PERCEPTION OF TEMPORALLY-SEGMENTED SPEECH*

A.W.F. HUGGINS

A common way of studying speech perception is to find transformations of the speech wave that drastically interfere with its intelligibility, and then to try to discover how the distortion has its effect. Presumably, the perceptual apparatus relies heavily on the information that has been destroyed.

In 1954, Cherry (Cherry and Taylor 1954) discovered that running speech can be made virtually unintelligible by switching it alternately to the left and right ears of listeners at about 3 cps. Higher or lower rates had little effect. Here was a transformation whose main parameter was DURATION, which had its most dramatic effect when the speech intervals reaching each ear of the listener lasted about the duration of a syllable. Further work suggested that intelligibility was destroyed because the speech reached each of the listeners' ears in bursts, separated by silence (Huggins 1964). The present experiments further tested this idea.

The 'temporal segmentation' of the speech was performed with the aid of a computer. The operation is equivalent to cutting a tape carrying the message into pieces, and splicing in a silent interval at each cut. Two sets of nine 100-word experimental passages of speech were cut into 'intervals' whose duration increased in nine log steps from 31 msec in the first passage in each set, to 500 msec in the last. Three experimental tapes were then made from each set of passages. In the three tapes, labelled 'short', 'equal' and 'long', the silent intervals were 41% shorter than, equal to, and 83% longer than the adjacent speech intervals, respectively. The ONLY difference between the tapes was the duration of the silent intervals that were spliced in.

The advantage of these materials is (1) the speech reaches the listener in bursts, as required, (2) no switching of attention is required (unlike alternation), (3) all the speech reaches the listener (unlike interruption), and (4) silent intervals and speech intervals can be independently varied.

One group of sixteen subjects shadowed the 'short' and 'equal' tapes, and a second group shadowed the 'equal' and 'long' tapes, with appropriate counter-balancing. Those subjects whose first exposure to the material was the 'long' tape showed a

* This research was supported by National Institutes of Health Grant No. NS04332.

learning effect over the first four passages. Therefore, the data from the first tape encountered by these subjects were discarded.

In Figure 1, intelligibility is plotted as a function of the duration of the SPEECH intervals. The left hand side of the three sets of data seem to lie on a single function, as if intelligibility progressively decreases as the duration of the speech intervals decreases. However, the minima occur at different speech interval durations for the 'short', 'equal' and 'long' functions. The recovery from the minimum occurs at progressively SHORTER speech-interval durations as the silent intervals are lengthened from one tape to another. Perhaps the recovery is also described by a single function depending only on the durations of the silent intervals. To test this possibility, the data are replotted in Figure 2 as a function of SILENT interval duration. This brings the RIGHT hand side of the curves into agreement, as if the intelligibility progressively increases as the duration of the SILENT interval decreases (the reverse from the speech case: a similar conclusion was recently reported in Powers and Speaks 1971).

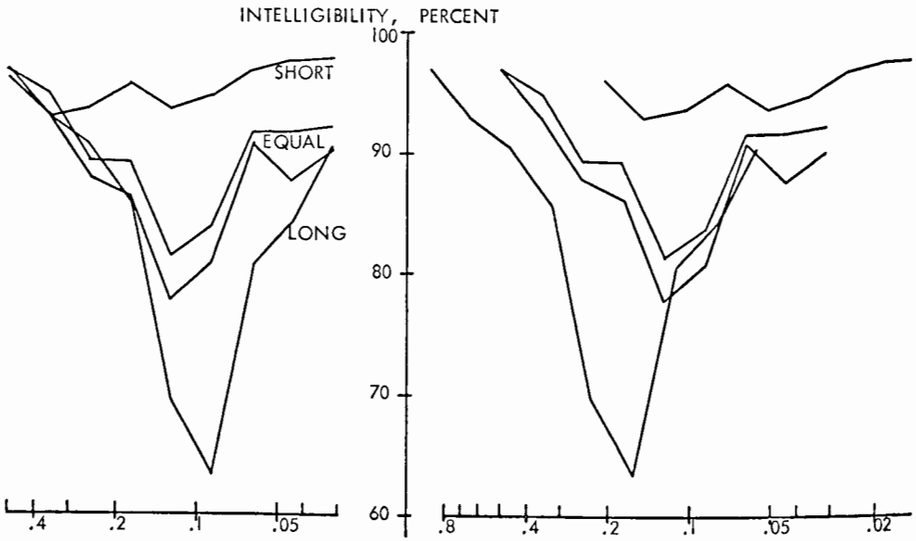

Fig. 1. Speech interval duration, seconds.

Fig. 2. Silent interval duration, seconds.

This analysis suggests that the minimum of the Cherry effect is only an artifact, resulting from the fact that in most experiments on alternated or interrupted speech, the speech-silence ratio is unity.

What might the two functions represent? Consider first the decline of intelligibility, at the left side of Figure 1. Here, long speech intervals are separated by long silences. As the duration of the speech intervals become shorter, they become less intelligible. But this is just a restatement of a finding by Pickett and Pollack (1964): brief excerpts of fluent speech become increasingly intelligible as they become longer. Their data

lie somewhat to the left of the data in Figure 1 — that is, longer excerpts were required for a fixed intelligibility — but their task was different, too. Their excerpts were presented in isolation, and always consisted of a small number of whole words, and their responses were also limited to whole words.

Extrapolating down the left hand side of the curve in Figure 1 points to a critical minimum sample duration of 60-70 msec for speech — or, more likely, one sound segment, since other work has suggested that it is the speech content of the interval rather than its duration that is critical (Huggins 1964).

