

## Word rate and intelligibility of alternated speech

ARTHUR WINGFIELD and JOHN L. WHEALE  
*Brandeis University, Waltham, Massachusetts 02154*

A continuous speech message alternated between the left and right ears retains generally good intelligibility, except at certain critical rates of alternation of about 3-4 switching cycles/sec. In the present experiment, subjects heard speech alternated between the two ears at eight different switching frequencies, and at four different speech rates. Results support an earlier contention that the critical intelligibility parameter in alternated speech is average speech content per ear segment, rather than absolute time per ear. Implications are discussed both in terms of critical speech segments in auditory analysis and in neural processing of binaural auditory information.

Broadbent's concept of channel separation and selective attention in simultaneous listening has undergone considerable modification in the two decades since its introduction (Broadbent, 1954, 1958, 1971). The original notion argued that perceptual analysis is limited to only one input "channel" at a time, with physical parameters of the speech signal primarily serving both to distinguish the channels and to guide their selection. This emphasis on physical parameters was weakened by demonstrations that semantic variables can override source of arrival in message segregation and selection (cf. Treisman, 1960). Syntactic and prosodic variables can also override source of arrival in perceptual segmentation of a single message switched periodically between the two ears (Wingfield & Klein, 1971). As the original notion of the channel has been questioned, so too has been the principle of a perceptual filter requiring a finite time to switch from one stimulus input to another (cf. Treisman & Geffen, 1967; Wingfield & Byrnes, 1972).

The perception of alternated speech is an interesting test of the initial filter hypothesis, since that model would make certain straightforward predictions in this singular case of intelligibility loss in the absence of competing inputs. When a single message is alternated rapidly between the two ears, intelligibility progressively declines with increasing rate, up to about 3-4 switching cycles/sec (167 to 125 msec per ear). Paradoxically, intelligibility then improves as switching rates are further increased beyond this point (Cherry & Taylor, 1954; Huggins, 1964; Schubert & Parker, 1955).

Cherry and Taylor's (1954) early explanation of this V-shaped function, although independently derived, can be seen as an exact extension of the principles of Broadbent's model to this phenomenon. At each point

of alternation, cessation of speech in one ear signals the subject to shift attention to the other ear where the speech now continues. If each shift of attention requires some finite time during which no usable information is available from either ear, the more frequent the interruptions, the greater will be the cumulative loss of acoustic information. Minimal intelligibility is reached when the subject's attention is completely out of phase with the shifting signal: by the time attention has been switched from one ear to another, the message has already returned to the original ear. At rates beyond this point, the subject begins to adopt a strategy of attending to the interrupted signal from one ear only, relying on the redundancy of the large number of small speech segments to accurately reconstruct the message (Miller & Licklider, 1950).

Huggins (1964) later used a variable-speed tape recorder to present alternated materials at two different speech rates. Increasing the speech rate produced minimal intelligibility at a shorter time-per-ear duration, suggesting that the critical factor is not switching frequency, so much as the size of the speech sample typically contained in each ear segment. Huggins related the point of minimal intelligibility to mean syllabic rate of the recorded speech, in line with more recent studies suggesting the critical role of syllable size units in ordinary speech perception (Liberman, Mattingly, & Turvey, 1972; Massaro, 1972; Savin & Bever, 1970; Warren, 1971). If correct, this would militate against any fixed-time account of alternated speech, whether based on time-dependent attention switching mechanisms (Broadbent, 1958, 1971; Cherry & Taylor, 1954) or on perceptual order confusions of temporal segments of a certain size (cf. Broadbent & Ladefoged, 1959; Warren, Obusek, Farmer & Warren, 1970).

