Some Observations on the Efficiency of Speech Sounds

A.M. Liberman*, F.S. Cooper, M. Studdert-Kennedy**, K.S. Harris, and D.P. Shankweiler

Haskins Laboratories, New York City

Our concern is with the question: how is it that on hearing the sounds of speech a listener perceives phonemes? Since the question is reasonable only if we assume that phonemes *are* perceived, we should say that our purpose here is not to justify that assumption(1); rather, we accept it and go on to ask how such perception might occur. But first we should say why we find the question a particularly interesting one. Then we will offer a brief outline of an answer.

There are two ways to see why the question is interesting. One is to note the several indications that phonemes could not be communicated by a simple cipher or sound alphabet. The other is to appreciate that phonemes are not, in fact, communicated in that simple way, but are rather recovered from a special code.

It seems clear that the ear is not well suited to the transmission of phoneme segments by means of an acoustic alphabet. We should consider, first, that speech can be perceived at rates

1. There is a considerable weight of evidence for the psychological reality of the phoneme; the ingenious experiments described by Kozhevnikov and Chistovich [Kozhevnikov, V.A. and Chistovich, L.A. *Rech' Artikuliatsia i Vospriiatie* (Moskva-Leningrad,1965)] provide another and recent example of such evidence. We are not assuming that the phoneme is always perceived, or that speech perception is always phonemic, only that the phoneme can be perceived and often must be perceived. The term "phoneme" is used here both in a linguistic sense and to denote the perceptual unit that is the nearest counterpart of the linguistic entity.

that require the listener to take in as many as 20 such segments per second. Given what we know of the temporal resolving power of the ear, we should suppose that 20 acoustic segments per second would merge, perceptually, into an unanalyzable buzz.(2) We might also anticipate difficulty in finding as many identifiable acoustic shapes as we need. There are approximately 40 phonemes in English, more in some other languages. The literature on auditory perception does not encourage the belief that it would be possible to find 40 or more highly identifiable acoustic signals of short duration that could be made to stand for those phonemes in a simple sound alphabet.(3)

We should also consider the more direct evidence that a simple cipher or sound alphabet on the phonemes is not likely to work. This evidence, by now considerable, comes from experience with non-speech ciphers on the language, including not only the familiar case of Morse code, but also the results of more than fifty years of research and development in the attempt to build reading machines for the blind. We all know how bad Morse code is, even after years of practice. The history of reading machines for the blind is even more revealing.(4) The difficulty there has not been to transform print into sound, but to find a set of non-speech sounds that can be

2. R.H. Stetson. Motor Phonetics. 2nd Edition. North Holland Publishing Co (Amsterdam, The Netherlands, 1951); M. Studdert-Kennedy and A.M. Liberman. Psychological considerations in the design of auditory displays for reading machines. Proc. of the International Congress on Technology and Blindness, I, 289-304 American Foundation for the Blind, (New York, 1963).

3. I. Pollack. The information of elementary auditory displays. J. Acoust. Soc. Amer., 24, 745-749 (1952); I.Pollack and L. Ficks. J. Acoust Soc. Amer., 26, 155-158 (1954); G.A. Miller. The magical number seven, plus-or-minus two, or, some limits on our capacity for processing information. Psychol. Rev. 63, 81-96 (1956); P.W. Nye. Aural recognition time for multidimensional signals. Nature, 196, 1282-1283 (London, 1962).

4. F.S. Cooper. Research on reading machines for the blind. In

identified rapidly and accurately. Many people have practiced for months, years, and even decades with a variety of sounds, yet top speeds are not better than can be achieved with Morse code and are less than a tenth of the rate we can attain with speech.

We should not be surprised, then, to discover that the sounds of speech are not an alphabet or substitution cipher on the phonemes, but a special and, compared with the non-speech ciphers mentioned above, highly efficient code.(5) In this brief paper we cannot even summarize all that is now known about the sounds of speech. We can only point to some typical findings and their implications.

The primary data come from experiments on the acoustic basis for speech perception. These data lead to at least two general conclusions indicating that speech *is* a code. One is that speech sounds are not segmented at the phoneme level. That is to say, there is no way to cut the acoustic signal along the time dimension so as to recover segments that will be perceived as separate phonemes: the acoustic representations of the phonemes overlap and intermix in units of approximately syllabic size. The relevant facts have emerged over the past fifteen

Blindness, P.A. Zahl, Ed. (Princeton, 1950); J. Freiberger and E.F. Murphy Reading machines for the blind. *IRE Professional Group on Human Factors in Electronics* (March, 1961); J.L. Coffey. The development and evaluation of the Battell Aural Reading Device. *Proceedings of the International Congress on Technology and Blindness*, *I*, 343-360. American Foundation for the Blind (New York, 1963); P.W. Nye. Reading aids for blind people -- a survey of progress with the technological and human problems. *Med.Electron.Biol.Engng.*, *2*, 247-264 (1964); P.W. Nye An investigation of audio outputs for a reading machine. National Physical Laboratory (Autonomics Division), Teddington, England (February, 1965).

