomenon' *Journ. Verb. Learn. Behav.* 5. 325-337.
- Carroll, L. (no date given) 'The hunting of the snark' in *The Complete Works of Lewis Carroll*, The Modern Library, N.Y.
- Chomsky, N. and Halle, M. (1968), *The Sound Pattern of English*, Harper and Row, N.Y.
- Cohen, A. (1966), 'Errors of Speech and their Implication for Understanding the Strategy of Language Users', *IPO Annual Prog. Re. I*, and in *Models of Speech, Intern. Cong. of Psychol.* Moscow.
- Freud, S. (1938), 'Psychopathology of Everyday Life' in *The Basic Writings of Sigmund Freud*, ed. A.A. Brill, Modern Library, N.Y.
- Fromkin, V.A. (1968), 'Speculations on performance models', *Journ. of Ling.* 4, 47-68.
- Green, E. (1969), 'Phonological and grammatical aspects of jargon in an aphasic patient: a case study', *Lang. and Speech* 12.2, 80-103.
- Hockett, C.F. (1967), 'Where the tongue slips, there slip I.' In *To Honour Roman Jakobson*, Vol. II, 910-936, *Janua Linguarum*, Series Major XXXII, The Hague: Mouton.
- Jakobson, R. (1964), 'Towards a Linguistic Typology of Aphasic Impairments' *Disorders of Language*, CIBA Foundation Symposium, London.
- Jakobson, R. (1966), 'Linguistic Types of Aphasia' *Brain Function*, ed. by E.C. Carterette, U. Calif. Press.
- Jespersen, J.O.H. (1922), *Language: Its Nature, Development, and Origin* (London: Allen and Unwin, 1950).
- Laver, J. (1969), 'The detection and correction of slips of the tongue' *Work in Progress No. 3*, Dept. of Phonetics and Linguistics, Edinburgh University, 1-13.
- Luria, A.R. and Vinogradova, O.S. (1959), 'An objective investigation of the dynamics of semantic systems.' *The Brit. Journ. of Psych.* Vol. 50, 89-105.
- Lashley, K.S. (1951), 'The problem of serial order in behavior' *Cerebral Mechanisms in Behavior*, ed. L.A. Jeffress, John Wiley and Sons, N.Y. 112-36.

- MacKay, D.G. (1970), 'Spoonerisms: the anatomy of errors in the serial order of speech.' *Neuropsychologia*.
- MacKay, D.G. 'Sound change and errors in speech.' to appear in *Historical Linguistics*.
- MacKay, D.G. (1969), 'Forward and backward masking in motor systems' *Kybernetik*, 6, 57-64.
- Marshall, J.C. and Newcombe, F. (1966), 'Syntactic and semantic errors in paralexia,' *Neuropsychologia*, Vol. 4, 169-176.
- Meringer, R. (1908), *Aus dem Leben der Sprache*, Berlin.
- Meringer, R. and Mayer, K. (1895), *Versprechen und Verlesen, eine psychologisch linguistische studie*, Vienna.
- Miller, G.A. (1962), 'Decision units in the perception of speech' *IRE Trans. on Information Theory*, IT-8, 81-83.
- Miller, G.A. and Nicely, P.E. (1955), 'An analysis of perceptual confusions among some English consonants.' *Journ. Acoust. Soc. Amer.* 338-352.
- Nooteboom, S.G. (1968), 'The tongue slips into patterns' *Nomen, Linguistic and phonetic studies*, The Hague, Mouton. (the references in the text are to a prepublication copy issued by the IPO).
- Nooteboom, S.G. (1967), 'Some regularities in phonemic speech errors' *IPO Annual Prog. Re. II*. (Instituut Voor Perceptie Onderzoek).
- Robbins, R.H. (1966-67), 'The Warden's Wordplay: Toward a re-definition of the spoonerism' *Dalhousie Review*, 46, 457-465.
- Sturtevant, E.H. (1947), *An Introduction to Linguistic Science*, Yale Univ. Press.
- Sturtevant, E.H. (1917), *Linguistic Change*, Univ. of Chicago Press, 1961.
- Wells, R. (1951), 'Predicting slips of the tongue' *Yale Scientific Magazine*, Dec. 9-12.

Wickelgren, W.A. (1966), 'Distinctive features and errors in short-term memory for English consonants.' *Journ. Acoust. Soc. Amer.* 39, 388-398.

Wickelgren, W.A. (1965), 'Acoustic similarity and intrusion errors in short-term memory.' *Journ. Exper. Psych.* 70, 102-108.

Wickelgren, W.A. (1965), 'Distinctive features and errors in short-term memory for English vowels.' *Journ. Acoust. Soc. Amer.* 38, 583-588.

Footnotes

1. Robbins (1966-67) suggests that the earliest literary example is found in Rabelais, in the following: 'Il n'y a point d'enchantement. Chacun de vous l'a vou. Je y suis mâitre passé. A brum, a brum, je suis prêtre Macé.' The contrived error of transposing the m and p in mâitre passé (past master) creates pêtre Macé (priest Macé), 'a monk whose name was synonymous with simple or foolish.'

In this same article, Robbins cites a 'near-spoonerisms' found in the Lives of celebrities by John Aubrey (1626-1697), who discussing a flirtation between Sir Walter Raleigh and a young girl, has the wench's protest 'Sweet Sir Walter' changed into 'Swisser Swatter'. (pp 457-458)

2. Most of the examples cited in the text will be from my own data. Those from other sources will be marked with an asterisk, with the source given. In the citation of examples, the '+' is to be interpreted as 'spoken as'; the single underlined unit -- segment, syllable etc. shows the sounds involved, and double underlining illustrates the error itself. The pronunciation of the utterance will be given in phonetic symbols, and within square brackets, only when the orthography may create ambiguous interpretation or obscure the actual speech errors. A series of three dots (...) represent a pause by the speaker; a series of hyphens represent the fact that no errors occurred in the intended words.

Thus,

'John dropped his cup of coffee' → --- cuff of coffee

is to be interpreted as an error in which the intended utterance was articulated as 'John dropped his cuff of coffee i.e., [kʌfəvɒfij]

3. Lewis Carroll, in his preface to The Hunting of the Snark discusses specifically his 'portmanteau' words:

'...let me take this opportunity of answering a question that has often been asked me, how to pronounce 'slithy toves.' The 'i' in 'slithy' is long, as in 'writhe'; and 'toves' is pronounced so as to rhyme with 'groves'. Again, the first 'o' in 'borogoves' is pronounced like the 'o' in 'borrow'. I have heard people try to give it the sound of the 'o' in 'worry'. Such is Human Perversity.

'...Humpty-Dumpty's theory, of two meanings packed into one word like a portmanteau, seems to me the right explanation for all.

'For instance, take the two words 'fuming' and 'furious'. Make up your mind that you will say both words but leave it unsettled which you will say first. Now open your mouth and speak. If your thoughts incline ever so little towards 'fuming' you will say 'fuming-furious'; if they turn, by even a hair's breadth, towards 'furious,' you will say 'furious-fuming'; but if you have that rarest of gifts, a perfectly balanced mind, you will say 'frumious.' (p. 754)

I have quoted extensively from Lewis Carroll, not only because it is always a delight to read or reread any of his comments, but because in this passage he states that his 'portmanteaus' or 'blends' are possible in natural speech and proposes a hypothesis as to how they occur. As we shall see below, however, these 'complete' blends are seldom found in just this way in 'normal' speech. In other words a blend of [fjurijəs] and [fjumɪŋ] is more apt to occur as [fjumijəs] or [fjurɪŋ], particularly because the first syllables are identical.

4. The speaker who made this error was one for whom [ʃ] followed by a consonant was not an unusual sequence in his dialect. In fact, this might represent a word substitution since [ʃtɪk] as a word exists in his dialect as well as smuck [ʃmʌk], stunk [ʃtʌŋk], shmo [ʃmow] etc.

5. Wickelgren's features were used by MacKay.

6. Chomsky and Halle in SPE are of course concerned with the grammatically determined aspects of the signal. The occurrence or lack of occurrence of speech errors involving phonetic features are in this article being discussed as they relate to a model of linguistic performance rather than competence. However, when one talks about 'the set of phonetic properties that can in principle be controlled in speech' (p. 295) it is difficult to find the clear separation between competence and performance. I am here, however, concerned with the features as performance units.

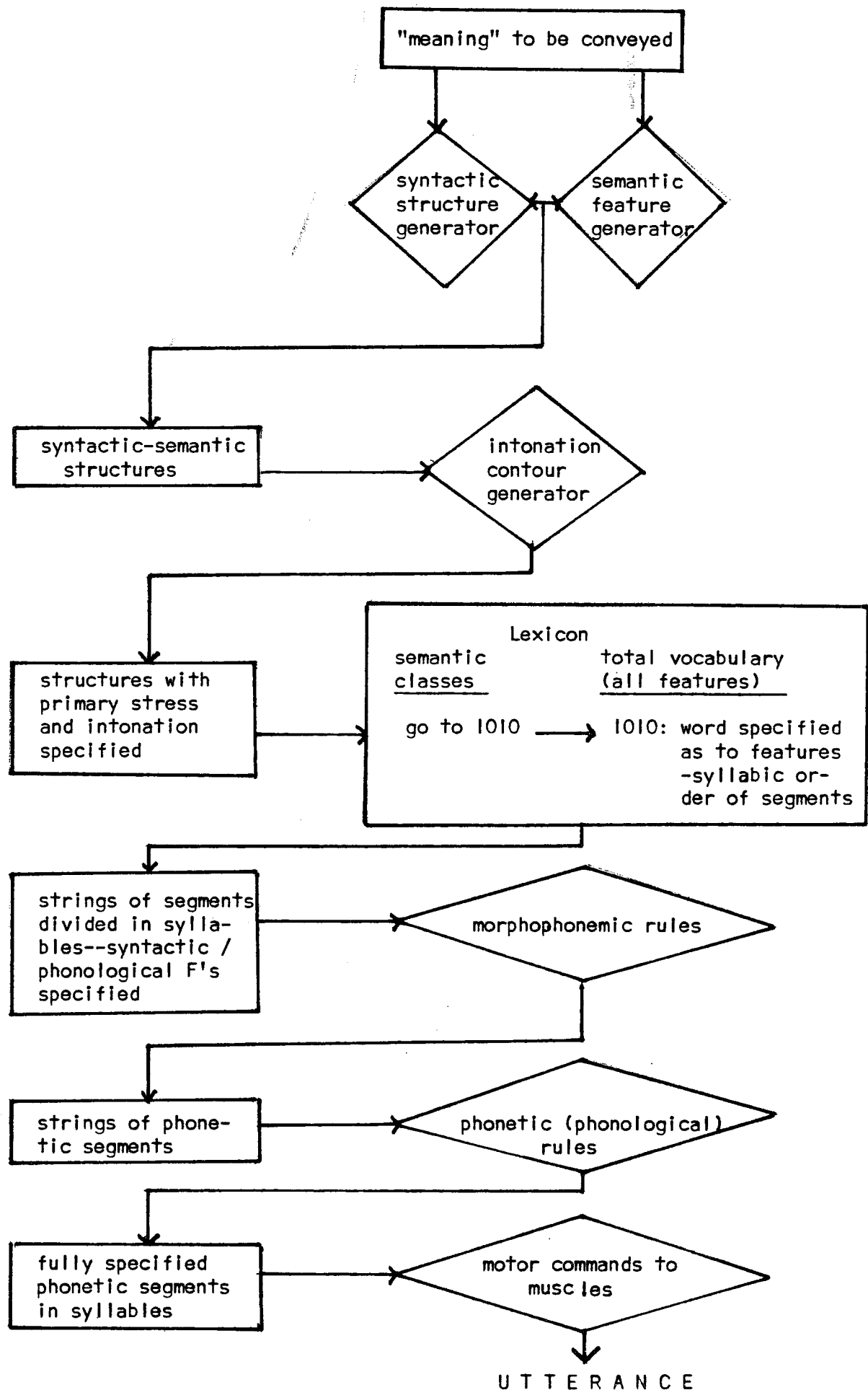


Figure 1: Utterance generator

****Additional Relevant References**

MacKay, D.G. (1970), 'Context dependent stuttering,' *Kybernetik*, pp. 1-9.

----- (1970), 'Spoonerisms of children,' *Neuropsychologia*.

----- (in preparation), 'Hybrids: The problem of synonymity in speech production.'

THE LOCATION OF RHYTHMIC STRESS BEATS IN ENGLISH:
AN EXPERIMENTAL STUDY*

George D. Allen

1. Introduction

Rhythm is the structure of intervals in a succession of events. It has long interested students of language and speech, perhaps because of the importance of poetry and music in our cultural heritage, or because of the compelling force of rhythm and intonation in their emotive, or expressive functions, or because the relationships between rhythm structure and both grammatical and phonological structure in language are not transparent.

Much has been written about the rhythm of language, and yet some of the most basic questions concerning language rhythm remain unanswered. Two of these are: (1) What are the "events" in speech that are the beats of the rhythmic structure we hear? (2) Where exactly do these events occur when we speak? The first question is important in light of the uncertainty as to whether syllables or stresses** are the rhythmic bases of language. Pike (1945), for example, distinguishes two kinds of languages on the basis of their rhythms, "stress-timed" languages (e.g. English) that tend to have equal time intervals between successive stresses, and "syllable-timed" languages (e.g. Spanish) that tend to equalize the time for successive syllables. Languages may not be clearly dichotomized on the basis of this distinction, however, since a given language might be partly stress-timed and partly syllable-timed,

* The research reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education, under the provisions of P.L. 83-531, Cooperative Research; and the provisions of Title VI, P.L. 85-864, as amended. The research was also supported in part by NSF Grants GS-82 and GS-1027, NIH Nr 04595, and ONR Nr N00014-67-A-0111-0005.

** The terms "syllable" and "stress" are notoriously undefinable, and I do not wish to sidetrack discussion by attempting to define them carefully. However, the reader deserves to know what I intend by them: I shall define "syllable" in the lexico-orthographic manner, i.e. syllables make up words, and words are divided into syllables according to rules involving morphology and spelling; I shall follow Daniel Jones by defining "stress" as "degree of prominence." These definitions are particularly appropriate for this paper, since the experimental subjects construed the term "syllable" in this "grammar book" sense and since an experimentally derived "stress score," to be described, must relate to the syllable's prominence.

or partly something else entirely.* The second question, concerning the locations of rhythmic events, is important in two ways. First, we must know where the rhythmic beats of speech are before we can measure precisely the time intervals that define the rhythm. But second, and more important, we want to relate the rhythmic events to speech itself, in hopes of discovering underlying mechanisms of speech rhythm.

In this paper I shall first discuss briefly the evidence for "syllable-beats" and "stress-beats" in English and previous empirical work directed at locating these beats in speech. Then I shall describe a series of experiments that show (1) that the beats of English speech are the stressed syllables and (2) that the beats are strongly associated with the onsets of the nuclear vowels of those syllables.

1.1 Syllables vs. Stresses as Rhythmic Beats

The most obvious place to look for evidence concerning the beats of speech rhythm is poetry, since we can paraphrase Orwell's pigs and say "Poetry is like speech, only more so." Poets use naturally occurring phenomena to their artistic advantage; one of those phenomena is speech rhythm, which they creatively select to fit some metrical scheme.

One metrical unit of English poetry is the syllable. We could take this fact to be an argument in favor of syllable beats in English until we note that another rhythmic unit is the foot, which is defined by the stress pattern of a small number of adjacent syllables. Furthermore, ideal metrical schemes (e.g. iambic, trochaic) involve successions of like feet, and the rhythm of a particular line of poetry obeys a given meter strictly just when the stressed syllables of the line fall on all and only the metric beats. Poetry therefore clearly implicates stresses as the beats of English speech.

We can also note the exclusiveness of syllables and stresses in the literature on poetic meter: no mention is made of any other kind of metric unit, except for punctuation, which functions *between* rather than within rhythmic groups.

As noted above, Pike (1945) classifies languages as stress- and syllable-timed and suggests that English is stress-timed. By this he means that the stressed syllables of English speech tend to occur at roughly equal time intervals. Many researchers (Shen and Peterson, 1962; Bolinger, 1965; Abe, 1967) have sought to test for this "tendency toward

* Williamson (1965) reports "tone-timing" in Ijò. However, she also states that in Ijò stress is predictable and highly correlated with high and changing tone. This "tone-timing" may therefore be essentially stress-timing.

equality" of interstress intervals, and all have been disappointed in not finding it in their data. I have pointed out elsewhere (Allen, 1968a, c) that these earlier studies may have failed in their search because they had no criterion for deciding when there was a "tendency" and when there was not; they could see only strict equality or none. Therefore, although previous studies failed to prove that English is stress-timed, the failure may be blamed as much on inadequate conceptions of what stress-timing can be and improper tests for its existence as on its nonexistence in English.

Stetson (1951) built a theory of speech production around "motor rhythms" of syllables and stresses, drawing parallels between his studies of muscle activity in speech and his earlier studies (1905) on the rhythmic patterns of other motor behaviors. He suggested that each syllable in speech is accompanied by a slight pulse of respiratory muscle activity in the thorax and that stressed syllables received greater action of these muscles along with pulses of activity from the abdominal muscles. It was important to him that these muscle actions be pulse-like, or "ballistic," since, according to his theory of motor rhythm, beats are always ballistic in nature. His concept of speech rhythm emerges as very close to the poetic notion, i.e. syllables grouped into feet, with the addition of specific muscle behaviors accompanying these units. It is hard, however, to guess whether Stetson viewed stresses as more central to speech rhythm than syllables, since his theory involved them both intimately.

Later work on respiratory muscle activity in speech might have changed his views, however. Ladefoged, *et al.* (1958) showed via electro-myographic recordings that Stetson's proposed syllable pulses did not in fact exist (i.e., in the EMG activity from the respiratory muscles they found no "peaks" easily associable with individual syllables), although they found some support for the notion of a stress pulse. Cooker (1963) drew similar conclusions from the time relationships between intraoral pressure peaks and movements of the chest wall during speech. He pointed out that Stetson had failed to allow for a phase lag between pressure buildups in the mouth due to articulatory closure and back-pressure effects on the chest walls. If this phase difference is taken into account, Stetson's records no longer show a pulse of activity just preceding each syllable, but rather a movement of the chest walls due to supraglottal articulations.

Hollister (1937) investigated the movements of subjects' hands as they tapped while reciting poetry from memory. He wished to elucidate the nature of the "syllable impulse" (his term) as a motor-phonetic unit and felt that hand movements would mirror subjective impulse feelings.* He found that when the only instructions to subjects were that they tap while reciting they tended to tap once for each syllable, this fact

* We cannot judge whether Hollister derived these notions from Stetson's work, since there are no bibliographic references in Hollister's paper.

suggesting that the syllable is the rhythmic base for English. However, the hand tapping movements tended to match the syllable character in "duration of hold" and "intensity of hit" (i.e. degree of stress accent), while passive and unconscious hand pressures similarly matched longer phrases of the utterance; speech rhythm emerges once more as a structure involving both syllables and stress.

Classe (1939) replicated an earlier study by Miyake (1902) in an attempt to find the precise location of the beats of the syllables in English. He had subjects tap on a key while they recited isolated lines of verse, the different lines chosen for their overall representation of different phonetic syllable structures. He concluded, consistent with Stetson and Hollister, that the beat of a syllable is "...*mainly* a subjective notion ... depend(ing) more on motor factors than on auditory ones" (emphasis his); but unlike them, he found that the overall task of tapping to all syllables is not easy. Unfortunately for the present discussion, he did not contrast tapping to stresses with tapping to syllables.

Newcomb (1960, 1961) wished to locate syllable *boundaries* by matching clicks to speech and found that "If these pulses were placed at any point other than the last consonant release (in each syllable) they seemed to lose all relationship with the speech ...," this relationship being that "... the pulses seemed to coincide with the rhythm of the syllables." In a later experiment, Newcomb's subjects moved a click around in an utterance, placing it at the location of "optimum sensation of coincidence (with the syllable)." It appears that instead of finding syllable boundaries, as Newcomb desired, his subjects were matching the clicks to the rhythmic beats associated with the syllables. Since Newcomb made no attempt to contrast stressed and unstressed syllables, we can conclude here only that syllables play some role in the rhythm of speech.