Although Huggins' results received early and wide acceptance (cf. Broadbent, 1971; Massaro, 1972; Moray, 1969; Neisser, 1967), the issue may have been prematurely closed. His study, like that of Cherry and Taylor, employed "shadowing" as a response measure. This involves repeating what is heard as it is heard,

This research was supported by PHS Grant NS-09767-01 from the National Institutes of Health to the first author. We thank Dr. Emerson Foulke, Center for Rate Controlled Recordings, University of Louisville, for assistance in time compression of the speech materials.

such that the subject must speak and listen simultaneously. The problems of overload and of peripheral masking complicate its use as a simple intelligibility measure (Underwood & Moray, 1971). Further, the effects of speech rate reported by Huggins were not found by Cherry and Taylor (1954). The argument remains an important one, especially in view of more recent work which raises again the possibility of interaural attention shifts to account for this and related phenomena (Guzy & Axelrod, 1972; Treisman, 1971).

The present study reexamines effects of word rate on perception of alternated speech, but with certain methodological improvements. First, four different speech rates are used. This is made possible through time-compression of the speech materials so as to accelerate word rate without major sacrifice of sound quality (Foulke, 1971). Second, the shadowing requirement used in previous studies is eliminated by inserting periodic silent intervals into the speech passages. This technique allows the subject time to repeat a message segment before the arrival of more speech, and also permits the use of higher speech rates than would ordinarily be possible with shadowing as a response.

The use of compressed speech thus allows one to vary independently the speech segment duration vs. the speech content in these durations. If the critical parameter is absolute alternation rate, the points of minimal intelligibility for all speech rates should be aligned on the abscissa at a single point representing the same time per ear. If the amount of speech contained in each ear segment is the critical parameter, the points of minimal intelligibility should shift on the abscissa by amounts approximately proportional to the degree of time compression. Thus, the position of the minima would vary with absolute speech rate, but remain constant with the size of the speech segments.

## METHOD

### Stimulus Materials

The stimuli consisted of 40 120-word passages selected from recent issues of the *American Scientist*. The passages covered a wide range of scientific subjects, and were selected to offer understandable, but high-information, speech samples. Proper names and technical terms were eliminated from the passages. Preliminary studies matched the passages for difficulty and intelligibility.

The passages were recorded by a speaker of American English at a rate of 185 words/min (192 msec/syllable). This recording was then compressed using the sampling method on an electromechanical compressor of the Fairbanks type. In this method, small segments of the recorded speech are periodically deleted and the remaining segments are abutted in time. The discard intervals were kept at 20 msec, and the degree of compression was varied by the frequency with which the tape segments were deleted.

Compressions were to 80%, 70%, and 60% of the original playing time, corresponding to speaking rates of 231, 264, and 308 words/min (154, 134, and 115 msec/syllable). Together with the originals at normal speech rate, a total of 160 passages at four

separate rates were thus available for the experiment. To eliminate the later need to respond while listening, each tape was processed to insert 4-sec silent intervals between every 10 words of the passages.

### Apparatus and Procedures

Alternation of the speech signal was accomplished by passing the output of a monaural tape recorder into a Grason-Stadler Model 829E electronic switch which operated to transfer the speech signal back and forth between the two earpieces of stereophonic earphones. The time phase of the switch was controlled externally by two Hunter Model 131C digital timers.

During the course of the experiment, each subject heard a total of 40 different passages, 10 at each of the four speech rates described. In each block of 10, the first 2 passages were taken as practice. The remaining 8 experimental passages were alternated between ears at the following periodic intervals: 250, 180, 160, 128, 112, 96, 62, and 31 msec per ear. Both the order of presentation of the compression ratios and the alternation rates were varied between subjects. The particular passages heard at each rate and ratio were also varied. By the end of the experiment, all passages were heard an equal number of times at each compression ratio and at each alternation rate.

The subjects were told they would hear speech samples in which the signals would alternate between ears, and to repeat as many words as they could, using the silent periods between each 10-word segment. The subjects monitored the passages over Sharp HA-10 dichotic earphones. The speech signal at the earphones was 75 dB SPL (re: .0002 dynes/cm<sup>2</sup>). Sixteen undergraduate volunteers served as subjects. All reported normal hearing, and all spoke American English as their first language.

## RESULTS

Intelligibility was taken as the number of words correctly reported from the middle 100 words of each passage. The first two and the last 10-word segments of each passage were not scored in order to eliminate end effects. Credit was given for all words reported correctly, regardless of order. No partial credit was given for incomplete words.