What about the recovery shown in Figure 2? Work with trains of pulses (Huggins 1969) has suggested that silent intervals shorter than about 100 msec are integrated as part of an acoustic event (or sequence), whereas longer intervals are not, but act to break the sequence up into separate events, separated by pauses. If a sequence of speech samples, each too short to be recognized in isolation, can be integrated into a single ongoing acoustic event, then the samples may again become recognizable (that is, when intervening silences are long enough that successive samples cannot be integrated into a single event, then they may remain unrecognizable). Extrapolating down the right hand side of the curves in Figure 2 points to a critical maximum silent interval of about 200-250 msec (which corresponds approximately to the duration of a syllable).

The foregoing analysis is supported by subjective impressions from listening to the tapes. When the silent intervals are short (i.e., at rates above the critical), the speech sounds as if it is being played at reduced speed. At the critical rate, the speech sounds very broken up. At slower rates, words and phrases are heard, separated by pauses.

The trouble with this explanation is that it should presumably apply to all sounds, not only speech, and thus represent a temporal parameter of the ear. However, as mentioned above, where speech is concerned, the critical parameter seems to be the CONTENTS of the speech sample (i.e., how many syllables, glottal cycles, etc., it contains) rather than its duration.

This conflict will have to be resolved by further work.

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts

REFERENCES

Cherry, E.C. and W.K. Taylor
1954 "Some Further Experiments Upon the Recognition of Speech, with One and with Two Ears", *Journal of the Acoustical Society of America* 26:554-559.

Huggins, A.W.F.
1964 "Distortion of the Temporal Pattern of Speech: Interruption and Alternation", *Journal of the Acoustical Society of America* 36:1055-1064.

Huggins, A.W.F.
1969 "Perceived Rate of Dichotically Alternated Clicks", *Journal of the Acoustical Society of America* 46:88(A).
Pickett, J.M. and I. Pollack
1964 "Intelligibility of Excerpts from Fluent Speech: Effects of Rate of Utterance and Duration of Excerpt", *Language and Speech* 6:151-164.
Powers, G. and C. Speaks
1971 "Intelligibility of Temporally Interrupted Speech", *Journal of the Acoustical Society of America* 50:130(A).

DISCUSSION

NOOTEBOOM (Eindhoven)
I wish to thank you for your paper. I think that the experimental technique you used is an improvement on previous ones. As you stated, however, that you hoped to learn more about the syllable from such experiments, one might think of a somewhat different technique, such as the one used by my colleagues A. Van Katwijk and J. t' Hart (Intelligibility of Syllable-Tied Interrupted Speech I.P.O. Report April: 99-102 [1967]). They prepared several versions of a slowly spoken text of over 1000 syllables in a way that in each syllable a gap was present, the position of which was related to the vowel onset. It was found that the intelligibility of the speech depended on the position of these gaps. Intelligibility was the worst when the CV transition was missing.

HUGGINS
Thank you for your comments. I too have done some work with speech that was switched at particular events in the speech wave, but my experiments were on alternated speech. (See A.W.F. Huggins, "Distortion of the Temporal Pattern of Speech by Syllable-Tied Alternation", *Language and Speech* 10:133-140 [1967]). I found insignificant differences in intelligibility between speech alternated at (1) every syllable boundary, (2) in the middle of every vowel, (3) at every CV junction, (4) at every VC junction, and (5) *both* at every CV junction AND at every VC junction. This result surprised me very much: the discrepancy with your colleagues results may be due to differences between interruption and alternation. I am at present running some experiments on syllable-tied interrupted speech using one channel only of my two-channel alternation tapes, but I have not collected enough data to permit me to comment.

LEHISTE (Columbus, Ohio)
I would like to ask you to clarify your methodology. You have talked about speech being switched from one ear to the other; you have experimented with speech from which certain time segments were systematically removed; and now you are talking about speech into which pauses were introduced, while no part of the signal was discarded. It is not clear to me how these three procedures relate to each other and the present experiment.

HUGGINS
In an earlier paper, I showed that intelligibility scores for ALTERNATED speech could be quite accurately predicted from the LOGICAL sum of the scores for the two complementary interrupted signals, which alternation subjects heard one in each ear. This showed that alternation and interruption had their effects on intelligibility for the same reason: that the speech reached the listener in bursts, separated by silence. The experiments I reported today achieved the same effect in a new way, and here too there is a dip in intelligibility at about 3 cps, quite similar to that shown by alternated speech. This similarity, together with the fact that temporally-segmented speech showed the predicted dip in intelligibility, leads me to think that all three distortions have their effect in the same way.

GERBER (Santa Barbara Calif.)
In our laboratory we have been concerned with the time compression of speech. In effecting time compression it is necessary to discard speech samples. We have the means to recover the 'discarded' segments and play them in one ear while playing the 'remaining' segments in the other ear. Even for speech compressed in time as much as fourfold, word intelligibility remains above 80%. The finding is consistent with the data of Dr. Huggins on interrupted speech and supports his conclusions.

HUGGINS
Thank you for your comment. A critical aspect of both interrupted and temporally-segmented speech seems to be the insertion of silent intervals which serve to delimit and separate the speech intervals. On the other hand, there are NO silent intervals in compressed speech, and (non-adjacent) speech segments are usually abutted in such a way as to conceal the fact that some of the signal has been removed. Thus, interrupted speech seems to differ significantly from time-compressed speech, and I would want to be very cautious before drawing parallels between them — but I must also confess that I am not familiar with the results you mention.