DOCUMENT RESUME

ED 059 200 TE 002 780

AUTHOR House, Arthur S., Ed.
TITLE Communicating by Language: The Speech Process.
PUB DATE [67]
NOTE 327p.; Proceedings of the Conference on Communicating by Language: The Speech Process (Princeton, April 26-29, 1964)

EDRS PRICE MF-$0.65 HC-$13.16
DESCRIPTORS Acoustic Phonetics; Articulation (Speech); Behavioral Science Research; *Communication (Thought Transfer); *Conferences; Discourse Analysis; Feedback; Group Discussion; Interaction Process Analysis; *Interdisciplinary Approach; Language Development; Language Skills; Models; Neurology; Perception; *Problem Solving; Psychology; Psychomotor Skills; *Speech; Stimuli

ABSTRACT

This document reports on a conference focused on speech problems. The main objective of these discussions was to facilitate a deeper understanding of human communication through interaction of conference participants with colleagues in other disciplines. Topics discussed included speech production, feedback, speech perception, and development of language and language skills. Five levels of discourse were dealt with--the acoustical, the neurologic, the articulatory, the psychologic or behavioral, and various model languages. A concept of the speech-production system discussed was a system that has an output of phonemes and an input of control instructions. The discussion of the concept of feedback revealed that, depending upon the level of complexity of the speech response that was under discussion, the conference participants had a difficult time settling on how many and how extensive had to be the feedback loops that would be involved. In the consideration of speech perception, conference participants again had difficulty in reaching a decision on the definition of the stimulus for speech perception. A review is given of the schedule of development of certain kinds of speech and language behaviors. Conclusions include: (1) Speech production is a general kind of complicated motor behavior; and (2) Time-variant characteristics of speech signals are less identifiable anatomically than are spectral characteristics. (CK)

COMMUNICATING BY LANGUAGE --

THE SPEECH PROCESS

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service

COMMUNICATING BY LANGUAGE: THE SPEECH PROCESS

Proceedings of the Conference on
Communicating by Language: The Speech Process
April 26-29, 1964, Princeton, N.J.

Chairman
FRANKLIN S. COOPER, Ph.D.
Haskins Laboratories
New York, New York

Editor
ARTHUR S. HOUSE, Ph.D.
Purdue University
Lafayette, Indiana

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service - National Institutes of Health
National Institute of Child Health and Human Development
Bethesda, Maryland 20014

TABLE OF CONTENTS

PREFACE

This book is the product of a conference entitled "Communicating by Language: The Speech Process," sponsored by the Human Communication Program of the National Institute of Child Health and Human Development, held April 26-29, 1964, at Princeton, New Jersey. The conference was organized by Dr. Norman F. Gerrie, Director of the Program, Dr. James F. Kavanagh, and Dr. Francis J. Kendrick. The proceedings were arranged by the Conference Chairman, Dr. Franklin S. Cooper, Director of the Haskins Laboratories.

The National Institute of Child Health and Human Development was established in 1963 and is the youngest of the nine institutes within the National Institutes of Health, Public Health Service, U.S. Department of Health, Education, and Welfare. Its role is to stimulate, support, and develop research into broad areas of human development, concerning itself with both normal and relevant pathological processes, and with the whole individual as well as with specific systems.

At the time this conference was held, the National Institute of Child Health and Human Development was, through the Human Communication Program, assessing the state of knowledge in the area of human communication for the purpose of revealing existing and potential directions of study, and to identify the roles which various disciplines can and do play in expanding that knowledge, both independently and jointly. This conference was one of several sponsored by the Human Communication Program to further these objectives.

Following a general reorganization of the National Institute of Child Health and Human Development in 1965, the activities of the Human Communication Program were integrated with those of the Growth and Development Program.

The speech process, which plays a central role in human achievement and human culture, is one important aspect of growth and development. Yet we are only beginning to understand how we communicate with each other by linguistic codes, how the codes themselves are organized, and how these skills are acquired by children and adults. It was our hope that the discussions during this conference might serve to let each of the participants view these problems through the eyes of colleagues working in other disciplines and so give deeper understanding of human communication.

James F. Kavanagh, Ph.D.
Growth and Development Program
NICHD

PARTICIPANTS

Members

Richard A. Chase
The Johns Hopkins Hospital
Baltimore, Maryland

Franklin S. Cooper
Haskins Laboratories
New York, New York

Peter B. Denes
Bell Telephone Laboratories, Inc.
Murray Hill, New Jersey

Dennis B. Fry
University College
London, England

Norman F. Gerrie
Human Communication Program, NICHD
Bethesda, Maryland

Ira J. Hirsh
Central Institute for the Deaf
St. Louis, Missouri

Arthur S. House
Purdue University
Lafayette, Indiana

James F. Kavanagh
Human Communication Program, NICHD
Bethesda, Maryland

Francis J. Kendrick
Human Communication Program, NICHD
Bethesda, Maryland

Eric H. Lenneberg
Children's Medical Center
Boston, Massachusetts

Alvin M. Liberman
University of Connecticut
Storrs, Connecticut

Kenneth N. Stevens
Massachusetts Institute of Technology
Cambridge, Massachusetts

Guests

Donald E. Broadbent
Applied Psychology Research Unit
Medical Research Council
Cambridge, England

Frank Fremont-Smith
Interdisciplinary Communications Program
The New York Academy of Sciences
New York, New York

Norman Geschwind
Veterans Administration Hospital
Boston, Massachusetts

Moise H. Goldstein, Jr.
The Johns Hopkins University
Baltimore, Maryland

John V. Irwin
University of Wisconsin
Madison, Wisconsin

Peter N. Ladefoged
University of California
Los Angeles, California

Joseph C. R. Licklider
IBM Research Center
Yorktown Heights, New York

John Lotz
Columbia University
New York, New York

Brenda Milner
Montreal Neurological Institute
Montreal, Canada

R. C. Oldfield
University of Oxford
Oxford, England

Irwin Pollack
The University of Michigan
Ann Arbor, Michigan

Arne Risberg
Royal Institute of Technology (KTH)
Stockholm, Sweden

EDITOR'S FOREWORD

There is some question whether the old saw, "Better late than never," is applicable to conference reports. The tardiness of this account of the Conference on Communicating by Language: The Speech Process is particularly unfortunate since there is no question that the meeting--by almost universal agreement of those who took part in it--was both instructive and stimulating.

Perhaps the informality of the meeting was at once a strength and a weakness. The discussion leaders were successful--sometimes by their own introductory comments and oft-times merely because of the provocative nature of the materials they introduced--in developing stimulating and free-wheeling discussions. The absence of formal papers for presentation and discussion contributed positively to the success of the conference from the participants' point of view, but has made the job of reporting the conference a problem.

From the options available to me I have taken the point of view that neither a verbatim account of our discussions nor an editor's synopsis of our discussions would constitute a proper report. This report is a compromise document and probably will please no one. In form it resembles the discussions that took place, but the exchanges and comments have been amended in a variety of ways. For example, I have attempted to make the discussions more intelligible to the reader who is without visual cues that show facial expressions and auditory cues of timing and intonation, and I have standardized the method of address to the use of surnames. In addition, each participant has had the opportunity of editing his own remarks, and this option has been exercised in a number of individual ways. In some instances at least, these editorial changes by participants have removed from the text a sense of excitement and adventure that permeated many of the original animated exchanges. This document, therefore, is not a completely faithful account of the conference, but it does give a picture of the scope of the materials considered and the directions in which the discussions moved.

The conferees have already profited by their active participation; this report, hopefully, may stimulate students and investigators who were not fortunate enough to participate in the conference. Any failure of the report to convey the spirit of the meeting to the reader must be attributed to my ineptitude.

The forbearance of the NICHD Staff is acknowledged with gratitude; without their cooperation my task could not have been accomplished. The clerical skills of Mrs. Carolyn Rifino were invaluable in preparing the manuscript; the report is dedicated to her first-born daughter whose "unscheduled" arrival contributed significantly to the confusion inherent in manuscript preparation.

Arthur S. House, Ph.D.

INTRODUCTORY REMARKS

COOPER I think all of you have heard colleagues complain about meetings at which there were too many papers; in fact, I expect some of you made the same complaint. The reason is that these papers tend to prevent the kind of discussion that takes place in the corridors--the kind that lets you find out what other people working directly along your lines are concerned with, or what people whose work is only marginally related but serves to give you a new look into your own problems are doing. One objective of this meeting, then, is to do away with papers, more or less in the spirit of the sign that was on a little motor scooter that occasionally parked in a small corner in front of our laboratory; the sign said, "Help stamp out Cadillacs."

There is also an indirect objective, namely, to have some of the kind of discussion that we complain about not having at regular meetings. There aren't many constraints on what we are free to discuss here. There is a topic, but I think we ought to consider it a rallying point to which we should return now and then rather than some kind of perimeter fencing us in. You might be interested in knowing, in general, how that topic was arrived at.

First, I was exposed to the broad objectives of the Human Communications Programs of the NICHD, and this certainly did not provide any kind of fence. They seemed to include anything that involved information, human beings, how that information is processed, and how the processor got that way. It did seem, in trying to arrange a conference, that it would be wise to restrict the topic a little more than that--to deal with one particular kind of information, namely, where the coded material for the information itself is structured, as it is in human language--and this is the basis for the choice of the first title, Communicating by Language. The intent is to put emphasis on communicating and the idea that the process involves an organized code, not simply anything that happens to occur, as would be the case, for example, in the processing of sensory signals. This involves information and involves human beings, but it is nature's code and not man's code.

The remainder of the title, The Speech Process, reflects a hope that we will deal, at least primarily, with the process and not all of its ramifications. You may have noticed in the agenda one exception to this, namely, "Man-Machine Communication and Models." There are some interesting parallels to be drawn from the way we communicate with machines and the way they inform us.

We have chosen not to consider man-to-animal and animal-to-animal communication, but this is a very reasonable extension for future discussions.

We have, then, a very wide area in which to carry out our discussion and, as each of you knows, he has not been invited to give a paper. Since nobody has been invited to give a paper, the form of the discussion is also very free. Of course, with that freedom comes responsibility, and the responsibility is to volunteer what you think might be of interest to the other persons, and to get right into asking questions and adding things to what somebody else is saying.

I was torn between two philosophies in looking for discussion leaders. There are a number of considerations, of course, and one is that the man ought really to know his topic and be quite prepared to be one of the major contributors. Another possibility is that the man should not be chosen who, you suspect, will have the biggest and longest contribution to make--such a person might be inhibited by being in the chair. I have used both philosophies, and you will have to decide for yourself when.

Some people have asked me, "What does a discussion leader do?" I have tried to be sufficiently vague so that they would try to figure out for themselves what they thought they ought to do; my guess is that this is the best thing they can possibly do. Since it is intended that the meeting be a free-wheeling one, the procedure that each discussion leader follows is to be about as free-wheeling as any part of it. This gives us a lot of liberty, a lot of flexibility. It also lays a responsibility on each of us to see to it that the discussion leader knows that we individually have something to say in his session.

Two people have taken on special assignments. Hirsh agreed to attempt to summarize for us or to bring up some of the unfinished business that we had not covered in our discussions for consideration and review in the Wednesday afternoon session. He will, no doubt, want to play this by ear, and do it the way he feels best. This will be, again, a time to see where we have been, what we ought to have covered and haven't, and to have one man's over-view of the discussions. House has taken on the chore of editing the transcript and trying to get it into form for circulation. In case any of you are feeling inhibited about talking while there is a stenotypist taking down everything you say, you can relax, because you will have a chance to see how bad it was before it goes into print.

Fremont-Smith will tell you a good deal more about the general pattern of this kind of discussion--some of you are familiar with it and some are not--and also some of the mechanics that we will be observing. For myself, I expect to enjoy these discussions, and I hope to go away with some new insights into my own problems and, maybe, even some new plans for research that have come out of these discussions. That really is the objective of the conference.

FREMONT-SMITH It is not speeches but communication that we are after. I like to contrast a speech or a lecture with a conversation. It seems to me that one way of looking at a lecture is to say that you have a captive audience. I except, of course, all those rare, beautiful and wonderful lecturers which just carry everybody with them. With the other kind of lecture, if the captive audience listens and if they have any interest in the topic, naturally, almost immediately or within a few minutes, they have a series of questions, doubts, ideas, comments or associations, all of which it would be such fun and really important to bring out. But you can't bring out these points because it is not polite to interrupt a lecturer. You repress these ideas, one after another. This is why people are so fatigued at the end of a lecture, because they are exhausted by the process of progressive repression and the progressive frustration that accompanies it. This is what happens, until one goes off into daydreaming or doodling or thinking about something else.

On the other hand, there is just one in the group who gets an awful lot of satisfaction out of a lecture, and that is the lecturer himself, because he is his own audience. He plans what he is going to say, he says it, and he hears himself saying it. The words come out very much as he planned them, and they are enormously reassuring to him. He gets progressive reassurance and progressive fulfillment out of the lecture.

Well, this is not what we are here for, and I would contrast this with a conversation, which, I believe, is the original method of communication developed by the human race very early in the game and somehow lost sight of to a large extent in any of our formal arrangements within the university or scientific community. We sort of lost it, and now we are trying to rediscover it and bring back into the fullness of its significance the value of conversation.

Conversation is a mutually corrective feedback system; the essence of conversation is interruption, and the mood of our conference should be, and I hope it will be, "Don't speak when I'm interrupting."

SESSION 1. The Perception of Speech

COOPER Our topic this morning is speech perception, and Dennis Fry is our discussion leader.

FRY I think, perhaps, I had better begin by making a couple of general remarks. When people take in speech, what they perceive are sounds. What I mean by this is that I don't think it's a very good thing to say that people perceive words or they perceive sentences or they perceive syllables. All these things are constructs on the result of what they perceive. I say this not in order to limit this morning's discussion, but because I believe, when we talk about the perception of speech, we really ought to keep this well in mind all the time.

FREMONT-SMITH You use perception of speech as distinguished from perception of meaning?

FRY I wouldn't use the words perception and meaning in the same sentence, but I guess that is a fair thing to say. I would say that the grasping of meaning is the result of the processing, and the processing begins with perception, but what we perceive are sounds.

FREMONT-SMITH And the perception is already an interpretation, isn't it?

FRY Again, I wouldn't want to use the word interpretation.

FREMONT-SMITH Well, can we say that we don't perceive first and then interpret secondly, but we bring in an interpretive aspect into the very initial process of perception?

FRY This, I think, is right. Our perception is certainly influenced by what are, in fact, from one point of view, subsequent parts of the processing.

FREMONT-SMITH I say this primarily to get a good excuse to interrupt immediately, because I promised I would do so.

FRY Well, the second general remark is simply this: when we look at speech we try to make observations of what is going on, and we just have to realize all the time that in speech there are many different ways of doing any one thing; there is not a single way of doing it. It's no use saying, in speech, this operation is done in this way and not in that way. This is the nature of speech, that everything can be done in different ways. It will be in different ways according to the individual, in different ways according to the circumstances.

I believe it is very important, when we start to talk about perception of speech, not to be bogged down in any kind of wrangle that things are done in this way but not in that way. The truth is that they are likely to be done in both ways at different times.

FREMONT-SMITH Or even simultaneously?

FRY Yes, this is partly what I mean; that one single operation is made up of component operations, and if we adopt mutually exclusive ideas, then, we simply get into mistakes.

Well, now, when we come to the various topics that we might take up, there are a great many of them, and I am not going to attempt to review them. We have a good deal of information about acoustic cues that are used in reception of speech and in the perception of speech. To my mind, one of the things which is lacking here is a great deal of information about how human beings deal with the quality of sound. There is a good deal of psychophysical information about pitch and about loudness, but, as far as I can make out, very little about quality.

In perceptual experiments, when we make variations in a particular physical dimension, we often are at a loss to know how to regard these differences from the perceptual point of view. For instance, in vowel studies with synthetic vowels, we make a certain change in the plot of the frequency of the first formant versus the frequency of the second formant (that is, the F_1-F_2 plot). We have no idea even what kind of scale is appropriate on the sensation side or what kind of scale is appropriate for differences of quality, and therefore we are embarrassed to know what relation this change in the physical dimension really represents on the perceptual side. Therefore, of course, we are hard put to it to evaluate changes in judgment which have taken place. This seems to me something that we really need information about and some work on.

Following on that, of course, we also have a good deal of information about the various kinds of cue in the perception of speech; I think it is extremely important now that there should be work on the combination of cues. We have had to do a lot of experiments so far with single cues. This was the only way to start, to take a particular physical acoustic cue and see what variation of this does to perception. But, clearly, in natural speech, as Fremont-Smith was suggesting just now, in a particular act of recognition, which is the next stage after perception, various cues are working at the same time, and we really do need hard experimental knowledge about what happens when you combine cues. It is clear that they do not simply add up, but interact in a fairly complicated way, and it seems to me that we need information and work along this line.

COOPER I wonder if you would go back to your first thought for the day, that perception is of the sounds. I wasn't quite sure whether you were stating a theory or giving a definition.

FRY I'm putting forward a point of view which seems to me to be the only solid one that we can use in speech, that is to say, to realize that constructs on the results of perception of sound begin very, very early in this process of taking in speech.

HIRSH I'm not really sure what you mean. In the individual case, does this mean, for example, that, as he looks at a text, one does not see or does not perceive the words, but perceives the letters, or, really, one does not perceive the letters, but rather the little lines of which the letters are composed?

FRY I'm not prepared to make a jump to the visual form of language.

HIRSH All right. Then, we'll stay with the auditory one and let me go in a different direction. I think this is what Titchener called the stimulus error (13). You are trying to find some definable stimulus aspect, and, since, in a physical way, you can't quite define a word, but you can define a sound, then you decide that the reference for the perceptual process shall be the sound.

FRY You mean you as the experimenter?

HIRSH Yes. You could say that what one really perceives are the nerve impulses going up the central nervous system tract to the brain. But I don't think that would get us anywhere, and I don't think "what one perceives is really sound" gets us any further.

LADEFOGED Could I save you, Fry, by digging in where you didn't want to, and pursuing the visual case.

Isn't the thing we should be wary of, as Fry was suggesting, thinking that we perceive in terms of any of the units or anything like syllables or whatever it might be, whereas, in the visual case, it seems fairly apparent that people's actions, as we observe them when they have certain eye movements and so on, do not correlate with how they must later on be interpreting the stimuli? As you read a page, your eye movements are not exactly random, but they are not directly correlatable with the words or with the features of meaning, in any sense, or the letter shapes and so on. Would that fit in with what you were indicating before?

FRY Perhaps the best thing here is to try to get some kind of example. Supposing I talk about some friend of mine, and I say this fellow is a cytologist. The question is, what does everybody perceive?

FREMONT-SMITH Doesn't everybody perceive a different thing, depending upon what psychologist means to them and also what you are saying, that he is a psychologist, means to them?

FRY Well, thank you very much. You see, what I actually said was that the fellow was a cytologist. (Laughter)

FREMONT-SMITH Oh, excuse me! That proves my point.

FRY It proves my point, too. (Laughter) The question now is, Hirsh, what is the situation about perception here?

HIRSH Well, at first, I thought that you said what one perceives is a succession of s, ai, t, and so forth.

FRY No, just sounds. I said you perceived sounds.

HIRSH Just uttered sounds?

LIBERMAN I thought he said something nobody could take exception to.

FREMONT-SMITH Yes. Doesn't this prove the need of this conference?

HOUSE I am confused at this point. I would like to ask you to show me what the implications of your assertion really are, because either you said something that is almost vacuous, or you said something that has some implications that are very deep. The implications that can be read into the statement run counter to my own interpretation of data about perception, but if you are merely saying that "I would like to open the discussion by saying we are listening to sounds, and this is what speech perception has to deal with," then, I will rest until you go further.

FRY I was simply saying something vacuous, but only because we so easily lose sight of this cardinal fact about perception. The taking of a cytologist as psychologist is rather crucial, I think. Did Fremont-Smith perceive the sound k or not?

BROADBENT Surely, this is a question about the use of the word perception.

FRY To my way of thinking, he simply has constructed psychologist, not through the perception of sounds, or certainly not through the perception of sounds in that sequence.

BROADBENT I should have thought that one could try to analyze these kind of statements by saying that you have physical events, and you have the chap having some kind of event inside him which corresponds more or less to the physical events. Now, are you going to say that he perceives something when you have identified what it is that he is responding to?

I think that we have been hopping back and forth from one of these usages to the other all the time. There is no doubt but that his perception is initiated by sounds, and equally that it may not correspond to what somebody else would perceive from the same sounds. But I don't think that it is terribly profitable to say that he perceives sounds as if this were an antithesis to his perceiving words.

FREMONT-SMITH May I make a comment here? We are in an inevitable dilemma, I think, which arises, in part, from the fact that we have quite a lot of information on what takes place in the central nervous system, and we have quite a lot of information about human behavior. The linkage between what takes place in human behavior and what takes place in the central nervous system is rather tenuous, but it is this bridge we are dealing with, because this is exactly where the problem is. Some of us locate our interest on the central nervous system process, and some of us are locating our attention on human behavior. I think we have to do with both.

Secondly, I would say I hope we don't get out of this hassle. It is exactly this kind of thing that is important. If we get out of the hassle because somebody makes a statement at this point, it's going to leave everybody where they were, with a feeling of discomfort and of lack of communication at this level. I think this hassle is exactly what we're here for, to try to struggle through with it, to go at it in depth, and finally find out, first, what agreements we do have, but, more particularly, to specify the nature of the residual disagreements as sharply as we can. Those disagreements, when they have been sharply looked at by a group of people of this sort, are likely to point to the next piece of research that has to be done.

I think, therefore, that the hassle right now is very good, and one of the reasons why we are inclined to have few discussion leaders and give them plenty of time is that we don't feel we have to hurry through or away from this kind of hassle.

KENDRICK To carry this a bit further, I have two closely related questions. One is, if words consist in part of sounds, what are the other components, or what is the other component? Secondly, do these other components exist before the sounds are perceived, or after, or during speech?

FRY Can I put that in a slightly different way? I was not saying that words consist in part of sounds. I am saying that when a listener recognizes a word, he does so by virtue of the fact that he has perceived certain features in the sound sequence that has come in to him, and on this kind of scaffolding he simply constructs his word, whatever it may be.

FREMONT-SMITH Which is just what I did wrong.

FRY No, you didn't do it wrong.

FREMONT-SMITH Well, I constructed something you hadn't said.

DENES But do you think these first primary and cardinal points could be pitch and quality, and could they be something more directly speechlike? To me, I think, this is the essence of this discussion. Does Fry think, on the primary level, even when listening to speech, you first of all perceive the features like pitch or loudness or quality of the sound, as if it were not a speech sound at all but just an acoustic stimulus. I think, perhaps, Fry denies this. Or is it that, immediately, knowing you are listening to speech, you are perceiving speechlike attributes?

FREMONT-SMITH Doesn't the question of metalanguage come in here? All kinds of signals, nonspeech signals, are given. You do it delightfully by your hands, you see, and so do you, Fry, with your face and facial expression. This tells the listener, when he is in face-to-face communication, something about the kind of message. This comes in through visual cues or other cues, or the tone tells us this is a humorous remark that is being made. I think, somewhere along the line, we will need to bring in the other aspects of communication besides merely the sound aspects, although they are involved in the sound aspects, too, in tone and quality.

GESCHWIND If I understand Fry's point, it is that it is perfectly conceivable that a receptor system could, in fact, not be able to distinguish pitch but be able to distinguish only changes in pitch.

I'm not saying this is the way the speech system works, but since such an arrangement is a possible one, you cannot necessarily say that pitch, loudness and quality are the fundamental stimulus attributes.

GOLDSTEIN But is it unreasonable to think that you don't go through the loudness-pitch-quality stage at all? Knowing that you are perceiving speech, you have a different frame of reference.

STEVENS I would like to know what experiments on loudness, pitch and quality have told us about perception of speech. I think they have told us very little.

FRY If I may take up Denes's point, I would say that we certainly do not perceive the sounds of speech in the same way as we perceive nonspeech sounds. I think there is absolutely no doubt about it.

GESCHWIND I wonder if it is necessarily true that we do perceive speech sounds differently. Even in listening to nonverbal sounds is it necessarily true that pitch, loudness and quality are the primary cues? Some other stimulus attributes may be the most important ones in all instances.

LIBERMAN It is certainly true, on the basis of what we know about the acoustic cues, that they are simply not changes in pitch in the ordinary sense of it. Change of loudness is of almost no consequence. But we do have some evidence, and I think a fair amount now, which suggests very strongly that the first decision a listener has to make is whether it is speech he is listening to. Everything that follows after that depends critically on how he has decided this.

There is evidence which leads us, at least, to the conclusion that when one perceives a particular acoustic variable in speech context, where that acoustic variable actually cues in a phonemic distinction, he hears one kind of thing, and when he listens to the most nearly equivalent, in nonspeech context, he hears something very, very different. One can't do the absolutely perfect experiment here. You can't get the perfect control, because it would be the same sounds in both cases, and you have to use either speech or nonspeech.

LADEFOGED I disagree with this, because I think that you can use the same stimulus. You say it has to be either speech or nonspeech. I've been involved in a kind of hassle with Harlan Lane over the perception of loudness (84, 85). The difference essentially comes because he has performed experiments where he has used speech stimuli, but they don't sound like speech stimuli any longer. When you just hear something going ba, ba (second sound higher in tone), repetitively, on and on, like that, it soon begins to sound just like a noise produced by a machine. This is my interpretation of why our results differ; people judge his stimuli not as if they were listening to speech, whereas I, using exactly the same noises but putting them always into a speech context, get different results.

This seems to be an indication of the kind of thing you are mentioning, that you can take the same noise, and if people know it is speech they are listening to, they behave in one way, and if they don't know, then, they behave in a different way. But, again, we are judging, now, by the way they interpret these sounds, and, perhaps, we have now pushed perception further than Fry would have wanted us to, in the first instance.

FREMONT-SMITH But the reverse of your instance certainly is true. I, and I'm sure, certainly, others, have heard a mechanical sound which they thought was somebody speaking. I have frequently thought somebody was speaking to me, and it was a sound in the house or something that had nothing to do with speech at all--it was made entirely mechanically. I think this is the reverse of what you are saying, that you get a mechanical sound which you then interpret as speech and you behave as if it were, at least for a moment or two.

FRY Yes. I think that one has to realize, with regard to Harlan Lane's experiments, this matter of repetition is a very specific thing with regard to speech. You will find that Chesterton says somewhere, if you keep on repeating to yourself the word telegraph, after a bit it sounds like snark or pobble. This is the effect he is referring to, and so I think it is a rather special case. (Laughter)

LIBERMAN I think it is important that we come to grips with some data at this point. I don't think it's as simple as all this. I think there are data available now which suggest that, in some cases, for some kinds of acoustic cues, for some kinds of phonemic distinctions, the perception of the acoustic variable which cues the distinction is very different in the speech case and the nonspeech case. There are other cases in which there appears to be no difference at all between the speech and nonspeech case, and I think that one of the most important things that anybody can do in this field at the present time is to get more information about this. Would you like me to offer an example?

FRY I think we could do with some examples, yes.

LIBERMAN Well, I can give some, if I may go to the board.

HIRSH There's nothing that frustrates discussion more than data. (Laughter)

LIBERMAN Many of you are familiar with these studies. Ladefoged, House and Stevens have done some of them, in one way or another, but I can describe some that have been done at Haskins, which, I think, will at least illustrate the kind of thing I am trying to get at. I'll discuss one that you can do in your kitchen; you don't need speech synthesizers, really, only a tape recorder.

This work was done chiefly by Bastian and some others at the Haskins Laboratories (5, 6). You record the word slit and you get a noise portion corresponding more or less to the s and a vocalic portion corresponding to the rest of the syllable. Now cut the tape so as to separate the noise from the vocalic portion, and begin to insert snippets of blank tape between the noise and the vocalic portions. The first signal has essentially a zero time gap between the noise and the vocalic portion. Then, you put in a 10-msec gap. That is about as fine as you can cut it.

FRY With the kitchen scissors, you mean? (Laughter)

LIBERMAN Yes; then insert 30- and 40-msec gaps and so on out to about 70 msec. You discover, when you listen to this series, that you hear all the signals up, perhaps, to the one with a 30-msec gap as slit and everything beyond that as split. It is quite compelling.

FREMONT-SMITH You say everything beyond that is what?

LIBERMAN It sounds just like split. That is reasonable since if you get a big enough gap, you can get a very potent manner cue. This says to you, in effect, "Close your mouth; shut up." When you close your mouth, you get a stop consonant. Why you get a p here rather than a t or a k is another matter that is really not relevant to this discussion.

HIRSH It certainly is. Those words--stlit and sklit--don't exist.

LIBERMAN I said "not for this discussion." That is really the point.

HIRSH Well, this is to perceive words. But go ahead. (Laughter)

LIBERMAN In any case, if one takes these signals now and tries to find out how well a listener can discriminate them on any basis at all--well, I had better back up and explain what I mean.

Here is the zero time gap (indicating on board), here the 10-msec, 20, 30, etc., and on up to the 60-msec gap. We want to know now, can the listener hear any difference between the zero case and the 10-msec case. To find out, we set this up in an ABX format, in which we would present, for example, the zero-gap stimulus and the 10-msec stimulus, the zero-gap stimulus being A and the 10-msec gap stimulus being B, and then X is either one or the other.

The subject's task is simply to tell us, then, whether the third signal, the X, is identical with the first or the second. If he cannot tell, he will guess and he will be right half the time, and, if he can tell very easily, he will be right 100 per cent of the time, and, if he can tell a little bit, he will be right, maybe, 75 per cent of the time.

If we do that for every pair, what we find--and we now plot per cent correct--is that the discrimination is very poor along here, and then somewhere in here (at 30 msec), just about at the phonemic boundary where we shift from <u>slit</u> to <u>split</u>, the discrimination curve goes up and then comes down again. The point, of course, is that all along one is discriminating acoustic changes which are physically equal. What we find, and not just in a <u>split</u>-<u>slit</u> experiment, but in other cases, too, is that the listener's ability to discriminate this difference is better or poorer depending upon where in this series the difference occurs.

Now, then, we started out by asking, what is the difference between the perception of speech and nonspeech? What we found is that perception of a 10-msec difference between two stimuli is very good when that difference straddles the phoneme boundary, and very poor when it falls entirely within a phoneme class. One asks now whether this is, perhaps, inherent in the perceptual system. Is this given behavior, or is this something that has been acquired in connection with our linguistic experience?

One way to find out is to ask how the listener discriminates essentially the same acoustic variable in a nonspeech context; that is, in a situation where this 10-msec difference in delay is not an acoustic cue for a phonemic distinction. What we do then is to produce a pattern that is noise, followed by a buzz, being careful to select the kind of noise that does not sound like a fricative, and a buzz that does not sound like a vowel.

We produce a series of these nonspeech signals which vary in essentially the same way that the speech signals vary; that is, we have a zero delay between the noise and the buzz, a 10-msec delay, a 20-msec delay, etc. We set up these nonspeech signals in exactly the same kind of ABX arrangement, and we undertake to determine, or Bastian in this case undertook to determine, how well a listener can discriminate these differences (6). What he found for the nonspeech case was that there was no peak in discrimination at that point which corresponds to the phoneme boundary in the speech case--a very different kind of situation.

FREMONT-SMITH Have you tried other languages, or other cultural differences?

LIBERMAN We are doing that now. The point is, in any event, there are all kinds of data which suggest that the perception of a given acoustic variable, when it cues a phonemic distinction, is quite different from the perception of essentially the same variable when it does not cue a phonemic distinction (90, 91).

Now, I said before it is really not this simple; at least, I don't think it is, because there are cases in which you don't get a difference. For example, Abramson (4), also at the Haskins Laboratories, did some work on vowel duration in Thai where it is phonemic; that is to say, I don't know whether blat means anything in Thai--I hope there are no Thai speakers here--but to say the same thing with a longer vowel means something very different. We now have a phonemic distinction based on vowel length. Notice, this is duration, again. We are talking about time, but it is a different order of time. We are not talking about milli-second differences now, really, but tenth-of-a-second differences.

Now, then, Abramson set up an experiment almost exactly comparable to this one I just described, and got Thai speakers to discriminate different durations of vowels in words like baat. He found, first of all, no peak in discrimination anywhere, and he found, secondly, that the data he got fit very well with perception of durations of nonspeech signals. He found further that there was no difference at all between Thai speakers and speakers of English. So one can only conclude about vowel duration in this case that when a listener perceives vowel duration he is perceiving nothing different from what he perceives when he listens to a pure tone.

To summarize, the point is, I think, that the perception of certain acoustic signals is different, depending upon whether the signal is in speech or nonspeech, but this is not always so.

LADEFOGED My own feeling is that if you are a Thai, you judge vowel duration relative to the duration of other vowels in the same sentence. Did Abramson test that?

LIBERMAN Yes, he did that, too.

LADEFOGED If you therefore do this experiment where you have an introductory sentence in Thai which says something like, "Say which word this is: baat," then, I would have guessed that the Thai speakers would produce a peak--that there did come a point when they could suddenly say, "Ah, yes, it has changed over from one word to another." I don't think the case has at all been proved. I would have thought that all acoustic signals were listened to in a different way when they were known to be speech sounds.

LIBERMAN Now, I wish I did have the data. Perhaps, Cooper has the answer to that. I do know that Abramson did vary the rate of the carrier, the rate of the manipulation, and found some shifts in where the tone boundary was. I simply can't recall whether he had baat imbedded in the carrier for the discrimination of sounds. Do you remember, Cooper?

COOPER My recollection is that the carrier was not used in the discrimination. But it was certainly true that the rate of the carrier affected the time duration at which the judgment switched from one meaning to another meaning, that is, from one word to another word.

GESCHWIND It seems to me that Liberman's experiments may have another interpretation. One could argue that instead of making a distinction between speech and nonspeech, you are making a distinction between familiar and unfamiliar. You chose a sound which you deliberately made as unfamiliar as possible, and therefore forced the person into a position in which he, in fact, could not make a categorical distinction.

Suppose you had a vocabulary of other sounds in the environment, such as the ticking of clocks or the mooing of cows, and were to do the same experiment; I wonder whether, in fact, you wouldn't find the same thing, that the person might suddenly say, "Oh, yes, that has changed from the noise a cow makes to that which a goat makes."

In your experiment the subject was specifically prevented from categorizing the nonspeech sound by your choice of a nonspeech sound which did not fall in the person's experience. It is possible that familiar sounds fall into familiar categories. Within speech those categories are the phonemes. We may be doing the same thing with familiar nonspeech sounds as we are doing with speech sounds. Your experiment is not designed to deal with this possibility.

BROADBENT There is a relevant point here. Of course, the drawing of the data is rough, but the way it looked was that the discrimination was better for the meaningless sounds than for the speech sounds.

LIBERMAN No, I didn't mean that; I'm sorry. No, there was a peak in the speech sounds and that peak did not appear in the nonspeech.

BROADBENT But where was the peak relative to the nonspeech sounds? You see the implication of what has just been said is that when you have a category, you discriminate better. I am suggesting that where you have a category, you discriminate worse.

LIBERMAN Oh, well, that may be true, too. Everybody is right here, I think. That's another story, you see. We have data which suggest that the vowels are perceived continuously (40, 127). You don't get this big peaking at the phoneme boundary. The general level of discrimination tends to be very high for the vowels. For the consonants, on the other hand, you can hear very few intraphonemic distinctions, but you hear very well across the phoneme boundary. For the vowels, on the other hand, you hear very well all the way through. Does that clear it?

BROADBENT Yes, I think so.

HIRSH Could we come back to the term perception? One of the reasons I don't like it particularly as a description of a process has, I think, been well illustrated in the minutes since you first introduced it. It seems to me that it consists at least of two processes and probably more. But at least at the psychological level, the process of discrimination and the process of recognition. I suppose, when people use perception alone, their meaning is closer to that of recognition than to discrimination.

It is clear, by what Liberman has said and even what you said before, that one can test for discrimination or discriminatory responses with speech stimuli. You could have said, for example, psychologist, cytologist,--are those two sounds same or different? I would suggest that whether or not the listener was correct would not enable you to say whether he had perceived correctly either one of those two words. The discriminatory response is given as a choice between the two or more alternatives, whereas the recognition response requires that a presently discriminated stimulus be compared with, associated with, some element in memory--I am saying badly what Broadbent has said much better--and that in between, as Liberman and others have shown, there are some highly overlearned discriminatory responses that can be affected by the number of categories within which one would ordinarily recognize.

Now, he says there are some speech cases where this is not true and some speech cases where it is true.

FRY What was that last point again?

HIRSH That you can set up discrimination experiments between two values on any acoustical dimension you like, and manipulate the values so as to get the discrimination data. You can also set up experiments in which you ask for immediate speech recognition. In the case of overlearned discriminations, by which I mean those values of the acoustic dimensions that are normally used by speakers and therefore have required discrimination at those particular values much more than at other values, you get an interaction between these two processes.

But I would suggest that, at least for the initial part of the discussion, particularly since there are no philosophers as such present, we might fruitfully abandon the more general term perception--I think it is more general in this context--and talk about either recognition or discrimination. I think we will find there are rather different sets of rules that apply to the two, although they do come together from time to time.

FREMONT-SMITH May I bring out another point? I realized something just now. I participate in a conference series on cellular dynamics, and if we had been in that conference series, I would have heard cytologist.

FRY Exactly.

FREMONT-SMITH But, last night, we were talking about psychology, and a number of people here are psychologists, and therefore yesterday's frame of reference carried over into today, and, I brought in something of my past to my interpretation of the sound that was given. How far back into one's past one has to go to see how this enters in, I'm not sure, but I think this is an element. You don't start off with a *tabula rasa*, you don't start off ever afresh, but you always start off in the light of your past.

HIRSH I should say your remarks pertain to the process of recognition, certainly, but if I asked you to tell me whether or not the words psychologist and cytologist sounded the same to you, I think that the preceding context would be irrelevant.

FREMONT-SMITH Much less relevant. I agree with you.

STEVENS I think, whether we talk about recognition or discrimination or perception, it still is important to answer whether we want to think about speech in a continuous signal space or in a discrete signal space. It is not going to be one or the other, certainly. But I think that which of these two should come in the fore is an important issue.

COOPER Do you mean a continuous signal space or a continuous versus a discrete reception space?

STEVENS Well, we're talking about the signals--they are the sounds. Is it more sensible to talk about the signals in such a way that we talk about certain discrete properties of the signals, or should we stick more to the continuum? It is true, when you talk about speech, you have to be bringing in a receptive aspect of this, by its nature, because, to have speech, there has to be a listener.

COOPER No, I meant a discrete recognition space. The stimulus is one thing, and measurably continuous. What is done with it during processing appears to be to put it into a succession of "boxes." But should we talk about that as the signal space?

STEVENS When we present a speechlike sound which may be so speechlike that it is speech to a listener, he automatically puts it in discrete categories. Then we have to say that the space for the sound should be a discrete space.

HIRSH No; you can always build a machine that will supply you with continuous variables, even though the listener operates on them as if there were sharp category boundaries.

LIBERMAN I think it's important to recognize here that the evidence so far suggests that for some classes of speech sounds the perception appears to be categorical or discontinuous, even though the variation in stimulus is continuous rather than categorical. So it is not all one or the other (40).

STEVENS It certainly won't be all one or the other, but I think that if we try to think of two steps, the first step being discrimination, where we work in a continuum, followed by another step which is recognition, where we work at a discrete level, we may be missing something. It may be that built into the system is a decoder that works in a discrete way for many of the stimuli that we would present to it.

LIBERMAN Oh, yes. By the way, that is not the distinction between discrimination and recognition that Hirsh was talking about.

CHASE One general point has been turning about in my mind as we took our point of departure from Fry's, to my mind, innocent opening remark. It seems that the difficulty began with reference to a point that Broadbent attempted to clarify when he stated that perception involves some kind of generation of a private event with respect to a stimulus.

When we come to the question of what are the important stimulus features, we have to define very clearly what the response categories are that we are concerned with, and, in a sense, there is no limit to the size of this class. We can impose limits on it, we can impose possible structure on it, but, in a sense, any member of this class is part of a perceptual system. Hirsh made a very helpful remark when he presented us with at least two distinct subcategories of this class, one of which is discrimination and another, recognition.

But I personally would not be disposed to a premature closure of this class. I think that there are many other ways in which we can think about the implication of an event in the outside world for the system it is presented to, and I think there are many languages for studying private events other than the behavioral techniques we know best, such as matching, which we say is giving us information about discrimination or recognition, depending on how you structure the experiment. When we enter the nervous system and define response characteristics in terms of electrochemical events, aren't we talking about perception also, in the sense that we are effecting a finer correlation of the structure of a public event, the stimulus, with the structure of a private event within the system we are presenting it to?

Isn't our whole discussion about speech perception one that requires a very articulate linkage of stimulus properties with a particular kind of response, and a particular set of tools for measuring response--isn't this an area in which we have to select and define the constraints?

COOPER You have raised one point that bears on the matter of a continuous signal space. We might recall one of the early difficulties that people had in processing speech. They were delving into the stimulus for the categories that are found at the behavioral levels, and these categories--the alleged invariants of speech--were nowhere to be found. I think you said, among other things, that we need to be quite careful to look at the different aspects of the process and give them different names, so that we can refer to them individually, without confusion; if so, Amen!

CHASE Right; very much so. It was learned that speech intelligibility did not seem to depend upon time or frequency or amplitude, but that it depended on certain time-frequency-amplitude patterns. So, at least for this case, we accept a new unit of stimulus organization, for which, to the best of my knowledge, there is still no verbal characterization other than identification of the amplitude-frequency parameters on the time axis. But isn't this really the problem we are speaking to when we ask: What are the significant features of the stimulus, with respect to a particular kind of implication? Don't we have to define very precisely what kind of implication, what kind of response, we are concerned with? Won't this really determine whether continuous or discontinuous stimulus properties are important?

GESCHWIND I wasn't objecting to atomism; I was objecting to the wrong kind of atomism. (Laughter) I think that a first derivative is just as atomistic as an absolute value. The problem is that of selecting the right set of physical values.

COOPER Now, there is an interesting implication here, namely, that there is a right set. You have thrown out--I think rightly--the physical set of dimensions of frequency, amplitude, and so on.

GESCHWIND Yes.

GOLDSTEIN But you also threw out that speech should be a special set, as I remember it.

GESCHWIND I didn't throw it out. I said that Liberman's experiment didn't prove that there was a special set of physical attributes distinguishing speech fron non-speech because his experiment just separates familiar from unfamiliar. He did not, in fact, show a difference between familiar nonspeech and familiar speech, which I think is the critical issue.

LIBERMAN May I respond by saying, first, that I did respond the first time, and I think I ought to do it now again. But we all recognize, of course, that perfect control is not possible here, because one would have to have exactly the same sound. It is relevant, though, that we have tried this over and over again with essentially the same acoustic variable--in a different context, to be sure, but the same variable--and we found a difference (6, 90, 91). Now, there are two aspects of the thing, and I think we are confusing them a bit.

One is the question whether or not there is an effect of learning, and the other is whether or not you get categorical perception. Even when we are dealing with speech signals, we don't always get categorical perception, regardless of how familiar people are with signals. For example, when we do this same kind of experiment--this split-slit kind of experiment--with vowels, we don't get tremendous peaks of phonemic boundary (40, 127), so it is not simply familiarity, you see. Whether or not one could get these peaks with nonspeech signals, I don't know. They have never been found. There is no precedent for this. But I don't think it is simply familiarity.

GESCHWIND We have, however, Ladefoged's point, that is, if you had several sentences spoken first by a native speaker you might have found that the subjects treated the vowels categorically, just as they did the consonants.

LIBERMAN We can set up conditions as between the stops and the vowels, which are, indeed, quite equivalent. Impressionistically, you can hear it. You listen to the stepwise stimulus changes that move the perception from slit to split, and you hear, in order, slit, slit, slit, and then suddenly the perception changes to split and stays there. Although the acoustic variation is continuous--or in small steps--the perception is quantal. If you listen to a series of vowels which are being changed in small steps, you don't hear that kind of thing at all. You hear every shade of difference and the perception seems very simply to keep in step with the variations in the acoustic signal.

GESCHWIND This could still be familiar-unfamiliar. Ladefoged has evidence that in order to know what the vowel is, you have to have the context of having had the talker speak first (80). In your experiment you didn't provide this information. The information derived from hearing the talker produce a sentence may not be necessary for the consonants, while it is necessary for the vowels.

LIBERMAN Well, yes.

LADEFOGED Since Liberman started by saying he would give some data, I will follow his example and mention a little more about this argument I have been having about the loudness of things (84, 85).

This is an example where people react to exactly the same stimulus. The first situation is when you know that a stimulus is a speech sound, and you are asked to judge its loudness when it is preceded by a sentence like: Compare the loudness of these two words, ba, ba (the second being louder). If the test is of that form, you then react and you compare the loudness, it turns out, in terms of how much effect you would yourself have put into making those words; in fact, your judgments reflect almost exactly the subvocal pressure--the air pressure below the vocal folds--necessary to produce these different sounds.

If, on the other hand, you don't give the subject an introductory sentence but, instead, just ask him to judge this sound relative to a scale of 10, and then present him with a series of noises like ba, ba, and so on, then, he doesn't judge them relative to the amount of effort he would have had to put into making them. If you like, he perceives them in a different way, and judges them according to the amount of acoustic intensity and according to the ordinary things about loudness judgments, exactly as if they were completely nonspeech sounds, as if they were made by a machine.

This has been my way of saying that you can use the same kind of signal; in one case, you react to it because you know it is speech in one way, and, in another case, in a different way.

IRWIN It is very tempting to relate the difference that Liberman reports between the vowels and the consonants, particularly stops, with respect to categorical differences, to production.

LIBERMAN I was hoping somebody would say that. (Laughter)

IRWIN Then, one could relate it either in terms of certain types of production making greater acoustic differences, which would be one explanation, or that production is part of perception, which, I suppose, he was also hoping I would say (laughter), which would lead to another mode of stimulation.

LENNEBERG I think there is another interpretation, though; namely, that vowels take very much longer than consonants, on the average, and you might argue that you have more time to make them.

DENES It depends entirely on what you mean by the length of the phoneme. I could make out the opposite case.

IRWIN I think, by common sense, you could define by measurements on a sound spectrogram.

DENES You really can't. The speech sound wave is a continuous event that is not easily divided into a succession of separate segments that corresponds to the discrete sequence of phonemes on the linguistic level.

IRWIN Do you really think _p_ lasts as long as _o_? Can't we agree that most consonants are shorter than vowels?

BROADBENT How about the stimuli in Liberman's experiments? The question of whether you can measure the length of consonants and vowels in natural speech may not come up, because we are talking about specific experiments with artificial speech, showing that you get the effect of consonants and vowels; in fact, where the vowels are longer than the consonants.

LIBERMAN Yes. I think that the vowels tend to last about 300 msec. I don't know exactly how long the consonants are. In the steady-state transition kind of thing, it is 40 msec or so.

HOUSE It should be pointed out, however, that you are using the one class of consonants for which Lenneberg's objection always holds.

STEVENS And that the duration you mention for a consonant does not include the gap that always occurs.

LIBERMAN Let me indicate how complicated this gets and how difficult it really is to answer Lenneberg's question. If you draw a synthetic pattern for _ba_, and if you draw exactly the same pattern but simply slow the transition with everything else the same, you hear _wah_. This, by the way, is a very distinctive difference. If, now, you erase the steady-state--and now I'm really getting into Stevens' kind of territory--what you hear is _wheep_ and _wheep_; neither signal sounds like anything but _wheep_, and they don't sound very different. So this sort of wraps the whole thing up in a bundle--you lose the distinctiveness, you lose the difference between the consonant and the semivowel, you lose everything. Given this fact, where was the consonant and where was the semivowel? I don't know.

LENNEBERG You can say this: Time distortions can be tolerated more easily on vowels than on consonants. You can lengthen the vowel quite a bit, artificially, and you don't lose any or very little of its acoustic quality, but you cannot do that with stop consonants.

LIBERMAN Yes, that is true.

GOLDSTEIN I think that an important feature of the experiments that Liberman mentioned to us is that the experimental design is such that it would tend to force out a continuity of response, if you possibly could. It seems to me that by setting up an ABX and changing only one parameter in as small increments as you practically can, you would expect to force out something pretty much continuous, unless there was almost an innate part of this system that is making a discrete decision on this.

LIBERMAN That's right. When we use the ABX procedure on some of the stops--that is, when we ask the subject to discriminate or to compare one stimulus with another--the results indicate that what he does, in fact, is to recognize or identify.

GOLDSTEIN Can you take that same split-slit test and in any way get him to get rid of the peak?

LIBERMAN Well, I don't know. I can't say that we have really tried to get rid of the peak. I would just like to add that when we ask the subject to recognize or identify the vowels, we find he tends to discriminate (32, 40). This is indicated, in part, by the fact that the response to a particular vowel will be very different depending upon what acoustic signal preceded it. From this we can infer that he is responding to the second signal somehow in relation to the first.

To get back to this distinction that Broadbent and Hirsh made between discrimination and identification, I think it is important to say that whether the subject does the one or the other depends not only on what the instructions to him are, but, rather, what he can do in the situation.

GOLDSTEIN And what the signals are.

LIBERMAN Yes.

HIRSH I would like to test whether a simple-minded generalization would be acceptable. If I may use some terms of descriptive phonetics, the vowels are different from each other with respect to what is called, for lack of better terms, place of, not quite articulation, but of the articulators, at least; and the consonants are distinguished from each other with respect to both that place and also what is called manner of production-- whether plosive, fricative, and so on. These discrimination-recognition differences are observed only with respect to manner of production, but not with respect to place of articulation. I don't find these sharp peaks in discrimination at the intervocalic boundaries that you find at boundaries that separate categories of certain consonant sounds.

Suppose that by the time you get through your next six or ten experiments, you will have ended up with a double set of consonant experiments, where some of these phoneme boundaries show sharp discrimination, and some phoneme boundaries do not. My suggestion is that the ones that do not, that is, that are like the vowels, are boundaries that are defined by place and that those that do are those that correspond to different ways of producing the consonant, that is, manners of articulation. Now, there have been six or seven people who say this is wrong.

HOUSE I believe it is wrong.

COOPER It is incorrect. The data are already in hand and they make your hypothesis untenable.

LIBERMAN He's offering it in the right spirit, though, I think. (Laughter)

DENES What data are there to contradict this?

HIRSH Wait a minute! I would have to pull out all those place examples where it is place plus the transition.

LIBERMAN That's why I said you offered this in the right spirit--it depends on what we mean by place.

At this point in the discussion the chairman declared a short recess. After suitable refreshment the conferees reassembled and the terminological issue raised by Fry's earlier remarks was reopened when Stevens asked whether reception was a useful term in discussing "the perception of speech."

STEVENS How about the word reception?

FRY Well, can this include so much more, or not?

STEVENS Isn't that what we are concerned with? My use of it includes more.

BROADBENT One advantage of reception is that it is not so likely to be used in the sense of whether one receives words or receives sounds. I don't think anybody is likely to think that there is a distinction between receiving sounds and words, whereas they might think this was so with the term perception.

IRWIN I would find reception almost as ambiguous as perception, myself. I think the dichotomy suggested by Hirsh might be useful.

CHASE It might be useful to leave the question open. I think we all agree that discrimination and recognition are useful ways of talking about the implications that acoustic stimuli might have, operationally, but there are others.

LOTZ Which others?

CHASE We spent some time this morning talking about speech signal properties that pertain to the capability of discrimination and the capability of recognition. However, we have been addressing ourselves largely to the question of intelligibility. Suppose we were to shift ground for a moment and ask what the speech signal properties are that give us information about the age or sex of the speaker, or the affective reference within which he is speaking. Doesn't this raise a whole new set of issues, which may really involve an entirely different set of pertinent signal characteristics?

HIRSH I'm sure that the acoustical dimensions that would be used or would be found to be important would be different, but I'm not sure that these two basic processes or the ways in which you ask listeners to tell you about these affective states or the age and sex of the talker would have to be organized in any more complicated way.

LIBERMAN Probably, less complicated, in fact, because the rate of flow of information at this level is so much slower than at the segmental and linguistic level. The thing that intrigues me about speech perception is that it is so efficient. But it is efficient only at the segmental level, really, where we are getting information at a very high bit rate. The number of decisions per unit time that can be made about the sex or the speaker or his mood is really very small. I wouldn't be surprised, therefore, to discover that it is, perhaps, a somewhat different problem psychologically and handled somewhat differently by the talker and listener.

CHASE Right. Actually, I raised the issue without making any commitment as to its complexity or lack of it, simply as a way of answering the question: What other kinds of processing of the information in the acoustic stimulus might lead us to a different kind of organization of significant stimulus properties?

GESCHWIND Liberman has again raised the question of the distinctiveness of speech. He cited certain examples, for example, the slowness of the blind to learn artificial sound systems or the slowness of following Morse code even by experts. I think, however, that these examples do not prove the point that speech is handled differently from nonspeech. It may only prove that those overlearned sound systems which you learned by the age of 10 years can be used very efficiently as against those which you learn in adult life.

The correct control is not an adult blind person whom one tries to teach a new sound system, but, possibly, teaching a bright blind child a new sound system. Similarly if children at the age of three or four years learned the Morse code, could they not develop rates of dealing with this which are comparable to those with which they deal with speech? I am not predicting the result of the experiment, but I at least raise the issue of alternative interpretations.

LIBERMAN I am prepared to make a prediction about Morse code and about some of these other systems, too. The prediction comes simply from what we know about what we might call the temporal resolving power of the ear. As long as you've got an acoustic auditory code which is segmented in the way that the phonemes of language are segmented, I think you are forever limited.

I don't see how you can possibly get around this, because, if you try to deliver these segments at the kind of rate at which people do deliver and receive speech, you're going to hear a buzz. You see, these phonemes, which are so nicely segmented, so discrete and commutable in the language, are not segmented in the acoustic stream. They are encoded at the syllabic level (128). At the very least, therefore, the nonspeech acoustic code which you are talking about, I think, would have to have that property. Morse code does not have it, and the signals that have been produced by the simple kind of reading machine for the blind do not have this property.

HIRSH We do have to add, I think, at least one higher-order kind of perceptual response. When you use the term recognition in context, so far this morning, most of us have in mind those experiments in which we present a relatively short item and ask the listener to identify it either by repeating it or by underlining a printed word or something like that. Certainly, there are cases in which many others here have done much more work than I myself have, in which the sequences become longer. Maybe you do not require repetition of a whole sentence or a piece of a paragraph, but you may require action based upon the content of those sentences, and here we must add to simple recognition, recognition plus storage mechanism of some kind, which, I suspect, is sufficiently different from simple discrimination and simple recognition to require that we set it off and add to it a long time.

LOTZ And comprehension and understanding?

HIRSH Perhaps, comprehension. That often implies something more complicated than I have in mind. I guess what I really mean is that we have been focusing on single words, without quite saying so, and there is the problem of the long strings, where individual words may, in fact, be lost, but where the entire string may be acted upon, in a comprehending way.

OLDFIELD I suppose we have to be careful to differentiate between recognition and discrimination, too, because recognition means recognizing something one has often seen before. That is the very point about sentences. Very often, you have never seen a sentence before, and yet you are able to behave in recognition of it. I think that recognition is an unfortunate term, because it has been used in context of this psychological experiment where you show people pictures and say, "Have you seen this before?" Whether or not they recognize it is a question of whether they saw it on a previous occasion. This is a very different thing from identifying it in the sense of relating it to a certain ensemble of other possible things that might have been, and so forth, and deciding which it is.

HIRSH But isn't that ensemble defined in terms of a set of stimuli that have either been learned before or are present, from which now one can choose?

OLDFIELD Well, is it so when it comes to a sentence you have never heard before?

BROADBENT Miller, Galanter and Pribram (103) have proved that it is logically impossible for this to be true.

FREMONT-SMITH Impossible for what to be true?

BROADBENT For you to have heard or to have experienced all the possible sentences that you can understand, because the number of combinations goes up so rapidly that it would take longer than a lifetime.

COOPER But we aren't planning, are we, to extend the <u>recognition</u> to the level of sentences? I thought that was part of Hirsh's point, that there are other processes, as yet unnamed by this group, which operate, once you get above simple identification or recognition of the small units.

OLDFIELD But, in that connection, I would suggest, where it comes to single words, what happens when you identify a single word is not merely that you perceive or become aware that you have heard that word before.

BROADBENT There is a danger of confusion here, but I must say it is something that never occurred to me before, which is the fact that I doubt if it is possible to perceive a word without having perceived it before, whereas it is certainly possible to understand a sentence without that.

LENNEBERG It seems to me, at one level, you do this all the time, because if you hear somebody speaking with a very heavy accent, particularly if you have never heard the word before--you automatically apply some kind of rule of transformation.

OLDFIELD It would enable you to identify the word, if you have never heard it before.

FRY What do you do with a new proper name?

BROADBENT Get them to spell it. (Laughter)

OLDFIELD You can see a word when it is written, or, rather, you recognize the statement when you have never heard it stated before but you have only seen it written before.

HOUSE But there is an alternative point of view that is current in the world today--that when you hear a new proper name spoken by a talker of your language, you recognize it without spelling it (63). But, if I may, I would like to harken back to what Chase said, because I really don't understand most of what has happened since he interpolated his remarks. (Laughter)

CHASE That puts a heavy burden on me.

HOUSE I suspect that what happened was that you started to make some statements about different points of view, different ways of looking at recognition or discrimination, which I didn't understand. Then Hirsh agreed about something that I thought was utterly incorrect at that point. I would like this situation to be clarified, because I think I understood Hirsh to say that there were different dimensions in the recognition of children's speech. I have never believed this to be true.

