OBSERVATIONS ON SPEECH RESEARCH: OBJECTIVES, STRATEGIES, AND SOME PARTIAL ANSWERS
Franklin S. Cooper
Recipient of the Fletcher-Stevens Award

Languages and Linguistics Symposium
March 30-31, 1977
Sponsored by:
Deseret Language and Linguistic Society
and
Brigham Young University College of Humanities
In conjunction with
Language & Intercultural Research Center
Speech research, about which I should like to talk with you today, has long been my major interest. In this, I can claim a certain kinship of the son-to-father kind with Dr. Harvey Fletcher. To be sure, his research was on different topics, used different tools, and had different goals. That was to be expected, given that speech research itself was, and still is, a dynamically changing field. Perhaps I should mention that Fletcher's book on "Speech and Hearing," published in 1929, was already a classic when I first discovered speech in the mid-40's. That holds also for the book by Stevens and Davis with its simple but comprehensive title, "Hearing." These two books were my starting point; I have had the good fortune to know both Dr. Fletcher and Dr. Stevens personally, though really only at grazing incidence.

So, for me, it is a very special and very personal honor to be asked to give the Fletcher-Stevens Lecture—and it is a challenge, too. I gather, from the general guidelines I have been given, that I am not expected to encapsulate all that has been learned about speech within a 40-minute message—rather, that I should talk about speech research as an enterprise that has both a history, which I have watched with interest, and also a burgeoning future. It is my hope that looking back to some of the earlier researches will help to highlight the continuing problems that were uncovered in the process of experimenting on spoken language. I think it is fair to say, though, that we have learned more about the problems than about the answers. I shall be content if the linguists among you, and the engineers and the psychologists, can get some sense of challenge to your own discipline, since your combined efforts will be needed if we are eventually to understand how it is that human beings are able to communicate with each other so effectively by these strange sounds we call speech.

Objectives

In turning to the objectives of research on speech, I ought logically to ask—as some linguists actually do—what does speech research have to do with, or for, linguistics? The answer to that question is my main theme, but just now I will ask your patience until I have commented briefly on the more obvious practical interests of the engineers.

What are some of the practical goals and problems in speech research? There are many situations in which face-to-face conversation would be desirable but is simply not possible. Thus, the need to talk at greater distances than the voice will carry led to the development of telephony, a field in which Dr. Fletcher made many notable contributions. Sound recording made it possible for man's voice to reach across time as well as across space. But in both recording and simple telephony, noise and distortion have always been major problems. When there is need to speak from city to city, rather than house to house, these problems become extremely severe. Initial solutions depended on putting high quality amplifiers (or repeaters) at regular intervals along the telephone line. Digital methods have made it possible to greatly simplify the design of these repeaters, and for an interesting reason: when the voice waveform itself is being transmitted, any noise that is added during transmission gets merged into the speech and cannot be sorted out later. Consequently, amplifying the signal to make it stronger only makes the noise stranger, too, and the deterioration continues to increase. But digital methods replace the speech waveform by a numerical recipe for reconstructing the speech. So when this numerical description is transmitted, the repeating station can regenerate a noise-free set of numbers, thereby avoiding cumulative deterioration.

Another problem for telephony is the comparatively large bandwidth (or the bit rate in digital transmission) that is required to transmit a voice message. In the late 1930's, Homer Dudley invented the Vocoder as an ingenious way to put ten conversations over the single wire that usually carries only one. The principle deserves comment. The incoming speech is first analyzed into a number of components, these are processed for transmission, and then a speech output is synthesized from the components. Vocoderes have not come into common use, in part because they are not yet cost effective in dollars even though they do save bandwidth, but by combining the vocoder principle with digital processing of the analyzed signal—for example, by adding digital noise before transmission and subtracting just the same digital noise at the receiving end—it is possible to achieve a considerable degree of message security.
SPEECH RESEARCH:

