Offprint from

CURRENT TRENDS IN LINGUISTICS

Edited by

THOMAS A. SEBEOK

VOLUME 12

Linguistics and Adjacent Arts and Sciences

1974
MOUTON
THE HAGUE • PARIS

ON TIME AND TIMING IN SPEECH

LEIGH LISKER *

However the student of language chooses to regard the object of investigation, whether as an assemblage of sentences[1] of indeterminate number or as a finite system of rules for sentence generation, he must at some point in specifying a language talk about elements and their arrangements. And however complex the network of relations that fix the elements within the sentence, the arrangement of these elements is the simple one of serial ordering. In the case of sentences 'actualized' as pieces of speech this serial ordering is necessarily one of temporal sequence, the elements of which are phonetic segments or 'speech sounds', i.e. the lowest-level phonological units susceptible of physical description. This description is in general resolvable into a set of specifications with respect to a certain number of parameters, so that any two phonetic segments are comparable as points located within some multidimensional physical space. These parameters, whether they specify states of the speech-generating mechanisms, acoustic properties of the speech signal, or even the instructions to the speech apparatus, are usually chosen for their usefulness in accounting for listeners' ability to decide consistently that some speech pieces are repetitions of a single sentence and that others are instances of different sentences. Thus for the linguist in particular the problem of describing the phonetic segments of a spoken sentence is limited, in that his interest is restricted to the potentially distinctive properties of those segments. His physical description of a segment is then partial and relational, constructed with an eye to the total ensemble of segments required by a universally applicable system of phonetic transcription. To be sure, the phonetician is in general also much concerned with the differential properties of segments, but his interest in speech behavior is not confined to those aspects directly relevant to the task of developing an efficient system for spelling or classifying sentences. For him it is not nearly enough to describe a sentence, here considered a class of speech pieces, as the mere encipherment of some particular graphical string, nor is he prepared to equate the physical description of a sentence with the sequence of physical specifications of the segments said to compose it.

* Much of the work of the Haskins Laboratories to which reference is made in the present paper, including that of the author, was carried on with the support of The National Institutes of Dental Research and of Child Health and Human Development.

[1] The term 'sentence' might be understood to include sequences of sentences which constitute partial or complete texts.

To the student of speech behavior the object of interest in its most directly observable form presents itself as an ensemble of auditory signals generated by complexly interrelated activities of body parts which control the movement of air into and out of the vocal tract. Whether considered in its acoustic or in its articulatory aspects, a speech piece is describable as a 'time function' of one or more dimensions of the kind called 'quasi-continuous'. By this is meant that the signal, although not devoid of discontinuities or abrupt changes with time, resists efforts to analyze it into segments which are simply related to the phonetic segments established by the linguist's auditory-phonetic analysis (Lisker 1957b; Fant 1962; Lüdtke 1969, 1970). For the linguist it has been said to occasion no difficulty that the speech signal does not yield physical segments matching those that underlie his phonological analysis (Chomsky and Halle 1968: 294). His interest in the speech signal stops effectively at the level of auditory analysis,[2] for this provides a sufficient basis for developing a useful spelling system by which to represent those facts about speech that engage his concern. Moreover he may justifiably assume that his phonetic segments are ultimately relatable to the physical signal, whatever the difficulties that the phonetician might have to contend with in establishing those relations. In assuming this onerous but not unrewarding assignment, however, the phonetician need not take it to mean that his role is merely one of validating the linguist's phonetic description, on pain of being declared linguistically 'irrelevant' in the event of failure. Beyond accepting the linguist's phonetic transcription as representing the result of some kind of perceptual-linguistic processing of the auditory signal, the phonetician is obliged neither to consider very seriously the physical content of the linguist's segmental description nor to adopt his alphabetic model as the most appropriate one for speech. Moreover the phonetician may well take issue with any implication that the investigation of speech behavior, except for the necessary minimum represented by the linguist's phonetic transcription, is devoid of serious linguistic interest (Mattingly and Liberman 1970).[3] If the aim of phonology is to give a coherent account of the linguist's auditory analysis, and its phonetic component is only physical enough in reference to serve as an 'objective' description of the elements of that auditory analysis, then the relation between the auditory unit and its physical specification is the simplest possible. It is also relatively uninteresting as a theory of how auditory signals recognized as speech are in fact identified with sentences of a language.

Underlying the linguist's graphical representation of a sentence is a model of the

[2] Whatever extra-phonetic information he brings to bear in deciding whether or not he has missed something in his phonetic description, that description itself is a matter of auditory decisions. The use made of grammatical information is a matter of 'discovery procedure'.

[3] The basis for this disinterest on the linguist's part may have been expressed most overtly by Hockett (1955 : 180), who supposed that no matter where in the communication channel we chose to study the linguistic signal, we should determine a set of elements that, despite very different physical properties, would show interrelations identical with those derived by the usual linguistic procedures. In short, the linguist would learn nothing he did not already 'know'.

speech piece as a temporal sequence of articulatory states, their acoustic resultants or their neural-command antecedents, which are themselves largely 'timeless' (Abercrombie 1967:42, 80–1), in that a particular segment is no more to be characterized by the time interval over which its defining physical properties are maintained than its graphical representative is by the space it may occupy on the line of print. Given the observed 'elasticity' of speech in the time domain (Gaitenby 1965), it is clear that the linguist's representation of a sentence must very particularly eschew any reference to absolute time intervals. What is invariably important about a segment in relation to time is not its temporal extent but its position in the temporal sequence of segments. The segments with their phonetic specifications might be thought of as samplings of the speech signal taken at a number of points in time, with the number for a given sentence being largely independent of the time taken for any particular performance of the sentence. On the other hand, however, it is recognized that speech events do in point of physical fact occupy some measurable time interval, for there are occasions when linguists include a reference to relative length in specifying a segment. Thus some vowels in a given language may be said to be shorter or longer in certain contexts than elsewhere, or they may show length differences in conjunction with differences in articulatory position. Such temporal differences are usually taken to be linguistically non-distinctive, an evaluation which the linguist may base on a particular phonological analysis, but which the phonetician seeks to derive from their phonetically redundant status as the physical or physiological consequences of other properties of the speech contexts in which they have been observed (Abramson 1962:109–110; Elert 1964:39–43; Hadding-Koch and Abramson 1964). There is an extensive literature along these lines, reviewed most recently by Lehiste (1970:18–41), which reports various kinds of 'conditioned' durational differences among classes of phonetic segment types in a number of languages. These differences are not only not significant linguistically, but many of those described in the phonetic literature are unreported in the linguist's phonetic description, and are presumably not detectible by ear as a temporal phenomenon, if at all. That temporal differences may sometimes figure more importantly in distinguishing segment types is suggested, for example, by statements to the effect that flapped consonants and semivowels 'admit of no duration' (Catford 1968: 330), presumably as distinct from stops and vowels respectively, and that consonants which may function as syllable nuclei do so chiefly by virtue of a significantly greater duration (Jones 1950:139). Finally, linguists sometimes include relative length as a phonetic feature of segments in recognition of the fact that certain linguistic contrasts depend essentially on a speaker's maintaining a particular articulatory posture for a shorter or longer time interval (Abramson 1962; Liiv 1961, 1962; Nasr 1960; Obrecht 1965). In this last situation, where a temporal dimension would seem unavoidably to belong to the set of segmental properties, a common practice of linguists is to analyze a 'long' segment into a sequence of identical 'short' ones (Bloomfield 1933:109–110; Hockett 1955:77), whereby the need to include length as a segmental feature is obviated. This is not invariably done, to be sure, in which case length is then recognized either as a segmental property or as a special kind of segment, which might be defined phonetically as the prolongation of the articulatory position of an immediately preceding element. The problem of choosing among alternative 'solutions' for representing a distinctive feature of length seems to be settled arbitrarily, by the application of criteria that are not overly compelling (Gleason 1961:282–284; Lehiste 1970:44–46), and in any case do not rest on a secure phonetic basis (Jones 1950:115–120; Hockett 1955:76–79). Nevertheless, depending on the choice made, a feature of relative length is or is not included among the set of segment-specifying features.

Despite the options available to the linguist in deciding how to define the phonological status of contrastive length, there would seem to be reason to consider a feature of this kind a foreign intruder within the set of dimensions used to characterize the phonetic segments. Phonetic segments are, after all, thought of as the building-blocks of a static-phonetic and not a dynamic-phonetic description. If a certain utility is conceded the first kind of description, which restricts itself to the physical specification of the segment inventory, and if at the same time one recognizes the need for a more dynamic phonetics not so closely tied to the segment, then it would seem that temporal features, insofar as they are interpreted perceptually as length, belong properly to the second type of description, whose domain is generally speech pieces of greater than unit segment composition. Unlike features of articulatory configuration, which by and large are considered characteristic of a segment in that they can be specified independently of context, temporal properties of a segment cannot for the most part be defined except as that segment participates in a longer piece of speech. The better way to put it, perhaps, is to say that length is not at all a property of segments as items in an inventory, but rather of longer entities which are in part describable as sequences of segments. In short, such temporal features as relate to length are prosodic in nature (Fry 1968:370; Lehiste 1970:2), in that their identification depends on the comparison of segments under conditions that exclude the possibility that the length difference have no sentence-differentiating function. Like features of differential pitch and prominence, relative length is also prosodic, according to Lehiste, in that it is 'a secondary, overlaid function of inherent features' (Lehiste 1970:2; also Peterson and Shoup 1966:114). But perhaps the three features are not equally prosodic. Pitch and prominence differences function in a more nearly orthogonal relation to the sequence of segments than does length, and this may help explain why it is the last which is often treated as just another linear segment. Then, in terms of a purely segmental description, the length of a speech piece is defined as the number of segments into which it is analyzed. If a linguistic contrast appears to depend on a length difference between two otherwise similar segments, the common linguistic practice of identifying the prolonged articulatory posture as a sequence of identical segments amounts to representing the length difference as a difference in the number of segments composing the contrasting speech pieces. Such an analysis would seem to recommend itself on the ground earlier mentioned, quite apart from whether or not it were 'allowed' by the phonotactics of the specific language.[4]

Articulatory postures prolonged to a significant degree are not always treated as sequences of like segments, so that a feature of relative length may have the status of a segmental property. As such it is of course not considered to represent a continuously variable quantity; rather does length come in quanta or 'morae' (Bloomfield 1933 : 110; Hockett 1955 : 61). A short segment differs from a long one, generally, in that it contains one mora as against two or, in rare cases, three morae.[5] The notion of the mora, dictated by the convention that speech be represented quantally, is consistent with and complementary to that of the 'intrinsic duration' of segments (Lehiste 1970: 18–9, 27–30), which, for a given overall speech rate, is determined by their specific segmental properties. Where the length of a segment claims the linguist's special attention this serves to alert him either to some contextually determined perturbation of its intrinsic duration or to the need for further segmentation of the sentence. Thus an interval of the speech signal for which no change of articulatory configuration may serve to mark a segmentation point is nevertheless divided into two or three segments on a purely temporal basis — a judgment that the interval of articulatory stability occupies, not the one mora its segmental properties 'entitle' it to, but two or three times that intrinsic duration. Whatever the phonetician may say about differences in segment size (Abercrombie 1967: 40), the segment as a unit of linguistic description is then most often specified without a temporal dimension, i.e. it is simply understood to be of unit length. The variability in segment length that may be observed in speech is then ascribed to a number of factors whose durational effects can be then separately determined (House 1961: 1174–8; House and Fairbanks 1953: 105–13).