HIRSH All I meant was that the acoustical dimensions, on the basis of which you would discriminate the speech of a child from that of an adult, or on the basis of which you would distinguish the speech of an angry man from the speech of a tranquil man, would not be the same acoustical dimensions as permit you to discriminate among the speech sounds.

HOUSE Do you mean that the absolute values are different?

HIRSH Loudness is one that is used in judging emotionality, and as somebody has said, it is almost irrelevant in most speech sound discriminations.

FRY I don't think it is irrelevant in phonemic distinction, not at all.

LADEFOGED Who said that loudness is irrelevant in speech?

LIBERMAN I did. (Laughter) I'm going to say it over again, too.

HIRSH Let's review the bidding.

HOUSE If this is so, I pass. If your remarks mean that you have to plug different constants into the process once in a while, I agree, but if you mean we have to shift our frame of reference in these situations, I can not.

HIRSH Let's go back to the very beginning, where Fry started. He and some others suggested that there are lots of acoustical dimensions being manipulated simultaneously, and that it is rare that a discrimination will be based upon a distinction in only one dimension. I am suggesting that there are clusters of those cues that account for speech-sound discrimination, and that there are still other cues, not particularly relevant for speech-sound discrimination, that become relevant when you are deciding whether a talker is a man or a woman.

FRY But did you mean other cues or other clusters? A different clustering, I would agree with, but a different set of cues, I would not agree with.

HOUSE I think that's exactly what I have been trying to ask, only it has been asked more clearly by Fry.

LIBERMAN It is clear that these are different dimensions. One can judge whether it is ba, and one can judge whether it is said by a man, by a child, or by a female. There is, presumably, one set of cues that we respond to for making a phonemic distinction; I think there is a different set of cues by which we discover whether it is a man, woman, or child, or whether the talker is angry or what not.

HIRSH Fundamental frequency, for exarple.

LIBERMAN Yes, fundamental frequency.

BROADBENT This is a very dangerous argument, thought, to say that because we make the same distinction between ba and da for a man, woman, or child, therefore, there must be some physical cues which are the same in each of these cases, or which are different from those which make the distinction between man, woman, and child. What you are saying when you say that the distinction between ba and da is one which one can make independently of the other distinctions is that our responses are organized in that way. It doesn't follow that the cues necessarily are.

LIBERMAN What I hoped I said was that I thought, since our responses are independent, we would take this into account in dealing with the cues and look for those particular invariants, for example, which permit us to make the phonemic distinction regardless of who says it, then look separately, as it were, at those parts of the acoustic signal which enable us to discover who said it, independently of what he said.

BROADBENT Well, I agree. There is a certain amount of independence of these two things--in that you can make some sense out of speech from which the intonation has been removed and so on--but it is not logically necessary, and I don't believe it is altogether true empirically. I think that there is a considerable amount of overlap. The clustering may be different to make one kind of response from what it is to make another kind of response, but there is some overlapping in the content of the cluster.

FREMONT-SMITH Is part of the problem here how we are using the word cue? Are we not using it as an isolated element sometimes, and sometimes as a cluster? A cluster may be a cue and a different cluster is a different cue. On the other hand, you may say that the items which make up the cluster are the separate cues.

BROADBENT Yes, I think so.

FREMONT-SMITH If we indicated which we are speaking of, this may help to clarify it.

FRY I think this might get clearer if we take up the example of fundamental frequency. Just what did you mean by this, Hirsh?

HIRSH I yield to the expert on prosodic features, Lotz, to my left.

LOTZ Isn't the word speech used in two different senses here? First of all, it is used as something pertinent to the linguistic aspect of the signal, and the other way is the rest, which includes also the locus situation. On the other hand, I think, in the stimulus, they are similar, because, after all, this is a single variable of time according to amplitude. From the point of view of function, however, they are different, and, possibly, different selections are made from the speech signal.

For example, take the formants. If it is true that the higher formants contribute to the distinction between man and woman, or various speakers, and so on, then, these belong, on the one hand, to the stimulus side, to the similar kind of phenomena, but on the other hand, on the evaluation side, to different phenomena. I think, for analyzing speech perception, one has always to distinguish very definitely between the discrete and better understood linguistic aspect and the rest of the phenomena.

CHASE Right. I think this is an important distinction; we are using speech in these different senses. When Fry opened the discussion, he said that when we talk about speech perception we are starting with sound. From this set of acoustical features that describe the stimulus, we concern ourselves, it seems to me, with the characterization of significant subsets, with respect to particular objectives. That is why I found Fry's initial remarks so innocent, because, in a sense, this is all we have to work with. This defines the information available in full, and, in a sense, the problem then becomes the specification of subsets from a total set--subsets that are essential for particular operations to take place.

As a point of information, I wonder whether the kind of work with synthetic speech that has permitted the specification of significant subsets of the total acoustic information necessary for intelligibility has been applied to the question of other judgments, such as affect, or the sex, or age, of the speaker? Wouldn't this be a reasonable empirical procedure, to start with the total to pare down, using any rules you want to use, and see what portion of the total set is essential and what is not?

COOPER There is a hidden difficulty here. One of the things that made it possible to move fairly rapidly with synthetic speech, in getting at the acoustic cues for the phonemes, was that one was dealing with a socially sanctioned code. This is what Lotz referred to as the language aspect. If you are going to communicate in English, you have to use that specific code.

Now, you may put a lot of affect into your speech and there are a lot of other things you can't avoid putting in, but the one thing you must do is to stick to the code used in your community, or at least close enough so other people can guess what you're trying to communicate.

With affect and the like, things are not so well regimented. True enough, a woman in our culture is expected to inflect her voice more than a man does; typically, she will, but how much more and precisely when, are aspects of her speech that are not codified. This distinction between the kinds of signals that fit into a set of structured information units--language--as distinct from all the rest of the sounds that we make with our vocal apparatus is one that we ought to keep very much in mind.

You asked why we could not use synthetic methods for studying these other aspects of speech. It has been done to some extent, as you know, but the experiments are very much harder, because when you come to ask the listener to tell you something about the stimulus, it is hard to know just what to ask him.

LOTZ I have another question in connection with this linguistic subset of the speech signal. I think that it is not only descriptively one of the many possibilities, but that is the purpose of the whole speech event. If I understand correctly, a very small amount of energy goes into this aspect of speech and most of it goes for other purposes--identification of the speaker, mood, and so on. If I am correctly informed, it is an extremely small percentage.

Would it be possible, first of all, to give information about how much of the speech signal goes for speech purposes, used for speech purposes, and how it is possible to utilize this extremely small amount of information in speech signal--information which is inundated by other kinds of information--for this particular purpose, for speech evolution and, in general, for communicating?

LADEFOGED I don't understand the question. The amount of energy?

LOTZ What part is used for communication? It obviously has various components, that is, the speech signal has various components which utilize various parts of the speech signal for communicative purposes. How high a percentage of this is used for communication and how high for other purposes?

LOTZ What part is used for communication? It obviously has various components, that is, the speech signal has various components which utilize various parts of the speech signal for communicative purposes. How high a percentage of this is used for communication and how high for other purposes?

LADEFOGED I don't see how you can quantify this in any sense.

LOTZ Can you quantify it in information theory or other methods?

LADEFOGED You can only quantify by information theory things which belong to discrete codes. You can't quantify the amount of information that is contained by the fact that, for example, there is a particular voice quality, because we have no way of quantifying how many voice qualities there might be.

LOTZ Couldn't you simply give a descriptive account for the various aspects of speech?

LADEFOGED I don't think you could ask questions of people that would make them reply in terms of any discrete number of things.

COOPER Could you, for example, use some one person's inventory of known voices--of friends, for example--and manipulate these to get some estimate of how many voices he can identify, and, in this sense, learn something about the information content of that part of the speech signal?

HOUSE As Lotz stated his question, however, are not these ensembles disjoint? He wants to separate out different <u>kinds</u> of information--suggesting that there is something you could call didactic information in the signal, and some other information that is cultural information, and other information that is emotional-state information, and so on. It seems to me that these things are disjoint, and, if they are disjoint, you cannot say how much of each one is contributing to the total information in an information-theoretic sense.

LADEFOGED I don't see how you can quantify this in any sense.

LOTZ Can you quantify it in information theory or other methods?

LADEFOGED You can only quantify by information theory things which belong to discrete codes. You can't quantify the amount of information that is contained by the fact that, for example, there is a particular voice quality, because we have no way of quantifying how many voice qualities there might be.

LOTZ Couldn't you simply give a descriptive account for the various aspects of speech?

LADEFOGED I don't think you could ask questions of people that would make them reply in terms of any discrete number of things.

COOPER Could you, for example, use some one person's inventory of known voices--of friends, for example--and manipulate these to get some estimate of how many voices he can identify, and, in this sense, learn something about the information content of that part of the speech signal?

HOUSE As Lotz stated his question, however, are not these ensembles disjoint? He wants to separate out different <u>kinds</u> of information--suggesting that there is something you could call didactic information in the signal, and some other information that is cultural information, and other information that is emotional-state information, and so on. It seems to me that these things are disjoint, and, if they are disjoint, you cannot say how much of each one is contributing to the total information in an information-theoretic sense.

HIRSH Bah! (Laughter) What kind of information did I just transmit, without a word?

HOUSE I don't know. But if you ask me to count your vocabulary and tell you that this is onemillionth of your vocabulary, I can; or if you want me to say that you have only two states--belligerent or less belligerent--I can. But the latter fact has nothing to do with the vocabulary count, as I see it.

LIBERMAN I don't think Lotz's remark was very innocent.

STEVENS He is simply saying, whatever measure you use--and you might not have a very good measure--much more information is in the linguistic aspect of the event.

LIBERMAN Of course, and this is carried by a relatively very small part of the speech signal.

HOUSE That's what I don't understand.

GOLDSTEIN What do you call a small part?

LIBERMAN I don't know how small a part, but we certainly know we can produce or reproduce the linguistic information much more simply. We don't need the fourth formant, we don't need the fifth formant, we don't need all the things that are normally present in speech signal. How accurately one can quantify this, I don't know. But it is certainly true, isn't it, that the linguistic information which Lotz is talking about and which, as Stevens says, comes along at a high rate, can be contained in a relatively small part of the speech signal.

STEVENS It seems to me, if you try to synthesize speech and you want to put in a certain emotional content or, perhaps, a particular identity to a talker, the number of new things you have to do to the synthesizer is really quite small.

LIBERMAN We're looking at this in different ways. You're talking about a good synthesizer like yours, and I'm talking about a poor one, like ours. As you remember, your synthesizer has got a lot of what we might call the stigmata of speech already in it, and so has the Stockholm synthesizer. Our pattern playback does not, so it is a question of how much more you have to have. All I'm saying is that it is possible, as you know better than I, to simplify the stuff. I think this is what Lotz is saying too.

GESCHWIND This just shows that the speech signal is very redundant in an information-theoretic sense.

LIBERMAN No, not redundant, but there are different kinds of information.

GESCHWIND But, at least in one sense, in terms of the linguistic information, it is highly redundant. If you can still know that somebody is saying ah, even though you cut out half of the frequency range, that is surely redundancy in the strict sense.

LIBERMAN Not necessarily, because what you are throwing away might give you information about who is speaking, but not what he is saying.

GESCHWIND Let's restrict ourselves for the moment to the linguistic informa'ion, that is, that which you could transcribe. In that sense the message is highly redundant. The redundancy not necessarily for the purpose of carrying another body of information. The redundancy may exist to ensure that one perceives the linguistic information.

LADEFOGED It is, though. This is the whole point, that the bit you can throw away, if you kept that and threw away the other half, you wouldn't have the linguistic information left.

FRY Oh, yes, you would.

DENES There are the well-known Fletcher curves showing the intelligibility of high-pass and low-pass filtered speech (38, 39). These data indicate that speech highpass filtered at about 1700 cps is roughly equally intelligible as when it is low-pass filtered at this frequency.

FRY Could I just hark back to my second general principle, which is exactly what we are saying now--that speech is redundant, in this sense, and there is no one single feature which is going to carry all this information? But when you throw away one part, you have what you need in the rest of it. This is exactly how speech works, I think, both for linguistic and for other purposes.

COOPER I agree in general. To the extent that speech contains other information that could serve the same purpose, speech is redundant. There is no question about that. But also there is a great deal of other "stuff" in speech that is useless, because you could not recover the word or the phoneme by using that particular part of the wave form; yet it gets into the information rate, as usually computed by the engineer. This is not redundancy--just rubbish. In other words, a good deal of what appears in the speech wave form is simply not useful and can be thrown out, without loss, insofar as the language code is concerned. Some of what is left is more than enough to carry mere identification, and so redundancy exists.

LADEFOGED Supposing you took parameters of speech and said you can characterize speech in terms of, say, seven parameters--three or four formants and so on--are you going to argue that you can throw away half of those parameters? I would have thought that it was quite plain that the first and second formants carried far more information than the third and fourth formants.

LIBERMAN Linguistic information?

LADEFOGED Yes, linguistic information, and this was quite indisputable. If you play the third and fourth formants alone you understand hardly anything.

DENES Oh, I think so.

FRY This is what the Harvey Fletcher curve says.

LADEFOGED But the third and fourth formants are well above the changeover point, which is 1600 cps from my memory, and not 2000 cps. The point is that you've got to have the second formant there.

DENES For most vowels, I think, 1600 cps would cut out the second formant as well as the first formant, so you are only relying on the third and fourth.

LADEFOGED I agree, I haven't done any formal listening experiments, but you must have tried in the same way I have, playing back only the third formant of a speech synthesizer or only the third and fourth formants of the speech synthesizer. I personally just can't understand anything the synthesizer says in these circumstances.

BROADBENT That means you don't know what the cues are in the third or fourth formants, or at least logically, that is all it entails.

LADEFOGED But when I play back only the first and second formants, I do know what it says.

BROADBENT You do know what the cues are in that case. You are saying you don't know how to synthesize, using only the third and fourth formants, in order to produce the speech, whereas you do know it in the first and second formants. That doesn't prove that in natural speech there is nothing in the third and fourth formants that conveys this.

HOUSE It doesn't say anything about either case, really. I suspect that all these detailed discussions about the acoustic domain inevitably lead to problems, since you must transform this acoustic information into another code. It is what happens in this other code that is much more interesting and, I feel, potentially more enlightening, even though I don't know the nature of the other code. When Lotz's problem is put into this framework--if we don't think about information in the acoustic signal but think of some sort of simple transformation--I fail to see where his question takes us. I don't think there is more information necessarily in one or the other of these things, right now.

FREMONT-SMITH Is your other code in the central nervous system?

HOUSE I don't care where you put it; I prefer to imagine some simple-minded transformation, either in the central nervous system or in the articulatory mechanism.

STEVENS I think it is important to emphasize that when we are talking about speech, we are not talking about any acoustic signal. We are talking about a subspace of all possible acoustic signals. In a sense, you made the transformation into a new space, the transformation that House is talking about.

LIBERMAN I can't disagree with that.

STEVENS When we talk about redundancy, we shouldn't talk about redundancy with respect to all acoustic space but rather about redundancy with respect to the speech space, which is some very small fraction of the total space, and which represents the signals that can be generated by a human being. Those are the signals with which we are concerned here.

I have been disturbed in the first hour or so of these discussions that we were talking about the perception of speech by human beings without reference to the fact that these organisms can also produce the stimulus. I think we can learn something about perception of speech by studying what constraints the generating mechanism places on the possible signals that you can receive. I am sure most people here know our point of view; we really feel that there is at least one aspect of the perception of speech in which the listener is making reference to the generative rules.

COOPER I think, you ought not to assume that people know this all that well. Why not lay it on and, if anybody starts yawning and gets tired, you can stop.

HOUSE Perhaps one reason why Stevens has been reluctant to jump into this is that because of our extremely biased point of view, wherever we look at your outlined program, we see the same thing.

COOPER So do I. (Laughter) For the same reason.

FREMONT-SMITH How do you mean?

HOUSE No matter where we start, we end up by saying the same thing. We have only one story, but it can be told from various directions. I'm not always sure just where the story goes.

COOPER Well, I'm not, either, but it you think this is an appropriate time, go ahead.

STEVENS Let me first talk generally about the kind of model that might be used to represent the performance of a device having the capability both of recognizing or categorizing patterns that are presented at its input and of generating such patterns at its output. After introducing this general

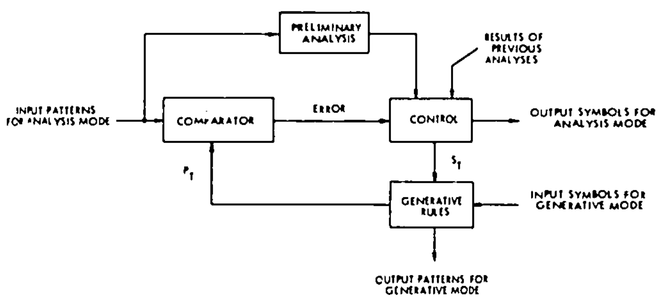

Figure 1. Diagram of a general analysis-by-synthesis model for speech processing.

model I shall suggest how it might be visualized as a model for speech perception. Let us assume that in the analysis or recognition mode of the model the input patterns are decoded into sequences of discrete symbols or categories; in the generative mode such strings of symbols are encoded into sequences of patterns.

A block diagram of the model is shown in Fig. 1. (This so-called analysis-by-synthesis model is discussed in more detail in reference 50.) The model is equipped with a set of generative rules which, in the generative mode, operate upon sequences of discrete symbols to produce output patterns, as indicated in the figure. Also shown in the figure is a component labeled preliminary analysis, which performs a preliminary, tentative categorization of input patterns in the analysis or recognition mode. This preliminary analysis provides information to a control component that oversees the operation of the model when it is in the analysis mode. The control component also makes use of results of analyses of input patterns that occur in the context of the patterns under analysis. The generative rules operate on the tentative or trial sequence of categories, S_T, to produce patterns P_T that are compared (in the comparator) with the patterns under analysis. A measure of the mismatch or error is provided at the output of the comparator. The control component makes use of this error information (together with data from the preliminary analysis and from the results of previous analysis of adjacent patterns) to make a further estimate of an appropriate sequence of output symbols S_T. A new set of patterns P_T is generated according to the stored rules, and these are again compared with the pattern under analysis. This process continues until a sequence of symbols that leads to the best match is obtained, and this sequence then constitutes the results of the analysis.

In common with other models for pattern recognition this method of analysis is essentially a pattern matching procedure, since a characteristic feature of the method is a comparison of input patterns with patterns in the analyzer. The difference between the present approach and the more standard pattern-matching technique is that in the latter an inventory of internal patterns is stored, as in a dictionary, whereas in the present model these patterns can be generated as needed, since the analyzer is equipped with the necessary generative rules.

If we now ask how this model could play a role in a model for speech perception, we observe that the generative rules for speech may operate at many different levels, including the syntactic, phonemic, articulatory and acoustic levels. If a model of the type we have proposed plays a role in speech perception, we would suggest that it could operate at several levels simultaneously. Thus at one level the generative rules could be the syntactic rules that govern the generation of sentences. At another level they could be the rules for operating on a phoneme sequence to produce a sequence of instructions to the articulatory mechanism. For purposes of discussion we shall focus on this latter level of analysis, and we shall consider the output of the analyzer to be a phoneme sequence. The input to the analyzer, then, could be considered to be auditory patterns that result from some type of peripheral analysis in the auditory mechanism. In our view, these auditory patterns bear a close relation to the pattern of instruction to the articulatory mechanism, which would constitute the output of the generative rules, and comparison between the two types of patterns could be effected at the comparator.

According to the model, then, the decoding of auditory patterns into sequences of phonemes would proceed in the model in the following way. The patterns would undergo direct, preliminary analysis or filtering to yield one or more trial phoneme sequences. These sequences then form the input to an internal generative process that calculates a corresponding trial sequence of patterns that are then compared with the original input patterns. An error score is determined and a decision is made as to whether or not the trial phoneme sequence was correct. Further trial sequences are then selected, these selections being based in part on data from the preliminary analysis and on observations of the error between previous trial patterns and the input pattern, and in part on the results of the analysis of previous patterns, taking into account the knowledge that is available to the listener concerning sequential constraints, possible syntactic structures, etc. The trial sequence that provides the best match in the comparator then constitutes the output of the analysis.

FREMONT-SMITH Output as phonemes?

STEVENS Yes, that's right. Note then that two types of analysis are going on at the same time. There is a direct analysis, and there is also some sort of feedback system.

HIRSH Are the rules part of the analysis or do they control the analysis?

STEVENS That is what I am proposing. The control component makes hypotheses about what the input sequence of phonemes was, and then makes use of the generative rules to see how well that internally generated signal now compares with the input signal. If you get an error, you make a new hypothesis and try again.

We suggest that this sort of process can be going on in addition to the direct so-called preliminary analysis. In a way, you might call it a duplex theory of speech perception. One can perform a direct analysis, study features or do pattern matching from a set of stored patterns, perhaps. In addition, however, you have available the generative rules, so that if, on the basis of direct analysis, you cannot get an unequivocal answer, you may make some hypotheses, or, let's say, make several hypotheses, and try out the generative rules until you minimize the error.

POLLACK The rules are with respect to what size unit of speech, as you visualize it?

STEVENS At the moment, we visualize them with respect to the phonemes, but we would consider a hierarchy of models of this type, in which larger units would be involved.

POLLACK How fast does an expert talker speak when he speaks rapidly?

STEVENS Perhaps ten to twenty phonemes per second.

POLLACK What is the usual considered neuromuscular loop time? Is there sufficient time to modify the individual units in this brief time? The schema seems to require some sort of auditory feedback for modification of the individual units.

STEVENS In the model it is not in fact necessary to regenerate the signal. A calculation goes on in the generative rules without overt movement.

LADEFOGED What do you generate? I'm not quite clear.

POLLACK This appears to be the crucial question.

STEVENS You calculate a set of instructions to the articulatory mechanism. You do not actually provide an output, but you do calculate the instructions that would be necessary if you wished to generate an output.

LADEFOGED In that box labeled comparator you are comparing two things and minimizing the error. Surely, the two things you compare must, in some essence, be the same; otherwise, you can't do it.

STEVENS That is true.

LADEFOGED So Pollack has a point. No one thinks you are literally suggesting that you generate muscular movements or anything like that, but what is it, what stage is it at which you are comparing one thing with another?

STEVENS You are not generating muscular movements, but you are generating instructions that give rise to muscular movements. The input patterns, however, are some sort of auditory patterns that result from peripheral processing in the auditory system. So in the comparator a comparison is being made between auditory patterns on the one hand and patterns of articulatory instructions on the other hand. We suggest, however, that a listener is capable of making transformations back and forth between articulatory instructions and auditory patterns. Thus comparisons between auditory patterns and articulatory instructions can in fact be made in the model.

LADEFOGED So neural control signals develop; is that it?

HOUSE I would avoid answering that question directly and merely point out that if you have an alternative pattern-matching kind of assumption in your thinking, this scheme is essentially no worse than any pattern-matching scheme. All we're saying is that rather than store in our heads all the messages we expect to run into in a lifetime, we are going to store a finite set of rules for generating matching messages. This is the crux of the model. If you want to throw out pattern matching as taking too much time in the nervous system, you can throw this model out along with it, because it suffers from the same limitation. But if you are willing to discuss pattern-matching, we feel that this is a more efficient model.

POLLACK I'm still not clear on the role of feedback for modification. Are you thinking in terms of acoustic or proprioceptive feedback?

HOUSE If you assume some sort of articulatory basis for these actions, there is no reason to demand that you have to talk in order to match. The control signals that are being generated on some neural level conceivably could be generated and matched on that level.

STEVENS The important thing here is that you are not storing a whole hierarchy of patterns, but you are storing the rules for generating those patterns.

OLDFIELD You've got a factory to make patterns as you require them, and you don't need to get a piece of punched tape and run through it.

STEVENS In addition, you don't have to generate all possible patterns every time, but you do some direct analysis and, of course, you base your strategy on what you have heard before and what you expect.

FRY May I ask you another question here. What triggers the action of the generative-rule component of the model?

STEVENS The control does, on the basis of some preliminary direct analysis.

FRY This was the point I was getting at. Therefore, there is already some processing of the direct analytical sort in order to make this thing go?

STEVENS Yes, I think there would have to be.

LIBERMAN You would have to make a decision of speech, for example. We would have to consider this.

GOLDSTEIN But is the generation of the articulatory movements necessary to the work of the system, or is--

STEVENS Overt generation? Certainly not.

GOLDSTEIN Well, I don't necessarily mean overt, but I mean at some time in the development of the system. How are you going to get these rules in the first place, if you had a system that never articulated?

GESCHWIND I would think, in order to learn it, you would have to, but, later you, you use the kind of feedback which is called the efferent model. This is the von Holst name for a mechanism that has a loop inside, precisely because it takes too long to produce an action and wait for a return (54). My guess would be that while, in learning, you cannot use the efferent model, you must have actually produced the act. But then, when you later on in life are using this internal model, you have a small loop, so that in some instances you can correct a word while you are in the middle of saying it. This would be impossible if you had to have the word come out and then correct it, because the loop time would be much too long.

POLLACK Can you test this notion operationally in a human being? I'm not talking about computer simulation of a person now, but how does one distinguish between these notions?

IRWIN At breakfast we were speculating about a crude, clinical representation of this. Thus, you might have a cerebral palsied child who from birth has not been able to produce recognizable, articulate speech. Yet, we can get evidence that this child apparently has normal recognition for speech. This example would seem to rule out the possibility of an original adequate production of the sounds at any time in the development of the system, and suggests that such development is not an essential.

HOUSE This kind of evidence and this kind of reasoning really doesn't settle the problem. It merely indicates that you are using a conceptual model of the nervous system that is very prevalent. Brain's description of aphasia, for example says you have a sensory phoneme in the nervous system which has to go into some central processor or lexicon to find things (15). After it has found them, it gives an instruction to a motor phoneme. The two things are separate, but it is not logically necessary to have a model in which a sensory phoneme and a motor phoneme are separate. We are saying that, at some level, they are the same, conceptually, in this model. The same equipment that is doing the identification is also capable of sending signals to a motor unit that will produce an output. The kind of objection you are raising indicates that there are people who have a fundamental difficulty that we don't understand, but, since we don't know whether they really have two different centers for language processing, we may not know whether this is evidence for or against this kind of model.

POLLACK But it does indicate the crucialness of the acoustic output in the feedback.

HOUSE I'm not sure it indicates that at all. You are pointing out that language is an acoustic phenomenon, and that we hear things through our ears.

LENNEBERG I think that it does somewhat limit the power of the model. It is an abstract model so to speak, which works but which must not be taken as an explanation of how children actually speak.

LADEFOGED For a particular child; but, as Fry said earlier, it doesn't assume that you can make these generalizations that all speech is perceived in this particular way. This seems to me the weakness in your argument; because a particular child can do this, it doesn't follow at all.

LENNEBERG I don't think that is true at all, because all children are ahead in their understanding compared to their production, and yet you get the impression that there is something in a child that allows him to make proper identification of phonemic signals.

POLLACK Let's not use up all our information for this afternoon's session. (Laughter)

FREMONT-SMITH It will be reused.

HOUSE Perhaps before we get too deep into a consideration of this model, we should immediately back off. I don't think that Stevens thinks of the model as the model to explain human speech perception. I believe we are presenting the model in a different sense. This is a way to build a machine, and, also, this is a model that is compatible with a great many of the things that we do know about speech processing. The model has a great deal of flexibility. For example, you can talk about hierarchical levels within this model; in other words, generative rules can be described at many levels. You can imagine, for example, that the generative rules are articulatory rules that relate very closely to the acoustic wave form, or you might speculate about a feedback loop in which you have a set of rules that describes the formation of phoneme sequences in the language, or still another set of rules for sentence production, for stress, for intonation, and so on.

Furthermore, there is a safety valve here, since there can be some direct analysis. You can go through feedback loops, or you can skip them. We feel that this model is compatible, for example, with some of the things that have been reported by Ladefoged and Broadbent (81) in identifying speech or nonspeech stimuli that occurred concurrently with speech. In other words, everything you hear doesn't go through all the feedback loops in the model, but decisions must be made at many points. A primary decision is whether the input is speech or not.

POLLACK What is the nature of a critical experiment?

HOUSE I don't know of a truly critical experiment.

LIBERMAN There are a lot of relevant ones.

POLLACK It seems to me that Liberman was presenting a critical experiment.

LIBERMAN Let me just say that we--Stevens, House and I--didn't conspire about this, but I was trying to leave the way open this morning for a duplex theory exactly like this. You may recall that I said I thought it was very important to compare the perception of speech signals and nonspeech signals, and what I had in mind was this kind of thing. That where one finds that the speech signal is perceived no differently from the nonspeech signal, I would suppose you might have a straight-through processing without reference to articulation. It is only where you find that the speech signal is perceived very differently from the nonspeech signal that you want even to consider this possibility of reference to the generative rules.

Beyond that, I would say that in those cases in which the speech signal is perceived very differently from the most nearly equivalent nonspeech signal, it does look as if it has gone through something like Stevens' reference to the generative rules, because the perception seems to go much more closely with the articulation than it does with the acoustic signal.

There are, I think, two parts to this. One asks, first of all, whether the speech and nonspeech are perceived in the same way. If they are, then, you forget about the loop and go straight through as in Stevens' system. If they are not, then, you need the motor reference.

BROADBENT I don't think this pays sufficient attention to the sort of parallels you get between the perception of speech and vision. I don't know of any evidence of perception of speech which shows that it is different from the sort of thing you get in vision. It is just as true in vision that the perception of a view of an object in certain illumination does not depend on the wavelength coming back from the object, but it is the context in which it is presented and so on. In that case, we do not generate ourselves an object in a certain view, and, clearly, there must be some mechanism which does not involve generation.

The evidence you have produced for the peculiar nature of speech merely shows that the kind of perception you get when a person is aware that this is speech is different from the kind of perception you get when he is not aware of it. That is also a general principle of perception. For example, if you present a pilot with information showing that his air speed is increasing, he proceeds differently when he knows he is flying inverted than when he knows he is flying the right way up. This is true of all kinds of perception.

I would just say that I know of nothing that distinguishes speech perception as a mechanism from any other kind of perception, except, of course, that it has its own particular set of outputs.

LIBERMAN Let me say that our motor theory of speech perception is not a general theory of perception. It is a theory about speech perception. I think there is a real difference between the auditory perception of language and the visual perception of language.

In the visual case, the only kind of system that works reasonably well is an alphabetic system, and there we have nearly a one-one correspondence between the phoneme, and the exteroceptive signal. We have already said you can't have this in a speech case. In fact, the wonder is that we didn't all learn to read and write before we learned to speak and listen. Because of the low temporal resolving power of the ear, the fact that the information has to be displayed in time, you've got to have an encoding of these phonemes into larger units. You don't need this in the visual case. If the auditory system were more like the visual system, we wouldn't have had to fall back on this kind of arrangement. I think that there is a special problem here. I don't think the two modalities are at all comparable for or in the perception of language.

BROADBENT In that case, what is the experimental evidence which makes you say that we do have a mechanism of this sort? The sort of evidence that we have had presented to us so far is merely that the percept or the reported output as to what the chap has heard is influenced by the statistical structure of the language and by various rules of this sort. This is just as much the case in visual perception of a nonspeech type, except that in that case, of course, the rules of the environment are different from the rules of language. I am thinking of the Ames work (1, 62), for instance.

LIBERMAN We can look at the acoustic signal, we can look at the articulation, and we can look at the perception of the linguistic structure. What we find, over and over again, in a variety of ways, is that the relationship between the acoustic signal and the perception is complex. On the other hand, the relationship between the articulation on the one hand, and the perception, on the other, is quite simple.

We think there are data which indicate to us quite clearly that there is a very close correspondence between the perception and the articulatory rules, and we are led, therefore, to assume, as Stevens and House are, that these articulatory rules get in there somehow and mediate this perception (28, 89, 93).

BROADBENT Yes, but I don't see why that requires a feedback loop. It could be just as true on an open-chain analysis, if the probabilities of the particular outputs were built into the open chain.

COOPER Yes, that is true. But you would still have the question whether the rules are organized essentially in motor terms or essentially in acoustic terms. This is one of the key points here, though not in the Stevens and House model.

One further thing about parallels with vision is that speech is quite special in the sense that practically everybody who speaks has had a lifetime of listening to himself; that is, there is a very tight and unavoidable feedback loop between what you do with your mouth and what happens in your ear. Your eyes do not have the same kind of feedback; the act of looking doesn't generate a visual stimulus.

FREMONT-SMITH But the baby does something with its ears long before it does anything of the same kind with its mouth, and the baby learns to recognize what the mother is saying; so it seems to me rather interesting, that Liberman said that it should be so much easier to read and write first. And, yet, the natural law is that language is learned first through the ear.

DENES Another difficulty is that when the child starts speaking, his muscular actions produce an acoustic effect which is quite different from the acoustic effect of the muscular actions of the adult. What is it? This is the kind of link here which has not been filled in. How does this gap, in fact, get bridged? Does the child try--in some way--to imitate the adults' articulatory movements or does he try to imitate the acoustic effects produced by the adults? The two are probably not the same--because of the very different size of vocal tract involved--and we don't know which kind of imitation takes place. Also if it is the first kind, how does the infant learn about how the adult moves his articulators?

CHASE The motor theory of speech perception shares with the model that Stevens has proposed the hypothesis that there is a very intimate relationship between the neural substrate for the reception and the production of speech. The point that we are trying to examine is: How can we move into the laboratory and look at the possible relationships between the neural substrate underlying productive and receptive capabilities with respect to speech?

FREMONT-SMITH Can I interrupt just there? You said production closely correlates with reception, but it seems to me that the ontogeny of speech does not show this. The baby can perceive long before it can produce in a comparable manner. It seems to me, therefore, quite clear that the perception is not dependent upon a productive capability.

CHASE I'm glad you raised the point because I am one of those who would like to leave open the question of what productive capabilities mean in terms of the nervous system. I'm sure you are speaking to the issue of understanding that is disproportionate to the ability to talk, which Lenneberg spoke about.

FREMONT-SMITH Right.

CHASE But is the ability to talk the only way we ought to think about productive capability? It may well be that there is a productive capability that is evolving and that has to reach a critical point before you can actually produce a motor command that will generate speech.

FREMONT-SMITH This is a very different idea of productive capability, isn't it?

CHASE Yes, I think it is. I think we need it, actually, in terms of this model.

FREMONT-SMITH But, perhaps, it is another way of saying that something on the perceptive side can be quite capable before the productive capability has really developed.

CHASE Oh, quite, and in this regard, some experiments come to mind on the ontogeny of vocalization in bird species--the work of Lanyon (86). He took a fledgling of one species of bird, exposed it to the adult vocalization pattern of another species at a point in its own maturation at which it was not capable of elaborating primary song, and then isolated it again. When the capability for motor production of vocalization occurred several months later, the young bird reproduced the adult vocalization pattern to which it had been exposed very briefly some considerable period of time before.

FREMONT-SMITH And not its own species?

CHASE And not its own species. In that way, the acoustical input had implications for the development of the structuring of productive capabilities.

FREMONT-SMITH A special form of imprinting that was across species boundaries.

CHASE Yes, and, yet, taking place without the benefit of auditory feedback for building the neural substrat of the productive gestures.

FREMONT-SMITH Yes, quite right.

OLDFIELD We have to be careful of this argument, because in some species this phenomenon takes place, and in others it does not.

FREMONT-SMITH Right; but the fact that it can occur in any species seems to me more important in this context than the fact that it doesn't occur in many.

OLDFIELD I think it goes to show that these things are a good deal more complicated than you would otherwise think. However, I don't think it gives you a typical knockdown point, one way or the other.

LENNEBERG I think the argument defeats itself in still another way, because here you have, supposedly, organisms that do have the reproductive capacity, but they definitely have not been shown to have the receptive capacity. There is no evidence whatever that parrots, who can articulate, can understand phonemes.

HOUSE But there is also no evidence that parrots can produce <u>phonemes</u>.

LIBERMAN Exactly.

CHASE I would like to return to the point that I was really aiming toward: How can we test this kind of model? The assumption was made that there is a close relationship between the neural substrate for receptive and productive capabilities--that the pattern against which we match for later recognition operations is built up out of sensory information available during early stages of learning.

Wouldn't this offer one of the more fruitful possibilities for experimental intervention--the determination of the sensory information available to the system at an early stage in its formation, out of which it might build a model for later receptive and productive capabilities? Don't you think, Liberman, that the patient we are going to present in some detail tomorrow, who has a congenital sensory deficit involving the lips, tongue and the palate, presents a natural experiment in the sense that this patient has been deprived of part of the set of sensory information available to the normal human being while learning speech. Can't we experimentally control the sensory information available during the learning of the speech motor gestures, and then test to see in what way this introduces characteristic and predictable limitations on later receptive capability?

GESCHWIND I can't help agreeing with Broadbent's view that, in the end, you must have made the sensory distinction in order to know that your articulation pattern matches it. Therefore, of what use is the articulatory pattern? The only use I can see is that once the child learns to make the articulatory match to the sensory distinction, he is then able to provide himself with an enormous amount of internal practice, because then he no longer needs to produce the overt sound and can practice on this inside loop. I think this provides him simply with a means of speeding up the learning of the process, but is not essential to it. Given practice in some other way, the child could still learn the patterns.

This mode does not predict what occurs in those aphasics who get lesions in which articulatory disturbances are prominent. Many aphasics with this type of aphasia may have no significant comprehension deficit whatever. Furthermore, if you take those aphasias in which limited lesions produce marked disturbances of comprehension, all of those forms are typically associated with retaining the phonemic articulatory structure of the language very nicely.

CHASE Wouldn't you think the developmental issue is really the critical one here? What kind of deficit does our patient have when he is learning in distinction to the kind of deficit he has when he is deprived of information after a set of rules has been structured?

GESCHWIND This evidence seems to show that anatomically the generating rules for articulation are differently localized from the perceptual rules.

HOUSE Are there any clearcut cases of aphasics who have no difficulty with production and do have difficulty with auditory perception, but can read language?

GESCHWIND Yes.

HOUSE Then, I don't see the other point as a crucial one.

HIRSH But these are pople who have their aphasia as the result of an accident sustained well past the time when language was completely learned.

GESCHWIND I'm sorry, House, I didn't quite get your point.

HOUSE I think the model is so deficient in details that I don't believe this kind of reasoning supports it or defeats it. You said, for example, that where you have a discontinuity between reception and production--when a person can receive but may not be able to produce--this case doesn't necessarily say this model is wrong. Isn't the stronger argument against this model the opposite case--when you can produce but cannot perceive. If you have learned to understand language, you can receive it through your ears, through your eyes, even through your fingers.

GESCHWIND When you understand language as an adult, whatever the modality, you are, I believe, in fact, translating it into auditory language. I believe that auditory comprehension is the final common path of all comprehension.

Now, there is a group of patients who have lost auditory comprehension and yet can read, and this would seem to be an argument against my thesis. It is not, because if you look in detail at the anatomy of those patients, in fact, the specific transducer which is part of the auditory association cortex is intact in those patients, but it is simply cut off from crude auditory stimulation. The point I wanted to stress is that the aphasic data don't suggest such a powerful link of generation with reception as Liberman's model would suggest.

CHASE How about a congenitally deaf person who has had to do a lot of organization without acoustic information and with very biased acoustic inputs? What happens when he falls into the category of having the same kind of lesion that you pondered?

GESCHWIND It is a very interesting question, which I simply can't answer since adequate information isn't available.

HOUSE In the case of disconnection--the auditory disconnection that you just alluded to--what about nonspeech processing? Are these people able to process any auditory signals at all?

GESCHWIND You are speaking of the cases of so-called pure word-deafness. Those people, at least in the cases where this has been studied, have an audiogram which is perfectly normal, as was first shown many years ago (92).

But I think, despite the existence of pure word deafness, it doesn't prove that you can, in fact, really separate the auditory step in language, as I think one could see from the anatomical arrangements. I think that in normal people there is a final step in language which is an auditory one.

BROADBENT I suggest an experiment one might do with patients of this sort, to confirm that there is some common path even in them. Conrad (27) has been showing lately that if you do visual memory experiments, using letters whose names confuse readily when heard through noise, you get the same pattern of confusions in the visual memory, although there seems to be nothing very much about the shape of the letters that produces it; I mean, if, for instance, I give you a memory span which consists of G-C-P or something like that, you are much more likely to make a mistake than if I say X-F-J or so on. Presumably, Conrad was able to show, if one was able to show further differences of this sort it would suggest that they were saying these letters over to themselves.

GESCHWIND Let me clarify. The final common path (in people who have learned language in the ordinary way) is Wernicke's area which is part of auditory association cortex. A lesion here produces disturbances in all modalities. If Wernicke's area is left intact, a lesion destroying the left Heschl's gyrus (primary auditory cortex) and which also cuts off the callosal fibers from the auditory region on the right side leads to a failure to comprehend spoken language while all other modalities are intact. Hearing is normal since the right Heschl's gyrus is intact.

HOUSE But isn't what you said in contradiction to this kind of concept? We already have the rules in there now. Suppose we put the rule in Wernicke's area?

GESCHWIND You don't get articulatory disturbance out of lesions in that location. That is the only point I am making. I am not arguing that there is not a long feedback loop, but at least perception and articulation are not in the same place anatomically.

MILNER I think, perhaps, you're pushing this auditory thing a little far. I have a lot of sympathy with the idea that the auditory processes are, perhaps, more funmdamental in language than other modalities, but I think the posterior speech area which goes a little further back from the main auditory area--not just Heschl's gyrus but the area in which you can get complex auditory responses to stimulation and so on--can just as efficiently be regarded as a more multimodal integrating area. To preempt it for audition is just a little unfair.

GESCHWIND I'm not preempting the area you have indicated for audition; I'm preempting only Wernicke's area. The main speech area you're talking about is the angular gyrus which lies behind Wernicke's area and which I certainly agree is not an auditory area. Wernicke's area is the exact analog of the auditory association cortex in the monkey, and is in the same location in man. I was not talking about the whole posterior speech area, but about Wernicke's area, specifically. You get different syndromes if you put the lesion further back and spare Wernicke's area.

MILNER You get different effects with the creation of the deficit. I agree with the great importance of the auditory processes in language. I just felt you were slightly overstating and giving a false impression of the simplicity of this. Also, I'm not too happy about this auditory association area, visual association area, and so on, as being so exclusively in the service of one modality.

FRY I wonder if I could come back to House to get in a question. Are you saying that in normal talkers and listeners it is quite possible for the whole reception of speech process to be carried out by this direct analytical circuit, or that people can receive and decode language which they cannot speak, if you like to put it that way?

HOUSE I see no reason why it must be assumed, in an incomplete model of this sort, that you have to be able to produce language to understand it, or, conversely, that you have to be able to understand it in order to produce it.

FRY Right; so I would like to follow that with a question to Liberman. If you get a situation in which somebody is undoubtedly taking the language in, without being able to produce it, do you still expect in such a case that you will not find this extra-sensitivity at the phoneme boundaries?

LIBERMAN I would have to say that a person who takes it in directly, without reference to the articulation, would perceive it differently. I would say, further, that I think he will perceive many of the phonemes less efficiently.

While I am on my feet, I'll spend a moment on the point Geschwind raised. He earlier raised the question: What does it avail the child to have the articulatory reference? I think he has already given us one answer--there is no question but that it helps the child to be able to practice. But I think it gives the child a leg up on the problem in several other ways, too. We must not forget that the speech signal is a difficult one to perceive. It does not bear a one-one relation to the phoneme. The engineer who tries to build a speech recognizer, a voice-operated typewriter, sees this immediately. He faces, for example, the problem of segmentation--the fact that these phonemes do not exist in a clear segmented form. And in other ways, too, there is an extreme lack of invariance between the acoustic signal and the phonemic structure of the language.

I think that to be able to mimic and match helps the child to decode this complex signal. In this way he finds the appropriate articulations, and these are very nearly invariant with the phonemic structure of the language. The child is helped to identify the signal by being able to mimic it, because he is able, thereby, to use his differential sensitivity. Having mimicked, all he has to do, then, is to judge same or different. This is easy. Having done that, he discovers that in order to mimic one signal, he has to use his lips and not his tongue. To mimic another, which is acoustically quite similar, he has to use his tongue, not his lips. Thus, he gets to know a difference that is distinctive; indeed, it is categorical.

LENNEBERG According to this, the child with the harelip or cleft palate should have a difficult time in discriminating between ds, bs, and ms, which is definitely not the fact.

HOUSE I don't believe you must make that assumption.

(The first session ended on this note of disbelief...)

SESSION 2. Part 1 - Speech Behavior and the Structure of the Linguistic Code

COOPER The next topic in our program is a continuation of the discussion on speech perception, under the heading, "Speech Behavior and the Structure of the Linguistic Code." Donald Broadbent will serve as our discussion leader. He can steer the discussion in any direction he wishes--and is able to manage.

BROADBENT Well, I have had to modify slightly the sort of thing I was thinking of saying, in light of where we finished up this morning. I suggest that the most profitable line to develop, perhaps, is to go along with this question of generating speech, starting, perhaps, with the relationship between utterance and perception, and then getting to more purely questions of utterance as the afternoon wears on.

As we heard this morning, there are clearly connections between speaking and listening. If you get somebody talking, he selects words, not, of course, at random, but in accordance with the sort of constraints which also affect the speech he has been listening to, and this makes it seem likely that there is some connection between the mechanisms involved.

In addition, there are a number of experiments now in which interferences between utterance and perception have been noticed. I would just like to draw attention to some of them. For instance, the work of Kalsbeek in Amsterdam now may not be known to all of you, but he has people writing compositions while reacting simultaneously to noise, in a choice-reaction task. As the load from the choice-reaction task goes up, so the composition gradually deteriorates, becoming rather more stereotyped and less elegant, and gradually disintegrating until it becomes nonsense, essentially, consisting of isolated elements without any interconnection (121).

A modification of this kind of approach, which has been developed by Baddeley (2), lends itself to quantitative analysis. You get somebody trying to utter, say, the months of the year or the letters of the alphabet in random order, and you find that the faster you try to get him to do this the less random he is. You can compute the information, for instance, in the digram structure of what he is uttering, and you find, as you press him to go faster, the information drops. You find that if you get him to handle some other information simultaneously--for example, sorting a pack of cards--then the information of his utterance drops; the more categories he has to sort the pack of cards into simultaneously the more stereotyped his utterances become.

In other words, there seems to be, again, some sort of interference between receiving, in one task, and uttering, in another; so that there is, I think, adequate evidence for something in common between the mechanisms in the two processes. Clearly, this is the sort of thing which the model we heard about at the end of the morning was intended to deal with. I agree, of course, that there is a great deal in the view that you have a series of events going on inside the person, which depend partly on a process rather like that which goes on when you utter speech yourself, and which takes into account the evidence that is coming from the senses, in order to try to match it.

I think, as I was just saying to Oldfield before we started, my main disagreement with the diagram that appeared at the end of the morning (Fig. 1) is that I am very doubtful about the error line that goes across from the matching back to the control. I suspect that, having got some information from the senses at one stage, you then use this to try to help you predict what is coming next, but, if your prediction is wrong, this is not necessarily then detected, so you do misperceive words because of strong prior prejudices.

Now, what sort of factors are controlling the sequence of controlled statements or central correspondences--or whatever you like to call them--the things which, perhaps, in the model this morning would have been instructions to articulate--although I am trying to avoid this, because I don't believe that part of it? Well, clearly, word frequency is one of the effects in question, because you do tend to utter words spontaneously which are more frequent in your experience in the past, and, indeed, this is almost tautologous, when you think about it closely enough. You tend to make statements in accordance with grammatical constraints. You tend, perhaps, to have subroutines or what have you built in so that you can transform the same kernel of meanings into various alternative statements, such as "I am bored by the discussion leader," or, "The discussion leader bores me," both of which are simply equivalent statements.

FRY But not true. (Laughter)

BROADBENT But one of them, anyway, can be translated or transformed into the other very easily. Many of you will be familiar with Professor Milliar's statement on this point (mispronouncing the name, Miller, and the word, familiar).

That, actually, leads me on to my next point, which is the planning ahead of the utterances, where the statement that you are about to make may have an effect upon the statement you are making at the moment. You noticed me mispronouncing Miller's name, which was coming later, into a word sounding similar to one that I was trying to utter at the moment. There are, of course, features of natural speech which show this kind of organization time being taken out from talking and so on, and show the process of planning ahead what the statement is going to be, interfering, possibly, with the momentary utterance. The sort of thing I am thinking of is Goldman-Eisler's work (46) where you may actually get a pause in the statement because what is coming is something which, in general, the talker would not predict. I say this with every intention of having somebody jump upon me.

Another point of considering a sort of long-term, before-and-after mechanism in utterance, is the question of short-term memory, which Hirsh raised a bit this morning. Again, to select the correct utterance at any point implies a memory for what has gone before, which may be limited and restricted, so, although certain statements may be in accordance with grammar, it becomes extremely difficult to make them. Again, Miller's work is relevant here. I expect many of you have heard the sample statement, which I shall probably not be able to repeat--"The race that the car that the people that the obviously not well-dressed man approached sold won was run last summer." This sentence is grammatical but, in fact, extremely difficult to utter, as I have just proved, or to perceive, because of the strain it places upon short-term memory. You have to remember the start of each clause until the final word appears later on, despite the fact that you are now dealing with another clause which nests within the previous one, and, when you get up to about four nesting clauses like this, then, the strain on short-term memory is too great and you lose it.

HIRSH How can Germans converse? They do this all the time.

BROADBENT I wonder. (Laughter)

GESCHWIND I don't think that the extreme syntactical nesting of which German is capable appears often in ordinary speech. This kind of thing is done in writing, and in the extreme written forms is difficult to understand, even for many Germans.

LENNEBERG I disagree with that. (Laughter) I think you can most certainly prepare yourself for elements that come later, and, in English, you do this just as much as you do it in German. It is just that the level of grammaticalization is different, or the grammars that we have available don't make this point quite clear. I think there are dependencies even in English that go over stretches as long as in German. I think there is, basically, no difference in English and German, except in a somewhat artificial grammatical model.

OLDFIELD Except that in German, you have to wait for the past participle.

LENNEBERG Yes, you do that, but in English, you can also start sentences and interpolate another sentence, which may be very long; in fact, you may have some twenty elements until you come to that dependent.

OLDFIELD Yes. What I was saying was that it is simply a matter of which part of speech you have to wait for. This applies to the Germans, who put the verb at the end, and not to us, because we don't have to wait for it.

BROADBENT This point is related to another kind of point which I thought we might fling in, which is the way by which short-term memory gets affected by the structure of the items contained in it. Going back to Miller's business of having six or seven chunks in short-term memory, if you have words which form cliche statements, of course, you will be able to remember more words, although, in fact, you may be dealing only with six or seven independent units, that is, units which could be interchanged and which hang together like counters without being broken up (99). This may, perhaps, have a relationship to the sort of structure you get in German, where the intervening statements, when you are going to have to hold on for a verb at the end, may be closely tied together by being familiar sequences, by being redundant once you have had the first word, essentially. This, of course, is only partially true.

The last point that I wanted to bring up was the relationship of monitoring feedback to utterance, which brings us, really, right back to the beginning and the relationship of utterance to perception, because it is so clear that there is some sort of feedback when you are speaking.

Even though I don't like the error line in the model that came up this morning, I remember Ladefoged doing a demonstration on me in front of a large audience once, of counting from nought to ten, with delayed feedback. This was the situation, you remember, where, although I was expecting it, it worked on me. You talk very fast. You start off and you say one, and you don't hear anything, so you say two, and then you hear one, so you realize that is where you've got to, and you say two again, and then you hear two, so you say three.

LADEFOGED Yes. You were a very nice example; you came out rather better than most. This was the standard kind of business of delaying feedback by about a quarter of a second. Broadbent is one of the people who actually do say, one, one, two, two, three, three, four, four, five, five, exactly like that. You can, in fact, superimpose records of his utterances. If you put them one underneath the other they fit exactly; he repeats the number after the one that he has just heard.

BROADBENT Well, this is just an example of the sort of way in which higher-level feedback is undoubtedly operating, and when you hear something which corresponds to a word rather than, perhaps, a sound, you know whereabouts you are in the utterance and use this as a control for what you are about to say. Of course, to the extent that you are doing this, you are bound to get the sort of interferences with perception that I spoke about earlier. Nevertheless, I would feel, myself, that there was more to the relationship between perception and utterance than simply this kind of monitoring feedback; that, rather, it is because the sequence of internal states which issues the control signals for the utterance is essentially the same as the sequence of states which follow when you are getting a series of input signals. This is where the common element between perception and speaking comes in.

Well, I don't want to talk too much. I will give just one last example of an experiment which is slightly relevant to this point, although I am sure it would not be regarded as a crucial experiment between various models of perception and utterance. This, again, is an experiment by Baddeley, who took advantage of the fact that on, at any rate, English Post Office teleprinter keyboards, the digits and letters are on the same keys, like an old-fashioned portable typewriter (2).

Now, it is true that typists can type more rapidly sequences of letters which are frequent in the language, just as many other types of perceptual effects depend upon experience of high-probability sequences. We know what the high-probability sequence of letters is in English, because we took a soap opera from the BBC and put it through the Post Office computer to find out. Using a teleprinter keyboard, it is possible, of course, to require people either to key out common or uncommon sequences of letters, or, alternatively, to key out sequences of numbers which actually use the same keys as common and uncommon sequences of letters. If you do this, you find experienced teleprinter operators, as Baddeley showed, do not show any advantage in typing the sequence of keys, which they have frequently typed before, when they represent numbers; that is, making a highly practiced sequence of movements is no help. It has to be a sequence of movements controlled by the sequence of control instructions corresponding to the letters.

Well, this really concludes the various topics that I thought we might discuss this afternoon. Going over them again, there is the relationship between the utterance and perception; the various features that come into controlling the utterance and also having an effect on perception, like word frequency, grammatical restraint, relationships between kernel meanings and various transformations; the planning ahead of statements; the role of short-term memory; and the importance of feedback.

I wonder whether anyone has done any studies on the effect of postprandial feedback? (Laughter)

COOPER Or feedforward? (Laughter) It seems to have a quieting effect.

We all know something about the effects of delayed auditory feedback, which is a kind of automatic process, but what kinds of experiments have been run on interruptions of a controlled kind, that is, interferences external to the person generating the behavior and aimed at finding the transient time of the system? How long after you interrupt the continuing flow of a skilled act is it before the effect of the interruption takes place?

BROADBENT There is quite a lot of information from which one ought to be able to extract this. I'm wondering if Chase has anything?

CHASE I have some information on this point. The effects of imposing sampling constraints, such as taking away sensory feedback information for a certain period of time are quite different from those which result from delayed auditory feedback. These are very different types of temporal intervention in the sensory accompaniments of motor activity.

To study this sort of question, we have shifted ground, as you know, and looked at some simpler kinds of motor activity. I would like to tell a bit about this story because I suspect it is pertinent to the speech case, and, parallel experiments probably could be done for speech.

The system we have worked with tracks positions of the index finger in space. The subject monitors his movement with respect to a target on an oscilloscope screen. We could delay the return of his visual feedback, but we have not.

However, I think the work of Smith (125) and others on delayed visual feedback leaves little question about the fact that when visual feedback of ongoing motor activity is delayed, qualitatively similar effects are observed in the ongoing motor pattern that we observe for speech under conditions of delayed auditory feedback.

In our system the subject is trying to keep his finger on target. He is not watching his finger but watching an oscilloscope screen. We have one beam fixed at the midvertical position indicating the fixed target, and the other beam of the scope is actuated by a rotary-motion potentiometer whose output voltage is passed through clocks which permit us to display the error signal intermittently (23). If we present 15-msec pulses indicating the direct analog of where the finger is with respect to target, and then make our subject wait one second before he gets another 15-msec pulse, there are changes in the pattern of movement, but very little impairment of the subject's ability to maintain his movement on target.

COOPER This would correspond, I suppose, to the chap who listens only now and then to what he is saying?

CHASE Yes. I think it indicates how much greater the tolerance is for going without information than it is for receiving incorrect information.

GESCHWIND I am interested in the wonderful sideshow demonstration which Ladefoged and Broadbent performed and which Broadbent mentioned earlier in his discussion. It seems to me to have implicit in it a theory of why the feedback is disturbing, a theory which almost suggests that it is not the feedback feature which is the disturbing one.

In the experiment which Broadbent cited he said that if he heard the word two come back to his ear just before he was about to produce the next count, you would then say three even if he had already said three once. This suggests that he could have run an identical experiment in which he did not feed back your own voice but simply fed in numbers from the outside. This would probably have interfered in the same way, that is, just at the moment when he was about to say something, a sound might come in from the outside, and therefore he would tend to produce what was essentially a response to that sound. In this case you would not need to introduce the concept of interference with a feedback loop. One could simply say that you were introducing another stimulus into the same system which you were using for speaking. This experiment, therefore, would be identical with those in which you introduce an interfering sound which is not controlled by feedback.

BROADBENT Well, I have particular feelings about this, but I think Chase has something to say.

CHASE This experiment has been done in the case of speech, and it is interesting to observe that the two situations are not comparable at all (18). They are so different that it poses issues of real magnitude. When you have a subject read a story and play that back while he is reading it, it does interfere. There are alterations in the pattern of speech-motor gestures, but not nearly as profound an effect and not qualitatively the same kinds of effects that obtain when you impose a fixed delay in the feedback.

COOPER Do you attribute that to time _per se_?

CHASE I am tempted to attribute it to time. The fact that the temporally distorted feedback is specifically linked to the unfolding motor gesture on the same time axis, without any ability to alter the phase relationships of the two, is important.

COOPER Suppose we could do an armchair experiment using Denes' vocoder, with which we could easily flip the channels from right side up to upside down and also delay the output so that the feedback would be alternately the speech delayed, or the same sound pattern delayed, though no longer speech because inverted. What would you predict about the effect on the person speaking into such a system?