[5] The term "code," as used here in contrast with "cipher," implies the restructuring of a message so that several units of the original are replaced or represented by a single unit of the encoded message, e.g., several successive phonemes by a single syllable of the spoken utterance.

years in many publications.(6) Among the more interesting treatments of this matter is that contained in the recent monograph by Kozhevnikov and Chistovich.(7)

The encoding of the phonemes into syllables is surely relevant to the unique distinctiveness or efficiency of speech, since it reduces by a factor of three or four the number of discrete acoustic segments that must be perceived per unit time. But it also poses an important question: how does the listener decode the signal and recover the phoneme segments?

The other general and related conclusion is that for many phonemes, including in particular those consonants that seem to carry the heaviest information load, there is no way to define the acoustic cues so as to have an invariant relation with the phoneme or with phoneme perception: the acoustic cue for the same phoneme (as perceived) is often vastly different in different contexts.(8) Moreover, as Lindblom has recently shown, the

6. C.M. Harris. A study of the building blocks of speech. J. Acoust. Soc. Amer., 25, 962-969 (1953); G. Peterson, W. Wang and E. Sivertsen. Segmentation techniques in speech synthesis. J. Acoust. Soc. Amer. 30, 739-742 (1958); and A.M. Liberman, F. Ingemann, L. Lisker, P. Delattre, and F.S. Cooper. Minimal rules for synthesizing speech. J. Acoust. Soc. Amer., 31, 1490-1499 (1959).

7. V.A. Kozhevnikov and L.A. Chistovich. Op. cit.

8. A.M. Liberman, P. Delattre, and F.S. Cooper. The role of selected stimulus variables in the perception of the unvoiced stop consonants. Am.J.Psychol., 65, 497-516 (1952); A.M. Liberman, P. Delattre, F.S. Cooper and L. Gerstman. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs, 68, No.8, 1-13 (1955); A.M. Liberman. Some results of research on speech perception. J.Acoust. Soc.Amer. 29, 117-123 (1957); F.S. Cooper, A.M. Liberman, K.S. Harris, and P.M Grubb. Some input-output relations observed in experiments on the perception of speech. Proc. of 2nd Intl.Cong. of Cybernetics, 930-941, Namur, Belgium (1958); K.N. Stevens Toward a model for speech recognition. J. Acoust. Soc. Amer., 32, 47-55 (1960); L.Lisker, F.S. Cooper, and A.M Liberman. The uses of experiment in language description. Word, 18, 82-106 (1962); A.M Liberman, F.S. Cooper, K.S. Harris, P.F. MacNeilage, and

invariance problem may be further complicated by variations in rate of articulation.(9) How the general lack of invariance relates to the distinctiveness of speech, and what it implies with regard to the decoding mechanism, will be considered later.

We should remark, incidentally, that the encoded nature of speech may not be limited to the segmental phonemes. There is evidence in the work of Lindblom(10), Hadding-Koch and Studdert-Kennedy(11), and Lieberman(12) that stress and intonation may present similarly complex relations between acoustic cue and perceived language. The abstracts of Lieberman(13) and Öhman (14) suggest that some aspects of this are to be presented at this symposium.

To appreciate that the speech code is special in some interesting sense, we may also consider the evidence for the existence of at least two modes of auditory perception, speech and non-speech. In several different kinds of experiments it has been found that the same or similar acoustic stimuli are

M. Studdert-Kennedy. Some observations on a model for speech perception. *Proc. AFCRL Symposium on Models for the Perception of Speech and Visual Form*. AFCRL (In press).

9. B. Lindblom. Spectrographic study of vowel reduction. *J. Acoust. Soc. Amer.*, *35*, 1773-1781 (1963).

10.__________. Personal communication.

11.K. Hadding-Koch and M. Studdert-Kennedy. An experimental study of some intonation contours. *Phonetica*, *11*, 175-185 (1964).

12.P. Lieberman. Intonation and syntactic processing of speech. *Proc. AFCRL. Symposium on Models for the Perception of Speech and Visual Form*. AFCRL (In press).

13.__________. On the structure of prosody. Abstract of paper to be presented at the Symposium of the Perception of Speech and Speech Mechanisms, XVIIIth Intl. Cong. of Psych., Moscow, August, 1966.

14.S.E.G. Öhman and J. Lindquist. Analysis-by-synthesis of prosodic pitch contours. Abstract of paper to be presented at the Symposium of the Perception of Speech and Speech Mechanisms, XVIIIth Intl. Cong. of Psych., Moscow, August, 1966.

perceived differently when, in the one case, they are important parts of speech signals and when, in some non-speech context, they are not.[15] Related to this is the evidence from studies of consonant perception that continuous variations in the acoustic cue are perceived discontinuously -- that in these cases the phonemes are categorical, not only in the abstract linguistic sense, but as immediately given in perception.[16] This, we should think, is the speech mode. Continuous variations in non-speech sounds and, indeed, in isolated steady-state vowels, are perceived continuously, which is, presumably, the normal non-speech mode.[17][18] The abstract of K.N. Stevens indicates

[15] A.M. Liberman, K.S. Harris, J. Kinney, and H. Lane. The discrimination of relative onset-time of the components of certain speech and non-speech patterns. J.Exptl.Psych., 61, 379-388 (1961); J. Bastian, P Eimas, and A.M. Liberman. Identification and discrimination of a phonemic contrast induced by silent interval. J.Acoust. Soc. Amer., 33, 842 (1957)(A). A.M. Liberman, K.S. Harris, P. Eimas, L. Lisker, and J. Bastian. An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. Language and Speech, 4, 175-195 (1961); A. House, K.N. Stevens, T. Sandel, and J. Arnold. On the learning of speech-like vocabularies. J. Verbal Learn and Verbal Behav., I, 133-143 (1962).