We see then that the linguist tends to minimize the temporal aspects of speech, treating it as an orderly sequence of evenly spaced units of fixed size. It may be said that this picture corresponds to the linguist's purpose in formulating his phonetic specification, which is to classify sentences. For this it suffices to NAME the segments as sequentially ordered components of sentence designations, using as names for the segments sets of distinctive features by which the segments may be located as points in orderly array within some physical, phonetic or phonological space. From the standpoint of the phonetician the linguist's purpose is an overly

[4] Thus, for example, Lehiste (1970: 44) allows this analysis in the case where a language has 'consonant clusters that function in the same manner as long consonants ... regardless of whether it is possible to demonstrate, phonetically, their geminate nature'. One is entitled to wonder why it could not be said of a language devoid of consonant, or vowel, clusters that the only allowable consonant, or vowel, sequences are those which involve identical elements.

[5] Bloomfield in fact violates the usual practice of dealing in only integral numbers of morae (1933 : 111) and thereby most strongly implies the quantal nature of the unit. But the mora also appears to be not clearly different from a 'degree of quantity' (Lehiste 1970 : 48), so that its quantal nature is not always evident. Jones' 'chrone' (1950 : 126) is, on the other hand, simply the actual measured duration, and as such continuously variable.

modest one, and the static representation that satisfies it is not to be considered an adequate description of either speech activity or the acoustic signals it generates. The segment properties which the linguist asserts to have distinctive function must for the phonetician be placed in the context of a more general understanding of the rules governing the operations of the speech-producing and speech-perceiving mechanisms. Not least among these may well be timing rules which determine the movements of the various articulators and their temporal interrelations, for there is good reason to think that, in order to yield the proper order and number of segments required to produce a given sentence in a normal manner, perhaps more than one hundred muscles must be coordinated to operate within tolerances of sometimes less than a tenth of a second (Lenneberg 1967:91–103). Failure to maintain relatively fixed phase relations among the component gestures of a given segment sequence will result in variation in the number of acoustic segments generated, and sometimes in the number and character of the phonetic segments perceived (Kantner and West 1960:264–269). There is also clear evidence that many kinds of deficient speech, including both humanly produced (Avakjan 1968; John and Howarth 1965; Lowe and Campbell 1965; MacNeilage et al. 1967; Shankweiler et al. 1968) and machine-generated varieties (Mattingly 1966), involve disorders of timing.

In the foregoing we have been considering the phonetic segment as an articulatory and/or auditory event that is relatively stable over some time interval. The temporal extension of the segment is also recognized in cases where its component features, theoretically co-occurrent, have certain manifestations that are noticeably not simultaneous. Here the temporal factor is reflected more in judgments of temporal order than extent. Thus a segment will be described as a glottalized stop, with closures of both the glottis and the oral cavity, but the two releases may not be synchronous, so that their temporal order must also be specified (Hockett 1955:35). It can be objected that such entities hardly qualify as phonetic segments, but the basis for this objection is by no means obvious, inasmuch as a reading of the literature reveals rather different notions of what the 'phonetic segment' ought to mean. Much opinion on the matter is flatly assertive and to a considerable extent controversial because, as in other areas within the linguistic sciences, it fails fully to distinguish between matters of definition and matters of physical and perceptual fact. Perhaps the most purely phonetic definition of the segment is that of Pike (1943:107), who described it as 'a sound (or lack of sound) having indefinite borders but with a center that is produced by a crest or trough of stricture during the even motion or pressure of an initiator'.[6] But in the literature generally a distinction between the

[6] The 'phone type' of Peterson and Shoup (1966:112) is perhaps even more stringently phonetic in nature, but so much so that it may not be synonymous with the phonetic segment of standard linguistic description. Thus, for example, while their system of definitions provides a place for 'complex' phones, these do not apparently include a class of affricates, which certainly belong to the set of phonetic segments available to the linguist.

phonetic segment as an articulatory state and as a unit whose size is determined essentially by phonological considerations is not clearly made. There has been no lack of definitions, ranging from the purely physical one of Pike's, over the purely phonetic or physico-perceptual (Abercrombie 1967:42), to definitions which make the phonetic segment purely phonological in nature (Gleason 1961:248–9). It is undoubtedly true that in practice the linguist's phonetic segments are not to be equated with 'prelinguistic', that is, purely auditory judgments of articulatory states. Certainly Chomsky and Halle (1968:293–5) are correct in describing the linguist's phonetic segment, not as a judgment of physical fact uninformed by linguistic knowledge or intuitions, but rather as fundamentally a phonological unit described by a set of physical features which are themselves selected on grounds as much phonological as physical. For them it is then a phonological unit, though it is at the same time unitary by some universally valid phonetic theory (1968:28). Perhaps because the phonetic segment, like the syllable, is one which linguists and other users of language seem to have strong convictions about, the definitions advanced have often taken the form of assertions as to what it IS, rather than being proposals as to the kind of definition that will yield segments appropriate to some stated goal. Of course, to complicate matters, it is not true that either linguists or other people are always quite certain about what the segments of a speech piece are, even when no definitional dispute is overtly involved. A notorious example is furnished by the history of dispute concerning the status of the affricates.

If the phonetic segment is characterized as a complex of simultaneous features, then there would appear to be no place in a phonetic representation for entities such as 'prenasalized stops' (Jones 1950:78–81; Ladefoged 1964:23–4) or 'occluded nasals' (Bauernschmidt 1965:477, 480–1), which can only be considered phonetically unitary if we can discover some basis for asserting that they constitute signal complexes which are integrated to a degree that ordinary sequences of nasal and stop consonants are not. Unlike entities such as the glottalized or aspirated stops, for which we may plausibly assume certain articulatory maneuvers to be executed in close if not precise synchrony, a prenasalized stop or an occluded nasal involves a sequence of two necessarily non-overlapping states of velar opening. If the segment is to be associated with a single articulatory state, it would seem that such entities ought to be considered two-segment sequences. In the case of glottalized and aspirated stops, where the nature of the temporal relations among the component articulations is neither one of simultaneity nor one of strict succession, there seems to be no firm phonetic basis for deciding whether to regard them as single segments or as sequences. If the treatments of such phonetic complexes that are found in Hockett (1955) may be considered representative linguistic practice, then clearly the notion of the phonetic segment as a purely phonetic entity is derivative; its limits are not so fixed that it can be said to represent a particular state of the vocal tract, but it is rather a phonologically unitary stretch of speech signal to which a phonetic description is attached. This phonetic description is most often, in fact, the specification of a single articulatory state, in which case the linguist's phonetic segment coincides with one to which the phonetician would as happily subscribe. But this need not be the case, and then it becomes necessary to describe certain segments as ordered sequences of articulatory states. That a speech piece will by and large be analyzed auditorily into segments that are very much the same no matter what the language is certainly true (Pike 1943:116), and these will be of the size of phonological units. There nevertheless remain certain phonetic events whose status as segments is ambiguous except by some phonological decision criterion.[7] Where this leads to the establishment of complex phonetic segments of the kind just mentioned, a feature of temporal order must be included in the overall inventory of features needed for the phonetic description of segments generally. Then one might define as phonetic segments such entities as preaspirated stops (Heffner 1950:168) along with the more usual postaspirated types, and suppose that these two classes are adequately distinguished by the different ordering of their constituent events. It would then seem, however, that the nature of this temporal feature is not different from the ordering relation among segments in longer sequences, and that therefore any notion of the phonetic segment as a purely phonetic unit is strictly untenable.

Let us now turn to the question of the affricates. Laziczius (1961:62–66) gives a fairly detailed resumé of the various opinions expressed concerning these items, and one gathers an impression of substantial agreement on their physical nature and endless wrangling as to their segmental status. They are produced, it seems fairly well agreed, quite in the manner of stop consonants so far as involving complete oral occlusion, but the time course of their releases is such that a noise interval results which is longer and identifiable as a fricative homorganic with the stop element. The affricates are simply then slowly released stops (Heffner 1950:120; Hockett 1955:81; Gleason 1961 : 248; Abercrombie 1967 : 147). Their stop and fricative phases must necessarily be homorganic, and presumably the stop component can have no stop-like release intervening between the occlusion and the fricative phases, since the latter is in fact the release. This slower release in the affricated as against the ordinary stop consonant produces an interval of noise that for some scholars is audible as a brief fricative (Heffner 1950:249; Abercrombie 1967:147), and for at least one other is 'too short to be heard separately' (Gleason 1961:249), at least, we may infer, by speakers of a language in which the affricate represents a single phonological unit. According to some accounts, moreover, the fricative phase of the affricate is distinguished from a true fricative again by a difference in duration (Gerstman 1956, 1957; Brooks 1964). Thus it appears that there is general agreement that the gesture of opening after a total occlusion is under temporal control to the extent that three rates of opening are distinguishable. As between the affricates and the stop + fricative sequences, it has

[7] These are Pike's 'fluctuants', which do not constitute a phonetic class apparently, but are simply those events about whose segmental status linguists are in doubt.

been supposed that the first are released with a single articulatory movement, while in the sequence 'the articulator goes through successive motions' (Hockett 1958: 81). However there seems to be solid evidence only for the durational differences, and none either to confirm or deny that these differences apply to articulatory movements that are otherwise identical.