The above evidence supports the poetic conception of English speech rhythm, i.e. a succession of "feet," each foot usually comprising one stressed syllable and some number (perhaps zero) of unstressed syllables. If this notion is valid, then whatever the rhythm of an utterance is, the stresses will be the down beats, the other syllables up beats. In the experiments reported below, this difference in rhythmic role between stressed and unstressed syllables was exploited by having subjects tap their fingers in time to the rhythm of *each* of the syllables of an utterance. As Classe's results suggest, since natural tapping behavior involves tapping to down beats, subjects were expected to tap more easily and therefore more reliably to stressed than to unstressed syllables. The high correlation of reliability of tapping with degree of stress obtained in the experiment validates this hypothesis.

1.2 Location of the Beat within the Syllable

I shall not seek to support further the notion that syllables and stresses are the rhythmic beats of English. Instead, I shall discuss

the precise location of the beat within the syllable. There is enough agreement among the findings of the various researchers as to the location of hypothetical syllable beats that we can say these beats really exist; for if they did not, we could expect considerable variation of syllable beat location from one study to the next, depending upon the investigator, the method of investigation, and the subjects used.*

The various studies of speech rhythm cited above all found the beat to be closely associated with the release of the initial consonants and the onset of the nuclear vowel of a syllable. Hollister (1937) found that subjects' hand taps occurred within 2 csec. either side of the onset of the nuclear vowel of "strong" syllables. Classe (1939) concluded that the "point" of stress comes "(a) at the moment of plosion of breathed occlusives; (b) after the explosion of voiced occlusives; (c) at the moment of maximum deviation of the recording-pen for fricatives; (d) just before the beginning of the vowel in the case of all other consonants, except /h/" (*op. cit.*, p. 45). He professed his figures to be in agreement with Miyake's (1902), which also showed the beat to precede the vowel onset, but there is in fact often a difference of up to 0.1 sec. for similar syllables in their work. The sources of these differences cannot be determined from their writings. Newcomb, as I have already mentioned, implicated the release of the last consonant before the nuclear vowel. More precisely, "In the case of voiceless obstruent consonants, the point of demarcation occurs at the release of the consonantal articulation. When semivowels separate syllables, the point ... falls at the beginning of the return from the extreme point of formant deflection toward the position of the following vowel" (1961, p. 3).

The consonant release/vowel onset is a natural candidate for the point of the syllable beat, since this point in the physiologic-acoustic speech sequence is associated with many motor activities and perceptual cues. There are generally large movements of the articulators as the last consonant before the nuclear vowel is released. As noted above, both Ladefoged, *et al.* (1958) and Cooker (1963) suggest that the stressed syllables are often accompanied by a pulse of chest muscle activity. If the initial consonants are voiceless, the onset of voicing requires the adduction and appropriate tensing of the vocal folds and the maintenance of sufficient air flow through the glottis. These movements of articulators, respiratory bellows, and larynx could be the source of many kinesthetic stimuli to associate with the rhythmic beat.

* In an attempt not to prejudge the nature of English speech rhythm, I use the somewhat confusing term "syllable beat." By this I do not mean that the syllable necessarily is a down beat in the rhythm of a particular utterance; rather I mean that, since each syllable probably plays some role in speech rhythm, we can consider it as being a down beat, up beat, or somewhere in between and try to find the location in time of its perceived down or up beat. In a waltz only one of the three beats in each foot is a down beat; nevertheless, the other two beats do occur, and at specific locations in time.

The acoustic wave, too, shows a great increase in energy and striking changes in spectral structure at the point of consonant release/vowel onset, and although these changes occur to different degrees and at different rates for different types of consonantal release, they could be strong perceptual cues in all cases. This point is therefore a clearly defined event for the listener as well as the speaker.

But there remain serious questions as to the exact mechanisms underlying speech rhythm, questions requiring a more precise specification of the location of syllable beat as a function of different syllable structures, different speakers, and the difference between speaking and listening. The experiments reported below do not answer all of these questions, but they do show that (1) stressed syllables are the down beats of spoken English, (2) both the nuclear vowel and the consonants preceding it in a stressed syllable are important determinants of the beat location, and (3) tapping the hand and matching a click give essentially the same answer to the question of syllable beat locations.

2. The Experiments*

As noted above, the validity of the notion "syllable beat" was taken for granted in this study, the primary purpose being to locate the beat within the syllable. It was therefore important that "natural" measures of location be chosen, that is, measures that could be intuitively related to syllable rhythm by theorist and subject alike. The subject is the more important person to cater to, since if he is asked to perform an unnatural task, he will act in a changeable, highly variable, and unmotivated fashion. Such action could not yield the precise measures desired.

Many rhythmically "natural" behaviors are imaginable that can be measured during speaking and/or listening, but finger taps and click matches were chosen for three reasons: first, these were the measures that previous studies had employed, assuring not only that these tasks were natural but also that some of the characteristics of the derived measures were predictable beforehand; second, the results of these experiments could be compared with those of the earlier studies for mutual validation; third, these measures are relatively easy to derive under precisely controlled conditions.

The click matching experiment, as it was first run, proved inconclusive in certain respects, as I shall indicate later, and so a second click matching experiment was carried out. There were therefore three major experiments of two types. The first experiment was the tapping experiment ("E1" hereafter); the second experiment was the first click

* Readers interested in more precise details should refer to Allen (1967).

matching experiment ("E2a" hereafter); the third experiment was the second click matching experiment ("E2b" hereafter).

2.1 Speech Materials

Precise statistical measurements of syllable beat location require repeated measures on the same syllable. It is possible to repeat an utterance many times, but there will be variations from repetition to repetition, adding articulatory variation to beat location variation. In these experiments the experimental tasks were therefore restricted to perceptual beat location, having subjects listen many times to one token of a given utterance.

Conversational speech probably has different rhythmic qualities if not different syllable beat locations from, say, reading or reciting from memory. In order for the experimental measures to be applicable to conversation, therefore, the stimulus utterances were chosen from a high quality recording of a conversation among myself and three subjects* seated in a soundproof recording room. Three utterances from each of the three subjects' speech were chosen from the resulting hour-long tape according to the following constraints:

(1) Each utterance was bounded at both ends by a major rhythmic boundary, i.e., a "long" period of silence or a linguistic pause. This rhythmic completeness was desired so that the utterances would be as rhythmically "complete" as possible and therefore contain as much information as possible about the locations of the rhythmic beats. The locating of rhythmic boundaries was necessarily judgmental, little being known of their acoustic nature.

(2) There was variation in rhythmic structure over the several utterances. The metric structure and the strictness of the rhythm again were subjectively measured.

(3) There was a broad sampling of syllable onset structures among the rhythmically accented syllables.

(4) Each utterance was fairly even in loudness, so that audibility was probably not affected in shifting attention from one part of the utterance to another. This criterion was not always easy to meet, especially in the speech of one of the subjects (TN). Overall speech levels were roughly equalized when tape loops were made for experimental purposes.

(5) Utterances were long enough to be rhythmically interesting,

* The three subjects spoke dialects typical of educated speakers in their home towns: Boston (subject JA), Cincinnati (subject JP), and Ann Arbor (subject TN).

but not so long that repetition via tape loop would be an inefficient use of the subjects' time.

(6) There was a fair range of the various phenomena of experimental interest (e.g. different rhythmic structures, different syllable structures) within the three utterances from a single speaker, allowing comparisons to be made between idiolects.

The nine utterances are shown in Table 1. Spectrograms and minigrams of the utterances are given in Allen (1967, Appendix A).

Speaker	Utterance Number	Utterance (stress score over syllable)
JA	1	10 0 6 0 0 10 0 1 6 0 1 10 Use the weight of the line to get more and more out.
	2	10 1/2 0 0 9 0 0 3 0 10 0 Spinners are particu(lar)ly good in currents.
	3	5 5 1/2 9 5 0 10 0 7 0 9 See when he reared back and fired it by a guy?
JP	4	10 1 0 6 0 10 2 After the play was over ...
	5	1/2 10 0 4 0 0 10 1/2 4 He wants to be a performer now.
	6	10 0 7 0 5 0 0 10 0 1/2 4 5 Talks in terms of getting a name for himself, so ...
TN	7	10 1 4 0 4 3 1 0 10 2 2 I would like to know how they construct these things.
	8	2 10 0 10 1/2 1 10 0 0 9 4 I like predicting; I love to predict things.
	9	9 0 6 0 7 0 0 0 10 4 1/2 0 9 What'll happen now and sort of go out on a limb.

Table 1. Utterances used in rhythm experiments: syllable stress score over the vowel of each syllable.

2.2 Stress Score

In order to determine the degree of stress associated with each syllable, I asked five colleagues of varying linguistic sophistication to transcribe the nine stimulus utterances twice each in sessions a week apart, according to the following instructions:

You will hear nine recorded utterances. Please transcribe these utterances according to whatever phonemic system(s) you know. You may transcribe the segmentals broadly, but be sure to transcribe both segmentals and suprasegmentals. The utterances are recorded on tape loops so that you may hear them as often as you wish. Signal the experimenter when you have finished with each utterance. Are there any questions?

The prosodies received primary attention in these transcriptions.

Four of the five subjects used Pike's (1945) system of three stress levels; the fifth used Trager and Smith's (1951) four level system. However, in agreement with Lieberman's (1965) findings, the subjects were not entirely consistent in their markings of stress, either with each other or with themselves between sessions. As a result, their sets of transcriptions acted as a "filter" for deciding the degree to which each syllable was stressed. The total number of "stress marks" each syllable received* from all subjects in both sessions is written above its vowel in Table 1.

Although syllable stress scores could range from 0 (no stress marks) to 10 (marked as stressed by all five linguists in both sessions), I noticed that with few exceptions what I felt were rhythmically fully accented syllables received stress scores of 5 or higher, syllables that were downbeats but subsidiary to the main utterance rhythm scored from 2 to 6, and unaccented, upbeat syllables scored 1 or less. (The reader unfortunately cannot test this observation for himself without hearing the utterances.) The stress scores thus show a strong link between stress and rhythm, at least in the nine utterances considered here.

2.3 E1 - The Tapping Experiment

2.3.1 Experimental apparatus**

Figure 1 shows in block diagram form the experimental apparatus.

* "Tertiary" or "weak" stresses in the Trager-Smith transcriptions were counted as 1/2 on each occurrence. All other stress marks counted as 1.

** The apparatus in all experiments was calibrated to the extent that at no time did apparatus-derived error constitute more than 1% of total experimental variance (i.e. apparatus error plus subjects' variance).

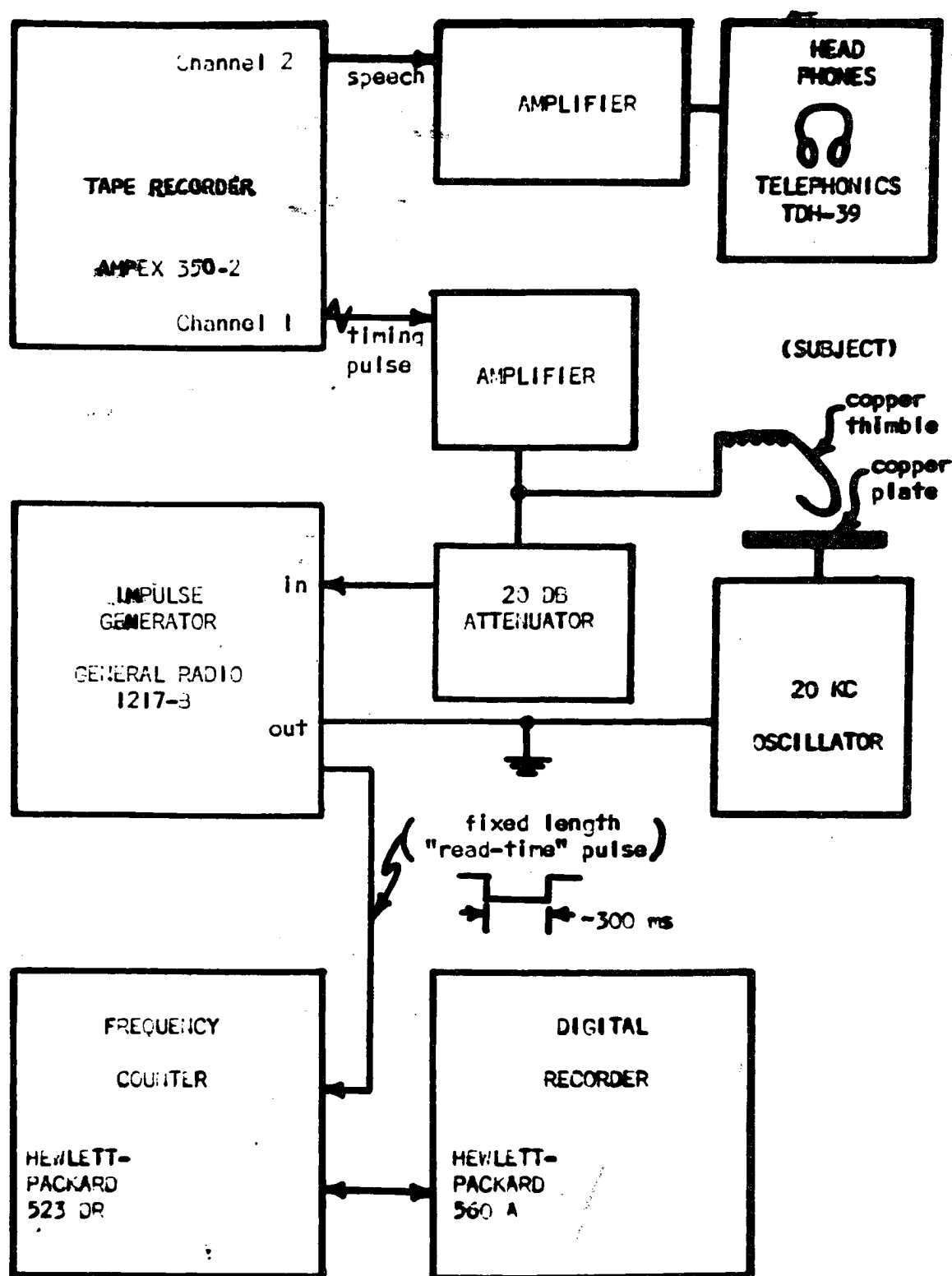


Figure 1. Apparatus for E1.

Each utterance was copied from the original tape onto one track of a two track loop of magnetic instrumentation tape approximately 30 inches long. On the other track of each loop, positioned a few inches before the onset of the utterance, was a reference pulse produced by touching the tape with a magnetized razor blade. Subjects heard the speech from track 2, but not the pulse from track 1. The form of each pulse was a single oscillation whose positive and negative peaks were greater than 1 volt d.c. in amplitude and approximately 1 msec. apart. This pulse was used to reset the frequency timer (hereafter referred to as the "clock") by initiating a fixed length pulse in a pulse generator. The onset of the generated pulse stopped the clock which was then read by a printer. The subsequent offset of the generated pulse restarted the clock. For this reason, this generated pulse will be called the read-time pulse. This scheme was used to stop and start the clock because the time taken by the printer to read and restart the clock was unknown and not easily measurable. The length of the read-time pulse could be calibrated, however, simply by reversing the polarity of the start and stop switches on the clock, starting at the pulse onset and stopping at the offset. On the finger with which subjects tapped they wore a copper thimble which was connected, in series with a 20 kHz* oscillator and a copper plate, across the input terminals of the impulse generator. The copper plate was fixed to the table part of the desk-chair in which the subjects sat, and when they tapped on the copper plate, the 20 kHz signal initiated an impulse in the impulse generator. The time for a single revolution of the tape loop, during which the subject tapped just once,** can therefore be described as the sum of four times: 1) the length of the calibrated read-time pulse initiated by the reference pulse on track 1 of the tape loop; 2) the time between the offset of the read-time pulse and the time of the subject's tap; 3) the length of the read-time pulse initiated by the tap; and 4) the time between the offset of this second read-time pulse and the onset of the read-time pulse initiated by the reference pulse for the next revolution. As far as this investigation is concerned, the time period of interest is the sum of the first two times, i.e., the time between the occurrence of the reference pulse and the subject's tap. These relationships are shown schematically in Figure 2.

2.4 Subjects

Because preliminary experiments had shown rather large inter-subject differences in both location and variability of tapping to

* This value was chosen to avoid a beat tone, audible through the earphones, that occurred near 10 kHz. and 100 kHz.

** Subjects tapped only once in each revolution of the tape loop because the electronic reader required approximately 300 msec. to read and print the time of the tap. Therefore not all the syllables could be tapped to on each revolution, and so only one was.

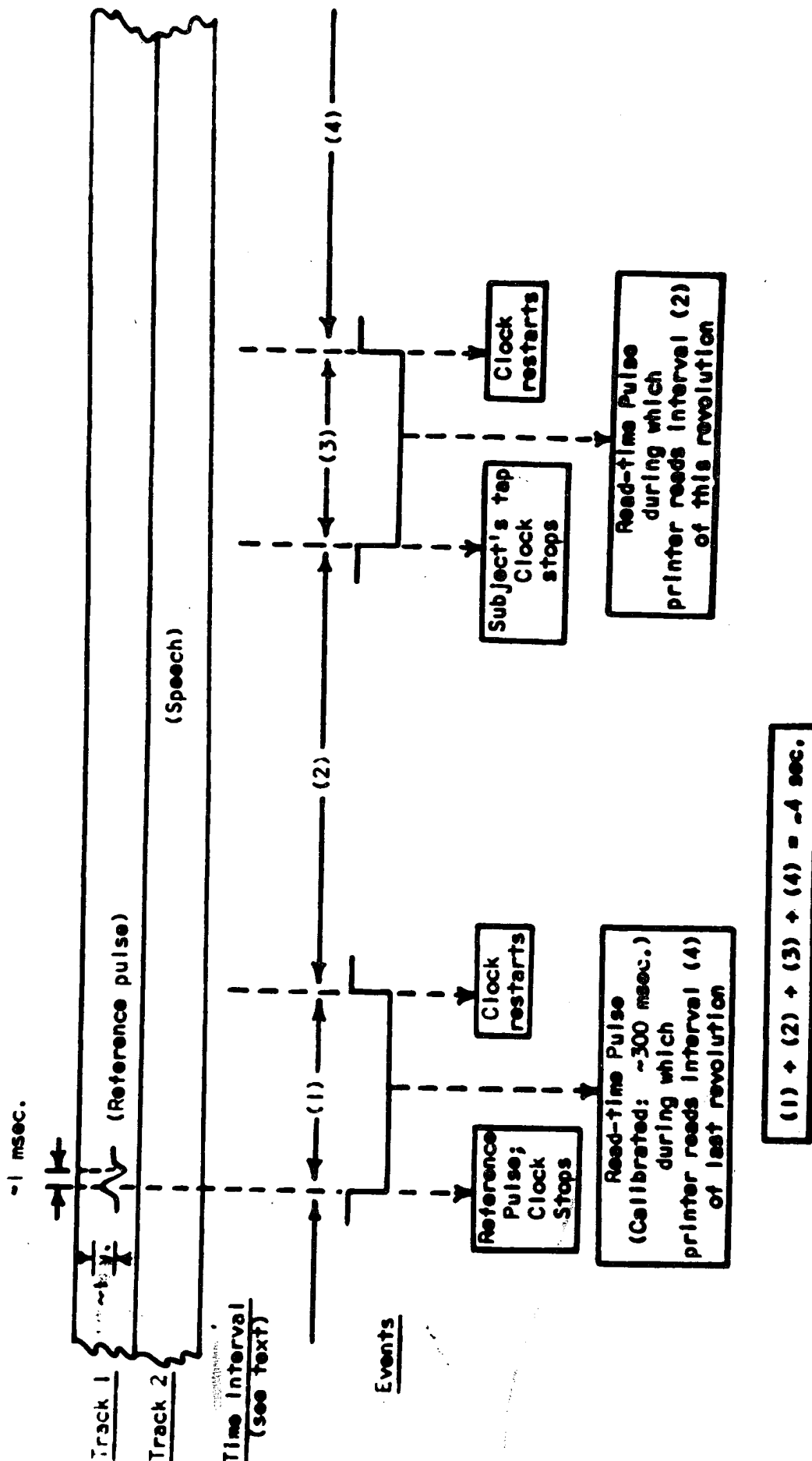


Figure 2. Schematic time diagram for each revolution of a tape loop in EI.

syllable beats, it was better to obtain a large amount of data from a few subjects rather than few data from many subjects. In this way each subject could tell where he felt each of the beats was, and data would not be grouped from subjects known to be different. The three speakers of the stimulus utterances were available for E1, and their use as beat locators allowed the further experimental refinement of having each respond to his own as well as others' speech. They were well acquainted with each other as well as with me, and I felt that the resulting increase in motivation to perform well overrode the possibility of my contaminating the experimental results by influencing their behavior.