[16] See all but the last of the papers referenced in Fn.15 and also: A.M. Liberman, K.S. Harris, H Hoffman, and B. Griffith The discrimination of speech sounds within and across phoneme boundaries J.Exptl.Psych., 54, 358-368 (1957); B. Griffith. A study of the relation between phoneme labeling and discriminability in the perception of synthetic stop consonants. Unpubl. Ph.D. dissertation, Univ. of Conn , 1958.

[17] D. Fry, A.S. Abramson, P. Eimas and A.M. Liberman. The identification and discrimination of synthetic vowels. Language and Speech, 5, 171-189 (1962); P. Eimas. The relation between identification and discrimination along speech and non-speech continua. Language and Speech, 6, 206-217 (1963); K.N. Stevens, S.E.G. Ohman, and A.M. Liberman. Identification and discrimination of rounded and unrounded vowels. J. Acoust Soc.Amer., 35, 1900 (1963)(A).

[18] Lane has written a critique of the assumption that speech perception requires a special mechanism, basing his argument very

that vowels in proper dynamic context are perceived more nearly in the fashion of the stop consonants.(19) These, then, would appear to be in the speech mode, and, according to the Stevens abstract, for theoretically interesting reasons.

Also relevant to the existence of these two modes are the now-emerging facts on the laterality of auditory perception. Several investigators have in recent years found small but reliable differences in the response to various acoustic stimuli depending on the ear to which the stimuli are presented.(20)

largely on a critical evaluation of the differences between categorical and continuous modes of perception. (H.L.Lane. The motor theory of speech perception: a critical review. Psychol. Rev., 72, 275-309 (1965). In our view, his criticisms reflect a misunderstanding of the relevant experiments -- which he has nevertheless replicated with our stimuli [D.V. Cross and H.L. Lane. An analysis of the relations between identification and discrimination functions for speech and nonspeech continua. In Experimental Analysis of the Control of Speech Production and Perception Progress Report, Vol. 6, 1964, Univ. of Michigan, Office of Research Administration, Ann Arbor, Mich.] -- an error in the application of the test for categorical perception, and the omission of other relevant evidence. We are preparing a detailed reply and correction. Cross, Lane, and Sheppard claim to have produced categorical perception in non-speech signals by applying the procedures of discrimination training. [D.V. Cross, H.L. Lane, and W.C. Sheppard. Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception. J.Exptl.Psychol., 70, 63-74 (1965)]. It appears to us that this claim is unsupported. [A.M. Liberman, M. Studdert-Kennedy, K.S. Harris, and F.S. Cooper. A reply to "Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception," by Cross, Lane, and Sheppard. Status Report on Speech Research, No.3 (1965) Haskins Laboratories].

19. K.N. Stevens. On the relations between speech movements and speech perception. Abstract of paper to be presented at the Symposium on the Perception of Speech and Speech Mechanisms, XVIIIth Intl. Cong. of Psychol., Moscow, August, 1966.

20. D. Kimura. Cerebral dominance and the perception of verbal stimuli. Canad.J.Psychol. 15, 166-171 (1961); Idem. Some effects of temporal-lobe damage on auditory perception. Canad.J.Psychol. 15, 156-165 (1961); B. Milner. Laterality effects in audition.

(Because of the greater efficacy of the crossed neural pathways, inferences can be made from such experiments about the relative contribution of the two temporal lobes in processing different kinds of auditory input.) When different speech materials are presented to the two ears simultaneously, thus creating binaural rivalry, the input to the right ear is more accurately perceived than that to the left ear. With brief melodies and sonar signals, the reverse is found. To discover, thus, that speech and non-speech sounds may be more effectively processed in different parts of the brain accords well with the hypothesis that such sounds are somehow processed differently.

In our own laboratory we have been carrying out similar experiments with speech stimuli of segment length. We have found that the perception of synthetic stop consonants seems to be more strongly lateralized than is the perception of synthetic, isolated, steady-state vowels.(21) This fact fits nicely with the evidence we referred to earlier concerning the different tendencies to categorical and continuous perception of the stop consonants and steady-state vowels. Extensions of this research,

In V.B. Mountcastle (Ed.). *Interhemispheric Relations and Cerebral Dominance*. Johns Hopkins Press, (Baltimore, 1962); M.P. Bryden. Ear preference in auditory perception. *J.Exptl.Psychol*., *65*, 103-105 (1963); D. Kimura. Left-right differences in the perception of melodies. *Quart.J.Exptl.Psychol*. *16*, 355-358 (1964); D.E. Broadbent and M. Gregory. Accuracy of recognition for speech presented to the right and left ears. *Quart.J.Exptl. Psychol*. *16*, 359-360 (1964); J.C. Webster, and R.B. Chaney, Jr. Information and complex signal perception. *Proc. AFCRL Symposium of Models for the Perception of Speech and Visual Form*. AFCRL (In press); D. Shankweiler. Effects of temporal-lobe damage on perception of dichotically presented melodies. *J Comp.Physiol. Psychol*. *62*, (1966) (In press).