CHASE I would predict that that would have virtually the same effect as the delay in real speech--again reinforcing my feeling that it is the timing that is critical here.

DENES Timing of what?

COOPER What, indeed, since we would have destroyed the articulatory tracking that some of us are still inclined to consider important in following speech?

CHASE Given two formants which are the same on the time axis, I will just define the very smallest two-dimensional space that I can fit them into. In other words, from the point of view of time, the stimuli are identical, and you are leaving unchanged the time relationship between ongoing speech and the displaced feedback. I think these are the critical parameters, accounting for the delayed auditory feedback effect.

DENES Would you think the same effect would happen if, instead of that lower spectrogram, you didn't have a speech-like spectrogram at all, but a voice-operated switch which switched on a delayed buzz?

CHASE Of exactly the same duration?

DENES Yes.

CHASE These ideas are coming dangerously close to an experiment you could actually do. (Laughter)

DENES I have long tried to do an experiment to find out something slightly different. I wanted to find out just what in the speech signal, on a kind of gross level, produced this delayed auditory feedback effect.

People here know that a vocoder is a special speech-processing device which analyzes the speech wave, and then synthesizes another sound wave, controlled by the sound features extracted from the original speech wave. The action of the vocoder enables you to separate those factors in the speech wave that are due to the action of the vocal tract, that is, movement of the tongue and lips, and those features of the speech wave which are caused by the action of the larynx, that is, the phonation mechanism.

I was trying to produce a delayed auditory feedback situation in which a vocoder is used, but only the phonatory part or the articulative part is delayed, to see if there is a difference between the effects and which one is bigger.

HIRSH Well, what did you find? (Laughter)

DENES I said I had long wanted to do this, but I never got around to it. (Laughter)

HIRSH We did an experiment. We don't have a vocoder, but we did low-pass filter the side tone. I think it was something of the order of 600 cps low-pass, and not a very sharp drop above. We got no reduction in the amount of disturbance. I think this accords with Chase's observations, and I suspect that it is the laryngeal part of speech that is interfered with or is interfering. You can introduce time distortion of an upsetting kind into the laryngeal signal without regard to what the articulators are doing, if that means anything.

FREMONT-SMITH But, on this theory, would you expect the same interference if you had the articulatory part alone?

HIRSH I don't think that is a proper question, if I may say so, because the articulatory part is a kind of modulation on a basic rhythm that is set up by vocal energy.

STEVENS There would be no interference because the articulatory part is silent. (Laughter)

CHASE I would like to tell a simpler story again, in the hope that some of the observations made here might be suggestive of parallel experiments with speech.

We were interested in whether the changes noted in speech under conditions of delayed auditory feedback were due to unique features of the auditory feedback control system for speech. To investigate this question we looked at a nonvocal motor task, as did Kalmus, Fry and Denes (65). We studied (21) a sequential motor task of the upper extremities--key-tapping--and looked at three kinds of sensory events with respect to the effect of delay.

The first, in analogy with delayed auditory feedback in speech, was delayed auditory feedback and key-tapping. The auditory event in this case was a click. The subject was asked to tap in groups of three taps, and he was trained for rate and amplitude. Subjects learned rate and amplitude very rapidly. By using a strain-gauge transducer, we can get an analog recording of the time-amplitude patterns of his tapping. Each time the subject taps, he functions as a switch, triggering the release of the click to his ears, at first synchronously with the motor event. Then, using a tape recorder as a delay line, exactly as we do for speech, we present the click 200 msec after the motor event has been initiated. The key-tapping under conditions of delayed auditory feedback shows an increase in amplitude, a decrease in rate, and repetitive errors, which are almost invariably single repetitions.

Now, what would happen if we were to have a light flash, either synchronously with the tap on the key or 200 msec after the tap? Suppose we activate a solenoid, which provides a tactile stimulus to some distant part of the body. Let's say our subject is tapping with his right index finger, and, each time he taps there is a synchronous tactile stimulus delivered to the left arm. What happens when we delay the tactile stimulus 200 msec with respect to the onset of the motor act, or when we delay the light flash? All of these delayed sensory events function in qualitatively the same way, and produce the same disturbance of key-tapping. This series of experiments impressed us with the fact that central processing functions were utilizing the various sensory events that we generated as time markers and not respecting modality (22). The specific distortion of phase relationships--so that some kind of decorrelation probably occurred--probably accounts for the impairment of regulation of the ongoing motor activity. We might expect that this would be the case for speech, also, in which case, there is a good chance that the fine architecture of the stimulus that permits us to recognize speech might not be as important as the time constants of the system here.

FREMONT-SMITH To what extent is the orienting reflex coming in here to interfere? Each time a new stimulus comes in you get an orienting reflex, and might this not in itself, as an attention-absorbing factor, interfere with the recognition of the speech? Maybe, I'm just saying the same thing in other words that you have been saying, but I wonder whether the concept of the orienting reflex doesn't belong in this situation?

OLDFIELD They would adapt the orienting reflex very quickly to a regular series of events.

FREMONT-SMITH It would depend on how much of an interval. If it came each time it was new, then, it became an orienting reflex. It is only when the orienting reflex becomes adapted that you lose it. But I don't know what the interval is.

OLDFIELD For adaptation, in a series like this, I think, the orienting reflex is supposed to adapt very readily.

FREMONT-SMITH Was this carried on beyond 60 msec?

CHASE Yes, considerably.

LENNEBERG Did you compare the delayed times in different experiments, one with one-fifth delay, one with other delays? What was the worst delay?

CHASE That is a very interesting point. Whereas most people make the same observation with respect to the delay in auditory feedback that produces maximal disturbance of speech by virtually any measure you pick--and get a value of between 150 or 180 msec--the people who have been studying key-tapping have been finding that, as you increase the delay in 100-msec steps from 100 msec up to 1000 msec, there is progressive distortion of the time-amplitude pattern of tapping. I don't know of any studies that used delays beyond 1000 msec.

HIRSH Is this for all three sensory inputs?

CHASE No, this is just for the delayed auditory case; it hasn't been studied as yet for the other modalities. But I think one issue here that has to be looked at very carefully is the correlation between the time of the significant unit gesture, the rate of elaborating unit gestures, the duration of the sensory event, and the actual delay time. If this could be done for speech and a nonvocal motor activity in parallel, some of the differences--such as the delay time producing maximal disturbance in speech and key-tapping--might be explained in terms of significant differences in other time constants of the system.

But let me just pose this question: Is it possible that doing this kind of experiment might shed light on what the significant unit of motor gesture is for speech with respect to control? I think this was implied in one of Cooper's earlier remarks.

HOUSE Isn't this conclusion one that people who work in this area from the speech point of view eventually always arrive at?

CHASE Most people conjecture that the delay time that produced maximal disturbance is related to syllable length.

HOUSE A syllable length which, in turn, can be related to articulatory activity.

CHASE Right. But it might be useful to structure speech-unit gestures of different orders-of-magnitude of time to see whether one could get variation in other time constants. If we were operating on a different order-of-magnitude of unit gesture--even if we just programmed our experiment that way--or if we altered the rate of elaboration of a fixed unit gestures, would we then get different delay times producing maximal disturbances?

OLDFIELD I think it might be well to introduce some experiments done by Treisman (131) at Oxford. She studied delayed auditory feedback with verbal material which was of the Miller-Selfridge (101) kind, or statistical approximation of various orders. It was quite plain that the amount of disturbance was a function of the order of approximation; so that, I suppose, one can't think of any very simple peripheral sort of loop as being the only one which comes in question in the disturbance. This evidently has gone up to the stage where questions of theme and context and grammar and so forth enter into it, and there were marked differences in the types of motor disturbance.

FRY All those who have spoken on delayed feedback, know the kind of generally disrupting effect it has on whatever you are doing. This wouldn't, to me, be quite an indication that the chief factor in this disrupting of speech might not still be the relation between the delay time...

OLDFIELD I think that is perfectly true. I'm saying this is not the only factor, because one can show further effects.

DENES The relation between delay time and what, Fry?

FRY Well, the time represented by unit motor gesture, whatever that means. (Laughter)

(At this point Denes remarked that he was surprised that the low-passed speech materials described earlier by Hirsh had a disruptive effect in delay experiments. After a rapid series of exchanges concerning whether or not syllabic or articulatory information was present in such a signal, the discussion continued.)

STEVENS Syllabification has to do with opening and closing of the vocal tract; which, in turn, has to do with increasing and decreasing the frequency of the first formant; which, in turn, it can be shown, has something to do with change in overall amplitude, even though the glottal source remains fixed in amplitude.

DENES Perhaps, the experiment would have been more significant if the cut-off frequency could have been even lower.

HOUSE Has anyone done any of these experiments with whispered speech? This seems to be the first experiment to do, as long as you are interested in doing this without elaborate equipment.

DENES I was going to do it with the excitation function of a voice-excited vocoder, where you have information only about the presence or absence of vocal-fold activity plus the pitch of the voice.

HOUSE Can you do it immediately without a vocoder--perhaps with a trained whisperer?

LADEFOGED Haven't you tried this? My own observation is that it has the same amount of effect on me whether I whisper or whether I phonate, more or less. Has nobody done this more systematically?

HOUSE Not that I know of.

CHASE How about delayed feedback of whistling? I think this gets more at this issue.

OLDFIELD Could I ask, among the people who have done experiments on delayed speech, how many individuals they find, shall we say, upon whom it has virtually no effect at all? We certainly find quite a lot of people who are unaffected by it.

FRY Completely unaffected, or who manage to carry on? Speakers of English?

OLDFIELD Yes. I have observed some ten under those circumstances.

FRY I think this thing is tied to language. There does seem to be, somewhere, a connection between the maximum effect of delay and the syllabification. I have never tried this extensively with French-speaking people, but I tried three subjects, and, really, there was virtually no effect of delay, partly for the reason that Lenneberg has just advanced, I think.

OLDFIELD You tried different delays?

FRY Yes, I tried a range of delays, but it didn't seem to produce the effect. But I wonder whether this may be partly due to the fact that in French, you do not have a strong tonic accent, so that it is possible that the syllabification comes up in a rather more regular cycle than it does in English? I wonder whether this has anything to do with the decrease in the effect of the delay?

GESCHWIND I have seen at least one speaker of English, who was not an Englishman, who was quite unaffected by delayed auditory feedback.

HOUSE This doesn't surprise me, but the one-in-ten surprises me. I have seen a number of subjects and very few of them were relatively unaffected. In every group you will find a small number of people who are terribly affected and cannot seem to adapt at all.

CHASE The variations in disturbance of ongoing speech under conditions of delayed feedback seem to me to raise two questions which overlap some of the considerations of the morning. Let's take the case, no matter how infrequent it may be, of the person who shows very little disturbance in his speech under conditions of delayed feedback. How might we explain this?

It seems to me there are two classes of explanation we could use. One is that this chap isn't doing the pattern-matching--isn't taking an error signal and comparing it to a standard--but is functioning very much in open-loop fashion. What goes out has been centrally programmed; it goes out and the sensory consequences are just not very important with respect to the programming of speech-motor gestures. Another explanation is that there is an hierarchy of importance of all the possible sensory channels functioning during speech, and that we have intervened in a sensory channel that is not very important for this speaker.

It seems to me that a consideration of these two possibilities is really quite important. I would like to make a few comments about the two without attempting to decide between them. With respect to the former, consider the possibility that there has been a central program and it is going on essentially without monitoring, in open-loop fashion. Is it possible that when speech is being learned it functions completely as a closed-loop control system, and that after speech becomes highly learned it is capable, at least under certain circumstances, of functioning in open-loop fashion with very little need to monitor? Indeed, we might think about learning in terms of progressive shifts from a closed-loop to an open-loop type of operation.

Another way of thinking about this point is that we have the option available at any time of closer scrutiny of the output, and the option of shifting from closed-loop control. If this is the case, then, an interesting question arises: When would we exercise the option to do one or the other?

With respect to the point that there might be a hierarchy of dependence on different channels of sensory feedback as a consequence of the way speech motor activity is learned, and that this hierarchy can be different for one person as compared to another, I would like to review briefly some experimental data (21). We examined the effect of delayed auditory feedback of speech, and delayed auditory feedback of clicks with respect to key-tapping, using the same population of subjects for both tasks.

The question in our minds was: Are there people who are vulnerable to delayed auditory feedback independent of the motor program, or are there individuals who are extremely sensitive to delayed auditory feedback of speech who are not vulnerable to delayed auditory feedback of clicks for a non-vocal motor activity?

HOUSE Can we choose sides before you tell us the answer? (Laughter) I fail to see any relationship between the two sets of activities and I don't think that one casts any light on the other. Of course, I'm basing my conclusion on Chase's earlier data from speech and nonspeech experiments where there didn't seem to be any relationship.

HIRSH But, essentially, there were disturbances shown in both.

HOUSE Yes, there were motor disturbances shown.

CHASE And they are qualitatively the same.

LENNEBERG Was the delay time different?

CHASE No, we used the same delay time.

HIRSH But he said that the effect increased with delay time.

LENNEBERG Yes, that's right.

HOUSE In one case.

CHASE Actually, I was reporting Rapin's experiment (114), and the thing that makes me reluctant to decide about the equivalence or nonequivalence of the two cases is the failure to control all the time constants of the system in a comparable way. For example, I don't know what the mean duration of the tap was for her subjects, or the rate of elaboration of the unit gestures for tapping. I think it may well be that the differences with respect to delay time that produce maximal disturbance are explicable in terms of differences in the duration of the unit gesture and the rate of elaboration of unit gestures.

GOLDSTEIN A second is certainly a very long delay. You could tap several times before the second was over.

CHASE Right; although the subjects may have been trained to a very slow rate of tapping, and this is exactly what hasn't been looked into.

COOPER Is my impression right that the effects of delayed feedback are qualitatively quite different, depending on whether you are talking ad lib or trying to tell a story, let's say, as against reading a story?

CHASE Yes, they are; that is correct.

COOPER This argues for interference at different levels, even for the same person, if the task is different.

CHASE Exactly. I think this is very pertinent to the issue of whether we can switch from open-loop to closed-loop operation, and what the contingencies might be which would direct us to do one or the other.

HOUSE I gather from what you just said in the past few moments that you can train a subject to be driven in a tapping experiment. Can you train a subject in such a way that you can get into the loop, as it were, and change his behavior?

CHASE By delaying a signal.

HOUSE This is very difficult to do with speech, although it can be done to a minor degree. People have tried to set up experiments to do this and usually have not been very successful.

CHASE Let me make sure I do understand you. What would an example of driving be, for speech and for tapping, as you understood that to happen?

HOUSE An experiment with speech would try to vary the delay time systematically and predict the fine structure of the changes in the speech of the subject. People have not been able to do this very successfully.

COOPER You mean, to interfere with speech?

HOUSE No, to interfere in a predictable way with speech.

COOPER I wonder how many of you have listened to your muscles growl when you were trying to talk? If you connect a myographic pickup to an amplifier, the muscle potentials can be heard as low-frequency noises that begin, typically, about a tenth of a second before you start to talk. It is an extraordinarily distressing phenomenon. This is not feedback from what has gone before, but it's just as disconcerting as feedback. Just when you are ready to say something, this sound beats you to it.

HIRSH Insofar as the usual direction produces stuttering, would you recommend this as a possible prosthetic device for those who do?

BROADBENT This is how Cherry found his effect of loud noise causing some stutterers to restart talking; explaining something of this sort (25).

HIRSH That was just sheer masking, wasn't it?

BROADBENT Not quite. It is a little more complicated than that, because the kind of subject he has got is normally not saying anything until you turn on the noise. Then, he starts talking. It may well be that they are anticipating that if they do talk, they are going to hear themselves after the delay, which is going to be very distressing. When they have the noise, it is an assurance that everything is all right. I have seen one or two of them do it, and it looks rather more as if, when the noise comes in, they are kind of pushed through some block in what they are doing.

CHASE Delayed auditory feedback often results in a cessation of stuttering.

IRWIN One of my colleagues, who is a stutterer, at times used a method of control which he calls high contact or high stimulation. This consists, essentially of paying a lot of attention to the lip-tongue feelings. When he is using this high stimulation, he says he feels himself talk rather than listens to himself talk. Under these circumstances, he is practically immune to any form of delayed auditory feedback. People trained to do this are also immune. I think this is compatible with what Chase has suggested, that there are hierarchies of control among the sensory system, and that not only may people differ, but the same person may differ at different times.

One of the tragedies of our age, with respect to feedback, as has been reflected in this discussion and in practically all our discussions, is our great concern with auditory feedback. We have so little information about feedback from within the mouth. We have a few deprivation studies, of reduced or absent contact, but almost nothing in terms of alteration of this type of pathway. I would anticipate that if we could have similar experiments in delayed tactile or delayed kinesthetic effect, the consequence would be more catastrophic than delayed auditory feedback, at least for skilled adult speakers of a language.

POLLACK Van Bergeijk and David (11) at the Bell Telephone Laboratories have tried to have people write, and have a system by means of which they could display the writing, after some delay. I believe the degradation was a monotonic function of the extent of the delay rather than, say, some critical delay such as, above that, there was no problem.

DENES I cannot remember what our results were, Fry, but I think we found the same kind of thing in London (65). As far as I can remember, there was a peak. Is that right?

FRY No, I'm sorry to differ with you, but I didn't find a peak. The longer the delay, the worse the activity got.

BROADBENT There is a point here, that writing without any information about what you are doing is pretty difficult, whereas we know that you can talk, to some extent, even though you are wearing noisy head phones and can't hear what you are saying. It is possible that the optimum period in speech may be because you go over from one mode of functioning to the other, or towards the same function as opposed to another, whereas, in writing, this may not be possible.

DENES We only tried writing simple figures like numerals. As far as I remember, the subjects did not find it too difficult to write in the dark. Under conditions of delayed visual feedback, however, they had the same kind of trouble as with delayed auditory speech feedback--they started to stutter. For example, they would make an extra loop on the number 3, or on a b.

OLDFIELD What kind of delay system did you introduce, may I ask?

DENES The system was basically a Telautograph, or distant writer, a device for transmitting the pen movements at one point for reproduction somewhere else. The pen used by the writer is mechanically coupled to the sliders of two electric potentiometers. Each potentiometer produces an output current proportional to the x or y coordinates of the pen movement. These currents can be used to control the movement of another pen in such a way that it follows the movement of the first pen.

In our experiment the electrical outputs of the two potentiometers were recorded and delayed in a similar manner to that used for obtaining delayed auditory feedback. The delayed signals controlled the second pen. The two pens--and writing surfaces--were placed one above the other, with the second pen on top and hiding the first one. The subject moved the first pen over the lower surface and saw the action of the second pen on the upper writing surface.

FRY Of course, there is a point about the auditory situation, I think, in that we do have practice with longer delays in reverberent places, and, certainly, we have never had this practice in the visual case of writing.

HIRSH But the delayed signal is never as strong as the undelayed.

FRY I would agree, it is never as strong, but my impression on delayed feedback, when the delay is long, is that it really feels like reverberation.

DENES Didn't you have some people speaking under delayed auditory feedback conditions whose lips and tongues were anesthetized?

LADEFOGED I did a similar experiment to the one Irwin has just suggested, and Ringel (115) went a little bit further. Both of us did the experiment with topical anesthesia, so that you couldn't feel whether your tongue was touching or whether your lips were touching anything, and, also, at the same time--or in different conditions, sometimes separately and sometimes not--using a loud masking noise so that you couldn't hear yourself. Ringel went a step further than I did, and used a deep anesthesia, so that you couldn't get any kind of steady sensory muscular feedback at all. Of course, under these conditions, when you couldn't hear and you couldn't feel, you were severely handicapped.

POLLACK But you could speak?

LADEFOGED Oh, yes, you could speak, and you could speak quite clearly. It sounds no worse than some severely affected patients whom one has heard.

GESCHWIND How does it affect the speaker? This is what is unclear to me from either of the experiments.

LADEFOGED Well, let me speak of mine, which I know rather better, naturally. When you couldn't feel where your tongue was and whether or not your lips were exactly touching, the typical difference was that you muddled up s and sh, f and p, and things of that kind; your articulations were not as precise as usual. My own observation was that you monitored your vowels fairly well on your auditory feedback, because it didn't seem to make any difference when you had topical anesthesia. My listening to Ringel's tapes indicates that the same is true, even under his kind of anesthesia--the vowel qualities do not seem to suffer so much under any kind of anesthesia as they do under deprivation of auditory feedback, whereas the consonant qualities are more affected by the anesthesia.

OLDFIELD The laryngeal muscles were anesthetized, too?

LADEFOGED In my case, I used a topical anesthesia, simply spraying on xylocaine. It didn't reach the laryngeal muscles at all.

OLDFIELD In his cases, did he?

LADEFOGED No, he didn't anesthetize the laryngeal muscles, either; only the articulators.

LENNEBERG It has been done all the time in bronchoscopy. There, the thing I have heard is that the patient on whom bronchoscopy is performed makes very odd noises.

OLDFIELD How far is the motor side affected in these various cases? What I want to know is whether the changes in articulation were due to the failure of the motor side.

LADEFOGED Not with topical anesthesia. In that case, quite clearly, the motor side is not affected at all. Perhaps, Irwin can report more on Ringel's study. I know from conversations with him that the motor side is not affected at all in his work, either.

IRWIN At least, he asserts this; that to the best of his observations, for example, with diadochokinesis, it was. When you are dealing with just topical anesthesia, one reason that has been hypothesized that the vowels do remain a little better is that they may be monitored by internal feel of the tongue, and this proprioceptive reporting is not destroyed, whereas many of the consonants are modified by tactile reporting, that is, contact. This does disappear when you spray on anesthesia.

HOUSE But, at the same time, isn't it a matter of dispute whether you really had any kinesthetic innervation in the tongue at all?

IRWIN Yes. I say, it has been hypothesized.

LADEFOGED It's the spindles in the tongue.

LENNEBERG Has that been demonstrated?

GESCHWIND There are muscle spindles in the tongue.

LENNEBERG But we don't know what they mean.

GESCHWIND There is pretty good evidence that the muscle spindles are deep muscle receptors.

LENNEBERG But the argument has been, has it not, that the spindles have been shown, but it is the same thing as with the eye muscles--that some people have claimed that the number of spindles which have been shown may not be sufficient to give good feedback.

GESCHWIND I think there is other evidence for eye muscles that visual control is more important then proprioceptive. But I would find it very hard to believe that someone was able to infiltrate the tongue with an anesthetic and just knock out sensory fibers and leave intact entirely all the motor fibers. A topical surface anesthesia would leave both motor fibers and proprioceptors unaffected.

LADEFOGED Of course, if I could just give the evidence, the kind of thing he cites as evidence--this is your field and I don't know how good the evidence is, but the kind of evidence is that, after he had done this anesthesia, the people could, nevertheless, move their tongues in any given way he wanted them to, by touching the tip of the nose or making any gesture that didn't demand a knowledge of the timing of the gestures, and producing them rapidly; so if they had ample time, they could do it. He inferred from this that there was no motor deprivation. Is that fair or not?

GESCHWIND I'm not sure.

LADEFOGED Can you explain to me why it is they have the ability to move into a given place?

GESCHWIND If some of the muscle fibers were paralyzed, the man might be able to make gross movements and still not be able to perform those delicate muscular movements which, presumably, give some of the delicate flavor to speech.

CHASE Do these experiments in which infiltration anesthesia was used involve controls in which a nonanesthetic agent was infiltrated, so we could assess the changes in the mechanical properties of the moving structures? Was a control done in which equivalent volumes of nonanesthetic agents were infiltrated?

IRWIN You mean, a placebo injection? Not to my knowledge, no.

HOUSE The anesthetic wasn't infiltrated; it was injected directly into the nerve.

GESCHWIND Did they infiltrate the muscle of the tongue itself with procaine or did they inject the nerve?

HOUSE The nerve.

GESCHWIND Which nerve? The tongue has its motor pathway through the twelfth cranial nerve and the sensory is through the fifth. If you infiltrate the fifth nerve in the right place, you can knock out sensation from the tongue and leave the motor functions completely intact.

HOUSE This is what was attempted.

GESCHWIND There would still be one problem in that case about getting all sensation because there is the possibility that some of the spindles might conceivably run with the twelfth nerve and not the fifth. The experiment you cite is a pretty adequate experiment. I had thought earlier that he infiltrated the muscles with procaine in which case it would be difficult to get loss of sensation alone.

BROADBENT On the feedback interpretation, I am a little bit hazy about what the time around the loop would have to be. Is it reasonable to expect disturbance of feedback to affect consonants that only last 30 or 40 msec?

LADEFOGED I'm merely saying what did happen. Whether or not it is reasonable for it to have happened, I don't know. (Laughter)

BROADBENT The point is that one is taking what happened with reference to a theory that this is because of the disturbance of sensory feedback. Now, we have had alternative explanations before, you see, and...

GESCHWIND For most ordinary muscular movements there is an obligatory feedback loop. The way that you normally innervate a muscle is to send an impulse out the so-called gamma efferent system to the small number of special muscle fibers inside the muscle spindles. As the result of the contraction of this so-called intrafusal fiber an impulse travels up the sensory nerve from the muscle spindle to the spinal cord and innervates the so-called alpha efferent system, that is, the efferent nerve fibers to the main bulk of the muscle fibers. Topical anesthesia should leave this entire loop unaffected since it shouldn't affect any nerve fibers inside the tongue.

HOUSE Isn't there an alternative response to this question, however? I don't really believe that we are talking about knocking out consonants that last 30 or 40 msec. You're not knocking out consonants; you're interfering with motor activity, and the activity is not measurable in 30 or 40 msec. When you say consonants of 30 or 40 msec you are talking about a measurement on a spectrogram. I object to this kind of identification, just as I objected to Lenneberg's talking about consonants in the same way.

BROADBENT Yes, I accept that. Of course, planning linguistic movement which may last a very short time may depend upon information acquired much earlier.

CHASE The issue that I raised earlier (see pp. 82-3) was one of a hierarchy of significance of sensory information coming along different channels with respect to the control of movement within different classes, leaving open for the moment the question of whether you could shift within a given motor system and rearrange this hierarchy under certain circumstances.

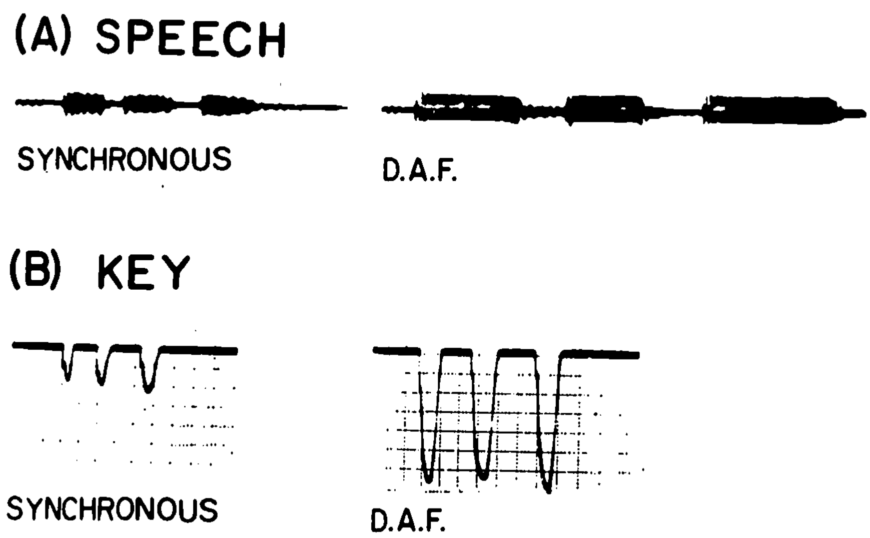

Figure 2. (A) Oscillograms of a group of three articulations of the speech sound b. The tracings to the left are under conditions of synchronous feedback; to the right, under conditions of delayed auditory feedback. In the delayed condition increases occur in vocal intensity, phonation time, and in the time between productions. (Errors in number also occur, but are not shown here.)
(B) Amplitude and time characteristics of a group of three key taps performed under two conditions. The downward displacement from the baseline is proportional to the pressure exerted on the key by the subject. In the delayed condition increases occur in pressure, in the time the key is held down, and in the time between taps. (Errors in number also occur, but are not shown here.)

Figure 2 shows the parallel qualitative changes in speech and key-tapping under comparable conditions of delayed feedback. The oscillogram of b, as in book, is at the top of the figure, and key tapping records are below. The comparable changes in motor performance for the two systems under delayed auditory feedback are: increase in amplitude of response, increase in time of the unit gesture, and a tendency to make repetitive errors (that is not shown in the figure).

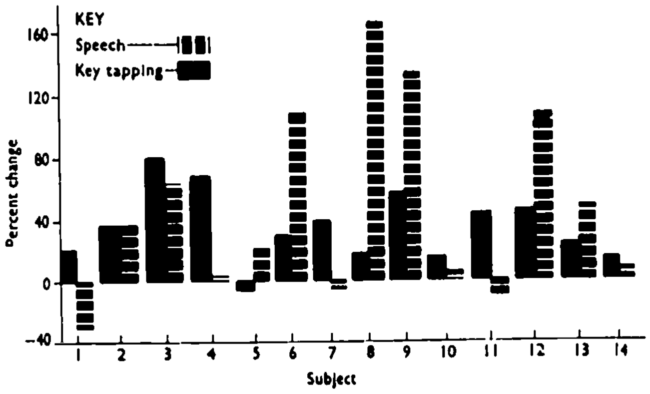

Figure 3. Percentage of change in mean unit-to-unit time for speech and key-tapping tasks, as indicated, for various subjects.

We examined the performance of 20 subjects, performing the task of elaborating groups of three speech sounds, under undelayed and delayed feedback, and elaborating groups of three taps under undelayed and delayed presentation of clicks. Figure 3 shows the percentage of change between units for speech sounds and tapping, going from control to delayed feedback conditions. The data are shown for each subject; the stippled bar is for key-tapping, and the striped bar is for speech. The absolute value plotted is the time from one unit to the next under undelayed auditory feedback, minus that value under delayed feedback, divided by the value under undelayed feedback--giving a percentage of change, moving from undelayed to delayed.

Most of the percentages of changes are in the positive direction, supporting the observation that the rate of the elaboration of motor activity slows down for both speech and key-tapping, but the order of magnitude of change for the same subject is quite different for the two motor tasks. There is good test-retest reliability for these data.

The interesting thing to us is that a subject who is markedly disturbed with respect to the temporal sequential release of motor units for speech may not be for key-tapping. We raise the question whether the determination of the hierarchy of importance of particular sensory channels with respect to the monitoring of motor activity, might depend upon the contingencies that surrounded early learning of different categories of motor activity. Perhaps when we are learning motor skills we have the capability of writing fairly individual programs which more or less set down the requirements for the relative importance of sensory feedback requirements for later control in different channels.

HIRSH Of those 14 subjects, if you rank them with respect to disturbance in speech and then rank them with respect to disturbance in tapping, what would be the magnitude of the correlation between the two rankings? I take it from what you said that it is not highly positive.

CHASE I think it is quite low.

A recess was declared at this point in the discussion.

SESSION 2. Part 2 - Language Skills: Development and Deficits

POLLACK Cooper's remarks this morning indicated that discussion chairmen were chosen on the basis that they were either experts in the field and therefore could skillfully guide the discussion, or else knew nothing at all about it and would not intrude. We might try to organize our ignorance concerning the development of language skills by asking four questions and see if we can talk to them.

The first question is: Do we have available detailed developmental schedules for the emergence of the various language skills? That is, is there a linguistic quotient that we might be able to identify any child by, based upon his performance on a set of standardized tests, which invites comparison with the performance of a large number of children at different age ranges? I hope Lenneberg will discuss this point eventually.

The second question is: Do we have corresponding norms with respect to normal neuromuscular development? There were a number of comments in a paper of Geschwind's (45) with relation to the role of neuroanatomical maturation and how this might relate to the emergence of the various language skills.

The third question is: What does the clinic tell us about the interruption of developmental schedules? In particular, what is the effect of early versus later nervous system impairment? Also, what is the role of early versus later deafness upon the development of various language skills?

The fourth question is: What is the role of second language learning upon the development of language skills? Until what age can a child begin to learn a second language without showing strong traces or accents of his previous language?

These are the four questions. Again, the first related to the problem of an objective specification of language skill; the second related to neuroanatomical development, as it might determine the emergence of various language skills; the third, related to the role of nervous system interruption in the development of the various language skills, and the fourth relates to the development of a second language. I call upon Lenneberg to start off with the first question.

LENNEBERG I think this is a very important point. Interestingly enough, the early observers did not consider the prelanguage period as relevant, and they started the language history with the advent of the first couple of words, which somehow obscures one of the most interesting facts, namely, that there is a development from birth on in vocalization. I don't think this could be called language, but there is a very rigid change in vocalization that can be studied and can be matched with milestones, the way we have motor milestones.

I can give you two examples of what these consist of. In the small infant, it is quite an event when the child begins both to smile and coo, and this very regularly occurs at an age of approximately two months, and is at a height in the middle of the third month and towards the fourth month. One can very mechanically stimulate this behavior in a baby by nodding one's head at it. The baby breaks out into a wide smile and immediately follows this first social response with cooing and clucking noises. This is one milestone which is very regular and correlates with motor milestones; the baby can hold his head easily, the tonic neck reflex is subsiding, etc.

Then, at about six to nine months, there is an emergence of some babbling sounds which are highly characteristic. It is quite easy to make judgments about a baby's age just by listening to these various noises.

I think the time of the first word is a very fixed milestone. Just about the twelfth month, something appears which everybody is quite willing to describe as words, even though I am aware that it is difficult to define this event more precisely.

FREMONT-SMITH What is the range of time for the onset of words? It certainly doesn't land precisely on twelve months.

LENNEBERG In a large survey, it was found that, roughly, 10 per cent of children lag behind these milestones. Something like five per cent of the population is early, and about five per cent is later.

FREMONT-SMITH But how early do the first words come? Are they roughly two or three months ahead of this schedule?

LENNEBERG Yes. For example, in a given sample of 100 children, we plot age in months versus the percentage of children that have at least one or two words, then, we get an ogive that has the most rapid rise around 12 months and falls off at, roughly, the 90 to 95 per cent level. Most of these curves turn out to have very much the same shape, that is, for different milestones of language development. This is really a normal distribution!

Twelve months is a turning point, though the onset of speech proper is later. The onset of speech is marked by an increase of words, and a sudden interest in language; suddenly words are put together into phrases, and this occurs just about the turn of the second to the third year. Here again, the function would look very much like an ogive--if you draw the curve in a different way, you would get a roughly bell-shaped but slightly skewed curve.

HIRSH What did you say happens at about 23 months? Is there a great increase in vocabulary size?

LENNEBERG The rate of increase in vocabulary suddenly changes; the lexicon is increased rapidly. There is a constant accretion of words from twelve months on, but, at first, very slowly. Then, at two years, suddenly, many more words are learned every day, and the child spontaneously puts these words together into new phrases. These are not phrases that are simply parroted, but, out of the stock of words that the child has, phrases are formed which characteristically are different from word sequences in the adult language. I think, this is a very important point.

Let me add that one can plot several milestones of this kind, and they are correlated with motor development. It doesn't mean that one is dependent on the other, but the whole process depends on maturational development plus proper stimulation.

POLLACK I don't know whether it was in a paper of yours or Geschwind's, but color naming and object naming appeared at different times. Could one of you indicate which is later for us?

GESCHWIND On the whole, color naming is later than object naming.

OLDFIELD What sort of objects? There is, in fact, a very considerable spread in the matter of object naming up to two to twelve years, according to what the object is, rather than color; whereas the colors come rather more constantly at about three years.

LENNEBERG That's right. You cannot say this is a milestone, but the ability to name colors probably is. You don't get this as regularly as some of the other things, because, obviously, there is a very important environmental influence on this. But it is a fact that at a time when children can name a great many objects and can make little sentences, they are frequently totally unable to get the color terminology correct. They are interested in the words, they know the words, but they can't match the word with the phenomenon.

With your permission, let me run through some slides. (At this point Lenneberg presented material that is reported in detail elsewhere (88).)

FREMONT-SMITH As I understood you, the infant's cooing behavior had to be evoked by parental activity? Couldn't this fit in with the ethological approach of an innate pattern behavior and a releasing mechanism?

LENNEBERG I think it is very much like it; at least, that is how I like to look at it. Later on, this is not so and the child begins to recognize faces and you cannot do this mechanical treatment.

FREMONT-SMITH The face becomes a releasing mechanism then, and the smile is a releasing mechanism, isn't it, for the baby's smile?

LENNEBERG At the early stage, it is, yes. Later on, he begins to discriminate faces.

FREMONT-SMITH But in the initial part, I mean, it is quite important to emphasize the innate pattern which is ready to be released and can be released by the appropriate mechanism?

LENNEBERG That is the way I look at it. I have just one final word. In another study where we followed deaf children of deaf parents over a longer period of time, the kind of babbling sound that these children make up to 12 to 18 months seemed acoustically very much like that of the hearing children. You can't make definitive statements on this because you can't really judge similarity in these very inarticulate sounds very well, but you hear very clear sounds such as mama, papa, papapa, in the congenitally deaf children. There are quantitative differences, to be sure, but both voice quality as well as babbling sounds are definitely heard in these youngsters.

FREMONT-SMITH And do the deaf children change their behavior as they get older?

LENNEBERG Well, they don't develop speech automatically.

FREMONT-SMITH But do they continue a babbling sound which is like the normal, beyond the normal age for babbling?

LENNEBERG They certainly continue it, but I can't give you quantities on it.

COOPER If I understood you right in this, I tried to follow some of the details and got lost on the main story, but, essentially, if you looked at the child in these early ages, you couldn't tell whether his parents were deaf or hearing persons. Is this it, roughly, that the behavior is about the same, or have I missed the point completely? I understood that the acoustic output from the child is about the same, regardless of whether his parents hear or don't hear.

LENNEBERG In the first three months, there is no quantitative or qualitative difference between those children who are born to deaf parents and to hearing parents, and even later on it is very hard to demonstrate clearly how the deaf children differ. Differences do emerge, but it is not a lack of capacity in the deaf children.

FREMONT-SMITH Don't the parents who are deaf have to use a different mechanism of evoking the cooing than the nondeaf parents? They don't use voice.

LENNEBERG That is right, but you don't necessarily need voice for it, at this early stage.

FREMONT-SMITH But you can use voice?

LENNEBERG Yes, you can.

FREMONT-SMITH So the parents who do not habitually use voice have to take on added activity of another kind to evoke the same amount of cooing.

LENNEBERG The most powerful stimulus at the age of three months is nodding the head.

DENES The implication of this is that it is really only at a relatively late age, say, six months or so, that the imitative action--the child trying to imitate what the adult is doing--has any effect on the child's speech activity.

LENNEBERG That's right, or these somewhat better controlled babbling sounds eventually do require both hearing and sensitivity to the environment. The absence of these leads to differences which emerge just about at six months or so. But the fact that you can hear infants utter sounds that are very much like the sounds that occur in English makes me wonder whether there isn't some rather deeply seated organization of muscles that allows the introduction of these sounds, and it is not entirely dependent on proper environmental treatment.

OLDFIELD But you say the sounds occur in English? You mean the sounds that occur in language, definitely. At this stage, they won't be able to produce words, will they?

LENNEBERG Yes; mama.

DENES These are the sounds, you would say, without any knowledge of speech at all? They are the sounds that children would make, merely by knowing that they can produce sound with their larynx, mouth, and so on?

LENNEBERG Yes.

POLLACK Without infringing on Chase's session, I wonder if Geschwind could tell us something about the neuroanatomical maturation that might go along with some of these milestones?

GESCHWIND I would like to discuss briefly some of the anatomical factors that may be involved in the development of language in man. I have discussed this elsewhere (44, 45) and will summarize briefly here without considering the evidence exhaustively.

Flechsig (37) stated the principle that the primary motor and sensory regions of the cortex have no long connections with any other part of the cortex. They have connections only with immediately adjacent regions of cortex which are called association cortex. Thus if we consider the primary visual cortex (in its strict sense, area 17 alone) it is clear that it has no long connections with any other region of cortex either in the same or the opposite hemisphere but has cortical connections only with the immediately adjacent visual association cortex. The visual association cortex in turn has three major sets of connections: (1) via the corpus callosum to the opposite side, (2) to motor association cortex in the frontal lobe. The largest single body of connections of visual association cortex is (3) connections running to the lateral and basal surfaces of the temporal lobe.

Why should the largest outflow of the visual regions be to the lateral and basal temporal lobe? If one examines the connections of this part of the temporal lobe one finds that its connections, insofar as they are known, appear to be predominantly with structures of the so-called limbic system--hippocampal gyrus, amygdala, dorsomedial nucleus of the thalamus.

The fact that the major outflow of the visual system is to the limbic system becomes understandable when you realize that the "visual" learning of a monkey consists predominantly of visual-limbic associations. Let me clarify this further. The limbic system appears to be involved predominantly in those motor activities related to preservation of the self and the species, and in those sensory processes related to these activities. If a monkey learns to respond with rage to a visual stimulus the pathway involved in this is presumably the visual-limbic pathway we have outlined above.

Consider the experiment in which you teach a monkey to choose between a circle and a cross. You teach the monkey to do this by rewarding the choice of one of these stimuli with, let us say, a pellet of food. You are in fact teaching the monkey to associate the visual stimulus to a limbic stimulus, that is, you are teaching a visual-limbic association. We can state this in an alternative way. We would never teach a monkey to choose a circle over a cross without reinforcing the choice of one stimulus over the other. But all the positive or negative reinforcements concerning which we have anatomical information have localizations in the limbic structures or in their connections.

Consider now the effect of removing the cortex of the lateral and basal temporal lobe. There are several experiments which describe the effects of this procedure and these are usually described as difficulties in visual discrimination. I would prefer to say that the animal fails to make a choice between the stimuli because neither stimulus is reinforced. The animal can no longer form visual-limbic associations.

POLLACK Excuse me, but, operationally, how would you distinguish between these two?

GESCHWIND In the first place evidence indicates that these animals have excellent vision. Monkeys with such ablations will pick up small pellets from the floor or will catch flies. Secondly, the range of visual disturbances in these monkeys are quite different from those seen in animals with lesions in the visual cortex proper.

MILNER I'd like to quarrel with the way you put this thing, that it should be related to the reward. Presumably, it is one aspect of pattern discrimination, learning and what have you. This is the usual interpretation of the cortical deficit, that there are many aspects of vision which can be disturbed and there is a difference in a peanut against a background and learning to discriminate.

GESCHWIND That's right. I realize that is the usual interpretation, and I think the usual interpretation has come, in fact, from the tendency of workers in the field to speak in perceptual terms.

MILNER It fits the human data, too.

GESCHWIND I don't believe it really fits the human data, but I will leave these aside for the moment.

MILNER This is just what you were talking about, really, when you said it was a visual-limbic association.

GESCHWIND Yes. I don't think it is a disturbance of pattern discrimination.

MILNER Pattern-discrimination learning, you mean.

GESCHWIND I know that this is what the experiment usually is called. I think, in fact, that it is an experiment in teaching the monkey to make a choice of one pattern over another pattern. I don't think these interpretations are the same, because of many other evidences that these monkeys can, in fact, pick out small things in a complex pattern, for example, protuberances on the wall of the cage.

HIRSH Some have thought that what the psychologists call discrimination learning was discrimination, but it isn't. It is much more complicated, as much more as you are making this out to be. I don't think these are so difficult.

GESCHWIND I have no disagreement on that. The only point I'm stressing is that the mechanism of this failure is that the animal is simply failing to make a choice because he is prevented from making the association between these visual stimuli and the limbic stimulus. I am offering a mechanism for the failure of the animal. Similarly, some of these animals are tame, but the evidence suggests that this is tameness to visual and not tactile stimuli.

MILNER Excuse me, but I really feel bothered about this, because the taming in the monkey is dissociatable from the deficit you get from the neocortex. Also, in man, you very definitely get deficits in learning to recognize unfamiliar visual patterns--where there is no question of giving the person a peanut if he is right--when the same person can do it quite well if we use a different kind of material. It seems very much the nature of the particular visual pattern we are giving him to learn which creates the difficulty. There are many visual tasks that these people can do, and it seems to me very important to discriminate among them.

GESCHWIND I believe that there may be some different explanations for the human material, but I will leave this discussion until later.

Let's return to the problem of the development of language and its maturation in man. You will notice that the monkey readily learns to choose between a circle and a cross. But it is extremely difficult for a monkey to do a task in which he has to make a cross-modal transfer or association between complex nonlimbic stimuli. Thus Ettlinger (33) showed that a monkey who has learned to choose between a pair of stimuli presented visually will show no evidence of learning when he is presented with the same pair tactilely. He must learn the task completely afresh. In the human by contrast this task is very easy.

FREMONT-SMITH You mean, the transfer is very easy?

GESCHWIND Yes. The usual argument as to why this task is easy for a human is to say that humans use <u>verbal mediation</u> in order to perform such cross-modal tasks. If you consider this question, carefully, however, you will realize that this is not really a solution since the very act of <u>verbal mediation</u> is itself dependent on the ability to form nonlimbic cross-modal connections. Thus when the child learns to name objects he is confronted with, let us say, a visual stimulus and given an auditory stimulus. Thus we show the child a glass and say, glass. The child must be capable of forming a visual-auditory association. So we see that we must turn the usual explanation around. In order to develop object-naming it is necessary as a prerequisite to have the ability to perform nonlimbic cross-modal transfers. Humans can form this type of association and hence can develop object-naming which is a necessary prerequisite for the development of language, but monkeys cannot do this.

Now why can the human form this kind of visual-auditory association which the monkey can not? Well, in the first place, if you inspect the connections of visual association cortex in the monkey and the auditory association cortex, you find, in fact, that neither by physiological nor anatomical methods has anybody ever demonstrated any significant body of connections running from the visual to the auditory association cortex. A few such connections have been found in the reverse direction, by strychnine and by anatomical methods, but this connection is very small compared to the large outflow running from visual association cortex to the lateral and basal temporal lobe in the monkey.

If the primate fails because he lacks the necessary anatomical connections, why does man succeed? We can approach this problem by considering the cortical areas of man. In the first place let us consider which areas in the human brain are the most developed when compared with the brain of a primate. The area which has grown most in the human brain is the posterior-inferior parietal region. This region, which corresponds roughly to the angular gyrus of man, has the hallmarks of a phylogenetically late region. Thus Flechsig found that it was one of the last regions to myelinate in the human brain. If you look at the region in question it can be seen that it lies in the meeting place of the visual association cortex, auditory association cortex, and somesthetic association cortex. It is thus strategically placed between the three nonlimbic association cortexes and is thus in the ideal location to mediate the nonlimbic intermodal connections necessary for the development of object-naming. Various bits of evidence from the study of aphasia, which I will not discuss here, fit in with this concept of the functions of this region.

LENNEBERG I really don't follow you at all. First, what does myelination have to do with this? My second point is, it is quite obvious that blind children have no difficulty in learning to speak, so the relationship between area 17 and other areas does not seem to be relevant.

GESCHWIND The blind child learns to name objects not by visual-auditory connections but by tactile-auditory connections; this is still a nonlimbic cross-modal activity.

Late myelination is an indicator of the phylogenetic lateness of the area in question. It seems reasonable that an anatomical area which is involved in the development of a function not present or only poorly present in lower animals should show the hallmarks of being a recent evolutionary acquisition.

LENNEBERG Your timetable does not coincide with the timetable of learning, because myelination still takes place at three or four years of age.

GESCHWIND Those systems which myelinate late come into activity late. This is not to say that the function and the myelination occur at the same time. Thus we know that many pathways are functional before they myelinate. It is, however, probably true that the later the myelination of a pathway, the later it comes into functional activity. The region we are speaking of is thus evolutionarily recent and it myelinates late. I have suggested that this region is probably involved in the types of cross-modal transfer which are necessary for the development of object-naming.

This region as I have noted, myelinates late. We know that boys tend to be delayed relative to girls according to most criteria of maturation. I would make the prediction that the myelination in this region occurs later in boys than in girls, and that this later myelination will correlate with the fact that language facility develops later in boys than girls and that disturbances of language learning--reading deficits, for example--are more common in boys than girls.

FREMONT-SMITH Do the girls talk earlier?

GESCHWIND I believe that is correct. Can you comment on this, Lenneberg?

LENNEBERG Statistically, yes, but it is just a couple of months.

GESCHWIND In school, don't girls generally do better in language in the early school years?

LADEFOGED This is possibly due to the fact that there are more women teachers than men. There is a study going on at UCLA, if I remember correctly, at the moment, showing that if you take pupils away from women teachers and substitute men teachers, then, the boys will do just as well as the girls.

GESCHWIND It is my impression that the specific deficits of childhood language occur four-to-one in boys as against girls. I don't think anybody has demonstrated a cultural group in which specific deficits turn out to be more common in girls than boys. I would suggest that at least part of the defects are the result of late maturation

POLLACK We are lucky enough to have a number of people associated with various clinics at this conference. I wonder if they would like to comment on the effect of early accidents, in particular childhood aphasia, with respect to this last point?

HIRSH I can't say very much about childhood accidents. If the child is already developing language and sustains the kind of accident that would normally produce the kind of aphasia that one sees in adults, then, we don't seem to see many. I suspect such a child is sent to a neurologist and to a speech pathologist, and would not normally be enrolled in the kind of school program where we see children with what has been termed by some aphasia.

I can, on the basis of a very remote connection with this school population, tell you some of the characteristics that, at least, I see in these children. Please understand that I don't live with them day to day, the way their teachers do and the way the original examiners do.

First of all, most of these youngsters have some hearing loss. This is a point that, I think, has been overlooked in many of the descriptive articles. By some hearing loss, I mean not enough hearing loss to account for the lack of language development. Now, having said this, I am not sure what I have said, because we don't quite know how much hearing loss is required before language development will be interfered with. We are just seeing evidence now, I think, in another group of youngsters who have hearing losses of the kind that involve a lack of sensitivity to the high frequencies, but very good low-frequency hearing. In other words, the kind of youngster you see who will turn around easily to a verbal command in an examining room, and therefore give one the impression that he has normal hearing. He turns around in response to a command which may be primarily low-frequency energy, an attention-getting device, if you like, but nobody really has examined the discriminative capabilities of these youngsters.

FREMONT-SMITH May I just throw in that when Luria (A. R. Luria, University of Moscow and Academy of Pedagogical Sciences, U.S.S.R.) was here a few years ago, he pointed out that they had found that quite a high proportion of their so-called mentally retarded children had been partially deaf during the period of language acquisition--a very minor degree of deafness in this period interfered with language so much that you got a mental retardation effect. They were changing their institutes for the mentally retarded in many instances to institutes for the deaf where they had been able to retrain many children.

HIRSH There is considerable evidence from a variety of sources and a variety of countries on the crucial nature of the kind of language learning and sheer auditory experience that takes place in these very early years, which, so far, at least, has not been capitalized on in normal educational practice, simply because the educators don't get hold of these children until they are at least three and many times not before they are five years old.

FREMONT-SMITH And also because we have denied it, or have not thought it was so. I believe this is a relatively new development--to the degree to which it is true--on the part of the Russians.

HIRSH In a degree, I would agree. I think you can find the same emphasis if you look through the literature on intelligence testing, where many youngsters show a sharp disparity between the results on the so-called performance scale and the results that you would get on a verbal scale of intelligence.

Well, to return just to this group, these are youngsters who seem to show relatively good response to items on a performance IQ test, and so one would not say they are mentally retarded. They have not developed language, and, naturally they fail badly on a verbal IQ test; in fact, they don't improve but continue to look poor, even after their education has become what many of the teachers would describe as successful.

One of the interesting aspects of these youngsters, I think, in the present context--and here I can report only specific examples rather than a statistical study--is that they can learn. (I should restrict my remarks to a particular educational system that is practiced at the Central Institute for the Deaf.) They can learn specific associations. They can learn, for example, to name objects, but, through a series of peculiarly tedious steps. You can teach one of these youngsters in a relatively few days to make a series of sounds, like the word ball, in response to your presenting to him the sound of that word. You can also teach him to select a ball from a group of objects when you say that word. But it is only somewhat later that he can turn around and say the word ball in response to the object, ball.

This kind of cross-modal transfer that you have been speaking of seems to be a crucial difficulty for these children--more difficult for them, I would submit, than it is for the normal deaf child, that is, the deaf child whose only problem appears to be deafness.

As far as the anatomy of these children is concerned, we know nothing. They are not particularly sick, and they will not be hospitalized for years.

FREMONT-SMITH Are these injured ones or just the ones who are slow?

HIRSH These children come to us because the primary complaint is that they have not developed language, and so we just don't know.

COOPER What age are they when you are talking about them, as you have just been?

HIRSH Anywhere from three years to, in some cases, six or seven. I should say that the older they are when they finally get referred because of a suspicion that there is a specific language problem, the more likely it is that they have been in some other institution first, being educated as if something else were wrong.

FREMONT-SMITH But they are not deaf?

HIRSH No, they are not deaf, but most of them are partially hard of hearing in terms of pure-tone audiometry.

COOPER You implied that the later you get them, the worse the prognosis for their learning language. What is the time scale on this? Could you indicate roughly how fast it cuts off?

HIRSH How fast the difficulty increases? No, I couldn't, but I'm sure it's monotonic.

LENNEBERG I have some relevant data that some of you may have already heard. I made a study of both published reports and actual hospital cases of traumatic aphasia in childhood. These are cases of children with acquired brain lesions that resulted in aphasia, and I studied their recovery. There are something like 30 published cases. In our hospital, we had a collection of some eight cases, with a very good correlation between age and likelihood for recovery. The younger the child, the less lasting were the symptoms; both in the literature and in our own cases there was a definite cutoff point at just about 10 to 13 years of age. It is a little hard to pinpoint it closer than that, but by this time, permanent deficits become much more frequent and very soon the incidence is the same as that in adult aphasics. Among adults about one-third of patients with traumatic aphasia never recover. In children, one begins to see that kind of incidence of permanent residues in the early teens. Apparently, in childhood one can make better adjustment to speech-disturbing brain lesions than later in life.

COOPER Is this better adjustment a transfer to the other hemisphere?

LENNEBERG I hate to call it transfer, but I certainly think the asymmetry has something to do with this phenomenon.

MILNER We have long been interested in the question of the critical age at which the other hemisphere can develop language. I would agree that I wouldn't call this transfer, because they are developing language for the first time. I think Rasmussen has found that somewhere around six or seven is the age, but we have individual variations. We would like very much to have more data on this. I don't know if Geschwind would predict that the boys would have a higher critical age.

GESCHWIND That is a most interesting point. I would suspect that boys would have a higher critical age than girls. I would also suspect for other reasons that left-handed children would have a higher critical age.

MILNER Yes, I would agree.

GESCHWIND I would in general agree strongly with the conclusion which I think both Lenneberg and Milner have drawn, that it must be very rare in adults for the right hemisphere to take over. I think, however, that there is at least one case in the literature that I know of, where it would be necessary to conclude that the right side did take over. This was a case that Dejerine (29) described. The patient was a left-handed woman who had sustained a left hemiplegia, became aphasic, was followed over the next four years and had a total recovery, or a nearly total recovery. At postmortem, her right hemisphere had been nearly completely destroyed, so that language must have shifted to the opposite hemisphere. The fact that she was a left-hander was probably involved in her ability to shift.

MILNER And the possibility that she had some bilateral representation.

GESCHWIND Yes, I think she had some bilateral representation.

MILNER What was the cause of her dysphasia?

GESCHWIND Originally, she had aphasia in all modalities of speech. The first thing that recovered was comprehension of spoken and written language, which, I think fits in with other observations. Then, she recovered spoken speech. Fascinatingly, even though she had a left hemiplegia and even though she had always written with her <u>right</u> hand (while all other activities were done with the left) the only thing which did not recover was writing. Otherwise the pattern of recovery is what you would expect; that is, the sensory components came back first. But I think this is really bilateral representation, which is unusual.

FREMONT-SMITH Is there a possibility of transfer other than to the other hemisphere? I am thinking of work where bilateral lesions were made in infant monkeys, who then recovered and showed that the transfer was made forward, to the frontal lobes. I wonder if there is any possibility that transfer in the same hemisphere is possible in humans for this speech situation?

MILNER Do you mean a series of bilateral lesions?

FREMONT-SMITH Yes, a series of bilateral lesions, with complete paralysis.

MILNER I think, within one hemisphere, this is what usually happens, that the neighboring tissue on that side takes over.

FREMONT-SMITH So the transfer doesn't have to be only to the opposite hemisphere.

MILNER I think it occurs very rarely in the opposite hemisphere. Only in a very young person is the opposite hemisphere able to develop this ability.

DENES Are there any visually observable anatomic changes which accompany this transfer, or is this purely a functional change?

MILNER I think an effort has been made to demonstrate anatomical changes, but the results are pretty scant.

GESCHWIND I think, again, you wouldn't expect to see an anatomical change in the child who had already developed language. I think that there are probably some anatomical differences between the left and right hemisphere, and we are hoping to demonstrate this fact.

POLLACK In the fifteen minutes remaining in this session, I would like to go to the fourth topic, the problem of second language learning.

DENES I have the feeling that many people think, rightly or wrongly, that you have to learn a second language at an early age, in order to be able to speak it fluently or without an accent. If this is true, this has some relevance to learning generally and language learning.

POLLACK I have heard the figure of age 12 with respect to accent.

DENES I have my doubts about how true this is. There is no doubt that most people who learn another language at a later age cannot learn it without an accent, but I wonder whether this is an innate disability or whether it is just lack of practice or some factors relating to social circumstances.