2.5 Experimental Design

Subjects sat in a desk-chair in a soundproof room and were asked to tap their finger in time to the beat of a given orthographic syllable, according to the following instructions:

This experiment concerns the rhythm of speech. You will be shown a printed version of an utterance. Some of the syllables are numbered. When you hear the utterance, tap your finger on the beat of the syllable numbered "1" until the speech is turned off. When it resumes, tap to the syllable numbered "2" until it ceases. Continue in this fashion until all of the numbered syllables have been tapped to. At the resumption of the speech each time, listen to the utterance once without tapping. Make your first tap on the second playing of the loop. Are there any questions?

"Syllable," "syllable beat," and the nature of the task were understood intuitively and immediately by all subjects. Once they started tapping to a given syllable, they continued to tap to that same syllable until the utterance stopped playing. During each revolution of the tape loop, therefore, the subject heard the utterance once and tapped once to the syllable in question. From the control room, in which the rest of the apparatus was located, I counted the number of revolutions and regulated the number of taps given by the subject for the syllable by turning off the sound.

Subjects were given a printed version of the utterances with all but the first few syllables numbered in some order. The early syllables in an utterance were not marked for tapping because a tap would be affected by the subject's reaction time and therefore would not mark the syllable beat.* Thus, if there were ten syllables in the printed version of the utterance and subjects were to tap to the last eight of

* The early syllables ("wants" in utterance 5 and "like" in utterance 8) had to be omitted from data analyses because of reaction time components in subjects' taps to them.

these, the numbers one through eight were written over the vowels of the last eight orthographic syllables in a sequence such that various order effects were controlled. The first marked syllables in the nine utterances were "weight," "are," "he," "the," "wants," "terms," "like," "like," and "hap-," respectively. The subject listened to the first playing of the utterance and then tapped once to the syllable marked "1" on each succeeding playing until the utterance was turned off. When the utterance was turned on again, the subject listened to the first playing and then tapped to the syllable marked "2" until the utterance again was turned off. Each experimental session consisted of a subject tapping to all marked syllables of a given utterance.

The order of the syllables was different for different subjects, and within the ordering of the syllables for tapping there was an attempt to alternate "rhythmic" (high stress scores) and "nonrhythmic" (low stress scores) syllables. I felt that a rhythmic beat in the utterance would probably be easier to locate and tap to than a "non-rhythmic" syllable, and that alternation of hard and easy tasks, if indeed they turned out to be so, would keep the subject more interested in the experiment and his tapping behavior more consistent. Furthermore, if the syllables were not different in difficulty, then this alternation could have no effect on the results.

The number of taps per syllable per subject was chosen as 50 or 100 with 50 as the more common. Preliminary data had indicated that a subject's 50 taps to a syllable would give a distribution whose mean was statistically stable enough to give adequate separation from the means of the distributions of 50 taps by the other subjects on the same syllable. However, I also wanted to observe the stability of tapping behavior over longer periods of time and to keep the experimental sessions, each of which involved tapping to all the syllables of a single utterance, reasonably consistent in length. Therefore, more of the syllables of shorter utterances were tapped to 100 times, keeping taps to an *utterance* fairly constant. If the numbers of taps to the syllables of an utterance are added, the result is 550 taps for all utterances except the fourth, which has a total of 500, and the ninth, with a total of 600. Each experimental session lasted approximately 45 minutes, with a rest break about half way through upon the request of the subject.

Each subject tapped to all the marked syllables of a single given utterance in one experimental session, and so there was probably little interaction among utterances. Nevertheless the presentation schedule of the nine utterances was different for the three subjects with no subject hearing any two utterances from the same idiolect in successive sessions.*

* That is, if a subject heard utterance A in one session and utterance B in the next session, this implies that utterances A and B were from different idiolects and that no other subject heard B in the session immediately following the one in which he heard A.

2.6 Results

Subjects' tapping behavior was used to investigate both the degree to which a syllable was a rhythmic beat and the location of the beat in the syllable. The measure of "rhythmicalness" was the reliability with which a subject tapped, i.e., the more of a down beat a syllable is, the easier can a subject tap to it and the more tightly clustered will his taps therefore be; the measure of beat location was the average tap location. In statistical terms, the obvious correlates of "reliability" and "average" are the variance and the mean of the distribution of taps. However, close inspection of the distributions themselves indicated that these statistical measures must be used carefully to preserve validity of interpretation. I shall first discuss tapping variability as a measure of rhythmicalness and then tapping average as a measure of beat location.

2.6.1 Speech rhythm and tapping variability

The discrete autocorrelation function of the 50 or 100 successive taps for all subjects on all syllables was positive for phase lags between one and ten and then slightly negative up to a lag of 25, indicating that the successive tap locations were not independent of each other. The autocorrelation at lag one is .17, indicating that subjects had some tendency to tap near where they tapped the time before, as though their notion of where the beat actually occurred oscillated over long periods of time (ten or more taps) about some central value. Such a positive correlation between successive tap locations tends to inflate the variance of their distribution, and so the so-called "mean square successive difference," symbolized d^2 (von Neuman, *et al.*, 1941), was used instead of the usual variance (s^2) as a measure of variability. In distributions with no serial correlation s^2 and d^2 are equal; if the serial first order (lag one) correlation is positive, d^2 is less than s^2 ; if the lag one correlation is negative, d^2 is greater than s^2 . The values of d^2 for the tapping distributions are highly correlated with s^2 but average about 20% less; correlations between subjects across syllables are also higher using d^2 than s^2 , indicating that d^2 is a more reliable measure of variability than s^2 for these data.

When the variabilities (d^2 values) of the different subjects' distributions of taps to the different syllables were grouped according to stress score, the average variability was seen to increase as stress score decreased, as expected. However, the range of the variabilities also increased with increasing variability; that is, the fully stressed, down beat syllables showed a narrower scatter of d^2 values than did the reduced stress, down beat syllables, whose d^2 values were in turn more narrowly scattered than those of the unstressed, up beat syllables. This correlation between average d^2 and variability of d^2 was removed by using the logarithms of d^2 , instead of the d^2 values themselves, a standard statistical transformation (Scheffé, 1959).

As mentioned above, the $\log d^2$ values were positively correlated

between subjects across syllables (Pearson r approximately 0.5), indicating that their scores could be grouped. Figure 3 shows medians and quartiles of the $\log d^2$ values, for each subject and averaged over the three subjects, for syllables with various stress scores.

Figure 3 is important in that it summarizes the proof that stressed syllables are the beats of English speech. A number of further comments are therefore appropriate in regard to it.

A one-way analysis of variance (ANOVA) showed the main trend of increase in variability with decrease in rhythmicalness to be highly significant, as summarized in Table 2. That this trend relates varia-

Source	Sum Squares	d.f.	Mean Squares	F-ratio	p
Stress score (A)	3.6162	2	1.8081	35.2	<.01
Subjects (B)	4.7264	2	2.3632	46.0	<.01
A X B	0.7393	4	0.0514	3.60	<.01
Within	11.7199	228	0.0514		
Total	20.8018	236			

Marginals:

Stress Score	
0-1	3.153
2-6	2.976
7-10	2.851

Subjects	
JA	2.934
JP	2.854
TN	3.191

Grand mean
3.038

Table 2. ANOVA of $\log d^2$ values in E1.

bility to stress can be seen from the negative correlation of $-.57$ between average $\log d^2$ for a syllable and the stress score of that syllable. That is, the more stressed a syllable is, the more reliably will a subject tap to the beat of that syllable. As I commented previously, the main rhythmic beats of the utterances had stress scores between 5 and 10, the less important beats had stress scores between 2 and 6 and the up beats had scores from 0 to 1. Therefore, the more stressed a syllable is, the more it will tend to be a rhythmic down beat,* and the

* In a theory of speech rhythm I would reverse the relationship, having degree of stress follow from rhythmic role.

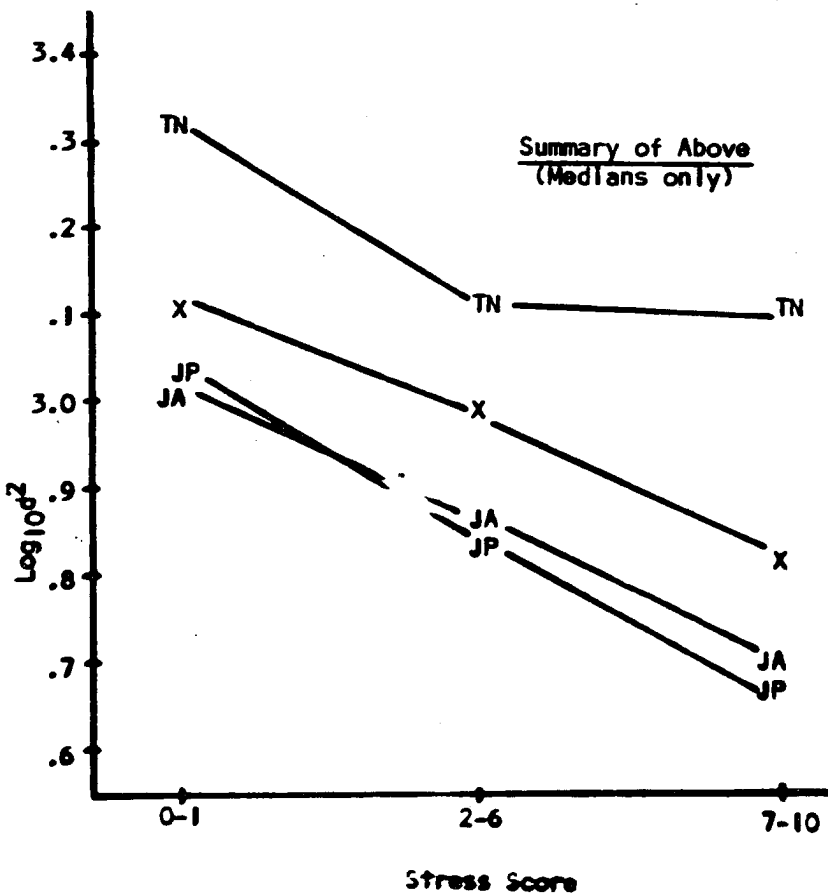
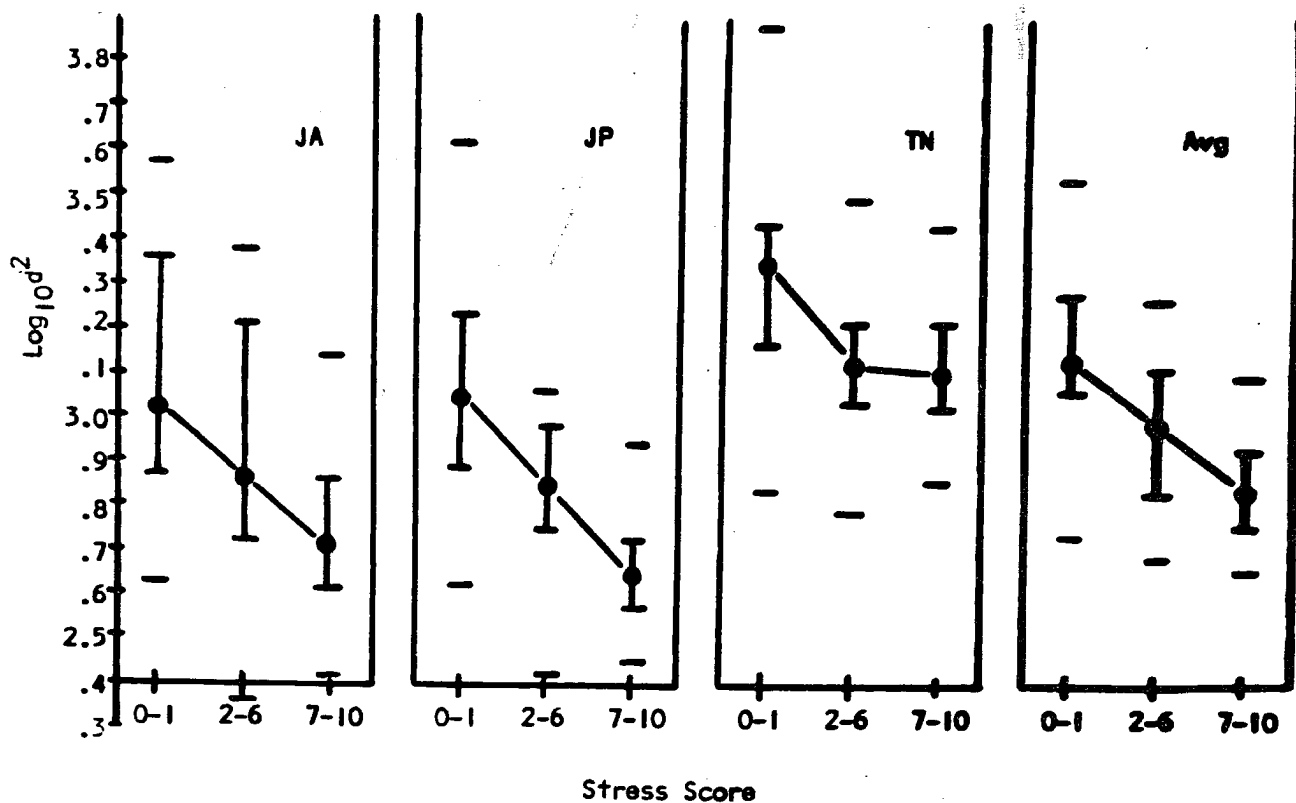
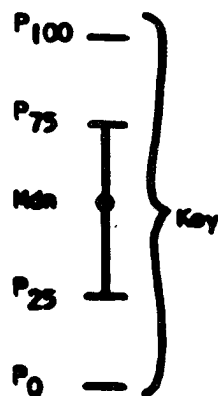


Figure 3. Median ($\bullet$) and zero, 25th, 75th, and 100th percentile points of $\text{log}_{10}d^2$ for tapping data of 3 subjects and the average (X) of these 3.



more reliably a person can tap to that beat with his finger.

These findings must be qualified, however, by the observation that the correlations referred to above between subjects and between average $\log d^2$ and stress score are only moderate in size, ranging between .35 and .6 in absolute value. These moderate correlations leave much of the variance in the tapping task unaccounted for, mostly due to low variabilities associated with unaccented syllables. That is, subjects are able to tap very reliably to some syllables that are but fleeting instants in the stream of speech.

Finally, I should point out a prediction from preliminary data that this experiment did not confirm: initial studies had suggested that subjects could use the beat location of a stressed syllable to tap more reliably to an unstressed syllable immediately following that syllable than to one immediately preceding the stress beat. No difference was in fact observed between the average $\log d^2$ for these two types of unstressed syllables.

Summarizing the results on tapping variability as a measure of the rhythmicalness of a syllable, we can say that the more rhythmically accented a syllable is, the more reliably (i.e. less variably) a subject tapped to it. Although this finding accounts for at most half of the variance of the subjects' taps, it is strong enough when combined with the high correlation between linguistic stress and rhythmic accent to prove that the sample of English used has a stress rhythm.

2.6.2 Beat location and tapping mean

The obvious statistic to use as the point at which the subjects were aiming in tapping is the mean of their distribution of taps. However, this choice is justified only if the tapping distributions are both unimodal and symmetric. In fact, the distributions of taps given by the subjects in this experiment were nearly normal in form, with the exception of those for the syllables "wants" in utterance #5 and "like" in #8. Therefore, the mean of a subject's distribution of taps to a syllable was used as the location of the rhythmic beat of that syllable for that subject. (The average locations in the speech wave of the subjects' taps are given in Appendix A of Allen, (1967).)

But other considerations force us to conclude that the mean tap location was not the point at which the subject was aiming. Inter-subject differences in preliminary testing suggested that different people have different styles of tapping, with associated differences in tap location. These differences may derive from motor-kinesthetic differences, or differences in rhythmic behavior, or both, but they are reasonably consistent within a subject and can therefore be calibrated. The rhythmic tapping biases of the subjects in E1 were calibrated by having each subject tap to the last three of a sequence of four equally spaced sharp clicks, 800 msec. apart, recorded on a tape loop, just as he would tap to the beat of a syllable. The resulting distributions of taps were similar in form to the distributions of taps

to syllables, i.e., almost normal with no skewing and approximately the same variance. The three subjects all tapped earlier on the average than each of the clicks, and the mean amount by which a given subject's taps preceded the clicks corresponded roughly to the mean amount by which that subject's taps preceded the vowel onset of the stressed syllables. The three subjects maintained an ordered relationship of tapping precedence, with subject TN's taps usually earliest and subject JA's usually latest. Table 3 summarizes these findings. Table 4 shows

3a	Subject	Variance (S^2) in tapping (msec. ²)	
		to clicks	to stressed syllables
	JA	826	767
	JP	768	580
	TN	1533	1775

3b	Subject	Mean precession in tapping (msec.)	
		to clicks	to stressed syllables
	JA	15	6
	JP	19	23
	TN	65	37

Table 3. (a) Variance (S^2) in tapping to clicks vs. tapping to stressed syllables in E1.

(b) Mean precession of taps to clicks vs. taps to stressed syllables in E1.

(Stressed syllables are those with stress scores between 7 and 10; see text.)

the amount by which the subject's mean tap locations preceded the vowel onsets of syllables with different initial consonant structures. The number of syllables varies from group to group, and the relative order of subjects wanders somewhat from group to group, but by and large the amount of precession preserves the ordering of subject JA latest, TN earliest.

The ordering of the groups of syllables is not accidental: the shortest initial consonants are in the syllables with no consonants at all and

Initial consonant	N	Subject			mean
		JA	JP	TN	
None	1*	-50	-32	9	-18
Voiced stop	6	4	32	23	19
Voiceless stop	2	22	128	56	60
Semivowel (/w/)	1	0	6	36	14
Sonorant (/l/)	4	-22	22	23	8
Nasal	5	11	34	38	28
Voiced fricative/sibilant	4	7	20	50	26
Voiceless fricative/sibilant	5	8	17	36	20
Cluster	3**	74	67	148	96
mean		16	39	53	36

* The syllable "out" is not included because its precedence values are far out of line with the other syllables.

**The syllable "cur-" was considered to be like a cluster because of the overlong friction following release of the stop. A close transcription would describe the initial consonant as a velar plosive followed by a velo-palatal fricative.

Table 4. Average amounts (in msec.) by which mean tap location for each subject precedes vowel onset in stressed syllables with various initial consonant structures. (N = number of syllables with that initial consonant type.)

the longest are in the clusters. The initial consonants tend to increase in length from the top of Table 4 to the bottom, and a concomitant increase in the degree of precession of tap to vowel onset may be noted. That is, as the length of the initial consonant(s) of a syllable increases, the amount by which a subject's taps precede the onset of a nuclear vowel of that syllable also increases. When the lengths (in msec.) of the initial consonants are correlated with the amount of precession, for the various stressed syllables, values of (Pearson's) $r = 0.43$, 0.04 , and 0.39 result for the three subjects. ($r = 0.40$ is significantly different from zero, $p = .01$; $r = 0.04$ is not significantly different from zero.) Two of the three subjects, then, show a significant positive correlation between tapping precession and initial consonant length. When the three subjects' precessions are averaged, to smooth out intersubject differences, and the resulting mean precession is correlated with initial consonant length, the resulting $r = 0.6$.