[21] D. Shankweiler. Laterality effects in perception of speech and other sounds. Haskins Laboratories *Status Report on Speech Research*, No. 3, (1965). D. Shankweiler and M. Studdert-Kennedy. Laterality effects in the perception of synthetic consonants and steady-state vowels. (In preparation).

and of the research reported in the Stevens abstract, are likely to throw much light on the characteristics of the speech and non-speech modes of perception.

There remains one important fact about the code that we have not yet mentioned: it is universally present in man. Consider again that the ear is apparently so poorly suited to the rapid transmission of phoneme segments that such transmission can be accomplished only by means of a special code; yet, for all its apparent complexity, that code is so well matched to man that linguistic communication by speech is universal. That speech perception is universal is the more interesting in the light of the fact that reading print is not.

We have so far been concerned to establish that speech is some kind of code on the phonemic structure of the language. We should like to turn our attention now to the perceptual process by which the phonemes might be recovered from the coded acoustic signal. For that purpose it will be useful to consider, first, where and how the encoding might have occurred.

In thinking about the successive steps in the process of producing speech, we can proceed -- as speech itself presumably does -- from meaningful units of sentence size through smaller, semantically empty units at the phonological level to the final acoustic waveform. This is the sequence shown in Fig. 1. Here the initial operations at the syntactic level,(22) currently the subject of widespread study and discussion, have been grouped under a single heading. A more detailed diagram would distinguish such components as phrase structure rules, transformational rules, and morphophonemic rules. Since our concern is primarily with the phonological phase, however, we shall skip the syntactic

[22] For a recent treatment and reference, see: N. Chomsky. Aspects of the Theory of Syntax. M.I.T. Press. (Cambridge, Mass., 1965).