The tenability of any purely phonetic basis for deciding on the number of segments composing an affricate is open to question no matter whether it is claimed to be temporal or articulatory or both together, for it is by no means certain that these aspects of speech activity can be quantified to yield just the classes to be defined. Thus some phonetic descriptions suggest that releases of oral occlusion may show as many as four durations, by which stops, affricated stops, affricates and stop + fricative sequences may be distinguished (Gimson 1962:153–154, 166–169; Malmberg 1963:43, 50–51). Moreover the usefulness of the articulatory argument based on the necessary homorganicity of the affricate components is reduced when one finds it applied as a basis for justifying the inclusion of both [pΦ] and German [pf], and English [tr̥] as well as [tš] among this class of items (Gimson 1962:31, 168), and when it is recalled that the effects of coarticulation, under which this 'homorganicity' may or may not be subsumed, have only just begun to receive quantitative treatment, both as to their magnitude and temporal extent (Chistovich et al. 1965: 126–132; Öhman 1966:151–168; Öhman 1967:310–320; Lindblom and Studdert-Kennedy 1967:830–843). If a phonetic description of speech activity must be the description of segments in linear order that the linguist deals in, then precise models of coarticulation are needed before the segment can be taken as a phonetic rather than a phonological unit.[8] Until the notion of a phonetic segment as an interval over which a fixed articulatory position is maintained can be better reconciled with the observation that at any instant one or more of the speech organs is in motion (Gimson 1962:42), it would be safer to pretend, following Gleason (1961: 248), that 'Phonetically there is no basis for the kind of segmentation which we customarily use', than to insist that the elements which a segmentation yields can all be specified by a set of fixed articulatory configurations. Neither point of view can be completely valid, however; there are certain kinds of acoustic discontinuities that perhaps are always interpreted as segment boundaries, but not every signal change has this perceptual effect (Pike 1943:46–47; Fant 1960:21–26).

Despite the fact that they may define certain auditory complexes as phonetic segments, some linguists and phoneticians nevertheless seem averse to abandoning utterly the idea that their segments are fundamentally units derived by a segmentation that is entirely divorced from phonological considerations. For them the fact that their phonetic segments converge generally with those derived by a language-specific segmentation based on phonological analysis is not only independently significant, but it impels them to suppose that ALL phonetic segments must be pho-

[8] Purely phonological considerations most obviously motivate proposals such as those of Hofmann (1967), who would classify all initial clusters in English as unitary segments.

netically unitary, despite any auditory complexity some of them might display. Thus, if considerations of coding efficiency indicate that certain stop + fricative sequences should be treated as units, they feel compelled to discover that phonetically their constituents are more intimately bonded than are stop + fricative sequences generally (Hockett 1955:164). This would not eliminate the need to include a feature of temporal order in the description of complex segments, but would allow us to distinguish phonetically, and most probably on a temporal basis, between these and sequences of stop + fricative segments. Thus, for example, a phonologically unitary sequence in Polish is distinguished from a sequence of segments presumably comprising the same phonetic events either on the basis of a difference between a close and an open transition from one event to the other (Bloomfield 1933:119), or by a difference in the durations of the fricative intervals (Brooks 1964). Similarly a one-segment occluded nasal $[n^t]$ differs from a sequence [nt] in that the nasal component of the first is longer, while the stop is shorter, less forcefully articulated, and more likely to be 'contaminated' by voicing than are the corresponding events of the segment sequence (Bauernschmidt 1965:480).

There is another strategy available for dealing with a sequence of phonetic events which the linguist would regard as phonologically unitary and therefore to be distinguished phonetically from a sequence of segments. Instead of referring the sequence of events to a sequence of articulatory states which may be especially closely integrated when viewed as parts of a single segment, he can suppose that they are generated by the interplay of several articulatory maneuvers which are essentially synchronous (thus Chomsky and Halle 1968:326–329). Though this mode of phonetic accounting may not always remove a temporal dimension from the set of segment-specifying features, and though it seems designed ad hoc to preserve the integrity of the phonetic segment as the linguist's basic unit of transcription, still it is of interest in that it allows for the possibility that the temporal order of acoustico-perceptual events is not precisely matched by that of corresponding events at some deeper level of the speech-generating process (Lenneberg 1967:93–98). It is intriguing to postulate that the perceptual unity of the phonetic segment, where it is acoustically a complex of successive events, reflects a unity at some level of the speech communication process that is less accessible to direct observation. In one sense linguists have presumably always been ready to believe this, i.e. that somewhere in the speaker's nervous system there are 'units of internal flow' (Hockett 1955: 5) corresponding to the linguist's discrete phonological elements and perhaps to the native speaker's intuitions as well (Jones 1950:79). This belief, however, amounts to little more than an assertion of the 'psychological reality' of such entities, and not necessarily to an assumption that the combination of certain specifiable articulatory properties or movements, or perhaps neural commands, generates the observed complex of acoustic events, and that the transformation of the segment as a complex of co-occurrent features at the deeper level into the not nearly so well integrated acoustic complex is in conformity with any generally statable rules (but see Pike 1943:114; Chomsky and Halle 1968:324). Assertions as to features at some deeper level are by definition less easily tested by physical observation, but with the development of new laboratory techniques in recent years hitherto inaccessible stages of the speech process are becoming vulnerable to study (see, for example, the papers by Harris and Sawashima in the present volume).

The role of a feature of temporal ordering is most obviously important in describing acoustically complex segments, but it cannot be assumed to be negligible in the case of the simple ones. This would amount to supposing that the nature of the articulatory movements and the acoustic signal produced during the first half of the interval 'occupied' by a phonetic segment is essentially symmetrical with that in the second. The only immediately obvious motivation for such an assumption would, I think, derive from the persistent tendency to view the graphical representations of speech as an adequate model of the speech process. Since the alphabetic signs of a phonetic transcription are freely transposed and 'spaceless', in the sense that no significance attaches to any asymmetry in their shapes, it might be supposed that the phonetic segments are themselves as freely movable. There is abundant evidence that this is just not so. Thus, for example, attempts to fabricate acceptable speech signals by cutting apart tape-recorded speech into 'segment-sized' bits and reassembling them in different orders have regularly failed expectations based on the assumption that all tape snippets bearing the same phonetic segments were identical in their combining behavior (C. M. Harris 1953; Liberman et al. 1967: 441). Nor can a speech piece consisting of auditorily simple segments be recorded and, upon being replayed with reversal in time, be heard as the same segments in a sequential order the reverse of the original (Harrell 1958). Recent experiments in the perception of dichotically presented speech stimuli also indicate that when these are perceived as single 'fused' speech pieces, the perceived order of segments is not generally predictable from the temporal order of presentation, but is significantly an effect of their non-uniformity in the time domain (Day 1970). Thus segments of the acoustic speech signal, selected as physical correlates of phonetic segments, are not freely commutable elements, for their shapes make them in general compatible only with their immediate neighbors in the speech piece in which they originated, plus perhaps certain others not predictable on the basis of the phonetic transcription (C. M. Harris 1953:962). In other words, the phonetic segments are not acoustically shaped like similar beads on a string, but rather like the pieces of a jigsaw puzzle.

So far in the present discussion we have been playing the phonetics game as the linguist plays it, taking as the primary data his auditory phonetic evaluations of speech and seeking out physical properties that might correlate significantly with them. Because auditory analysis, both with and without the light of linguistic knowledge, gives us speech in the form of phonetic segments in temporal order, we look for discontinuities in the acoustic signal that will mark off time intervals matching the phonetic segments in some sense. If a piece consists, let us say, of x number of phonetic segments, then we seek x + 1 points (including those of onset and termination of the piece) at which to cut the physical signal. The resulting x time intervals are then paired with the same number of phonetic segments in a way that we feel represents an optimal matching of the two kinds of elements. The method involves a knowledge of the acoustic correlates of certain articulatory events that can serve as temporal reference points in the acoustic signal, e.g. the brief noise 'burst' of a stop release, together with an ability to follow variations in the acoustic signal that reflect the cyclic changes in oral aperture at syllabic rate, as well as the gross differences which correspond to alternations in voicing state. But this matching procedure, however 'optimal' it be judged, does not guarantee the significance of the pairing relation it establishes. The assumption that somewhere in the acoustic signal are to be found properties which can be shown to correspond to those ascribed to the phonetic segments by no means implies that the acoustic signal is necessarily to be segmented along the time dimension as described above, for this procedure presupposes that the signal must be parcelled out among the segments with neither overlap nor residue. Every bit of acoustic signal must, in short, be assigned to one and only one segment. Since there is an abundance of evidence to refute any such simple view of the relation between the phonetic evaluation of speech and its acoustic properties, the meaning of the kind of segment matching we have just described would seem at best to be not transparently clear. There is no problem about segmenting the acoustic signal, since, contrary to the once fashionable emphasis on the continuous nature of speech, we usually find more acoustically reasonable cutting points than we know what to do with (Fant 1962:30). At the auditory level, however, it is more questionable that we may speak of boundaries 'between' phonetic segments (Pike 1943:107). If one nevertheless accepts a view of speech as a flow of discrete segments that follow one another without hiatus, assumes this view to be applicable without qualification to the acoustic signal, and performs the matching operation, there still remains the far from trivial matter of deciding what aspects of each segment are to be isolated as physical correlates of the matched phonetic segment. This problem is of course central to the field of acoustic phonetics, and knowledge of the phonetically significant features is by now detailed (Delattre 1968). The determination of these features has been a fairly straightforward matter, particularly since the development of techniques for the controlled synthesis of speechlike auditory signals, for the set of elements to be specified physically is well defined by auditory judgment and lingustic function. The physical specifications in so far as possible are correlates of the static-phonetic descriptions of the phonetic segments — i.e. vowels and certain of the consonants are described by fixed 'target' values of particular acoustic dimensions, and the values exhibited by these segments in running speech are viewed as more or less systematic deviations from those targets (Stevens, House and Paul 1966; Lindblom and Studdert-Kennedy 1967). The precise location of acoustic segment boundaries is not crucial to this enterprise, at least with regard to the specification of the simple phonetic segments, although the classification of a segment as 'simple' or 'complex' would naturally depend on the placement of those boundaries.