Some Practical Goals and Accomplishments

---

Voice Communication

- At a distance: telephony
- At another time: recording
- With less noise: repeatering & digital methods
- With less bandwidth: vocoder
- With better security: privacy & secrecy systems
- With computers, for greater versatility:
  Answer-back systems
  Data retrieval
  Text-to-speech conversion
  Automatic translation
  Speech understanding systems

Aids for the Handicapped

- For the blind: text-to-speech reading machines
- For the deaf: hearing aids, tactile vocoders, cochlear or neural implants
- For quadriplegics: communications boards

Figure 1
But why stop with digital processing of such a limited kind? Why not put a whole computer between analyzer and synthesizer and then program it both to recognize what was said into the microphone, and to give an appropriate answer through the synthesizer? That is to say, why shouldn’t humans and computers talk to each other? It turns out in practice, as I am sure most of you know, that computers have a great deal of difficulty in understanding human speech. It is only a little less difficult to teach them to use natural language—English, for example—even when the input and output are in typewritten form.

The extreme difficulties of teaching computers to use ordinary human speech has prompted serious questions of whether there is any real reason why computers ought to learn the use of speech. One attempt to answer this kind of question was described in a delightful article by Chapin in the Scientific American about two years ago. Chapinis asked two people to cooperate in the task of putting together a mail order kit: the other person, in a different room, had the instructions for putting it together. They could communicate in various ways, but only one way, or combination of ways, was permitted in any one experiment. How long it took to assemble the device depended, as you might suppose, on the kind of communication that the experimenter allowed. The gross result of experiments on many modes of communication, some with speech and some without, was that when the task had to be done without voice it took about twice as long. In one respect, this is a gross underestimate of the difference, since the communication rate (as measured in words per minute) was about ten times as high with speech as with any other words. However, the teams wasted most of this advantage by using five times as many words when they talked as when they wrote, thus leaving a net factor of only two. The moral, for speech understanding systems, seems fairly clear: not only do people prefer to write a letter rather than to send it by voice, but they write much faster at it. Given man’s insistence on his own convenience, we can expect that spoken input/output to computers will eventually replace present methods, just as the telephone replaced the telegraph despite greater cost and complexity. The question of why speech can be so fast is one to which we shall return.

If we repackage this device-oriented system, we have a model such as psychologists often use in talking about human behavior (Fig. 2). Sometimes they ignore the central processor entirely and deal only with input-output relationships, though I do not suppose they would go so far as to deny the logical necessity for a certain amount of machinery between ear and mouth. Moreover, many present-day psychologists would insist on having at least the components indicated in the figure. But we can see what the problem is in using a model of this kind, whether for dealing with a speech understanding system or a human being: so many of the important components are totally inaccessible.

Perhaps two heads are better than one (Fig. 2c), since then the message is out in the open where we can capture and study it. But that is only one of the reasons for using this as a model for talking about strategies in speech research. The other reason lies in the dual nature of speech. It is this which prompts us to probe experimentally into the processes of reception (to the right) and the processes of production (to the left). At no other point in the total communicative process do we have such rich experimental opportunities.

Dual Nature of Speech

But why attribute a dual nature to speech in justification of these more or less obvious ways to approach perception and production? Do I mean to imply that the speech signal is unusual, i.e., that it is special in some important sense, in the degree of its dependence on the mechanisms by which it is produced and by which it is perceived? That is exactly what I do wish to imply, namely, that speech is not just a set of acoustic signals, but rather that it is an encoding of the message, in the cryptographic sense of the term. Moreover, the nature of the code is dependent on two sets of properties and characteristics: those of the processes that produce the signal and also those of the processes that receive and decode it. It follows from these considerations that our best hope for finding the message within the acoustic signal lies not in studying the signal itself but in trying to find out how it was encoded in production and how it is decoded in perception.

This is a point of view that has evolved in consequence of research on speech; certainly, it was not how speech was viewed as of the late ‘40’s and early ‘50’s. Then, the emphasis was on the acoustic signal. The objective was to locate acoustic invariants that could characterize phonemes and/or distinctive features. Much of this interest was sparked by the sound spectrograph, which emerged in the mid-‘40’s from war-time research at the Bell Telephone Laboratories. The spectograms revealed in full detail the complex patterning of speech sounds.