SCHEMA FOR PRODUCTION

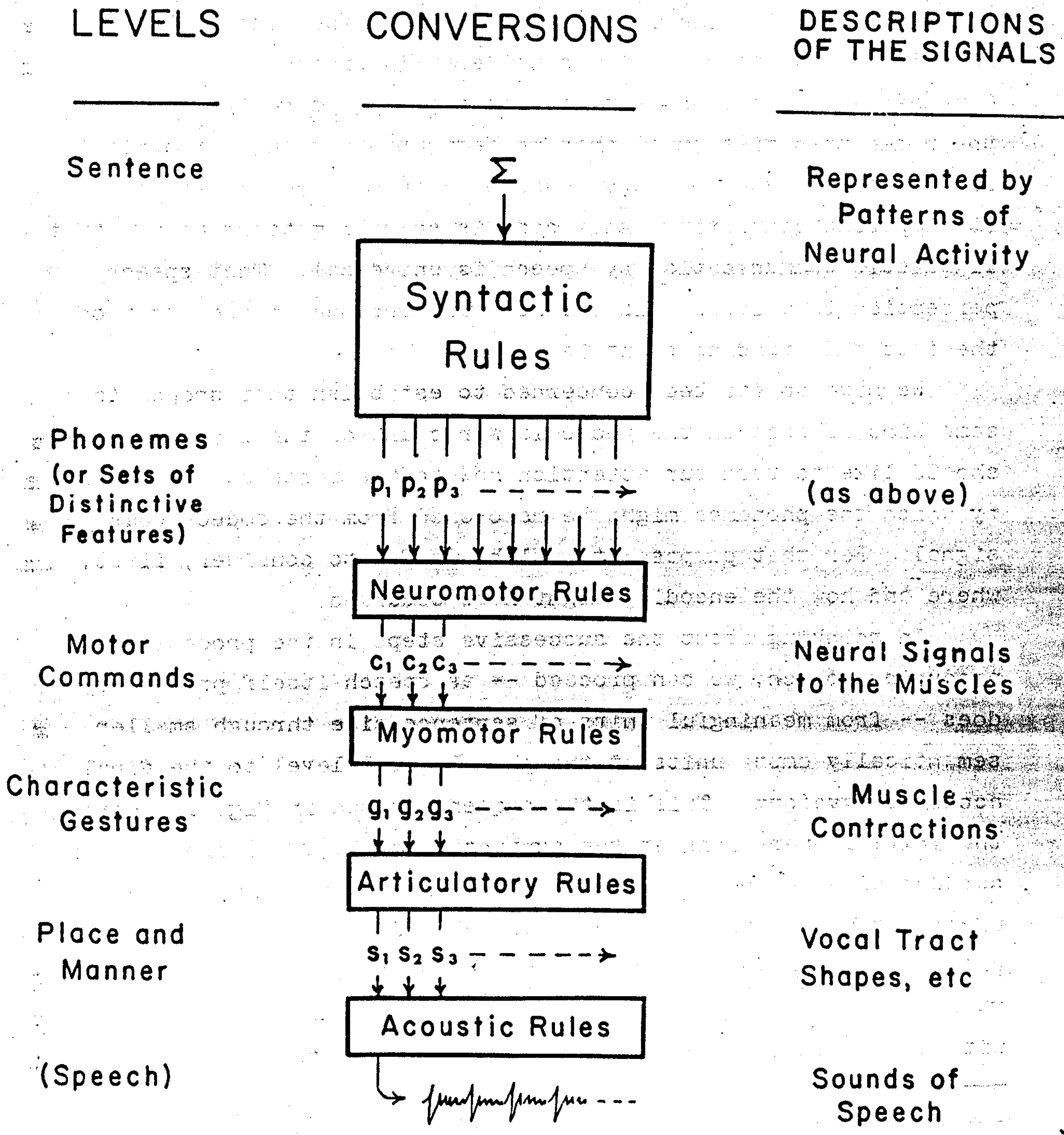

FIG. I

operations and start with the message in the form of a phoneme sequence (alternatively, sequential sets of distinctive features), taking this sequence as the input to successive converters that operate by neuromotor rules, myomotor rules, articulatory rules, and acoustic rules to yield, eventually, an acoustic stream.

In the first of these operations, the neuromotor rules serve to convert the ordered string of phonemes into a temporal sequence of neural signals to the muscles of articulation. The extent to which each phoneme (or its set of features) is uniquely represented by one or more of these neural signals is a point to which we shall return, for it is central to a working hypothesis that is guiding a major part of our own experimental work.

The relationships seem simpler and clearer at the second stage in the production process, namely, the conversion of neural signals into muscle contractions on the basis of myomotor rules. The signals map directly onto the muscles and control their contractions, whether via the intrafusal system or directly; moreover, the muscular events are observable by electromyographic techniques.

But simplicity has disappeared again at the next stage: the conversion from muscle contractions to vocal tract shapes (and related expiratory movements) by way of articulatory rules. The complexity introduced by this conversion is of two kinds. One follows from the bone and muscle structure of the articulatory system, with its intricate mechanical linkages and the spatial overlap of its muscular actuators. The shape that the tract will take in response to a particular set of muscle contractions is, one supposes, fully predictable, though extraordinarily difficult to compute. But what will happen if a new set of contractions (for the next phoneme) begins before the last set has had its full effect? Clearly, this introduces another kind of complexity, one that arises from temporal overlap of the incoming instructions. The shapes that result will no longer stand in one-to-one correspondence with the phonemes, but will reflect at each instant the interacting influences of several phonemes. This merging of effects constitutes an encoding operation in every sense except that the relationship of input to output is not arbitrary; it may be multivalued, however, and will be especially complex when temporal and spatial overlap occur together. Since overlap is the rule in speech, we can summarize the effects of the conversion from contraction to shape by saying that it is complex at best and almost always introduces an encoding of the sequential input units into output units of about syllabic size.

The final conversion, from continuously changing shape to a modulated acoustic stream, is by now rather well understood, thanks in considerable part to the able efforts of one of the organizers of this conference, Dr. Gunnar Fant(23). The application of the rules is complex in a computational sense and may give unequal acoustic prominence to various aspects of the changes in shape; nevertheless, the rules operate on an instant-by-instant basis and yield (for the most part) one-to-one relations between shape and sound, so that it is appropriate to consider this step from shape to sound as an enciphering rather than an encoding operation.

There are, then, between the phonological input and the acoustic output, at least four distinguishable conversions. We can say with assurance that the third step necessarily introduces an encoding of about the kind we observe in the output sound stream. And so we see how and why it is that speech, as it exists out in the air, must be a coded message.

23. Fant's monograph on this subject offers a comprehensive treatment and references to other related research. [G. Fant. Acoustic Theory of Speech Production. Mouton. (S'-Gravenhage, 1960)].

But how does the listener decode it? Let us consider the total process of speaking and listening as it is sketched in Fig. 2 Please ignore for the moment the dashed lines and consider only the U-shaped sequence. The schematic account it gives of events in reception -- the righthand side of the U -- is just what one would expect if the perception of speech had no special connection with its production. This is, indeed, the obvious and seemingly simple view of how speech is perceived, namely, that acoustic units of appropriate size are learned as separate neural patterns that are stored in the central nervous system, with as many neural patterns as there are words (or perhaps syllables) in the language. This would solve the invariance problem, albeit by an extravagant use of storage, but it leaves unexplained the listener's known ability to perceive on a phoneme-by-phoneme basis. If we are to explain how the listener retrieves the phonemes, we must postulate a decoder with functions that include, at the very least, an inverse of the encodings imposed by bone and muscle on the acoustic stream. All this is the "simple" way to listen to speech!

But can we not account for the perception in some less extravagant way? We think so. We think there is a mode of perception uniquely fitted to speech and responsible for the high efficiency of its signals. In the most general terms, this mode takes advantage of readily available mechanisms that allow perception to operate by reference to the motivating events of articulation. (24)

24. Some of the bases for this view and its development into a motor theory of speech perception are to be found in earlier publications from our laboratory See especially: A.M. Liberman, P C Delattre, and F.S. Cooper. The role of selected stimulus-variables in the perception of the unvoiced stop sonsonants. Am. J. Psychol., 65, 497-516 (1952); A.M. Liberman. Some results of research in speech perception. J.Acoust.Soc.Amer. 29, 117-123 (1957); F.S. Cooper, A.M Liberman, K S. Harris, and P M. Grubb.