Matters are rather different when we come to consider the question of matching acoustic and phonetic segments with respect to the time dimension. In the first place, as we have seen, the feature of length is only marginally considered to be a property of the phonetic segment. If an articulatory position seems to be maintained for different durations in linguistically distinct speech pieces, the linguist does not promptly consider length a segmental feature; he is more likely to look for some other phonetic difference by which to explain the time difference. Whereas segments are classified regularly with respect to such features as voicing, nasality, tongue height and lip posture, there is no purely phonetic classification of segments on the basis of their perceived durations. Only where such a classification is forced by the recognition of length contrasts in a particular language is this done, and then only after the linguist has searched in vain for other feature differences that might account for the temporal difference. This reluctance on the part of linguists to recognize a feature of temporal control at the segmental level is most strikingly exemplified in Chomsky and Halle (1968: 317, 318, 327). Thus in general, phonetic segments can hardly be said to have a well defined status with respect to the dimension of time. Of course, since we here speak of phonetic segments we are thereby considering time in the sense of a perceptual dimension, one that is independent of phonological considerations. In the second place, the determination of segment boundary locations in the acoustic signal is of obvious importance if one wants to make physical time measurements that can be meaningfully related to the length, or perceived duration, of segments or longer stretches. Moreover, unlike the situation with respect to the indisputably segmental features, it is not generally possible to specify target values for the duration of a phonetic segment, since its so-called intrinsic duration can only be one factor, and most probably not a major one, in determining its duration in any given context, as produced by a given speaker, in a particular text, and under particular social and physiological circumstances. It is of course possible to define classes of acoustic segments as physical correlates of the phonetic segments, and to proceed to determine their mean durations, but it seems farfetched in the extreme to suppose that these various values are targets in the same way that the formant frequencies of isolated vowel sounds are construed to be (Lindblom 1967: 4–5).

It is perhaps useful here to emphasize a distinction generally made, and which we have been observing more or less consistently, between duration as a physically measurable attribute and length as its closest single analogue in the domain of perception (Malmberg 1963: 122). This distinction, which parallels others drawn between physical dimensions and their perceptual correlates (e.g. fundamental frequency vs. pitch) is perhaps less carefully observed than those others because the relation between duration and length is felt to be so much closer. Fundamental frequency, the number of fundamental periods per second, is in speech a measure of glottal pulse rate, but pitch is not simply a subjective estimate of that rate. Pitch is rather an auditory property of a pulsed acoustic signal whose frequency is such that the individual pulses are not separately perceived. Length, on the other hand, is an auditory judgment as to the duration of a sound signal, where this duration presumably might be more accurately measured by instrument. Pitch is, by its nature, not subject to instrumental measurement. Of course, even with respect to duration and length, particularly when dealing with speech signals, we can only say that the accuracy of statements about length may be determined simply by instrumental measurements of duration if we accept the listener's description of his length judgment as an estimate of temporal duration. Now while we should in fact expect the duration of an auditory signal to be the major determinant of its length, the possibility exists that the relation is nonlinear (Fry 1968:386), and that factors other than simple duration are of significance. Moreover there still remains the 'problem of segmentation', which in the present context is one of deciding what acoustic segment or sequence of segments it is whose duration should be defined as the physical correlate of the length of a given phonetic segment. This requires some decision if phonetic segments are to be said to have temporal extension in the sense of physical duration. Failure to define explicitly the physical intervals makes for difficulty in understanding exactly what is meant by published statements giving precise (within tolerances of a few milliseconds apparently) durations of various segment types. One cannot simply carry out measurements of the durations of selected intervals in spectrograms and report these as durational values for vowels, stop consonants, or other segment types; we do not SEE vowels or stop consonants in the spectrogram, we only find evidence of them. The durations reported are of acoustically defined sections of the acoustic signal, and not strictly of phonetic segments, even as measurements of fundamental frequency are not determinations of pitch. These considerations, obvious enough, have prevented neither instrument manufacturers from marketing 'pitchmeters' nor phoneticians from devising conventions that allow them to refer to the durations of phonetic segments.

In adopting measuring conventions for defining, and then measuring, the duration of vowels, which is the class of segments that has enjoyed most of this kind of attention, it has not, as far as I know, been clearly stated what the basis for selecting the acoustic markers of vowel onset and termination should be. The onset of vowels has been fixed sometimes at the point where the glottally excited full formant pattern is established following silence or the release of a preceding consonant constriction (House and Fairbanks 1953:107; Abramson 1962:29), sometimes at the point of the consonant release (Peterson and Lehiste 1960:694), or at a point where the 'intensity curve rises sharply' (Fischer-Jørgensen 1964:182), or where the formant frequencies judged appropriate to the vowel are attained following the transitional movement from a preceding consonant (Delattre 1964). Hadding-Koch and Abramson (1964:98) included transitions within the acoustic interval defined as the vowel. Fónagy and Fónagy (1966:15) adopted the arbitrary procedure, but really no more arbitrary than any other, of grouping doubtful intervals with the immediately preceding acoustic segment. The vowel termination in most cases was located at a point corresponding to the end of the transitional movement to the constriction of a following consonant, though Delattre (1964), consistent with his view of transitions as essentially consonant cues, determined vowels at a point where the formants enter the transition phase.

Where the object of acoustical measurements is to specify 'consonant duration' the conventions are in general complementary to those for the vowels. The literature on measurements associated with consonant length is not large, and perhaps almost the whole of it identifies as the appropriate objects of measurement the acoustic segments that correspond to intervals of constriction. In the case of certain of the languages for which durational data have been reported, i.e. English, German, Swedish and Danish, all languages which include voiceless aspirated stops in their segment inventories, the decision as to the status of the aspiration phase might well affect the magnitudes of reported consonant durations by as much as 75% (Lisker and Abramson 1964).

In view of the somewhat different measuring conventions followed by various researchers, to say nothing of the fact that some reports fail to describe any in satisfactory detail (Parmenter and Treviño 1935: 129–133; Fintoft 1961: 119), the comparability of their data is seriously to be questioned (Laziczius 1961:119), certainly so far as the absolute magnitudes reported. Aside from the matter of measuring conventions, it is known that individual speakers show large differences in overall speaking rate, and that at higher rates it is not the case that segments are eliminated to the point where those retained preserve the durations exhibited at slower rates (Gaitenby 1965:3.4).

The variability in the criteria applied in segmenting the acoustic signal, which derives from the fact that the number and location of 'natural' segmentation points in the acoustic signal do not conform in detail to an intuitively satisfying segmentation based on auditory judgments ('t Hart and Cohen 1964:35), suggests the need for some criterion, not necessarily acoustic, by which to agree on a segmentation procedure. One possibility might be to select intervals for measurement that best match listeners' judgments of relative length. Thus, for example, on the question of whether the acoustic segment identified with aspiration should be considered a phase of the vowel or of a preceding stop, the decision would depend on which assignment yielded physical time values in better agreement with judgments of the relative length of the vowels following aspirated as against unaspirated stops. It might, alas, turn out to be a case of the halt leading the blind, if listeners simply proved unable to make consistent durational judgments when given this task. From figures given by Peterson and Lehiste (1960:701) for American English it would seem that the durations at issue do not differ by much more than the general just-noticeable-differences established for speech sounds (Lehiste 1970: 13). Moreover when the durations of vowels following aspirated stops are determined both with and without aspiration as part of the vowel, the average of these two is almost exactly equal to the mean duration of vowels after /b,d,g/. In other words, given the likely event that listeners do not regularly report differences of vowel length that can be correlated with the aspirate-inaspirate difference in the preceding stop, we cannot simply settle the matter by showing that one assignment of aspiration will yield more nearly equal durations for the two sets of vowels than will the other.

One might return to the static-phonetic mode of description and elect as appropriate intervals those portions of the acoustic signal that, by some more or less arbitrary definition, can be considered to be 'essentially steadystate' (Lehiste and Peterson 1961:272; Gay 1968:1571). The measured durations of such acoustic segments would then be defined as the durations of particular phonetic segments. If this procedure were consistently followed, the picture presented in the literature as to the relative durations of different segment types would require serious revision. Thus statements to the effect that the vowels in English take up a disproportionately large part of total speech time (Parmenter and Treviño 1935:130–131) would certainly turn out not true, given the likelihood that many vowels, the unstressed without doubt, are by any reasonable definition steadystate for at most half the durations often ascribed to them (Bernard 1970:92; Lehiste and Peterson 1961: 275; Lindblom and Studdert-Kennedy 1967:831). The well-established practice of equating the consonant with a relatively steadystate signal interval, e.g. the closure silence of a stop, while taking as the vowel an entire interval of voiced formant pattern, would seem to reflect the application of some highly arbitrary segmentation criterion, or at any rate one that, justifiable or not, has not been put into an explicit form. By the usual assignment of acoustic segments to phonetic we have vowels that may be described as relatively long but rather unstable, and consonants that tend to be shorter but with relatively fixed acoustic characteristics. If one takes the extensive data on American English supplied by Peterson and Lehiste (1960:702) and by Lehiste and Peterson (1961:275) to derive durations of steadystate stretches within the total intervals assigned the vowels, one arrives at values much more in line with those reported for consonant durations.

At this point it is appropriate to ask whether it is necessary to follow the linguist's precisely segmental model of speech in the description of its temporal aspects. As Laziczius (1961:117) put it, 'die Laute einer Lautreihe reihen sich nicht wie die Perlen eines Halsbandes nebeneinander, sondern greifen ineinander ein wie die Glieder einer Kette', and this has been amply demonstrated by the acoustic analysis of speech signals (Fant 1962:30) and by experiments in their synthesis (Liberman 1970a:309–315). We may believe, in accord with the segmental model, that the speaker really 'wants' to produce a string of neatly discrete sound or articulatory units, and that only certain linguistically irrelevant constraints on his ability to execute these at the rates actually achieved in the production and transmission of phonetic segments interfere (Heffner 1950:204–207; Fónagy and Magdics 1960;

Osser and Peng 1964; Lenneberg 1967:90; Liberman 1970a:306). There is reason to believe, however, that the speech decoding mechanism which is presumed to operate on the auditory level could not handle discrete units at those rates (Liberman et al. 1967), and this can be true at the same time that it is also true that speech is normally slower than it need be for comprehension (Fairbanks et al. 1957; Fairbanks and Kodman 1957; Klumpp and Webster 1961; Endres 1968; Sticht and Gray 1969). Even if the rate of segment production hovers about the lower frequency threshold for pitch perception, say between 10 and 20 segments per second (derived from Heffner 1950:204–206), the fact that time compression of as much as 50% may not seriously reduce intelligibility ought to bring the rate of segment identification above that threshold, perhaps to as much as 30 per second (Liberman et al. 1967:432). If at the same time we bear in mind that listeners appear to be unable to judge accurately the temporal order of discrete auditory stimuli until their individual durations are as much as 700 msec (Warren et al. 1969:587), in the case of nonspeech sounds, and that sequences of vowel sounds are not correctly ordered until their durations are more than 100 msec each (Thomas et al. 1970), it seems impossible to reconcile all this with the strictly segmental model of speech. Moreover it becomes difficult to know what inferences to draw from the data published on the durations of the various classes of phonetic segments, or to suppose that these durations have any very close connection with the real units of speech production and/or perception.