You might enjoy seeing what is, so far as I know, the very first spectogram ever made (Fig. 3). It was published by John Steinberg in 1934. One of the reasons the method did not
MODELS

A
Speech Processors

B
Speech Behavior

C
Speech Research

FIGURE 2
catch on then was that the Fourier analyses for
this one sentence required several hundred hours
of hand measurement and computation.

Early Research on Speech Perception

My own discovery of the wonders of speech
had a different origin more nearly related to the
remarkable speed with which speech sounds are
perceived. My colleagues and I had been trying
to build a reading machine for the blind. We
used photocells to convert the shapes of letters
(on the printed page) into distinctive acoustic
shapes, but no matter how hard we tried to find
an optimal set of sounds over long our subjects
worked to learn them, about the best they could
do was ten to twenty words per minute. That is
barely a tenth of normal rates for speaking and
listening. It took us a long time to understand
that the real problem was not why our signals
were so slow, but why speech was so fast.

We built the Pattern Playback as a way to
find the acoustic cues in the speech signal,
since it seemed obvious that there must be some
underlying pattern in the sound spectrogram
that served to carry most of the information.
The basic idea of the method was quite simple: one
used a paint brush to draw a simplification--
a caricature--of the real spectrographic
patterns. Then, by playing back these patterns,
that is, by reconverters them into dynamically
changing sounds, one could judge by ear whether
the visual simplification had in fact caught the
phonetic content. In spite of crudities, this
proved to be a powerful research method.
Initially, we worked with sentences. The original
spectrogram of such a sentence, and a simplified
pattern of the kind we often used, are shown in
Figure 4. (Recordings were played of the speech
synthesized from these patterns.)

The playback itself is, as I have said, a
rather simple-minded device. The diagram in
Figure 5 shows how it works: a line of light is
modulated by a tone wheel and is imaged on a
photocell that is connected with an amplifier and loudspeaker. The modulations that are selected for
conversion into sound depend on where the painted
areas are along the modulated line image from the
tone wheel; the low frequencies are at the bottom
of this line, the high at the top. The tone
wheel is a motor-driven disc about 20 inches in
diameter with the harmonics of 120 Hz, from 120
to 6000 Hz, recorded photographically in variable
density mode on fifty-tenth-inch bands. The
entire device is mounted on three lathe beds
bolted to a heavy table, and deserves characterization as American Gothic for the stark simplicity
of its construction.

I would like to draw your attention to some
interesting characteristics of speech that are
implicit in the fact that the playback is so
crude, and yet talks intelligibly. Obviously,
naturalness and good voice quality are not
essential to intelligibility, though maybe they
would help. Likewise, pitch inflections are non-
essential, since playback speech is a flat
monotone. This is entirely consistent with the
underlying principle of Homer Dudley's Vocoder,
namely, that the voice carrier, including its
pitch, serves one function, while the modulations
imposed on it by articulation serve another.
Linguists and phoneticians make a corresponding
distinction, though they may not attach as much
significance as they might to the near-indepen-
dence of the segmental and suprasegmental aspects
of speech. A third point, and perhaps the most
interesting of all, is the ability of these
highly simplified patterns to carry all of the
message, or, more accurately, as much of it as
we expect to find on a printed page.

Let me review for you, very briefly, some
eyearly experiments we made with the playback.10
Figure 6 summarizes several series of studies of
the acoustic cues for stop and nasal consonants.
In the experiments, we used a wide variety of
second-formant transitions, nasal resonances, and
first formant transitions and "cutbacks". The
nine patterns shown in the figure are the best
we could find for the nine consonants when paired
with the vowel [a], and they are here arranged
in rows and columns according to distinctive
acoustic characteristics. You can see that the
same 3 x 3 arrangement would have resulted if we
had used the familiar phonetic dimensions of
place and manner. But you should not assume that
the acoustic characteristics I have mentioned are
accurate invariants for these consonants because,
if they were, then one should find them intact
when the same consonants are paired with different
vowels. This is not so, as you can see in
Figure 7, where the transitions are quite
different from vowel to vowel, most notably those
of the second-formant for [a].