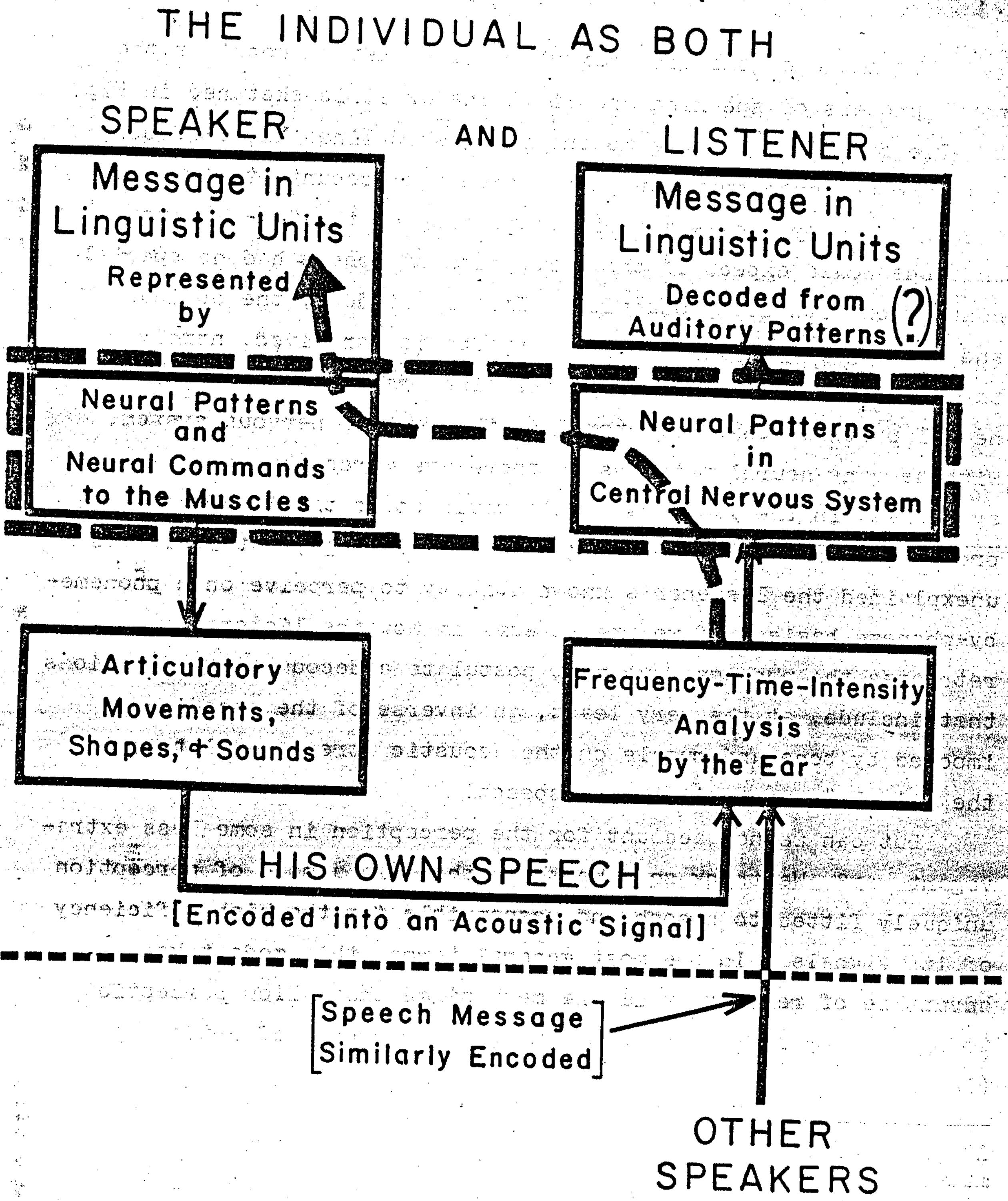
THE INDIVIDUAL AS BOTH
SPEAKER
AND
LISTENER
Message in Linguistic Units
Represented by
Message in Linguistic Units
Decoded from Auditory Patterns (?)
Neural Patterns and Neural Commands to the Muscles
Neural Patterns in Central Nervous System
Articulatory Movements, Shapes, + Sounds
Frequency-Time-Intensity Analysis by the Ear
HIS OWN SPEECH
[Encoded into an Acoustic Signal]
Speech Message Similarly Encoded
OTHER SPEAKERS

FIG. 2

A possible model (25) finds its basis in overlapping activity of the nerual networks that supply control signals to the articulatory apparatus and those that process incoming neural patterns from the ear. We know that temporal overlap of these activities exists as an ever-present consequence of the fact that people listen while they speak. If we assume also (a) that there is functional overlap at the neural level so that *both* motor and sensory networks respond (in ways that are characteristic of the activating event) when *either* is activated, and (b) that information can be passed in either direction through these neural mechanisms, then there exists a path from ear to perceived message that is not dependent on an *auditory* decoder and does not require a vast store of auditory patterns. It is this pathway, and the assumed areas of overlap in neural structure and function, that are indicated by the dashed lines of the figure.