If we follow Fant in the exercise of analyzing the spectrogram of a short piece of English (the piece is the short form 'Santa Claus', Fant 1962:30), we isolate nine components, one for each phonetic segment alleged to constitute the piece, which partially overlap in such a way that they yield twice the number of segments at the acoustic level. Most of these acoustic segments comprise acoustic features associated with two phonetic segments; of the remainder most provide cues to either three or four segments simultaneously. If we further ask how much time these nine acoustic components corresponding to the phonetic segments would occupy if they were purely sequential, so that instead of eighteen acoustic segments there were only nine, the answer is that the total duration of the piece would be increased by more than a third. If we were to make the extreme assumption that each of the overlapping components has the minimum duration required for auditory detection and proper identification of the corresponding phonetic segment, assuming that in fact they were clearly identifiable in the temporal order in which they were actually produced, then the durations of these components would seem the appropriate items to measure, since they represent more nearly the durations over which the acoustic reflexes of the separate gestures characteristic of the piece were present in the signal. There are certain further observations that may be made about this sample analysis, which in effect treats the phonetic segments as features of the acoustic signal. While the briefest acoustic segment is of the order of 10 msec in duration, and the mean segment duration is about 115 msec, the shortest component is almost

60 msec long, and the mean component duration is greater than 150 msec. The sequential ordering of the phonetic segments is reflected in the fact that each component either begins or terminates before its successor, and never begins or terminates after its successor. Each component co-occurs, in its initial part, with one or more of its predecessors, and, later on, with one or more successors. Thus the time interval for a given component is acoustically far from steadystate, and may even contain sharp discontinuities. We do not know how much of this complexity is actually required in order that the signal pass as intelligible speech, but some at least is indispensable. This means that the listener is capable of tending to more than one component simultaneously presented in the course of decoding the linguistic message (Liberman 1970b:242–244). If Fant's sample analysis of 'Santa Claus' correctly represents the way in which the acoustic cues are temporally distributed in speech generally, then one must say that only exceptionally does the listener encounter an acoustic segment which includes only a single component. In Fant's example the initial and final acoustic segments are of this type, and otherwise only the phonetic segments [k] and [ɔ] are represented by components which are not always accompanied by components associated with other phonetic segments. The durations of the two one-component acoustic segments are about 55 and 200 msec for [k] and [ɔ] respectively, values that are rather smaller than some phonetic segment durations reported in the literature (e.g. Kent and Moll 1969:1550; House 1961:1174), but quite as large as some others (Sharf 1962:28). On the other hand, the total durations occupied by the components representing the [k] and [ɔ] segments are more like 200 and 340 msec respectively, values that are quite a bit larger than any reported for such segments. However, as we have noted before, a variety of factors affect the durational properties of speech, some drastically, so that it is difficult to evaluate the significance, or sometimes the credibility, of a report of absolute magnitudes of duration without certain information as to the nature of the speech sample measured and the conditions under which it was elicited.

If we elect to equate each phonetic segment of a speech piece with an acoustic segment in one-to-one fashion, choosing steadystate segments, and then define the duration of the acoustic segment as the physical correlate of phonetic segment length, we thereby say in effect that transient or transitional segments make no contribution to perceived duration. If, on the other hand, we make the component, i.e. the stretch of signal over which the acoustic cues to a phonetic segment are spread, the basis for determinations of duration that are to be related to length, we are led to the apparent absurdity of supposing that a single acoustic segment might contribute at once to the length of as many as four phonetic segments (as in Fant's example cited above). Either measuring convention will lead to difficulty at some point. If measurement is restricted to steadystate segments, we must suppose that perceptually the duration of a speech stretch is determined by the summed durations of the steadystate segments, and this clearly is untenable. If, on the other hand, we choose the component, then we should reckon with another possibility, namely, that variation in degree of encoding or coarticulation, with consequent change in total performance time, might not be perceived as a variation in length at all, provided there were no change in the durations of the individual components. As a possibility this latter is not so much implausible as a matter of unverifiable conjecture. But in the case where only steadystate segments are to be measured, it is difficult to believe that transitions make no contribution to length. It seems far more likely that they figure importantly in the length of vowels, for the degree of steadiness of an acoustic stretch associated with a vowel depends in part on the particular consonantal context, even where no length difference is perceived (see, for example, the spectrographic patterns in Öhman 1966:160–162 and in Liberman 1970a:309). Only if one believes that because transitional segments are generally more important for consonant than for vowel identification (Delattre 1964, 1968), they must therefore be entirely assigned to the consonant, can one suppose that they contribute nothing to the perceived duration of the vowel. With respect to duration the contrary is more likely true: so far as place and manner of articulation are concerned, the transition is primarily consonantal, but so far as length goes it is part of the vowel. Even so far as the duration of the transition is concerned, this serves not as a cue to the length of the consonant, but to its membership in one or another manner class of such segments (Liberman et al. 1956). If one may draw a conclusion from the published studies concerned with consonant duration (Denes 1955; Lisker 1957a, 1958; Liberman et al. 1961a; Fischer-Jørgensen 1964; Obrecht 1965), there is a general conviction that consonant length is tied exclusively to the duration of a steadystate acoustic segment.

The discussion so far has been exclusively of the temporal dimension as a segmental property, primarily because it is at this level that difficulties of the kinds mentioned are most apparent, difficulties that make it hard to assess the significance of a good deal of the numerical data available. The literature also includes studies of the temporal properties of entities both smaller and larger than the phonetic segment. Items of the first kind are, for example, the individual periods of the voice fundamental (Lieberman 1963) and the timing relation between stop articulation and voicing onset (Liberman et al. 1961b; Lisker and Abramson 1964, 1967, 1970; Abramson and Lisker 1965, 1970). Speech stretches longer than the segment for which temporal properties have been determined are the syllable, the word, the phrase, and the entire speech piece itself. The durations of these various units are of course not independent of one another, and the literature concerned with these larger units is devoted to describing, not their absolute durations, but their durations relative to the durations of units of a different, usually higher, order of magnitude. Thus, for example, Chistovich and her colleagues (1965:92–95) determined durations for selected syllables in relation to the durations of both the words and sentences in which they were contained, their purpose being to determine regularities from which to derive a model of the temporal organization of speech behavior generally. It might be said that their method of investigation effectively bypassed the segmentation problem, for they did not look for temporal relationships among segment durations, but rather went directly to certain accessible parts of the vocal tract to obtain recordings of the articulatory components involved in the production of their test utterances. They were then able to determine the temporal relations among these components of the articulatory complex. From their work and that of MacNeilage and DeClerk (1969:1233) the conclusion has been drawn that the syllable, or perhaps a 'basic syllable' consisting of a consonant and following vowel, is the unit of temporal organization of the articulatory gestures executed in producing what is ordinarily thought to be a sequence of segments. The various articulatory gestures are not adequately defined by membership in a class of segments, for their execution depends significantly on the particular combination of gestures which constitutes the given syllable (Chistovich et al. 1965:161). Work by these same authors, and by Daniloff and Moll (1968), demonstrates that the articulatory gestures attributed to a particular segment are not in general executed in phase, and that the interval of co-occurrence is that of a syllable or more. The syllable as a phonetic unit has long been controversial among linguists, and owes whatever recognition it enjoys with them to its establishment as a convenience in talking about phoneme distribution (Haugen 1956). Phoneticians, however, have recognized that certain segmental properties are associated with position within the syllable. Thus Abercrombie (1967:40) states that releasing or syllable-initial consonants are of very short duration, while syllable-final or arresting consonants are longer, and in presenting durational data phoneticians generally are careful to specify the location within the syllable of the segment in question (e.g. Malécot 1968). Failure to sort out segments by position within the syllable and in relation to pause can in fact lead to statements of dubious value, as for example those of Parmenter and Treviño (1935), on the relative length of voiced and voiceless stops in English.

One kind of evidence advanced in support of the syllable as a unit of temporal organization of speech behavior is that certain acoustic segments said to belong to the same syllable show durations that are negatively correlated (Chistovich et al. 1965:105). Thus it has been observed repeatedly that in English and some other languages vowels tend to be shorter before voiceless consonants than before voiced, and that at the same time the constriction durations of voiceless consonants are greater than those of their voiced counterparts (Laziczius 1961:120). Nor have convincing counter-examples to this generalization been attested. The relation between the durations of vowel and following voiced as against voiceless consonant is such that the total durations of the two kinds of sequences tend to approach equality, given a constant overall speaking rate. From data for English trochees produced in isolation (Sharf 1962:28) it would seem that the combination of a given vowel with either one of a homorganic pair of stops has a duration that is independent of the voicing characteristic of the stop. We might then be prepared to believe that the syllable, or at least the stressed syllable of English, has an inherent duration so far as the vowel and following consonant are concerned, one that varies only with change in overall speaking rate. This, however, is open to question. If we consider total durations of all vowel-consonant combinations it is certainly not true that these are independent of the particular vowel constituents, nor is it true of the consonants in any respect other than the voicing characteristic of the stop and possibly also the fricative consonants. From Sharf's data for English, as well as from Elert's for Swedish (Elert 1964:158–163), it appears that the variability in stop closure duration is not sufficient to make up for vowel differences ascribable to their inherent durations (Lehiste 1970:18–19). Moreover, if we consider vowel-consonant sequences containing consonants of diverse manners of articulation, we find that their components are not so correlated in duration as to yield a fixed syllable duration (Chistovich et al. 1965:111). Thus, given the generalization that low vowels are longer than high ones, and that fricatives are longer than stops (Laziczius 1961:120), it is not the case that a sequence of low vowel and fricative has a duration anywhere near equal to that of a sequence of a high vowel and a stop consonant. For this reason Chistovich and her colleagues conclude that if rhythmic patterning of speech activity is to be established, it must be on the basis of temporal regularities at a level higher than the syllable (Chistovich et al. 1965: 110).