In working with these acoustic cues we often
observed a striking perceptual characteristic of
consonants that one might not have expected.
Thus, when we listened to a series of patterns
that progressed in small steps along one acoustic
dimension—for example, the extent of the second-
formant transition—we noticed that the perception
remained the same for several sounds, then
changed abruptly to a different perception,
and again to a third. This impression was well
founded because, when we asked our subjects how
well they could tell the difference between two
adjacent patterns, the answer was that they
could hardly do it at all (as between patterns
that were in the same category, that is,
patterns that had the same label), although they
were quite aware of an equal stimulus difference
at the boundary between categories. George
Miller once characterized categorical perception
of this kind by saying that phones, like coins,
rarely land on edge.

In describing the Pattern Playback, you may
have noticed without my mentioning it the
resemblance to the device-type model of Figure 2a.
For the Pattern Playback, the central processor
was a paint brush operating on a spectrogram and
the synthesizer was a tone wheel plus an
electro-optical system. At a later stage in our
work, we built and used an exact counterpart of a
<table>
<thead>
<tr>
<th>Frequency</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ga</td>
<td>da</td>
</tr>
<tr>
<td>ka</td>
<td>ta</td>
</tr>
<tr>
<td>ja</td>
<td>na</td>
</tr>
<tr>
<td>ba</td>
<td>pa</td>
</tr>
<tr>
<td>ma</td>
<td></td>
</tr>
</tbody>
</table>
vocoder and called it a Voback. Figure 8 shows
the Voback's own kind of spectrogram between ana-
yzer and synthesizer. The principal virtues of
this device were that it gave better quality
speech and it let us manipulate pitch as well as
spectrum. More recently, we have put together
a computerized version of this vocoder-type
playback. Figure 9 shows how it works and why
we built it: neither the Pattern Playback nor
Voback made it possible to work with real speech
with any degree of facility, but this Digital
Playback does. We have tried to retain the
simplicity and immediacy of the original playback,
insofar as the operator is concerned, though at
some considerable cost in terms of instrumental
complexity, as you can see in Figure 10, which
shows the major components of the system.12

I have mentioned only a few of the ways in
which speech perception can be studied by
manipulating the speech signal and then trying
to infer from the perceptual consequences what
the perceptual processes might have been. This
is a lively area of current research, as many of
you know.

Speech Production

Let me turn now to research on speech pro-
duction. You can understand what motivated us
to look in this direction by recalling the 3 x 3
array of stop and nasal consonants shown in
Figure 6. Originally, the patterns for the nine
consonants were arranged that way strictly on
the basis of acoustic regularities but, when this
was done, the row and column headings turned out
to be the familiar names for manner and place of
articulation. This is not really surprising, of
course, since one would expect production and
perception to be compatible. What is strange,
is that the acoustic features were not really
acoustic in nature—at least, they did not differ
along the usual acoustic dimensions of frequency,
intensity, duration, and the like. Indeed, very
few of the acoustic cues we found as we studied
the various sounds of English seemed to have any
intuitive relationship to acoustic dimensions or,
indeed, to be perceived as acoustic events that
happened to have phonetic names; rather, they
were phonetic events in the first instant. On
the other hand, all the cues could readily be
rationalized along articulatory dimensions. This
set us to wondering if perceptual decoding of
the speech signal might best be understood by
studying the encoding operations of speech
production.

We thought this approach might also help us
to understand why speech can go so fast. You
will remember that this was what lured us into
speech research in the first place. If one looks
with a fresh eye at the articulatory system, his
first impression would have to be that even Rube
Goldberg wouldn't have built a plumbing system
like this: the airway to the lungs and the food-
way to the stomach actually cross each other and,
because they do but must be kept functionally
separate, there is a whole array of valves to
control the traffic. But notice what this implies
for speech. If these valves—more generally,
these articulators—are controllable independently
then the individual components do not have to
move rapidly in order that the state of the system
(as determined by all of them collectively) can
change many times per second. Thus, correspond-
ing events in the acoustic signal can also
change rapidly, and we have the miracle of rapid
speech produced by slowly moving articulators,
simply because there are several of them and
they can operate in parallel. One could make the
same point about touch typing as compared with
the single-finger hunt-and-peck method. An even
better analogy is stenotyping.