FREMONT-SMITH It might be that learning a second language in French or German might be different from English, or vice versa.

LENNEBERG I'm sure there is a difference, but I think it is a strange coincidence that children of immigrant families invariably pick up the new language without accent if they are under ten years or so; invariably, the older people have an accent. I just can't quite imagine that all of this would be just convention.

DENES But the question is, can you, in fact, acquire another language, or more than one language, without an accent, regardless of the age at which you start learning the second language?

LENNEBERG I do know that throughout history it has been a matter of life or death for people to speak a language without an accent: hence the <u>shibboleth</u>.

DENES But can one, in fact, speak two languages without an accent?

FREMONT-SMITH You mean, no matter what age? Oh, yes; three languages, I believe.

LENNEBERG If you learn them early enough, you can speak five languages without an accent.

DENES I can speak three languages fluently, all three of which I learned before starting school. Just the same, I can't speak any language without an accent. I wonder whether this is due to the fact that you can either learn several languages fluently but not without an accent, or just one language without an accent, regardless of your age.

LENNEBERG There are many cases, I think, of persons who speak two languages without an accent, who learned those languages in childhood, and still another language acquired after 12 years of age, spoken with an accent.

POLLACK Are there clinical cases of dissociation such that one language remains and the other somehow or other drops out?

FREMONT-SMITH This has happened in people who have learned two or three languages at different ages and have had an aphasia, an accident, and they recover. This is the story that is told--the early language first, and the last-learned language last.

POLLACK Our neurologists are shaking their heads.

MILNER There are many factors here. Any of the patients whom we have seen--and we see a lot of people who speak French and English, for example, in Montreal--who spoke more than one language, we certainly found to be aphasic in all the languages they spoke. This depends a little bit on the skill of the observer in detecting it.

An English-speaking nurse might say, "Mrs. So-an-So is terribly aphasic in English but in French, she's fine," or a French-speaking nurse might say, "She is very aphasic in French but not in English." I think this is true of writing and so on.

With the recovery of language they may make more progress in one or another, and there may be many factors--the situation in which they are, or the people around them speaking one or the other language--influencing these recovery rates (20). There are extreme cases, where one language may not be used--for emotional reasons, perhaps. But I am very skeptical about a pure dropping out of one language and retention of another.

FREMONT-SMITH But isn't it true that the longest dropping out is of the latest learned language?

MILNER Not necessarily, no.

GESCHWIND No. In fact, Lampert's studies (83) done in Montreal showed that, on the whole, in this French-Canadian population, the language best preserved when the patient became aphasic was the language which he was speaking most and best at the time when he became aphasic. Now, I have a suspicion on the basis of some of my own experiences that many of these patients who appear to be better in one language than in another, would, if you gave them a few days' practice in the unpracticed language, turn out to be about equally aphasic in both.

HOUSE I suspect that, so far, most of us are expressing some folk biases about this subject, and the most refreshing thing I have heard today was Denes' revelation. This may be the key to the whole problem. Maybe, you always learn a language with an accent and we must try to specify what accent you are learning. If you learn two or three languages at the same time, you may not have an opportunity to develop an appropriate phonemic system in any one of them.

GESCHWIND But I think it's clear that there are people who are perfectly bilingual.

HOUSE I'm not denying that at all. I'm merely saying that many hypotheses can be advanced because we really don't have enough evidence. I don't want to reject the hypothesis that you can learn a second language at any age; I think that with the proper instruction and the proper motivation you can do it.

LENNEBERG I think there are many examples where people's lives did depend on it, and people were killed in cold blood because they couldn't learn it. This was even a fact during the last world war. I know of several instances.

HOUSE These events usually depend on relatively local pronunciation or knowledge. An example is stopping a soldier during the Battle of the Bulge and saying, "Who won the pennant last year in the American League?"

LENNEBERG I think this is not quite true. There are people who escaped from prison camps, and it was vital that they adopt the right kind of language around them.

DENES Just a moment! Nobody is saying that in 24 hours or so, you can acquire this.

LENNEBERG All right; then, take movie stars, where livelihood does depend on proper pronunciation. There are many examples of people who came to this country and could not adapt themselves. I think the Pygmalion story is precisely the fairytale that we have to combat. There is not a single example that can be quoted, that this has ever been achieved, up to a certain age.

OLDFIELD Take the example of professional singers who acquire exceptional perfection of pronunciation. If you ask many of them to speak spontaneously, even if they did speak that language somewhat, I think you would find they did not have perfect diction. I think that a voluntary act of speaking to acquire perfect pronunciation of something one knows by heart is quite another thing.

HOUSE You are saying then that they fail to learn to speak the language; they learn merely to sing in the language.

LENNEBERG Yes, there is already some distortion.

HOUSE I have the impression, when I meet somebody who speaks a large number of languages--and occasionally you meet linguists who make this claim--that he will speak all of these languages in the sense that Denes claims he speaks his. He won't speak any of them correctly. (Laughter)

FRY If you ask yourself what it is in the speech that makes you recognize a foreign accent, a kind of summary answer to this is that the time scheme is not right. When people sing, the time scheme is given to them by the music, and this is why you have the impression that they are singing with a perfectly good accent. If there is an opera in which there is any dialogue, you soon realize your mistake.

MILNER Doesn't this fit in with many other skills, such as dancing and skating and so on, where, very clearly, you have an advantage in learning quite young? This is only a mark of a highly developed instance of the same thing.

COOPER You are speaking of "quite young." There is a question of how young is "young," when it comes to learning certain things. We have had the ages of six and seven for total recovery, if I remember correctly, and of 10 or 12 for second-language. But aren't both of these figures quite late for the development of language skills by hard-of-hearing youngsters?

HIRSH I think that the age that is critical for picking up first-language skills is much earlier than any of these figures. I think the superimposition of a second language on language is rather different and a less complicated affair.

POLLACK I would like to ask Liberman: Have you done any of your identification discrimination tests on children of varying ages, with the notion that there are some sounds we do not use in English that young children might be able to identify or discriminate in a noncategorical fashion in contrast with adults?

LIBERMAN The answer is no.

CHASE We have been thinking about the question of the capability for acquiring a second language, and wondering whether this involves a plasticity that is chronologically linked, and which becomes less and less available. I'm wondering about considerations raised in the materials which Risberg distributed yesterday with respect to this point.

Here, the question is not the capability for learning different motor gestures on the basis of a linkage with different kinds of acoustical inputs, but rather the effect to learn a unitary set of motor gestures, utilizing recoding of the essential acoustical information in one language structure into other modality presentations. Is there any information, Risberg, whether the capability of recoding speech from acoustical into non-acoustical sensory displays as a way of teaching speech to the hard-of-hearing also involves the issue of critical periods in learning, and requires plasticity that is present at an early age and that is not present at a later age?

RISBERG I'm afraid that there has been very little done in this field. We have no data at all about when transformations should be made or put to work, but it is probable that it ought to be done as early as possible.

CHASE I wonder what the consensus is, in general, about whether the kind of plasticity, in terms of neural organization underlying the capability of second-language learning, overlaps the issue of plasticity required for learning speech on the part of a congenitally deaf child utilizing a recoding to another modality?

RISBERG I think this is very probable. The work that has been done has been done on adults, and the results are, as a rule, not very good.

POLLACK On this point I think we can stop.

SESSION 3. The Production of Speech

COOPER This morning, we come back to the underlying essentials of the speech process--speech production and the acoustics of speech. These two topics are essentially indistinguishable, I suppose. House has agreed to be our discussion leader for the first part; Stevens for the second part; they can divide the time between them as they see fit.

HOUSE I want to provide a few interruptable, "on the house" comments to get started this morning.

It is to our advantage to think of the peripheral processes of speech production as being located in a black box. The output of the box is the acoustic end product that we call speech, and the input to the box may be likened to a set of discrete control signals or instructions.

We discussed the output signals a little bit yesterday. They are the messages we try to discriminate and recognize, and I am sure that more will be said about them today. This morning the nature of the input instructions is our primary concern.

We know quite a bit about this particular black box. From an anatomical point of view, it consists of portions of the digestive and respiratory systems, as well as associated neural elements. In simple terms, the respiratory system provides energy which generates an acoustic disturbance, and this sound excites a system of cavities that are varying their sizes and shapes in time. The number of muscles involved in these operations is large and their individual actions during speech production have never been described in great detail.

The neural innervation serving these muscular systems is also quite complex and is not fully understood. Although there has been some interesting work done on muscle action--

some by people who are present here today--the prospect of writing input signals on the level of muscular activity seems remote and writing input signals on a level of neurologic description may be even farther away.

COOPER I'm not as sure as you are about the remoteness of the first of those two prospects.

LADEFOGED I agree with you, Cooper.

HOUSE If the input instructions on the muscular level are fairly discrete, I think we may still have a little trouble in writing them.

COOPER I was thinking of indiscrete ones, in a sense; namely, that you don't try to describe everything in high-fidelity, but you try to describe the essential signals that carry the information. These are two very different descriptions.

HOUSE I hope that a better description of muscular events will develop during the course of the morning.

To continue, the kind of input description that would be very economical and which we seem able to handle is a stream of phonemic symbols. This input would be a very appropriate one, but in many senses it is at a very high level of abstraction and its derivation may prove to be a more difficult task than that of the other descriptions.

Where these kinds of instructions are stored or how they are manifested is not particularly clear. Our understanding of articulatory activities in terms of gross movements of the masses of tissue surrounding the cavities we desire to excite is much greater. The literature of general phonetics is ancient and extensive, and, even today, is being refined by new ways of looking at things and new ways of measuring things.

It has proven useful to view the cavities formed by the movements of the articulators as an acoustical system. Particularly in the past twenty-odd years, the study of speech production in terms of the acoustics of this system has been extremely fruitful (26, 35). It is very instructive to discuss the acoustics of speech concurrently with speech production, since the transformation from articulatory configurations to acoustic representations seems to be understood in great detail today.

FREMONT-SMITH Would you identify the main cavities? I am very ignorant of this and, perhaps, not everybody knows them.

HOUSE We can think of the acoustical system as a tube that starts at the level of the larynx and extends upwards to the level of the lips, with a side branch at the velopharyngeal port leading out through the nostrils.

FREMONT-SMITH That is one cavity?

HOUSE It can be thought of as one long tube, with a side branch.

FREMONT-SMITH And below the larynx?

HOUSE Below the larynx, we have other structures and other cavities, but we will ignore them in our acoustical model today.

HIRSH Did I understand you to say that these articulatory-to-acoustic transformations are reasonably well known?

HOUSE Yes. These last comments about studying articulation in terms of acoustic manifestations were made with the explicit approval of my colleague, Stevens. We got our heads together last night and anticipated that the discussion of speech production could occur under either topic, so, at any time, he can jump into the discussion or you can address your comments directly to him.

STEVENS May I say something that, perhaps, is not in direct accord with what you have said? This view of the speech-production mechanism, as we both know, is that of an open-loop system. We talked about phonemes, and then these act as the input to a subsequent stage, until finally after several more stages sound is generated. Yesterday, we talked in some detail about feedback processes in speech production. There is one group that is studying speech production from the point of view of looking at articulator positions and looking at controls to the muscles. There is another group that is studying auditory, kinesthetic, and tactile feedback. It seems to me that these two views are not quite consistent.

If you are looking at the signals within the loop of a feedback system, they do not necessarily bear too direct a relation to the output or input signals. Consider, for example, the heating system of a house, which is a very simple feedback system. What it is supposed to do is keep the temperature constant, regardless of the outside temperature. In examining the characteristics of such a system one should not devote one's entire attention to the study of how often the furnace turns on and off in trying to keep the temperature constant. The important thing is how well the temperature is controlled under the influence of outside temperature fluctuations.

COOPER What the thermostat is doing, in other words?

STEVENS Right. What you are doing in examining the on-off behavior of the furnace is getting inside the feedback loop.

LADEFOGED On the contrary, one is trying to find out what kind of gas is being burned in the furnace.

STEVENS I only wish to make the general comment, that we are, from one point of view, viewing the speech-production system as an open-loop system, and from another point of view, as a closed-loop system. Perhaps, we should get together.

COOPER Well, if I understand you right, and if you are saying that the motor commands--that is to say, the shorthand for the neural operations or neuromuscular operations that cause the articulators to move--are inside the feedback loop, then you are talking about a feedback loop that includes the outside acoustic space, that is, the loop that one uses to monitor his own speech.

STEVENS And, possibly, tactile feedback.

COOPER And possibly tactile feedback--yes.

FREMONT-SMITH Wouldn't you almost say surely tactile? Certainly for every articulation there are afferen impulses, proprioceptor impulses, which are possibly subconscious.

STEVENS Yes, I'm sure there must be.

FREMONT-SMITH So that the tactile feedback would be there inevitably, not just possibly.

STEVENS Yes; and can't one view these neural control signals to the muscle as signals that are, in a sense, reducing some error to zero?

COOPER But haven't we got two feedback loops in the models we're talking about?

The diagram you were describing earlier had a feedback loop in it, that is, from a controller which put out phonemes, or something like them, back through the rules for generation and into an error detector. I took this to be completely inside the skull and a fast-acting loop. Then, there is the one that goes around the outside--through the air--that may very well have something to do with how a youngster learns language in the first place, and has to do with how we monitor our speech. It operates on a rather crude and somewhat slow basis; the feedback studies suggest that something of the order of 0.2 sec is an appropriate time constant. These are two distinct feedback loops that we ought not confuse.

FREMONT-SMITH They are probably linked to each other, aren't they?

COOPER Yes, they certainly are.

FREMONT-SMITH So there would be three.

STEVENS Yes.

HOUSE Might it not be appropriate, Cooper, to make some comments at this time about current research on muscular activity, so that we can better understand the problem under discussion?

CHASE Before we do, can we review some pertinent assumptions we seemed to willing to make yesterday? They seem to be as follows: that the speech motor system is capable of functioning in varying degrees between pure open-loop and rigid closed-loop control; that not only are there developmental issues that determine the relative amount of closed-loop control involved, but the contingencies under which the adult speaker is speaking will determine the extent to which he is going to monitor.

Another set of assumptions concerns the ways in which we monitor. These will determine which of a fairly large set of potential feedback loops may come into play. In the most limited sense, we may want to be speaking with great precision under circumstances in which we are formulating spontaneous speech. Under these circumstances, we may be more dependent upon the utilization of afferent activity.

COOPER I think I would agree.

DENES I think you forget that the two feedback systems that we were discussing were, one, the feedback that was postulated in connection with analysis-by-synthesis yesterday--which was a feedback loop involving only the motor images in speech but no execution, and the end-product of any action like this would be a motor instruction going out, say, from the cerebellar level or something like that--and, two, the complete monitoring feedback including acoustic.

COOPER May I return to House's question about research on muscular activity and its relation to speech perception? Let me describe a simple experiment that we could do easily, if anyone has doubts about the results. I would ask one of you to repeat back to me a few words or a short sentense, and the rest of you to listen and decide whether or not the repeated message was the same as the original. Your answer would almost certainly be affirmative, although you would be quite aware that the two sets of sounds, judged simply as sounds, were not at all the same. This would be considered a remarkable phenomenon if it were not so familiar.

We can account for the results of the little experiment I have just described by assuming that the incoming speech is sorted into little bins that correspond to the elemental units of the language. Likewise, the motor instructions for repeating the message are drawn in the same sequence from a similar set of bins.

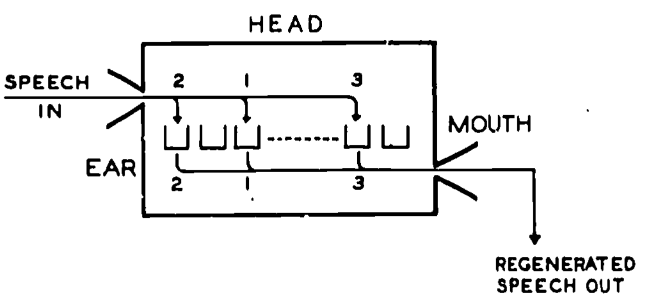

Figure 4. Block head with phoneme bins into which, and from which, phonemes flow in the indicated sequence when a message is repeated.

But are there separate sets of these phoneme bins, one for reception and one for production, or does a single set--perhaps located over on the motor side of the system--suffice? We think that the single set--as I have drawn it--is enough and we are trying to characterize these bins, both in terms of how sharply categorical they are and in terms of the muscle contractions that correspond to them when one speaks.

HOUSE Isn't it generally accepted that when you are talking about linguistic behavior, you're talking about a process of categorization? On the other hand, if you wanted to push the sound wave through a person, you could hook him up like one of Galambos' cats and use his ear as a microphone. In this case you wouldn't have any linguistic processing--the listener is just a transducer. In the more normal situation the input is being categorized--by definition, as it were.

POLLACK I think you're saying more than by definition.

COOPER Yes, indeed.

OLDFIELD By studying the errors made, surely, one can distinguish between these possibilities.

LENNEBERG I think there is empirical evidence that something like that must happen. If you hear, as a nonspeaker of Chinese, some Chinese words and are asked to repeat them, you cannot reproduce them, either phonemically or phonetically, whereas, if you hear something in English which is odd, you can repeat it and you reproduce something that is phonemic. I think nobody would deny this.

LADEFOGED If you were looking at me, in that case, I would deny it.

LENNEBERG Do you think you can repeat such materials?

LADEFOGED I think that Fry and a number of us here would say that if somebody said something to us in a language we didn't necessarily know, it would be an overstatement to say that we could not repeat the words phonemically or phonetically.

LENNEBERG But that is a meagre statement that you are making. I'm saying, if you don't know anything about the language, say, Old Gulish (an invention of mine), the very first time you hear it, you don't even know that this sentence exists. I'm going to speak to you in Old Gulish and you haven't had a chance to make any study of phonemes. In this case, I think, your production of Old Gulish would not be very good.

LADEFOGED But, in the case of Old Gulish, presumably, it is another Old Gulish speaker who will decide.

POLLACK I think you're asking us to assume the identity of the phoneme, and some of us would like to approach speech without making this assumption.

HIRSH Would you accept one restriction on your statement, Cooper--that is, after the learning of the language?

COOPER I was speaking of a practiced speaker performing in his own culture, of course. What I wanted to get at was the mechanism by which speech reproduction can happen; so, too, did Stevens and House. Yesterday, they drew on the blackboard (see Fig. 1) a system that had a feedback loop in it--one that operated through generative rules to perform the categorizing operation that I was talking about a moment ago. Also we talked yesterday about the feedback loop that operates between a man's mouth and his own ear. If you introduce a time delay into this loop, you cause him all kinds of trouble. These are the two feedback loops that we want to distinguish--the little one on the inside, and the big one around the outside. No doubt, there are other loops, too, but these particular ones are our present concern.

IRWIN Would you say once more what you think the little fast-acting loop monitors and reports on?

COOPER As I understood Stevens description, it operated past an analyzer that gave you something comparable to linguistic units and generated something comparable with this information, and then put out speech.

IRWIN Neurally rather than muscularly, is that right? At a neural rather than a muscular level?

HOUSE No, since it is a conceptual model, it has no tissue identification as yet.

IRWIN That's what I was thinking.

FREMONT-SMITH But do you think it is outside the brain, conceivably?

HOUSE Oh, no! I think of the operations as being at a high level in the nervous system, but I don't understand the brain enough to tell you just where they are taking place.

FRY I must say you've got me confused now. This thing you were talking about yesterday was used in reception, was it not?

STEVENS I'm confused, also (laughter), because, as I understand it, you're talking about speech production here. The feedback that we talked about yesterday had to do with speech reception.

LIBERMAN Well, this is reception, too.

COOPER I'm talking about both processes--reception and production--I would like to try not to pull them too far apart, if you don't mind.

HIRSH I thought I understood our illustrious lecturer to say that he was describing this inside feedback loop with respect to the process whereby the continuous input speech wave form got converted into categorical receptions.

COOPER I was interrupted before I got that far, but go ahead.

HIRSH Therefore, this same loop you were using yesterday is in this process?

COOPER In the previous discussion, I described a little experiment that illustrates the fact of categorization in dealing with linguistic units, both in reception and production, but I didn't get far in talking about the processes by which categorization might be achieved. One of these is analysis-by-synthesis, complete with feedback loop. My major difficulty with this mechanism lies in imagining a neural embodiment. I find it easier to think in terms of Hebb's (51, 104) neural networks. Without being too specific, let us imagine a region in the brain that is aroused into patterned activity by both the motor activity of speaking and the sensory inflow from the speaker's own ears. Presumably different patterns (different cell assemblies, in Hebb's terms) correspond to different linguistic units. (We shall call them phonemes for convenience, but without committing ourselves too firmly to this particular unit.) If the activation of a particular cell assembly is always associated with both the production and reception of a particular phoneme, then it may come to pass that the sensory inflow from reception alone will be enough to activate the same assembly and with it the "recognition" of the phoneme. Thus, the same neural machinery would serve for both production and perception, and there would be no mystery about the close link we find between them or in the fact that acoustic cues seem to be organized along articulatory dimensions.

The primacy of articulation, and the location of these phoneme bins over on the motor side of the system are not at all surprising if one thinks of speech as a process for communication rather than as a signal in the air intended to be caught by the ear. Indeed, if one follows through the successive operations that constitute production, the wonder is that any acoustic invariants referring back to the initial phonemes could survive in the acoustic signal. And, lacking these one-to-one correspondences between sound and linguistic unit, we can hardly expect speech perception to be as simple as an auditory sorting of the incoming speech stream.

Let us look at the successive stages in producing speech: we start with an intended message and end with an acoustic wave form--the problem is to infer what operations must lie between them and what effect these operations will have on the form of the message. Whatever the high-level processes used in generating a message, I think we can assume that it exists on some level as an intended phoneme sequence; at least, this assumed sequence is a reasonable starting point for these speculations about the productive process. This, I suppose, is about the level at which we would look for the cell assemblies I mentioned a little while ago.

At a later stage, the neural activity corresponding to the phoneme sequence would have become sets of neural impulses flowing to the articulatory muscles. Does a one-to-one correspondence still exist between phonemes and these motor commands? Perhaps it does, at least in the sense that a particular muscle or set of muscles will always contract when a particular phoneme is called for, regardless of context, and conversely, that a particular combination of muscle contractions is diagnostic for a particular phoneme. I should like to return to these assumptions, since they are the ones we are trying to test in our electromyographic studies of speech articulation. All we can really say at this point is that the relation between phoneme and neural signal might still be simple at the motor command level.

We can be reasonably sure, though, that the muscle contractions at the next stage of production will correspond, one-to-one, to the neural signals. Whatever simplicity of relationship remained at the motor command level will be preserved at the level of muscle contractions. This is important because we can reasonably expect to investigate these contractions by observing the electrical phenomena that accompany them.

But we can be equally sure that the relation between muscle contraction and the resulting shape of the articulatory tract will not be simple. Indeed, the new shape will depend not only on what muscles are active at the moment, but on the configuration that resulted from the preceding commands. In this sense, the articulatory shape is an encoding--as distinct from an item-by-item encipherment--of the motor commands. We do not need to assume that segmentation was preserved up to this stage, but assuredly it will not survive this operation. As an aside, I might say that the special advantage of electromyography over x-ray movies lies in circumventing this particular encoding operation.

Then, there is the final conversion of articulatory shape (and respiratory activity) into sound. The relationships between shape and spectrum are complex, to say the least, and changes in shape are often hidden in silence. The relationship is, in principle, computable on an instant-by-instant basis in the forward direction (35), though not necessarily in the reverse direction, without considering the preceding events.

In summary, then, production involves at least a conversion from phoneme sequence to motor commands and muscle contractions (which may or may not preserve one-to-one correspondences), a complex encoding from motor commands into articulatory shapes and respiratory movements, and then a further none-too-simple encipherment of this code into an acoustic stream. This is what we give the ear--and expect it to pull out the same phoneme categories that we started with! It seems an almost impossible task, unless we let the ear have access to the same coding machinery that we used in producing the speech in the first place.

FREMONT-SMITH Can you avoid the feedback in this description? It seems to me, the moment the muscle begins to contract, we know there are feedback loops which tell the brain how far it is contracted and control the contraction, so it should be the right amount. Therefore, I would say, for every single one of these multiple contractions that are taking place, there are multiple sequential feedbacks which direct this and control it. I raise this in connection with your question about increasing complexity. My guess would probably be that it could be described as decreasing complexity, as the complexity has to be built into the thing from the very beginning.

GESCHWIND I believe there is evidence that some movements can be made quite accurately in man even in the presence of deafferentation.

FREMONT-SMITH You mean, with complete deafferentation? You mean, from habit, then?

GESCHWIND Lashley (87) showed that a man who had a deafferented leg was able to make highly accurate movements.

POLLACK How did he show this?

GESCHWIND He had the man simply move his leg to touch a specific point. Essentially, it was the kind of thing you do in a finger-nose test in a clinical examination.

BROADBENT I've been wanting to make the point that you can shift from a closed-loop to an open-chain system. There is any amount of psychologic evidence on this, as you get more practice.

FREMONT-SMITH But you have to have the practice, and you have to have the feedback in the very beginning.

GESCHWIND I agree. You need the feedback to learn but you don't need it after you have learned the sequence well.

COOPER May I interrupt the people who interrupted me to finish one comment? I would accept the existence of the feedback, but not necessarily its role in normal speech operations. A comparable non-speech example would be playing rapid scales on a piano, where time simply does not allow detailed feedback control. There may very well be side loops which send information about the feedback or about the initial signals to some place, to ask, in effect: "Is this thing going all right? Am I matching up with what is to be expected as this skilled performance runs off? If not, let's call a halt to the whole thing, but if so, let it run."

OLDFIELD There is no doubt there is a kinesthetic feedback. But this does not ensure continuous performance, or, at any rate, does not in the absence of acoustic feedback. If you make people sing tunes with what we call their acoustic feedback knocked out, then, for a few notes they are reasonably accurate, after which the performance goes awry. You can carry on only for a short period of time on the basis of kinesthetic feedback.

LADEFOGED It depends on how well you know the whole thing. You mustn't sing a song that they have sung over and over.

OLDFIELD You can get people to go quite wrong on "God Save the King" or something equally as familiar to Englishmen.

LADEFOGED I would maintain that I can sing "God Save the King" under conditions of deprivation of ordinary feedback, and most of the subjects I have tested on this can do the same, if it is something that they really know quite well and are in the habit of doing.

OLDFIELD Well, my observations have not been the same, but it's a matter of small import.

CHASE I think the remarks about what a human being can do with a deafferented extremity are very germane to the several questions we are discussing. Cooper has reviewed the several ways in which we have been discussing feedback loops, and has suggested that the pattern-matching that underlies the receptive capabilities for speech bears very striking similarity, if, indeed, it is not identical to the kind of pattern-matching operation underlying the utilization of feedback for the productive capabilities of speech.

HOUSE This seems to be almost the error that some of the discussion got into yesterday, by identifying these two things as one, or as a feedback loop. The comment yesterday was critical of feedback loop times when the feedback loop then under discussion was a conceptual one.

CHASE This is actually not what I want to focus on, but, before I get to that, I wanted to outline what I have appreciated as something of an evolution in our discussion of feedback loops--that they are used for both productive capabilities and receptive capabilities. Let's leave open the question, if this is controversial, of the extent to which the componentry and functional operations overlap or do not overlap. But I think Cooper was making comments suggesting stronger overlaps than any that had been made up to this point.

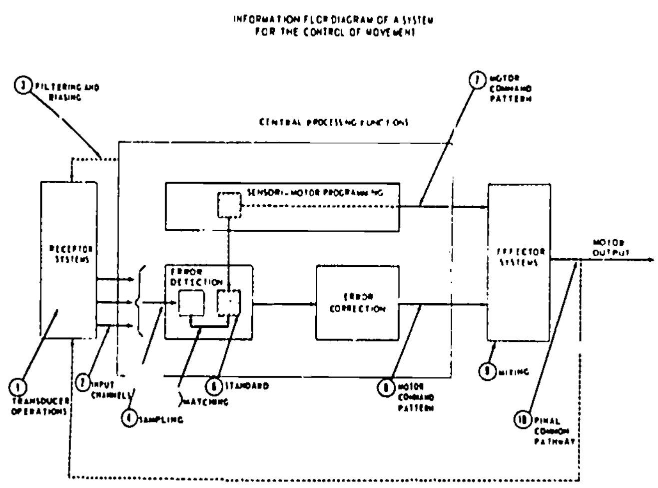

Figure 5. Information-flow diagram of a system for the control of movement.

Could I comment on the deafferentation? A flow diagram of some of the functional components that I would posit to be part of any generalized control system utilizing feedback is shown in Fig. 5 which represents a general system for the control of movement.

At the left of the figure we see the receptor systems that are able to generate information pertinent to the motor activity under control. The several arrows indicate that we can use different kinds of sensory information for the control of any particular motor output. Any control system utilizing feedback requires that errors be detected, and error detection requires that the sensory feedback can be compared against some standard so that a discrepancy between the sensory representation of the output of the system with respect to some standard can be detected. Furthermore, if it is an adequate control system, this discrepancy can be corrected by an appropriate revision of the motor-command pattern that ultimately will be translated into movement.

There are two points that I would like to speak to in considering the deafferentation experiments that have been done in primates. The first is that there are several kinds of sensory information available at any one time. Yesterday, we spoke to the issue of possible hierarchies of importance of these channels with respect to any given type of motor output.

The other point--one that I don't think has yet come into sharp focus in our discussion--is the issue of what we are trying to control? After all, in the most formal schematic model of a control system, we are trying to maintain some steady state of the output. In our discussion thus far we have been talking about accurate motor reproduction of what we hear, or accurate motor generation of speech, but isn't it possible that there are sets of steady-state programs--aren't there several different kinds of steady states? In other words, I think we have been assuming that there is only one, or, at any rate, nobody has chosen to indicate more than one.

STEVENS A steady-state analysis of speech would be doomed before it got started.

CHASE Well, let me speak to the experimental observations that at least raise this question in my mind. Geschwind said that a human being with a deafferented extremity can perform movements with normal control; I don't agree with that. There has been a great deal of work done on the monkey, showing that if you completely deafferent one extremity, by sectioning all the dorsal roots to the brachial plexus such that there is no sensory return from that extremity, the limb is not used for voluntary motor activity (70). But if, in the prior history of the animal you have conditioned a flexion-avoidance response, then following the deafferentation the flexion-avoidance response to the conditioned stimulus will persist.

I think that this set of observations brings into rather sharp focus the fact that the program for the motor activity functioning in open-loop terms is available, and, indeed, the system can operate in open-loop terms for some kinds of movement, but, for others requiring closed-loop operation we see a profound functional deficit. Might there not be parallels for the speech case?

GESCHWIND The experiment Chase has just cited is most interesting and was not known to me. Sherrington's experiments on deafferentation of the limbs of monkeys are often cited (123). I don't believe, however, that it has been shown that deafferentation in man produces the defects which Sherrington found in the monkey. Certainly some humans with gross deafferentation don't show the flaccid paralysis which Sherrington described.

HIRSH How does gross compare with complete?

CHASE In the primate the roots in the brachial plexus are C 5-8, and the first thoracic. If you spare even a single root, there is minimal impairment of movement, so this is not a matter of the motor deficit being proportional to the amount of deafferentation. Apparently, total deafferentation is the issue here.

GESCHWIND Hirsh's question as to whether gross deafferentation may not really be comparable to complete deafferentation is very well taken and it may well be that this explains the discrepancy between the human cases where complete deafferentation probably doesn't occur, and the animal experiments.

Whether the same principles which apply to the limbs also apply to speech movements still remains a problem, and I would agree with Oldfield's views. Speech movements may rely heavily on acoustic (rather than somesthetic) feedback just as there is much evidence that it is visual control which is essential to accurate eye movements.

HOUSE Could I interject a comment here? The Ringel and Steer (115) study that was mentioned yesterday showed that interference with the auditory feedback gave the primary effects, while interference with tactile feedback did not have as great an effect. Is the disagreement that we have at the moment about certain specific neurologic effects? Does our application of these ideas to speech show that we are agreed that both closed-loop and open-loop systems are operating in speech production?

COOPER As far as I'm concerned, no.

FRY No, I don't think I would agree with that. I really think you always have closed-loop operation, in some sense or other.

LENNEBERG It seems to me that we're back in the same wrangle we were in yesterday. We followed House's discussion yesterday by asking: How does the anatomy or physiology fit into the model? At the end of that time, we have decided that we were talking about very abstract feedback systems. So now the discussion of the anatomical situation seems quite out of date, because we have overcome it yesterday when we decided that the anatomy and physiology are really not relevant to the discussion.

FREMONT-SMITH But you can't dismiss anatomy and physiology very long from this discussion. You've got to come back to them again and again.

LENNEBERG We have very much the same problem in perception and in neurology. You can talk about perception in psychological terms but you really can't talk about it in terms of neurology, except in very elementary ways.

FREMONT-SMITH This is the tenuous bridge between neurophysiology and behavior. In a sense, this is what we are trying to build, so I think we must come back to it again and again.

LIBERMAN I wonder whether we are all agreed that at least one of these feedback loops can be almost infinitely short, in the sense that the person reads off directly the motor commands? In this connection, I would like to return to the point that Geschwind was making, about the extraocular muscles. He said that, apparently, there are no proprioceptors, or very few there, and that evidence suggests that we control this movement extraceptually, or visual cues.

Consider the fact, or what I think is a fact, that convergence is a cue for the perception of depth and distance. I think it has been reasonably well isolated now and known to be so. If, for the sake of this discussion, you assume, first, that there is no proprioceptive or tactile feedback, and, second, that convergence is a cue, then, you must assume that the subject has the ability to read off the motor commands correctly. Obviously, you've got to have information about what the convergence angle is. Is this correct?

GESCHWIND There are proprioceptors in the sense that there are spindles in the eye muscles. Their function, however, is not primarily to give information about the position of the eyes.

LIBERMAN All right, I'm accepting that and saying that if you make this assumption, and if you also assume that convergence is a cue for the perception of depth and distance, then you can assume only that the person who uses this cue is somehow getting information directly from the motor command. I mean he is just taking a loop at the level of the commands which are causing his eyes to converge at a particular point. This is a very short loop, isn't it?

GESCHWIND I'm not sure that you don't get visual information from convergence. You can tell when you are converged by the fact that you have double vision beyond the convergence point. You may not be actively conscious of your double vision beyond your point of convergence, but you may still be using this information.

LIBERMAN It tells you that you are converged, but does it tell you how far?

GESCHWIND I don't know.

LIBERMAN In a way this is irrelevant. One question is whether we actually have this kind of evidence. But a more important question is whether there is any reason we couldn't? Is there anything we know about the human being which says the feedback loop could not, in fact, be short?

HIRSH I think this is a very interesting notion--the source of feedback signal being the command itself. I just wanted to compliment you on bringing it up at this particular time, the hundredth anniversary when the doctrine was first announced by Helmholtz, under the title of sensations of innervation (14). (Laughter and applause)

OLDFIELD I want to go back to this question of various types of feedback loop, because I think that one important point there is that in some sense they interact, and I think the more elaborate higher-order loops, if disturbed, can react back upon the more physiologic ones, producing disturbances. If, for instance, I am trying to express a difficult proposition in a foreign language, this may reflect back upon my powers of articulation. It is also true with dysphasics that the more difficult the problem, the more difficult the statement made, or the more difficult the series of commands controlled by more feedback loops, the more liable they are for their articulation to go wrong. I think we have to consider, probably, a whole hierarchy of these loops, with the important ones interacting backwards on each other.

When one is trying to make a statement, there is a very long-term loop which is trying to ensure that, as I produce my words, I am saying what I want to say, and I alter my behavior when I perceive the sentence has gone wrong in such a way that it brings me back. We have, therefore--obviously, I don't know how many there are--five or six loops, and I think the important point is that they can interact backwards towards the physiologic. I don't think we get very far by taking any one of them and assuming that we've got the feedback loop fixed for the phonetic aspect and have therefore explained speech; there are others that may be required to react upon this one.

HOUSE I believe this wrangle started with Stevens' comment about feedback loops in reference to experimentation on the muscular aspects of the articulatory processes. I'm not quite sure that I understand what has happened in the discussion since that point.

STEVENS It is certainly true that if feedback does not play an important role in speech production, then observations of the musculature--EMG signals--will, indeed, bear a very close relation to the speech units. This relation may not be so direct, however, if feedback does play an important role.

COOPER May I say with all deference to Oldfield's point, that it may nevertheless be useful to talk about the feedback loop or whatever mechanism it is that accounts for the categorical conversion into and out of units of the general size of phonemes. This is the mechanism that is operating when somebody receives a message into his ears and produces the same message, but not the same waveform, out of his mouth. There is some machinery inside the head that does this, and if we all agree to talk about this particular operation, we might restrict somewhat the range of our discussion.

POLLACK There is an interesting extrapolation to consider. If some of these feedback loops are being shared between reception and production, perhaps we hear ourselves in self-instruction in the same manner as we hear information in the environment.

GOLDSTEIN We've talked a lot about feedback loops but I think we have to agree that we don't know much about the feedback loops represented by the signals that Cooper put inside the box (see Fig. 4). Maybe, it is worth considering the coding question. I think, if we ask: What is a simple description of speech--speech at the level that Cooper just mentioned--we might get somewhere? If we keep going to the higher levels there are other feedback levels coming in, and those, we know even less about.

I think we've some insight into the anatomy of the nervous system involved in the question of higher-order feedback, but we just don't know enough detail, and we certainly won't get much further with it today.

Now, what about the question of the simple description? We certainly know that it is not the pressure wave, since this is a very poor description of the speech sound. There has been a lot of use of the speech spectrographic type of representation--patterns of spectral energy in time. I would contend that the spectrographic representation is tied quite closely to the coding of signals in the afferent hearing process. There must also be a relationship to the efferent process, speech. In speaking, we are controlling cavities which change spectral energies in time. This close relationship between the physiological processes of speech and hearing is one of the reasons that the spectrographic display has been a useful way to look at these signals.

Now, the question of segmentation. You can, again, look at both physiological processes, hearing and speech.

Actually, we are more interested in what is in the middle than in what can clearly be called afferent or efferent. We can look again at the afferent side, which probably means knowing more about the coding of speech sounds or speechlike sounds in the auditory system. I think this has to be done at the cell-by-cell level, looking at the coding of sound in the system. On the other side, it might certainly be worthwhile to look at it somewhere, and it is probably hard to look at the neural signals that control the speech cavities.

Two descriptions of the units have been mentioned, and this is exactly where I would like to pick up a point, if I may. We can look at the sound spectrogram and hunt for counterparts of the units of language there; or, we could look at the description of the motor commands, that is, what muscles are moving and in what time sequence, and look for relationships between that description and units of language.

COOPER The view of speech perception that Liberman and I share, which grew out of an attempt to do the first of these, leads us to hope that we can succeed in doing the second, and to think that that is where the closest correlations will be found. The nature of the intervening process is open to a lot of speculation. My own bias is to put the emphasis on the main open loop, with some detection of whether or not the process has gone astray--but this is only a feeling.

HIRSH Production--speech production?

COOPER Production or perception. I find it difficult--at the level of events in my kind of model--to keep them separate. They are essentially the same operation--a patterned neural activity.

STEVENS Suppose we consider the production of the consonant b in terms of these ideas. Those of us who have studied the acoustic signal recognize that the acoustic counterpart of b can be quite different, depending upon whether it follows a vowel or procedes a vowel or is in a consonant cluster. But it seems to me that the simplest description of a b is that the lips are closed, and if one tried to look at the muscular activity necessary to make the lips close, you would again come to a more complex description.

LADEFOGED This is an assumption, again, that the latter thing would occur, that you would come to a more complex one. Isn't it at least possible that the simpler commands or the simpler description is in terms of the motor commands.

STEVENS That's the question I'm raising.

LADEFOGED I would have thought that you had a small amount of evidence. It was MacNeilage's (94) work which indicated that, although the acoustic signals varied very greatly and the spectrogram varied very greatly, the articulatory gesture was more or less the same.

STEVENS Not the articulatory gesture, but the muscular movement.

LADEFOGED Yes; the action of the muscles was more or less the same.

STEVENS I'm looking in between these two and looking at the articulatory gesture and asking whether this is equally simple or, perhaps, even simpler.

DENES I think the experiment showed the opposite. The motor command is much simpler than the actual muscular movement, because, as far as I remember, the duration of the frictional part of the f changed considerably, depending on context, whereas the motor command was there for the same length of time, regardless of the context of the f.

HIRSH Can we differentiate motor command from muscle movement? Are there two operations involved? Are they more distinct in motor command as opposed to muscle movement?

DENES Well, I can only answer what I gathered from the Haskins papers, and they seem to show that the action potential has a much closer relationship to the linguistic units--in this case, the phoneme f--than the muscle movement as expressed by the duration of the frictional element in the acoustic output from the spectrograms.

COOPER Well, if I may give an operational answer to Hirsh's question, there are three things that you can do in the laboratory: You can put an electrode on the muscle and measure the potentials, which indicate when the muscle is contracting. You can record the sound and make a sound spectrogram. Or you can take high-speed movies, focusing on the lip action.

I think Stevens' point was that the high-speed movie would give a simpler description than the myogram. We used the spectrogram as a reflection of what was going on, and found that the spectrogram and, by inference, the movie, was more complicated than the myogram. That was the result of the experiment.

CHASE I have one question--a point of information. Several types of analysis of the speech motor gesture have been outlined. I wonder whether you are disturbed at all about whether some of the differences in description are related to differences in sampling--in the sense that the acoustical analysis of the spectrogram gives you all the acoustical information, whereas the EMG description gives information from a finite set of electrodes, and, even when we use x-ray techniques we are looking in a two-dimensional space?

COOPER I think this is an important point, and one really ought to compare spectrograms with myograms taken from electrodes all over the articulators, in which case you would have the task that we all faced when we first looked at spectrograms--the problem of deciding what carried the information. At a later stage, and assuming that we did a reasonably intelligent job of sampling for the electromyograms, what we ought to compare would be these myographic traces and the acoustic cues from simplified spectrograms.

GESCHWIND It seems to me that Ladefoged's view on stress was that, in fact, the acoustic correlation of stress is a very complex one, while the physiologic one is a simpler one.

FREMONT-SMITH Isn't the term <u>simpler</u> almost meaningless? Don't you have to say simpler with respect to something?

GESCHWIND Simpler with respect to the number of components you have to specify in order to get the answer. To describe stress in acoustic terms is very complex, but it correlates very simply with the subglottal pressure.

LADEFOGED This is where I have to agree with Stevens. The results of the muscle movement are easier to specify than the muscular movement itself. The irrelevant factors of how much air there is in the man's lungs and things like that mean that certain muscles are involved in one case and not in the other case--these are completely irrelevant to linguistic stress which depends simply on the subglottal pressure produced by the muscles (79, 82).

GESCHWIND If, in studying stress you simply looked at which muscles were acting, this would correlate poorly with stress, because the stress would also depend on the amount of air in the lungs at the moment. How about the electromyographic commands? Were they simple or were they complicated in the same manner?

LADEFOGED They are complicated in the same manner--in the sense that the activity of any one muscle will increase with respect to the amount of air that is going out of the lungs. If I pull a lot of air into my lungs and I let it out until I have hardly any air in my lungs, I can still stress things--from the point of view of linguistic stress, you will hear the same degree of stress. In the case where I've got a lot of air in my lungs, I am using the intercostals minimally, while with very little air in my lungs I am using the intercostals maximally. Actually, in both cases, you can say that stress can be correlated with an increase in the electromyographic activity of the internal intercostal muscles.

LIBERMAN I think it's clear in any case that what all these people are saying--that is, Ladefoged, House, Stevens and the Haskins people--is that there is a simpler relation, a more nearly one-to-one relation, between some aspect of the articulation and the perceived language than there is between the acoustic signal and the perceived language.

Now, we can go on from there to consider what aspect of the articulation. I'm not suggesting that this is a trivial question, but I think the more important question is whether, first, we accept this general point, and, whether, second, we assume that, in decoding the complex signal, the listener somehow makes reference to the articulation. On this, we seem to agree.

STEVENS On your first point here, we can't necessarily agree. For example, I can describe a vowel by three numbers in the acoustic domain, and that will give me essentially a complete description of the vowel--three resonant frequencies of the vocal cavities or formants.

HIRSH I can describe it more simply than that (opening mouth). Isn't that simpler?

STEVENS It's hard to write down a description for that. I would say that the description is simpler, perhaps, if it is in terms of formant frequencies. For consonants, I would agree with your statement, however.

(A short recess was taken at this point in the discussion.)

HOUSE I am a little concerned about the clarity of some earlier points. I am calling upon Stevens to clarify some of his earlier remarks.

STEVENS What I have to say bears on two questions that may have seemed unrelated, but which I think are related: One, that of feedback, and the other, that of which level of description is the simplest.

Perhaps, I can talk in terms of an example. Suppose you as a talker are about to generate the consonant t. Your tongue is initially in a certain position, say the position for a vowel preceding the t; you then give instructions to the tongue, and these instructions are dependent upon how far you have to go in order to make the proper closure for the t. In other words, the motor instructions could well depend upon the error signal, the distance necessary to go from the vowel to the consonant, or, if you like, they could depend upon the difference between your intentions and your present state.

This would then indicate that, depending upon what the preceding vowel was, you would have a different motor command, because you have a different error or distance to go.

This reasoning would suggest, therefore, that a simpler description than the motor command might be the articulatory position--in this example, the fact that you have to make contact between the tip of the tongue and the hard palate.

LIBERMAN I would simply say that we would certainly agree, but we would put it somewhat differently. I believe we have been thinking in terms of what Cooper said before: the increasing complexity as you go downstream from the neural signals in the brain--which somehow represent phonemes--to the speech wave itself. We would certainly agree that the simplest form must certainly be these intentions, if you will.

The question, then, is whether the relation to linguistic structure is still fairly simple by the time you get down to the final common path. You were saying that, conceivably, it is not, and I think that it is a very reasonable position. You were, moreover, if I understood you correctly, saying that you can construct a model which would work from the intentions to the movement. But we are agreed that either of these, in a sense, is simpler than the complex acoustic signal (28, 93).

STEVENS Yes; all right.

HIRSH I find these remarks very encouraging, if I may interject a comment, because I assume that by the time you chaps get ready to complete these descriptions in articulatory terms, then, you can tell me, and I can tell our teachers, exactly what to tell a deaf child to do. This does not seem to be the case at the moment.

HOUSE I would like to disenchant you. (Laughter) I have the feeling, with reference to your present comment and the earlier comments you made about being able to program the production of a particular sound, that the description you seek is completely without acoustical reference. I think the so-called myographic descriptions we have today are slightly misleading because we already have a phonetic description of the activities. Most myographic studies go to a point in the vocal tract where we know activity is critical and get evidence that the activity is actually going on. Cooper has already pointed out that we don't have a myographic description of the activity along the entire system.

HIRSH <u>Myographic</u> is not what you mean by <u>articulatory</u>, is it, Stevens?

STEVENS That's right.

COOPER To answer House's implied question: It is partly true that measurements have been made just where activity could be expected, and for the obvious reasons. This is not entirely so, even for the preliminary studies. For example, in investigating velar closure with the consonants <u>p</u>, <u>b</u>, and <u>m</u>, electrodes were put all over the upper and lower sides of the soft palate, on the faucial pillars, all over the back wall of the pharynx, and around the whole area known from phonetic considerations to be involved in nasalization. These areas were all investigated, but in only one area was there comparatively simple, one-to-one correspondence of activity with oralization as distinct from activity that was present, more or less, whenever there was speech.

It is this kind of one-to-one correlation that characterizes a <u>diagnostic</u> gesture. The conclusions to be drawn from this study of <u>p</u>, <u>b</u>, and <u>m</u> were that we oralize <u>p</u> and <u>b</u> but we don't do anything for <u>m</u>--just let the velum hang. Thus, it is only the velum that is either actively contracted or passively dropped. This is a little more of a description than one had before. It is a description which meets, I think, all of your requirements for having looked at everything--in this case everything in a plausible region--before deciding what is significant.

GESCHWIND Did you say that the bursts of action potentials in the muscles may precede the production of the sound by 100 msec?

COOPER They regularly do just that.

GESCHWIND That seems important to me because it means that the gaps between sounds are not as long as we think they are, and there may be an inherent limitation to speech rate.

HOUSE What gap between sounds did you have in mind?

GESCHWIND I didn't really mean gaps between sounds. I just meant that the rate of speaking must have some limitation on it.

COOPER It was certainly interesting to us that the rapid changes you see in sound spectrograms do not seem to have their counterparts in myograms; that is, the time scales are much slower in myograms, and there is much more of a shingle-roof effect, that is, overlapping, than you find at the acoustic level.

LENNEBERG Somehow, this bothers me, because everything you see in the spectrogram must be muscularized somehow.

COOPER What you see is the consequence of an encoding operation on the muscular operation.

LENNEBERG But couldn't it be that you simply haven't tapped enough muscles and don't have the whole picture, or that your signal is too gross?

COOPER This is always a possibility, but consider how the encoding might go: If you are making a sound as you close the lips, the sound can get out with little impairment until the very moment when the lips close; thus, the turn off can be very much faster than the gesture that caused it. Really, all I am saying is that when you look at the gestures, you are surprised at first that they seem rather slow.

HOUSE At this point, can we create a bridge to go over into the discussion of acoustics? I think a way to do this is to point out that part of our trouble in describing speech production is that we have to pick a level at which to make a description, and that at almost any level we decide to make this description, there is great complexity. Our major task at every level is to reduce the complexity to as simple a picture as possible. I'm not quite sure that we must assume that, as you go downstream things are always more complicated. Often, the complication on one level can be reduced to be as simple as descriptions at any other level. Here we have been using articulation in more than one sense--in talking about the activity of the muscles, and sometimes in talking about the general shapes of the acoustical system. Therein lies another story.

STEVENS Perhaps, I might spend a moment or two in reviewing what is known about the relation between the articulatory and the acoustic domain. This work is primarily due to the efforts of our friends in Sweden (35).

My first comment would be, reinforcing House's statement, that we feel the transformation between the articulatory domain and the acoustic one is now fairly well understood. It is true that we often do not know in detail the shapes of the vocal cavities; I am assuming that articulation is known. I feel, however, that once the shape of the vocal cavities can be described, we can then predict for the most part the sound output, and the problem comes down to what shape the vocal cavities have.

LADEFOGED Of course, I would disagree with that, because I have been looking at, perhaps, some different languages (77), and I can't predict the sound when somebody says something in a language with clicks like Zulu, or many of the other unusual sounds in African languages.

STEVENS Part of this, though, is not knowing what the articulation is. I think, if you put the acousticians to work on the job, and if you describe the articulation to them completely, they could come up with a predictive sound output. The clicks, however, do represent a little bit of a problem, I will admit.

LADEFOGED I have quite a lot of data that I know fairly well--what the articulations are and what the sequences of movements are--but I haven't yet got as far as dealing with the more complicated simultaneous articulations. I just don't know what happens in many languages where you have a k and a p simultaneously. This situation where I can tell you exactly what the relations of the articulations are and where one articulation occurs with reference to the other, I don't think we yet know what the sound should be.

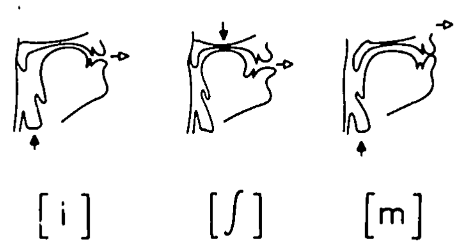

Figure 6. Sketches of articulatory configurations in the midsagittal plane for three speech sounds, as indicated. The solid arrows show source locations and the open arrows show output locations.

STEVENS Maybe I should have qualified my statement by saying that one question which we can't answer completely is: What are the acoustic characteristics of the sources? Perhaps, I should back off a little and review what I mean by source and what I mean by articulation.

One often looks at the articulatory structures in a midsagittal section. Examples of such sections during the production of three speech sounds are shown in Fig. 6. It is recognized that you can make a fairly clean dichotomy between the sources of sounds in the vocal tract, and the resonators that exert an influence on these sources. In the case of the production of a voiced sound such as a vowel, shown at the left of the figure, the source is created by vibration of the vocal folds into these cavities. This glottal source has a certain spectrum, and this spectrum is modified depending upon the shapes of the cavities. The characteristics of the source are fairly independent of the shapes of the cavities.

FREMONT-SMITH Does sound pressure include the rate of flow of air?

STEVENS No, the sound pressure includes only the acoustic components, not the direct current.

FREMONT-SMITH Do you need the movement of air, also?

STEVENS You need the movement of air through the glottis so that you have something to interrupt, so that you can create an AC from a DC.

When I say that this process is understood, I mean that if we know what the shape of this cavity is, we can predict more or less what the sound output will be. You might have to use a computer to determine it, but it is now known how to program the computer. The possible exception is that we don't know everything about the characteristics of the glottal source. It varies from one individual to another, and it will depend upon emotional factors, and so forth.

In the case of the nasal consonant shown in the middle of the figure, there is coupling to the nasal cavities. Provided that one knows the shape of the nasal cavities, one can, again, predict what sort of sound output one will get from this kind of configuration. It is a spectrum that is somewhat irregular and will depend, of course, upon where the closure in the vocal tract occurs, that is, upon which nasal consonant is being generated.

There are other sounds in which the DC air flow through the cavities is caused to pass through a constriction or over an obstruction and to create turbulence, and therefore noise, in the vocal tract. This is another type of source whose properties are not thoroughly understood. This source is, in turn, modified by the shapes of the cavities around it, and you obtain the sound output at the mouth. The spectrum of this output can be predicted theoretically.

One can say, then, that the transformation relating sound and articulation is known, although there are some details that still need study. As a matter of fact, one can assert that this transformation from articulation--or at least from articulatory instructions--to sound is something that we all know unconsciously, because, after all, we hear the consequence of the modifications we impose on our articulations.

HIRSH Would you modify your general conclusion about this transformation to the extent that it is certainly better known for certain classes of speech sounds than others, and that, in fact, the weakest link in this transformation is what the phonetician calls manner of production?

STEVENS Yes, that is correct; the role played by the place of articulation is better known at present. Place of articulation has to do with the cavities--what we call the transfer function--and the manner of production has to do more with the sources and their time characteristics, and that is not quite so well known.

DENES May I ask you two questions? First, you mentioned the shape of the vocal tract as the key to the transformation. Your diagrams in Fig. 6 show only the mid-sagittal plane of the vocal tract. Does the cross-sectional area matter as well, and if so, can you deduce it from the sagittal dimensions? Is it the cross-sectional area alone, or rather the variation of this area along the length of the tube, or does the shape of this cross-section matter also? For example, assume two tubes both with a constant cross-sectional area along their entire length; one tube has a constant shape, say circular, and the other varies its shape, bulging at one point whilst contracting at another. Would these two tubes have the same resonances?

The second question is whether the transformation from articulatory shape to acoustic properties is reversible.

STEVENS The answer to the first question is that something is known about the relation between the midsagittal dimensions and the area function. What you need to know, as far as predicting the sound wave is concerned, is the cross-sectional area of the vocal tract at each point along its length. Given the anatomy of one individual, you can predict this cross-sectional area reasonably well from observation of the midsagittal plane. As far as the acoustic behavior is concerned, it doesn't matter whether a given section has a wide narrow area, a thin wide area, or a fat round area.

DENES So you really answered two questions. One is that it is only the cross-sectional area and not the shape that matters, and secondly, that for any one individual you can predict the area from the midsagittal plane.

STEVENS Yes. I am suggesting that this is possible, although it has not been done for many individuals.

DENES This is only for vowels, or is it also for consonants?

STEVENS For consonants, too.

The second question related to the inverse transformation: Given the sound wave, can you predict what the articulation is? This requires, I think, knowing more about the constraints on articulation than we know at present. There are many vocal-tract shapes that could give rise more or less to the same sounds. Not all of these vocal-tract shapes can be generated by a human being.

DENES Are these alternative shapes possible ones in view of what we know about the human vocal tract or are all but one shape impossible?

STEVENS I don't know the answer to that. There might be more than one.

COOPER Don't you have to consider two kinds of transformation? One is the instant-by-instant transformation, and the other is the dynamically changing one. I am speaking of dynamics over something of syllable length or thereabouts.

STEVENS Yes. I'm only talking about the transformation between the articulation and the sound wave at one instant of time or a short interval of time. I'm not talking about the dynamics of the situation at the moment.

COOPER You will probably have to invoke that if you want the inverse transformation, however.

STEVENS That's right. It has been demonstrated experimentally that the kinds of mathematical models people have used for making this transformation are indeed valid, because people have built speech synthesizers that, in a sense, mathematically describe these transformations, and these synthesizers will generate speech sounds. The studies of the acoustics of speech production have not only led to descriptions of this transformation, but have also led to a description of ways of synthesizing speech signals, that is, of specifying what parameters are appropriate for describing speech signals.

In essence, it has been suggested that if you know what the formant frequencies are, or what the resonant frequencies of the vocal cavities are, this is really all you need to know. You can predict the rest of the sound spectrum just by knowing these resonant frequencies. This is a slight oversimplification, but it is more or less true. So if one can measure the resonant frequencies--the bars that you see on a spectrogram--this is the whole story for vowels. For consonants, you may have to measure more than three; for vowels, you generally need to measure only three.

HOUSE Is it possible that you cannot measure some of these things on a spectrogram?

STEVENS Sometimes the spectrographic display does not have the resolution to measure certain features.

DENES The source spectrum may also influence the final output. The source spectrum might have its own peaks and valleys on which the transfer function of the vocal-tract impedance is imposed. The final result that you pick up and that you see on a spectrogram is the product of these two. How can you tell which are the formants?