The observed negative correlation between vowel and succeeding stop consonant that has been reported for a number of languages — e.g. English (Sharf 1962), French (Belasco 1958; Delattre 1962), Russian (Chistovich et al. 1965:99), Norwegian (Fintoft 1961:35) and Tamil (Lisker 1958) — has been the object of several kinds of speculative explanations. The generally greater closure durations reported for voiceless stops has most often referred to a feature of relative force of articulation, the fortis-lenis dimension. Thus it is claimed either that voiceless stops and fricatives involve a greater expenditure of articulatory energy and hence longer closures, or that fortis obstruents have as a consequence of their more forceful articulation longer closures that are usually voiceless (Jakobson, Fant and Halle 1952:26, 36, 38; Chomsky and Halle 1968:325). The vowel preceding a fortis stop is shorter than one before a lenis one because the speed of articulation is greater for the former (see, for example, Fujimura 1961), and consequently the vowel is abbreviated. Alternatively it has been supposed that there is a tendency for vowel-consonant sequences to be produced with a constant total expenditure of effort. If effort is linearly related to duration, then the greater effort required for the fortis stop calls for a reduction in vowel duration (Durand 1946:172; Belasco 1958). Still another explanation has been advanced for the greater duration of vowels before voiced stops, namely that in order to maintain voicing during the articulatory closure a special adjustment of the larynx must be completed before the closure, and vowel durations of the magnitude found before the voiceless stops do not allow sufficient time for this laryngeal maneuver (Chomsky and Halle 1968: 301). In view of the fact that vowels, in English at least, are as long or longer before other consonants for which no special adjustment of the larynx seems necessary for voicing, and that moreover the intrinsically long vowels may well be longer before voiceless stops than the intrinsically short ones before the voiced, this last way of accounting for durational differences between allophones of the same vowel phonemes in the two positions is far from convincing. Nor can it be said that attempts to explain durational differences among the consonants of differing manner of articulation merit any more immediate acceptance.

Of speech stretches greater than the syllable, comparatively little information is available until we get to the speech piece as a whole. For certain languages, English for example, it is said that there is a more or less regular alternation of prominent syllables and stretches of one or more syllables of comparatively low audibility. These are the so-called 'stress-timed' languages (Abercrombie 1967:97–98). By contrast, the 'syllable-timed' languages presumably exhibit no temporal organization of syllables in larger rhythmic units (Abercrombie 1967:97). In languages of the first type, the durations of syllables and their component segments that fall between two prominent syllables depend considerably on how many of them there are in the particular piece. The total duration of the piece also appears to determine strongly the durations of component segments and/or syllables. Thus Fónagy and Magdics (1960:192) report that short phrases in Hungarian are produced at a slower rate than are phrases consisting of eight or more segments (*op. cit.* p. 182). Extensive data for English have been presented by Goldman-Eisler (1954, 1956, 1961, 1967, 1968) to show that speech rate, defined as the number of syllables per minute, may vary over an extremely wide range for pieces of less than about 100 syllables, becoming stable for longer pieces at a rate of about 225 syll./min., though with considerable intersubject variation. She determined, however, that the degree of variability for short pieces, from about 60 to 600 syll./min., was not one in speed of articulation, but one in the amount of pausing within the piece (Goldman-Eisler 1968:26). On the other hand Laziczius (1961:123) cites durational values for a particular Hungarian vowel in a series of successively longer (in number of syllables) words, and the vowel undergoes progressive shortening. This effect is perhaps in part to be explained as a junctural effect, since proximity to pause seems regularly to be accompanied by an increase in segment duration as part of a reduction in the rate of higher-level units (Lehiste 1960; Hoard 1966).[9]

The durations of the various units into which the speech piece may be divided are more or less mutually dependent, and consequently the precise significance of absolute time determinations is in general subject to question. The nature of the

[9] It is difficult to believe, however, that uninterrupted pieces are not sometimes judged to have been produced at different speaking rates. There is no certainty, if indeed this is true, as to whether a measure of phonetic segments, syllables, or perhaps words per unit time would best match listeners' judgments of relative rate. Experiments by Osser and Peng (1964) in which phonetic segments per minute were determined for English and Japanese, failed to turn up a significant difference between speakers of the two languages. Since each group judged the other to consist of rapid talkers, the meaning of the negative finding is not clear, nor would a discovered difference be easy to interpret.

temporal dependencies, both among units of different levels and coordinate units, is of much greater significance for the establishment of temporal regularities in speech behavior. Exceptions to this are to be found, however, in studies aimed at determining particular time constants of articulation or perception (Lehiste 1970: 6–9). Duration is often to be correlated with the perceptual dimension of temporal length, but by no means invariably. As an attribute of the acoustic segment it is sometimes interpreted as length (Abramson 1962; Obrecht 1965), sometimes as a cue to consonantal manner of articulation (Gerstman 1956, 1957; Liberman et al. 1956), force of articulation (Chomsky and Halle 1968:324), voicing state (Denes 1955; Lisker 1957a, 1958; Sharf 1962), stress (Fry 1968:370), and perhaps speech rate (Huggins 1968). At the level of the syllable little data on duration have been collected, apart from segmental measurements involving the comparison of syllables of differing degrees of prominence. The syllable and the word figure often, however, in discussions of speech rate, the determination of which requires only the decision as to how many of one or the other units are contained in a speech sample, and a determination of the total duration of the sample. But with the discussion of speech rate the focus of interest tends to become the individual speaker, and thus peripheral to the central concerns of phonetic research.

REFERENCES

ABERCROMBIE, DAVID. 1964. Syllable quantity and enclitics in English. In honour of Daniel Jones, pp. 216–22. London, Longmans.

——. 1967. Elements of general phonetics. Chicago, Aldine.

ABRAMSON, ARTHUR S. 1962. The vowels and tones of standard Thai: Acoustical measurements and experiments. RCAFL-P 20.

ABRAMSON, ARTHUR S., and L. LISKER. 1965. Voice onset time in stop consonants: Acoustic analysis and synthesis. 5th International Congress on Acoustics, Liège, ed. by D. E. Commins, A51. Liège, Imp. G. Thone.

——. 1970. Discriminability along the voicing continuum: Cross-language tests. PICPS 6.569–73.

AINSWORTH, W. A. 1968. Perception of stop consonants in synthetic CV syllables. L&S 11.139–155.

AVAKJAN, R. V. 1968. Mimicking the duration of the vowels in aphasia. ZPhon 21.190–93.

BAUERNSCHMIDT, AMY. 1965. Amuzgo syllable dynamics. Lg 41.471–83.

BELASCO, SIMON. 1958. Variations in vowel duration: Phonemically or phonetically conditioned? JAcS 30.1049–50.

BERNARD, J. R. L-B. 1970. On nucleus component durations. L&S 13.89–101.

BLACK, JOHN W. 1961. Relationships among fundamental frequency, vocal sound pressure, and rate of speaking. L&S 4.196–99.

BLOOMFIELD, LEONARD. 1933. Language. New York, Henry Holt, BOLINGER, DWIGHT L., and L. J. GERSTMAN. 1957. Disjuncture as a cue to constructs. Word 13.246–55.

BRICKER, PETER D., and S. PRUZANSKY. 1966. Effects of stimulus content and duration on talker identification. JAcS 40.1441–49.

BROADBENT, D. E., and P. LADEFOGED. 1959. Auditory perception of temporal order. JAcS 31.1539. (A)

BROOKS, MARIA ZAGÓRSKA. 1964. On Polish affricates. Word 20.207–10.

CATFORD, IAN C. 1968. The articulatory possibilities of man. Manual of phonetics, ed. by Bertil Malmberg, pp. 309–33. Amsterdam, North-Holland.

CHISTOVICH, LYUDMILA A., V. A. KOZHEVNIKOV, V. V. ALYAKRINSKIY, L. V. BONDARKO, A. G. GOLUZINA, YU. A. KLAAS, YU. I. KUZ'MIN, D. M. LISENKO, V. V. LYUBLINSKAYA, N. A. FEDOROVA, V. S. SHUPLYAKOV, and R. M. SHUPLYAKOVA. 1965. *In* Rech': Artikulyatsiya i vospriyatiye, ed. by V. A. Kozhevnikov and L. A. Chistovich. Moscow and Leningrad, Nauka. Trans. as Speech, articulation and perception. Washington, D.C., Clearinghouse for Federal Scientific and Technical Information, JPRS 30.543.

CHOMSKY, NOAM, and M. HALLE. 1968. The sound pattern of English. New York, Harper and Row.

COHEN, A., J. F. SCHOUTEN, and J. 'T HART. 1962. Contribution of the time parameter to the perception of speech. PICPS 4.555–60.

COMPTON, ARTHUR J. 1963. Effects of filtering and vocal duration upon the identification of speakers, aurally. JAcS 35.1748–52.

COOPER, FRANKLIN S. 1950. Research on reading machines for the blind. Blindness: Modern approaches to the unseen environment, ed. by Paul A. Zahl, pp. 512–43. Princeton, N.J., Princeton University Press.

CREELMAN, C. D. 1962. Human discrimination of auditory duration. JAcS 34.582–93.

DANILOFF, RAYMOND, and K. MOLL. 1968. Coarticulation of lip rounding. JSHR 11.707–21.

DAY, RUTH S. 1970. Temporal order perception of a reversible phoneme cluster. JAcS 48.95. (A)

DELATTRE, PIERRE. 1962. Some factors of vowel duration and their cross-linguistic validity. JAcS 34.1141–43. (L)

——. 1964. Change as a correlate of the vowel-consonant distinction. SL 18.12–25.

——. 1968. From acoustic cues to distinctive features. Phonetica 18.198–230.

DELATTRE, PIERRE, and M. HOHENBERG. 1968. Duration as a cue to the tense/lax distinction in German unstressed vowels. IRAL 6/4.367–90.

DELATTRE, PIERRE, and M. MONNOT. 1968. The role of duration in the identification of French nasal vowels. IRAL 6/3.267–88.

DENES, PETER B. 1955. Effect of duration on the perception of voicing. JAcS 27.761–64.

DENES, PETER B., and E. N. PINSON. 1963. The speech chain. Bell Telephone Laboratories, Inc. Baltimore, Waverly Press.

DURAND, MARGUÉRITE. 1946. Voyelles longues et voyelles brèves: Essai sur la nature de la quantité vocalique. Paris, C. Klincksieck.

EEK, ARVO. 1970a. Articulation of the Estonian sonorant consonants, I. *n* and *l*. ETAT 19.103–21.