As to speech perception, we have seen already
in the 3 x 3 array that the acoustic cues are
describable in terms of a set of more-or-less
independent features that are used in combination.
Thus, the speed with which perception operates
can also be understood in terms of the parallel
processing of slowly changing features encoded
into the sound signal. In short, if perception
somehow tracks production then we should be able
to understand several things about speech: why it
is so fast, why the encoding operations do not
place an intolerable burden on perception, and
why the acoustic cues for speech seem so
strangely unacoustic in their makeup.

But I started to consider strategies for
research on speech production. Here is a diagram
of some of the stages in the process (Fig. 11). We
know that the speech that comes out of the
articulatory process is encoded, though we do not
know the size or nature of the units that are
involved in that operation. We can guess, from
linguistic theory as well as from research on
speech, that segments of something like syllabic
length are near the upper limit for unitary
speech gestures and also approximate a lower
bound on the operations that belong in the
domain of linguistics.

Since we wished to learn as much as we could
about the organization of speech gestures, and
how and where speech blends into phonology, we
chose to work with electromyography because
it provides information about stages that lie as
high in the production chain as we can reasonably
expect to reach experimentally, and because it
lets us by-pass two low-level stages where we
can be sure there is no encoding. This
coding comes about because the movements of
the articulators depend in a complex way on the
pattern of muscle contractions, and this is true
not only on a moment-by-moment basis but across
some span of time so that the phonetic segments
are, in effect, merged into each other. Likewise,
the shape of the vocal tract at any instant
depends not only on how the articulators are
moving at that moment, but where they have been
and where they are going. These effects are
commonly "explained" as coarticulation, though
the emphasis this puts on phone-sized segments
may actually complicate the task of accounting
for the ongoing parallel operations of articula-
tion. But coarticulation can offer little by
way of explanation for components of a gesture
that anticipate their segmental roles in it—for
example, lip rounding of all the consonants
in an initial cluster before [j], though not
CHANNEL VOCODER

FIGURE 8
* Organization of the Speech Gesture(s)

* Neuromotor Commands to the Muscles

EMG

Muscle Contractions

* Articulatory Movements

X-Rays

Vocal Tract Shape & Excitation

Spectrograph & Ear

Speech Output

Note: *Levels at which description is most desired.

FIGURE 11
before [1]. Such things point to the need for a neural organization of the speech gesture that occurs above the level of muscle contractions and also above the neuromotor commands that these contractions reflect. A major question for research on production is, then, to discover what span of attention to attribute to the operation labeled "Organization of the Speech Gesture," and also to learn how much, or how little, reorganizing goes on there.

If a schema of this kind really does describe how we produce a spoken sentence, then linguistics has the problem of providing speech production with the kinds of input signals it can use. Whatever we can learn about such units by direct experimentation can serve as a guide and constraint to linguistic theorizing; also, it can help to provide a model of sorts for the linguistic operations required to assemble lexical items according to semantic needs and then put them into linear form so the sentence can be spoken. If, alternatively, the sentence is to be written, the requirements are somewhat different and so is the linguistic product.

At the left of the schema, I have indicated how different research methods can help us to understand the successive stages in articulation. These are not the only available methods by any means, though they are the most versatile. Constant efforts are being made to develop new methods; also comparatively new is the use, in combination, of EMG, X-rays, and speech analysis. The first two, in particular, need to be used together because muscle activity at any instant will depend on where the articulators already are at that instant and, conversely, the effects of muscle contractions are really not predictable from EMG signals but must actually be observed. Several laboratories across the country are toiling up to do just this kind of research.