RISBERG This might be a problem. I think that the interaction between the source and the vocal cavities can be considerable sometimes but not enough to make the identification of the sound difficult. The interaction is of two kinds, first a change in the over-all spectrum shape and second a change in resonant frequency during the excitation cycle. You can measure this change in resonant frequency but if you synthesize speech without taking this factor into account, you get a very small change in quality. We have done some experiments with this in our laboratory. This interaction has also been introduced in some vocoders.

HIRSH Is this an interaction that is different for different persons but for all vowels, or is it an interaction that changes with the vowels as well, within one person?

RISBERG The amount of interaction changes from person to person and with how you speak. You have one resonant frequency in the vocal tract when the glottis is closed, and another resonant frequency when it is open; these can be measured using the inverse filtering technique.

They are quite obviously different, but the perceptual importance of the difference is not very great.

OLDFIELD Do you know what happens with song? You have a great variety of fundamental frequencies. Does one produce the consonants in the same way, on the basis of this carrier each time, or are they separate; or does one keep the consonants separate from the vowels and really sing the vowels?

RISBERG I don't know. The source is one of the real problems in the description of speech at present. The techniques for studying the source are not so easy to handle as those used to describe the influence of the cavities.

OLDFIELD But you have a technique for doing it, don't you?

RISBERG Yes. There are several techniques now, but none of them is very convenient. They are all very time-consuming.

POLLACK Could you give us an idea of the range of variation in terms of the change in the resonant frequencies, due to the fact that the glottis is open or closed?

RISBERG The amount is about 10 per cent, I think.

COOPER Is the effect primarily one of change in frequency or change in damping? Both are involved, I am sure.

RISBERG Yes. We have not measured this so much. We have been more interested in the source function and not measured the influence on the resonant frequency.

DENES The first time I had this really dramatically demonstrated was many years ago at Walter Lawrence's laboratory where he had a time-domain inverse filter. He could manipulate his filters until the glottal wave appeared reasonably smooth over one part of the glottal cycle, but not during the remaining part of the glottal cycle, or vice versa. The two parts of the glottal cycle probably corresponded to its open and closed phases. He had to manipulate both the frequency knob and the bandwidth knob in order to rebalance when going from the open to the closed phase, showing that both the frequency and bandwidth of the formants were affected.

RISBERG Yes, exactly.

HIRSH If I sing a scale on the same vowel, are you saying that the formant frequencies that characterize that vowel change as I go up the scale?

RISBERG Yes, I think so; certainly, but not a very great change.

HIRSH About 10 per cent?

RISBERG It depends. Your formant frequency will oscillate between the vocal periods, so you have two resonant frequencies, one in one part of the period and another in another part. What you hear is some kind of average frequency, so a 10 per cent change would not amount to much.

FRY But you are saying this 10 per cent change is within the two parts of the glottal cycle. Hirsh is asking whether you also get a change as you change fundamental frequency, is that right?

HIRSH Yes. If I sing a scale from C to C using one vowel during that time will the first and second formants remain at the same frequencies or will they change?

RISBERG The change is not very apparent in the spectrogram.

HIRSH It is smaller than you could see on a narrow-band spectrogram?

RISBERG Oh, yes.

LADEFOGED This assumes that Hirsh is a good speaker, rather better than most of us. When we sing a scale over an octave, in order to stiffen the vocal folds, we move the whole glottis up or down. Of course, as you move the glottis up, this makes the vocal tract shorter, and this raises the formant frequencies.

RISBERG Of course, a sound spectrograph is not a very good instrument to measure these very small changes.

LADEFOGED There would be quite a considerable change, and I can't give you a figure. Fry, do you have any figure as to how much?

FRY I'm not sure, but I agree that it changes.

LADEFOGED These are two different things. One is the change in the open-closed ratio within the cycle with a given vocal frequency; the second thing is a change in vocal frequency, forgetting what goes on within a given cycle. I am suggesting that, although people could be trained so they did not move the glottis up or down, for most of us, high notes are produced with the glottis higher up, in order to reduce the tension on the vocal folds and therefore have slightly higher formants.

STEVENS This rise would be one or two cm, and that would, again, produce a shift of, roughly, 10 per cent. Well, while we're talking about problems of the glottis and the subglottal system, I believe that those are the areas where acoustical work is being centered these days, and I think what I cavalierly said is, perhaps, not true. We do not know as much as we should about the source. I wonder, Risberg, if you have any other comments to make about what you and your colleagues are doing in Stockholm?

RISBERG What we are doing at present is to study and develop techniques for looking at the source. We are working with three different techniques. The first method is direct observation of the movements of the vocal cords by means of high-speed films or measuring the light-flow through the glottis. We have not made any high-speed films yet, but we have worked with the light-flow method together with Sonesson, who has developed the method (126). You put a lamp on the outside of the throat below the glottis and measure the light through the glottis by means of a photocell in the mouth.

The two other methods work from the acoustic signal. In the first a spectrum-matching technique is used and we get the frequency spectrum of the source (129) and in the second the acoustic speech signal is passed through a network that is the inverse of the vocal-tract transfer function. In this case we get a curve showing the volume velocity through the glottis. We have compared the light-flow method and the inverse-filtering method and the results agree quite well (36).

COOPER How good an area function do you think you get from the transillumination?

RISBERG I would say it is quite good. It depends a lot upon the subject, of course. It is not easy to get very good pictures, and, in all these studies, the surrounding noise is a very great problem. For instance, in the inverse-filtering technique, if somebody opens a door in the building somewhere, you get a disturbance in the recording.

DENES How do you get a faithful original recording? This must be a great problem.

RISBERG You must have a recording going down to zero frequency, so you use an FM tape-recording system.

DENES What kind of microphone do you use?

RISBERG A condenser microphone that goes down to about one or two cycles. Then, the whole inverse filter responds down to very low frequencies. But this means that all the low-frequency noise will get into the system, and this is very, very troublesome.

KAVANAGH Where is your light source?

RISBERG In Sweden, we have an external light source below the glottis, and the light goes through the tissue on the midline below the larynx.

KAVANAGH You are working with the human, and not a cat?

RISBERG Oh, a human! It has also been tried, reversing the lamp and the photocell (at the Haskins Labs). You put the light source in the mouth and pick up the light that emerges here (indicating lower part of the throat). It seems to work fairly well. I don't know if there are any results from these tests, though.

GOLDSTEIN Did you indicate that in speech discrimination, the cavities below the glottis are less important, or the shape of them is less important, than the ones above?

RISBERG Oh, yes. I think, of course, the cavities below the glottis will probably influence the source waveform. How they influence it, I don't know exactly, but this has been discussed by van den Berg (10) and others.

STEVENS While we are talking about the glottal and subglottal systems, it seems to me that we lack physiological data on these systems. It is not yet understood what the mechanism of vibration of the glottis is, and how this is brought into vibration by the subglottal system and by the various muscles in the larynx. I wonder if you have anything to say on this, Ladefoged, or, perhaps, you could lead into some discussion of your own work?

LADEFOGED I really haven't worked on how this is made to vibrate, at all. I assume that the work of Faaborg-Andersen (34) which has shown how the muscles work in order to pull the vocal folds into the different positions necessary, does indicate fairly well what happens. The vocal folds get drawn fairly close together, and there is a current of air there that sets them vibrating due to a Bernoulli effect. Is this an oversimplification that worries you?

STEVENS These days, there seems to be some discussion that the muscles of the larynx do not act to bring the vocal folds together, but, as a matter of fact, to keep them apart some distance, at a resting position. Then the air flow begins the vibration of the vocal folds through the Bernoulli effect, bringing them together.

LADEFOGED My impression of the Faaborg-Andersen data was quite clearly that during the rest position, the vocal folds were a little way apart. He asked his subjects to phonate; he gets some electromyographic potentials out; the vocal folds come a little bit closer together, not completely closed, but distinctly closer than in the ordinary position for expiration. When they are this little bit closer together, then, the Bernoulli effect can take over and the air flow going through them can set them into vibration.

STEVENS I guess my point is, this has not been set down in mathematics, and no one has predicted what the output will be, taking all these factors into account.

HIRSH This sounds very simple, too simple. I thought that the action of the vocal folds was more complicated. When you say the Bernoulli effect takes over and the folds are set into vibration, is this free vibration that you mean? I thought that there was a rather alternating action between a pressure buildup, so to speak, on an almost closed glottis that threw them apart and so on. Is this conception, which I learned for lecture purposes three or four years ago to be supplanted by a new one? (Laughter)

STEVENS As I understand it, there is some question as to the role of the subglottal pressure in performing the separation of the vocal folds.

HIRSH Let's say the muscles perform the separation: What happens after they come nearly together?

STEVENS Well, the Bernoulli effect will bring them together, and then they will sort of coast the rest of the way because of their mass. They will become closed, and then either the subglottal pressure or the muscles will pull them apart again.

DENES The shades of Husson (60, 61) now are rising again. (Laughter)

HIRSH Are they pulling apart so many pulls per second?

STEVENS Oh, no.

HOUSE There is a problem in trying to put some sort of mathematical model onto this operation to make some description of the kinds of forces that are operating at the larynx during the phonation. In the classic description there is a loose statement that says: You need a little more pressure below than you have above, and then this forces the folds apart allowing an equalization of pressure above and below the glottis, and things come together. In more recent years, the Bernoulli forces that are operating in the larynx have been added to the description, but these two ideas have not really been well integrated. For example, myographic activity is usually interpreted as supporting the idea that the musculature is trying to come to the midline, but it could equally be interpreted as evidence that there is muscular activity present that is balancing out the Bernoulli forces that are exerted toward the midline.

COOPER But I would have thought this would take different muscles, and that the myographic signal would have been found in different places, depending on whether you were pulling together or pulling apart. As I remember Faaborg-Andersen's work, though it was very far from a complete description, it was consistent, as far as it went, with the pulling-together hypothesis. Also, the myographic activity happens a long time ahead of the vibratory acoustic phenomena.

DENES May I ask, since I am not familiar with the latest work, is there a return now to a possibility of what is called a neuro-chronaxic theory?

HOUSE No, we're not suggesting anything like that. We're merely questioning whether things are approaching the midline through muscular action primarily, or going away from the midline primarily through muscular action. There is no denial that there is muscular activity.

FREMONT-SMITH But isn't the fact that paralysis of the vocal cords leads to closure relevant here? I think this is correct. It may lead to suffocation.

HOUSE I don't think these cases bear on this particular problem. This is more a problem of how to make a physical interpretation of the behavior.

GESCHWIND It is certainly true that the vocal folds, if you cut the tenth nerve, come together. But there are two reasons why they might do this. They might do it because the only muscular force which was normally exerted on them was one which separated them, or it might be that although the muscles both brought them together and separated them, the elasticity of the system brought them together when total paralysis occurred.

FREMONT-SMITH But if the elasticity of the system brought them together, it means that there must be a constant tension pulling them apart any time they are separated.

GESCHWIND That wouldn't answer the question as to whether the folds are being brought together in normal conditions by muscular activity or by the drop in pressure because of the air moving through.

FREMONT-SMITH Or both.

GESCHWIND I agree.

LENNEBERG I would like to report on some very recent studies in Germany on the fine anatomical structure of the vocal muscles. Berendes (9) found that the histology of the vocal muscles, electronmicroscopically, is rather different from that of other muscles in the larynx. I think the main features there were a very specific increase in mitochondria, which might be relevant to this discussion, although I don't know exactly what it means. The usual teaching is that the concentration of mitochondria has to do with energy mobilization; that, presumably--although this is all speculative--indicates that these muscles can exert mechanical actions that other muscles would be much slower in doing.

FREMONT-SMITH Or, maybe, that they have to resist greater forces.

LENNEBERG It could be. We don't know what it means, physiologically, but I think it is relevant to this discussion, that these muscles are highly specialized.

The other specialization found--and this is work by Rudolph (119) and Paulsen (109, 110)--was the finding of spindles, but spindles which were very modified; in fact, they showed morphological similarities with structures in the eye muscles. They were not identical with them, but they were closer to the structures in the eye muscles than any other skeletal muscle.

FREMONT-SMITH Is this relevant to the fact that many of the vocal muscles and the respiratory muscles and the eye muscles have both voluntary and involuntary controls?

LENNEBERG I can't answer that. It does seem to be established now, and other people have confirmed it, that the vocal muscles by themselves are histologically rather different from other muscles. They are different from other laryngeal muscles.

FREMONT-SMITH But you relate them to the eye?

LENNEBERG Well, Rudolph mentioned that, and there are some pictures on this available, as references.

GESCHWIND Since the evidence on the whole indicates that it is the relaxation phase rather than the shortening phase which uses energy in the process of muscle contraction, the high concentration of mitochondria in the vocal muscles may be related to a need for rapid relaxation of these muscles.

FREMONT-SMITH But your statement about the contraction in emergencies is not really complete, is it? You're not suggesting that when a muscle is thrown into tension and maintained that way, there is no energy being used?

GESCHWIND The shortening phase involves the passive use of energy, that is, it involves no metabolic expenditure. In fact prolonged shortening characterizes rigor mortis.

FREMONT-SMITH It also happens when you're in rigor liveliness. (Laughter) They are two forms of energy. One is the building up of energy for future contraction--or am I wrong?

GESCHWIND The muscle is equivalent to an extension spring. The energy that you have to put into the system from without is the energy that you use in stretching it. The energy of contraction in this case is just a downhill process and is not metabolically linked. Metabolic energy is used for extending the muscle, but is not used in contracting it.

It may be possible that the vocal muscles that Lenneberg mentioned are designed for rather rapid relaxing.

STEVENS While we're talking about these physiologic problems, do you want to say something about your physiological work, Ladefoged?

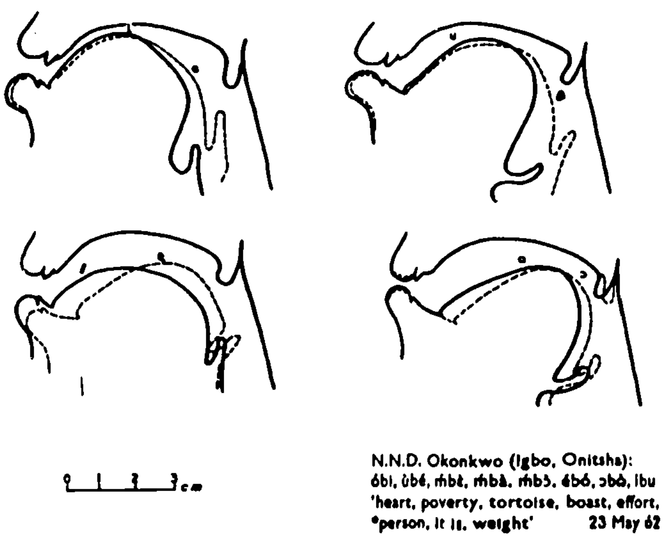

Figure 7. Tracings from a cineradiographic film showing vocal-tract shapes in two sets of contrasting vowel sounds in Igbo. From (77).

LADEFOGED I don't think I have anything particularly relevant at the moment, except to say that I have been preparing studies that try to look at the muscular action, not in the respiratory system, which, I think, we now know quite a lot about, nor in the larynx, which I thought I would bypass because other investigators are working there, but I thought I might come and join the Haskins group--not that they aren't worthy investigators (laughter)--but because there is a great deal to learn about what happens in the muscles of the tongue.

I think that my starting-off point is, perhaps, a little different from theirs. The problem is posed by the data shown in Fig. 7. This is the kind of thing many of us have been worrying about. These are some vocal-tract shapes characteristic of some of the vowels of Igbo, a language of Eastern Nigeria.

This language has two sets of vowels. The difference between the two sets of vowels, physiologically, seems to be that the whole of the pharynx is more contracted in the set shown with the dashed lines; the difference is clearer in the upper pictures than the lower pictures. The two sets of vowels are linguistically distinct, in the sense that they never occur in the same words. It is one of these cases where the words are largely composed of two syllables--see the text at the bottom of the figure--and there seems to be some overriding feature, a property of the words as a whole, to have the pharynx, if you like, contracted or not contracted.

Well, our kind of problem is to try to say what muscular actions are doing this, and this is the same kind of problem as Cooper and the Haskins people are doing. I'm trying to get at a description of the muscular actions that can be responsible for this kind of thing.

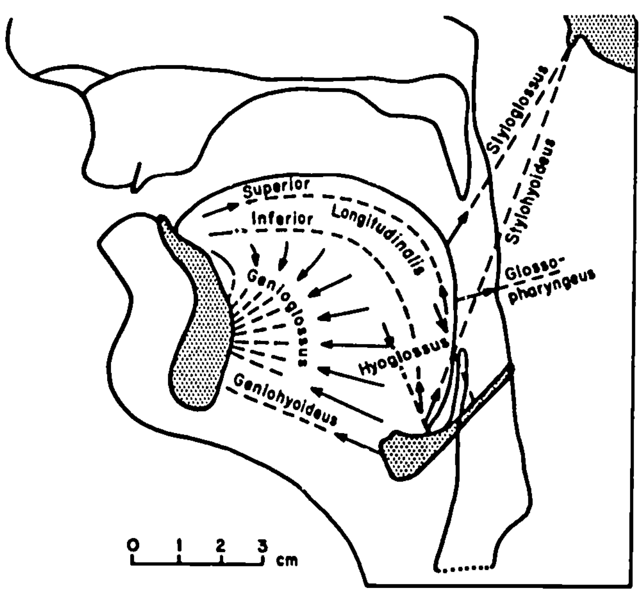

Figure 8. A scale diagram based on x-ray data and direct measurements showing the directions in which forces might be exerted by the tongue musculature. From (78).

The way I have been approaching this, and I would value any comments on it, is to take, not a speaker of Igbo, which was the previous language, but myself, because I can get better radiographic data on myself. I was able to find out fairly exactly where a number of forces might be exerted. One can find the styloid processes on a lateral radiograph; one can see, of course, the hyoid bone; one can see the part of the mandible that is relevant, and, just about where the genioglossal muscle attaches. On the x-rays that I have, you can see the line separating genioglossus and, probably, geniohyoideus, but I wouldn't press that one too far. But I do get fairly good pictures of where the extrinsic muscles of the tongue are, so that I can say that the system of forces acting on my tongue is as shown in Fig. 8.

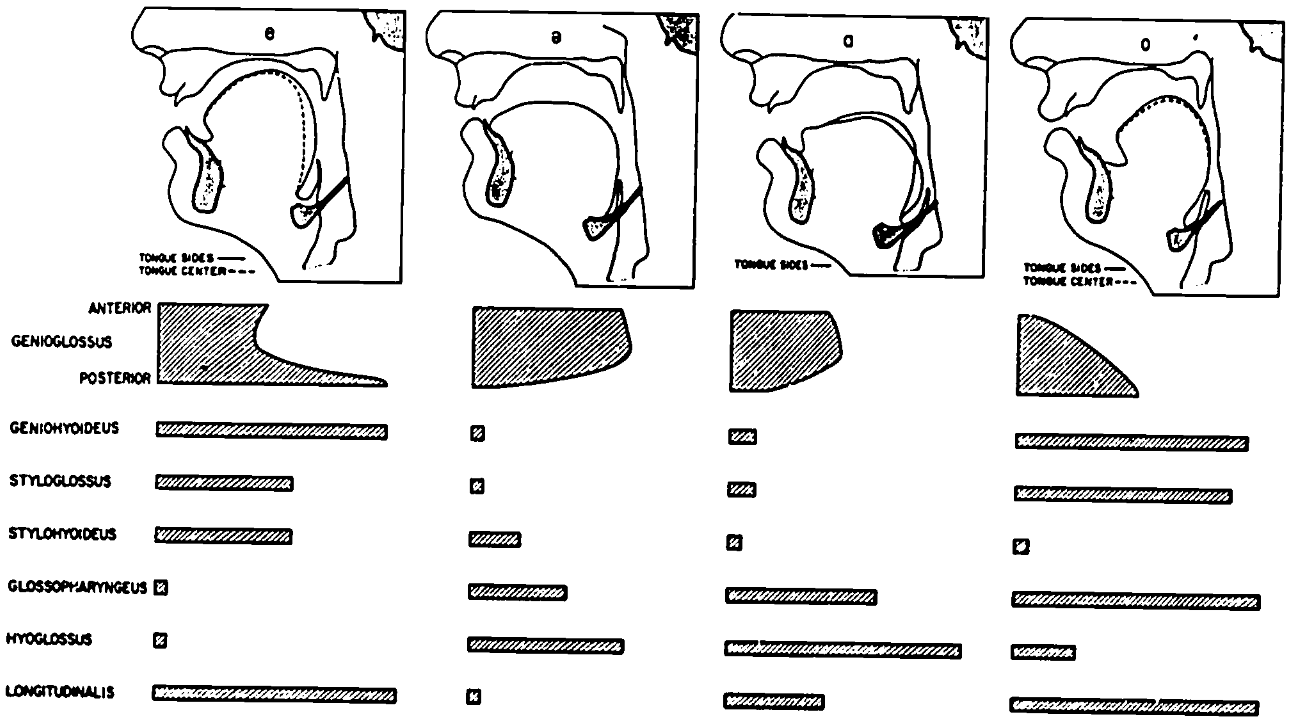

Figure 9. The relative forces probably exerted by the tongue muscles in the formation of various vowels. From (78).

What we are now trying to do is to express this system of forces in terms of what happens in a number of different vowels. The top part of Fig. 9 shows the x-ray positions for a number of vowels, and the bottom part shows an estimation of what the muscular actions must have been in order to have produced that vowel, based on the fact that we know where the muscles are. We can say to what extent the muscles are stretched or contracted.

Now, this is what I would value some guidance on. To what extent, through the knowledge of how much a muscle is stretched or contracted, is one entitled to estimate the degree of force which it must be exerting? On that kind of basis, I have made the estimations, and you can see that these agree, really, with Stevens' point. But I come up with a fairly complicated kind of statement about possible muscular actions, introducing those shapes.

COOPER May I add that MacNeilage and Sholes (95) at Haskins Laboratories have been working on very much the same kind of problem, using electrodes along the tongue just to one side of the center line, from the tip to as far back as you can get. They used 13 positions, as I remember, roughly a centimeter apart at the front and about two centimeters apart farther back. They were taking muscle potentials as a function of time during utterances of various vowels, all for one person to get comparable data. They interpreted their data in much the way you have, except that they were asking about electrical activity, and would also have given the x-ray shapes that were observed.

As to the details, you will have to talk to them, but a good deal can be learned from these two approaches. Whether this gives a simple description or not is a matter of opinion, but it has to be compared with a cross-sectional description at many points along the tract. How many points would you say, Stevens?

STEVENS Say, twenty.

COOPER So the shape description is not a very simple one, either. One other point was that MacNeilage and Sholes concluded that in some cases, a single muscle had to be considered as operating functionally in more than one part. For example, the forward portion of the genioglossus was active at some times and the remainder of it not so active; at other times, the part that attaches into the root of the tongue was active but not the forward part. What the anatomists describe as a single muscle does not seem, from their results, to operate as a unit. This, of course, may not surprise anybody but me!

LADEFOGED I agree with you that, quite clearly, Fig. 9 shows that the genioglossus is not going to be operating as a whole. This particular one I split up into a continuum, going from anterior to posterior, showing how the anterior fibers, possibly, are behaving in a different way.

On the other hand, a number of the other muscles, straightforward things like the styloglossus, or even the stylohyoideus--you can see both its origin and its insertion very clearly on an x-ray--seem to function as unit muscles. I think it is only genioglossus that does function in the way you suggest.

CHASE Do you think there might be a role for different kinds of transducers in the description of the speech motor gesture? Would it be helpful to measure force directly rather than to infer it from the EMG data?

LADEFOGED Well, there are two points on that. Firstly, I agree, of course, that speech is a dynamic process and that, ultimately, I would like to study its dynamic aspects. For the moment, life is so complicated that I would like just to stick to the steady-state things and ask: How do you maintain a given position? Well, of course, you could talk about how you maintain the articulatory position for the production of a given speech sound in terms of forces.

The line of approach that we are currently working on is to try to build a replica of the human vocal tract in materials with comparable elastic properties. We are trying to work from there to see what forces we must exert on this model in order to push it into the shapes that we observe. I don't know how to answer your question; we have only gotten started on this kind of work.

STEVENS I am concerned a little bit about the non-uniqueness of this solution, and I guess you are very much concerned, also; that is, there are a number of ways in which you could have solved this problem. For example, if you have a beam that represents the tongue, and you contract the muscle on one edge of that beam, it will curl to one side. You could have said that curling was due to a contraction of the longitudinal muscle or a contraction of the muscle that actually displaces a local portion of the beam.

LADEFOGED I don't think there is as much non-uniqueness over the gross structure as you suggest. There are many muscles in a very intricate kind of relationship to one another, but some of them have quite straightforward actions, and they would seem to be operating in the way we suggest. I can't see many alternative ways of producing that same shape, given the fact that there are muscles which exert pulls in the directions shown in the previous diagram.

COOPER As I understand you, then, there are three effective constraints about which you have information: You know, roughly, what it is that contracts, as one member of a limited set; you know the resulting shape; and you can get measures of electrical activity that must agree with your assumptions about exactly what muscle was contracting.

GESCHWIND Why do you want to know the forces?

LADEFOGED In order to find out whether I can make an articulatory description in terms of what the forces are. Would this not be an articulatory description, at least in the steady-states?

GESCHWIND It seems to me that the problem of specifying the forces in a muscle like the tongue is a formidable one. You'll not be able to get it just by taking the EMG and summating it since the summation EMG will correlate poorly with the total force. Strain gauges probably can't be used since it is difficult to see how they would be placed.

LADEFOGED I warn you, it would probably be my tongue. (Laughter)

COOPER Might we switch the discussion a bit? We have talked about such things as Bernoulli forces at the glottis and subglottal effects on the spectrum--all very much 'grass roots' processes. Could we ask some of the people here who know something about what goes on 'inside the head' to comment on organization of speech gestures there. People who work in speech typically think about the buccal articulation or the glottal or subglottal functions as three separate classes. Presumably, this gross division has its counterpart in the nervous system; perhaps there are counterparts at less gross levels. How simple a picture could we have that would still account for <u>something</u>--not how finely can we look at details?

IRWIN Before we switch topics may I make one comment that will be relevant to what Ladefoged was saying. Kydd, a dentist at the University of Washington, has developed a process which he calls 'continuous palatography' (75). In this process he has mounted some ten electrodes in an artificial palate and has a neutral electrode mounted, at present I believe, on the arm. These electrodes do not measure force, that is, they are not strain gauges or strain-resistance points, but simply measure contact. Then, he has lights arranged in the same pattern as the electrodes are arranged on the palate, so that while the subject makes lingual contact--not in static position, necessarily, but during on-going speech--an array of lights that reflects his lingual contacts is activated. As I have read his reports, the method is still experimental but apparently is actually functioning.

STEVENS It's nice to have a method of measuring the movements in speech without having to irradiate the subject. A student of mine also has been developing instrumentation for doing this (116).

LADEFOGED With what success, Irwin, or haven't you seen it?

IRWIN I have seen it demonstrated; it worked precisely as described. Apparently, it is functioning now and they are using it for normal work with Dr. John Palmer, a member of the Speech Department.

CHASE Do any of the other reports that were mentioned discuss projects in which not only contact but also force is being recorded?

IRWIN Kydd has done other work in which force is measured (74, 76). Indeed, I suspect most of the work in force has been done over in orthodontia. The term tongue thrust has not been mentioned here, and one wouldn't expect it to be mentioned in a meeting of this type, but, to the orthodontist who sees his work collapsing around about him, tongue thrust becomes a very important question. The orthodontist has been using the strain gauge as a means of measuring lingual thrust in swallowing and in speech.

The gauges, as arranged, however, are not really set up to measure speech forces very well. They do better on swallowing, particularly on forward pressure of the tongue in the swallowing action. It certainly seems to be modified in speech.

CHASE It seems that if the contact electrodes on Kydd's instrumentation system are replaced by a matrix of load cells force could be measured.

HOUSE There is some current work on palatography using pressure-sensitive devices. I know, for example, of a graduate thesis (97) project under way at Purdue University under the direction of J. D. Noll.

STEVENS Our physiological friends have had a few minutes to think of answers to Cooper's question.

GESCHWIND Could you restate the question?

COOPER It wasn't a question, rather a vague desire for information. I feel that we could easily look in too much detail at what goes on in the speech process. How gross a look can we take at the neural machinery--the central nervous system and the nerves that carry commands out to the muscles--that would still correlate with the principal kinds of speech phenomena that we observe, namely, that there is comparatively slow activity in the chest area that provides pressure, there is a complex set of fast-acting muscles around the larynx to control vocalization, and there is another set (or several sets) of muscles that shape the tract and provide formant modulations. Is there something about the neuroanatomy that corresponds to these main functional divisions?

GESCHWIND I don't know that the neuroanatomy necessarily helps in this case.

COOPER From what you know of the neuroanatomy, how many independent components should we have to deal with in attempting a description of speech in gestural terms?

GESCHWIND I think that's very hard to answer.

(At this point in the discussion Milner gave a short account of the role played by various portions of the brain in human language behavior.)

GESCHWIND I agree with Milner's views on the role of the anterior speech area in sequencing.

HOUSE Both of you have said something about sequencing. Can you make it clearer what sequencing is under discussion.

GESCHWIND Of motor sequences, of well-learned sequences.

MILNER I was thinking of a habitual series. For example, all of these patients can count forwards very easily, and they can all count backwards very easily. They do this as soon as they see us appearing. This is really a rather insensitive test for stimulation of the posterior temporal area--it is not nearly as good as naming. The patients, after temporal lobectomy, do not have this kind of trouble. They have difficulty in naming and they have other quite interesting linguistic difficulties, but just these habitual sequences are quite well preserved.

COOPER How about the more closely articulated sequences, like consonant clusters? Do these give them trouble?

MILNER Not utimately, but in the early postoperative period, you may get a lot more.

HOUSE Can we assume then that this is not an articulatory level, as we have been discussing it, but a higher level than that--a higher level of linguistic organization, in some sense?

HIRSH Word to word.

MILNER Yes, I think so.

GESCHWIND Why wasn't the articulation as badly affected as you might expect?

MILNER It is, initially, of course.

HIRSH May I remember this discussion as having to do specifically with the speech musculature, that is, this isn't sequencing in other motor modalities?

GESCHWIND Was writing not involved?

MILNER His writing was all right. We haven't got systematic data on this. We're hoping to get them, because we have very few patients and we are getting intrigued by this.

GESCHWIND I would think that the longer sequences tend to show their impairment longest, simply because these are going to suffer worse on the basis of length and complexity than the short articulatory ones. Hence, recovery may first appear in articulation, and the long sequences will be impaired longer.

COOPER Is there, perhaps, another reason why they are more noticed? Is it because they may be of more clinical interest?

MILNER I think not. They are awfully handicapped, at first. No, I don't think so.

LADEFOGED Does everybody say <u>oot</u> for <u>too</u> or <u>out</u> for <u>one</u>?

MILNER Oh, no!

LENNEBERG It does happen occasionally. I have a tape recording of a patient, who is quite a different type of patient from Milner's. She was a stroke patient who had a propensity for Spoonerism, and the Spoonerisms occurred both on a phonemic level as well as on the larger segments. She would constantly anticipate sounds that were yet to come.

LADEFOGED That's different. A Spoonerism is different from saying oot for too. I mean, did you ever get reversals, actually wrong orders, as opposed to anticipation of sounds still to come?

LENNEBERG This is true, but she did have reversals as well as Spoonerisms.

FRY I have heard, actually, one example of a reversal which was very striking, in an English-speaking patient. In attempting to say the word brush, he got the vowel and the r reversed, although he hadn't got a postvocalic r in his English. To me, this is a very significant reversal.

GESCHWIND Was he aphasic?

FRY Yes.

HOUSE The intent of my earlier question about what was being sequenced was to determine whether or not we were getting some information about articulatory processes. Most of the descriptions we have heard so far have been contaminated by problems of aphasia and related disorders. I'm not quite sure whether we are talking about linguistic control signals that are disturbed, or whether or not the organism is capable of producing, if asked the proper question, something that we would call a phonetic act. Have we had an answer to this question yet?

GESCHWIND If you put bilateral lesions in the lower end of the motor cortex, you will get a permanent disturbance which is clearly articulatory and in which there is no aphasic component. This is a purely phonetic disturbance.

SESSION 4. Part 1 - Disorders of Speech Production

COOPER In our discussions yesterday and today, we have had good examples of the production of disordered speech. We now turn, in a formal sense, to the subject of disorders of speech production. Irwin has agreed to lead the discussion.

IRWIN What I would like to do this afternoon, to get started, is to present a classification of speech disorders that is now, I think, in fairly wide use in this country. I will go to the blackboard to do this, not because this classification is so good, but because it is so bad. I hope, seeing this on the board may stimulate you to some thinking along these lines.

This is not a textbook classification that I am going to present, but this is the kind of classification that I have derived on the basis of lots of site visits, lots of contacts with clinics, and looking at records. These seem to be the kind of diagnoses that are actually being made.

Basically, if you organize the field of speech disorders on an input-output basis, somewhere about halfway through--and nobody has really drawn this line--we have made a division. On the input side, is audiology, and on the output side, is speech pathology.

Recently, in our own professional literature, and particularly the literature that reflects the thinking of the American Speech and Hearing Association, a new word has appeared, and this new word, oddly enough at this conference, is language. Our Association is now beginning to recognize by specific statement that there are language disorders. The conventional classification of speech disorders that I am going to use will have a little bit of language in it. I will not be surprised if Hirsh also deals somewhat with language, but I do not have any commitment from him on this point.

I would like to suggest to you that, rather unwittingly and, perhaps, witlessly, the classification that has developed in speech pathology has really been based around two groupings. One of these groups has been, at least roughly, related to speech output--let me label one, speech output. The other major group really hasn't been labeled, but if one were forced to label it (and that is the position I find myself in today), I would call it over-all condition.

The disorders of speech production that fall under this heading of speech output can be listed as follows--and these are the actual labels that you will find on clinic cards. First, you will find articulation--not as a process in the sense that we have been dealing with it, but as a kind of disorder, referring to the speech of an individual who can't handle phonemes.

A second disorder of speech output that we recognize is concerned with voice. Here, voice, again, would mean a problem of voice in the carrier sense. This problem can exist along several dimensions, with pitch, loudness, quality, and absence of, being typical subclassifications of voice.

A third clinical classification concerned with output is stuttering, and this term does suggest an abnormality by its very name. This is the person with progressive differences, which I shall not try to define.

The fourth condition of output is delayed speech. As I am using this term it is a blanket classification that covers severe variation in articulation and, perhaps, in voice, but refers primarily to a reduction in output, whether measured by vocabulary, sentence linking, talking time, or other such measures.

Here, then, are four headings--articulation, voice, stuttering, and delayed speech--that all relate to what the speaker says. But, if you continue to look through the files of clinics, you will find another kind of heading. You will find the child who is labeled, cleft palate. If it is in the speech clinic, this usually refers to the speech of the person with palatal cleft, but it is interesting that the diagnostic label is not usually _speech_, but rather _cleft palate_, an over-all condition.

Another similar term is _cerebral palsy_. A third term that appears commonly is _aphasia_. A fourth term that does not appear too commonly but is certainly present is _laryngectomee_, referring to an individual who has had his larynx removed.

Now, the anomalous situation that I would like to call to your attention is that the cleft-palate child may also have articulatory or voice disorders, he may even stutter, and his speech may be delayed, but if he happens to have a cleft palate, he typically is classified by condition, rather than by speech behavior. This would also be true of cerebral palsy or aphasia or the laryngectomee.

FREMONT-SMITH Is there a classification by therapeutic needs, to some extent?

IRWIN I wish I could answer that. It is a classification that just seems to have grown. It does partly reflect therapeutic needs. I think it also partly reflects the different origins of this field, that is, we have some medical backgrounds--and I think that these appear here--we have some educational backgrounds, we have some psychologic backgrounds, as in stuttering, and so these groupings have just grown by convenience and without conscious planning by any particular group.

In addition to this concept of groupings, we also tend to make assertions about etiology, but these are, in my judgment at least, equally peculiar, and I would like to call your attention to the relationships.

For articulation and voice, we tend to say that they may be organic, psychologic, and/or imitative conditions. Most of the people in our field might use any of these words etiologically for either articulation or voice. But when we come to stuttering, we have a different conception. Here, the etiology is usually considered to be psychologic or organic, and I am using the _or_ to indicate that, by and large, people who would hold to the psychologic diagnosis would not make the organic one.

FREMONT-SMITH You mean psychosomatic has not come into use?

IRWIN I was really halfway including that under psychologic.

FREMONT-SMITH But the word soma is in psychosomatic, and this is the interrelationship between psyche and soma.

IRWIN That's right, and if that is included, it is included more in psychologic than in organic.

FREMONT-SMITH If you say and/or, you will be all right.

IRWIN I really prefer not to, since by and large no one person would use both. That is the impression I am trying to avoid.

FREMONT-SMITH Stanley Cobb would.

IRWIN Would he in the strictly neurologic sense?

FREMONT-SMITH Oh, yes, I think he would classify it as psychosomatic.

IRWIN But he would put it in the organic category, then.

FREMONT-SMITH You know, there is something here. Organic factors, no matter how severe, do not protect an individual from the physiologic effects of psychologic factors.

HOUSE The most interesting thing here, I believe, however, is that people in the speech and hearing field who say something is psychosomatic usually want to place it in the psychologic category rather than the organic category, because the site of the lesion in the former case is a little nebulous.

FREMONT-SMITH I think this is a misuse of the meaning of psychosomatic, which really means the interrelationship of psychologic features with physiologic or physio-pathological ones.

GESCHWIND But Irwin is trying to represent to us, and I think quite accurately, what people in the field actually do.

IRWIN I am just trying to convey an impression of how, as a field, this is likely to be done, and this is likely to be an either/or type of situation.

Now, delayed speech, means severely delayed in development; indeed, this may even go down to almost an absence of speech. It is a very broad term, a wastebasket term, I think, in a very literal sense. We have four classic etiologies, and here, instead of an either/or type of arrangement, these are likely to be cumulative, either in a particular child or in the use of a particular diagnostician. These, not necessarily in order of importance, are: mental retardation, hearing loss, emotional disturbance, and some type of central nervous system damage (sometimes abbreviated as brain damage). Therefore, a child with delayed speech, if diagnosed successfully, might be accurately pinned down to one of these or might be vaguely indicated as all of these.

By and large, with the four remaining conditions here, we tend not to make any etiologic specification, and I assume you can see why. In a very real sense, the etiology, at least the basic etiology, is suggested by the label, although, as I hope will come out in the discussion, this is really not a very accurate substitute for true etiology.

COOPER Is that true also for aphasia?

IRWIN As it is used, particularly with adults--and I'll get to that problem in a moment--a presumption of neurologic deficit is implied.

HIRSH As the term <u>aphasia</u> is used with children, however, all four of those can be used as etiologic statements.

IRWIN Which, of course, is why I avoided the term here, because it does have so many connotations.

A third type of specification that we tend to make concerns age, at least roughly. For voice and articulation, we are usually satisfied--and by we, I mean in our diagnostic cards--to specify child or adult. In other words, we recognize two kinds of problems or two kinds of therapies, as having a great deal in common. In stuttering, on the other hand, the presumption in our field is so strong that this is a disorder of early childhood that, usually, no age specification is made. If it does happen that a typical speech pathologist deals with a stutterer who began to stutter, for example at age 16, he is likely to record atypical as his designation of the age factor. May I repeat that I am not trying to defend these practices, but trying to call them to your attention; so a designation such as atypical in an age connotation would mean so late that it is quite conceivable that he is not the usual stutterer.

As far as delayed speech is concerned, ordinarily, no age differential is made, the presumption being that it is concerned with children. When we get to cleft palate, an assumption of existence at birth is made. If the condition appears later, the designation, acquired, is used. Examples are the child who had been shot through his palate or the adult who developed palatal cancer and who became cleft through surgical intervention. With cerebral palsy, usually, no age specification is made. Again, the presumption is made that it is associated with the perinatal period and an adult who becomes cerebral palsied, as demonstrated by behavior, frequently is not labeled cerebral palsy in our field. Aphasia, as Hirsh has suggested, was traditionally linked with the adult aphasic. The original meaning probably implied loss of, and so it was applied to the adult. But it has been used increasingly with children, with the distinction that aphasia, is sometimes applied to the adult, and congenital aphasia, is reserved, then, for the youngster with this problem. By and large, in laryngectomee, no age specification enters into the diagnosis, this typically being a condition of middle age or later.

My sole purpose in subjecting you to what comes dangerously close to being a lecture, is to get you to see how unfortunate this classification is. It is not consistent; sometimes it describes in terms of output, sometimes in terms of condition; it is not mutually exclusive. The cleft-palate child may have problems of articulation or voice which are not related meaningfully to etiology or to age.

The tragedy is that even though this scheme of classification has not been a conscious product, it has had a very powerful influence on the development of our field. At the risk of being a little too extreme in my statements, I think that I can almost assert that, over the country, speech pathology programs and their courses, their clinical organization, their research and their therapy, tend to be structured by this unfortunate classification. My hope was that this afternoon, you people might be able to suggest some more appropriate ways of looking at speech disorders.

FREMONT-SMITH Would you indicate where this classification fails to meet the needs of the field? You say it is more unfortunate; unfortunate with respect to what?

IRWIN I would be happy to indicate that. Perhaps, some of the other speech and hearing people present would like to do it. Would you like to, House?

HOUSE I agree that the curriculum in many training programs tends to be structured around this outline. You find courses labeled, "Speech Therapy for the Cerebral Palsied," for example. Because teachers are human and students are the same, after a period of time, in spite of all good intentions, such a course becomes instruction primarily in what the medical or paramedical conditions are, and usually, in the last week of the course, you suddenly realize that you haven't really discussed speech. This happens in other courses as well. Sometimes when talking to clinicians I have the uncomfortable feeling that they are overly concerned with the condition and don't know enough about speech processes.

FREMONT-SMITH They don't know enough in order to apply therapy wisely or don't know enough for what purpose?

HOUSE Probably a great many clinicians who come out of programs that have this structure are very well trained in clinical sensitivity. They know how to approach people, they know how to handle the emotional problems of cerebral-palsied children, etc.--but they get into trouble when they try to deal with the speech itself, particularly its articulatory aspects. They may, on the other hand, assume that they must work on some articulatory aspect of the speech, when they could more profitably be working on the emotional problems associated with the speech. Very often, it seems, people aren't quite sure which field they are in, and appear to know more about cerebral palsy than they do about speech processes. From my point of view, this seems unfortunate; clinicians should know something about both of these subjects.

FREMONT-SMITH They don't know enough to apply appropriate therapy to the situation in hand; is that it?

COOPER Or is it that they ought to be taught "cookbook" methods of handling these people, rather than what they are taught?

HOUSE No. I am suggesting that it should be possible to know enough about the process of producing speech and perceiving speech so that you can make some logical deductions as to what you ought to do in a given situation. I find that this is the kind of training that is lacking.

GESCHWIND I would feel that the didactic courses should concentrate on basic knowledge and that much of the material now taught in these courses should be covered in bedside teaching. The didactic courses ought to be directed to understanding mechanisms and principles and not to trying to teach practical things. I think this is a problem in many fields. An excessive stress on practical technique is usually bad. There is a problem which one sees in other fields as well, for example, psychology or medicine.

HIRSH I don't disagree, particularly with the statement that this field is not so different from psychology or medicine in the inconsistency among the criteria that are used, for example, in seeing and describing patients or in making diagnostic statements. I also don't think that it is particularly profitable for this group to examine courses and training of speech pathologists and audiologists, except to point out that, in general--and I think psychology and speech pathology are closer in this to each other than either one is to medicine--these training programs have been organized within colleges where there are departments of other things, whereas the medical professional curriculum has been organized within a single faculty, and you don't have to do this much in one school and that much in another school and satisfy requirements of a lot of other kinds of people. The curriculum is rather well laid out, and there are relatively few electives compared to those in more general educational institutions. So I am not sure the comparisons can go much further.

As I look at the training programs in this area--and I don't look at them as often as Irwin does--I find, first of all, that there are practical technique courses, where people do get taught how to do what one does and do get taught in rather practical situations, with patients or clients. The difficulty has been in the specification of courses like pharmacology. What are the preclinical science courses for this curriculum? Some of them have already been indicated, and some of them have come about, I think, through historical accidents. Every speech pathologist along the way takes a course in phonetics and in the anatomy of speech and voice, and these are, by and large, good things, I suspect.

My guess is that for these programs the closest analogy to pharmacology in medicine is something having to do with the psychology of learning, which is required in some training institutions but not all. Some of the higher-level phonetics that we were talking about this morning and yesterday afternoon--I know we didn't get very high--have not, by and large, found their way into the preparation of these kinds of clinical workers, and I suspect that, after they do, this kind of organization will change. I suspect that there will be some rather hard looks at some of these complicated models of the speech communication process, with an attempt to delineating diagnostic categories along the lines in that model.

You may want your discussion to proceed, for example, by trying to line up various disorders as they are perceived here, with breaks or lesions in the systems that we have described so far or talked about so far in the conference. As Irwin suggests, I think, we would begin to fail after about the second or third category.

CHASE It seemed to me that in the classification Irwin presented, as with most classifications, there is really a history lesson. Nobody could really defend the classification, because, in a sense, nobody was at enough points in time or enough points in space to be accountable for this kind of structure, which, in a sense, is a product of historical evolution.

The top group, or the top set of groupings, seems to me to be very close to the kind of complaint the patient brings to us; these terms are more economical, however. The second set involves a superimposition of some specific diagnostic categorizations, some of which are not cleancut but are used as heuristics in medicine. I don't think cerebral palsy has ever come into sharp focus etiologically, or even descriptively, as a diagnostic entity.

But there is one question, I think, that has to be kept in mind when one looks at a classification, particularly when invited to suggest a rearrangement, and that is: Why do we make classifications? Whom are they supposed to help, and how are they supposed to help them?

When we look at these questions, I think that the clinician has one kind of requirements, and the investigator, another. In terms of basic investigation we would like to think of categories that effect as close and accurate a linkage of a deficit in behavior with its neurologic substrate. In a sense, the classification that would be most satisfactory to us scientifically is that which will emerge from the kind of material we were discussing this morning and, to some extent, yesterday. It would be linked to an emerging specification of the speech process as such, and its underlying neural substrate. But would the same kind of classification be most useful to the clinician? I am not sure; I suspect, ultimately, yes. At this point in time, it seems to me that the most unfortunate consequence of this kind of classification is that it serves as a useful way for the clinician to categorize things which we can't categorize for him much better. So, in a sense, it helps him to deal with the practical problems of the waiting room full of patients.

But the most unfortunate consequence implied in your comments and those of House are that the same constraints which emerged in terms of meeting practical contingencies in a clinical environment, flow over into the academic field, and stand in the way of cultivating more precise understanding of the speech process and its abnormalities. We might, at this point in time, do well to think about these two problems separately.

IRWIN I would certainly agree, particularly with your latter statement, that, perhaps, the most serious consequence of this at the moment is not to the client or the subject receiving therapy, but to the control that it exercises on the thinking and the training of a profession. This is why I was wondering if, in terms of the kinds of things we have been talking about, there might be a pinning down, as Hirsh phrased it, of the breakdown points in these models, that might provide us with a more meaningful way of looking at disorders of speech.

FREMONT-SMITH In terms of studying them rather than in terms of therapy?

IRWIN At least for the moment, yes, although I would like to think that, ultimately, this would include therapy, and, perhaps, not too ultimately, either.

FREMONT-SMITH Of course. One way which we deal with this kind of problem in other conferences and which sometimes has worked very well is, after the discussion and in the light of the transcript, to get each one of the conferees who is willing to do so to put up an alternate classification, tentatively, and get this into the record.

HIRSH I don't think we're very far away from this, in general, in terms that we have used before at this conference. For example, voice disorders have to do with source characteristics, and this is rather a clinical entity, whether it is the laryngectomee or paralyzed single cord or badly used voice or whatever other descriptions are used. This is a group of disorders about source. We have been using <u>articulation</u>, I suspect, to comprise at least two kinds of disorders in this classification. What you mean by articulation in the clinic has to do with how precisely the articulators are placed, at particular regions in the vocal tract, whereas stuttering and some similar phenomena, have to do with the kinds of timing motions that one goes through in these articulatory placements. Whether we would want to combine those or keep them distinct, I'm not sure, but I would suggest that at least the logic of the two parallel systems is fine, as long as we are dealing with what people describe as speech as contrasted with language.

Now, when you come to describe a language disorder, then, it gets very difficult, partly because of the fact that when a person has a language disorder, the description of his trouble and, indeed, the therapy that he gets depend upon whether, in the formal course of referrals, he happens to get to a speech pathologist, a neurologist, or a psychologist.

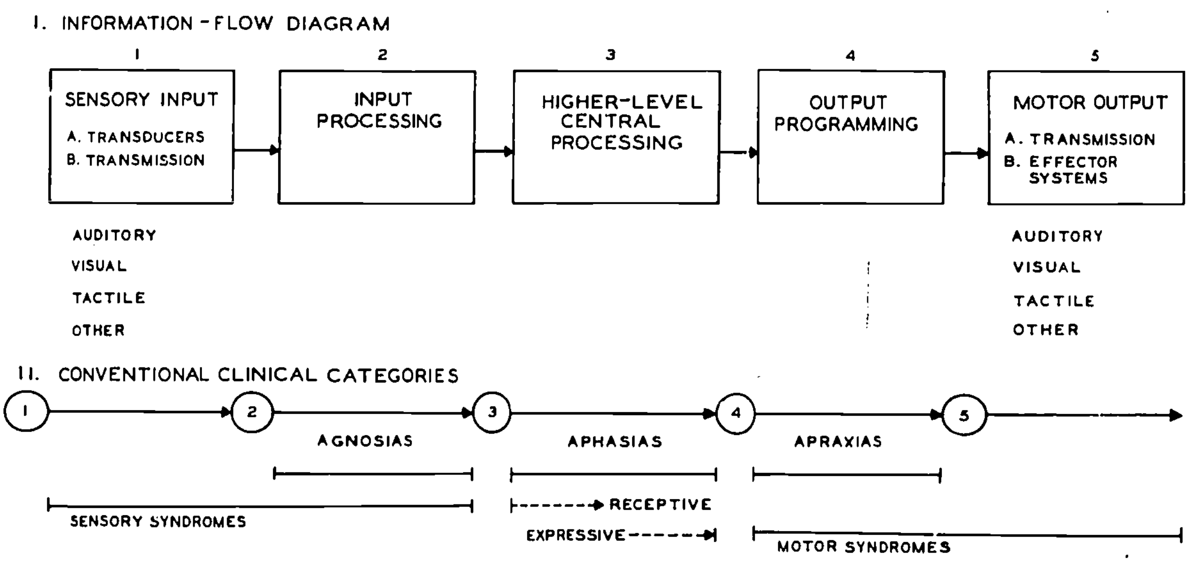

Figure 10. Classification of communication disorders.

RISBERG We have talked a lot about feedback here--both internal and external feedback--and to some extent, I think, some of these disorders can be attributed to either internal or external feedback not working. This might be one possibility for classifying the speech disorders.

IRWIN I might interject here that we have used feedback a great deal at this conference. One interesting way of looking at a therapist is that he is someone who injects himself into a feedback loop of the client. With certain types of disorders, this may be the most effective way of looking at the therapist--as a participant in an external feedback circuit--and, to the extent that he can supply the matching and correcting function indicated in Fig. 1, or help the client do this, he has been effective.

CHASE Figure 10 outlines a classification of communication disorders in information flow terms. I wonder with respect to Hirsh's point, whether, aside from the practical issues of the assessment of a patient with a speech disorder as opposed to one with a language disorder, a lot of the underlying issues in terms of what kind of classification might be best are not really quite comparable?

If looking at the input end of the system, we have a lesion involving the transduction of acoustic information into neural activity or the transduction of information about speech motor gestures into neural activity, then, we are not even getting into the system the information upon which it has to operate. I think, in this sense, the consequences are most severe at an early stage of development. A child with congenital deafness--whom I would classify as having a transducer deficit if he had a cochlear lesion, and as having a transmission deficit of the most peripheral sort if he had bilateral eighth-nerve lesions--is deprived of sensory input that undergoes processing in the learning of speech motor gestures.

These kinds of deficits, as we know, can occur at any stage of life, but the consequences are quite different for an acquired compared to a congenital deficit. I am positing that the information going into the system undergoes a hierarchy of processing before it can actually be translated into a correct pattern of response.

Let's move to the far end of the figure (i.e., Fig. 10). Any lesion of the motor systems can give rise to a speech disorder, just as it can give rise to abnormality of control for any kind of voluntary movement. We find dysarthria in association with cerebellar system lesions and basal ganglia lesions, as well as specific abnormalities of sequential programming of rate and order of motor units in the case of basal ganglia lesions. These deficits reflect themselves in speech as they do in the programming of walking, for example.

The componentry comprising the highest-level organization of sensory input appropriate to the structuring of an output in a given context of communication is shown in the center of the figure. This is the planning, if you will, which draws upon processed information from the periphery, which knows it has available to it a full set of capabilities for structuring response, and deals with the decision of: Now, what do I want to do? Here is where I think a lot of the so-called aphasic disorders would be best considered.

What I am suggesting, then, in most general terms, is that this broad information-flow model (shown in Fig. 10), in which we are concerned about the processing of sensory information, sequential hierarchical processing systems, and appropriate transmission of information, might serve as a reference for classification of speech and language disorders.

COOPER May I ask you something at this point? Doesn't this imply that you must know a good deal about what is going on inside in order to apply this kind of classification? And what, as a practical matter, do you use in the meantime?

HIRSH But those boxes (in Fig. 10) aren't pieces of anatomy, are they? (Laughter)

CHASE Hopefully, pieces of anatomy can be plugged in, but I think this is the point Cooper is challenging me on, because they are not yet pieces of anatomy.

HIRSH But they need not be.

CHASE No, they need not be. They are subsystems, if you will.

HOUSE There may be a real problem in making use of the information that is being presented. In my own educational experience, I seem to remember that most of the categories that Irwin listed were derived from a diagram similar to this one--a general communication model with the talker on the left and the listener on the right. The model itself really doesn't solve the problem.

GESCHWIND If you have basic information about what is known about phonetics from the point of view of more advanced research, it's a shame not to put it into the courses and to neglect this at the expense of kinds of techniques which people could pick up rather easily. I don't feel that classifications are too important. In the end our knowledge is always so incomplete that our classifications will be sloppy.

CHASE But what kind of classification gives us the maximal opportunity of learning new things?

GESCHWIND Classifications too often don't order knowledge but give a false impression that they have done so.

CHASE I still want to make the point that I'm not getting through to you on, that is, that there is a problem of classification with respect to the superimposition of some order on that which we know so that it becomes more useful in practical application, and there is the problem of classificatory schemes that are helpful in guiding our investigation. The scheme (Fig. 10) is not one that I am recommending for clinical use, and it is not one in which I would like to attempt a fitting of all kinds of clinical observations, since I am a proponent of sloppy organization as well. But I do propose it as possibly pertinent to the question of the kinds of loose fitting of terms needed to give the modicum of organization necessary to make an observation. I still think there is a confusion about whether you are using a classification to give answers or whether you are using a classification to ask questions. This is a classification that I'm using to ask questions.

COOPER Is this, perhaps, a classification that you want to use in applying the observations about disorders of production, to find out something about the process of speech in its more normal manifestations?

CHASE Right.

COOPER This was one of the things we had in mind in bringing disorders of production into the discussion at all--to ask what kinds of insights we could get about the speech process per se from a consideration of disorders, quite aside from the question of how to deal with those disorders as interesting and important human phenomena in their own right.

GESCHWIND Doesn't Hirsh's classification fit in with something physiological which does give rise to questions?

HIRSH Acoustic, as well.

GESCHWIND I think that most of the useful questions arise from concern with details.

HIRSH But this is one of the attractive things about Chase's diagram, that he does distinguish on the end between the output mechanism itself and those patterns that tend to organize commands to the output mechanism. Without such a structure, without such a plan, there are certain kinds of truths that will never be discussed. Risberg, for example, has a notion that many of these problems have to do with feedback. I don't think that the concept of feedback could have emerged from clinical observations. In fact, with the concept of feedback, you cannot make your diagnoses more precise, nor can you recommend a different kind of therapy. At the moment, it is a term that is useful only in another realm of discourse.

GESCHWIND But feedback did emerge from clinical observations; for example, the first explanation for paraphasic speech in patients with comprehension disturbances, in 1874, was a feedback explanation (113).

CHASE There are isolated examples, but, as a structural model which permitted armies of people to see new things, this didn't happen often in the nineteenth century.

LENNEBERG Maybe the remark I want to make is a little bit past the discussion, but I will make it just the same. I have rubbed shoulders with speech therapists for the last five years, and they use some classifications that are very much like this. I have been impressed, however, with something else. When you look at the clinical record, you see some diagnoses or some classifications made, sometimes much more fanciful than this, and when you look into the therapy that this has triggered off, I have always been under the impression that the therapy is pretty much the same no matter what the classification.

The system that is used most often is this, you start with muscle strengthening, and this was done for mentally retarded children, for delayed speech, for cleft palate. It consists of tongue exercise, to put the tongue here and put the tongue there, and blowing. I think nobody has shown that the muscles are really weak, and that this has any relevance to speech. The next step in this program is, by and large, exercise in pronouncing certain sounds. This has been done for all of these people that I have known. Perhaps mine is a very local experience, but one has the impression that the therapy that follows is not at all ruled by the classification.

KAVANAGH Apparently, Irwin's book (132) hasn't been read.

LENNEBERG Is there a wide choice of therapies?

KAVANAGH Yes, I believe there are several.