2.7 Summary of E1

We can summarize the findings on beat location, then, by saying that (1) subjects generally tap before the onsets of the nuclear vowels of rhythmically accented syllables, by an amount that is a function both of the consonants preceding the vowel and of the subject himself, but (2) the onset of the nuclear vowel is a good first approximation as "the" location of the rhythmic beat.

2.8 E2a - The First Click Matching Experiment

The problems about syllable beat location that remained unresolved in the light of E1 suggested the need for further experimentation. First, intersubject differences in tapping behavior obscured the exact beat location by adding a partially calibratable bias to each subject's taps. This bias is inherently part of the tapping task and can be removed entirely only by using some different beat location task. Second, it would be important to know if the reliability of locating a syllable beat is strongly related to the degree of stress and rhythmic function of that syllable in a task that does not involve muscular activity. We are quite used to "keeping time" with our hands and feet, making the finger tapping task an obvious one for rhythmic beat location; but can we keep time as well in our heads?

For these two reasons, then, the second beat location task involved matching an auditory click to the rhythmic beat of a syllable. The time of occurrence of the click relative to the speech was under control of the subject but involved no rhythmic action on his part; and the judgment of whether the click matched the beat was an auditory perceptual one.

2.9 Speech Materials

E2a used a more time consuming beat matching process than did E1, and so only a subset of the speech material used in E1 was used in E2a. One utterance was chosen from each of the three idiolects, with an attempt to obtain different rhythmic and syllable onset types as before. The utterances chosen were numbers 2, 5, and 7 (cf. Table 1).

For two reasons, a separate tape loop of each entire utterance was made for use in responding to each different syllable of that utterance. In E1 each tape loop was played approximately 2000 times, and some decrease in the signal to noise ratio was noted, a signal degradation that was felt might prove important in the more strictly auditory task of E2a. Through the use of separate tape loops for each syllable the total number of revolutions for a single syllable loop was kept between 221 and 644, with a mean of 554; no signal degradation was noted, by either the subjects or myself.

The second reason for separate tape loops related to the time in-

terval limits of the matching click. In order to give subjects a satisfactory degree of control over click placement, the total click location range was set at approximately 1.5 sec. With a greater range of placement, subjects were not able to control local accuracy adequately; with a smaller range, a relatively large turn of the control knob (described below) produced no perceivable change in location. Since the utterances were greater than 1.5 sec. in length, not all of the syllables in an utterance would fall in the possible range of click locations; with a different loop for each syllable, however, a timing pulse could be placed so that any given syllable fell well within the limits of the 1.5 sec. range.

2.10 Experimental Apparatus

Figure 4 shows in block diagram form the apparatus for E2a. Subjects heard an utterance from track 2 of the tape loop. The timing pulse on track 1, placed by magnetic razor blade as in E1, started a variable length pulse in the circuit labeled "Time Variable Click." The offset of the variable length pulse produced a sharp click (rectangular, 0.2msec. in duration) which was fed through an amplifier both to the subject's earphones (Sharpe HA-10) and to the STOP input of the clock. The length of the variable pulse was adjusted by turning the knob on a box in front of the subject. The subject could thereby move the audible output click around in a syllable until he was satisfied that the click matched the beat of the syllable. The interval between the timing pulse on track 1 and the audible click was recorded by the printer at each revolution of the tape loop, and in this way the successive locations of the click were measured relative to the timing pulse. The timing relationships in E2a are shown schematically in Figure 5.

2.11 Subjects

Two of the three subjects of E1 (JA and TN) served in E2a, the third being unavailable.

2.12 Experimental Design

Subjects sat at a table in a soundproof room and were asked to move the click until they judged that it coincided with the syllable beat, according to the following instructions:

You will be shown a numbered list of the printed versions of some of the utterances you heard in the first experiment, with one syllable underscored in each line. You will hear the utterances in the order of the list. Along with each utterance you will hear a click which you can move forward or backward in the utterance by turning the black knob on the gray box in front of you. Place the click

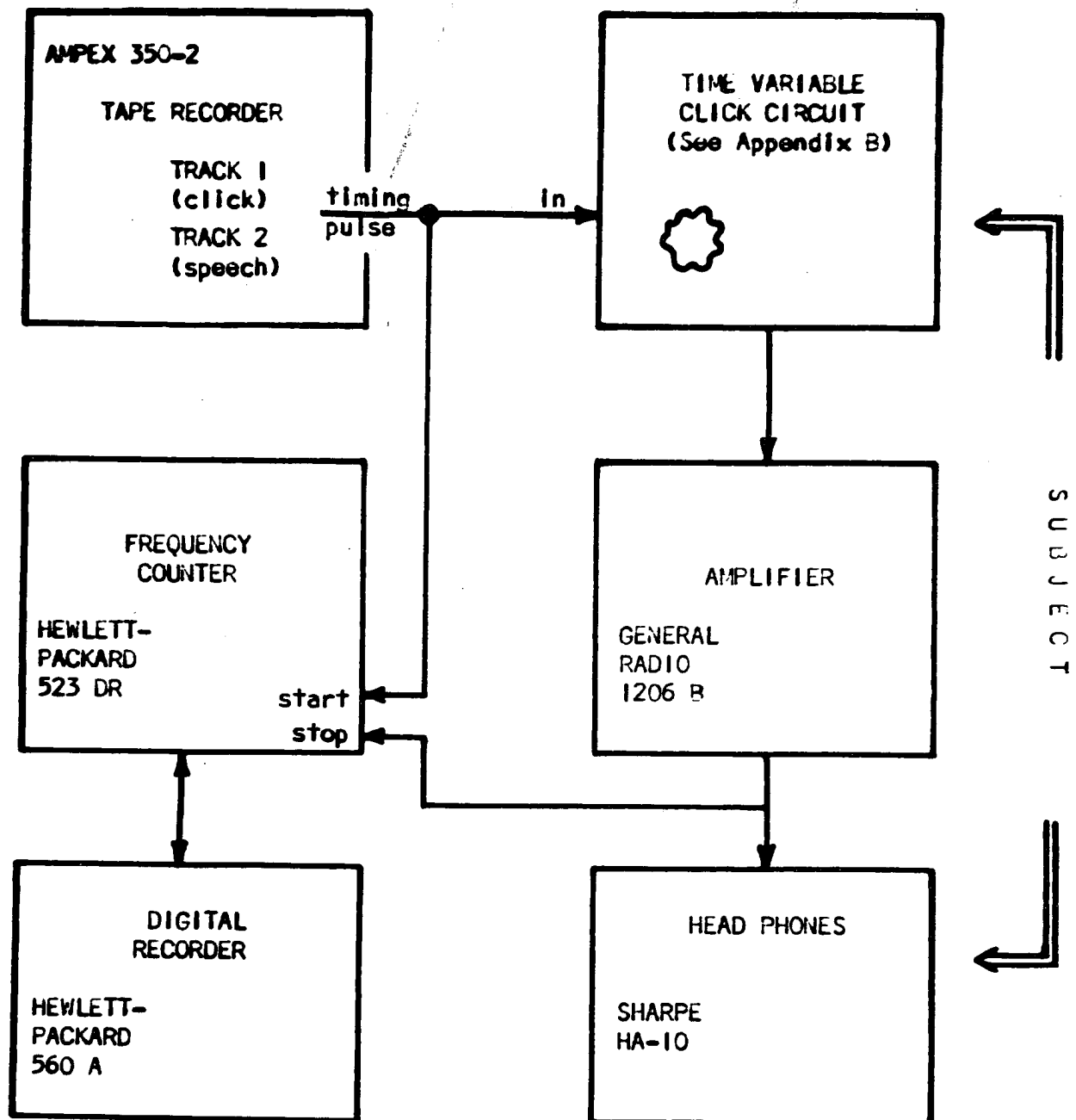


Figure 4. Apparatus for E2a.

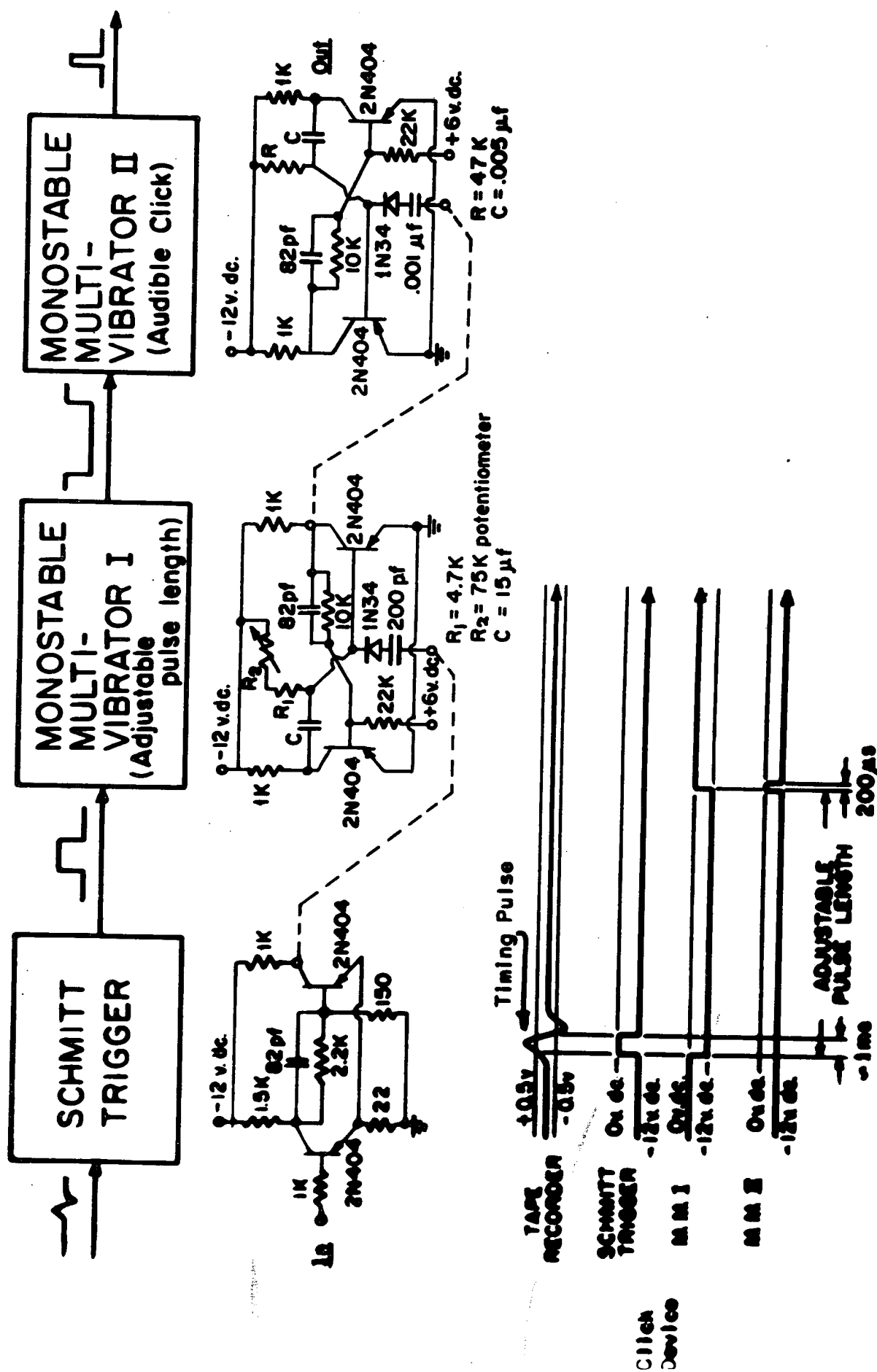


Figure 5. Apparatus and Timing sequence for E2a.

on the beat of the underlined syllable in each utterance in turn through the list. You may hear an utterance as often as you wish until you are satisfied that the click matches the syllable beat. Signal the experimenter when you are satisfied with the placement of the click. In the first few hearings of the utterance move the click so that it both clearly precedes and clearly follows the marked syllable. Are there any questions?

Since this task was auditory in nature, there was no reaction time component to the click placements, and so all syllables of an utterance could be responded to, rather than just the later syllables. E2a therefore derived measurements of the beat locations of the early syllables that could not be measured in E1. Subjects heard an utterance as many times as they wished, moving the click between hearings, until they were satisfied with the click's location. Syllables and utterances were again presented in an order that alternated presumably hard and easy matching tasks to keep subjects interested while controlling for order effects.

Three different conditions of stimulus presentation were used in an attempt to control for perceptual asymmetries of the two ears. Fodor and Bever (1965) noticed in a somewhat similar task small statistical biases in click placement related to differences between right and left ears, namely that clicks were perceived to occur earlier when they were heard in the *left* ear and the speech in the *right* ear than when the ears were reversed. The three conditions were, therefore, both click and speech in both ears (*BOTH*), click in left ear and speech in right (*CL-SR*), and click in right ear, speech in left (*CR-SL*). Note that the perceptual matching process in the *BOTH* condition, in which click and speech form a single signal, may be fundamentally different from that of the *CL-SR* and *CR-SL* conditions, in which the click and the speech come in different ears and must therefore be matched at some higher center.

Five click placements were obtained from each subject for each syllable under each of the three conditions, making a total of fifteen placements per subject per syllable.

2.13 Results

As before, and for the same reasons, I shall first discuss click placing variability as a measure of rhythmicalness and then click placement as a measure of syllable beat location.

2.13.1 Variability of click placement

There was no sequential dependency between successive click placements, and so s^2 (variance) was used to measure variability of placement.

If the condition of stimulus presentation were to have an effect on the variability of click placement, we would predict that placements under the *BOTH* condition would be less variable than placements under either the *CL-SR* or *CR-SL* conditions, because the latter two conditions would add a stage of perceptual matching beyond that present in the *BOTH* condition. Table 5 shows the cumulative ranks of the click placing variabilities under the three conditions. Subject TN shows the *CR-SL*

		Subject JA			Subject TN		
		Condition			Condition		
		BOTH	CL-SR	CR-SL	BOTH	CL-SR	CR-SL
Rank of s^2	Lowest	10	9	11	8	9	13
	Middle	10	10	10	10	11	9
	Highest	10	11	9	12	10	8

Table 5. Number of syllables on which each condition of click presentation had a variability of response (s^2) that ranked lowest, middle or highest among the three conditions, for each subject in E2a.

condition to have the lowest variability, against the above prediction; neither table approaches statistical significance, however ($p > .5$ in both cases). Therefore the variabilities under the three conditions of presentation were averaged to give a single estimate of the variability of click placement for each subject on each syllable.*

The next comparison of variances was between subjects, to see if they were equally variable from syllable to syllable and across syllables. When the two subjects' variances are correlated across the 30 syllables,** a value of (Pearson) $r = 0.08$ results. This low correlation shows that the two subjects were not more and less variable on the same syllables, as they had been in E1. The two subjects' scores are therefore treated separately.

* The variances were averaged, yielding a combined variance with 12 degrees of freedom, rather than computing a single variance from the fifteen pooled observations under the three conditions, with 14 degrees of freedom, because any difference in *mean* click placement due to difference in presentation condition would inflate the variance artificially.

** The syllable "-u-" in "particu(lar)ly" was omitted from this and most other data analyses because of a poorly placed timing click.

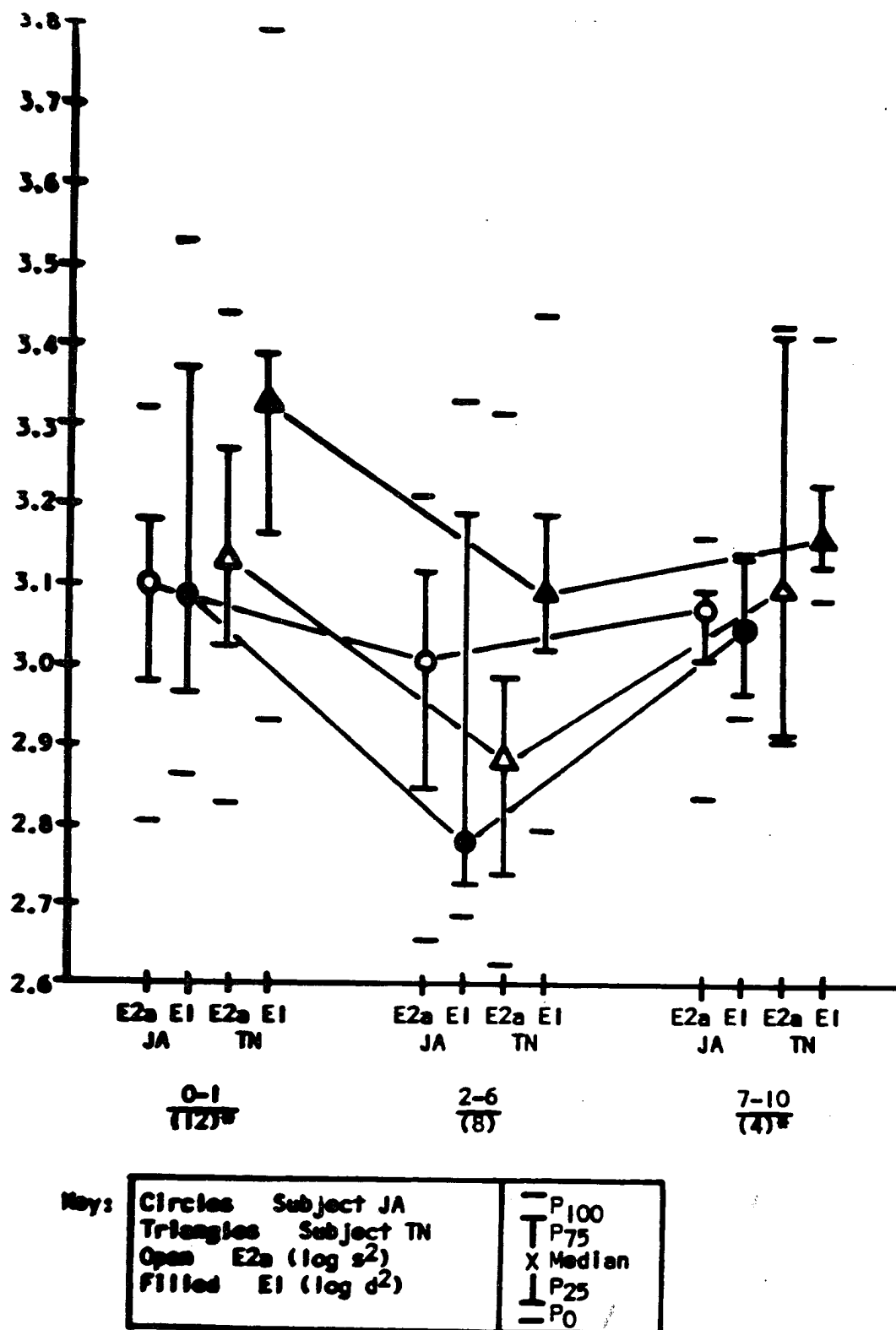
When each subject's variabilities are grouped and averaged according to syllables with the same stress score ranges compared in E1, as shown in Figure 6, we can see the reasons for the above low correlation between subjects' variability of click placement. Subject TN shows a decrease and then increase of click placing variability with increasing stress score, while subject JA's variabilities are virtually unchanged across the stress score range. Furthermore, TN's variabilities on the four syllables with stress scores in the 7-10 range are very widely scattered. The results of E1 are also included on Figure 6 to show that the major difference between variances in the two experiments is the relative variability on the two classes of stressed syllables: the variability of tapping to syllables in E1 decreased *on the average* as stress score increased; the variabilities of the moderately stressed syllables (2-6) used in E2a were less than the variabilities of the strongly stressed syllables, however, in *both* experiments. The two experiments therefore agree, but the picture they present is not so pretty as that of Figure 3.