——. 1970b. Some coarticulation effects in Estonian. Academy of Sciences of the Estonian S. S. R. 6.81–85.

ELERT, CLAES-CHRISTIAN. 1964. Phonologic studies of quantity in Swedish. Stockholm, Almqvist and Wiksell.

ENDRES, WERNER K. 1968. On the experimental evaluation of the interior redundancy in speech. The 6th International Congress on Acoustics, Tokyo. B111–B114.

ESSEN, OTTO VON. 1957. Überlange Vokale und gedehnte Konsonanten des Hochdeutschen. ZPhon 10.239–44.

FAIRBANKS, GRANT, N. GUTTMAN, and M. S. MIRON. 1957. Effects of time compression upon the comprehension of connected speech. JSHD 22.10–19.

FAIRBANKS, GRANT, and L. W. HOAGLIN. 1941. An experimental study of the durational characteristics of the voice during the expression of emotion. SpMon 8.85–90.

FAIRBANKS, GRANT, and F. KODMAN, JR. 1957. Word intelligibility as a function of time compression. JAcS 29.636–41.

FANT, C. GUNNAR M. 1960. Acoustic theory of speech production. The Hague, Mouton.

——. 1962. Sound spectrography. PICPS 4.14–33.

FANT, C. GUNNAR M., J. LILJENCRANTS, V. MALÁČ, and B. BOROVICKOVA. 1970. Perceptual evaluation of coarticulation effects. STL-QPSR 1/1970, 10–13.

FAY, WARREN H. 1966. Temporal sequence in the perception of speech. The Hague, Mouton.

FINTOFT, KNUT. 1961. The duration of some Norwegian speech sounds. Phonetica 7.19–39.

FISCHER-JØRGENSEN, ELI. 1954. Acoustic analysis of stop consonants. Miscellanea Phonetica 2.42–51, ed. by A. C. Gimson. London, International Phonetic Association.

——. 1963. Beobachtungen über den Zusammenhang zwischen Stimmhaftigkeit und intraoralem Luftdruck. ZPhon 16.19–36.

——. 1964. Sound duration and place of articulation. ZPhon 17.175–207.

——. 1969. Voicing, tenseness and aspiration in stop consonants, with special reference to French and Danish. Annual Report III. Inst. Phon. Univ. Copenhagen, pp. 63–114.

FÓNAGY, IVAN, and J. FÓNAGY. 1966. Sound pressure level and duration. Phonetica 15.14–21.

FÓNAGY, IVAN, and K. MAGDICS. 1960. Speed of utterance in phrases of different lengths. L&S 4.179–92.

FRY, DENNIS B. 1955. Duration and intensity as physical correlates of linguistic stress. JAcS 27.765–68.

——. 1968. Prosodic phenomena. Manual of phonetics, ed. by Bertil Malmberg, pp. 365–410. Amsterdam, North-Holland.

FUJIMURA, OSAMU. 1961. Bilabial stop and nasal consonants: A motion picture study and its acoustical implications. JSHR 4.233–47.

GAITENBY, JANE H. 1965. The elastic word. Status Report on Speech Research 2–3.1–3.12. New York, Haskins Laboratories.

GAY, THOMAS. 1968. Effect of speaking rate on diphthong formant movements. JAcS 44.1570–73.

——. 1970. A perceptual study of American English diphthongs. L&S 13.65–88.

GERSTMAN, LOUIS J. 1956. Noise duration as a cue for distinguishing among fricative, affricate and stop consonants. JAcS 28.160. (A)

——. 1957. Perceptual dimensions for the fricative portion of certain speech sounds. Doctoral dissertation. New York University.

GIMSON, A. C. 1962. An introduction to the pronunciation of English. London, Edward Arnold.

GLEASON, H. A., JR. 1961. An introduction to descriptive linguistics. (Revised edition.) New York, Holt, Rinehart and Winston.

GOLDMAN-EISLER, FRIEDA. 1954. On the variability of the speed of talking and its relation to the length of utterances in conversations. BrJPsych 45.94–107.

——. 1956. The determinants of the rate of speech output and their mutual relations. Journal of Psychosomatic Research 1.137–43.

——. 1961. The significance of changes in the rate of articulation. L&S 4.171–74.

——. 1967. Sequential temporal patterns and cognitive processes in speech. L&S 10.122–32.

——. 1968. Psycholinguistics: Experiments in spontaneous speech. London and New York, Academic.

HADDING-KOCH, KERSTIN, and A. S. ABRAMSON. 1964. Duration versus spectrum in Swedish vowels: Some perceptual experiments. SL 18.94–107.

HANHARDT, ARTHUR M., D. H. OBRECHT, W. R. BABCOCK, and J. B. DELACK. 1965. A spectrographic investigation of the structural status of *Ueberlaenge* in German vowels. L&S 8.214–18.

HARRELL, RICHARD S. 1958. Some English nasal articulations. Lg 34.492–93.

HARRIS, CYRIL M. 1953. A study of the building blocks in speech. JAcS 25.962–69.

HARRIS, KATHERINE S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. L&S 1.1–7.

——. 1974. Physiological aspects of articulatory behavior (see this volume).

HAUGEN, EINAR. 1956. The syllable in linguistic description. For Roman Jakobson, pp. 213–21. The Hague, Mouton.

HEFFNER, R-M. S. 1950. General phonetics. Madison, University of Wisconsin.

HENDERSON, A., F. GOLDMAN-EISLER, and A. SKARBEK. 1966. Sequential temporal patterns in spontaneous speech. L&S 9.207–16.

HENRY, F. 1948. Discrimination of the duration of a sound. JExPsych 38.734–43.

HIRSCH, IRA. 1959. Auditory perception of temporal order. JAcS 31.759–67.

HOARD, J. E. 1966. Juncture and syllable structure in English. Phonetica 15.96–109.

HOCKETT, CHARLES F. 1955. A manual of phonology. IJAL Memoir 11.

——. 1958. A course in modern linguistics. New York, Macmillan.

HOFMANN, T. R. 1967. Initial clusters in English. MIT-QPR 84.263–74.

HOUSE, ARTHUR S. 1961. On vowel duration in English. JAcS 33.1174–78.

HOUSE, ARTHUR S., and G. FAIRBANKS. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. JAcS 25.105–13.

HUGGINS, A. W. F. 1964. Distortion of the temporal pattern of speech: Interruption and alternation. JAcS 36.1055–64.

——. 1968. The perception of timing in natural speech I: Compensation within the syllable. L&S 11.1–11.

JAKOBSON, ROMAN, C. G. M. FANT, and M. HALLE. 1952. Preliminaries to speech analysis: The distinctive features and their correlates. Technical Report No. 13, Acoustics Laboratory. Cambridge, Mass., M.I.T.

JASSEM, WIKTOR, J. MORTON, and M. STEFFEN-BATÓG. 1968. The perception of stress in synthetic speech-like stimuli by Polish listeners. Speech analysis and synthesis I. 289–308. Warsaw, Państwowe Wydawnictwo Naukowe.

JESPERSEN, OTTO. 1932. Lehrbuch der Phonetik. (5th ed.) Leipzig and Berlin, B. G. Teubner.

JOHANSSON, KURT. 1969. Transitions as place cues for voiced stop consonants: Direction or extent? SL 23.69–82.

JOHN, J. E. J., and J. N. HOWARTH. 1965. The effect of time distortions on the intelligibility of deaf children's speech. L&S 8.127–34.

JONES, DANIEL. 1950. The phoneme: Its nature and use. Cambridge, W. Heffers and Sons.

KANTNER, C. E., and R. WEST. 1960. Phonetics. (Revised ed.) New York, Harper and Brothers.

KENT, R. D., and K. L. MOLL. 1969. Vocal-tract characteristics of the stop consonants. JAcS 46.1549–55.

KIM, CHIN-WU. 1970. A theory of aspiration. Phonetica 21.107–16.

KLUMPP, R. G., and J. C. WEBSTER. 1961. Intelligibility of time-compressed speech. JAcS 33.265–67.

LADEFOGED, PETER. 1964. A phonetic study of West African languages: An auditory-instrumental survey. West African Language Monographs 1. Cambridge, Cambridge University.

——. 1968. Linguistic aspects of respiratory phenomena. Proceedings of the Conference on Sound Production in Man, Nov. 1966, pp. 141–51. New York, New York Academy of Sciences.

LAZICZIUS, JULIUS. 1961. Lehrbuch der Phonetik. Berlin, Akademie Verlag.

LEHISTE, ILSE. 1960. An acoustic-phonetic study of internal open juncture. Suppl. to Phonetica 5.1–54.

——. 1970. Suprasegmentals. Cambridge, Mass., M.I.T. Press.

LEHISTE, ILSE, and G. E. PETERSON. 1961. Transitions, glides and diphthongs. JAcS 33.268–77.

LENNEBERG, ERIC H. 1967. Biological foundations of language. New York, Wiley.

LIBERMAN, ALVIN M. 1970a. The grammars of speech and language. CogPsych 1.301–23.

——. 1970b. Some characteristics of perception in the speech mode. Ch. 16 of Perception and its disorders, research publication of the Assoc. Research Nerv. Mental Dis. 48.

LIBERMAN, ALVIN M., F. S. COOPER, D. P. SHANKWEILER, and M. STUDDERT-KENNEDY. 1967. Perception of the speech code. PsychRev 74.431–61.

LIBERMAN, ALVIN M., F. S. COOPER, M. STUDDERT-KENNEDY, K. S. HARRIS, and D.P. SHANKWEILER. 1968. On the efficiency of speech sounds. ZPhon 21.21–32.

LIBERMAN, ALVIN M., P. DELATTRE, L. J. GERSTMAN, and F. S. COOPER. 1956. Tempo of frequency change as a cue for distinguishing classes of speech sounds. AJExpPsych 52.127–37.

LIBERMAN, ALVIN M., K. S. HARRIS, P. EIMAS, L. LISKER, and J. BASTIAN. 1961. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. L&S 4.175–95.

LIBERMAN, ALVIN M., K. S. HARRIS, J. A. KINNEY, and H. LANE. 1961. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. JExPsych 61.379–88.

LIBERMAN, ALVIN M., F. INGEMANN, L. LISKER, P. DELATTRE, and F. S. COOPER. 1959. Minimal rules for synthesizing speech. JAcS 31.1490–99.

LIEBERMAN, PHILIP. 1960. Some acoustic correlates of word stress in American English. JAcS 32.451–54.