HIRSH I think Lenneberg's observations are correct, but, again, I would emphasize they are not unique to the field of speech pathology. I am thinking particularly of psychotherapy. (Laughter)

BROADBENT I think a lot of these comments are very similar to some I heard made by a chap the other day, who had been looking into the relationship between temporal lobe damage and certain diagnoses, either of neuroses or of schizophrenia, where, very frequently, the same patient is turning up at different times. If the doctor concerned had seen the EEG, the diagnosis was temporal lobe damage, and, if he hadn't seen the EEG, it was neurosis or something of that sort. It is really just a generalization, I'm afraid. (Laughter)

LIBERMAN I would like to ask whether we have any clinical data on the basis of which we might be able to tie together some of the things we talked about this morning and yesterday with some of these problems. More specifically, in the area that you have indicated, Irwin, as articulation. What more do we know about this? Are there any analytic studies? Do these things pattern in some way? If so, can we then make any connections between articulatory disorders and the kind of thing we talked about this morning? What are the data? What are the prospects for doing this?

IRWIN I think it is the kind of field which now has relatively little precise clinical data to supply. It may come closer, as Cooper suggested, in occasionally testing, as an exception, a general model and providing the data for a general model.

LIBERMAN I mean, are there any data which suggest, for example, some reasonable kind of way of classifying articulation disorders? Does it ever happen that a person has difficulty in making place distinctions but not manner distinctions, or manner distinctions and not place distinctions, or any kind of particular manner distinctions and so on? Do we know that this does happen? Do we have information that it does not happen? What is the situation?

IRWIN Well, one example would be an individual who is unable to raise the velum, perhaps, after polio; there you have almost a pure case of velar inactivity.

CHASE Before we leave this general topic, may I play a brief sample of speech from one of my patients, and see how people would classify it?

(Tape recording of a speech sample obtained from a 41 year-old male, admitted to the National Institute of Neurologic Diseases and Blindness for evaluation of motor abnormalities following carbon monoxide poisoning.)

CHASE I would like to add two comments. When you ask this patient to count from one to ten at a constant rate, he counts faster and faster until the whole temporal program disintegrates. If you ask the patient to tap on the table at a constant rate, he taps faster and faster until there is disintegration of the temporal program. One of his complaints is that he cannot control the rate of step placement in locomotion. He shows the festinating gait that we often see in patients with Parkinson's disease. Lenneberg has told me he has seen a similar problem.

LENNEBERG Actually, it is Geschwind's patient and we were studying him together. The speech is exactly the same.

CHASE I wanted to see what other people thought, but I don't think this patient can be fitted into Irwin's classification. But he can be fitted into a scheme in which there are some notions about how motor activity is organized, for this patient shows a problem of sequential release, both in terms of rate and order, which we see running through broad categories of voluntary motor activity, of which speech is one.

LIBERMAN Geschwind's point is still well taken, I think.

GESCHWIND Chase has described this patient as having "sequential programming difficulty." One difficulty with this sort of nomenclature is that this is a quite different kind of disturbance in sequencing from the one that Milner described this morning. It sounds to me rather like the type of speech disturbance that is seen with stimulation of appropriate areas in the thalamus.

LENNEBERG I have some relevant information on precisely the point that Geschwind has made. In 1962, Guiot, Hertzog, Rondot and Molina (48) reported that they electrically evoked this type of speech response in patients undergoing neurosurgery for correction of Parkinsonism. The surgeons were introducing an electrode into the thalamus; as they came to the latero-ventral nucleus they could often evoke a change in the rate of speech. In half of the patients the change was a slowing of counting, and in the other half, a speeding up. They published the rate of speeding up, which seems very much to confirm what we just heard, and conforms, certainly, to the patient that Geschwind and I saw together. I think, here in this case, even the pathology is quite consistent with this, and he may have some loss of large cells in the thalamus.

MILNER He brought the tape recording to Montreal, and as soon as I heard it, I recognized it. I haven't heard the slowing down but I heard the speeding up.

LENNEBERG The speeding up was reported to be a maximum of five to six digits a second, which is, I think, quite a significant figure. That was the maximum rate.

IRWIN Chase has been kind enough to offer to show a film--as near flesh-and-blood as we can come to in this kind of circumstance. Most of you people, I understand, have avoided any flesh-and-blood relationship to speech as far as possible, so we're going to force a little of it on you here today. This is an example with enough deprivation so that certain essential inputs would be of theoretical interest to some and of clinical interest to others.

CHASE This film (24) was prepared to demonstrate some of the findings of neurological examination of a patient with a congenital sensory syndrome, and abnormal speech development.

The patient is a 17-year-old, right-handed, white female schoolgirl who has been studied extensively at the National Institute of Dental Research because of difficulty swallowing, chewing and speaking from infancy or very early childhood. She was a full-term infant, delivered by forceps with some difficulty, following a prolonged labor. Pregnancy had been complicated by intermittent bleeding through the second and third trimesters. There were no evidences of neurological impairment in the neonatal period, however.

Difficulty in sucking and swallowing was noted during the first months of life, and there has been a lifelong history of difficulty in chewing and swallowing. Marked drooling was noted at one year of age, and has persisted to the present time.

The patient did not develop normal speech motor activity, and has been restricted primarily to the production of vowel sounds most of her life. With the aid of speech therapy during the past few years, minimally inteligible speech has been developed. There is a life-long history of minor traumatic injuries, primarily contusions and burns.

She walked at eleven months, and her general motor development was within normal limits. However some clumsiness of fine movements of the hands has been observed from early childhood. There has never been evidence of intellectual impairment, and despite her marked communication handicap, she has done reasonable work in a regular public school. The patient is one of three siblings. Neither of the other siblings, nor any other members of the patient's family have similar symptoms.

This girl is presented as an example of a congenital sensory syndrome with specific motor control abnormalities involving lips, tongue, pharynx, and hands. We feel that the motor control abnormalities are the result of inadequate sensory feedback information needed for the normal developmental organization and later control of movement. (The motion picture film was shown at this point.)

Kavanagh knows this patient and Liberman has seen here as well. Kavanagh asked the patient to compose an essay about the school she went to, and he was good enough to give me a copy of what she produced. The essay consists of about 75 words arranged in about eight sentences--simple and compound--and sentence fragments. The syntax could be described as poor, and quite a few words were misspelled. (A typed version of the essay was projected for viewing by the discussants.)

HIRSH That seems inconsistent with your previous report that she was getting along reasonably well in school.

CHASE Right, but we accepted that inconsistency because she was getting along in a regular public school. She performed on the Wechsler-Bellevue at a dull-normal level and, in terms of her social adjustment, has done quite well. This is an example of her written language. This is one of the problems that Kavanagh raised with us and which he is probably interested in having comments on here: How would you explain this kind of performance?

GESCHWIND I wouldn't be able to explain any of this girl's performance. I find this very mystifying. First of all, your sensory examination seemed to me to indicate that she had practically pure pain loss. You pointed out that position and vibration were intact all over the body and light touch was intact everywhere. Am I correct in assuming that this girl could not tell sharp objects from dull ones, or was it that she simply did not find sharp things painful?

CHASE She could not distinguish between sharp and dull.

GESCHWIND What you have here is somebody who seems to have a fairly isolated pain loss. The fibers which are involved in proprioceptive control are those carried in the posterior columns and are much more closely involved with position and vibration. The patient who has trouble carrying out movements with his eyes closed need have no pain loss. Patients with syringomyelia, with marked isolated loss of pain sensation carry out movements perfectly with their eyes closed. I find her speech difficult to understand in terms of the type of sensory loss that she demonstrates.

KAVANAGH Chase wanted to say that we too are mystified by the whole picture.

GESCHWIND I think that this girl has more than one lesion, and I don't think her sensory loss explains the motor disturbances in her delayed speech. Lenneberg has a patient with delayed speech who is similar and he has no sensory loss.

LENNEBERG Some of you have seen the film of my patient. This child has no speech but perfect understanding of spoken language much the same as the girl just discussed. He is slightly retarded.

COOPER What about manual skill in writing?

LENNEBERG It is good.

COOPER That is one of the things that is striking about Chase's patient; when she wrote a composition on the blackboard, she didn't do as well as a normal 16 year-old, but, even so, she did very well.

HIRSH The film showed that some manual responses were improved--mainly coordination--with improvement under visual control. Does her speech also improve when you allow her to speak in front of a mirror?

KAVANAGH We observed some improvement when using a mirror.

HIRSH As much improvement as this (demonstrating)?

KAVANAGH No. But there was slight improvement in her articulation following direct auditory stimulation <u>and</u> when with a mirror she was encouraged to compare her facial and lingual movements with those of the examiner. There did not, however, seem to be any carry over.

DENES Her speech disorders were not of the kind which would have improved by vision, were they? Most of her deficiencies were vowels and the lingual sounds, consonants like <u>r</u> and <u>l</u>, and the only relatively visible sound that she had difficulty with was <u>sh</u>?

HIRSH Oh, no, there were many more than that. (General agreement was expressed.)

DENES But <u>kr</u>, <u>pr</u>, and other clusters were pretty good.

LIBERMAN Chase said a moment ago that I had seen this girl and tested her; it was my first contact with anything clinical. I am not clinical; in fact, some of my friends accuse me of being anticlinical, but it is true that several of us from Haskins did work a bit with this girl. We did two things. First of all I would like to say that, like Cooper, I am interested in the fact that she can write reasonably well but can't speak nearly so well, even though whatever it is that's wrong with her seems to be quite general. Does this suggest certain interesting differences between the way we produce language when we write it and when we speak it? One very obvious difference is that the rate is no problem at all when you write; you can go at any rate you want to. Your rate may be critical in the speech case.

We tried her on a stop-consonant discrimination task, and it turns out that she is quite normal--her responses peak at phoneme boundaries. The level of discrimination is very nearly normal, as nearly as we can judge by comparison with other subjects.

Just to save the theory, we asked her to produce quite a variety of monosyllabic words, the idea being that we wanted to present these words to trained phoneticians for phonetic transcription. We have done this, or, rather, Thomas Rootes at Haskins Laboratories has done this. We don't have a definite answer yet, but I am, nevertheless, intrigued by the fact that a pattern may emerge here, and that this could be done on a larger scale. For example, it does seem to be the case with this young lady that she cannot handle voicing distinctions for stops. She does much better on place distinctions, as judged by our three trained phonetician listeners. This is not to say that she does a perfect job of placement, but she does reliably make a different noise when she is instructed to make a b from what she does when she is instructed to make a g.

FRY I think we ought to underline explicity that her reception of speech is quite good.

LIBERMAN Oh, yes, quite normal.

FRY But her production of speech is not. I think it's worth saying that.

LIBERMAN Absolutely.

COOPER But, Fry, to say her speech is very good in the sense that she responded freely and easily in an interview situation is one thing, but to say it is normal would imply that she could handle all the distinctions that the rest of us handle, that she could handle them against noise interference, and so forth. I don't believe these points have been tested.

FRY The fact is that not only could she receive in the interview situation, but I was passing remarks to other people, not to her, when she was at the front of the small lecture room, and she got them and turned and smiled. I call this pretty good. I don't think there's much wrong with her reception.

LENNEBERG Would you agree that the child I showed you in the film had good speech reception? (The film referred to here is not the film screened for the present session.)

LIBERMAN Well, it's hard to tell from the film. There are gestures and all kinds of things involved in reception. I'm not saying he didn't; I just can't be sure.

LENNEBERG Let me say what is in the film, for the benefit of the others. This child was told a story and he could nod yes or no. The story lasted three or four minutes and, after the story, he was asked questions such as: Was the milk drunk by the nice lady? Did the cat run out of the house? There were something like twelve questions of this kind. To all of these, with one exception, I think, he nodded correctly, and, if I am a fair judge, it is quite reasonable to assume that he could understand the questions.

LIBERMAN There are two problems here, though, Lenneberg. One, is it true that this child really does not talk? I understood you to say that he doesn't talk much...

LENNEBERG No, you heard sounds there. He is much worse than this girl. This is not my interpretation, because he has been around.

LIBERMAN The other thing is--you know about the horse called Clever Hans (111), of course? I mean, is this controlled here at all?

LENNEBERG We were showing the film to show that the child did not have to watch tiny movements of the examiner. He doesn't even look at the examiner. In addition to that, Clever Hans had a perceptual acuity that, normally, people do not have. I think we are now imputing an acuity to this child that is just unheard of.

LIBERMAN Well, I do want to come back to the point that Cooper made before, which is that the motor theory doesn't say that such a person could not hear speech or could not discriminate speech. It does say that he would do it differently from a normal person, and, I think, not as efficiently. Certainly, if it were true of this patient that she could not reliably make different gestures for, let's say, <u>ba</u>, <u>da</u>, <u>ga</u>, then, the theory would have to predict she would not show peaks in a discrimination function. It so happens she shows peaks, but it also happens that she reliably makes different gestures for <u>ba</u>, <u>da</u>, <u>ga</u>.

I would like to know more about your patient. Perhaps, we ought to try to apply more sensitive measures to how quickly he can respond, or how efficiently, in comparison with normal people.

LENNEBERG This is somewhat of a crucial point for the argument we had yesterday. I think it is quite common to find children who will say nothing, or very little, or make very odd noises indeed, but who have no obvious impairment of reception. It is an entity that is well known.

COOPER And <u>no</u> losses of reception?

LENNEBERG As far as one can tell. One can ask them to do very complex things, such as, "Touch your left ear with your right big toe," or something like that, and they will promptly go to it, or try, at least, to do this sort of thing. (Laughter)

GESCHWIND Your case, Lenneberg, had no sensory loss elsewhere, on neurologic examination?

LENNEBERG No sensory loss of any kind has ever been demonstrated.

HIRSH A point of information, Mr. Chairman. For which argument yesterday was this observation crucial?

LENNEBERG I'm coming to that. When I heard about the motor theory of speech perception, I had my doubts, because here is a film that shows a patient who does not have the motor part, but perceives speech. We played the film. Now, we have seen another one, and yesterday we were busy throwing out the entire argument, I think, with good reasons. But we still have lots of emotional feelings about it, both Liberman and I. (Laughter)

KAVANAGH May I add one or two more comments about this patient. First, she was unable to recognize any of a group of plastic objects--such as a cross, square, and circle--when they were placed in her mouth. She could identify them visually, however.

FREMONT-SMITH She fails tactilely?

KAVANAGH Yes--in the mouth.

FREMONT-SMITH Can she identify such objects in her hand?

KAVANAGH She has trouble recognizing objects in her hand, too, but she is particularly handicapped in the mouth. Secondly, she is not as retarded as one might conclude. Her mean score on the Wechsler-Bellevue is 91.

MILNER How about her verbal score?

KAVANAGH Her verbal score is lower. There was a 16-point differential between the performance and verbal scores.

(At this point the discussion was terminated.)

SESSION 4. Part 2 - Disorders of Speech Perception

COOPER Our next topic is disorders of speech perception, and Hirsh will lead the discussion.

HIRSH It is a little illogical for us to take up disorders of speech perception now, I think, because we haven't really had a good go at speech perception yet. Feedback got into discussion some time yesterday, and we didn't quite get back to some of the basics. Let me just open the discussion by using Chase's model (see Fig. 10, p. 186), if I may.

I would point out to you that, in general, in clinical measurement, one is concerned with three aspects of the hearing process, in estimating the degree of disorder. One has to do with sensitivity, which is the most common and the most generally used--how many more decibels of energy does a patient need compared to a normal in order to detect something. This, I assume, has something to do with the many auditory functions that are included here.

Input processing, in general, is not measured clinically. Instead, we come to a third type of clinical measurement. We have had one that is done usually, one that is not done at all, and now we come to a third type of clinical measurement which involves what, I think, is some higher-level processing, by throwing words at listeners and asking them to repeat the words that they hear.

This third type of measurement, or what is called discrimination testing or speech discrimination testing or speech testing, is both too general and too specific. It does not, in general, include those central processes that would be described as having to do with receiving language intelligibly. Ordinarily it does not test for the hearing of sequences of words. On the other hand, neither does it test selectively those aspects of input processing that we think may be relevant to the various acoustic cues in speech perception.

In describing hearing disorders, one resorts to degrees of hearing loss, and here, I think, the most common reference to degrees of hearing loss is made with the decibel (dB) scale; in other words, on the basis of sensitivity alone. In this country and also in Western European countries generally, there appear to be four degrees of hearing loss that are widely recognized. From 0 to 30 dB loss, which is not considered to be a handicapping loss; from 30 to 50 dB, roughly, a mild loss; from about 60 to 80 dB, a severe hearing loss; and more than 90 dB is deafness. You may ask, what are these decibels of? In general, they are hearing losses averaged for frequencies like 500, 1000 and sometimes 2000 cps, but not always. You will notice that I left 10-dB gaps between the last three regions; I did that on purpose. There is some fuzz in the system and nobody tries to make the lines very sharp.

In the last two decades there has been superimposed on this characterization of hearing disorders something about the hearing of speech. With reference to the kind of widespread clinical practice of which Irwin spoke in connection with speech, this is most often the percentage of monosyllabic words that can be repeated correctly.

These two quantities, decibels of sensitivity loss and percentage of monosyllabic words that can be repeated correctly, are almost, but not quite, independent of each other. That is to say, you can have a fairly severe hearing loss of the order of 60 dB with no so-called discrimination loss. That is, if you amplify the speech by 60 dB so that it is relatively as strong as for a normal listener, the patient may repeat correctly 100 per cent of the words.

The converse is also true. There are some patients, but they are rare, who have a hearing loss of 0 dB, that is, have no shift in sensitivity, but who cannot repeat monosyllabic words correctly. These are, from our point of view, I suspect, even more interesting than the others; one wonders what is wrong.

The ways in which what is wrong has been studied are two. One is an attempt, or several attempts, at what we might call clinical psychoacoustic measurement, where one is interested in the way in which this auditory system, normal or not with respect to sensitivity, processes signals. One studies things like the loudness function, discrimination (usually of tones or noises) with respect to intensity, frequency, bandwidth. There are also some other psychoacoustic tasks, about which we know in the normal, which have not yet been applied to patients.

The second way of knowing about what is wrong in the case of discrimination loss is to look at it as a kind of dysphasia, and this is even less typical of present clinical studies, to my knowledge, and brings me to a somewhat more general mention of language disorders from the receptive point of view. I cannot, should not, and will not talk about adult aphasia or dysphasia. I would like to talk a little bit about two aspects of language disorder in children. One is what we have called congenital aphasia or dysphasia, or the extreme degree of what Irwin has called delayed speech, where there is no speech at all or no language at all for a couple of years.

Yesterday afternoon, when I was describing some of these youngsters, I may have given the impression, by reference to these transmodal associations that were difficult, that the primary difficulty with these youngsters was an associative one. I am not sure that that is true. It may be true of some of these youngsters, but there is at least one subgroup--and I don't want to make it any more general than that--for which the difficulty appears to be almost entirely the perception of sequential information, information that is presented sequentially in time.

I would like to mention one study in which we took a group of these deaf children and a group of normal children and measured the visual memory span (137). We were interested in measuring the number of objects that the child could immediately reproduce by selecting from a much larger bin of cards on which the same images were represented. We flashed a row of images on the screen, and he had to reach in and, from more bins than those that were shown, select those that had been presented, and also reconstruct them in order. This was for several different kinds of vocabulary; in one case silhouettes of real objects, in another case geometrical shapes, and, in a third case, what are called nonsense shapes.

The interesting thing is what we were looking for and found was that the way in which you presented the material would be important in distinguishing the groups of children. The aphasic children, the deaf children, and the normal children were not different with respect to their ability to reconstruct these sequences when the sequences were presented simultaneously as a row of images. But, in addition to this, we presented them, sequentially, one image at a time, as if they were sounds, if you like, to test the proposition that it was sequential processing that was wrong, whether or not the information was delivered to the auditory channel. This is a deficiency, we hypothesized, that was temporal in character but not necessarily auditory in character.

We had expected that the deaf and the aphasic children would both be deficient. What we found was that the deaf and the normal children were alike, and the aphasic children were severely retarded in their ability to reconstruct the sequences. This is at least a second dimension of this receptive disorder, at least as it is found in children.

The second point about language disorder in children has to do with the relation between uncomplicated hearing loss, and the development of language. We are all accustomed to the notion that the hearing of language is a requisite for the development of language, and the clearest evidence for this is the speech of the deaf child, which is poor at best and absent in cases where he is not taught by one or another quite specific technique.

What is becoming more clear, and has not been suspected, at least quite evidently, is that relatively minor hearing losses can interfere with the development of language. By minor, I mean an audiogram, not unusual in children, that shows relatively good hearing in the low frequencies--10 or 20 dB loss--and rather bad hearing for the high frequencies. The case that I alluded to yesterday, for example, would not be detected on superficial examination where you call the child's name and he turns around, because he responds to low-frequency energy. These youngsters do develop language--but their language retardation is not just misarticulation, which normally goes with some of these high-frequency hearing losses, but a genuine retardation. They may be set back by some years in the development of the higher-order language constructs--grammar, vocabulary, and the volunteering of spontaneous speech.

I trust that I have mentioned a large enough number of areas so that you will want to discuss disorders in speech perception. Where do you want to begin? Risberg, do you want to tell us what you do about reconstructing speech when you don't have any hearing?

RISBERG We have started a program to study if it is possible to send information about speech through other senses than the auditory, or in case of high-tone hearing loss, if you can take information from the high frequencies and transmit this also in the low-frequency part of the spectrum. The experiments have just been started, and we have no results from any real tests yet. We are now testing a tactual device, where we transmit speech through ten vibrators, one on each fingertip. We make an analysis of the frequency spectrum of the sounds by means of a filter-bank, so you get the lowest frequencies on the smallest finger of the right hand and the highest frequencies on the smallest finger of the left hand, giving a continuous frequency scale over the 10 fingers.

This has been tested before in this country (42, 134) and also in Sweden (112) and the results show that you can use this in connection with lip-reading. Some sounds are difficult to distinguish with lip-reading--for example, _m_, _p_, and _b_--but then you can get added information through this tactual device. Most of the tests have not been made on real speech material, that is, single words and not sentences have been used and it remains to be proven that you can use this information even in connected speech.

Of course when you are trying to use this yourself, for the first time, you are very bewildered. You don't feel anything. You can't distinguish between vowels and fricatives very easily. But it seems that you can learn it reasonably quickly, and we have now a deaf-and-blind subject on whom we will try this, we will also study if she can improve her speech with this device.

We have made a test with a small device and she discovered the difference between long and short vowels very easily. She was not aware of this difference before. This subject has language and can read Braille.

COOPER Excuse me, but are you using this both to provide information about normal speech and to provide feedback about her own speech?

RISBERG Yes, this is attempted, in the experiment with the blind-and-deaf. It is intended for use both in perception of the speech of others and their own feedback speech. The devices can be built small, so there is no objection to them in this respect.

COOPER What kind of physiologic limitations would you expect to find on these?

RISBERG There are many physiologic limitations on the signal types you can transmit through the tactual sense. You can't for example use vibration frequency as one of the dimensions, as the skin is very insensitive to frequency variations. There are also many other limitations of this type. The psychological limitations we know very little about. This has to be found out in tests of different types.

STEVENS In having the person learn to use this device, does he learn faster if he can talk into himself, or if he listens to other people? Do they learn faster if they can also use it and sort of build up a set of patterns for themselves?

RISBERG It is quite probable that the best method to learn the tactual patterns is to try to say the sounds or words yourself. For a deaf person this of course, means that you also have to teach him the correct sounds.

HIRSH What evidence do you have that the upper eight channels are being used at all?

RISBERG Of course, the question is, what can you transmit through this tactual device? In many schools for the deaf they have used a simple vibrator to feel the speech rhythm. Probably most of the information we get through our tactual device is the rhythm or syllable structure, but I think that even if it is only the syllable structure we transmit, we can find a better method to transmit this structure than with a single vibrator. We don't know if the subject can really use the information in all ten channels, however, that has to be tested.

KAVANAGH Are these oscillators something like a bone-conduction oscillator?

RISBERG Yes, bone-conduction transducers. The vibration frequency is constant--about 200 to 300 cps.

KAVANAGH I didn't hear what the frequency range was for each finger.

RISBERG It is the same frequency for all fingers, 200 to 300 cps in this case, and only the amplitude of the vibrations varies.

BROADBENT For any one finger, what frequency range of sound is delivered to that finger?

RISBERG That depends upon the particular finger. The lower frequency channels are about 200 cycles wide, and the highest channel is about 2000 cycles wide.

DENES Perhaps this is a good time to say something about the vocoder device that I have been trying out. We are in the middle of the experiment, so there are not many results to report at this time. The purpose of the experiment is to obtain some additional information about the much-discussed motor theory of speech perception. At the same time, the experiment may well have some relevance to the problem of re-coding the acoustic information in the speech wave for use by the severely deaf.

As far as the motor theory is concerned, I wanted to see if experience in producing speech helped at all in learning to recognize it. In the experiment, naturally produced speech was processed--re-coded, if you like--into another sound wave which preserved much of the articulatory information of the original wave yet sounded sufficiently unlike normal speech that relearning was required to make it intelligible. Two groups of subjects were then made to learn this re-coded speech; in one group each person heard his own re-coded speech, whilst those in the second group only heard processed speech produced by others.

The re-coding consisted of a form of spectral compression. If these compressed speech signals were learnable, then the same device could also be used as a hearing aid for those whose residual hearing covers a spectral range smaller than that of the normal speech spectrum. As you know, the speech spectrum extends from about 100 cps to about 6000 cps,

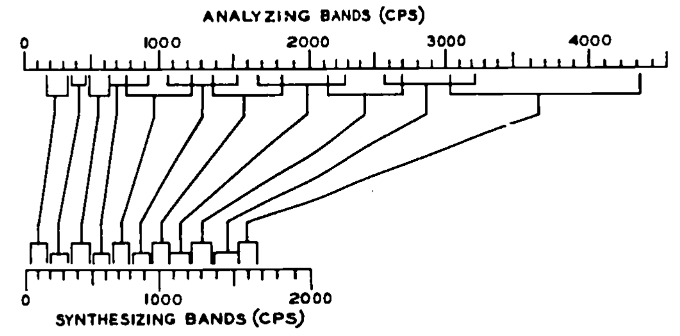

Figure 11. Schema illustrating vocoder method of transposing the frequency spectrum of speech. The energy levels in various bands of the original speech are measured by the analyzing filter set, as shown above; these measures provide control signals for generating the synthesized "speech" signal in different frequency bands, as shown below.

although a number of deaf people are sensitive only to the lowest 1000 cps of this range. The apparatus used in the experiment presented some of the information of the wider range in a narrower 1500-cps wide, band. The general arrangement is shown in Fig. 11. The speech frequency spectrum was divided into 11 bands. The energy in each of these bands was measured and used to control the output of another set of 11 filters, shown in the lower part of the figure. Each of these synthesizing filters covers a narrower frequency range than its corresponding analyzing filter and the combined bandwidth of the 11 synthesizing filters is about 1500 cps. The vocal frequency of the speech input was also determined and was used to control the frequency of a buzz generator that served as the input to the 11 synthesizing filters. The frequency of this buzz signal was always one-third of the measured vocal frequency of the speech input. The excitation of the synthesizing filters changed to hiss whenever the speech input was aperiodic. As Fig. 11 shows, the device compressed the original speech spectrum in the ratio of three to one. Its effect was to maintain the shape of the original spectrum, but displaced and compressed towards the low frequencies.

In conducting the experiment on the motor theory, I first listened to the output of a conventional 11-channel vocoder which used the same analyzing filters as the device I have just described. The output was reasonably intelligible and served to show that the speech spectrum, as specified by the 11 filter outputs, had enough information to produce intelligible speech. I then switched to the compression device and found that its output was not immediately recognizable, showing that the same information which previously was recognizable became difficult to understand once it was re-coded. The output of the compression device, therefore, was well suited as the basis of the learning experiment I outlined a moment ago.

In the experiment, two groups of subjects listened to the output of the device again and again, trying to learn to understand the speech. One group was allowed to hear only the pre-recorded speech of other people. These materials were played into the device from a tape recorder and the subjects heard the re-coded output via earphones. The other group of subjects not only heard this set of utterances but spoke the same words through a microphone input to the compressor. The members of the second group, therefore, heard their own individual utterances processed by the device while they were producing them.

Every session consisted of a 20-minute learning period followed by a test period. During the learning period the subject had a printed list of the words he heard over the earphones, and during the test period he had to write down the words he recognized. Responses during the test period were scored and served as an indication of learning. Comparison of the learning of the two groups indicates how much hearing your own voice while you produce speech helps in learning to recognize speech.

HIRSH Excuse me, but how do they avoid hearing their own high frequencies?

DENES We hope that the sound delivered by the earphones is loud enough that it will mask the higher frequencies.

KAVANAGH That would have to be mightly loud.

DENES It is pretty loud. It is arranged so that it is something like 10 dB below the threshold of feeling.

POLLACK But how do you get around the objection that was raised, that when somebody else is speaking and when, perhaps, there is a visual word to tell what is said, the person isn't setting up the motor responses in his own throat?

DENES Because he doesn't hear it through here. The whole point is that motor action is effective only if it is fed back through the auditory apparatus.

LIBERMAN What happens if you present--in the case of the people who are permitted to speak through this system--one of these words which they can't identify, and you then let them proceed by trial and error to try to find out what it is; can they do it? How long does it take? Have you got to limit the number of times?

DENES We don't let them do that. We wanted to control the amount of exposure each subject has to the sounds through this system, so all they are allowed to do is to listen and then say the same word they have heard through the device. They are not allowed to "babble" through the device.

LENNEBERG When they first try it, does it change their voices? Does ones voice begin to sound funny when you listen to it?

DENES No.

LENNEBERG They are not affected at all, even when it is altered feedback?

DENES Not as far as one can hear.

HOUSE Do you have a group of large people--that is, people with long vocal tracts--and a group of small people?

DENES They are all high school students.

LIBERMAN There would also be a difference in the experiment, to apply this to some of Hirsh's children who have high-frequency deafness.

HIRSH Is he working with a 1500-cps low-pass signal? These kids and normal listeners do well with amplified speech through a 1500-cps pass band.

DENES Yes, but, really, the question here is--and I can't give the answer to this because I haven't analyzed the data yet--whether they learn to perceive, to recognize the sounds whose energy is in essentially the higher-frequency region? The results at the moment are fairly unsatisfactory, because the learning rates--as measured in terms purely of over-all recognition scores and not in terms of what types of sounds are lost or not lost--are running more or less parallel for those who listen only and for those who listen and speak.

POLLACK They both learn?

DENES Oh, yes, very much so. We had to start a second series because although the words heard in the first series were initially only 50 per cent recognizable after four sessions, they were soon up at the 80 to 90 per cent recognition rate. So I increased the vocabulary from which the words were randomly selected to 150. The recognition rates then started at around 20 per cent and are up around 50 and 60 per cent now.

KAVANAGH Are these results for monosyllabic words?

DENES Yes, they are.

GOLDSTEIN But you tried this only with the slowed-down pitch rate, or have you tried it with the normal pitch rate?

DENES No, with the slowed-down pitch rate.

GOLDSTEIN Why did you do that?

DENES Because we wanted to get enough harmonics into each of these filters to define the spectral envelope fairly well.

STEVENS This would mean some short vowels of about two pitch periods?

DENES I suppose so. What is the point?

STEVENS Well, that isn't very many pitch periods to have in a vowel.

BROADBENT I don't quite understand what will happen to the signal here. Presumably, if you take, say, the top filter in the input end of the device, its output varies in time. You then use that signal to control it again on the top filter in the other device, and then you pass through that a set of harmonics which are spaced at a different interval. What is the output from that filter going to look like?

DENES Basically, the highest-frequency filters will be excited by hissing sounds only, and, under those conditions, on the synthesizing side, also, it is a hiss generator which is exciting the filters.

BROADBENT Well, take any of the others, then. My problem is still that I can't envision what the envelope of the output of your second bank of filters is going to look like. Isn't it a mixture of the original envelope plus the envelope due to the new exciting function?

DENES You mean intensity against frequency?

BROADBENT No, I'm sorry; I mean intensity against time.

DENES It would be exactly the same as in the original.

BROADBENT Will it be?

DENES Well, apart from the time constants of the filters themselves, the variations in time of the energy output will be the same as for the analyzing filters. That's just the point. Apart from the time constants of the filter, you can think of the spectrum, including the harmonic spacing, simply being compressed, but unchanged in the time dimension.

GOLDSTEIN When you mentioned work with a similar device, Risberg, did you say it didn't work very well?

RISBERG Yes. It has been tried at the Bell Telephone Laboratories before, by Guttman (49), I think.

DENES Yes, and even before that. What they used was a different device, called the Vobanc, which does a similar thing; the compression is based on three filter channels, dividing the whole of the speech spectrum into three channels rather than eleven.

RISBERG And there was work in Germany by Oeken (107).

DENES Also by Koenig (71, 72) and by Bertil Johansson (64).

RISBERG Yes, but that is another type of device. But Oeken, using frequency division by means of a tape recorder with rotating heads, obtained signals resembling yours. He trained hard-of-hearing subjects with low-frequency hearing with this device. He got improvement in the results, but when he gave the same amount of training using ordinary amplification the improvement was much greater--about two or three times better--so he concluded that this technique is of no use to improve speech perception.

DENES As far as I'm concerned, the main interest is a theoretical one concerning the motor theory. From the point of view of hearing aids, the theoretical interest is whether deaf persons can be trained to read or assimilate information presented in this way.

LIBERMAN But in terms of either the theoretical point of view or the practical point of view, it seems to me that it might be very appropriate and relevant to ask what happens here if you start in with much younger children and give them much more practice. What you know here, in any event, is that the group that is permitted to speak into this system surely will do better than the other group, if both groups have got to find out what is being said. These people at least have a chance to find out what it is that is being said.

GESCHWIND What age, Risberg, were the hard-of-hearing subjects on whom this device was tried?

RISBERG In the German material, I think, they were about 30 to 50; they were adults.

GESCHWIND Weren't they too old for this type of relearning?

RISBERG Maybe, too old.

FRY May I ask a question which I think is pertinent? Is anybody trying to train deaf children with hearing aids?

LIBERMAN You almost spoiled the conference with that question! (Laughter)

COOPER What is the point?

FRY Hirsh is going to tell us what the point is.

HIRSH I think the timing of the conference relative to this development is a bit unfortunate and premature, in the sense that the vocoder kind of output has just now been engineered in these various ways. Let me first round out the picture and point out for your memories that Denes' device, when you let the outputs of the filters be illuminated spots on the oscilloscope was tried as "visible speech." Another system, also from Stockholm--but on the other side of town from Risberg's laboratory--divides the frequency band into two, at about 1000 cps or something like that. The frequencies above 1000 cps are heterodyned, so as to be brought back into the system in the low-frequency region--now, the spectrogram has been cut in two, and the top part simply flipped down so that the energy, as energy, is now down in the same low frequency with the vowel formants--and when the s sound occurs between two vowels all three sounds have energy in the same region.

The same engineering developments have caused other workers to look again at the kind of information that people are trying to reconstruct here, and say, "Have we ever tried to capitalize on this information in other forms?" as Fry suggests. The information may be available through ordinary amplification, where you know that you won't have any energy above 500 cps, in some cases, or even at the fingertip--as I was suggesting to Risberg earlier--just by using a single vibrator and have that vibrator put out the energy to which it would normally be attuned. These are massive low-frequency devices, and they would put out something of the order of 200 or 300 cps.

As Fry suggests, some of these clinical studies have turned up with results that are as good as those that you get after all this fancy engineering (12). Perhaps, we should go back to some of our discussion of yesterday, and point out that--in the extreme--what is being done in the simple amplifying system, or the single vibrator system, is to present the time-varying information of a buzz and noise without any or at least very little of the spectral information.

COOPER Is there enough information for intelligibility of connected text?

HIRSH Well, because I'm conservative now that I'm older, I think that there is enough, when combined with lipreading, so that one should expend some effort on seeing what could be done with appropriate learning, before one invests both time and large size in the other systems. But I would hate to see those efforts terminated.

FRY Our experience is that there _is_ enough information so that, if you supply the child with it from a very early age, he manages to learn a system--that is, a linguistic system--good enough for him to receive speech from other people and good enough to enable him to control his own speech so that he is highly intelligible to other people. I'm not saying there is no child who fails to do this, but I'm saying that a great proportion of children manage to do it.

Now, if I could ask Hirsh one or two questions--you see a lot of very young children at CID, presumably?

HIRSH Yes.

FRY Can you give any sort of guess in what proportion of cases you can just not discover that they have any auditory sensation?

HIRSH There are no children in whom we get no response to acoustic stimulation, but I don't know whether that response is mediated by the skin sense or by the auditory system.

FRY Would this matter much? I'm not sure.

HIRSH Well, you asked me about the sensation of hearing--that's why I turned it the other way. There are sensory responses to acoustic stimulation in all of the children that we see. I can say that.

FRY Shouldn't you first establish for whoever it is who is learning this, that he can't learn it by some other more readily available method?

DENES Well, are you suggesting that an ordinary hearing aid, or, let's say, Hirsh's single buzzer or vibrator, would teach a completely deaf person or a person with a cut-off at, say, 500 or 1000 cps, fricative sounds?

FRY You're really in difficulty in defining what the hearing loss is.

DENES Somebody who has no hearing in the part of the spectrum in which we know that the energy is produced would be able to learn either to perceive or make a certain response with the device you are suggesting.

FRY This means, when you get to the point when you can do pure-tone audiometry, he has a hearing loss in excess of 90 or 100 dB.

DENES Say, a 90-dB loss or under. It doesn't matter.

FRY He will learn to produce some kind of noise which is quite passable to us as a fricative, and he can fill in the gaps well enough to know when you are using a word with s in it.

CHASE A question I would like to speak to concerns the general point of what kind of information has to go into a re-coded display. It concerns the point in time in which the experiment is being done, and its objective.

It seems to me that at least two broad classes of issue have been raised here, one involving the person who has learned speech and sustains some kind of sensory deficit involving a functioning feedback pathway. The question of prosthesis here is how to give information he needs for control purposes, along another pathway. This is the conventional way of thinking about sensory prosthesis. However, when we spoke yesterday about what the minimal logical requirements are of the control system, we considered that, whether we were talking about reception or production, pattern matching is needed--that the information coming into the system that will undergo utilization for control purposes has to be matched against some standard.

It seems to me the other issue we are speaking to here concerns the critical sensory information that has to be obtained and linked to the motor activity during critical stages in learning, so that standards can be organized for pattern matching.

The critical information requirements for these two purposes may be different. In addition, consider the child with congenital hearing loss, who goes beyond the age at which there is optimal speech learning for the normal child. The information requirements for teaching him speech may be vastly different from what they would have been at an earlier age in time, when there was greater plasticity for neural organization. I think, therefore, there are several questions here that might give rise to different answers concerning the kind of critical information which should go into a re-coded or altered display.

POLLACK What proportion of the deaf children are deaf from birth and what proportion become deaf at an age after which they usually learn to speak?

LENNEBERG In a recent survey of two schools for the deaf with a total population of 160 children there were only eight who, with reasonable certainty, had lost their hearing after three years of age. There were many stories of parents who claim that their child was born with hearing and then had a fall, but we discounted most of those because they seemed so incredible.

FRY I think it is a very small proportion, but I can't say how many.

RISBERG I should like to come back to this device that we spoke about--developed by Johansson in Stockholm--where you take only the fricative energy and transpose it down to lower frequencies. This seems to me to be very simple. Johansson has been testing this type of device for a couple of years now and I think he has about thirty subjects. Only one of them has not profited from this device. The amount of gain in correct perceived monosyllabic words is from zero to about 25 per cent. I think he will publish the results from these tests soon.

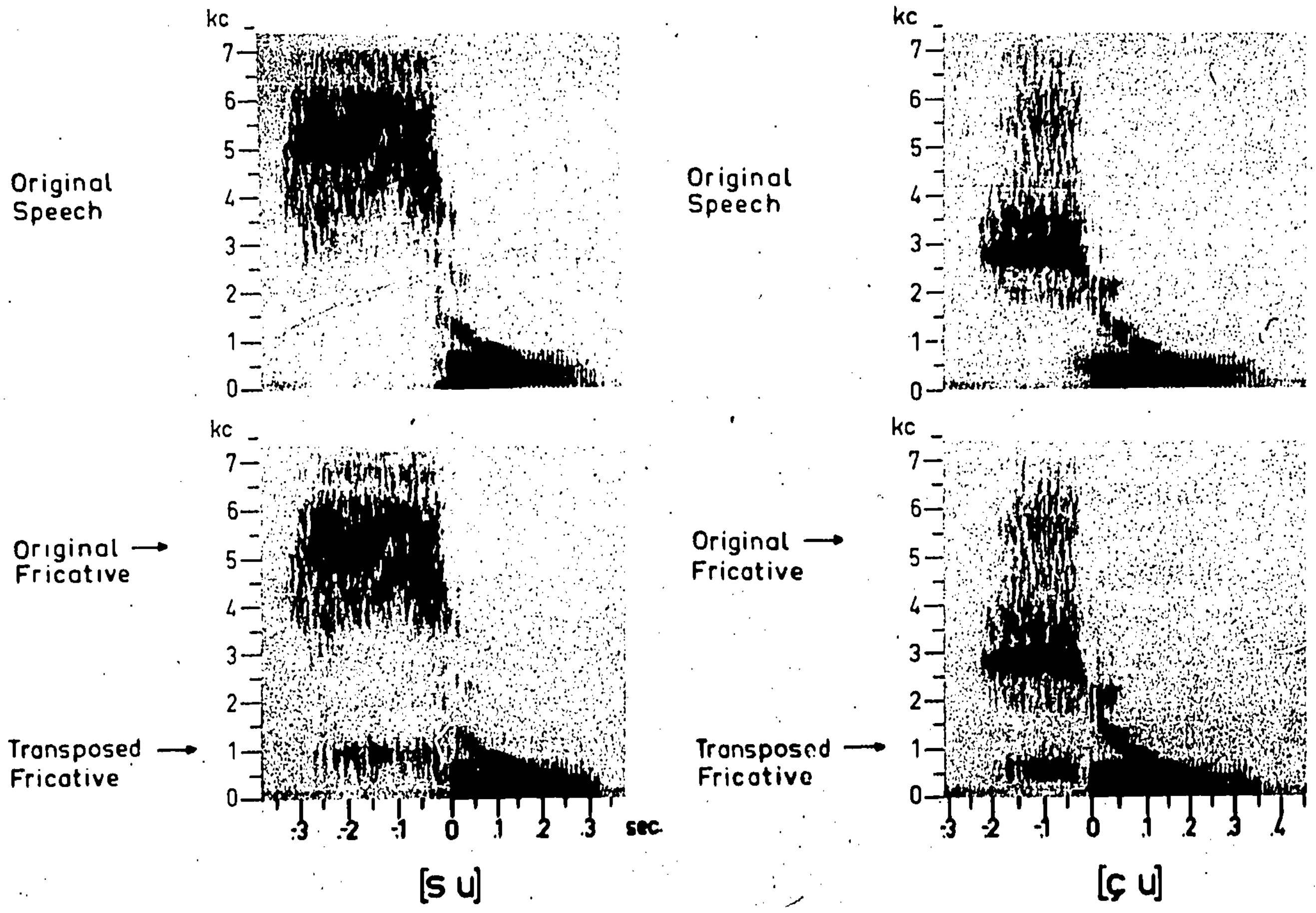

Figure 12. Examples of transposed fricative spectra. Energy from the spectrum of the fricative consonant in the example to the left has been transposed to a frequency region around 900 cps; in the example to the right the energy has been transposed to a region around 600 cps.

We have also started tests on a similar device. There are many different techniques that can be used, and we have tried two. We have not tested them yet on a hard-of-hearing subject. In one of these devices, only the information fricative is transposed down to the low-frequency end, but you can't hear any difference between the fricatives. In the other device, we try also to transmit place-of-articulation information.

We use three filters in the fricative region of the spectrum and transpose the energy in these filters down into three different frequency regions in the low-frequency part. Examples of the transposition are shown in Fig. 12.

You can easily hear the difference between the fricative sounds. I think we can hear them on the tape here. (A tape recording was played intermittently during these remarks.) You heard only that it is a fricative, there is no information about place of articulation. Now you will hear a second type of transposition and then two sentences, the first just low-pass filtered at 1000 cps with no transposition of fricative energy and then the same sentence with the transposition of fricative energy using the device that also transmits the place of articulation.

LADEFOGED What language is that? (Laughter)

HIRSH How can we know what language it is with a 1000 cps low-pass? That's another question. I heard something about a boathouse, that's all.

RISBERG The sentence is, "Which police first caught the wolf champing rotten zebra bait near my goat house?"

LADEFOGED The other one, I think, is, "Yes, judges do treasure very thin soiled T-shirts,"--if I remember correctly from reading it in a book some years ago.

HOUSE I have a feeling that, in many experimental situations the subjects are getting the kind of training that they should get in a learning or clinical situation but don't. For that reason, whenever you do some work with someone who needs help, you always seem to get some improvement. I am always skeptical about these matters because the thinking that has led to transposition and compression experiments can kill the kind of model that I'm interested in. I don't believe that you listen as though the materials are just pieces of sound. There is something more basic in the kind of speech processing that humans do besides just listening to coded sounds--you've got to relate the sound structure to something else.

I agree with the suggestions that Hirsh has made--that all or most of the information that is going to be helpful in retraining, or in original learning, is present below 1000 cps. You are missing some information, but a lot of the information that you think you are missing is there in some form. A great deal of information about temporal patterns is present in the sequences of sounds in the language; a great deal of information in amplitude change is made available as the speech flows along. These are just some of the redundant cues that tell you about the presence of consonants that seem not to be there. Therefore, I would argue that if you have amplification of low frequencies up to, say, 800 or 1000 cps, plus intensive training--the kind of training that you give people when you are introducing new devices, perhaps even using visual cues and putting a good therapist in the feedback loop--there is hope in using amplification *per se*.

RISBERG Yes, but one of the subjects in the group has had this device for, I think, five years, and two or three others, too, have been using it for one year, every day, and they seem to like to have it. They don't like to use anything else.

HOUSE But I remember, also, reading how one member of an experimental group learned how to read moving speech spectrograms (113) while most of us who work with sound spectrograms of speech can't even read them when they are standing still (laughter), and so I take this kind of report with a grain of salt. It goes against all my theoretical expectations, as well.

RISBERG I can give you some grounds for this. We ran a short test to see if normal-hearing subjects could use this information. We masked the subjects with a low-pass filter, 1000 cps, and used a white noise about -15 dB relative to the speech. Then we trained them for a total of two or three hours, to see if they could use the transposed information.

There were three groups: one used only the low-pass filter; one used the equipment that transposed only the information *fricative*; and the other used the device with three filters. There were some differences between the groups. The fricative information went through much better with frequency transposition. For example, in words like stain, where you have s and t together, the group that listened to only low-pass filtered speech tended to miss the s but the other groups put in a fricative. But if you compare the number of correct words received, then the group with only the low-pass filter was better than the other groups. These results agree with those of Oeken (108).

Of course, the training was very short and normal-hearing subjects were used. There was quite substantial evidence that this new information tended to destroy other information, but that they could gradually adapt to this.

HIRSH May I ask the conference whether you wish to continue our discussion of the receptive disorders of speech, with respect to this re-coding of information for the deaf child or adult, or whether there are other aspects of the problem to which you want to give some attention before we close this afternoon?

LIBERMAN I would like to say something that relates to this and also to the points that were made yesterday. In his introduction to the first session, Fry asked whether we knew anything about combination of cues or whether we ought to try to find out something about how cues combine.

Now, the fact is that there are several acoustic cues for any given phonemic distinction, and these often lie in very different frequency regions. I would suggest, therefore, that we might be able, on this basis, to help the deaf. But we need more analytical information about which phonemes deaf people hear and which they do not, and still more analytic information as to which acoustic cues they hear and which they do not. If one should find that a particular deaf person is having difficulty with one of the more prominent cues and not with a less prominent one, there are various kinds of things that might be done to call attention to the less prominent cues, including, for example, using synthetic speech.

But, quite apart from this, I am really asking my usual question, and I think I always raise this question with you, Hirsh: Do we have this analytical information yet, and if not yet, when are we going to get it?

HIRSH I would mention, since we last talked, that there is considerable information in a thesis by Rosen (118). The only difficulty with interpretation is that he did not use a closed message set, but he used a typical 50-word list of monosyllabic words. But there is considerable information on the kinds of errors that are made.

DENES What is the essence of Rosen's experiment and conclusions?

HIRSH With regard to consonant articulation, one generalization that is available is that the kinds of errors that are made are no different from those that were reported by Miller and Nicely (100) in connection with normal listeners and low-pass military communications systems; that is, the manner of articulation rides through but the place of articulation gets lost.

POLLACK I believe this is also Schultz's finding at Michigan (122).

LIBERMAN This is for what kinds of listeners?

HIRSH Rosen's listeners all had sensori-neural hearing losses. There are particular vowel confusions of interest to the phoneticians, particularly which vowels are most often confused with which.

DENES This is also interesting from a completely different point of view. You remember that speech statistics paper of mine (30), where, again, the functional load is very much heavier on manner-of-production distinctions than place-of-articulation distinctions.

POLLACK What is "functional load"?

DENES Functional load is the extent to which a language depends on a particular feature--in this case, articulatory feature--to make a minimal distinction between two words. For example, if a language frequently uses the plosive-fricative opposition to distinguish words from each other, than you would say that there is a high functional load on the plosive-fricative opposition. If, on the other hand, nasal-plosive oppositions, for instance, are rarely used to distinguish words from each other, then, you would say there is a low functional load on the nasal-plosive opposition.

HOUSE Are these functions specific to a given language?

DENES Oh, yes. The snippets I have seen in other publications, however, seem to indicate that in most European languages the frequency distribution of articulatory classes is similar to what I found for English.

LIBERMAN These distinctions with the heavy functional load are categories, and this is why they are efficient. They can carry the load.

HIRSH How far can this notion be carried? It comes preciously close to the old notion that I heard at lunch, that consonant distinctions are more important than vowel distinctions for general intelligibility in English. What about that generalization--does it still stand, or does it require some subdefinition?

LIBERMAN I was just talking about this as alleged--this is the assumption, this is the superstition. But I do believe it.

DENES If you write down a sentence, indicating correctly the identity of each consonant but only marking the positions of the vowels by asterisks without specifying what they are, and then do the reverse, identifying the vowels but only marking the positions of consonants without putting in what they are, you will find that the former sentence will be completely intelligible, and, by and large, the latter sentence will be completely unintelligible.

GESCHWIND Hebrew is generally written without the vowels.

BROADBENT In that case, you remove rather more letters.

GESCHWIND Not necessarily.

BROADBENT Well, let me put it this way: If you delete all but the ten most frequent letters, you get something that is reasonably intelligible, whereas, if you delete nothing but the ten most frequent letters, you delete the message. The ten most frequent, of course, include most of the vowels. So I would rather doubt this story, which, I must admit, I used to believe, Denes.

LIBERMAN It's not just a question of frequency of letters; it's the number of distinctions and their frequency. The kind of load you want to measure should take account of both of these things, shouldn't it?

DENES Yes, it certainly should, and it was done in my paper (30). I should have thought that omitting the least frequently occurring sounds would have had a more serious effect.

BROADBENT No, it has a less serious effect. You see, you leave more sounds in the text. This is connected discourse, of course.

DENES Oh, I see.

BROADBENT It is connected discourse that one is usually interested in rather than isolated words. You might have some trouble with isolated words.

GESCHWIND In written Hebrew there is a nearly straightforward consonant-vowel alternation, with a few exceptions. By cutting out all the vowels in written Hebrew, you are cutting out half the symbols, and yet the script is clearly perfectly comprehensible. If you did the reverse, that is, included the Hebrew vowels but removed the consonants, I'm sure that it would be completely incomprehensible.

BROADBENT I was so surprised by the result I described that I suggest one try it even with Hebrew.

HOUSE In all these cases, however, you're talking about letters, not about articulatory processes, so far as I can see. I find it very difficult to see how you can remove the consonants from the vowels to make the test.

DENES I don't understand. Are you now talking about removing these from acoustic sequences?

HOUSE I don't know how to remove them from an acoustic sequence.

DENES I quite agree--this is impossible. In discussing the functional load carried by consonants and vowels I was speaking about the distribution of linguistic units and not of articulatory or acoustic segments.

HOUSE I know how to remove them from an orthographic sequence, and I suspect that neither one of these cases bears directly on the point.

LIBERMAN Well, one can only approximate this, of course, but it surely must be true that different phonemes and different classes of phonemes carry different loads. This may be very difficult to measure, indeed, but the kind of thing Denes has tried to do is certainly relevant.

GESCHWIND Now, all that these transposition devices are doing is creating another language in which none of the phonemes use the high frequencies. The child need only learn to match his articulation to the new phonemes.

HOUSE Herein lies the problem. I'm not sure that human language behavior can be elicited by matching to just any arbitrary signal--we seem to match to signals that we can produce.

LIBERMAN But what if you have the kind of feedback that Denes provided?

DENES You wouldn't be learning another language, but you would be learning speech in a different physical medium.

GESCHWIND He would be learning another phonemic system.

HOUSE I then suggest that he could match just as easily to some low-pass filtered material. We have some scant evidence, for example, that adults don't learn auditory displays that are speechlike in some arbitrary sense as well as they learn displays that are really speech, or that are very much unlike speech (55).

GESCHWIND The problem here is that you used adults. After all most English-speaking adults have trouble handling actual speech in foreign languages.

LIBERMAN In these cases, I'm not sure that I want to go back to the child, because these results do fit with the adult.

LENNEBERG Allow me to change the topic just a little bit. The trouble in speech perception that seems hardest for me to understand is the temporal resolving power that becomes evident in speech perception. One can do a very fast job in identification of a sequence of signals, faster, maybe, than one can do in visual perception. I'm not sure that this is true, but it almost looks that way.

I was wondering, coming back to Risberg's tactual perception, if it might not be interesting to find out whether research on temporal perception has been done, and whether the tactual system is anywhere as fast as the auditory. If this were so, this would break down the notion that there is something special about speech perception.

OLDFIELD I think Geldard (43) has done some work on tactual perception of very short intervals between the same kind of stimulus, a square pulse delivered to the fingers. I forget what the time intervals are, but they are exceptionally short, down to 100 msec or so.

CHASE I wonder whether the way in which we examine the receptive capabilities of patients who present abnormal speech development should not include some tasks like those we have been speaking about in the past few minutes, which, rather than dealing exclusively with some of the complexities of discriminative abilities for speech perception, also permitted us to get ideas about things like temporal resolution capabilities _per se_?

HIRSH Some of the acoustical ones themselves are not very revealing; that is to say, if you take Miller's interrupted noise (102) and look for a fusion frequency, for example, you will find that in the patient who has great difficulty recognizing words correctly, the fusion frequency which is critical is no different from what it is in a patient who has no difficulty recognizing words (53).

LENNEBERG That is the relevant test.

CHASE I would like to invite comments about the question of what the armamentarium of techniques for the clinical assessment of receptive capabilities ought to be.

HIRSH For speech perception?

CHASE Yes, pertinent to speech perception and the evaluation of the child with the speech or language disorder. In your opening comment, Hirsh, you mentioned, again, I think, as a historical anomaly, that we have a lot of information about pure-tone hearing acuity, and we have information about sensation level and other simpler aspects of the reception of complex stimuli, like words. However, there is a big gap in terms of our ability to detect processing deficits, and, in a sense, thinking over the material of the afternoon, we think primarily about disorders of speech production--they are fairly apparent. We may not understand them, but they are obtrusive; they bring people to the hospital. They don't come with complaints about their receptive capacities, for the most part, but they do come with complaints about their productive capacities. When it comes to the evaluation of the receptive capacities, we are back to the problem that Broadbent defined yesterday, with respect to perception, which is the description of what, in a sense, is a private event.

I wonder what the consensus is about what should be added to our rather meager armamentarium of techniques for the assessment of receptive capabilities pertinent to evaluating speech perception capability and the development of speech productive capability?

HIRSH Are there suggestions?

GESCHWIND Hirsh mentioned earlier that there are children with moderate deafness who show rather marked learning deficits. Are there children normal with respect to language acquisition with the same type of audiometric deficit?

HIRSH My guess is, in the normal circumstances for learning language, you can trade, I don't know how many points, of IQ for so many decibels of hearing impairment. We started testing a few dimensions by asking some of our deaf children about their ability to discriminate duration, about their ability to count number of sounds in a sequence, and about their ability to recognize or to distinguish vowels that are distinguishable on the basis of first formant alone. The list should be longer, and the suggestions should come from the acoustic phoneticians, for, surely, they know what cues are important for the perception of speech.

FRY Can I put a word in here on the philosophy of this subject? We must not really think of some sacrosanct set of cues. This is what Geschwind was talking about. The child will learn to develop the necessary set of cues, and they are not necessarily the same as the set of cues that I use when I take in speech.

POLLACK There has to be discriminatory information, though.

HIRSH There have to be come cues that you can demonstrate are available to him. Having measured his sensitivity, what do you know about his discriminatory capacity for any cue?

LIBERMAN There is no real problem here, because you can find out what cues the normal person uses and what cues the abnormal person uses. One needs not only to know, as far as speech is concerned, what phonemes or what phoneme classes, or what dimensions, are knocked out and which are not. Once he has that information he has got to probe still more analytically into the situation, using the techniques that are now so readily available in speech synthesis, to find out which cues can or cannot be heard. The perception of the individual cues should be studied first in a speech context and then in a nonspeech, purely psychophysical, arrangement.

GOLDSTEIN It certainly seems possible that you could have a close-to-normal audiogram and still have a lot of trouble in the peripheral side of the afferent pathways. Maybe, it is worth testing for that first, before getting to sounds that are quite close to speech.