As a final instance of research strategy, I may remind you of the rationale for the Pattern Playback: you control the synthesis and then observe the perceptual result. If we use an articulatory synthesizer in this way—moving the individual articulators in ways that test various theories about articulatory targets—we may be able to conduct a search for the articulatory cues that are effective in perception. This would be strictly analogous with the earlier search for the acoustic cues. It seems to us an exciting prospect.

Some Current Concerns

I have touched on some of the current concerns of speech research, though undoubtedly I have omitted many more. Let me comment briefly on three major questions. First, what is the nature of the relationship between perception and production? There is a great deal of evidence and general agreement about the existence of a significant relationship of some kind. The question is, what kind? I have tended here to emphasize the articulatory aspects of speech, and have at least implied that perceptual operations somehow make use of knowledge about how speech is encoded in order to decode it. I am sure you know that there are other highly respected points of view. Kenneth Stevens, for example, sees a relationship between the quantal states of production and the feature recognizing capabilities of perception and proprioception. In his view, these cooperate to localize the interpretation of speech signals within the sensory domain without need to involve the motor domain. It may, of course, turn out in the end that it is meaningless to ask whether the special perceptual mechanisms needed for speech are primarily sensory or primarily motor. For the moment, though, a lively exploration of differing points of view is underway.

A second question, closely related, has to do with the units in which the speech signal is organized in production and dealt with in perception. Put another way, it is a question of how much preplanning—and how much carryover—"belongs" in the speech component of language, and where the boundary is between the domains of speech, narrowly defined, and the rest of linguistics. A third question to which I have not so far alluded has to do with how language is acquired in the first instance. Are there operations and mechanisms that are important in language acquisition, though they may no longer be needed by adults for the everyday use of spoken language?

Summary

Let me recapitulate the main points that I have tried to make. The most general one is that speech is an integral part of language and therefore a proper concern of linguists as well as of people who call themselves speech scientists or speech researchers. Remember, please, that speech is a continuous process from sentence formulation through production and perception to comprehension by the listener. There are, at each stage in passing a message down this chain, encoding operations that have much in common, if indeed they do not have a common pattern. The efficiency of spoken language, as well as the challenge it offers to us as scientists, follows from the fact that we are dealing at every level with coded messages.

At the level of the speech signal, this code can best be understood in terms of the processes that have shaped it in production and that, in perception, must recover the message. This is the basis for an overall strategy that uses speech as a target of opportunity. From this vantage point, research can proceed downstream toward perception, making good use of the kind of analytic-synthesis methods typified by the Pattern Playback; also, research can be carried upstream quite some distance, mainly by psychological methods.

The goal of all these strategies and tactics is to understand the nature of the message in its spoken form and how this relates to the overall linguistic processes within which speech
is embedded. It is an exciting field for cooperative research. So let me end by inviting you to share in it.

May you have as good luck in the future as Dr. Fletcher and Dr. Stevens—and I, too—have had in the past.

---

**Figure Captions**

**Figure 1:** Some engineering aspects of speech research.

**Figure 2:** Models and strategies. (A) A general-purpose model for communications devices. (B) Adaptation of (A) to the human as a communications entity. (C) Communication between two humans and research strategies such a model suggests.

**Figure 3:** The first spectrogram of the sentence "Joe took father's shoe bench out," as published by John Steinberg in 1934.

**Figure 4:** Spectrograms used with the Pattern Playback, about 1950. At top, an original photographic spectrogram; at bottom, a hand-painted simplification for the same sentence.

**Figure 5:** Functional diagram of the Pattern Playback. The Light Collector, connected to an amplifier and loudspeaker, could be positioned either below the belt carrying a spectrogram (as shown) for use with transmission spectrograms, or above the belt for use with the hand-painted reflection-type spectrograms that were used in most of the research. The tone wheel, driven by a synchronous motor, provided fifty light beams modulated at harmonics of 120 Hz to 6000 Hz.

**Figure 6:** Nine two-formant patterns found to be optimal for the stop and nasal consonants when paired with the vowel [a]. The arrangement by rows and columns is according to similarities in the formant transitions; the individual patterns have the usual spectrographic dimensions of time and frequency.