The variabilities of click placement were smaller, on the average, in E2a than in E1. This is shown in Figure 6 by the fact that the averages for E2a variances always lie below the average variances in E1 for the same subject on syllables in the same stress score range. In E2a subjects heard the utterance several times for each click placement and were satisfied that the click then matched the beat. In E1 they were not consulted as to whether or not a tap had hit the beat, and so the taps were therefore slightly less reliable as a group.

A minor comment that must be made about the click placing variances concerns the form of the distributions from which they were derived. Whereas the distributions of taps in E1 were essentially normal in form, the distributions of click placements were decidedly flatter than normal. Perhaps the reason for this is that when subjects moved the click around in the syllable, they tended to move it by small amounts in a single direction on a given trial; i.e., if a subject started moving the click toward the syllable beat from early in the syllable he would move it later and later by small amounts until he was satisfied with its location; similarly, he would move the click earlier toward the beat by small amounts from late in the syllable. Subjects tended to stop moving the click at the first acceptable location, which therefore was often at one end or the other of the range of acceptable locations. The resulting distribution of placements has too many responses at the extremes of the range and too few in the center, making it flatter than normal in form. This flattening of the response distribution has been referred to by Smith (1957) as a "bias in the method of limits," and was one of the motivations for the different experimental design employed in E2b.

2.13.2 Location of click placement

Whereas there were no differences in variance of click placement under the three different conditions of stimulus presentation, the mean click placement was different for the three conditions, but in different ways for the two subjects (Table 6). Subject JA's average placements under CL-SR were consistently later than her average placements under



* 12 of the 15 syllables in E2a with stress scores between 0 and 1 and 4 of the 7 syllables in E2a with stress scores between 7 and 10 were responded to in EI.

Figure 6. $\log s^2$ and $\log d^2$ for groups of syllables responded to in both E2a and EI, for 2 subjects. Values shown are median and zero, 25th, 75th and 100th percentiles.

CR-SL, with her condition 1 placements lying intermediate. This result

Rank of mean		Subject JA			Subject TN		
		Condition			Condition		
		BOTH	CL-SR	CR-SL	BOTH	CL-SR	CR-SL
		Earliest	7	18	6	9	8
Middle	14	11.5	5.5	14	13	4	
Latest	10	1.5	19.5	8	10	13	
		$\chi^2_{d.f.=4}=28.0, p<.001$			$\chi^2_{d.f.=4}=9.10, p\sim.05$		

Note: The syllable "-u-" was included in this analysis. Hence the rows and columns sum to 31, not 30 as in Table 5.

Table 6: Number of syllables on which each condition of click presentation had a mean response that ranked earliest, middle or latest among the three conditions, for each subject in E2a.

is in agreement with Fodor and Bever's finding that subjects perceptually prepose clicks more under the CL-SR condition than under CR-SL; a subject's average click placement under CL-SR should therefore be *later* than under CR-SL to compensate for this bias. As Table 6 indicates, however, subject TN showed a different and rather uninterpretable bias for the three conditions, namely that CR-SL placements were usually earliest or latest. An ANOVA of the click placements, given in Table 7, shows a nonsignificant main effect for conditions ($p > .5$) but a highly significant interaction of subjects with conditions ($p < .0005$), in support of the above conclusions.

The ANOVA shows the subjects not to be statistically different in mean click placement ($p \sim .1$). From the marginals in Table 7 we can see that the average difference between the two subjects in E2a was 4 msec., far smaller than the 170 msec. difference for the same syllables in E1. Nevertheless, when the differences are considered syllable by syllable,* an interesting pattern emerges. On those syllables where the difference between JA and TN's mean tap location was greater than average (i.e. JA tapped much later than TN), their mean click location difference was also greater than average (i.e. positive); likewise, when their tap location difference was less than average (i.e. small, or even negative),

* The syllable "wants" has again been excluded from this comparison because of reaction time components to its associated taps; with the further exclusion of "-u-", there are therefore 24 syllables compared between E1 and E2a.

Source	Sum Squares	d.f.	Mean Squares	F-ratio	p
Conditions (C)	16600	2	8301	.41	>.5
Syllables (S)	80.05	29	--	--	--
Listeners (L)	3337	1	3337	2.48	~.1
C X S	79190	58	1365	1.01	~.5
C X L	39650	2	19825	14.7	<.0005
S X L	140300	29	4839	3.60	<.0005
C X S X L	85880	58	1481	1.10	>.25
Residual	968000	720	1345		
Total	1333115	899			

Marginals for Conditions by subjects (msec.)

	Subj. JA	Subj. TN	Total
C1 (both ears)	588	590	589
C2 (CL-SR)	603	581	592
C3 (CR-SL)	578	586	582
Total	590	586	588

Table 7: Mean click placement ANOVA.

their click location difference was also less than average (i.e. negative); the correlation over the 24 syllables of difference in mean tap location vs. difference in mean click location is (Pearson's) $r = 0.6$, ($p < .005$). This result suggests that added to the subject-related bias in tapping is a subject-by-syllable interaction bias whereby one or the other subject responds extra early or extra late on certain syllables. Both E1 and E2a show this interaction bias (as evidenced by the high correlation, above), but only E1 shows the tap bias, E2a being purely perceptual.

When these differences in mean response location for E1 and E2a are compared for syllables grouped according to stress score, as in Table 8, we can observe that JA's responses were later than TN's less often and by smaller amounts on unstressed syllables than on stressed; although this effect is not significant for either E1 or E2a alone, the combined effects of the two experiments together are significant ($\chi^2_{d.f.=1} = 7.0$; $p < .01$). One interpretation of this result is that

Stress Score = 0-1			Stress Score = 2-10		
Syl	$\Delta E1$	$\Delta E2a$	Syl	$\Delta E1$	$\Delta E2a$
are	32	-2	tic	44	-6
par	-32	-32	good	55	22
ly	113	34	cur	120	50
in	-7	26	be	5	23
rents	89	29	for	-26	-9
to	-42	-48	now	-14	-16
a	-66	-20	like	56	21
per	3	-37	know	53	-21
mer	53	31	how	43	44
to	125	-9	struct	65	13
they	65	-27	these	137	29
con	45	-3	things	85	17

Table 8. Differences in mean response location (msec.) between subjects JA and TN in E1 and E2a for syllables grouped according to stress score. Positive values mean subject JA's location was later in the speech than TN's.

TN tapped later on unstressed syllables than on stressed ones, or that JA tapped earlier on stressed syllables than on unstressed, or both of the above. Comparison of JA and TN's mean tap locations with those of JP on these syllables reveals the first of the above possibilities to be the more probable, since JA and JP remain in the same relationship on both stressed and unstressed syllables, while TN's taps precede JP's on stressed syllables and succeed them on unstressed ones. Clarification of intersubject differences was another motivation for E2b.

Of greater theoretical interest than intersubject differences in click placement is the relation of the average click locations to the speech wave. Table 9a shows the amount by which the two subjects' average click locations (and tap locations, for comparison) precede the onsets of the nuclear vowels of the stressed syllables in E2a. Table 9b gives the rank order correlations (Spearman's rho) between these columns and shows that, in agreement with the results of E1, initial consonant length is a good predictor of mean click precession for the stressed syllables in these utterances. (Recall that the correlations between initial consonant length and mean tapping precession for *all* the stressed syllables of E1 were somewhat lower: .43 for JA and .39 for TN.) The higher intersubject correlation for the E2a data ($\rho = .53$) as compared with the E1 data ($\rho = .14$) suggests that the E1 data have an intersubject error component not present in the E2a data; the source of this error is not evident.

(a)

Syllable	Length of Initial Consonant	$\bar{T}$ -v.o.		$\bar{C}$ -v.o.	
		JA	TN	JA	TN
tic	40 msec.	36	80	76	70
good	60	6	61	11	33
cur	140	41	161	75	125
be	30	28	33	37	60
for	130	55	29	64	55
now	70	38	24	39	23
like	25	-44	12	-4	17
know	80	41	94	76	55
how	105	18	61	31	75
struct	130	50	115	77	90
these	80	-3	134	39	68
things	40	12	96	44	61

(b)

$$\bar{T}\text{-v.o.} \begin{cases} \rho_{JA, TN} = .14 \\ \rho_{L, JA} = .65 \\ \rho_{L, TN} = .51 \end{cases} \quad \bar{C}\text{-v.o.} \begin{cases} \rho_{JA, TN} = .53 \\ \rho_{L, JA} = .51 \\ \rho_{L, TN} = .60 \end{cases}$$

Table 9. (a) Length (L) of initial consonant and mean tapping and click placing precedence for comparable accented syllables in E1 and E2a (msec.). ($\bar{T}$ -v.o. = mean tapping precedence; $\bar{C}$ -v.o. = mean click placing precedence; v.o. stands for vowel onset.)

(b) Correlations (Spearman ρ) between various pairs of columns. (A correlation of .50 just exceeds the .05 level of significance for $n = 12$, assuming ρ to be an unbiased estimate of the Pearson product-moment correlation.)

2.14 Summary of E2a

The results of E2a may then be summarized as follows:

- 1) E2a agreed in general with E1, in that click placing variability was lower on accented syllables than on unaccented ones, at least for one of the two subjects, and in that the clicks were placed near the onset of the nuclear vowel, but preceding it by an amount correlated positively with the length of the initial consonant(s) of the syllable.

- 2) E2a gives more reliable indications of syllable beat location, both because a given subject's responses are less variable and because there is less difference between subjects in mean click placement.
- 3) One of the two subjects placed the clicks later in the speech when she heard the click in the left ear and the speech in the right than when the sides were reversed, her click placements lying intermediate when both click and speech were heard in both ears. Although this finding agrees with a similar statistical bias noted by Fodor and Bever (1965), the other subject's placements did not show the same effect.

2.15 E2b - The Second Click Matching Experiment

Two methodological matters related to the matching of clicks to syllable beats required clarification. First, the matter of intersubject differences in click location was unresolved by E2a; on the one hand, there was only a nonsignificant trend toward a difference between the two subjects, but this trend was complicated by a significant interaction between subject and syllable, probably the result of subject TN's placing the click earlier than expected on stressed syllables and later on unstressed syllables. E2b was therefore designed to highlight intersubject differences, principally by using a large number of subjects and by having them place the clicks on syllables which would yield low response variances. Second, the response distributions in E2a were flatter than normal, a fact attributed earlier to bias in the method of limits. That explanation does not distinguish, however, between the possibility that the syllable beat is a broad slur, anywhere within which the click may be acceptably placed, and the possibility that the syllable beat is a single point in time, perhaps different for each subject, which the subject can identify in time only approximately. Both possibilities will yield a range of acceptable click locations, but in the first case the range is a function of the speech, and in the second it is a function of the subject. (E1 argues for the second possibility, in that tapping to a sequence of clicks gave results very similar to tapping to syllable beats.) E2b used a task that gave click placements uniformly distributed over the syllable beat, and therefore was not subject to the bias of the method of limits.

2.16 Speech Materials

Utterances 2, 5, and 7 were again chosen for E2b, but subjects matched the clicks with the beats of only five syllables, namely "-tic-" and "cur-" of #1, "-for-" of #2 and "how" and "-struct" of #7. The reduction in E2b of the number of syllables matched was again the result of an increase in the time spent on each syllable over that of E2a.

2.17 Experimental Apparatus

The apparatus for E2b is diagrammed schematically in Figure 7. Utterances were recorded on one track of a two track magnetic tape

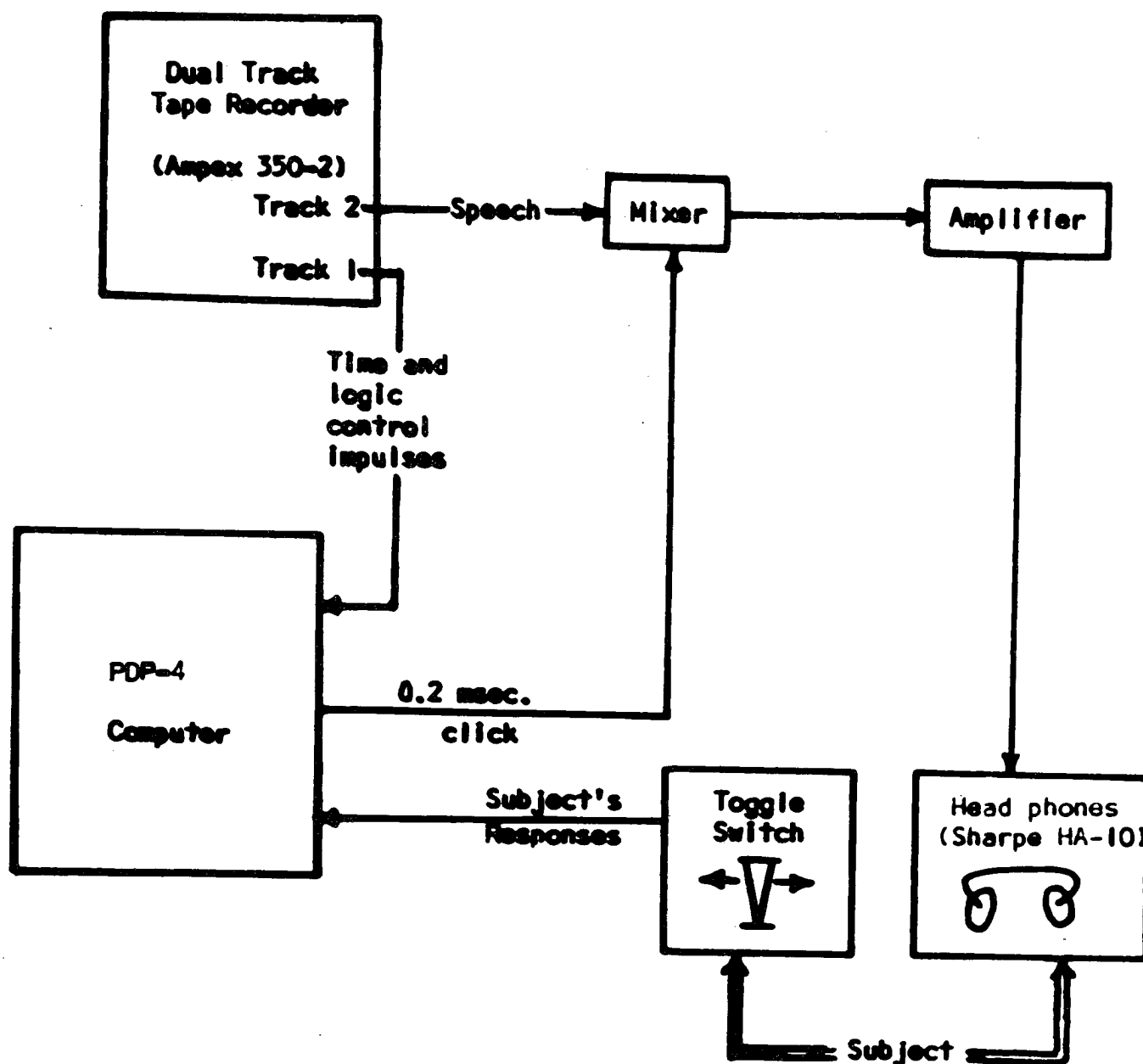


Figure 7. Apparatus for E2b.

(Scotch #111) and played to subjects on an Ampex 350-2 dual track tape recorder at a tape speed of 7 1/2 inches per second. On the other track of the tape were three pulses for each utterance, produced by touching the tape with a magnetized razor blade, as in E1 and E2a; these pulses served to initiate logical operations on a PDP-4 computer. The first and last pulses came before and after an utterance on the tape and were used to start, stop, and rewind the tape under program control. The middle pulse was placed on the tape a short space before the syllable to which subjects were to react, this pulse being used to start a msec. time delay counter within the computer. The computer was programmed to start the tape, wait for the middle pulse and then, upon sensing that pulse, wait a certain number of msec. before sending a click (rectangu-

lar, 0.2 msec. in duration, as in E2a) to the subject. By having stored within it the proper set of waiting times, the computer could send a click to the subject at any time in the utterance from shortly after the middle control pulse to just before the last (stop) control pulse. Waiting times were in fact given to span the beat of the syllable in question. At the occurrence of the audible click, the computer's msec. counter was restarted at zero.

Subjects sat in a soundproof booth in front of a two-way toggle switch. To the left of the switch was written "HIT," to the right, "MISS." Subjects heard both the speech and the superimposed click in both ears through headsets (Sharpe HA-10) and responded according to the following instructions:

This experiment concerns the rhythm of speech. You will hear recorded sentences repeated several times. Each sentence has a rhythm, and some of the syllables are the downbeats of that rhythm, just as some of the notes are the downbeats of a musical rhythm. You will be asked to pay particular attention to these downbeat syllables.

Imagine that you are tapping your finger or your foot in time to the rhythm of the sentence. Each time the sentence is played, you will hear a click superimposed on the speech on or near one of the downbeats (you will be told which syllable the click will be near). If you think the click "hits the beat," as you would hit it with your tap, press the switch in front of you to the *left*, toward "HIT." If you think the click misses the beat, press the switch to the *right*, toward "MISS."

You will be shown a card on which will be typed a sentence with one syllable underlined. You will hear the sentence 40 times in a row (about five minutes), and the click will be on or near the underlined syllable each time. After these 40 trials you will be shown another sentence with another syllable underlined. You will hear that sentence 40 times with the click on or near the underlined syllable, and so on. The entire session will last about 45 minutes.

There is no right or wrong in this experiment. *Your* feelings about whether the click hits the beat are important. If you feel that the click is at all off the beat, answer "MISS."

Your primary task is to tell whether or not you think the click hit the beat, so make sure you press the switch in the proper direction. The time it takes you to respond is also of interest in this experiment, *so press the switch as quickly as you can*, making sure that you press it in the correct direction, of course. If you make a mistake, tell the experimenter immediately by means of the intercom. Then listen for the next trial.

Are there any questions?

Their responses and response latencies (reaction times) relative to the audible click were recorded automatically by the computer and stored for later processing. If the third control pulse was encountered before the subject had responded, the tape was stopped and held until he responded or until 3.3 sec. after the audible click, whichever came first, at which time the tape was rewound and a new cycle begun. The time and logic sequence for a single playing of the utterance, recording of the subject's response, and rewinding of the tape (hereafter referred to as a "cycle") is diagrammed schematically in Figure 8.

2.18 Subjects

Because one of the motivations for E2b was to investigate inter-subject differences, many more subjects were used than in E2a. Sixteen subjects, mostly undergraduates at The University of Michigan, were drawn from the subject pool of the Center for Research on Language and Language Behavior, where the experiments were conducted, and were paid the standard rate for their services. However, as will be pointed out in the *Results* section, not many of these subjects could successfully perform the experimental task, and so that data for E2b *in fact* derive from but *three* subjects, two of which were linguistically naive employees of the Center, the third being myself.

2.19 Experimental Design

Since the aims of E2b were methodological (specifically, to seek intersubject differences while avoiding response biases) rather than theoretical or speech related, only a small sample of the speech stimuli was used. As described earlier, a total of five syllables were chosen from the three utterances, namely "-tic-," "cur-," "-for-," "how," and "-struct."

Preliminary testing indicated that subjects would accept as HITs approximately a 300 msec. range over the syllable beat, and so in order to roughly equalize the numbers of HIT and MISS responses, a 600 msec. range was chosen for the clicks. The center point of this 600 msec. range was chosen to be the average of the click placements for the two subjects of E2a.

The standard error of click placement in E2a had been about 30 msec. for stressed syllables; the 600 msec. range used in E2b therefore represents a distance of plus or minus 5 "E2a standard deviations," a much greater range of errors. The greater inability of subjects in E2b to determine whether or not a click matched the beat probably derives from the fact that the subject did not know where the click was to occur with respect to the beat on a given cycle; his perceptual "window" was open wider in E2b than in E2a, and his accuracy within that window was therefore less.