HIRSH Broadbent, I think that a comment of yours got lost in the general noise.

BROADBENT I was trying to fill a gap. I was trying to turn this question back to Chase, because he had this general model or description of the speech process which was going to guide investigation, and I wondered what tests of performance this model suggested.

CHASE That's fair. (Laughter) This is a hard question to answer.

To give you an example of what you might do in a single case, I'll take the question of pattern matching. It has come up in the context of receptive capabilities in the model that Stevens and House presented; it has come up in the context of control capabilities, with respect to feedback monitoring of speech. If I lose my ability to detect and analyze the sensory representation of my own speech or somebody else's, so that now I am getting a mismatch because I have misprocessed, this would serve as one example of how I might get into trouble, and it points in the direction of kinds of experiments that this information-flow model suggests.

In this regard, the studies on temporal pattern discrimination interest me a good deal, because it seems to me that if you've got difficulties in temporal resolution, you are in very serious trouble with respect to pattern matching, and any of the other processing operations you can consider to underlie speech discriminative capability and, certainly, speech productive capability.

The kind of experiments which have been mentioned are, I think, quite germane. Hirsh spoke about the documentation of receptive disability for sequential stimuli in aphasic children. Efron (31) has reported difficulties in the correct identification of sequential order of pairs of visual and auditory stimuli in adult aphasic patients, and suggests the kind of thinking implicit in the information-flow model--that some of the profound deficits of speech reception and speech production capabilities that we see in the adult aphasic do not necessarily represent high-level deficits of the programming of the motor gesture on the productive side, but, rather, low-level deficits in temporal resolution.

GESCHWIND I believe that Efron's experiments are simpler than the description just given. They showed very simply that if stimuli are delivered simultaneously to the two hemispheres the stimulus reaching the left hemisphere is judged to be the earlier one. In a more general way he showed very elegantly that the judgment as to which of two stimuli precedes is made in the left hemisphere.

BROADBENT Any stimulus, or speech?

GESCHWIND Any stimulus. In fact, he didn't use speech in these experiments. He was using flashes of light and touches. These judgments are made on the left side. They are markedly impaired in aphasics, but interestingly enough not in aphasics with receptive difficulties but rather in those with production difficulties. The simplest explanation in my opinion from Efron's results is not that there is any complex sequencing mechanism but rather that judgments of simultaneity are made verbally and hence use the anterior part of the speech area.

CHASE I certainly accept the controversy about it. I am not disturbed by the differences he found in terms of receptive and productive aphasic deficits, because I think the question of what you are doing on the receptive side and what you are doing on the productive side is a slippery one. I think that it is pretty hard, in most of the issues in which this has come up in the past two days, to differentiate clearly the neural processes pertaining to one and the other. There are so many points of juncture that are not clearly defined that I would like to leave this question open.

SESSION 5. Part 1 - Neural Mechanisms and Models

COOPER We come, in a formal sense, to the question of neural models and mechanisms this morning. This is, as you remember, the session that was set aside for whatever we wanted to discuss, whether it had been programmed or not. I assume that one of the things you may want to do is to hear from Chase, since he was crowded out of yesterday's sessions.

CHASE The general problem we want to start discussing is the question of neural models, and since a model, whatever else it may be, is the product of the imagination, there are any number of forms these models might assume, and any number of practical contexts in which you might undertake to fashion them. And so I would like to make a few introductory remarks about guidelines that might be kept in mind, as we think about neural modeling.

It seems to me that what we would like to do when we fashion a neural model is to effect as close a fitting of what we have come to feel the functional capabilities of the system are with what we have come to know of its componentry. It seems also that one of the very important functions of a model is to be productive of further insight into both the definition of other functional capabilities that we do not yet understand, to permit us to discover componentry that we have not yet adequately defined, and, ultimately, to continue through this sort of counterpoint process to refine progressively the fitting of functional capabilities and the underlying hardware and processes that define them.

There is a lot of information about the nervous system, and, just as in the case of some of the classificatory schemes we reviewed yesterday, there are certain conventions that have arisen, historically, as ways of organizing insights into structure and function in the nervous system. There are certainly kinds of mapping which have come to be accepted conventionally, and we go to these systems of mapping to get a lot of our information, both about structure and function. One type of mapping simply shows the general location of neural pathways and the direction of information flow along them. Conceptually, it is entirely comparable to the transit map for a large city. Another kind of mapping is simply the identification of a place within the nervous system at which some significant event has occurred. This kind of information does not give us a model, but it gives us pieces that might be fitted into a model.

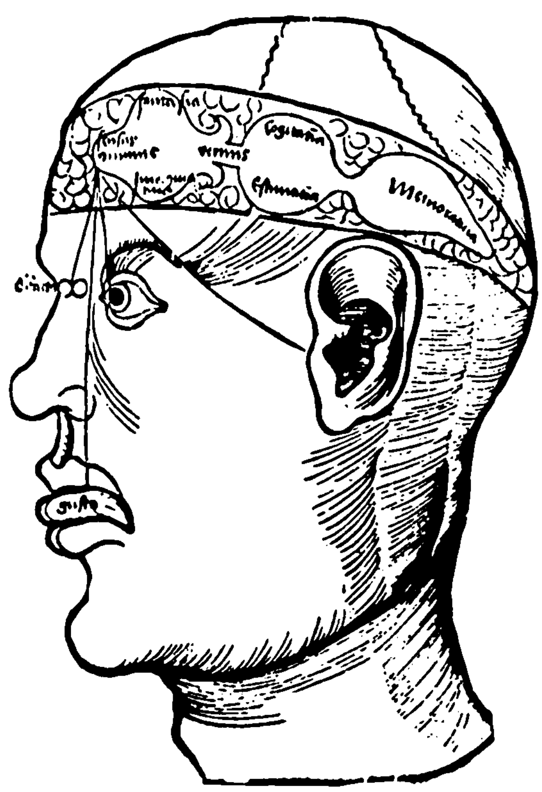

Figure 13. Early woodcut showing the cerebral ventricles and the faculties considered to reside within them. From G. Reisch, Margarita Philosophica (Strassbourg, 1504).

Figure 13 shows an early example of something that, I think, qualifies as a more adequate model of neural function, although I'm sure most of us would consider this a premature and, in many respects, anatomically incorrect characterization of structure-function relationships in the nervous system. But, this, as you know, is one of the familiar examples of medieval neurophysiology, showing the sensory receptors as channeling their information to a common juncture which was placed in the most-anterior ventricle of the brain. It then posits progressively more-posterior processing of this information, under the broad categories of thinking and imagination and judgment; memory is tucked away in the most-posterior ventricle.

It seems to me that, as we approach the question of neural modeling with respect to speech, what we want to do is not unlike what we see in this figure. That is, to obtain progressive definition of the functions that we think underlie the speech-communication process--to collate, or, at any rate to catalog, the componentry that we think is pertinent from the classical fields of neuroanatomy and neurophysiology, and attempt some kind of functional fitting together of the two. At this time I will list some of the major functional capabilities that have been reviewed during the past few days, point to some of the componentry that I think is pertinent, and then invite discussion.

We know that a great deal of tissue in the precentral cortex is associated with the musculature used in elaborating the speech motor gesture. I suppose we don't know whether all of this tissue simply represents the tremendous range of capabilities for programming the musculature of the vocalization system which later becomes categorized into much simpler gestures, but, at any rate, this is certainly componentry that we are concerned about. I invite discussion from those of you who have studied the speech motor gesture and have come to conclusions about the economy and the categorical nature of the elaboration of motor gestures. Perhaps someone will give us speculations about this very generous componentry which seems to suggest that a tremendous number of possibilities for programming the musculature used in speech might ultimately be subject to the kind of constraints that seem to be suggested by the observations about the categorical-economical nature of the set of simple motor gestures that define speech capability within a given linguistic system?

POLLACK Does there exist a precise cortical mapping of the vocal system, or only a gross mapping?

MILNER I don't think it has ever been possible to map it out in detail. I don't mean it doesn't exist, but it has not been done.

CHASE When we are engaged in speaking, there are important receptive aspects of the process, whether I am monitoring my own speech or whether I am behaving as a link in a communication system. And so, in either event, whether we are focusing on productive or receptive aspects of speech, we have to start at the peripheral transducers and consider mechanisms for the shaping and early processing of sensory information.

We know that many sensory channels give information pertinent to the monitoring of speech and potentially for the recognition of speech as well, and I think one of the points that we should consider as we move to central processing is just how identification is made. It seems that the capability of intermodal transfer implies that some part of the nervous system is able to recognize common morphologic features of input, independent of the modality along which it was presented.

The error-detection and error-correction operations which are shown in Fig. 5 in the context of control, have come up repeatedly in our discussion in the context of perception and in the context of the ability to regenerate an input. And so, I think we should give some consideration to what componentry in the nervous system might be involved in the shaping of the standards against which a sensory pattern can be compared for a matching recognition operation, and whether this componentry is the same whether we are monitoring our own speech, recognizing the speech of another, or trying to reproduce speech acoustic input in terms of our own motor translation.

There are a lot of things that are not drawn in the large box under <u>central processing</u>. One of the most important, certainly, concerns the way in which speech motor gestures are planned. We might look at the simple structural organization of a message, which, at the very least, requires that we go to our catalog of motor gestures, pick the right ones, put them together in the right way, and get that message out to the effector systems. Or, we might look in a more complex direction, and think about these ordering operations taking place in the context of some communicative objective. In either case, we are concerned about central processes involved in the organization of the message. I invite Milner's comments on the extent to which she feels the work on cortical function is telling us about the planning and the organization of speech-language activity.

As we move progressively over to the right portion of Fig. 5 we once again come upon componentry that is not unique to the speech motor system. We know that the message once organized, has to filter through the componentry involved in the organization to motor activity per se, and so, when we see abnormalities, for example, of the cerebellar system, involving, as it does, the programming of voluntary motor activity, we are surprised when we are confronted with dysarthria of a characteristic sort. This may be a trivial matter with respect to unique processes involved in the organization of speech, since it may just mean that the speech motor program has to pass through a number of toll gates, if you will, which represent a higher-level common level pathway for motor activity.

When we consider the basal ganglion system and observe clinical abnormalities of the sequential release of speech motor gestures in common with the abnormalities of sequential release of other motor gestures, I think that we have to question seriously whether some of the componentry of the basal ganglion system is involved in sequential programming of motor activity, and whether this system is implicated in the initial planning of speech motor activity. The more central the concern we have, and, certainly, programming and pattern matching are fairly central operations, we see so much componentry that could be used that we have to use the description of the functional operation as a guide to what might be significant. The problem is not dissimilar, in my mind, to the problem that Cooper and his colleagues faced at one time, in looking at the total spectrogram and wondering what was essential for a particular kind of operation.

COOPER I wonder if we do have, in fact, so much componentry that we must look at function in order to understand what is happening. Couldn't we look at the componentry and make some statement about what functions could or could not be performed? I believe Geschwind tried to do this with respect to object-naming in humans.

GESCHWIND The answer is simple--one must do both. If you don't look at the componentry, i.e., the anatomy, you are simply throwing away one important source of information. I certainly don't believe that there is so much componentry that it is impossible to make sense of it. Knowledge derived from anatomy has been fruitful and should certainly continue to be very useful.

LENNEBERG I would like to add something here, and it is something like a criticism of terminology. When we talk about componentry, I conjure up a vision of individual pieces that are put together, and this is clearly not the case in the brain. There is no single piece or component which is independent and which is silent at one time and noisy at other times.

As far as we know, the brain is a completely tight-knit unit, and activity involves all the so-called components, at all times. I think that we ought to look at the various interferences that result from disturbances in various parts of the brain. This is not to say that we should conjure up a model in which part of behavior is manufactured up here, namely, in the cortex, and something else is added in one other piece. Probably, the entire brain is involved in speech.

COOPER But aren't we tempted to assign functions to specific areas by the fact that when one cuts away some parts of the machinery that were active, nothing happens with respect to the functions?

LENNEBERG The only parts of the brain you can cut away with impunity are certain portions of the telencephalon, the upper part of the brain. Nothing else that I know of can be cut away. You can't cut away the midline structures, as far as I know, or the lower structures. The brainstem is an extremely tight-knit unit, and almost any kind of lesion there is catastrophic.

GESCHWIND Perhaps all of the brain is involved in speech to some extent. The important point, however, is that the contributions of different regions differ grossly from each other.

MILNER Referring to this question that Cooper raised, I didn't interpret it as asking whether there is a place for a Skinnerian position. I interpreted it, perhaps wrongly, as meaning we got our cues and our starting point for investigation from disorders of function, but then we might look into the structures and see if we can find the main systems which are involved. But I think I interpret his question to ask whether we do that or whether we start looking at the details of the componentry and the details of the anatomy. You only get your ideas of what to look for from the disorders, from the function, and then you proceed, I think to map out systems.

GESCHWIND We really have to work both ways. All through the history of aphasia important advances have been made from the study of function, but on the other hand many very important advances have grown out of a study of the anatomical factors.

MILNER I do agree with that, but I think it would be wrong to start with too many details.

GOLDSTEIN In a problem as complicated as the role of the brain in speech behavior, to take one path would certainly be a mistake. We have to remember that these lesions usually involve thousands or more single nerve cells, and that they are not all doing the same thing even if they are all in an area that is classified as functionally homogeneous. In our work with animals other than men--which makes it difficult to discuss language or speech behavior--we are getting some ideas about the coding of sound stimuli by single cells in the auditory pathways. There are those who say we shouldn't study the single cell in groups until we understand the single cell alone, but I would say that I'm glad there is work on the physiology of the single cell alone. I am also glad that there is work on groups of cells as well as work on cells in the whole brain of humans.

CHASE I think that it would be perfectly in order, after this discussion of the difficulties that surround this kind of experimentation and interpretation, if we could review some of our successes, modest as they are. I wonder if Goldstein would be willing to begin with some comments on the coding of sensory information?

GOLDSTEIN I can mention some work--not my own--on vision in cats, and, maybe, after I have outlined it, we can discuss whether it is germane to speech. I would like to focus attention on what Chase called input processing. Some people in neurophysiology think of the cortex as being very central, but, in the context of the results I shall discuss, the primary cortex of the animal is involved in input processing.

The work that I will describe is mainly that of Hubel and Wiesel, and it in turn relies on earlier work. I am sure some of you are quite familiar with it and so I will rapidly go over that part which is reported in Scientific American (56) and then present some more recent data which may be quite appropriate to our topic.

They are working on anesthetized cats. The cats' eyes are not moving; the cats sit with eyes focused on a screen and spots of light and other patterns are projected onto the screen. A very small electrode is placed near enough to a single nerve cell so that it records the activity of only that single nerve cell.

In the cat the most peripheral level at which you can look at single nerve cells is that of the ganglion cells which are not first-order cells in the visual system. What you see at the level of the ganglion cells is as follows. If, say, the cat is looking at a given spot, you will find an area in its visual field where a cell responds. The cell usually is firing all the time in a more-or-less stochastic pattern. If you shine a spot of light in a small circular area, the cell will fire more actively and you get what is sometimes called an on response to the spot of light, while in the surrounding field you get inhibition. In other words, you get a decrease in firing at the beginning of the presentation of the spot, and an increased rate of firing when the spot is turned off. These areas are quite small--maybe a couple of degrees of the visual field. For some cells there is a pattern of excitation in the center and inhibition in the surround; for others you get the opposite--inhibition surrounded by excitation. For an on center cell, if you make the spot bigger and bigger, you get more activity, as long as you stay within the central region. If you go over to the surrounding region there is inhibition as well as excitation and you get less activity.

An interesting thing about the ganglion cells in the cat is that, although the position of these receptive fields will vary in size, they are all about the same shape. They all have these simple concentric circle patterns. You see pretty much the same thing at the level of the lateral-geniculate, with, perhaps, the balance between excitation and inhibition being a little more delicate.

The patterns are very different at the cortical level. You find more or less linear patterns or line patterns rather than concentric circles, and the patterns tend to be larger. A typical pattern might be a line of excitation, surrounded by inhibition (57). Another pattern might show two lines of excitation and a region between where shining the spot gives inhibition. Furthermore, there are cells that do not respond too well to a spot of light, but respond very well to a slit of light. I will call this case the slit cell, and there are others more analogous to split cells.

LIBERMAN I didn't think my data could be so easily explained. (Laughter)

GOLDSTEIN The cells in which you can map out regions of excitation and inhibition are called simple cells by Hubel and Wiesel. Besides the simple cells, there are cells in the cortex which respond, say, to a light in a certain orientation. It can be a slit of light, and it can be in any place in a certain field, but it must be in a certain orientation. For these cells, you can't locate inhibition and excitation in the field. This whole class of cells is called the complex cells. These cells respond well, say, to an orientation of a slit in a field, or to an edge in a field; some even respond to a corner (57).

Some of the recent work by Hubel and Wiesel (58, 135) is quite interesting. I won't say how closely related it may be to speech because, you see, this is still in cats; in fact, it is the work they have done on kittens. The first thing they did was to look at the cortex of newborn kittens to see if they could see the sort of patterns of receptive activity just discussed. What they found is a picture of coding that is in many ways similar to that seen in the grown cat. The response patterns are more sluggish, and they seem to find fewer cells in a penetration through the cortex. There is less activity and more sluggishness, but the patterning seems to be similar.

I want to bring in one more point, and that is the point of response of cortical cells of the two eyes. When you can focus the two eyes so that they are looking at the same field, you can ask whether a given cell responds to stimulation of both eyes or to only one or the other. Hubel and Wiesel have done some work on this problem. In the normal cat, when they observed a few hundreds of cells, the distribution of response was about 10 per cent to only one or the other eye, and the other 80 per cent to both, in varying degrees. Of this latter group, about one-quarter responded pretty much equally to both eyes. But the point is that 80 per cent of the cells responded to both eyes.

These experimenters have also done some work with animals where one eye is sutured as soon as the kitten opens its eyes; that is, they suture the lids of one eye, or they put a translucent occluder over one eye. This allows the kitten to be reared normally, but getting all its visual experience from the normal eye while the other eye is getting only light and not getting patterned visual experience.

The interesting thing about these kittens, which they have studied at an age of about two months or a little longer, is that after this special experience 83 out of 84 cells studied responded to the normal eye only. The 84th, when it responded to the eye that had been occluded, responded in a very abnormal fashion. Here, then, is rather complex coding, which seems to be there almost at birth, and, yet, certainly, seems to be plastic to the extent that it depends on the visual experience of the kitten.

COOPER But is this a change in the cells that responded initially in one way, or a recruitment of cells that were inactive at the beginning but are later organized by experience?

GOLDSTEIN I think the guess that Hubel and Wiesel would make is that this is a sort of atrophying of the connections from the eye that is not used, rather than a complete dissolution of everything and then building it up again.

LENNEBERG They have done the histology and shown that the failure to grow works in young kittens, and the cells that correspond to the occluded eye fail to grow as the others do. That is the point.

GOLDSTEIN I think you are referring to the lateral geniculate--the failure of cell growth being reflected in smaller layers in the lateral geniculate body.

LENNEBERG Yes, you're right. This was in the lateral geniculate, whereas, in the cortex, they found no histologic changes whatsoever.

GOLDSTEIN Since the cortical cells usually receive afferents from both sides, while the geniculate cells usually receive afferents either from one side or the other, you might expect that this deprivation would be more serious for geniculate cells than for cells in the cortex. I think the key here is that the functional properties of the cells are definitely different according to the experience of the kitten; and, also, behaviorally, the kitten can work very well at placing its paws and walking around with the normal eye, but he cannot do so with the eye that has been occluded.

The point I am trying to make is that the input coding--and I still look at this as input coding--may be fairly complex, and, certainly, is plastic and may be changed by early experience. Unfortunately, I don't think we have a similar story for the auditory system. We do have some data on coding at various levels of the auditory system--studies with single units--which I would be glad to discuss.

COOPER If you call this input coding, where, in a functional sense, would you draw the line between this kind of operation and the next stage in the process?

GOLDSTEIN I guess the next step would be the association areas. I don't think there is much of a story on the coding in association areas. What the single-unit physiologist has found there is usually more of a convergence of modalities.

LENNEBERG May I add one thing? They did find, interestingly enough, if they did similar deprivation studies on older cats, this change in plasticity did not take place. It could be demonstrated only in the neonate kittens; there was a definite age gradient or maturation gradient, from which they concluded that plasticity is present during formative stages and then stops.

GOLDSTEIN They even showed this to be graded. I think they took some kittens at a few weeks and found some but less asymmetry in the response patterns of the two eyes. Then, in an adult cat, as Lenneberg said, there is no change.

LENNEBERG As a matter of fact, to make one other comment, even the histology changes. The kittens did catch up with their undersized cells in the geniculate eventually, but if the cats were sacrificed at a later time, those cats showed no histologic changes, in contrast to the ones that had been sacrificed earlier.

POLLACK Has there been a corresponding series of studies with auditory deprivation? Have the results been negative or do we not have any results?

GOLDSTEIN The auditory system is a little more difficult to study in some ways here. I should answer this briefly and say no. (Laughter)

POLLACK There was a recent note published, as you probably noticed, on the effect of auditory deprivation on the pinnal reflex in the guinea pig (7). This is the first note I have ever seen which suggests that auditory deprivation in the young animal will result in auditory dysfunction at a later date.

HIRSH It is not quite true that there is no information. Some is indirect, of course. The reason for this extreme lack is that auditory deprivation that will be later restored is almost impossible to arrange. You can excluse external patterned sound by a continuous masking sound--essentially an analog of Hubel's translucent cover where you have light stimulation. But unless that is very strong light or strong sound, you will have patterned body noises that cannot be excluded, so that auditory deprivation even in these general sensory deprivation experiments, has just not been successfully obtained.

The indirect evidence that we do have concerns the deaf child, where you may have a moderate loss measured at or near birth, and then no auditory stimulation by way of amplified sound and the kind of instruction that Fry was calling for yesterday. This is typically followed by a much more difficult time teaching that child at the age of seven or eight years when he shows up in school than one who has been so stimulated with sound.

GOLDSTEIN One problem that will make the auditory system a bit difficult to study, I think, in the sense that Hubel and Wiesel have studied the visual system, is the question of timing--information structured in time versus information structured in space.

For the visual system the coding seems to be laid down in anatomical patterns within the nervous system. Thus Hubel and Wiesel are able to work with animals under anesthesia, which quiets things down. And it is true that they see more or less the same responses in the unanesthetized cat, but there is so much activity that it makes it hard to study things.

In audition there is no question but that much of the coding involves time, and that, as soon as we anesthetize the animal we lose the ability to study this, at least at the higher levels of the auditory pathways. This is a major technical difficulty in studying single-cell units in the auditory system.

I would say that for temporal coding, the auditory system stands alone. It is very hard, as I understand it, in the tactile system, to push skin in such a way that you get a single discharge from an afferent fiber. On the other hand, in audition, it is very easy to present a single click and look at a single fiber in the eighth nerve and see that fiber fire once in response to the click (69).

POLLACK Does this hold true in the lateral line organs, too? Won't they preserve the temporal integrity very well?

GOLDSTEIN I don't know. The animals I'm thinking about don't have the lateral line.

COOPER What about temporal patterns in audition as compared with spatial patterns in vision? Would you expect a different organization of the input data when time is the variable, between vision and audition? In these two modalities, one seems to be coded in time (for hearing), and the other in space (for vision). Would you expect a different kind of input coding at the cortex when time is the dimension as between the modalities?

GOLDSTEIN I would expect it, but I can think of very little direct data to back this up. I think we have some data from gross electrode recordings that indicate that the auditory cortex can follow repetitive stimulation of the auditory end-organ to rates up to about 200 cps (47). I believe this is probably exceeding what the striate cortex can do in response to interrupted light. We also know that, psychophysically, it is well below the fusion frequency of chopped sound.

Could I say a few things about what we do know about the coding of single units in the auditory system. Most of us think about the eighth nerve as being connected to the cochlea in a functional sense. Fortunately, thanks to Bekesy (8), we know much more about what is happening in the end-organ than workers in vision know about what happens in the rods and cones of the retina.

Especially recently, a good deal of work has been done on the eighth nerve, or at the level of the eighth nerve (66, 67, 69, 106, 120, 130). The first thing most of us do when recording from single cells in the eighth nerve is to get what is called a tuning curve--that is, to find those frequencies to which that cell will respond. The tuning curves in the eighth nerve tend to be quite sharp. In fact, they are sharper than the Bekesy data would indicate (8).

GESCHWIND There is considerable evidence that similar mechanisms play a role in sharpening vision.

GOLDSTEIN Certainly; some people feel that the mechanism for the sharpening in vision is the pattern of excitation with an inhibitory surround.

HIRSH Isn't there some evidence for that?

GOLDSTEIN There is a sort of gathering body of data that this happens also in audition. The work is on monkey (67, 106), cat (120), and bat (41). The inhibiting frequency ranges tend to be on the two sides of the excitatory curve; this, I think, is quite a nice analogy to the visual system.

CHASE I wonder if I could invite some comment at this point about the two sets of data that you have introduced, extracting one generalization from the review of Hubel and Wiesel's work? You have shown some of the features of coding information in the visual system, but, more importantly, that the actual way the coding system functions is contingent upon experience.

In the case of many of the problems in speech that we have been talking about, it seems that both the receptive and the productive functions of speech represent a transition from an early plasticity and capability to the ultimate definition of an economical system, involving categorical operations. Could it be that some of the efferent componentry, some of the componentry that permits the bypassing and filtering and shaping of information at the front door of the auditory system, is another way of exercising the categorical functions, both receptive and productive, with respect to speech? When thinking about some of the work demonstrating the ability to alter the acoustic message, both at the cochlea and at higher stations--such as the extra-reticular pathways that can inhibit the response of the cochlea nucleus to click--I was reminded about some of the experiments that Broadbent has done, pertaining to how we recognize one speaker from another. I wonder whether he feels that these capabilities for selectivity in acoustic inputs are pertinent?

BROADBENT Yes, There is certainly a possible mechanism which will allow selection of one group of pathways rather than another, and, therefore, would give greater weight to information coming in by some paths rather than others.

On the other hand, this is not quite sufficient to explain some of the effects one gets. The most difficult sort of experiment to explain, I think, from this point of view, is the kind of thing where you put in two speech messages into one ear, and the same two speech messages into the other ear, but produce time delays between the two so that they are apparently localized in different places. In this case, it is possible to select one message or the other for response, and this is very much easier if you do not have the time delays in (17).

This suggests that the selection of which channel you are going to use takes place at points beyond which or at which localization takes place. Of course, it is still possible that localization takes place much lower down than people imagine, but, nevertheless, it does suggest that the selection takes place after the interaction between the two ears.

HIRSH When you say <u>selection of paths</u>, are these paths channels in an information system or are they neural paths? At which level of discourse are you speaking?

BROADBENT I'm talking at the information-theoretic level of discourse, and by a <u>channel</u>, I am simply implying that if you take any stimulus event which has a number of features, and all stimulus events that possess one feature in common which they don't share with any of the remaining stimulus events, they form a channel. Thus, for instance, all sounds arriving on one ear can be regarded as a channel, and, in that case, it is equivalent to neural pathway.

You can, however, have all sounds arriving first by males as contrasted to a female speaker, in which case, you are using, probably, the same sense organs to pass the both types of signal.

MILNER I think this might be a good moment to point out which channel, in this sense, is more effective depending upon the verbal or nonverbal nature of the material. (Milner continued with an extended discussion of experiments demonstrating that verbal and nonverbal materials are processed in different parts of the brain.)

CHASE The time for this session is approaching the end, and there were a few things we definitely did want to include before we finished. I would just like to reextend our invitation to Milner to review, in any fashion she sees fit, the general catalog of work on cortical speech areas.

HOUSE Could I make an anticipatory interruption here? (Laughter) I remember, some years ago, Ladefoged and Broadbent (81) did some experiments in which clicks and other sounds were presented along with speech. I wanted to ask whether those experiments were done monaurally or binaurally, or both?

BROADBENT They were done with loudspeakers.

HIRSH May I add a detail by way of describing a related phenomenon? These tests of Broadbent have to do with the interference with one side on the reception on the other; you don't need to go to the cortex, apparently, to demonstrate at least something about where integration from the two sides takes place. I am thinking of the observations of Matzker (96), for example, who divides the speech spectrum into two filtered bands, and finds that there are certain central nervous system lesions that do not permit the patient to integrate, say, a low-frequency spectrum in one ear with the high-frequency spectrum in the other ear.

These lesions are very high up if you are an otolaryngologist, and they are low down if you are a neurologist.

Most of them have to do with tumors of the cerebellar-pontine angle. It is curious that the functions that seem to go with the pathological material are really accidents of who's working where. Neurology departments often employ psychologists and some of these function tests get designed out of the context of psychological research. The otolaryngology departments see some of the tumors that are lower down, and, by and large, are employing audiologists, and their orientation, as far as designing of function tests is concerned, has to do with the traditional armamentarium that we discussed yesterday afternoon.

OLDFIELD I think, in this particular case of Matzger's messages being separated. there is very good evidence that this function may be able to be performed at almost all levels, as Broadbent mentioned. Whether origin in different ears is to be regarded as a lower function, I don't know, but, certainly, pitch differences as between male and female voices would. Treisman (131) has also shown in two messages which are precisely similar as regards all those properties but different only in the degree of contextual constraint within the material, the separation is still dependent on the differentiation factor; and, as between passages of different orders of approximation to English, it seems clear that the separation must have been performed at a pretty high level--at a level at which there is function in relation to grammatical sequences and continuity themes. I think it would be wrong to search for any particular level, or even two levels, at which such distinctions and discriminations are achieved.

BROADBENT I quite agree about the number of levels operative, though I do think there are differences in the type of selection at different levels. We've got some tasks which show there are functional differences between them. Now, fusion between the two ears of the high-frequency and low-frequency parts of the spectrum depends upon the time characteristics of the envelope. Therefore, if, at any point, the channel from one ear were passing through some disorder which upset these time characteristics, this would stop fusion taking place and this could happen at a very low level.

I think that this gets back to a more general point, because time-varying characteristics of the message, which do seem to be used as a common feature to allow fusion or selection of time information as being necessary, are precisely the things that do not show up in the anatomy. These, perhaps, are cases in which looking at the anatomy will not suggest useful tests.

(The remainder of the session was devoted to Milner's presentation of recent materials on the locus of dysphasic lesions.)

SESSION 5. Part 2 - Man-Machine Communication and Machine Analogs of Human Communication

COOPER One of the people I was anxious to have with us in this conference was J. C. R. Licklider, primarily because our discussion about communication between human beings has parallels in how you communicate with machines through some kind of organized code. This is an area that has fascinated Licklider for some long time, and one in which he has been working actively, so I think he can tell us something about the current state of thinking in this rapidly growing art. He has agreed, also, to serve as a discussion leader for this session.

LICKLIDER I am very sorry not to have been present at the earlier parts of the discussion. As many of you know, I have been on vacation from the laboratory, separated from real things by a kind of paper curtain, and I'm not sure that I'm going to be able to contribute to this discussion effectively. I would like to try, however, because communication between men and machines--mainly digital computers--is going to have increasing significance for the general topic of discussion here.

Until three or four years ago, there was not much interaction between people and computing machines. The use of the computer, except in a few instances such as the Sage system, was almost always a matter of writing a program, getting your cards punched, and getting them down to the computer center. There they would get into the batch processing schedule and sometime--like tomorrow or next week or, in a real good place, a few hours--you would get back a stack of printouts from the computer, and you would read those. They might tell you that you hadn't done something just right, so you could try it again the next day or the next week.

But now there are some computer systems with which people really hold conversations, and, in the process of developing those systems and the language and techniques that go with them, a real lore of man-computer interaction that has some significance for our topic here is starting to develop.

In Chase's diagram in Fig. 10 the left-most portions deal roughly with speech recognition, on the right with speech generation, and in the intermediate portions there is interpretation and, perhaps, generation from stored information. This group would be extremely good at and probably much interested in building the devices and developing the theory that would let a person literally speak to the computing machine, and, indeed, I think some people here have worked a little on that problem. This group also would be extremely good at and interested in developing the techniques and devices which would take some kind of formal string of symbols and turn it into acoustic speech, or, perhaps, handwriting or displayable signs and symbols.

Those areas are not very well developed yet in computer technology and the technology of man-computer interaction; I think this is so largely because they are very difficult, and secondarily, because the computer people assume that the group here assembled can, whenever it gets to be feasible to run the whole thing, build these parts very quickly. They will think back to some of the Haskins papers on exactly how to do this, and say, "All right, we'll build that just as soon as we have a good stream of formal language coming out of the device."

Well, I think it's in the intermediate portions of the system that most of the interesting story that we can get from the computer field lies--largely in the realm of formal language. In connection with computer programming and in connection with mathematical linguistics, there has been quite a revolution in the understanding of formal languages, and it is possible now, if you give a competent language designer a good description of what you want to do, to obtain a fairly good formal language for handling that class of problem.

This has several interesting implications. For one thing, almost everybody in this field can do anything, can handle any class of problems, and, in order to do so, starts off by saying: To work on that problem, I need a language for talking about a particular class of problems. What properties should it have? Let's get it designed and then I'll go and work on the problem.

Almost conspicuously, neurophysiologists and audiologists, and speech experts, and so on, have not in the past thought that way. I am becoming convinced that constructing the language with which to solve problems is probably the right way to go at research.

What about these languages? Do I mean FORTRAN and ALGOL? Well, no, not really. Computer programming was a problem in and of itself for a long time, and so languages arose to facilitate computer programming. Indeed, most of the man-computer interaction languages that exist now, in some sense, facilitate programming. But I like to think of the user as somebody other than the programmer. There is an interaction through programming languages, and an interaction through user-oriented languages.

Actually, to get anywhere with computers at this time, the user has to have a little programmer in his blood. But things are moving in the direction of letting the user interact with the machine, which has been, to a large extent, programmed by the programmer. It is the characteristic of user languages that is probably most interesting for us here.

COOPER Would you put the computer, in this sense, in the role of a trained laboratory technician whom the programmer has trained and has left with a program of how to clean test tubes, inoculate, and so on--to put the analogy in a biological frame?

LICKLIDER If you look into the computer, the part that is really of interest is not so much the processer or the core memory, the things you buy, but it is the computer viewed as a <u>data base</u>, plus a language and facilities for interacting with the data base.

COOPER I'm not sure I understand what you mean by data base. Is it, again to take the human counterpart, encyclopedic knowledge as distinguished from skills.

LICKLIDER Well, let's go back. Data base is jargon; information base is probably a little better. The information base is an organized repository of information in machine-processable memory. A data base that is typical of systems that are coming into existence consists of language facilities which are essentially compilers. These will not take a lot of input and translate it into machine code and then run. Rather such compilers will take this input sentence by sentence, translate into machine code, and run sentence by sentence or statement by statement. But the data base will include many different language facilities, designed for and oriented toward different purposes--giving a lot of processing algorithms and a lot of noise. One way to put it is that everything is contained in noise, including the language facilities and the algorithms.

COOPER If I can persist in my interruptions here, and in my desire to make the analogy I started with, I would put the first two operators, that is, the translater and compiler, with the sets of algorithms as skills, and the files as knowledge of the literature, let's say.

HIRSH In language analogy, this becomes lexical as opposed to structural memory.

COOPER Yes, exactly.

LICKLIDER Well, let's go directly to a point I wanted to make, which may make this a little clearer. One of the things that has been learned about the design of language facilities--the programs that implement languages--is how to get the essence of the language out of computer programs written in the classical way of just writing down a serial list of instructions followed by an instruction to do those things. That is to say, the essence of the language is crystallized out and put off by itself in the form of a list of instructions. When input to the computer is processed, however, you cannot a priori make a distinction between tabulated data and active computer program, because programmers are learning to take the essence out of the active computer program and put it into the form of tabulated data.

This is a development of considerable power, because if you let me think of a long sequence of instructions and things to operate upon as a program, and a table as being what we think of when we look into a journal and see a matrix with marginal headings, these ordinarily are taken one after another in the program, unless one of the instructions says, "Jump some place else." What I am saying is that, while you can represent almost any process you can describe succinctly in just a program, imbedding the data and everything else into it, alternatively, you can try to make this a general thing that doesn't have much specific content, but put the specific content into a table. If you do that, this becomes clearer and you start to think of a language for talking about the contents of the table.

In the case of the compilers that are being designed now, one can describe the syntax of the language for which the compiler is being developed in a formal language designed to describe language syntax, and can put a formal description of syntax into a table. A general-purpose program operating upon input strings and the syntax defined in a formal instruction described in language, therefore, can effect the syntactic part of the translation. I think this may have some significance even, say, for interpreting neurologic or neurophysiologic things about speech.

HOUSE I don't understand why the first analogy that Cooper brought up is not an adequate analogy to what you said. An alternative analogy to him might be being able to say to the computer, Go, when Go is in a code, that is, has linguistic relevance. This appears to be an alternative to saying to the computer, "Move your mouth parts and your muscles in a particular way," or, "Move your nervous system in a particular way, so that you articulate the word Go." Isn't this somewhat analogous to saying to the laboratory assistant, "Wash the beaker," rather than saying to him, "Innervate the muscles of your right arm and pick up the glass, etc."?

LICKLIDER In the sense that if it is suitably programmed you can give the computer a very terse, crystallized stimulus and get out an organized complex response, certainly, that is true.

Let me go to a different point of the discussion, and say that one of the aims is to make a large computer system, particularly a very large memory system, available to several or many people simultaneously, on the ground that human interaction and teamwork can be greatly facilitated by working through the superior communication and coordination facilities of the machine. If we think about the data base for such a computer system, it will make a good correspondence to a technician, but it will have parts specialized for various technical activities. There will be a part that does the mathematical processing; it will have a symbolic integration program, and it will be able to do a lot of the different things that you ask a graduate student in mathematics to help you with, if you need some help. But, equally well, it will be able to parse various languages. You might say, well, it is also a linguistic technician. It will be able to facilitate your efforts to simulate processes. I don't know whether there is a technician who can do that, but the computer will have a number of different capabilities.

What I think is the most important thing for us today is to get at some of the concepts that are shaping up, such as this one about tabulating the essence of a program, and then devising or having devised a language for talking about the contents of the table, and then tabulating the essence of the program.

COOPER What is in the table? I can imagine, in old-fashioned types of programming, that the table might contain subroutines, for one thing, and that it might also contain numerical data, or at least the format into which numerical data could be put. But I suspect you have something else in mind here.

LICKLIDER Yes, I have something else in mind. Let's come back to the syntax of a language, where we are trying to make a compiler to translate into machine code statements written in a formal language such as ALGOL. One can think of each individual situation that may arise in working with the language, and try to make a subroutine to handle it. Early in the development of compilers, that is approximately what was done.

With the developments in mathematical linguistics, one can now write out a series of what are called constructions. A simple example of this would be that if we have, say, an auxiliary verb followed by a main verb, that sequence is equivalent to a verb. In other words, one can define all of the ways one can put together subcategories of the syntax to make larger categories. Sometimes, you don't even go up one in the echelon of categories. Frequently, there will be several different constructions, so that the materials called back in normal form are accompanied by a symbol which means <u>or</u> followed by various alternatives. A program, operating upon such a description and an input string, can handle the problem, that is, it can handle all of the syntactic aspects of the interpretation just as well as if it were all built into a dynamic program form.

I think that when you move to a description in a formal language, of another language, you at least have the opportunity to see its structure revealed in a different way from the way you are used to. I have a feeling that several languages at the first level--say, all the different natural languages--may look enough as we described in a formal language that it is then possible to operate on that formal language to good effect, and one really needs a language for doing that.

I was much impressed by the neurological discussion here this morning. I have the feeling that the way we have profit from the work in other parts of the technology is just to get a lot of these little pictures in our minds, and then things will be more meaningful when we come across a particular result of the neurophysiological sort. It will trigger off a line of thought that might not otherwise come.

HOUSE When you say <u>formal</u> language and <u>natural</u> language, do you mean that so-called formal languages have some of the attributes of what we call natural language, and that that is why they are formal languages, or is it simpler than that?

LICKLIDER Well, by and large, formal languages are designed by man and are extremely ruly, and there aren't any parts in them that the designer doesn't understand, to begin with. He may see something in a different way later on, a way different from what he thought of when he made it. But, in the natural languages, there are many parts that we don't understand at all, really. We discover things about them. They are open languages, in two senses: the user can keep creating new parts, and, they are a little like a large geographical area that is not thoroughly explored.

HOUSE Is there an implication here that when you try to make a formal language more efficient or useful you modify it in the direction of natural language? Is there something about natural languages that constitutes a model for the writers of formal languages, or are these two things totally unrelated? Is this greater efficiency that you are finding in these operations that you have described to us a result of better understanding of natural language?

LICKLIDER One of the lines of development in user-oriented computer languages is toward natural language, but this is only one of the lines. There are many people who think that, for any given set of problems it is possible to design a more-effective language different from the existing natural ones. But people, or users, start off knowing at least one natural language, and many of them--military commanders in particular--say, "What we want to be able to do is to control that machine in (ordinary) language," so there is a considerable effort to develop the facilities that will let a man, if not talk, at least write to the computer in essentially unconstrained language. I think there are some interesting developments along that line, and those are probably the ones that are most interesting for this group, rather than the formalization.

HOUSE The real intent of my question, however, is not on that level. I am more interested in the feelings you have about the formalization. Is there anything in the structure of a formal language--not its manifestation in an acoustic or other physical sense, but in the structure of a language--that approaches natural language, as a desirable attribute?

LICKLIDER Yes. In the study of natural language--in mathematical linguistics--the syntactic aspects of language have been studied extensively to the point where most of the people in the field are now thinking that we have to get to the problem of handling the semantics. There is really not much further to go in the study of syntax.

In between the syntactical-grammatical aspects and the most clearly semantic aspect are what are called data structures, and we do, in natural language, make very extensive use of data structures. If you ask a technician to print out a table of the presidents of the United States and their terms in office, you are saying, table, which tells him what kind of table structure to pick--essentially, a matrix with certain marginal headings. Some of these data structures are known to most of us. A boss and his secretary pretty soon get over the hurdle of how to see by inflection of voice, almost, and indicate which data structure he means when it is ambiguous. Well, the interaction of a word like president with a word like table defines and clarifies a fantastic lot of processing, and, also, terms in office definitely implies a pair of dates connected by a hyphen or something of the sort.

It is possible to build into the language an appreciation for such data structures, so that in the input string you don't have to say any more than that you want the table of presidents with terms of office. The language then will have in it a set of priorities about which substructure to use for president, how to put the things together, where to put the headings, how to adjudicate problems that arise if the list is too long for the piece of paper it is going to be written on, and so on. In computer languages, therefore, there is a great deal being done at present in an effort to match the data-structure implications of words to their syntactic categories.

In my example, I did not indicate anything about the interaction between these. President and terms in office are both clearly substantive. But if you have a thing that can be either a noun or a verb, such as table, its implications for data structuring are entirely different in the case where it is a noun and you are going to "make a table," from what they would be if it were a verb interpreted in the sense of "don't continue this discussion now." If you can add to the syntactical analysis the machinery for working in the built-in knowledge about data structuring, then, you get rid of an awful lot of the detailed programming of the machine.

This is intermediate between clear syntactics and clear semantics. It is very hard to tell whether these data-structure properties are syntactic properties or are not. Whenever we deal with numbers in a class on grammar, we have a very hard time saying what parts of speech pure numerics are, and, in the world of computing, of course, you have many different kinds of numerics.

I guess I want to say that there is under way a big development of this category of things that was a little bit in limbo before, but now considers the semantic significances of words. The computing field has not gotten terribly far with this, but it has done a little. In the first place, it is pretty clear that verbs have something to do with subroutines, and that nouns have something to do with entries in a data base, and the function words help; that is, subroutines require arguments. A subroutine is a thing that operates on arguments to produce a function of the arguments, and so a subroutine needs arguments, and these arguments have to have particular data-structure characteristics.

The arguments have to have special characteristics, many times, besides data-structure ones; for instance, a subroutine might need an argument that can be an active agent with volition of its own. Whenever you deal with people, you come across these. So there is starting to be a little understanding of how to take advantage of particular constraints on the meanings of individual words, so that the machine does not grind out nonsense sentences but makes appropriate, possibly meaningful, sentences.

In all of this, the thing that seems significant for the neurologic part of the discussion is the way in which the computer people are finding it possible to have a program that operates on several different tables, and pieces together the component contributions that deal with syntax, data structure, and semantic meaning. I think this bears on the problem that has been with us for a long time, that is, how is it that when we generate a flow of speech, somehow, the ideas are coming up and determining a lot of what is said, but there is also a grammatical machine there, trying to get the words put together in the right way.

FRY I think the implication here is not only to the neurologic organization but also to the linguistic organization, because this was your original question. In the natural languages, as in these formal languages, you have units which both operate as parts of the formal structure and which have meaning, and you have to manipulate these units from both the semantic and the syntactic points of view, so that the output makes sense.

LICKLIDER Is pragmatics still acceptable as a third branch of this general subject? When I went to school there were syntactics, semantics, and pragmatics, and the whole thing was called semeiotic. I understand this went through a period of unacceptability, and you're pretty old-hat if you mention pragmatics. But in this computer technology, the pragmatics are getting quite a lot of attention, but, somehow, not in the same group of people as worry about the language problems.

In the simulation of cognitive processes and artificial intelligence, there is much writing of programs in which there are payoff functions or value matrices, in which decisions are made in such a way as to maximize some kind of utility--the utility of the system of programs, the utility put there, of course, by the program. But it seems to me that it would be a reasonable thing to bring the lower of the computer decision processes into the language picture.

Of course, people aren't very much interested in getting computers to do things to maximize the utility for the computers; they want their own utility maximized. But you could get a system of programs that would not only speak grammatically and make sense in a semantic way, but also say things carefully calculated to maximize the utility function of the program system.

COOPER I was interested in one other aspect of what you said, namely that your military commander wants to be able to talk to this computer as if it was another human being. Then he can say a thing one way or say it in quite a different way--in terms of the words he uses--and still have the same operation performed.

This runs very much counter to the way the machine itself operates when you dig all the way down to the transistors, because the machine has a fixed, rigid pattern of going about things. This has seemed to a lot of people to be one of the great differences in hardware organization between computers and brains; that is, there seems to be a wide gulf between the highly precise processing in the computer and the parallel, but highly imprecise, processing we assume in the brain. You want to get this imprecision into the process somewhere ahead of the hardware unit, if I understand you correctly.

LICKLIDER I'm not sure that it is imprecision. It is that there are large and rather subtle equivalence classes. But I don't think much is to be gained by making the thing imprecise.

OLDFIELD Do we really suppose the brain is imprecise at the same place that the computer is imprecise? I would have thought we would have supposed, in the last resort, these logical operations have to be performed in a certain sort of coding, and all the equivalents have to be reduced to this.

COOPER I was thinking of imprecise in the sense of some of Lashley's very early experiments. If you go in with a scalpel and make a series of slices through the computer, it doesn't work nearly as well as the brain does if you do the same thing to the cortex. (Laughter)

GESCHWIND I would tend to agree with Oldfield on this. A redundant system is not necessarily an imprecise one, so that if you have enough redundancy in this system, you are going to be able to go in and shoot a couple of holes in it and still have it function well. But this does not mean imprecision; it just means you can damage the machine and still have it work. Also, it depends where you put the small amount of damage in the brain. There are many areas where a small amount of damage is absolutely overwhelming.

LICKLIDER Let me mention something about computer design. There is a trend now to have multiprocessers, which is to say, several or many processing units, perhaps a little simpler than some of the fanciest that have been made. There is a trend to break the primary memory up into blocks which are independently addressable, and to have a lot of input-output channels which can run essentially simultaneously, and have a component that you might call an arbitrating circuit, which makes assignments to each incoming string of characters, or makes an assignment that such-and-such a processer will work for you in this step. Say that you want access to memories 37 and 438; this processer will get you hooked up with those. Your little step of computing will be done with those facilities. But, the next time you need something done, with the train of thought that is coming in from some user, you may have entirely different hardware working for you.

The people who build military computers are interested, of course, in the same level of operation, local damage notwithstanding, so they say, "Let's check the performance of these units all the time," and any time a unit isn't working well, it gets a little flag put up that says, "Don't use me." This brings you right back to the Lashley kind of situation.

OLDFIELD I suppose there might be situations in which a quick way to do something with a computer might depend on large storage capacity. If the computer knows it has got its storage capacity damaged, it has to do this the long way round, and this could be used as an instruction to it to always do this. This is the sort of thing, I take it, that happens to the brain.

LICKLIDER There is much redundancy in the use of memory already. In that project, working with multiple-access computers at MIT, I think, they are still dumping the disc file about every four hours. But if anybody gets into enough trouble, he can always go back and pick up from where he was.

At the Livermore Laboratory, they are aiming at a fantastically big memory, precisely for keeping track of various stages of calculations, but, in the interim, they put this out on a line printer, which runs so fast that they don't count pages, but they measure the output in edge feet. (Laughter) Literally, the paper comes out of that thing just as fast as paper can flow. Nobody could conceivably read it all, but it's there in case you want to go back and pick up your calculation at some particular point or in case of error.

COOPER But isn't a lot of the necessity for that due to the fact that if the calculation has gone wrong because some small loop in the program wasn't quite right, or something of the sort, the whole thing is wrong? In other words, it's the catastrophic effect of minor damage to the machinery which sets the computer apart from the human.

OLDFIELD As these things become more complicated, and these possibilities of alternate modes of function and reduction of higher-order conceptions within the machine occur, will not the mistakes it makes, when it does make mistakes, more nearly approximate the sort of mistakes that are made by human beings with speech disorders? Instead of making completely ridiculous statements, one might suppose that it made statements that were incorrect in a grammatical sense or in a syntactical sense.

I also wonder about what the results are when there is damage. Also, as the program progressively refines the system of programming in such a way as to make--in a sense, what Licklider has been talking about, the whole thing more efficient--one might compare this to a child learning language, and ask whether it goes through phases in which you don't like its output because, although it is fairly reasonable, it has got inelegancies in it?

LICKLIDER Well, I think there is very much in what you suggest. It is now a fairly common experience to see a computer print-out that almost makes sense.

OLDFIELD Instead of being so ridiculous that you can tell straightaway that something is wrong?

LICKLIDER It looks a little like the word hash from an aphasic. I think, as the language mechanisms get to be hierarchical--well, as you suggest, an error now is not going to ruin everything, but will just ruin one aspect of the program--it might be, say, that the function words don't get inserted properly.

OLDFIELD It might make a figure mistake, I suppose, or a vague statement or something of that kind.

GESCHWIND I suppose that Cooper was trying to get at the problem of reliable systems that von Neumann (105) discussed. For the sake of reliability you may abandon the attempt to make the system error-free. Instead you can deliberately design the system in such a manner that it will, in fact, make more errors, but they will not be critical. Perhaps, humans are efficient in the sense that they make more errors than machines, but, as Oldfield pointed out, they are not such terrible errors. This is a problem of designing your system in such a way that it makes some errors, but keeps them small, instead of trying for perfection, in which case you may get larger errors. I think this is one of the problems that von Neumann addressed himself to although I don't think he gave a very satisfactory solution to the problem.

COOPER As I recall, von Neumann simply used more components in parallel in order to get reliable operation of the ensemble.

GESCHWIND That's right, but I suspect it is a highly inefficient means of achieving reliability.

BROADBENT It need not be an inefficient means. Let me call attention to the Winograd and Cowan (136) monograph on building reliable computers with unreliable components, because they have taken up this problem and extended an analogy from error-correcting codes to the case of taking a simultaneous set of elements at a sense organ or an input, and translating them into a further set of elements later on.

They go into the sort of way in which you couple the various input elements to your various later stages, in order to produce something like an error-correcting code which would do reliable computations. They get an analogy to the Shannon capacity theorems for computers, which works out more efficiently than even von Neumann's system. It has one or two interesting properties from the point of view of analogy with the brain, in that, as an error-correcting code, the best way of doing it is to make every symbol in the code dependent to some extent on almost every one in the input matrix, so there is a lot of lateral cross-connection, as there is in the senses. Also, you have to build your modules with an equivalent of presynaptic inhibition in them. McCulloch (98) has applied these ideas to communication disorders.

DENES Is this error-correcting of the language or of the data?

BROADBENT This is correcting errors in the transmission system, not errors in the original input.

DENES I wonder whether there are any compiler programs now available which are instructed to go on and make the best of imprecisely designed input statements?

LICKLIDER This whole discussion operates on many different levels--there is as much work as indicated here in trying to make computers reliable through redundancy at the component level. Some of the most interesting things for our purposes, though, come from the design of an interactive man-machine language, which leads you to a kind of conversational way out when an impasse arises.

One of the nicest of these, I think, is at The Rand Corporation, a language called JOSS. This is a very small thing, on a very old and honorable computing machine, but eight people can use it at once from typewriters in their office. When they do something against the rules, it tells them what they have done wrong, and it is very forgiving. In fact, it may type over some of their stuff for them and get them on the right place again. If you try to divide by zero, for example, you are simply reminded that you're trying to divide by zero, which is not a good thing to do.

There are, in this system and a couple of others, reminders that "I am assuming so-and-so, that you mean so-and-so." This, in compilers, is fairly standard. The good ones don't stop when you make a mistake. But still better compilers are incremental or differential ones, which work with you as you write the program and say, "No, it's better to do it this way." As you know, it's better to catch these errors when they arise rather than permeating the whole program with them, because the chances of anybody making the right assumptions fifty times in a row are very low. The secret of all this, therefore, seems to be to write your program with the machine, let the compiler respond to you every time you write something, and then you can see immediately and can practice on real data, simple data, so you can see what is shaping up.

GESCHWIND I think Broadbent's comment on the monograph of Winograd and Cowen is interesting, and I was trying to think of an analog to the technique of the authors in sense data. If each frequency band were used as a phoneme, then, if that frequency band were knocked out, that phoneme would be lost forever. But the fact is, I gather, that most of the phonemes use nearly all of the available frequency bands and, as a result, no one frequency band carries all of the load of a single phoneme.

BROADBENT Yes, I think this is part of it. Of course, there is the additional part, really, that you are not simply repeating the same signal in every frequency band, but, rather, the signal is the combination of what is happening in all frequency bands. But this does make the thing very resistant to removing one particular part.

COOPER But wouldn't you say that this corresponds more nearly to the paralleling components rather than to an error-correcting code which operates over an interval?

BROADBENT I was jumping from one to the other, because the thing which excited me in the monograph, that I hadn't realized, was the analogy between simultaneous presentation in a number of different frequency bands, and the successive coding of a long binary digit, which one is more used to think of as error-correcting.

CHASE I would like to ask a question about another kind of analogy that Licklider drew between machine language capabilities. Speaking of the machine case, as I understand it, a set of processing capabilities was outlined with respect to a hierarchy of programs, and one hierarchy that you specified was a program of syntax, a program for data structure, and a program for semantic processing.

Of course, there are parallels in the case of the human, and I think that most of us think about the immature nervous system as representing the componentry and the capabilities for handling many programs and that specific programs come to be written in all of these classes. But one of the problems that eludes us is specification of the inputs that are critical for shaping specific programs in these categories.

Here, I think, the machine case might afford some unique possibilities for suggesting directions for work in the human case. I wondered what generalizations Licklider could make about the kinds of information you have to feed into the machine to permit it to perform its syntactic, its data structure, and its semantic processing operations, and to what extent the corresponding classes of inputs have unique or overlapping features?

LICKLIDER These are the inputs of data to the store which would be used during translation, say?

CHASE Yes. I'm really talking about the education of the machine.

LICKLIDER Obviously, I can't say very much here. There seem to be two approaches: to crystallize the material, and then feed it into the machine; and alternatively, to give the machine some heuristic guidelines about how to profit from experience, and then feed it lots of texts and keep telling it when it does well and when it does not. There has been much more progress in the first of those two ways than in the latter, but it might be that the latter is more powerful in the long run.

If I may make a comment about one thing slightly orthogonal to that--when one tries to represent symbolic information for economic storage in a computer, one sees that the characters or the codes for characters may not be just what you want. A lot of the drive to get a better representation in digital computers--in a fixed word-length machine--stems from something that may be wholly irrelevant to the nervous system, which probably does not have anything like a fixed word length of 48 bits. But there are some conveniences, anyway, to getting a representation of a string of characters, which is not just the concatenation of their codes.

There is a development, for example, called hash coding. It is possible, with the aid of an algorithm which puts the first character in and then the second and then the third, to calculate a code for any sequence of characters. If you work hard enough on designing this algorithm, you can minimize the probability of having two input strings generate the same representational code, or you can minimize that almost as far as is logically possible. Working with these things has a greatly facilitating effect on programming.

You can't have thought about the nervous system and its not-very-transparent way of representing things, and also hash codes, without getting a feeling that what the nervous system is doing is something very much like hash coding. I have a feeling that this little analogy can stand a lot of scrutiny by neurophysiologists. Indeed, people apparently make mistakes for only the reason of this collision of the representations of actually disparate things.

You see, also, we take in information so much more facilely than we put it out. If you work with hash codes, you have this quick way of calculating representation. We also seem to be able to think of something about a complicated thing, like a very familiar phrase or sentence, or if we memorize a poem, somehow, we can think of that whole poem without going through all the intermediate learning stages, or all the components. This gets a little irregular, but, in hash coding, one typically makes hash codes corresponding to words and then, calculating from those hash codes, makes a hash code for a clause or a sentence, and then, working with sentence hash codes, makes a hash code for, say, a paragraph. Finally, one has one representation of, perhaps, 36 bits, that stands for this great, big, long string of stuff.

COOPER Actually hash or jargon?

LICKLIDER These are called hash, because there is no resemblance to the original, and, yet, the whole thing is eminently digestible.

COOPER Digestible by the computer, I suppose?