**Figure 7:** Two-formant patterns for the voiced stops with each of three vowels, showing how the transitions of the second formant differ for the same consonant when paired with different vowels.

**Figure 8:** A channel vocoder, showing how the information flowing from analyzer to synthesizer is representable as a spectrogram and therefore modifiable for experimental purposes, just as with the Pattern Playback. A device called Voback implemented this approach, using photocells to "read" hand-painted spectrograms into the vocoder synthesizer; separate control channels carried information about pitch and buzz/hiss switching.

**Figure 9:** The Digital Playback is designed to serve the same experimental functions as the original Pattern Playback, but to do it conveniently for real speech.

**Figure 10:** Components and control facilities for the Digital Playback. Real-time analysis equipment stores a few seconds of speech spectrum and waveform in computer core, whence it can be called for faster display as a spectrogram or to recreate the speech by synthesis. Easy modifications can be made to the contents of core memory, and the results can be seen in a comparison spectrogram or heard, as speech synthesized from that spectrogram.

**Figure 11:** Schema for speech production, with notations (on the right) about the extent to which successive transformations encode the message, and (on the left) about major experimental approaches to the speech process.
Notes and References

1. Fletcher, H. (1929) Speech and Hearing
   (New York: van Nostrand).

   Bell System Techn. J. 19, 495-515.

3. See, for example, Klatt, D.H., Review of
   the ARPA speech understanding project,
   to appear in J. Acoust. Soc. Am.; also
   Medress, M.F., et al., Speech Under-
   standing systems; report of a steering
   committee, to appear in SIGART Newsletter
   (ACM), no. 62 (April 1977).

4. Chapanis, A. (1975) Interactive human com-

5. Cooper, F.S. (1972) How is language conveyed
   by speech? in Language by Eye and by Ear,
   edit. by J.F. Kavanagh and I.G. Mettingly

6. Steinberg, J.C. (1934) Application of sound
   measuring instruments to the study of
   phonetic sounds. J. Acoust. Soc. Am.,
   5, 16-24.

   machines for the blind. In Blindness:
   Modern Approaches to the Unseen Environ-
   ment, edit. by P.A. Zahn (Princeton, N. J.:

   speech analysis and synthesis.

   search of the acoustic cues. In
   Mélange à la mémoire de Pierre Delattre,
   (Available also in Haskins Laboratories
   Status Report on Speech Research,
   SR-19/20 (1969), 9-26.)

10. Cooper, F.S., P.C. Delattre, A.M. Liberman,
    Some experiments on the perception of
    synthetic speech sounds. J. Acoust.
    review.)

11. Liberman, A.M., F.S. Cooper, D.P. Shankweiler
    and M. Studdert-Kennedy (1967) The per-
    ception of the speech code. Psychol.
    Rev., 74, 431-461. (A mid-term review
    and interpretation; references.)

    Phonetic perception. To appear in
    Handbook of Sensory Physiology, Vol.VIII,
    "Perception", edit. by R. Held,
    H. Leibowitz, and H.L. Teuber (Heidelberg:
    Springer-Verlag). (Review and references.)


    P. Mermelstein and T. Montlick (1976)
    A digital pattern playback for the
    analysis and manipulation of speech
    signals. Haskins Laboratories Status

15. Cooper, F.S. (1965) Research techniques and
    instrumentation: EMG. In Proceedings
    of the Conference on Communicative
    Problems in Cleft Palate. ASHA Reports,
    1, 53-168.

    of Articulatory Behavior. In Current
    Trends in Linguistics, 12, edit. by

17. Harris, K.S. Physiological Aspects of Speech
    Production. In Implications of Basic
    Research in Speech and Language for the
    School and Clinic (working title: to be

   (The papers by Harris can be found also
   in Haskins Laboratories Status Reports
   on Speech Research, SR-23 (1970), 49-87,
   and SR-48 (1976) 21-42.)


[The demonstration recording played at the end of the talk was excerpted from a soundsheet that was bound into IEEE Transactions on Audio and Electroacoustics, Vol. AU-21, No.3, (June 1973)].