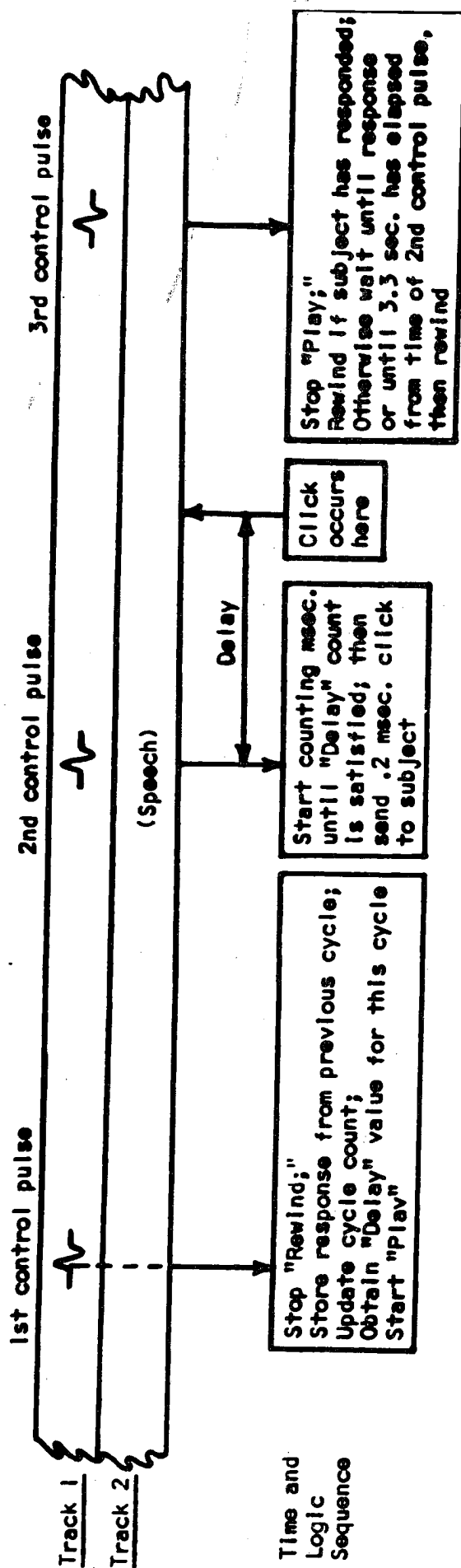


Figure 8. Time and Logic sequence of one response cycle of E2b.

Forty click placements were generated for each of the five syllables. Each click placement occurred on a separate cycle, and since each cycle required approximately 7.5 sec., a subject heard the same utterance for about five minutes. Forty was chosen, therefore, as the greatest number of consecutive responses that a subject could give to one syllable while retaining some interest in the task. (The forty playings of an utterance will hereafter be referred to as one *trial*.)

The forty clicks were located in a random order at 15 msec. intervals across the syllable beat allowing subjects to judge click locations at both the center and extremes of the acceptable range. Each subject responded to the forty click placements on each of the five syllables on three different occasions, at least a day apart, making a total of 600 responses per subject. Each subject responded to all five syllables in a different order on the three different occasions, and no subject heard the syllables in the same order; nor did any two subjects hear the clicks in the same random order on the same syllable. Furthermore, no subject responded to syllables in the same utterance on successive trials; i.e., if a subject heard one of the syllables from utterance #2 on one trial, he would hear one of the syllables from #5 or #7 on the next trial. Interference between syllables was thus minimized, although it will be noted that subjects heard utterances #2 and #7 twice as often as #5, since #2 and #7 had two syllables in them for responding, #5 only one.

Each experimental session lasted approximately 45 minutes and consisted of five trials (one trial of forty click placement cycles for each of the five syllables), plus calibration time between trials. No additional rest period was allowed subjects beyond the calibration time between trials, which lasted about five minutes.

2.20 Results

The most striking (and disconcerting) result of E2b was that most of the subjects could not perform the task asked of them. In contrast with the preliminary subjects, whose responses had formed patterns of HITs surrounded by MISSES, only two of the 16 experimental subjects were able to thus perform. The rest of the subjects either responded with almost all HITs, or else a random pattern of HITs and MISSES across the entire 600 msec. range. Since one of the objects of E2b was to investigate intersubject differences, I therefore grouped for analysis the data from the two capable subjects with data from myself as a subject.

In accord with the analyses of E1 and E2a, I shall first examine the ranges or variabilities of subjects' HIT responses and then the average HIT location. The "range" of hits for a given trial was determined in the following way (cf. Figure 9): first, the "beginning" was found by looking backward in time from the median HIT until the first MISS was encountered, then looking further until the last HIT, and finally taking half the distance between the first MISS and last HIT; the "end" was found analogously by looking forward in time from the median

HIT; finally, the "range of HITs" was then defined as the time duration between the "beginning" and "end" of the pattern of HITs. The ranges of HITs for the various trials are given in Table 10a.

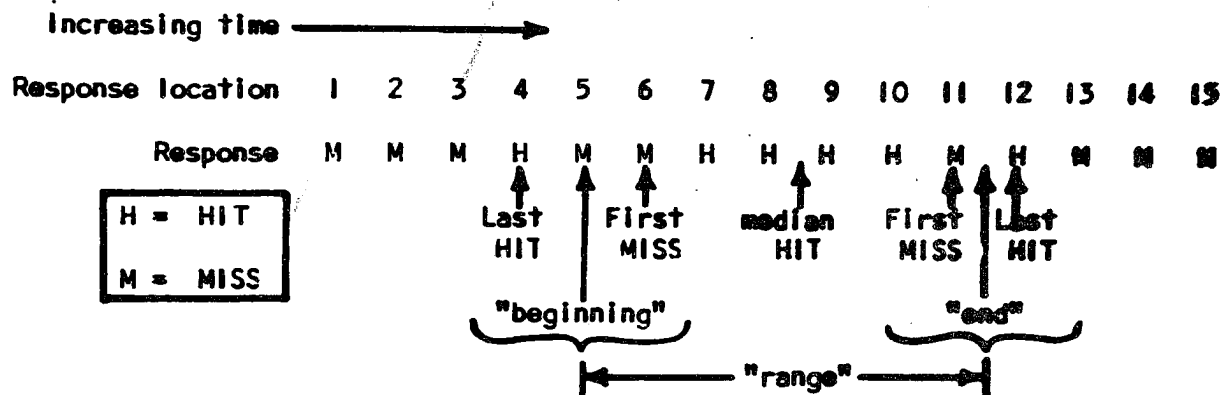


Figure 9. Schematic diagram illustrating the median HIT and "range" of HITs in E2b. (15 responses are used instead of 40, for simplicity.)

The variabilities of the three HIT ranges for each subject on each syllable are reasonably homogeneous (Cochran's C for 15 cells, 2 d.f. per cell = .205; $p > .05$), and so a two-way ANOVA, three observations per cell, was performed. The results of this ANOVA are given in Table 10b. The difference between subjects is highly significant, and from Table 10a we can see that subject AR had considerably wider ranges than GA or SH. We therefore cannot consider that there is a single well defined range of beat locations for a syllable, but that the range of acceptable click matching locations depends upon the subject.

The sensibility of the notion "range of HITs" may be checked by the response latencies. If the syllable beat is perceptually a slur, then subjects should be able to respond more quickly when the click is either in the middle of the slur or well outside it, and less quickly when the click falls near the boundary (Cartwright, 1941), yielding a bimodal distribution of latencies across the 600 msec. range of click locations. Latencies are notoriously variable, and so latencies from different trials had to be combined in some way to increase the reliability of comparison. No fixed location for the boundary of the range of HITs could be assumed, since analysis had shown the subjects to have different ranges of HITs, probably as a result of different perceptual abilities and, perhaps, motivation.* But the "beginning" and "end" of the pattern of HITs had

* Even with my considerable training in the task and obvious motivation, my HIT range was not the smallest of the three subjects'; motivation can therefore not be the only factor at work.

(a)

Syllable	Subject			Syllable mean
	AR	GA	SH	
tic	215	155	180	183
cur	275	208	220	234
for	250	213	195	219
how	247	180	188	205
struct	287	217	170	225
Subject mean	255	195	191	213

(b)

Source	Sum Squares	d.f.	Mean Squares	F-ratio	p
Subjects	201.35	2	100.67	11.0	<.01
Syllables	46.412	4	11.603	1.27	>.25
Sub X Syl	37.322	8	4.665	0.51	>.5
Within	274.33	30	9.144		
Total	559.41	44			

Table 10. Analysis of HIT ranges of 3 subjects on 5 syllables in E2b.

(a) Average HIT range in 3 trials for 3 subjects on 5 syllables (msec.)

(b) ANOVA of HIT ranges in all trials.

been determined, and so the latencies associated with the responses to clicks at the five locations both immediately before and after the "beginning" and the "end" of the range were combined over the 15 trials for each subject. The (logarithmic) averages of these sets of latencies, plus the average latencies at the median HITs, shown graphically in Figure 10.

The variances associated with each of the points in Figure 10 were significantly nonhomogeneous between subjects and between "beginning" and "end" of the range of HITs, and so a three factor ANOVA was not justified for these data. Instead, six one-way ANOVAs were computed for the ten click locations around each boundary for each subject. The results of these ANOVAs are summarized in Table 11c; from the fairly constant size of the F-ratios in Table 11c, the effect of the boundary

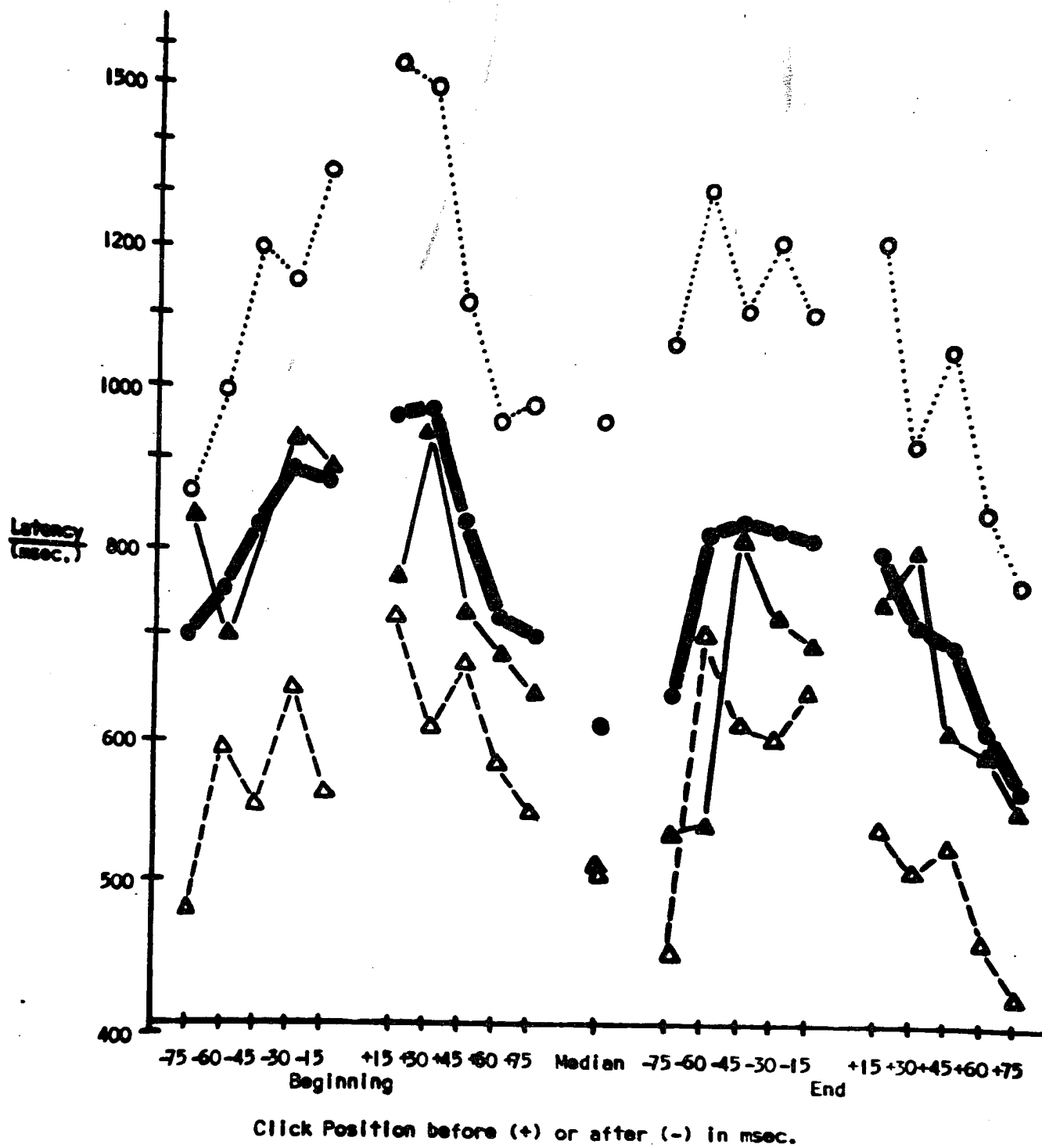


Figure 10. Latencies at click positions before and after "Beginning" and "End" of HIT range, and at Median HIT position, for 3 subjects in E2b. (N = 15 observations per point, in most cases.)

(a)

	AR	GA	SH
beginning	785	593	1157
median	496	496	949
end	652	541	1046

(b)

	AR	GA	SH
Average latency at beginning - average latency at end Range/2	133 123	32 98	111 95

(c)

Subject	Location with respect to range of HITS	Source	Sum Squares	d.f.	Mean Squares	F	p
AR	Beginning	Location Error Total	253.75 2037.25 2291.00	9 139 148	28.20 14.66	1.92	<u>-.05</u>
	End	Location Error Total	314.00 2319.50 2633.50	9 139 148	34.89 16.69	2.09	<u><.05</u>
GA	Beginning	Location Error Total	216.50 1565.00 1781.50	9 140 149	24.06 11.18	2.15	<u><.05</u>
	End	Location Error Total	413.00 2707.50 3120.50	9 139 148	45.89 19.48	2.36	<u><.05</u>
SH	Beginning	Location Error Total	598.75 2904.75 3503.50	9 137 146	66.53 21.20	3.14	<u><.01</u>
	End	Location Error Total	464.00 4219.75 4683.75	9 140 149	51.56 30.14	1.71	<u><.1</u>

Table II. Analysis of latencies in E2b.

- (a) Mean latencies at "beginning," median, and "end" of HIT ranges, for 3 subjects in E2b (msec.).
- (b) Comparison of mean difference in latencies at the beginning and end of HIT range with half the length of the hit range, for 3 subjects in E2b.
- (c) One-way ANOVAs of latencies near the "beginning" and "end" of the HIT ranges for 3 subjects in E2b.

upon decision time is seen to be statistically the same for each subject and position. Figure 10 shows the effect to be a slight increase in latency at the "beginning" and "end" of the HIT range, supporting the notion that these are indeed perceptual boundaries. We can also note from Figure 10 and Table 10a that the latency in response to the median HIT is low in comparison to the latencies near the ends of the HIT range, suggesting that the median HIT is in the middle of the range, not at a boundary.

From the nonhomogeneity of the within cells variances, both between subjects and between "beginning" and "end" of the range of HITs, we can conclude (albeit indirectly, from the fact that the means and variances are correlated) that the mean latencies for different subjects are different and the mean latency at the "beginning" of the HIT range is different from that at the "end" of the range. Inspection of Figure 10 shows the obvious differences between subjects, GA reacting faster than AR, and AR faster than SH. The difference between "beginning" and "end" is not so obvious, but is in the direction of lower latencies at the end of the HIT range. One possible explanation for this relates to the way the latencies were measured in E2b, namely as the time between the occurrence of the click and the subject's response to that click. If the click occurred *before* the syllable beat, the subject's latency would include the time he spent waiting for the syllable beat to occur, in order to match the click to the beat, whereas if the click came *after* the beat, the subject would not have to wait for the beat in order to decide the result. The fact that the mean difference between latencies at the beginning and at the end of the range is approximately half the mean HIT range for each subject, as shown in Table 10a, lends credence to this explanation, assuming that the syllable beat "occurs" at the center of the HIT range.

In summary, then, we can conclude from the properties of the range of HIT responses in E2b that (1) the area within which a click matches a syllable beat differs in width for different subjects, and that (2) the decision as to whether the click matches the beat or not is made quickly both in the center of this area and well outside it but more slowly at the area's boundaries. E2b therefore makes the syllable beat look like a broad slur, whose width depends upon the subject.

Now let us consider the results of E2b as they apply to syllable beat location. The beat location measure appropriate to this task is the median HIT location, that is, the time of occurrence in the speech of the click whose response was the median of all the HIT responses given on a single trial. Figure 9, above, shows schematically the median HIT of a hypothetical set of responses for a trial of 15 cycles.

Each subject had three trials on each of the five syllables. The variances of the three median HIT locations for all subjects on all syllables are reasonably homogeneous (Cochran's C for 15 cells, 2 d.f. per cell = .234; $p > .05$), and so another two-way ANOVA with three observations per cell was performed. The results of this ANOVA are given in Table 12, which shows that there was no difference between subjects in

syllable beat location. (The large effect due to syllables is a function

(a)

Syllable	AR	Subject GA	SH	Syllable mean
tic	815	790	792	799
cur	1665	1650	1678	1664
for	870	857	870	866
how	1100	1070	1103	1091
struct	1572	1595	1555	1574
Subject mean	1205	1193	1200	

(b)

Source	Sum Squares	d.f.	Mean Squares	F-ratio	p
Syllables	5758480	4	--	--	--
Subjects	1104	2	552	1.08	-.5
Syl X Sub	5840	8	730	1.43	>.1
Error	15296	30	509.9		
Total	5780720	44			

Table 12. Analysis of median click placements for 3 subjects on 5 syllables in E26.

(a) Mean of the median click placements in the three trials for each subject on each syllable.

(b) ANOVA of the median click placements.

of location on the tape of the second reference pulse and should be ignored.) Likewise, the interaction between subjects and syllables is not significant in this analysis. These findings of no differences are not likely to be real differences obscured by large error variance, since the within cells variance in this ANOVA is approximately one third the within cells variance in the similar ANOVA of the beat locations in E2a cf. Table 8 above). There were therefore probably no great differences between subjects in their perceptions as to the locations of the syllable beats, either in E2a or in E2b. Click matching can therefore be considered to be a reliable measure of the perceived syllable beat.

Since there were no differences between subjects in syllable beat location in E2b, all the locations for all the subjects on each syllable were averaged to obtain a single measure of the location of the beat of that syllable. These locations are within the range of locations derived from E1 and E2a, and show the same precession effects that the previous locations showed. Table 13 compares the lengths of the initial consonants of the five syllables with the amounts by which the average beat locations precede the onsets of the nuclear vowels of those syllables; the (Spearman) rank order correlation between consonant length

Syllable	Length of Initial Consonant	Average Length of Syllable Beat Range	Precession of Mean Syllable Beat Location
tic	40	183	70
cur	140	221	100
for	130	219	105
how	105	205	40
struct	130	225	90

Table 13. Comparison of initial consonant length, average syllable beat range, and precession of average syllable beat before nuclear vowel onset in 5 syllables on E2b (all values in msec.).

and click precession is 0.72. Also included in Table 13 are the ranges of HITs for the five syllables, averaged over subjects and trials, to show that the acceptable click range is greater for syllables with longer initial consonants. These average HIT ranges correlate 0.50 with initial consonant length and 0.88 with click precedence. This difference between HIT ranges for the five different syllables is in partial conflict with the nonsignificant main effect for syllables in the ANOVA of HIT ranges; even if in Table 10b the subject-by-syllables interaction variance is pooled with within cells variance to obtain a more reliable estimate of error variance, the syllables main effect barely achieves a probability level of .25. We can say, then, that the difference between HIT ranges for different syllables is not as great as the difference between subjects but is known to exist because of its correlation with other known differences, namely initial consonant length and median HIT precedence.