Some input-output relations observed in experiments in the perception of speech. *Proc 2nd Intl.Cong.Cybernetics*, 930-941 (Namur, Belgium, 1958); L Lisker, F.S. Cooper, A.M. Liberman. The uses of experiment in language description. *Word*, *18*, 82-106 (1962); A.M. Liberman, F.S. Cooper, K.S. Harris, and P F. Mac Neilage. A motor theory of speech perception. *Proc. Speech Comm. Seminar*, Royal Institute of Technology (Stockholm, 1963); A.M. Liberman, F.S. Cooper, K.S. Harris, P.F. MacNeilage, and M Studdert-Kennedy. Some observations on a model for speech perception. *Proc. AFCRL Symposium on Models for the Perception of Speech and Visual Form*, AFCRL (In press).

25. The model sketched out in this paper is intentionally nonrestrictive as to physiological mechanisms. The reader will detect, however, an obligation to Hebb and Milner that the authors are glad to acknowledge. [D.O. Hebb. *The Organization of Behavior*. Wiley (New York, 1949); P.M. Milner. The cell assembly: Mark II. *Psychol. Rev.*, *64*, 242-252 (1957).] The abstract for this Symposium submitted by Tappert seems to have much the same orientation as the model discussed in this paper. [C.C. Tappert. A model of speech perception and production. Abstract of paper to be presented at the Symposium on the Perception of Speech and Speech Mechanisms, XVIIIth Int. Cong. of Psychol., Moscow, 1966.]

Clearly, it will be important to know the linguistic level at which the message units are recovered, and the model we have described does not speak to this point. Before we turn to one that does, it may be useful to make some further observations about this very general model for speech perception by reference to production: (1) Even so general a model as this permits useful inferences, since it implies recovery of the speaker's own message -- or his analysis as a listener of the messages of other speakers -- in terms of the same linguistic units that enter into production. Thus we can see, in a general way, how the listener is able to decode complexly encoded messages into their linguistic components without having to make use of a special auditory decoder. (2) The reference to motor activity does not -- as in some older motor theories -- call for reference to the peripheral muscle activity and its proprioceptive consequences. This kind of operation is not excluded by the model, but is no part of it. (3) The model deals primarily with the nature of the decoding mechanism, not with how it was acquired by the species or the individual. The relative contributions of intrinsic structure and learning, as well as their interaction, raise interesting but separate questions. (4) Reference to production provides a pathway for perception, but not one that is obligatory -- that is, the existence of this pathway does not preclude direct auditory processing of speech patterns by the same means that are used for recognizing animal cries, traffic noises, and the like. The special pathway would be used, we suppose, whenever it facilitates perception, as it would in recovering linguistic units that lack invariant acoustic counterparts, but when it is not needed, it may not be used. (5) Finally, the special processes that permit reference to production are not necessarily restricted to reference at the phonological level or, for that matter, to only one level at a time.

So much for this very general model. It is, as we have noted, noncommital as to the level at which reference and recovery occur. Since our present concern is with the phoneme and how it is perceived so well and so fast, we shall try to make the model more specific in ways that bear on that question. One possibility is to postulate that the reference occurs at a level in production where the neural patterns represent the individual phonemes, i.e., at the input to the first of the phonological converters of Fig. 1, the one that operates by neuromotor rules. This would "explain" why it is the phonemes that we perceive -- but it accounts for very little else.

A more productive assumption is that the neural overlap is at, or just above, the output of this neuromotor converter. Here we would expect to find separate neural patterns that send actuating signals to the separate muscles (or closely related groups of muscles) of the articulatory apparatus. But independent control of the component parts of the mechanism means that the information flowing through it becomes multidimensional in physiological coordinates; that is, the serially ordered phonemes come to be represented by muscle activities that proceed in parallel, and that can persist and overlap as the segmental phonemes cannot. These muscle activities will leave their traces in the sound stream, though here the original dimensions no longer exist in independent form, but only in the implicit form of subphonemic features. If, now, the neural patterns of reception, which likewise contain the "traces" of the original muscle activities, overlap and link into the motor patterns that actuate the independent muscle groups, then reception, too, will be multidimensional, and in the original terms.

The original terms, however, were those of multidimensional motor control signals that had somehow come to represent the phonemes. But let us be more explicit about how this happens.

The strongest assumption would be that each individual phoneme of a language has its own characteristic set of neural patterns all the way down to the motor nerves that go to the muscles. This is, in fact, the assumption we are testing in some of our experimental work.

The working model containing this latest assumption does not imply, however, that the *total* neural and muscular activity will be in one-to-one correspondence with the phoneme, but implicates only one or a few component parts, perhaps even the contraction of a single specific muscle in the extreme case. The neural signals for this characteristic component of the total activity are what we have referred to as *motor commands*(26); the term was chosen to distinguish these characteristic, or invariant, signals from all the other neuromotor signals (needed in well coordinated gestures) that may be present at the same time. The objective of the experiments is, then, to find muscle contractions (corres-

26. Motor commands are essentially the same as the "action patterns" we discussed in an earlier description of this model. (F.S. Cooper, A.M. Liberman, K.S. Harris, and P.M. Grubb. *Op. cit.*).