LICKLIDER When you try to make output, suppose you apply the hash code for a paragraph to the output mechanism; it cannot just calculate back that great, big, long paragraph from this hash code, but it has to go to tables and look up things. The table may occupy a lot of memory, but it can look up this hash code and say, "Oh, yes, this is for a sentence, and the words in it are the words with such-and-such hash codes." Then going back to each of those, it will say, "Oh, yes, these are for words and the characters for such words are such-and-such, so let's put it out."

It is rather eerie to see a page come out of a computer typewriter when all you put in is 36 bits, but it is perfectly possible, and it has many characteristics that are reminiscent of language behavior.

HOUSE It is very interesting to me that, without being here for the past two days, you have been able to summarize the discussion. (Laughter) We have been saying that there are a number of levels of description of the activities in which we are interested, and they all reduce, we think, one to the other, in different codes, even at a very mundane level. (As I am talking, my code is being reduced by the stenotypist.) Some of these codes are transformable back and forth, and with some facility, but usually we lose some of the description when we do this.

We have already talked about being able to go from an articulatory description to an acoustic description, but not being able to go back, necessarily, unless we know a very good set of rules; similarly, if the stenotypist has a good set of rules, she will be able to come back into natural language and, if she forgets these rules, it's just hash.

It seems as if all the operations we are talking about are very much like this, and it suggests that the kinds of things that have been described about computer development are really logical extensions of what we know about natural behavior. In a sense, the people who are working on computers are trying more and more to make the computers do things the way people do them. This, at least to me, seems to be a reasonable interpretation of what Licklider has been saying. I don't think he has said yet that the computer methods, in essence or in principle, that have been developed are sufficiently powerful today to cast more light on what we are doing, except in the sense that we have to examine what we are doing in order to develop the computer method.

LICKLIDER I didn't say it, but I disagree that there isn't something for us to learn from all of this. What the computer technology gives us is, first, a language--richer and much more precise than any we have had before--for describing things in which we are interested, and, second, a way of making dynamic interrelations of the complexity of the things we work with.

The true value in all the computer technology for people like us, I think, is what now is being called dynamic modeling, with which we may express a theory or hunch or hypothesis or model in computer program form, and then have a very compact representation. But even more important, the representation can then run or unfold before your eyes, literally talk to you, draw a graph for you or whatever you like--and if there is something wrong in your formulation it will surely appear, because it is a very difficult thing to make this thing run without having formulated it correctly.

DENES But would you use sophisticated compilers for building your models or for getting the model to communicate its output to you, the experimenter?

LICKLIDER Well, most of the people now, I think, who make simulations, use one or another of these languages in a general-purpose simulation system. It is the difference between writing simulation in such a language and working it out in machine code, or like having the results tomorrow and having the results three years from now.

DENES Some special compilers may be more efficient than the simpler ones, but you have to learn to use them--is the amount of learning worth the increased efficiency? In scientific work, each problem is very different, and you would have to learn as many different compilers as you have problems to deal with, whereas, if you use the less sophisticated language, you just become very fluent in somewhat less efficient language.

SESSION 6. A Summary and Its Discussion

COOPER Several people have carried a large share of the responsibility for the conference thus far. One other person who accepted a major assignment was Hirsh. He will try to bring together, in whatever way he sees fit, some kind of summary of where we have been and, quite possibly, where we ought to have been, or might want to go another time.

HIRSH Let me mention one of the chief reasons why it is difficult to summarize, and that has to do not with language but with languages. I think, perhaps, I should get this off my chest at the very first.

We have dealt primarily with five languages or five levels of discourse--the acoustical, the neurologic, the acticulatory, the psychologic or behavioral, and various model languages. Of these--and I suppose this is because I am not an acoustician--the acoustical seems most clear and most unambiguous.

The neurologic offers us a difficulty, because results that are pertinent to the problems of speech, language, and hearing appear to be mostly neuroanatomical; that is to say, one speaks about places or regions, whereas the data in other realms of discourse, in acoustics or in hearing, for example, have more to do with process. Therefore, it would seem that the most logical correlate in the neurologic realm should be process. This should be neurophysiology, and, as you all know, most of the neurophysiologic data that we have available come from animals, where many of the phenomena that we have been talking about, at least by consent of this group, do not exist.

The articulatory language presents still other difficulties. In one sense, it is like the introspective private language of an older psychologist. Every phonetician knows what he means by an articulatory description of a sound, and, thus, at a linguistic level, this seems satisfactory. But people outside of phonetics and linguistics get the idea that articulatory means muscular, and I believe it is true--at least it has been evident here--that even the best phoneticians cannot always transform what is meant by an articulatory description into a muscular one, whether that muscular one be electrical potentials from motoneurons or motor endplates or even movements and forces of muscle groups.

The model languages offer still another problem. The most frequent kind of model language that has been employed here has some relation to one or another aspect of engineering and, in most cases, such model languages offer advantages--often heuristic, as Chase pointed out this morning--of at least clarifying issues. In other cases, the models appear to do nothing more than transcribe unsatisfactory terms into new terms that are similarly unsatisfactory but have a better sound. Thus it is that such terms as control signals, I believe, are a 1964 translation of intentions.

For the summary itself, I have chosen to ignore the order in which we took up topics, but I have tried to organize the summary around the topics as indicated in the plan of the conference. I would like to begin with the process of speech production.

We were treated to one concept of the speech-production system as a system that has an output of phonemes and an input of control instructions. It was claimed that the anatomical bases of many of the component parts of the speech-production system are reasonably well known, but that the sequential program that causes these parts to interact with each other over time is much less well understood.

The concept of feedback was a very important topic of discussion, particularly with regard to speech production. It became quite clear that, depending upon the level of complexity of the speech response that was under discussion, one had a difficult time settling on how many and how extensive had to be the feedback loops that would be involved.

A rather different view of speech production, which will come up again in the discussion of speech perception, has to do with the nature of the control information that is stored, either for the purpose of receiving speech or producing speech. Until recently, most of the storage has been considered in terms of items that were like the items that would become the units of whatever level of discourse had preceded the discussion of speech production; for example, the phoneme is one such unit.

Different from this is the conception presented to us yesterday, that what was stored was not so much patterns that corresponded to individual phonemes, as rules that would be used either to generate phonemes in the output of this system or, indeed--as came up in another context--in receiving speech-sound information and converting that information into these same units of speech.

The levels of description of speech events appear to be reasonably well laid out and separable, but are not always coincident one with another. Thus, the phoneticians' articulatory description of individual phonemes does not accord directly with the muscle potentials or the muscle gestures that comprise a speech event, nor do they accord always with the acoustical result of these gestures. And so, we were in some difficulty about which of these levels of description of the speech event was the most appropriate to use as the physical specification of a speech sound.

We were told that the pressure waveform in itself is inadequate in the sense that it contains too much information, that is, more information than is required for identification within the speech system, but that the spectrogram of a speech sound probably corresponded more to the information that was processed by the auditory system, at least containing that information that was required to identify phonemes. The myogram appears to be not complete enough, but in certain nonlinguistic features of speech, the prosodic features and particularly stress, the myogram appears to be more relevant than some of the acoustical characteristics.

Some of our specific treatments of the acoustical level of speech brought us to some detailed information, particularly about the acoustical features of the source, shifts in conceptualization of the mechanism involved in producing the glottal periodic tone, and the influence of cavities on this glottal tone. It seemed clear, at least at the acoustical level, that the description of a speech sound in terms of the formants was complete, but only for certain classes of speech sounds, in particular the vowels. This conclusion was offered a hundred years ago by Helmholtz (52).

As we turned our attention backwards in time to the process of speech perception, we found that we were still troubled by a decision as to how to describe the stimulus for speech perception. One contention was that the stimulus to speech perception was sound, but it became clear as we analyzed the auditory receiving system that if the stimulus was, indeed, sound, then one of the very first steps in auditory processing was a conversion from a continuous signal space into a discrete, categorical signal space. Signals, at least at the next level in the auditory system, were categorical in nature and the process would have to involve both the information that was in the sound and information that was stored. Some of us were reluctant to accept the definition of the stimulus to speech perception as sound particularly because those acoustical dimensions that had been used traditionally to describe sounds, and even the psychoacoustics that has been built up on those dimensions, appear often to be irrelevant to some of the phenomena that we know about in speech recognition.

It was also suggested that the definition of the stimulus for speech perception might depend upon the kind of response that was called for, and that if an acoustical phonetician wanted to test for auditory discrimination along certain acoustic dimensions suggested to him in his own work as being important, the definition of stimulus in that case might involve a different kind of specification from one where speech identification was required.

The role of feedback returned to this discussion of speech perception, and I think that it was a new idea to some of us to find a feedback loop, much like the one that one often finds in the speech-producing system, in the perceiving system. The role of the feedback loop here is to correct the analysis that is being performed by the auditory system on incoming acoustic information. In this sense, then, the auditory processing of sound information is continuously corrected by stored rules in one case, or stored patterns in another.

We spent only a little time on the perception of nonlinguistic features in the speech stimulus, and I sensed that the general conclusion was that even though our knowledge concerning identification of speech was inadequate, we knew even less about the recognition or identification of talkers, and about the identification of affective state of the talker, and so on.

Following these early discussions, we talked about both speech production and perception within the context of a linguistic frame of reference. In connection with speech production, for example, it seemed reasonable to consider how production was modified by self-perception of the lexical constraints on this production, and of grammatical, longer-time constraints, and also, short-term effects covered under monitoring or delayed feedback.

We took a side branch for a time and considered some of the interesting phenomena that indicate that delayed feedback, as a possible source of distortion on a motor output, is not restricted to the speech case, but can be demonstrated in other motor modalities. Indeed, it seemed that some of the features of the disturbance that was imposed by a delay in auditory, tactual, or visual feedback resembled those disturbances that we are more familiar with in the case of speech. The group was not agreed and, in fact, was a bit up in the air, concerning whether or not this was a general phenomenon or whether, even though there was motor disturbance brought about by delayed feedback and other modalities, there was something rather special about speech. The suggestion, for example, that the amount of disturbance brought about by delayed feedback can be manipulated by varying the order of approximation to English is one important point in separating off the speech disturbance from a more general motor disturbance.

In the case of the normal talker and listener, we gave some attention to what it is that gets stored and how it is that this memory store--whether of rules or of patterns--gets built up. Particularly, questions were directed to the necessity of an auditory input or the possible substitution of a tactual or kinesthetic input with consequent storage in--what shall we call them--tactual and kinesthetic images, if you like? There seemed to be some sentiment that here, again, an auditory input was rather special and rather necessary, and, as will come up again when we talk about some of the neurologic matters, it was this auditory input and its possible connection to other sensory and motor systems that seemed to separate off the human brain from the brains of lower animals.

At least one of us tried to look at the input and output mechanisms in a single coherent system, a system that involved a receiver that takes sounds from the outside or from itself, and segments them in time and quantizes them in terms of phonemic or other linguistic units. This early categorization into bins appeared to be necessary because it is only at the linguistic level that one finds a kind of invariant relation between stimulus and response, no matter what the level of discourse.

When we came to talk about the development of language and language skills, our leader called for a discussion in four areas, but only one of them, I think, was responded to adequately and perhaps this suggests that some more discussion is required. We did have a review of the schedule of development of certain kinds of speech and language behaviors--an early schedule, I should say--and we found that as far as anatomical development was concerned, at the neuromuscular level, there was rather a void. In neuroanatomical terms, however, there were things to say, particularly if we emphasized the dependence of language on cortical connections.

Our discussion of neural mechanisms with regard to language and language development could not proceed without consideration of pathologic material. It seems, if speech and language are, indeed, a peculiarly human phenomenon, and if one is interested in the relevance of various neural structures to this phenomenon, then definite information is available only in the form of pathologic material.

The exception, of course, is basic gross anatomy.

This consideration of pathologic material, I will put aside for just a moment and bring it back into the context of the discussion of disorders of speech production and speech perception. Here, we found that an attempt to classify these disorders is based almost entirely in history as opposed to logic. It has accommodated various clinical problems as they have come along and as they have been identified separately. Language disorders, in general, have not been clearly classified in terms of symptom--I suspect because they have so often been classified by people who were looking at the nervous system. Thus we find that the language disturbances--more, perhaps, than peripheral speech disturbances or hearing disturbances--are classified with reference to missing structure rather than symptomatic description.

In discussing the implications of some of these disorders for localization of function, we ran across a few both new and old concepts. I think that most of us are used to and found again in our discussions this morning and yesterday the idea that the more complex a phenomenon, the less likely it is to be localized in a particular place. This, indeed, carries us back to some of the notions put forward by Lashley about the relation between rather specific functions and those involving complex associations for learning. But as we listened to some of the reports on pathology associated with language disturbances, we began to find that there are functions that most of us would have called very complex--not just the articulation of a speech sound, but, rather, disturbances in the sequencing of speech sounds, differentiation between adequacy of ordinary verbal responses and the likelihood that one is going to get spontaneous speech--associated with lesions in rather particular parts of the brain. To my way of thinking, this kind of evidence appears to be exceptional to this more or less traditional view that the more complex the behavior under study, the less likely is one to find a definite localization.

Let me suggest conclusions or, if you like, points of agreement--and also a couple of problem areas that seem to me still wide open. We seem, first, to be pulling away from global statements about speech sounds in general, and we appear, rather, to be making two kinds of generalizations: one about speech information that is spectral, and another about speech information that has to do with time-varying characteristics. It was suggested that this dichotomy, perhaps, also corresponded to phoneticians' distinctions between manner of production and place of articulation, but this suggestion was not accepted widely, nor was it permitted to go too far.

There is another general conclusion about the nature of speech production as an example of a more general kind of complicated motor behavior. The conclusion appears to me that it is that, plus something more--that is, speech behavior appears to share with other kinds of complicated motor behavior certain general properties, but these properties, even when completely described, will not fully describe speech behavior. The linguistic code, for example, seems to bring in another dimension of control that requires further kinds of explanation. In other words, speech reception is not just a special case of auditory perception, since even if one could write all the laws about auditory perception, until the rules imposed by a language code were taken into account, such an auditory-perceptual theory could not satisfactorily explain the perception of speech.

There is another conclusion that I should, and will, make the first problem. There seems to be a suggestion in this dichotomy between time-variant and spectral characteristics of speech signals that the former will probably be the least identifiable anatomically, involving as they do, important sequencing, and, along with this, a difficulty that they appear to contain more information for intelligibility.

Another very important problem has to do with the notion that is now appearing to be current, that the categories imposed by the linguistic code do themselves modify modes of production and perception. An important problem here, in connection, for example, with the motor theory of perception, is whether or not this is a special case imposed by a linguistic code or whether it is, in fact, a special case that serves as an example of the more general phenomenon where discriminability and mediation of complex phenomena can be aided by labels. That is, is there anything within the linguistic system itself that makes it special, or does, for example, the phoneme category merely provide a label that sharpens discrimination?

Finally, there seems to be a necessity, before we can go much further, to characterize the development of speech, language, and the perception of speech at several levels of discourse with respect to several orders of function. I mean we not only need to know, at a behavioral as well as an anatomical level, something about the development of speech and language, but we also need to be able to categorize that information with data obtained at successively more-and-more complex levels of behavior. We have some information on phonemes, on the first couplets and triplets, and we have some information from the later school years about the development of rather sophisticated language skills, but the transition between these two kinds of information is not at all clear, and I suspect that the data are not now available.

I am sure that your discussion will add to those parts of this summary that have misrepresented, or omitted, important parts of our last two-and-a-half days.

COOPER Thank you, Hirsh. Your outline came up to my expectations; it was a good summary. You have raised some problems and, no doubt, inspired some objections. The floor is open for any comments.

LIBERMAN I want to say that I thought Hirsh put one question very well indeed, when he suggested that there are two possibilities in regard to the categorization that we observe in speech. One is that it is a consequence of certain special aspects of the linguistic code--I would say articulation--and the other is that it is more generally a consequence of a time-variant act. I would bet on the former rather than the latter, but it is a very reasonable question.

I am not absolutely certain that I understand exactly what is meant by a distinction between spectral characteristics and the time variants. Is this, for example, the difference between vowels and any of the consonants?

HIRSH There are spectral differences that are important for distinguishing certain consonants, such as fist from fished, and I would include these with spectral differences. In general, by time-variant characteristics, I mean those where the phoneme identification depends upon distinction between a rapid as opposed to a slow rise in amplitude of something--where the recognition is dependent upon the rate at which the transition is made, and so on.

LIBERMAN I would like to suggest, then, that I think this distinction is very highly correlated with another distinction which I tried to make earlier. It goes something like this: there are some acoustic cues and some speech sounds that are perceived very differently from those most nearly equivalent nonspeech acoustic counterparts, and there are others that are not perceived differently. I would suggest, therefore, as something for investigation, that we try to get more information about this. When I look at the information we now have, I think, it fits rather nicely into these two categories.

HIRSH There are two aspects of prosodic features. One is that we didn't know very much about them, and the other is that, for them, the muscle information appeared to be more relevant than, for example, phonemic distinction, or the kind of things Ladefoged reviewed for us.

LADEFOGED You are using prosodic in a rather special sense. I am tempted to quibble a little bit, perhaps, thinking of perception of vowels. You plainly want to have information about what the other vowels are before you can perceive a given vowel. In that sense, you have to take account of the stream of time.

CHASE I would like to question two generalizations that were made--less out of a conviction that they are wrong than out of a conviction that I would like to leave these issues somewhat more open than they have been considered during the past few days and in Hirsh's summary. They are, one, the extent to which the language capabilities in the human reflect unique and, therefore, implicitly, qualitatively distinctive capabilities in terms of a hierarchy of biological systems; and, two, that human speech motor activity represents unique requirements of which the two that have been discussed the most are temporal resolution capabilities and dependence upon acoustic information. I am not really convinced about any of these points. I think that very interesting data have been brought up in support of these points, but, in a sense, I think that too premature a decision about them might tend to make us not look in areas in which more looking might be fruitful.

Is it the case that speech learning requires acoustic information as a uniquely important kind of information, or is the issue, rather, that this just happens to be the way, under normal circumstances, in which I get most of my early information about the patterning of speech motor gestures? I am concerned about the patients with whom Risberg is working and the general problem of the education of the congenitally deaf child, which could certainly be hampered by a strong bias in terms of the unique importance of acoustic information. I wonder whether a very young child living in a world in which he is getting most of his information about speech motor gestures through the visual pathways or tactile inputs--as he would living in the world of Risberg's laboratory--and excluded, by virtue of his pathology, from sharing in the rich acoustic environment of speech transacted by other human beings, might not do equally as well.

HIRSH The congenitally deaf child is hampered because he has no acoustic input. You may say this is just a coincidence of the fact that most of the information that is normally provided in the child's environment is acoustic--to which I would have to agree--but that coincidence is rather more important than that word implies, for two reasons. One, we suspect, or at least I do, that the language system has been built up because it is a spoken system, and the rules of language that the child is supposed to be inducing from his experience are rules about a system that involves the speech mechanism. Secondly--and I don't know which is chicken and which is egg--the ears in the normal person are not shut, ever. Of all the sense modalities that one could choose, I suppose, this becomes the most important, because of this fact, that it is the system by which one remains in some kind of sensory contact with environment, always.

COOPER I would like to come back to your point, Hirsh, that there is something a little special about the time dimension--although that is not quite the way you put it--and also the comment you made that the most of our discussion about neurophysiology had been in terms of place. I wonder if we shouldn't consider the possibility of time patterns in neural function?

HIRSH Well, you did not misunderstand me. Let me state the case in the most extreme form, just for argument's sake.

By 'place in the nervous system' I suppose I have in mind a multiple-channel system of transmission that can be required, for example, for the recognition of speech. Whether these channels are spread out along the basilar membrane or along a strip of auditory cortex, I don't care. It is entirely conceivable to me that someone may have an auditory system that is defective with respect to this spread in space; he is then rendered a one-channel listener--he cannot recapture the kind of spectral information that would normally get transformed into spatial information--but he would still get along reasonably well in communicating on the basis of time-varying features alone. What I am saying is that response--which in many cases is a twentieth-century equivalent for the mind--may react to time-varying neural stimulation directly.

COOPER Do you mean patterns that exist in time as well as in space?

HIRSH I mean patterns that don't get transformed into space.

COOPER Well, just to your left is Licklider who thought at one time that they do get transformed into space and get handled that way.

LICKLIDER I suppose, if the response occurs at a particular time, and if it is a response to a sequence of events in the past, then all of those past events--except maybe the current one, if you will allow it--have to get represented in some nontemporal way.

HIRSH Periodicity pitch is one example--low-frequency pitch, where you don't have the spatial conversion.

COOPER But I thought Licklider's model did convert temporal phenomena to place phenomena.

LICKLIDER Even reverberating memory techniques require some kind of a spatial separation. You can't get anywhere with merely temporal variations. I cannot remember anything that you have said, except that one primordial blob, if I don't have some kind of spatially distributed memory in my head.

HIRSH Can't you have it temporally distributed in the same cell?

LICKLIDER Not at one moment, I cannot.

BROADBENT I'm quite happy about that. I'm sure that if, say, you alter refractory periods of one element, you may then alter the temporal patterns in which it will or will not take part. If you can think of some structural change which will alter the refractory period, say, which will then ensure that in future whenever it is stimulated it will respond to one temporal pattern and not to another, this does not necessarily mean that the trace is localized in the sense of being in one unit and not in another unit.

LICKLIDER My reaction was essentially on a different level of argument. I was just being markovian in saying that if response is only to record in the nervous system at time T, then there had better be something else besides temporal distinctions being made because there aren't any temporal distinctions being made at time T.

GOLDSTEIN I would like to make a brief comment on this subject. We may make a big mistake by trying to make such a clear cleavage between time and space. For one thing, the little bit of detailed neurophysiology that we do have about the auditory systems of animals shows the spectral pattern being presented and re-presented three times in the cochlear nucleus and on up (117), and at least three times at the cortical level (138).

It may be important to look at the temporal patterns playing over this space. We can think of the auditory system as feeling out the speaker's vocal tract. Such a concept would fit with the carrier theory of speech perception. There is evidence now that single cells will respond to a moving pattern, in a way that they would not respond to a pattern that did not move. I think, maybe, there are cases of this, and, maybe, that is the sort of thing that we can keep looking for as we look at the responses to more complex patterns of sounds.

LIBERMAN I want to return to the question that Chase raised a little while ago, about whether there is something very special about the auditory aspects of language and, if so, what are the implications of this for the deaf child? I believe that there are certain characteristics of language which are not modality specific. These reflect cognitive constraints, if you will. One of them, for example, is the essentially phonemic structure of language. But the phonemic segments are not directly represented in the acoustic signal, and for very good reasons which I think we discussed the other day--namely, that the temporal resolving power of the ear could not possibly handle these segments at ordinary speech rates. The acoustic signal represents a fairly elaborate encoding of these phonemes into units of approximately syllabic dimensions.

This raises the problem, as a practical matter, that if you present the linguistic information in this encoded acoustic form to some modality other than the auditory one, and if the person hasn't got a chance to do the kind of articulating that normally helps him to decode the complex signal, then, you've got a serious problem. Conceivably, it would not be nearly so difficult to get the language into the child by some other modality, with one of Risberg's devices, if, indeed, you could present it in phonemically segmented form. Then he would not have to face from the very outset what I conceive to be an extremely difficult decoding problem.

CHASE Implicit in your remarks, Liberman, if I understand them, is that not just the information be presented in the segmented form, but also that, at the critical learning period the situation be structured such that the child must converse with the synthetic or prosthetic display. Whether or not you subscribe to the motor theory of speech perception, the evidence has been abundant in the past few days that the child learns speech as a function of active participation. Indeed, the unique import of acoustic information might be the laying down of the instructions to the child for his pattern-matching task, which becomes progressively more refined, more economical, and which many of us suspect involves the very process by which the patterns or rules for categorical receptivity and productive operations become structured.

I think that we may find--looking at work like Risberg's--that, if these instructions for the dialogue within which you are inviting a child to shape his motor gestures so as to come in closer and closer correspondence to yours use nonacoustical displays of an appropriate sort, this may well turn out to be a very close parallel to the normal acquisition of speech by a hearing child.

HIRSH May I just mention one factor? We can't do more than mention it because the data are not available in this connection. The exposure to auditory stimulation of nonspeech, and auditory stimulation by speech including monitoring, is continuous in childhood. It occupies at least twelve hours a day, and perhaps we should count twenty-four hours a day--even during sleep. This is a different kind of limitation on putting the information into another modality; not just that the modality is different, but that the child has to look or has to be touching in order to get the information. The total amount of time is several orders of magnitude different. If you could code for the visual or the tactile system, the input channel must be left open as long, say, as the ears are open.

CHASE There are several points here. Could it be that most of what the child hears is not relevant for the acquisition of language? Furthermore, if one could isolate the optimal learning period for speech acquisition, the process of teaching speech using nonacoustic instructions might be a fairly economical procedure.

HIRSH What do you suppose is happening to the analogs of these Hubel hookups (59) in the auditory system, even during that nonspeech period?

CHASE I really don't know. I am assuming, as I guess most of us would, that the prelinguistic vocalization patterns of the child reflect a primitive catalogue of plastic motor gestures which become progressively refined. One of the questions that concerns me a great deal is what the critical information is that imposes this refinement in economy in all the things the child could say, such that he says only certain kinds of things. I am not quite sure what the implication would be in terms of the neural substrate.

HIRSH I am implying an importance for more than speech stimuli in keeping this system developing.

GOLDSTEIN But we are always touching somewhere. That system is working all the time, twenty-four hours a day.

COOPER In the visual case, aren't we learning to categorize those things that move versus those things that do not move, or those visual patterns that remain intact as against those that dissociate? This would be categorization by example, if you will. Speech differs from this, only because the same sorts of categories that come in from the external world also reappear from the internal world, so that a matching operation is available, in addition.

GESCHWIND The point you are making is, then, that a child can readily reproduce spoken speech but cannot so easily reproduce written language or tactile language.

RISBERG It seems obvious, anyhow, that this is a very important experiment for our understanding of speech perception. If it works we will learn quite a lot about what goes on, and, if it doesn't work, we will still learn quite a lot.

COOPER We have had at least two reasons given why it might not work. One is Hirsh's point that it would be hard to arrange enough exposure, although the rapid rate of learning in children is in our favor. The other is that, in experimenting with adults, we may be working with people who quite literally are too old to learn. The question of plasticity may be central to a test of the basic hypothesis.

RISBERG Yes. Then this test must be made on young people. One thing, of course, is that many of the deaf--well, not many, but some of the congenitally deaf--can learn to speak very well with little exposure or almost no exposure at all to any sound situation.

HIRSH They are usually very bright, however.

GESCHWIND Hirsh mentioned that the traditional view was that the more complex functions of the nervous system were less localized, while the less complex ones were more localized. I think it depends on whose traditional view it is, and I would suggest that the traditional view that he is talking about, in fact, is a very recent one. Before 1860, the view was held by the French physiologist, Flourens, of the nonlocalizability of higher functions, but this was rapidly rejected and, in fact, it did not have any significant hold on people's thinking. That view did not reappear until about 1900; in fact, it had its heyday predominantly in the United States after 1920 and is now declining again. In fact, it was only a rather brief period in which people seriously took to the view about complex functions not being localizable.

LENNEBERG I think that relevant to this is work by the late E. von Holst (54). The last thing he did was something that most neurophysiologists frowned on. He took very fine electrodes and inserted them into the brainstem of live chickens without any attempt at determining the exact location. He made these fine wires worm their way through live tissue and delivered electrical stimuli to the animals at random. He found that the chicken could be made to perform extremely complex patterns of behavior; they were species specific patterns such as roosting, scratching, threat postures, etc. I think, as a matter of fact, this is relevant to speech, because some neurologists feel that there are structures in the brainstem that are relevant to motor integration for speech, and that there are lesions that interfere with speech on that low level of the neuraxis.

COOPER Would you expect to be able to get stimulation that would produce speech?

LENNEBERG This is impossible to speculate on. However, some experiments have indicated that the periaqueductal gray matter was particularly relevant to vocalization and integration for vocalization (3, 19, 68, 124). They used cats and recently dogs; no work has been done on monkeys. It was found that this location had something to do with integration of vocalization.

POLLACK One important aspect of communication that we have not discussed is that communication is typically a social act. Perhaps, this is an item for another group to discuss. Some place along the line, however, we might have addressed ourselves profitably to this aspect of the communication problem.

HOUSE Can I raise a few mild objections to some of the interpretations that Hirsh has put on the proceedings. It seems to me that a conclusion that says less global statements rather than more global statements have come out of the discussion is a misinterpretation of what has happened. Hirsh's attempt to separate spectral information from time-varying information has been relatively unsuccessful--for this group at least. I still believe there is a lot of time information in the spectral display. I also believe there is spectral information in some of the time-varying parameters seen in a spectral display. Furthermore, I do not believe that Helmholtz understood vowel perception in the sense that we understand it today. Equating modern formant theories--or more strongly, modern acoustical theories--to the phenomena that he talked about seems incorrect to me. Today we try, for example, to use the same concepts in talking about all speech activity--vowel production and consonant production.

When the problem of the categories of linguistic code is raised, I think of the productive and perceptual processes we have been talking about as descriptions of the production and perception of a linguistic code--not as processes that are modified by the code.

I find Hirsh's last observation about behavioral levels extremely interesting. Many people think you can generate models that will take you from one behavioral level to another; there are structural linguists, for example, who feel that rules of syntax can be extended down to the phonological level, and also can be extended into rules on some semantic level.

LIBERMAN I understood Hirsh to say that we were agreed that the relations between linguistic units and acoustic signals, for example, were complex, and that the relationship is also complex as between linguistic units and articulation. I want to say for the record that we don't all agree on that. There are some people who think that while the relationships between, let's say, the control signals in the common path and the linguistic units may not be one-to-one, nevertheless they are simpler in some sense that the relationships between the acoustic signal and the linguistic unit. This is an important question and it is a point you made, about which, I think, we don't all agree with you.

HIRSH If the purpose of a summary is to evoke further discussion, I am sure I have achieved it. If the purpose is to set down, among other things, generalizations on which all would agree, then, I am afraid the page would be blank.

I would like to respond to the point that House made about a distinction between time-varying and spectral information that reflects my interpretation of some things that have been reported here. Insofar as these are differently affected or effective, in setting up linguistic constraints, one would expect, for example, that only certain kinds of discriminations would be modified by categorization and not others.

LICKLIDER Was it made explicit in the earlier discussion that I missed that the same thing can be both time-varying and spectral, depending upon the point of view, and that the important thing here is what echelon of the hierarchy you are looking at when you say this? Presumably, the things that are time-varying and also spectral in the vowel domain are varying so fast in time that you don't want to think about that in this particular discussion, but the vowel spectra themselves may be in the next echelon of the description, time-varying.

HIRSH I don't think we have discussed this point as much as we should have. It really has to do with the point Fry made earlier, that there are almost always several kinds of cues available for a discrimination, and that you can't always count on a particular cue being the most important in all contexts.

HOUSE In this regard, if I were called upon to modify either one of these so-called classes of cues in a sound spectrogram, I'm sure I would be confused by my task. If I looked at a sound spectrogram for the first time, I believe that I would be similarly confused--confused enough to identify things as being either time or spectrally varying.

STEVENS I confess that earlier in the conference I was forced into a yes-or-no answer to a question relating to the dichotomy of temporal versus spectral properties of speech signals. I would, however, prefer not to use the labels <u>time</u> and <u>spectral</u>, but rather to think of dichotomies in articulatory terms. Whatever comes out as sound may not be easily categorized as time varying or spectral.

HIRSH Something like place and manner?

STEVENS Possibly, or, perhaps, open vocal-tract tone versus constricted vocal-tract tone.

LIBERMAN Possibly, vowel versus consonant.

LICKLIDER I would like to suggest a problem, or ask a question, whichever way you look at it. I remember seeing a machine that makes spectrograms in real time. Has anybody exploited the use of such real-time spectrograms in the effort to teach children who are deafened how to talk clearly? It seems to me that there is the way to get almost all those properties of dynamic display and feedback that the ears provide, except the visual system doesn't have the dynamics of the auditory one. But I would bet this would work in a handsome way.

DENES The Bell Telephone Laboratories has taken another look at the old visible speech apparatus and has now produced a more modern version of it. The new device is essentially the same as the earlier one, except for its more up-to-date- and less-costly internal construction. It uses a bank of 24 filters that are scanned by an electronic switch whose output is displayed in the conventional spectrographic manner on the face of a rotating cathode-ray tube. The frequency-intensity-time spectrogram appears instantaneously as you speak into a microphone and displays the spectrum of about the last three seconds of applied speech.

We are actively considering how this device could be used to help the deaf. We would like to know whether the spectrographic patterns are learnable, whether they could usefully supplement deficient hearing the way that lip reading does, whether they can help in developing useful speech in deaf children, and so on. Even more basically, we would like to know which acoustic speech features are to be displayed to provide the maximum effect in any of the problems just enumerated.

HIRSH There is some work going on in Detroit (73) under somewhat different principles. There the teacher has been trying to abstract out for the child the principles that should be attended to in the spectrogram--rather than letting the child ferret them out for himself.

LICKLIDER It seems to me that you could fairly easily instrument a system in which you have a display showing target sounds which are like those the child should generate, a display of what the child actually generates, and the difference between the two, so the child's task is just to reduce that last display to nothing.

RISBERG We tried this with a display that shows only frequency and amplitude. The way this is used in teaching is that the teacher says the sound and you get a pattern. Then the teacher presses a button and the pattern is stored, and you can map the pattern onto plastic sheets in front of the tube. Then the child can try to produce the same sound. This works quite well for fricative sounds where the teacher's frequency spectrum is almost the same as the child's, but with a vowel, with the great difference in head size and fundamental frequency, it is not possible to work this way. But this can probably be overcome. We have been testing this device for a month or so and it seems to work rather well, at least for the fricative sounds.

Our intent now is to test this, first, as a general test to see how the children react and how the teachers react, and then to see if they can learn to produce the fricative sounds. We have no measurements of the improvement yet. We have just started the training, but we will get the first results back soon.

DENES This is just what the Kopps in Detroit have been doing. The children are sitting watching the visible speech device in a room with a blackboard. The teacher produces the sound and shows the resulting pattern on the screen of the cathode-ray tube. She then draws the salient features of the pattern on the blackboard while pointing to the immediately displayed spectrogram, and then she invites the children to do the same thing. The children cotton on to this very quickly and they control each other's behavior. If one child is asked to produce a syllable like ta and the second formant transition isn't appropriate, all the other children start to shout, throwing up their hands indicating a need to raise the formant transition, for example, and showing that they appreciate the relationship between the articulatory movement and the acoustical pattern.

LENNEBERG Has this training improved articulation in these children?

DENES Well, that is a different question; I don't know.

LADEFOGED Isn't there in the Kopps' report (73) a claim that children have improved their articulation significantly by using this technique, as compared with a matched group of children who had equal amounts of training but were not exposed to this method and who didn't improve to the same degree?

GOLDSTEIN Have there been any attempts to present directly to these very young children information about the articulatory configurations of either themselves or the person who is speaking to them?

RISBERG Yes, there has been one attempt, I think, at Northwestern University, where they measure the capacity between the tongue and the velum (16).

HIRSH Displayed visually?

RISBERG Yes, I think it is displayed visually.

COOPER We have one minute until the time I promised we would stop. I should like to use that one minute, if I may, to thank the NICHD for being our hosts at this conference, for having made it possible. We should thank, also, the members of the NICHD staff who have hovered in the background and anticipated all our needs--Miss Betty Barton and Mrs. Meryom Lebowitz; also, Dr. Fremont-Smith, who had to leave early, and Mrs. Betty Purcell of his staff; and in particular, Miss Edna Meininger, who has been taking down all the words of wisdom that we, hopefully, have been producing. To those of you who have served as discussion leaders, I owe a personal debt; I hope and believe you found the experience rewarding. Shall we give the NICHD a rising vote of thanks? (Applause)

REFERENCES

1. AMES, A. Jr. Visual perception and the rotating trapezoidal window. Psychol. Monogr. 65, Whole No. 324 (1951).

2. BADDELEY, A. D. Unpubl. PhD thesis, University of Cambridge (1962).

3. BAILEY, P. and DAVIS, E. W. Effects of lesions of the periaqueductal gray matter in the cat. Proc. Soc. exper. biol. Med. 51, 305-306 (1942).

4. BASTIAN, J. and ABRAMSON, A. S. Identification and discrimination of phonemic vowel duration. J. acoust. Soc. Amer. 34, 743, A (1962).

5. BASTIAN, J., DELATTRE, P. and LIBERMAN, A. M. Silent interval as a cue for the distinction between stops and semivowels in medial position. J. acoust. Soc. Amer. 31, 1568, A (1959).

6. BASTIAN, J., EIMAS, P. D. and LIBERMAN, A. M. Identification and discrimination of a phonemic contrast induced by silent interval. J. acoust. Soc. Amer. 33, 842, A, (1961).

7. BATKIN, S. and ANSBERRY, M. Effect of auditory deprivation. J. acoust. Soc. Amer. 36, 598, L (1964).

8. BEKESY, G. v. Experiments in hearing. New York: McGraw-Hill, 1960. P. 454.

9. BERENDES, J. u. VOGEL, W. Kehlkopfmuskeln im elektronenmikroskopischen Bild. Arch. Ohren-, usw. Heilk. 176, 730-735 (1960).

10. BERG, Jw. v. d. Ueber die Koppelung bei der Stimmbildung. Z. Phonetik, 8, 281-293 (1954).

11. BERGEIJK, W. A. v. and DAVID, E. E. Jr. Delayed handwriting. _Percept. motor Skills_, 9, 347-357 (1957).

12. BOREL-MAISONNY, S. Audition et perception: Intelligence du langage. _J. franc. Oto-Rhino-Laryngol._ 8, 47-64 (1959).

13. BORING, E. G. _History of experimental psychology_. New York: Appleton-Century, 2nd ed., 1950. Pp. 417-419.

14. BORING, E. G. _Sensation and perception in the history of experimental psychology_. New York: Appleton-Century, 1942. Pp. 527-529.

15. BRAIN, R. _Speech disorders: Aphasia, apraxia and agnosia_. Washington: Butterworth, 1961.

16. BRANNON, J. B. Jr. Visual feedback of glossal motions and influence upon the speech of deaf children. Unpubl. PhD thesis, Northwestern University (1964).

17. BROADBENT, D. E. The role of auditory localisation in attention and memory span. _J. exp. Psychol._ 47, 191-196 (1954).

18. BUTLER, R. A. and GALLOWAY, F. T. Factoral analysis of the delayed speech feedback phenomenon. _J. acoust. Soc. Amer._ 29, 632-635 (1957).

19. CAIRNS, H. _et al_. Akinetic mutism with an epidermoid cyst of the III ventricle. _Brain_, 64, 273-290 (1941).

20. CHARLTON, M. H. Aphasia in bilingual and polyglot patients: A neurological and psychological study. _J. speech hear. Disord._ 29, 307-311 (1964).

21. CHASE, R. A. _et al_. Studies on sensory feedback, I: Effect of delayed auditory feedback on speech and keytapping. _Quart. J. exp. Psychol._ 13, 141-152 (1961).

22. CHASE, R. A. _et al_. Studies on sensory feedback, II: Sensory feedback influences on key-tapping motor tasks. _Quart. J. exp. Psychol._ 13, 153-167 (1961).

23. CHASE, R. A. et al. Studies on sensory feedback, III: The effects of display gain on tracking performance. Quart. J. exp. Psychol. (in press).

24. CHASE, R. A., CULLEN, J. K. Jr., SULLIVAN, S. A., BOSMA, J. F. and KAVANAGH, J. Motor control abnormalities secondary to congenital sensory deficits: A case presentation. Motion picture produced at the Neurocommunications Laboratory, The Johns Hopkins Hospital, Baltimore, Md.

25. CHERRY, E. C. et al. Experiments on the complete supression of stammering. Nature, 176, 874-875 (1955).

26. CHIBA, T. and KAJIYAMA, M. The vowel: Its nature and structure. Tokyo: Tokyo-Kaiseikan, 1941.

27. CONRAD, R. Acoustic confusions in immediate memory. Brit. J. Psychol. 55, 75-84 (1964).

28. COOPER, F. S. et al. Some input-output relations observed in experiments on the perception of speech. Proc. Second Internat. Cong. Cybernet., 1958. Namur, Belgium: Assoc. Internationale de Cybernétique, 1959. Pp. 930-941.

29. DEJERINE, J. Contribution à l'étude de l'aphasie chez les gauchers. Rev. neurol. 24, 214-226 (1912).

30. DENES, P. On the statistics of spoken English. J. acoust. Soc. Amer. 35, 892-904 (1963).

31. EFRON, R. Temporal perception, aphasia, and déjà vu. Brain, 86, 403-424 (1963).

32. EIMAS, P. The relation between identification and discrimination along speech and non-speech continua. Lang. Speech, 6, 206-217 (1963).

33. ETTLINGER, G. Cross-modal transfer of training in monkeys. Behaviour, 16, 56-65 (1960).

34. FAABORG-ANDERSEN, K. Electromyographic investigation of intrinsic laryngeal muscles in humans. Acta physiol. Scand. 41, Suppl. 140 (1957).

35. FANT, G. Acoustic theory of speech production. 's-Gravenhage: Mouton, 1960.

36. FANT, G. and SONESSON, B. Indirect studies of glottal cycles by synchronous inverse filtering and photo-electrical glottography. Quart. Prog. Status Rep. 4 (1962), Speech Transmission Laboratory, Royal Institute of Technol. (Stockholm). Pp. 1-3.

37. FLECHSIG. P. Developmental (myelogenetic) localisation of the cerebral cortex in the human subject. Lancet, 2, 1027-1029 (1901).

38. FLETCHER, H. Speech and hearing. New York: Van Nostrand, 1929. P. 280.

39. FRENCH, N. R. and STEINBERG, J. C. Factors governing the intelligibility of speech sounds. J. acoust. Soc. Amer. 19, 90-119 (1947).

40. FRY, D. B. et al. The identification and discrimination of synthetic vowels. Lang. Speech, 5, 171-189 (1962).

41. FURMAN, G. G. and FRISHKOPF, L. S. Model of neural inhibition in the mammalian cochlea. J. acoust. Soc. Amer. 36, 2194-2201 (1964).

42. GAULT, R. H. A partial analysis of the effects of tactual-visual stimulation by spoken language. J. Franklin Inst. 209, 437-458 (1930).

43. GELDARD, F. A. Cutaneous channels of communication. In Sensory communication, W. A. Rosenblith, Ed. New York: Wiley, 1961. Pp. 73-87.

44. GESCHWIND, N. Disconnection syndromes in animals and man. Brain, 88, 237-294; 585-644 (1965).

45. GESCHWIND, N. The development of the brain and the evolution of language. Monogr. Series Lang. Linguist., Georgetown Univ., 17, 155-170 (1964).

46. GOLDMAN-EISLER, F. The distribution of pause durations in speech. Lang. Speech, 4, 232-237 (1961).

47. GOLDSTEIN, M. H. Jr., KIANG, N-Y. S. and BROWN, R. M. Responses of the auditory cortex to repetitive acoustic stimuli. J. acoust. Soc. Amer. 31, 356-364 (1959).

48. GUIOT, G. et al. Arrest or acceleration of speech evoked by thalamic stimulation in the course of stereotaxic procedures for Parkinsonism. Brain, 84, 363-381 (1961).

49. GUTTMAN, N. and BERGEIJK, W. v. Frequency division in real time as a means of counteracting high-frequency hearing loss. Proc. Fourth Internat. Cong. Audiol., 1958. Lyon: Audin, 1959. Pp. 255-256.

50. HALLE, M. and STEVENS, K. N. Speech recognition: A model and a program for research. IRE Trans. Info. Theory, IT-8, 155-159 (1962).

51. HEBB, D. O. The organization of behavior. New York: Wiley, 1949.

52. HELMHOLTZ, H. L. F. Sensations of tone. New York: Longmans, Green; trans. A. J. Ellis, 1912. Pp. 103-114.

53. HIRSH, I. J. Unpublished observations.

54. HOLST, E. v. and SAINT PAUL, U. v. Electrically controlled behavior. Scient. Amer. 206, 50-59 (March 1962).

55. HOUSE, A. S. et al. On the learning of speech-like vocabularies. J. verb. Learn. verb. Behavior, 1, 133-143 (1962).

56. HUBEL, D. H. The visual cortex of the brain. Scient. Amer. 209, 54-64 (November 1963).

57. HUBEL, D. H. and WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160, 106-154 (1962).

58. HUBEL, D. H. and WIESEL, T. N. Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. J. Neurophysiol. 26, 994-1002 (1963).

59. HUBEL, D. H. and WIESEL, T. N. Studies of the visual system in light-space and form-deprived kittens. In Physiological basis for form perception, a symposium of the Visual Sciences Study Section, National Institutes of Health. Providence, R. I.: Hunter Laboratory of Psychology, Brown University, 1964.

60. HUSSON, R. Étude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée. Thèse, University of Paris (1950).

61. HUSSON, R. Physiologie de la phonation. Paris: Masson, 1962.

62. ITTELSON, W. H. and AMES, A. Jr. Accomodation, convergence, and their relation to apparent distance. J. Psychol. 30, 43-62 (1950).

63. JAKOBSON, R. and HALLE, M. Fundamentals of language. 's-Gravenhage: Mouton, 1956.

64. JOHANSSON, B. A new coding amplifier system for the severely hard of hearing. Proc. Third Internat. Cong. Acoust., 1959. Amsterdam: Elsevier, 1961. Pp. 653-657.

65. KALMUS, H., FRY, D. B. and DENES, P. Effects of delayed visual control on writing, drawing and tracing. Lang. Speech, 3, 96-108 (1960).

66. KATSUKI, Y. et al. Electric responses of auditory neurons in cat to sound stimulation. J. Neurophysiol. 21, 569-588 (1958).

67. KATSUKI, Y., SUGA, N. and KANNO, Y. Neural mechanism of the peripheral and central auditory system in monkeys. J. acoust. Soc. Amer. 34, 1396-1410 (1962).

68. KELLY, A. H., BEATON, L. E. and MAGOUN, H. W. Midbrain mechanism for facio-vocal activity. J. Neurophysiol. 9, 181-189 (1946).

69. KIANG, N-Y. S. *et al*. Stimulus coding in the cat's auditory nerve. *Ann. Otol. etc., St. Louis*, 71, 1009-1026 (1962).

70. KNAPP, H. D., TAUB, E. and BERMAN, A. J. Effect of deafferentation on a conditioned response. *Science*, 128, 842-843 (1958).

71. KOENIG, G. u. EICHLER, F. Ein neues Verfahren zur Hoerverbesserung bei hochgradiger Schwerhoerigkeit durch Hoerapparate. *Arch. Ohren-, usw. Heilk*. 165, 327-331 (1954).

72. KOENIG, W. F. Histologische Untersuchungen ueber die sensiblen nervoesen Endigungen im Stimmband des Menschen. *Arch. Ohren-, usw. Heilk*. 174, 92-102 (1958).

73. KOPP, G. A. and KOPP, HARRIET G. An investigation to evaluate the usefulness of the visible speech cathode ray tube translator as a supplement to the oral method of teaching speech to deaf and severely deafened children. Final Report, Grant RD-526, Off. Voc. Rehabil., Department of Health, Education, and Welfare, Washington, D. C., 1963.

74. KYDD, W. L. Maximum forces exerted on the dentition by the perioral and lingual musculature. *J. Amer. dent. Assoc*. 55, 648-651 (1957).

75. KYDD, W. L. and BETT, D. A. Continuous palatography. *J. speech hear. Disord*. 29, 489-491 (1964).

76. KYDD, W. L. *et al*. Tongue and lip forces exerted during deglutition in subjects with and without an anterior open bite. *J. dent. Research*, 42, 858-866 (1963).

77. LADEFOGED, P. *A phonetic study of West African languages*. Cambridge Univ. Press, 1964.

78. LADEFOGED, P. Some possibilities in speech synthesis. *Lang. Speech*, 7, 205-214 (1964).

79. LADEFOGED, P. Subglottal activity during speech. *Proc. Fourth Internat. Cong. Phonet. Sciences*, 1961. 's-Gravenhage: Mouton, 1962. Pp. 247-265.

80. LADEFOGED, P. and BROADBENT, D. E. Information conveyed by vowels. J. acoust. Soc. Amer. 29, 98-104 (1957).

81. LADEFOGED, P. and BROADBENT, D. E. Perception of sequence in auditory events. Quart. J. exp. Psychol. 12, 162-170 (1960).

82. LADEFOGED, P. and McKINNEY, N. Loudness, sound pressure and subglottal pressure in speech. J. acoust. Soc. Amer. 35, 454-460 (1963).

83. LAMPERT, W. and FILLENBAUM, J. A pilot study of aphasia among bilinguals. Canad. J. Psychol. 13, 28-34 (1959).

84. LANE, H. L. Psychophysical parameters of vowel perception. Psychol. Monogr. 76, Whole No. 563 (1962).

85. LANE, H. L., CATANIA, A. C. and STEVENS, S. S. Voice level: Autophonic scale, perceived loudness, and effects of sidetone. J. acoust. Soc. Amer. 33, 160-167 (1961).

86. LANYON, W. E. The ontogeny of vocalizations in birds. In Animal sounds and communication, W. E. Lanyon and W. N. Tavolga, Eds. Washington: American Institute of Biological Sciences, 1960. Pp. 321-347.

87. LASHLEY, K. S. The accuracy of movement in the absence of excitation from the moving organ. Amer. J. Physiol. 43, 169-194 (1917).

88. LENNEBERG, E. H., REBELSKY, F. G. and NICHOLS, I. A. The vocalizations of infants born to deaf and to hearing parents. Vita humana, 8, 23-37 (1965).

89. LIBERMAN, A. M. et al. A motor theory of speech perception. Proc. Speech Communicat. Seminar, 1962. Stockholm: Speech Transmission Laboratory, Royal Institute of Technol., 1963. np.

90. LIBERMAN, A. M. et al. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. Lang. Speech, 4, 175-195 (1961).

91. LIBERMAN, A. M. *et al*. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J. exp. Psychol*. 61, 379-388 (1961).

92. LIEPMAN, H. *Ein Fall von reiner Sprachtaubheit*. Breslau: Schletter, 1898.

93. LISKER, L. et al. The uses of experiment in language description. *Word*, 18, 82-106 (1962).

94. MacNEILAGE, P. F. Electromyographic and acoustic study of the production of certain final clusters. *J. acoust. Soc. Amer*. 35, 461-463 (1963).

95. MacNEILAGE, P. F. and SHOLES, G. N. An electromyographic study of the tongue during vowel production. *J. speech hear. Research*, 7, 209-232 (1964).

96. MATZKER, J. Ein neuer Weg zur otologischen Diagnostik zerebraler Erkrankungen; der Binauraltest bei raumverengenden endokraniellen Prozessen. *Z. Laryngol. Rhinol. Otol*. 36, 177-189 (1957).

97. McCORMICK, G. P. Intraoral myometric palatography via electrical pressure sensitive transducers. Unpubl. PhD thesis, Purdue University (1965).

98. McCULLOCH, W. S. Reliable systems using unreliable units. *Disord. Communicat*. 42, 19-28 (1964).

99. MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev*. 63, 81-97 (1956).

100. MILLER, G. A. and NICELY, P. E. Analysis of perceptual confusions among some English consonants. *J. acoust. Soc. Amer*. 27, 338-352 (1955).

101. MILLER, G. A. and SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *Amer. J.* Psychol. 63, 176-185 (1950).

102. MILLER, G. A. and TAYLOR, W. G. Perception of repeated bursts of noise. *J. acoust. Soc. Amer*. 20, 171-182 (1948).

103. MILLER, G. A., GALANTER, E. and PRIBRAM, K. H. Plans and the structure of behavior. New York: Holt, 1960.

104. MILNER, P. M. The cell assembly: Mark II. Psychol. Rev. 64, 242-252 (1957).

105. NEUMANN, J. v. Probabilistic logic and synthesis of reliable organisms from unreliable components. In Automata studies, C. E. Shannon and J. McCarthy, Eds. Princeton Univ. Press, 1956. Pp. 43-98.

106. NOMATO, M., SUGA, N. and KATSUKI, Y. Discharge pattern and inhibition of primary auditory nerve fibers in the monkey. J. Neurophysiol. 27, 768-786 (1964).

107. OEKEN, F-W. Frequenztransposition zur Hoerverbesserung bei Innenohrschwerhoerigkeit. Arch. Ohren-, usw. Heilk. 181, 418-425 (1963).

108. OEKEN, F-W. Untersuchungen ueber das Sprach und Sprachlautverstandnis bei Frequenztransponierung. Phonetica, 10, 92-109 (1964).

109. PAULSEN, K. Ueber die Bedeutung von Muskelspindeln und Schleimhautrezeptoren bei der Phonation. Arch. Ohren-, usw. Heilk. 173, 500-504 (1958).

110. PAULSEN, K. Ueber Vorkommen und Zahl von Muskelspindeln in inneren Kehlkopfmuskeln des Menschen. Z. Zellforsch. 48, 349-355 (1958).

111. PFUNGST, O. Clever Hans. New York: Holt, 1911. Trans. by C. L. Rahn.

112. PICKETT, J. M. Tactual communication of speech sounds to the deaf: Comparison with lipreading. J. speech hear. Disord. 28, 315-330 (1963).

113. POTTER, R. K., KOPP, G. A. and GREEN, H. C. Visible speech. New York: Van Nostrand, 1947. P. 25.

114. RAPIN, I. et al. Effects of varying delays in auditory feedback on keytapping in children. Percept. motor Skills, 16, 489-500 (1963).

115. RINGEL, R. L. and STEER, M. D. Some effects of tactile and auditory alterations on speech output. J. speech hear. Research, 6, 369-378 (1963).

116. ROME, J. A. An artificial palate for continuous analysis of speech. Unpubl. SB thesis, Massachusetts Institute of Technology (1964).

117. ROSE, J. E. Organization of frequency sensitive neurons in the cochlear nuclear complex of the cat. In Neural mechanisms of the auditory and vestibular systems, G. L. Rasmussen and W. F. Windle, Eds. Springfield, Ill.: Thomas, 1960. Pp. 116-136.

118. ROSEN, J. Phoneme identification in sensorineural deafness. Unpubl. PhD thesis, Stanford University, (1962).

119. RUDOLPH, G. Spiral nerve-endings (proprioceptors) in the human vocal muscle. Nature, 190, 726-727 (1961).

120. RUPERT, A., MOUSHEGIAN, G and GALAMBOS, R. Unit response to sound from auditory nerve of the cat. J. Neurophysiol. 26, 449-465 (1963).

121. SCHOUTEN, J. F., KALSBEEK, J. W. H. and LEOPOLD, F. F. On the evaluation of perceptual and mental load. Ergonomics, 5, 251-260 (1962).

122. SCHULTZ, M. C. and BERMAN, M. Speech discrimination in normal and pathological ears. J. acoust. Soc. Amer. 36, 1050, A (1964).

123. SHERRINGTON, C. S. and MOTT, F. W. Experiments upon the influence of sensory nerves upon movement and nutrition of the limbs. Proc. roy. Soc., London, 57, 481-488 (1895).

124. SKULTETY, M. F. Experimental mutism following electrolytic lesions of the periaqueductal gray matter in dogs. Trans. Amer. neurol. Assoc. 86, 245-246 (1961).

125. SMITH, K. U. *Delayed sensory feedback and behavior*. Philadelphia: Saunders, 1962.

126. SONESSON, B. On the anatomy and vibratory pattern of the human vocal folds. *Acta oto-laryngol.*, Suppl. 156 (1960).

127. STEVENS, K. N. *et al*. Crosslinguistic study of vowel discrimination. *J. acoust. Soc. Amer.* 36, 1989, A (1964).

128. STUDDERT-KENNEDY, M. and LIBERMAN, A. M. Psychological considerations in the design of auditory displays for reading machines. Paper presented at International Congress on Technology and Blindness, New York, N. Y., 1962.

129. TAPPERT, C. C., MARTONY, J. and FANT, G. Spectrum envelopes for synthetic vowels. Quart. Prog. Status Rep. 3 (1963), Speech Transmission Laboratory, Royal Institute of Technol. (Stockholm). Pp. 2-6.

130. TASAKI, I. Nerve impulses in individual auditory nerve fibers of guinea pig. *J. Neurophysiol*. 17, 97-122 (1954).

131. TREISMAN, A. M. Reading rate, word information and auditory monitoring of speech. Med. Res. Council (UK), Psycholinguistic Res. Unit Rep. PLU 64/8 (1964).

132. VAN RIPER, C. and IRWIN, J. *Voice and articulation*. Englewood Cliffs, N. J.: Prentice-Hall, 1958.

133. WERNICKE, C. *Der Aphasische Symptomencomplex*. Breslau: Max Cohn and Weigert, 1874.

134. WIENER, N. *et al*. Felix (sensory replacement). Quart. Prog. Reps., Research Laboratory of Electronics, Mass. Institute of Technol., Jan. 1949-April 1951.

135. WIESEL, T. N. and HUBEL, D. H. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol*. 26, 1003-1017 (1963).

136. WINOGRAD, S. and COWAN, J. D. Reliable computation in the presence of noise. New York: Wiley, 1963.

137. WITHROW, F. B. Jr. Immediate recall by aphasic, deaf and normally hearing children for visual forms presented simultaneously or sequentially in time. Unpubl. PhD thesis, Washington University (1963).

138. WOOLSEY, C. N. Organization of cortical auditory system. In Sensory communication, W. A. Rosenblith, Ed. New York: Wiley, 1961. Pp. 235-258.

INDEX TO NAMES

☆ U.S. GOVERNMENT PRINTING OFFICE : 1967 O - 243-541