The results are summarized in Figure 1, which shows mean intelligibility scores as a function of alternation rate for each of the four speech rates tested. The abscissa values are plotted on a logarithmic scale. The small arrows under the curves show the obtained minima for normal speech rate (128 msec per ear) and the predicted minima for the three compression curves. The latter values were derived by taking 80%, 70%, and 60%, respectively, of the obtained value for normal speech rate.

Although there is some random variation in intelligibility for the four curves, it is clear that (a) overall intelligibility declines with increasing time compression, especially beyond 80% compression, and (b) the points of minimal intelligibility show a systematic shift to the right with increasing compression, with the obtained and predicted minima corresponding as closely as possible given the alternation rates tested.

Analysis of variance was performed following an arcsin transformation to stabilize the variance of the percentage scores. The overall effect of compression ratio was significant,  $F(5,60) = 50.70$ ,  $p < .01$ , as was the effect of alternation rates,  $F(9,58) = 3.01$ ,

$p < .01$ . The shift in the minima of the intelligibility curves was reflected by a small, but significant, Compression by Alternation interaction,  $F(21,294) = 1.66$ ,  $p < .05$ .

## DISCUSSION

Cherry and Taylor's (1954) notion of attention shifts and corresponding loss of acoustic information to account for intelligibility loss in alternated speech receives little support from these data. With this, we must also reject any other explanation based entirely or primarily on a fixed-time hypothesis. While overall intelligibility might drop with compression, the position of the minima for each curve would remain unchanged on the abscissa. This would be so because the number of alternations, and hence attention shifts per unit time, would be the same regardless of word rate. This is clearly not the case.

These results do support Huggins' (1964) position, that the critical determiner of intelligibility is the speech content contained in each alternation segment. As compression increases the amount of speech content per segment, the curves shift systematically along the axis by an approximately equivalent amount. Although these results hold without reliance on shadowing as a response measure, the typical V-shaped function becomes marked only at the higher compression ratios. Thus, it would appear that the subject must be under some load for the intelligibility loss to occur; if the burden of shadowing is removed, one must increase the input rate to sustain the effect.

It may be possible to account for the main results without recourse to attention switching mechanisms and attendant loss of acoustic information with alternation. Such an explanation would lie in the neural processing of binaural auditory information. The perception of alternated speech must involve, at some stage, a central integration of the separate ear segments presumably in the dominant speech hemisphere (Milner, 1962; Studdert-Kennedy & Shankweiler, 1970; Zurif & Sait, 1969). If the contralateral hemispheres to each ear conduct some preliminary analysis of their own input prior to integration, this would carry two implications. First, left-ear inputs in right-handed subjects would necessarily receive some analysis in the nonspeech hemisphere. Second, analysis of each ear segment would be independent of the surrounding phonemic context known to be of critical importance (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Either of these assumptions would lead to higher level processing based on attempted left-hemisphere integration of incompletely, or erroneously coded, segments. What has been described as a V-shaped intelligibility function might thus be better described as a single log-linear dimension: the closer alternation rates approximate critical segmentation, the poorer will be the overall intelligibility of the speech signal.

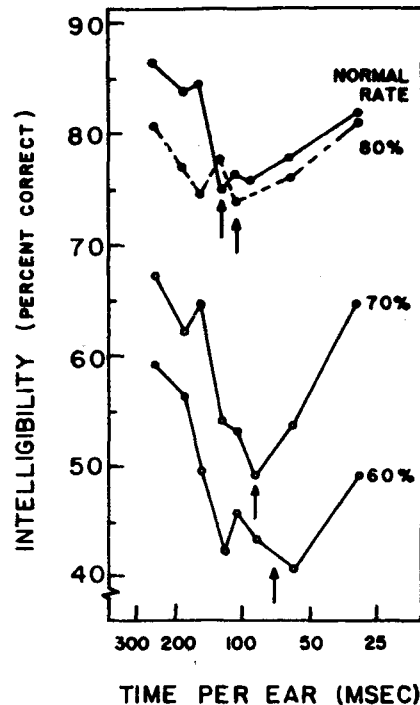


Figure 1. Intelligibility as a function of time per ear in alternation for normal and time-compressed speech. Small arrows show predicted intelligibility minima based on speech-dependent hypothesis.