27. Little can be said here of experimental methods and early results. See, however, K.S. Harris, M.M. Schvey, and G.F. Lysaught. Component gestures in the production of certain final clusters. *J Acoust.Soc Amer.* *35*, 461-463 (1963); P.F. MacNeilage and G.M. Sholes. An electromyographic study of the tongue during vowel production. *J. Speech and Hear. Res.* *7*, 209-232 (1964); F.S. Cooper. Research techniques and instrumentation: emg. *Proc. of the Conference on Communicative Problems in Cleft Palate ASHA* Reprints, No. 1, (April, 1965); K.S. Harris, G.F. Lysaught and M.M. Schvey. Experimental studies of the production of oral and nasal labial stops. *Language and Speech*. (In press). Electromyography is a most useful tool since it permits very direct inferences about the neural commands to the muscles, but x-rays and a variety of physiological measures are valuable also, even though the information they give directly is about the encoded consequences of the nerual commands. Ohman has described a model that computes vocal tract shapes resulting from VCV coarticulation.[S. E.G. Ohman. Numerical model for coarticulation, using a computer-simulated vocal tract. *J.Acoust.Soc Amer.* *36*, 1038 (1964) (A)]

ponding to motor commands, or simple combinations of them) that are present whenever a particular phoneme is present in a message and are not present when the phoneme is not(27).

It may turn out that Nature will not endorse so simple a model. She provides many examples of intricate motor coordination in skilled movements, and it may be true of speech that significant reorganization of the neural patterns occurs above the level at which nerve impulses are sent to the muscles. A likely case would be the production of clusters that overlap spatially; these might come to be treated as "ligatures" in motor command terms. The model we have proposed is not too inflexible for such eventualities; it would, though, be less useful in helping us to understand the processes and to predict their consequences if the neural reorganizations were so extensive and so far upstream in the nervous system as to be inaccessible to experiment. But even if this should happen, the more general model for perception by reference to production would remain, with much to recommend it.

We began these observations by asking how speech can be so

[28] Both similarities and differences may be noted between the model discussed here and Steven's model for analysis-by-synthesis. [K.N. Stevens. Toward a model for speech recognition. J. Acoust. Soc. Amer. 32, 47-55 (1960).] Both involve interaction between productive and receptive processes, with a comparison of the two sets of signals and linguistic choices determined by this comparison. The orientations are rather different, with more of the imagery of electronics and computation in the one case and of neurophysiology and adaptive networks in the other; also, the analysis-by-synthesis model (at least in its early forms) implies that comparison is done on the receptive side, i e., that the incoming speech patterns are matched with implicit auditory patterns generated by a motor mechanism, whereas the model presented here calls for a comparison of implicit motor patterns, and on the other side of the system. It is obvious, of course, that neither model will be found under the skull, and it may be that even the attempt to localize the comparative function on the one side or the other has no meaning in neural reality. Preference, then, will rest on predictive power -- and so on further research.

rapid -- how it can evade the rate limitations of the ear -- and how segmental phonemes can be recovered as well-perceived segments from the unsegmented and usually encoded acoustic stream We have found an explanation in terms of a model(28) that refers the acoustic and auditory features of an incoming message back to the multidimensional events of actual or implicit articulation Simultaneous tracking of the features can then lead to recovery of the equivalent multidimensional descriptions, and so of the phoneme string. Thus, the phonemes are disassembled and reassembled during production and reception, respectively, and are transmitted in a parallel mode as features(29) that refer to neuromotor events.

The transmission is in terms of several slowly changing features, and so falls easily within the speed limitations of the ear-brain system. But how do we recover the high-speed timing and proper sequencing of implicit motor events, as we must since correct timing is as important in production as the correct selection of actuators? Perhaps this is the role of phonological constraints and a justification of the phoneme's existence(30), for if the neuromotor events are allowed to occur only in particular combinations and sequences, then there is a basis for re-

[29] These might have been called "distinctive features" (in the Prague School sense) and this usage would also have acknowledged our deep obligation to Roman Jakobson and his colleagues. We have not done so just to avoid confusions between Jakobson's usage and the usage here. Whatever the overlap may prove to be, the defining operations that we use are different from those used by Jakobson et al., and so call for different terms. (See: Roman Jakobson, Morris Halle, and C. Gunnar M. Fant. *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. M.I.T. Press, 2nd Ed., 1963.)

[30] In this view, the segmental phoneme is a unit that specifies in quantal form *both* selection (i.e., the features) and relative timing, and so facilitates -- at the very least -- the conversion from the parallel mode employed in transmission to the serial order of message units as they exist on the phonological level.

synchronizing -- and so for imposing phoneme boundaries on -- the incoming multidimensional message stream.

In short, we have assumed that the encoding of the phonemes occurs below the level of the neural patterns that send actuating signals to the articulatory muscles. By referring the incoming sound to those patterns, the listener can track the features and recover the invariant relation to the phoneme. One can suppose, moreover, that the many and distinctively different dimensions of the neuromotor events give the listener a basis for identifying the phonemes absolutely, and for doing this far better than he can with an equal number of acoustic signals that are not in the speech mode. Finally, we see that by encoding the message so as to put the phoneme segments through in parallel (as features), we avoid the limitations on rate of discrete segment perception that are set by the temporal resolving power of the ear. In general, then, we find that the sounds of speech are uniquely well perceived because they are a special and especially efficient code, processed by a special and readily available mechanism.

* Also, The university of Connecticut
** Also, Barnard College, Columbia University