2.21 Summary of E2b

E2b successfully clarified the methodological questions raised by E2a. The three subjects in E2b were not different in their perceptions

as to the location of the five syllable beats studied; and they reacted as though the syllable beat were a broad slur, approximately 200 msec. in duration, anywhere within which an audible click would sound as though it matched the beat. Therefore, when these results are combined with those of E1, we find both similarities and differences in the answers they give concerning the nature and location of syllable beats in English. Both clicks and taps are placed by subjects in the general region of the onset of the nuclear vowels of the stressed syllables but before the vowel onset by an amount correlated with the length of the consonant(s) preceding; and both clicks and taps are more reliably placed on stressed syllables than on unstressed syllables. This means that English has a stress rhythm and that the rhythmic beats are somehow associated with the consonant-to-vowel transition in the stressed syllables. But tapping locations are different for different subjects, and these tapping biases make this behavior less suitable as a beat locating device than click matching, which does not exhibit intersubject differences. On the other hand, clicks are heard to match the syllable beat over quite a broad range, whereas taps seem to be aimed at a single point in time; the broad range of click matching is largely but not simply a result of subjects' perceptual inabilities, however, as the acceptable range of click locations is correlated with the initial consonant length. The nature of the syllable beat is left somewhat unclear by these conflicting results.

3. Discussion

The results of the experiments described in this paper speak to two general problems of rhythm and stress in English. First, they describe the locations of rhythmic stress beats in ongoing speech, and second, they illustrate the nature of stress in English as an element of a performance or speaking grammar as opposed to a competence grammar of English.*

3.1 Locations of Rhythmic Stress Beats

Previous studies on the rhythmic beats in speech implicated the onset of the nuclear vowel (Classe, 1939) or the release of the last consonant before the nuclear vowel (Newcomb, 1961) as the location of the beat associated with a stressed syllable. The present study combines these results, showing that the beat location precedes the vowel onset by an amount positively correlated with the length of the initial consonants. (The actual locations as determined by the experiments may

* A competence grammar of a language deals with what a speaker knows intuitively about how to compose and understand sentences in that language. A performance grammar is concerned with the speaker's knowledge about how to produce and perceive actual utterances in the language (cf. Chomsky, 1964).

be seen in Appendix A of Allen, (1967).) Although initial consonant release and vowel onset appear to be the major determinants of beat location, however, a rule for precise prediction of beat location must also include information concerning the articulatory structure of the initial consonant sequence, the degree to which the syllable is stressed, and tempo. Finally, a more recent experiment (Allen, 1968b) has shown that the quality of the nuclear vowel and the rhythmic structure of the utterance must be taken into account as well.

The locations of the stress beats in speech give us some insight into the processes involved in speech rhythm. In the introductory section of this paper we noted that large amounts of motor activity and spectral change are associated with speech at the time of release of the initial consonants into the nuclear vowel and that Stetson (1951) has characterized this part of the syllable as its "ballistic" phase. The ballistic pulses of stressed syllables therefore are probably the fundamental events around which the speaker organizes the rhythm of his speech. The listener then perceives these rhythmic beats, and thereby decodes the rhythmic structure, either directly from the acoustic cues of amplitude and spectral change or indirectly from kinesthetic-like feelings about the internally generated signals by means of which he perceives the speech. The latter, indirect, process is the more likely, in the light of present day explanations of perception as an active process. We should note that the subjects in the present experiments agreed on beat location only on stressed syllables, suggesting that the speech signal lacks necessary cues for the locations of the beats associated with unstressed syllables, and so these locations are unreliably generated in the listener. This interpretation is supported by Ladefoged, *et al.* (1958), who state that unstressed syllables have no marked muscular activity associated with them beyond that required for articulation; there are no ballistic pulses on unstressed syllables, and so there are no well defined rhythmic beats.

The fact that the present studies deal exclusively with perceived beat locations suggests the need for parallel studies of the locations of rhythmic beats as the speaker *produces* the utterances. We could in this way get closer to the precise rules the speaker uses in generating rhythmic events in speech. One possible experiment would have the speaker tap his finger while he speaks, thus giving us the speech production correlate of the responses in E1, the tapping experiment of the present study. We could tell from this experiment, for example, whether Hollister (1937) was right in suggesting that subjects tend to tap to all syllables, or whether they tend to tap only to stressed syllables. Another experiment, to derive the production correlates of the responses of E2a, the click placing experiment, would have the speaker time his speech rhythm to a sequence of clicks that he hears. A pilot study along these lines (Allen, 1968b) gave results on stress beat locations in good agreement with the experiments of the present study. These two experiments could further be combined by having the speaker tap his finger in time to the utterances he is attempting to speak in time to a sequence of clicks.

There seem to me to be two main uses for the information on beat location derived from the present experiments. Most important, these results give us a certain amount of confidence in the use of the onsets of the nuclear vowels as rhythmic beat locations, and we can therefore proceed to higher order investigations of speech rhythm with the knowledge of how accurately we can pinpoint the beats. This knowledge will allow us to measure changes in the lengths of rhythmic intervals in speech as a function of changes in lexical, phonetic, or prosodic structure within and around those intervals. (For suggestions as to how such changes may be used to test notions about stress-timing, see Allen, 1968c.)

Precise knowledge of rhythmic beat locations will also allow us to investigate certain aspects of syllable structure. For example, Kozhevnikov and Chistovich (1965) give evidence that in describing the syllable structure of Russian the final consonants of one syllable should be grouped with the initial consonants of the next syllable to form one consonant cluster preceding the vowel of the second syllable. I suspect that in English, however, only the syllable's initial consonants are important. If the rhythmic beat of a syllable precedes the vowel onset by an amount that is positively correlated with the length of the initial consonant cluster, then, English and Russian subjects should show different behavior in locating the beats in syllables preceded by syllables with and without final consonants. For example, the Russian speaker or listener should locate the beat in about the same place with respect to the stressed /ɪ/ vowels in /streɪˈtʃu/ (stray tissue) and /streɪˈtʃu/ (straight issue), both of these placements being closer to the vowel onset than in /streɪˈtʃu/ (straight tissue), because of the additional /t/ in the latter sequence. A speaker of English, on the other hand, should locate the beat similarly in "stray tissue" and "straight tissue," both of these locations being farther from the vowel than the location in "straight issue."*

It is possible, however, that the above comparison between Russian and English speakers would result in uninterpretable data, since speakers of different languages may use completely different rules for rhythmic beat location. Yoruba syllables, for example, have a maximum of one consonant followed by exactly one vowel, and so the length of the initial consonant may not be as important a determiner of beat location in Yoruba as it is in English. Likewise, few languages have as wide a range of stress variation as does English, and the rhythmic beats of stressless Japanese, for example, would be predicted on the basis of a rule that ignores stress. The experimental results of the present paper are therefore not generalizable beyond English.

* A properly controlled experiment would require Russian subjects to respond to Russian speech and English subjects to English; the example is illustrative only and not intended as a design for an experiment.

A final comment on beat location concerns the different experimental methods. The present study showed that the different methods of locating rhythmic beats are not all equally useful, and I would suggest that the method of E2a, that of matching a movable click to the perceived rhythmic beats, is the most generally applicable of the three methods investigated. It is relatively free of bias, and what bias it does have appears to be independent of the listener. Furthermore, it is more easily instrumented than the others, requiring only a movable click, the electronics for which are shown in Figure 5, above, and a two track tape loop, onto which both the utterance of interest and a timing pulse can be recorded. The speech plus matching beat can then be examined spectrographically, oscillographically, or by online computer techniques.

3.2 Stress and Rhythm in English

Stress plays a dual role in English. On the one hand, it is part of the competence grammar, carrying grammatical meaning by signaling the different syntactic meanings of sentences made up of the same words in the same order. On the other hand, it plays a basic role in the performance grammar, as a unifying marker around which the rhythm is organized. Because stress plays this dual role, the syntactic patterns of English interact with the rhythmic patterns in real speech, the resulting composite stress patterns often being difficult to describe. Both linguists and experimental phoneticians have often been confused about which aspect of stress they are describing.

Recent linguistic writings (Chomsky and Halle, 1968) have shown that much of the complex syntactic stress pattern of an utterance is predictable at least at the level of words and short phrases. To say that stress is predictable is to say that the speaker or listener knows what the stress pattern is simply by knowing the meaning of the utterance, which he need not hear at all. A difference in meaning between two utterances using the same words in the same order and with the same intonation therefore implies a difference in stress levels somewhere in the two utterances. Speakers of English know the rules for such complex stress pattern differences as part of their knowledge of English.

Speech rhythm is also rule governed and forms part of the performance grammar of English, that is, what a speaker of English knows about how to speak English; unlike the complex patterns of syntactic stress, however, the rhythms of English follow rather simple rules: speakers organize the sequential production of stressed syllables so that most of the rhythmic patterns in normal conversation are successions and alternations.* Rhythms of succession stem from the many monosyllabic content words in English, which will all receive some stress in any sentence in which they appear. For example, the common adjective-noun-verb sequence, if

* For a discussion of the psycholinguistic evidence for this simplistic view of speech rhythm see Allen, (1968d).

all three words are monosyllables, will result in such a rhythm of simple succession (e.g. "The big boy ran."). Alternating rhythms derive from two sources: English has many monosyllabic function words which tend to alternate with content words and which are always unstressed (e.g. "the boy," "to hear," "I ate a piece of cheese."); there is also a strong tendency for polysyllabic words in English to alternate stressed and unstressed syllables (c.f. "Alternating Stress Rule," Chomsky and Halle, 1968, p. 78). I am not arguing that *all* rhythmic patterns in English are successions or alternations, and, in fact, an example that I shall use later shows that accidental parallelism in stress patterns can result in more complex rhythmic structures. Nevertheless, most rhythmic patterns are simple in form.

The pressures of rhythm and syntax on each other keep the two kinds of stress in close alignment, and much of the time it is difficult to separate them. Bolinger (1965) has cataloged an impressive variety of examples both illustrating this agreement and showing the kinds of accentual changes that occur when rhythm and syntax conflict. He shows, for example, how the position of a word in the sentence or the locations of strong stresses in neighboring words can affect the stress pattern of the word: the degree of stress on the last syllable of the word "unknown" changes according to whether it is followed by a stressed syllable, as in "An unknown land," or whether it is at the end of the utterance, as in "quite unknown."* In the first phrase, the syllable "-known" does not receive primary stress; in the second phrase it does. It may be argued that the stress shift is somehow related to the different syntactic environment of "unknown" in the two phrases; however, the syntactically identical phrases "a largely unknown land" and "a quite unknown effect" show the same shift of stress.

The above example shows that the stress pattern of an utterance can change as a function of rhythmic constraints on speech, in this case a constraint to produce alternations between stressed and unstressed syllables. There was, however, no change in meaning as a result of the shift of stress. The following example will show that rhythmic forces can produce a change in the stress pattern of a phrase that results in a meaning that the speaker does not intend.

Among the many readings of the sequence "the man who came to dinner," two possible ones are as the by now idiomatic title of a play and as the answer to the question "Who was he?" In my dialect these two readings are distinguished by the degree of stress on the word "came": with about equal stress on "man" and "came," the phrase answers the question "Who was he?"; with significantly less stress on "came" than on "man," the phrase is the title of the play. If we now embed the phrase in the two environments "I saw ..." and "I went to see ...," the rhythmic structure of the leading words imposes itself on the test

* This example is due to Jones (1956), who also treats rhythmical variations in English, though not to the same depth as Bolinger.

phrase. My reading of "I went to see the man who came to dinner" has less stress on "came" than on "man," probably because "man who came" is rhythmically parallel to "went to see," and the sentence therefore is an answer to the question "What play did you go to?" In the sentence "I saw the man who came to dinner," however, the rhythm is one of simple iambic alternation, and the stress on "came" is about the same as that on "man." This sentence is therefore an answer to the question "Whom did you see?"* The point of this example is not that the same sequence with different stress patterns can mean different things; that much is an obvious fact about English. Nor is the point that I *can* read the phrase "the man who came to dinner" with a fixed stress pattern, independent of context. I mean to suggest rather that my reading of the sentence "I saw the man who came to dinner" will not change, in normal conversational circumstances, depending upon whether I am talking about the man or the play. In normal speech there will be ample context to resolve any ambiguity, and so the rhythm will have its way.

Some linguists have denied the relevance of behavioral measures in linguistics, and at least at some level of linguistic theory this denial is appropriate. The stress scores used in the analysis of the experiments reported in this paper are derived partly from the rhythmic structure of the utterances studied, and so the results of these experiments are therefore probably not directly applicable to such linguistic analyses of stress in English as that described by Chomsky and Halle (1968). However, there are many other levels of linguistic theory to which these results *should* apply. For example, the data from which theories of the development of grammar in the child are derived are behavioral in nature, simply because the linguist does not have as direct access to the child's intuitions about language as he does to his own. A theory of the development of stress patterns in children's language must therefore be constrained by the rhythms of speech production; without this constraint the linguist cannot know what in the child's stress patterns is linguistic and what is psycholinguistic.

The present study shows that lexico-grammatical stress and rhythmic stress are highly correlated in English, adding evidence to the ever increasing documentation of the meaning of stress-timing. A careful refinement and extension of the behavioral measures used in these experiments can perhaps show precisely where syntax and speech rhythm diverge.

* C.f. Bolinger, 1965, for an entire arsenal of examples, and Vanderslice, 1968, for a coherent theoretical synthesis of the interplay of rhythm and syntax in English.

References

- Abe, I. (1967), "English sentence rhythm and synchronism," *Bul. Phon. Soc. Japan* No. 125, pp. 9-11.
- Allen, G.D. (1967), "Two behavioral experiments on the location of the syllable beat in spoken American English," *Stud. Lang. and Lang. Behav.*, 4, pp. 2-179.
- (1968a), "Towards a description of stress-timing in spoken English," *Proceedings of the Conference on Language and Language Behavior*, Ann Arbor, Michigan, E. Zale (ed.), pp. 267-282.
- (1968b), "Experiments on the rhythm of English speech," *Working Papers in Phonetics* No. 10, UCLA, pp. 42-46.
- (1968c), "On testing for certain stress-timing effects," *Working Papers in Phonetics* No. 10, UCLA, pp. 47-59.
- (1968d), "The place of rhythm in a theory of language," *Working Papers in Phonetics* No. 10, UCLA, pp. 60-84.
- Bolinger, D. (1965), *Forms of English*, I. Abe and T. Kanekiyo (eds.), Harvard Univ. Press.
- Cartwright, D. (1941), "Decision-time in relation to the differentiation of the phenomenal field," *Psych. Rev.* 48, pp. 425-42.
- Chomsky, N. (1964), "Current issues in Linguistic theory," in J.A. Fodor and J.J. Katz (eds.) *The Structure of Language*; Englewood Cliffs, N.J.: Prentice-Hall; pp. 50-118.
- Chomsky, N. and M. Halle (1968), *The Sound Pattern of English*, New York: Harper and Row.
- Classe, A. (1939), *The Rhythm of English Prose*, Oxford: Blackwell.
- Cooker, H.S. (1963), *Time Relationships of Chest Wall Movements and Intra-Oral Pressures during Speech*, Ph.D. Dissertation, State University of Iowa, University Microfilms #64-3357.
- Fodor, J.A. and T.G. Bever (1965), "The Psychological Reality of Linguistic Segments," *J. verb Learn. verb Behav.*, 4, pp. 414-420.
- Hollister, R.D.T. (1937), "Relation between hand and voice impulse movements," *Speech Monogr.*, 4 (1), pp. 75-100.
- Jones, D. (1956), *Outline of English Phonetics*, Cambridge: Heffer.
- Ladefoged, P., Draper, M.H., and D. Whitteridge (1958), "Syllables and stress," *Miscell. Phonet.*, 3, pp. 1-14.

- Lieberman, P. (1965), "On the acoustic basis of the perception of intonation by linguists," *Word* 21, pp. 40-54.
- Miyake, I. (1902), "Researches on rhythmic action," *Stud. from the Yale Psychol. Lab.*, 10, pp. 1-48.
- von Neuman, J., R.H. Kent, H.R. Bellinson and B.I. Hart (1941), "The mean square successive difference," *Annals of Math. Stat.* 12, pp. 153-162.
- Newcomb, W.B. (1960), *Some Tempo Manifestations of the Terminals of English*, Ph.D. Dissertation, University of Wisconsin, University Microfilms #60-1013.
- (1961), "The perceptual basis of syllable boundary," unpublished manuscript, General Dynamics Corporation, Rochester, New York.
- Pike, K.L. (1945), *Intonation of American English*, Ann Arbor: Univ. of Michigan.
- Scheffé, H. (1959), *The Analysis of Variance*, New York: Wiley and Sons.
- Shen, Y., and G.G. Peterson (1962), "Isochronism in English," *Stud. in Ling. Occasional Papers* 9.
- Smith, J.E.K. (1957), "The precision of the psychophysical method of limits," *Amer. Psychologist*, 12, p. 469 (abstract).
- Stetson, R.H. (1905), "A motor theory of rhythm and discrete succession," *Psychol. Rev.*, 12, pp. 250-270, 293-350.
- (1951), *Motor Phonetics*, Amsterdam: North-Holland.
- Trager, G.L. and H.L. Smith, Jr. (1951), "An Outline of English Structure," *Studies in Linguistics Occasional Papers*, 3, Norman, Oklahoma: Battensburg Press.
- Vanderslice, R. (1968), "Synthetic elocution," *Working Papers in Phonetics* No. 8, UCLA.
- Williamson, K. (1965), *A grammar of the Kolokuma dialect of Ijo*, New York: Cambridge University Press.

Electromyographic Study of Lip Activity in Swedish

CV:C and CVC: Syllables

Kerstin Hadding, Minoru Hirano, and Timothy Smith

The present study is an extension of a previous EMG experiment concerning Swedish bisyllabic nonsense words (to be published). In the previous study surface electrodes were placed at the vermilion border of the upper lip, some 5 mm below the vermilion border of the lower lip, and on the tip and blade¹ of the tongue. Six front vowels and one central vowel were investigated, the vowels being in the frame [t ____ tən]. There were four long and three short vowels. Since in Swedish short vowels are always followed by either a long consonant or a consonant cluster, the utterances would be either [tV:tən] or [tVt:ən].

The earlier study provided the following data pertinent to the present study: (1) the five rounded vowels [y:, ø:, ʊ:, œ, ʊ]² showed a consistent difference between the EMG activity recorded from the upper lip, sufficiently characteristic to make a distinction between them possible; (2) the electrode placed on the blade of the tongue showed, in most instances, a double peak of activity for the [t:] following a short vowel. This double peak was preliminarily interpreted as a gemination phenomenon; (3) it was not clear which muscle was being recorded by the electrode placed below the vermilion border of the lower lip. In most respects the EMG data from this electrode did not agree with those taken from the upper lip. It was assumed that the lower lip electrode was picking up activity from more than one muscle, presumably mainly from the depressor labii inferioris. It is difficult to interpret EMG recordings made with surface electrodes for this very reason. In the present experiment, an attempt was made to clarify some of these earlier data.

¹ The term "blade" is used in the sense of Abercrombie and Ladefoged (more fronted than "middle" and "back").

² The symbols ʊ and ʊ are taken from the Swedish dialectal alphabet. The vowel [ʊ] can be described with reference to the IPA vowel [ʊ], which is a close, central rounded vowel. The Swedish sound is less close and, at least in the dialect investigated, more fronted. It has a characteristic rounding, similar to that for whistling. The vowel [ʊ] is a central, half-open, less rounded vowel. It is sometimes transcribed with the IPA symbol ɵ.

One subject was used in this pilot study (the same subject as in the earlier experiment). Nine long and eight short vowels were investigated. They were produced in the frame dɪ'd ____ d(:)ɪn. In addition, four long and four short vowels were investigated in the context dɪ'p ____ p(:)ɪn. All the utterances are nonsense words in Swedish, and all were pronounced with the "acute" word tone (accent 1). The frame dɪ'd ____ d(:)ɪn was chosen to provide a neutral context for the lip articulation of the vowels. The frame dɪ'p ____ p(:)ɪn, on the other hand, was chosen in order to test whether the articulation of the second [p] as indicated by the lip electrodes, would show some indication of gemination following short vowels, similar to that suggested by the articulation of [t:] in the earlier study.

The Swedish vowels used in this study are shown below, with key words given in Swedish orthography.