——. 1963. Some acoustic measures of the fundamental periodicity of normal and pathological larynges. JAcS 35.344–53.

LIIV, GEORG. 1961. Estonian vowels of three degrees of length. Institute of Language and Literature of the Academy of Sciences, Estonian S.S.R. 1–8.

——. 1962. On the acoustic composition of Estonian vowels of three degrees of length. ETAT 1962/1.271–90.

LINDBLOM, BJÖRN E. F. 1963. Spectrographic study of vowel reduction. JAcS 35.1773–81.

——. 1967. Vowel duration and a model of lip mandible coordination. STL-QPSR 4.1–29.

LINDBLOM, BJÖRN, E. F., and M. STUDDERT-KENNEDY. 1967. On the role of formant transitions in vowel recognition. JAcS 42.830–43.

LINDNER, GERHART. 1969. Einführung in die experimentelle Phonetik. München, Max Hueber Verlag.

LISKER, LEIGH. 1957a. Closure duration and the intervocalic voiced-voiceless distinction in English. Lg 33.42–49.

——. 1957b. Linguistic segments, acoustic segments, and synthetic speech. Lg 33.370–74.

——. 1958. The Tamil occlusives: Short vs. long or voiced vs. voiceless? IL 19, Turner Jubilee I. 294–301.

——. 1972. Stop duration and voicing in English. Papers on linguistics and phonetics in memory of Pierre Delattre, ed. by Albert Valdman. The Hague, Mouton.

LISKER, LEIGH, and A. S. ABRAMSON. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20.384–422.

——. 1967. Some effects of context on voice onset time in English stops. L&S 10.1–28.

——. 1970. The voicing dimension: Some experiments in comparative phonetics. PICPS 6.563–67.

LISKER, LEIGH, F. S. COOPER, and A. M. LIBERMAN. 1962. The uses of experiment in language description. Word 18.82–106.

LOWE, AUDREY D., and R. A. CAMPBELL. 1965. Temporal discrimination in aphasoid and normal children. JSHR 8.313–14.

LUBKER, JAMES F., and P. J. PARRIS. 1970. Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. JAcS 47.625–33.

LÜDTKE, H. 1969. Die Alphabetschrift und das Problem der Lautsegmentierung. Phonetica 20.147–76.

——. 1970. Sprache als kybernetisches Phänomen. Bibliotheca Phonetica 9.34–50. Basel and New York, S. Karger.

MACNEILAGE, PETER F. 1963. Electromyographic and acoustic study of the production of certain final clusters. JAcS 35.461–63.

——. 1970. Motor control of serial ordering in speech. PsychRev 77.182–96.

MACNEILAGE, PETER F., and J. L. DECLERK. 1969. On the motor control of coarticulation in CVC monosyllables. JAcS 45.1217–33.

MACNEILAGE, PETER F., T. P. ROOTES, and R. A. CHASE. 1967. Speech production and perception in a patient with severe impairment of somesthetic perception and motor control. JSHR 10.449–67.

MALÉCOT, ANDRÉ. 1968. The force of articulation of American stops and fricatives as a function of position. Phonetica 18.95–102.

——. 1969. The effect of syllabic rate and loudness on the force of articulation of American stops and fricatives. Phonetica 19.205–16.

——. 1970. The lenis-fortis opposition: Its physiological parameters. JAcS 47.1588–92.

MALÉCOT, ANDRÉ, and P. LLOYD. 1968. The /t/:/d/ distinction in American alveolar flaps. Lingua 19.264–72.

MALMBERG, BERTIL. 1963. Structural linguistics and human communication. New York, Academic Press.

MARTIN, JAMES G., and W. STRANGE. 1968. The perception of hesitation in spontaneous speech. P&P 3.427–38.

MATTINGLY, IGNATIUS G. 1966. Synthesis by rule of prosodic features. L&S 9.1–13.

MATTINGLY, IGNATIUS G., and A. M. LIBERMAN. 1970. The speech code and the physiology of language. Information processing in the nervous system, ed. by K. N. Leibovic, pp. 97–117. New York, Springer.

MORTON, JOHN, and W. JASSEM. 1965. Acoustic correlates of stress. L&S 8.159–81.

NASR, R. T. 1960. Phonemic length in Lebanese Arabic. Phonetica 5.209–11.

OBRECHT, DEAN H. 1965. Three experiments in the perception of geminate consonants in Arabic. L&S 8.31–41.

ÖHMAN, SVEN E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. JAcS 39.151–68.

——. 1967. Numerical model of coarticulation. JAcS 41.310–20.

OSSER, HARRY, and F. PENG. 1964. A cross cultural study of speech rate. L&S 7.120–25.

PARMENTER, C. E., and S. N. TREVIÑO. 1935. The length of the sounds of a Middle Westerner. AS 10.129–33.

PETERSON, GORDON E., and I. LEHISTE. 1960. Duration of syllable nuclei in English. JAcS 32.693–703.

PETERSON, GORDON E., and J. E. SHOUP. 1966. Glossary of terms from the physiological and acoustic phonetic theories. JSHR 9.100–20.

PICKETT, JAMES M. 1965. Some acoustic cues for synthesis of the /n,d/ distinction. JAcS 38.474–77.

PICKETT, JAMES M., and I. POLLACK. 1963. Adjacent context and the intelligibility of words excised from fluent speech. JAcS 35.807. (A)

PIKE, KENNETH L. 1943. Phonetics. Ann Arbor, University of Michigan Press.

RUHM, HOWARD B., E. O. MENCKE, B. MILBURN, W. A. COOPER, JR., and D. E. ROSE. 1966. Differential sensitivity to duration of acoustic signals. JSHR 9.371–84.

SAWASHIMA, MASAYUKI. 1974. Laryngeal research in experimental phonetics (this volume).

SCHOUTEN, J. F., A. COHEN, and J. 'T HART. 1962. Study of time cues in speech perception. JAcS 34.517–18.

SCHWARTZ, MARTIN F. 1967. Syllable duration in oral and whispered reading. JAcS 41.1367–69. (L)

SCOTT, ROBERT J. 1967. Time adjustment in speech synthesis. JAcS 41.60–65.

SHANKWEILER, DONALD, K. S. HARRIS, and M. L. TAYLOR. 1968. Electromyographic studies of articulation in aphasia. Archives of Physical Medicine and Rehabilitation 49.1–8. (American Congress of Rehabilitation Medicine, Chicago, Ill.)

SHARF, DONALD J. 1962. Duration of post-stress intervocalic stops and preceding vowels. L&S 5.26–30.

——. 1964. Vowel duration in whispered and in normal speech. L&S 7.89–97.

SINGH, SADANAND, and J. W. BLACK. 1965. A study of nonsense syllables spoken by two language groups in varying conditions of sidetone and reading rate. L&S 8.208–13.

DESIVERS, FANNY. 1964. A qualitative aspect of distinctive quantity in Estonian. Word 20.28–34.

SLIS, I. H. 1970. Articulatory measurements on voiced, voiceless and nasal consonants: A test of a model. Phonetica 21.193–210.

SLIS, I. H., and A. COHEN. 1969. On the complex regulating the voiced-voiceless distinction II. L&S 12.137–55.

SODA, TOYOJI, Y. NISHEDA, and H. SUWOYA. 1967. Intraoral pressure change in Japanese consonants. Jibi to Rinsho, = Otologia Fukuoka 13, suppl. 1.34–43. Fukuoka.

STETSON, R. H. 1951. Motor phonetics. Amsterdam.

STEVENS, KENNETH N., A. S. HOUSE, and A. P. PAUL. 1966. Acoustical description of syllabic nuclei: An interpretation in terms of a dynamic model of articulation. JAcS 40.123–32.

STICHT, THOMAS G., and B. B. GRAY. 1969. The intelligibility of time compressed words as a function of age and hearing loss. JSHR 12.443–48.

STUDDERT-KENNEDY, MICHAEL, and A. M. LIBERMAN. 1963. Psychological considerations in the design of auditory displays for reading machines. Proc. Intern. Congr. on Techn. and Blindness, pp. 289–304.

SUBTELNY, JOANNE D., and J. H. WORTH. 1966. Intraoral pressure and rate of flow during speech. JSHR 9.498–518.

SUZUKI, H. 1970. Mutually complementary effect of rate and amount of formant transition in distinguishing vowel, semivowel, and stop consonant. MIT-QPR

96.164–72.

'T HART, J., and A. COHEN. 1964. Gating techniques as an aid in speech analysis. L&S 7.22–39.

THOMAS, IAN B., P. B. HILL, F. S. CARROLL, and B. GARCIA. 1970. Temporal order in the perception of vowels. JAcS 48.1010–13.

VIEREGGE, WILHELM H. 1969. Untersuchungen zur akustischen Struktur der Plosivlaute. (doct. diss.) Binn, Rheinische Friederich-Wilhelms-Universität.

WARREN, DONALD W., and S. B. MACKLER. 1968. Duration of oral port constriction in normal and cleft palate speech. JSHR 11.391–401.

WARREN, RICHARD M. 1968. Verbal transformation effect and auditory perceptual mechanisms. PsychRev 70.261–70.

——. 1970. Perceptual restoration of missing speech sounds. Science 167.392–93.

WARREN, RICHARD M., C. J. OBUSEK, and R. M. FARMER. 1969. Auditory sequence: Confusion of patterns other than speech or music. Science 164.586–87.

WESTIN, KJELL, R. G. BUDDENHAGEN, and D. H. OBRECHT. 1966. An experimental analysis of the relative importance of pitch, quantity, and intensity as cues to phonemic distinctions in southern Swedish. L&S 9.114–26.

WICKELGREN, WAYNE A. 1969. Context-sensitive coding, associative memory, and serial order in (speech) behavior. PsychRev 76.1–15.

YANAGIHARA, NAOAKI, and C. HYDE. 1966. An aerodynamic study of the articulatory mechanism in the production of bilabial stop consonants. SPhon 4. 70–80.

ZEMLIN, WILLARD R., R. G. DANILOFF, and T. H. SHRINER. 1968. The difficulty of listening to time-compressed speech. JSHR 11.875–81.

ZIMMERMAN, SAMUEL A., and S. M. SAPON. 1958. Note on vowel duration seen cross-linguistically. JAcS 30.152–53.

ZWIRNER, EBERHARD, and K. ZWIRNER. 1966. Grundfragen der Phonometrie. (2nd ed.) Bibliotheca Phonetica. Basel and New York, S. Karger.