<u>Symbol</u>	<u>Key Word</u>	<u>Symbol</u>	<u>Key Word</u>
i:	rita (to write)	ɪ	ritt (ride)
e:	reta (to tease)	- 3	
ɛ:	räta (to straighten)	ɛ	rätt (right)
ɣ:	ryta (to roar)	ɣ	rytt- (rid- as in "rider")
ø:	röta (to rot)	œ	rött (red)
u:	ruta (square)	ũ	rutt (route)
u:	rota (to poke about)	o	rott (rowed)
o:	Rota (a Norn)	o	rätt (raw)
ɑ:	rata (to refuse)	a	ratt (steering- wheel)

[ɣ:, ɣ, ø:, œ, u:, o, u:, ũ, o:, o] and perhaps [ɑ:] are considered to be articulated with some degree of lip-rounding.

³ Swedish is often said to have 18 vowel phonemes, 9 long and 9 short. However, in many dialects, for instance Central Swedish, there is no distinction between [e] and [ɛ], which are both pronounced [ɛ]. This merging seems to become more and more common also in Southern Swedish. Since the speaker (who speaks a South Swedish dialect) did not make any distinction between the two sounds, [e] was excluded from the vowel list.

Electromyographic data for these utterances were recorded from three points: the upper and lower parts of the orbicularis oris muscle and the depressor labii inferioris. Bipolar thin wire electrodes were used, a technique reported among others by Basmajian and Stecko (1962) but modified for speech research by Hirano et al. This technique involves the use of two very thin coated wires, which have been pulled through a hypodermic needle, the ends being bent backwards over the sharp (beveled) end of the needle. One wire is made slightly shorter than the other in order to avoid short circuiting. When the needle is inserted into the muscle and withdrawn again, the "hooked" electrodes are left behind in the muscle. They remain in place during articulation and phonation but are quite easy to remove by pulling the wire. For detailed description and illustrations of the method see Hirano et al. (1967 a, b, and c, 1969).

In the present study one pair of hooked wires was inserted on the midline of the upper lip, just above the vermilion border. Another pair was similarly placed just below the vermilion border of the lower lip. The third electrode was inserted between the vermilion border of the lower lip and the edge of the chin, between the midline and the corner of the mouth. Various gestures were performed and recorded, as closing, opening, protruding and everting the lips to test the position of the electrodes. The same gestures were performed and recorded at the end of the run as a control that the electrodes were still in the same position.

The electrodes were connected with a preamplifier. The electromyographic activity and the audio signal were recorded on tape simultaneously. The tape recorder had an extra playback head and amplifier to be used for computer processing of data. The EMG and audio data of 20 samples of each utterance were rectified, averaged, and smoothed using the LINC-8 computer as described by Harshman and Ladefoged (1967).

The findings of this preliminary study can be summarized briefly as follows: (1) In terms of onset time, offset time, and location of peaks, the EMG data from the upper and lower parts of the orbicularis oris muscle were in reasonable agreement. One curious but interesting fact, quite consistent throughout the data, is that the EMG from the lower lip always led the EMG from the upper lip by about 40-50 msec (again in terms of onset peak, peak location, and offset time). No explanation for this is offered at this time. No such consistent difference in timing occurs in the production of the non-phonatory gestures. (2) The averaged amplitude of the EMG from these two parts of the orbicularis oris was not the same for different vowels. Figures 1 through 3 show the averaged EMG and audio signals for the utterances [dɪ'dy:dɪn], [dɪ'dw:dɪn] and [dɪ'du:dɪn]. Here one can see that while the signals for the lower lip for each vowel are quite similar, the upper lip EMG data differ considerably, depending upon the type of vowel involved. Thus the different types of rounding were

manifested in a difference in the EMG of the upper lip, but of the upper lip only. (3) In figures 4 and 5 the averaged EMG and audio signals for the utterances [dɪ'pɪ:pɪn] and [dɪ'pɪp:ɪn] are shown. In the case of the former it is clear that there is only a single peak associated with the second [p]. In the latter case there is a high initial peak followed by a prolongation of the activity at a lower amplitude after a small "dip". Although a second peak could be observed in several individual samples, two clearly defined peaks with a marked dip between were not consistently found which could be associated with a gemination of the consonant following short vowels. The picture can probably equally well be explained as a relaxation and following pick-up of activity due to the extra effort necessary for the prolonged articulatory activity for [p:]; this activity is however clearly very different from that for [p].

The electrode assumed to be located in the depressor labii inferioris gave results that were particularly difficult to interpret. Earlier results with similar samples by the same speaker but with surface electrodes, where one electrode was placed ca 5 mm below the vermilion border of the lower lip, were interpreted as showing activity of the orbicularis oris muscle contaminated by activity of the depressor. This time an effort was made to separate the two muscles using two pairs of needle electrodes. The data from the lower electrode might be interpreted as showing the activity of the depressor contaminated by the orbicularis oris. In [dɪ'de:dɪn] only the (assumed) depressor is active (figure 6). In [dɪdɔ:dɪn] (figure 7) activity is recorded from both lower lip (L) and depressor (D) while the upper lip (U) is inactive. This may perhaps indicate that (L) is contaminated by (D) instead of the other way round. Another muscle which may have been involved, is the depressor anguli oris (or triangularis). According to Öhman et al. (1965) the spread vowels are associated with an activity of the depressor anguli oris (p. 7).

In samples with strongly rounded vowels as [dɪ'du:dɪn] and [dɪ'dw:dɪn] (L) seems rather to show similarities to both (U) and (D) (see figures 2 and 3). On the other hand (D) might have been expected to be comparatively relaxed when the lips are in a more closed position (acc. to Hirano). See however the large peaks at (D) on the articulation of the p's of, for instance, [dɪ'pɪ:pɪn], occurring simultaneously with those at (U) (figure 4). (Activity for the [i:] is only recorded at (D), however.) However, Öhman has found (1965 and 1966) that an electrode inserted at a point rather close to the point (D) of the present study, and identified as representing the depressor labii inferioris, shows activity particularly during the articulation of consonants. The production of a bilabial stop is said to involve both the depressor anguli oris and the depressor labii inferioris (as well as the mentalis) (1966, p. 2).

Hirano, who feels confident that the electrodes were indeed placed in the depressor muscle, now reports from Japan that contrary to his

expectation from reading the literature the depressor may be involved in the lip-closing gestures. Thus, some of his patients with facial palsy show depressor activity when closing the lips, although most subjects do not. Later findings indicate that some speakers normally activate the depressor when articulating labial stops; this activity may be seen quite generally in connection with tight closure of the lips and/or extreme protrusion (personal communication). Some of our findings may however be due to an individual way of articulating or electrode (D) may not have been placed in the optimal position. The experiment will be repeated with the same speaker and also with other speakers for control and comparison.

The difference of timing in the activity at (U) and (L) should also be examined further. The orbicularis muscle consists of a great number of interlacing fibers which do not necessarily act simultaneously, as one unit. This too will be investigated in greater detail, together with a more extensive electrode mapping of the facial muscles involved in the articulatory gestures.

References

- Basmajian, J.V., and Stecko, G. (1962), "A new bipolar electrode for electromyography," *J. Appl. Physiol.* 17, p. 849.
- Hadding-Koch, K. (1968), "Lapp-EMG," *Nord. Tidsskr. f. Tale og Stemme* 28, p. 1.
- Harshman, r., and Ladefoged, P. (1967), "The LINC-8 computer in speech research," *Working Papers in Phonetics* 7, UCLA, p. 57.
- Hirano, M., and Ohala, J. (1967a), "Use of hooked wire electrodes for electromyography of the intrinsic laryngeal muscles," *Working Papers in Phonetics* 7, UCLA, p. 35.
- Hirano, M., and Smith, T. (1967b), "Electromyographic study of tongue function in speech: A preliminary report," *Working Papers in Phonetics* 7, UCLA, p. 46.
- Hirano, M., Ohala, J., and Smith, T. (1967c), "Current techniques in obtaining EMG data," *Working Papers in Phonetics* 7, UCLA, p. 20.
- Hirano, M., Vennard, W., and Ohala, J. (1969), "The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation," *Working Papers in Phonetics* 10, UCLA, p. 111.
- Öhman, S., Leanderson, R., and Persson, A. (1965), "Electromyographic studies of facial muscles during speech," *STL - QPSR* 3/1965, p. 1.
- (1966), "EMG studies of facial muscle activity in speech II," *STL - QPSR* 1/1966, p. 1.

Figure 1 dɪdɪdɪn

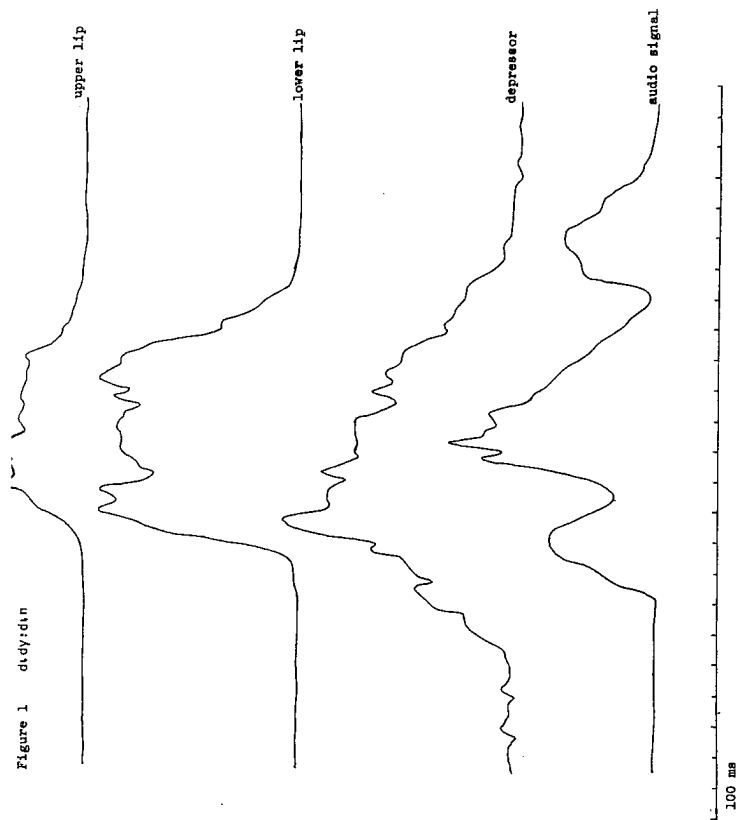


Figure 2 dɪdɪdɪn

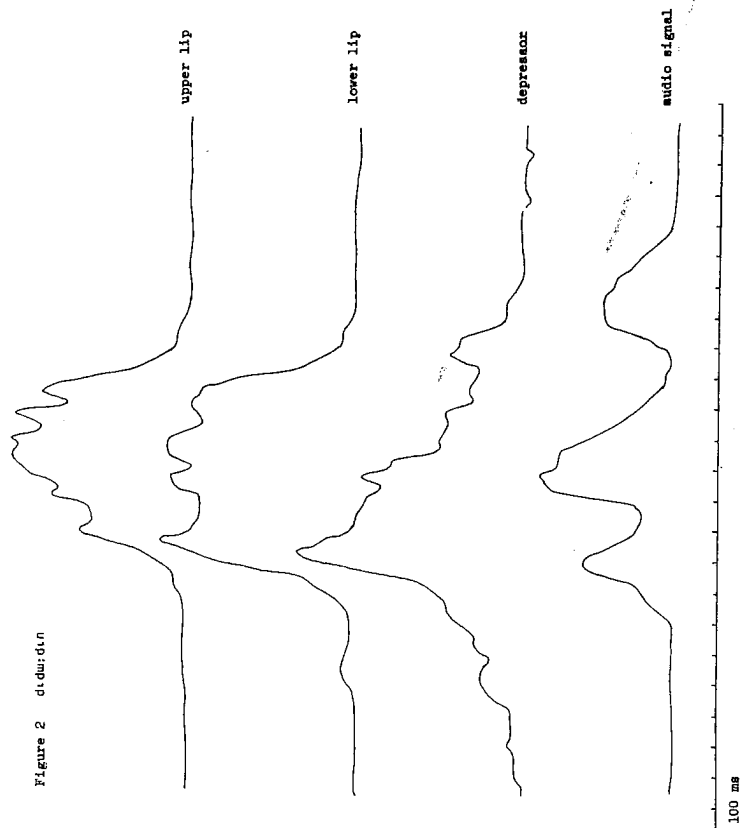


Figure 3 dɪdɪdɪn

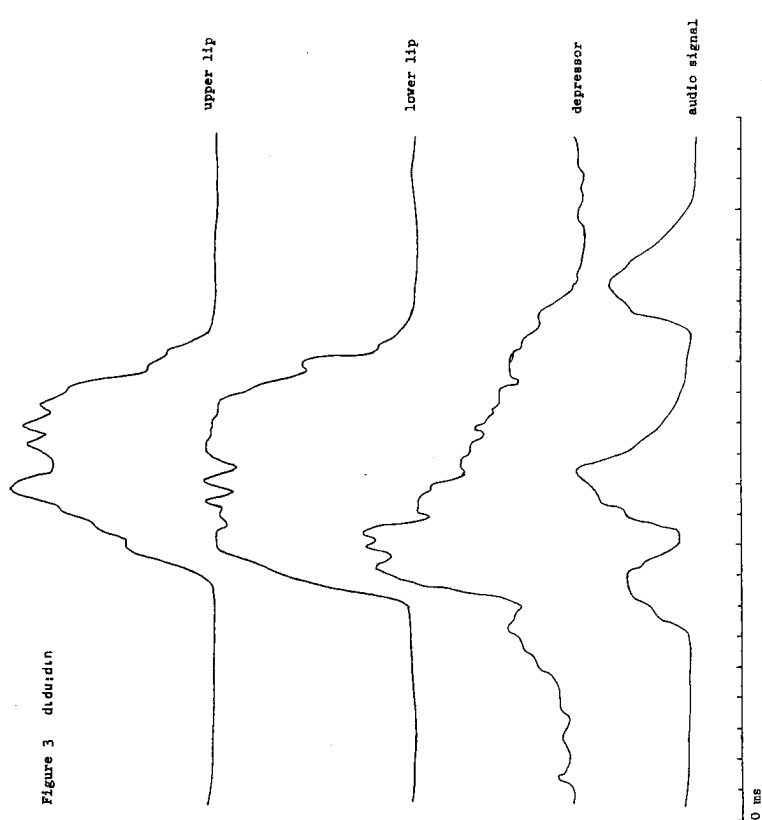


Figure 4 dɪpɪpɪn

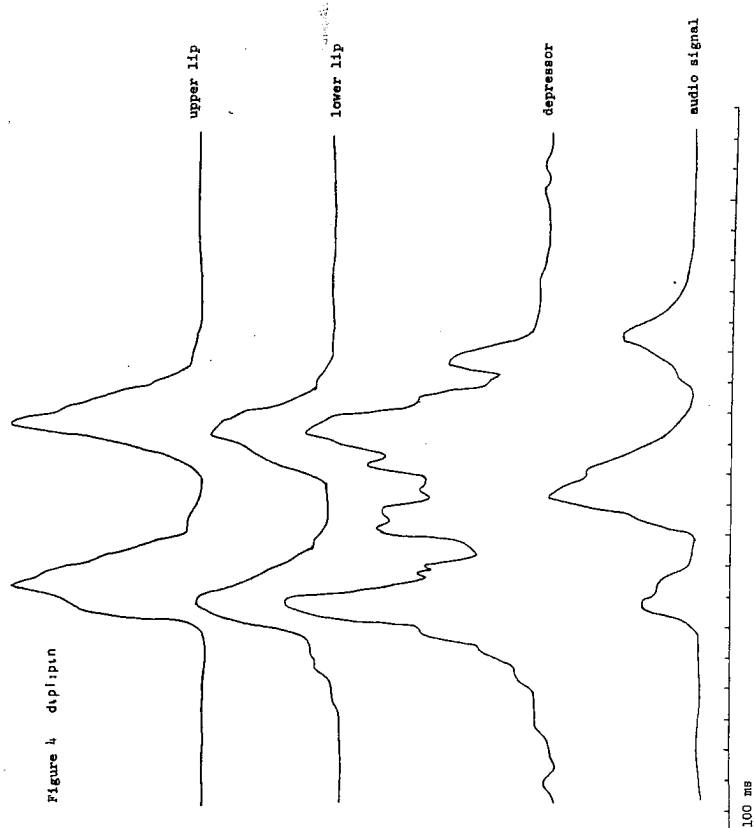


Figure 5 didadidin

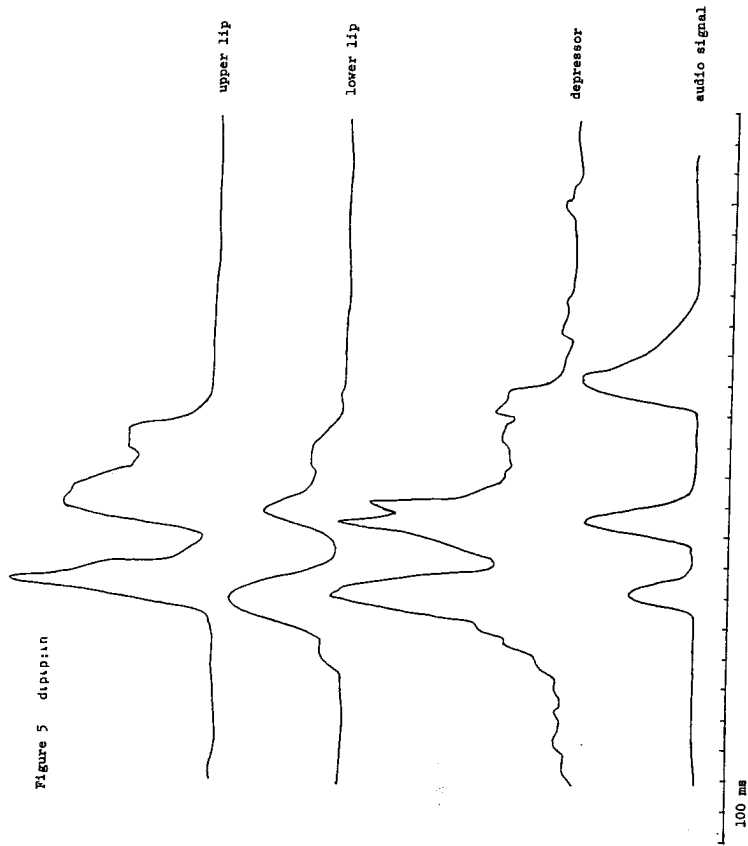


Figure 6 dido:dn

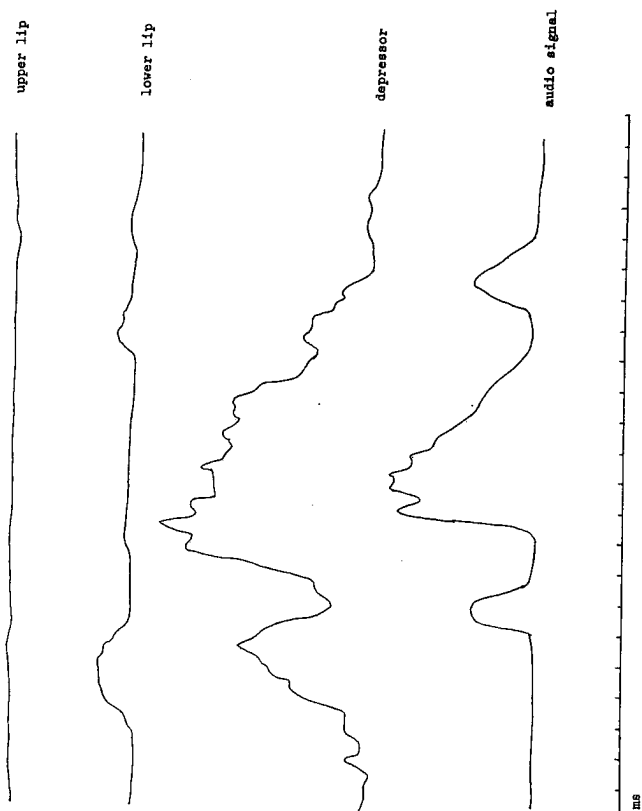


Figure 7 didardin