The view that the size of the speech segment per ear is the critical variable suggests a reanalysis of prior studies in terms of speech content, rather than time per ear. The point of poorest intelligibility for the data of Cherry and Taylor (1954), Huggins (1964), and Schubert and Parker (1955) all lie around .36 words, or about .60 syllables (Huggins, 1964; Neisser, 1967). Our own figures are close to this. The mean values across all speech rates correspond to .39 words (range = .32 to .43) and .66 syllables (range = .54 to .73). While testing at larger numbers of alternation rates would help to define the duration of the critical segments with greater accuracy, there is an upper limit on accuracy bounded by the ability to define a "syllable," or any other speech unit, in temporal terms. Certainly, past attempts at applying alternated speech to this end have met with mixed success (cf. Huggins, 1967; Rupf, Hughes, & House, 1971).

While the nature of the "unit" of speech perception remains at issue, two more fundamental questions arise from the assumption that alternation operates to transform the speech pattern into two parallel streams of isolated speech segments. The first question is why silent periods of certain durations cannot be bridged by the perceptual system (Huggins, 1972; Miller & Licklider, 1950), and second, the implications of possible independent channel analysis for the ordinary binaural processing of speech. It is to these issues that further research must be addressed.

## REFERENCES

- BROADBENT, D. E. The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 1954, **47**, 191-196.
- BROADBENT, D. E. *Perception and communication*. New York: Pergamon, 1958.
- BROADBENT, D. E. *Decision and stress*. New York: Academic Press, 1971.
- BROADBENT, D. E., & LADEFOGED, P. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 1959, **31**, 1539.
- CHERRY, E. C., & TAYLOR, W. K. Some further experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, 1954, **26**, 554-559.
- FOULKE, E. The perception of time compressed speech. In D. L. Horton & J. J. Jenkins (Eds.), *The perception of language*. Columbus, Ohio: Merrill, 1971.
- GUZY, L. T., & AXELROD, S. Interaural attention shifting as response. *Journal of Experimental Psychology*, 1972, **95**, 290-294.
- HUGGINS, A. W. F. Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, 1964, **36**, 1055-1064.
- HUGGINS, A. W. F. Distortion of the temporal pattern of speech by syllable-tied alternation. *Language and Speech*, 1967, **10**, 133-140.
- HUGGINS, A. W. F. Perception of temporally-segmented speech. In A. Rigault & R. Charbonneau (Eds.), *Proceedings of the Seventh International Congress of Phonetic Sciences*. The Hague: Mouton, 1972.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431-461.
- LIBERMAN, A. M., MATTINGLY, I. G., & TURVEY, M. T. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. New York: Wiley, 1972.
- MASSARO, D. W. Perceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, **79**, 124-145.
- MILLER, G. A., & LICKLIDER, J. C. R. The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 1950, **22**, 167-173.
- MILNER, B. Laterality effects in audition. In V. B. Mountcastle (Ed.), *Interhemispheric effects and cerebral dominance*. Baltimore: Johns Hopkins Press, 1962.
- MORAY, N. *Listening and attention*. Baltimore: Penguin Books, 1969.
- NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
- RUPF, J. A., HUGHES, G. W., & HOUSE, A. S. Speech-synchronized versus periodic interaural switching of speech. *Journal of the Acoustical Society of America*, 1971, **49**, 608-610.
- SAVIN, H. B., & BEVER, T. G. The non-perceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 295-302.
- SCHUBERT, E. D., & PARKER, C. D. Addition to Cherry's findings on switching between the two ears. *Journal of the Acoustical Society of America*, 1955, **27**, 792-793.
- STUDDERT-KENNEDY, M., & SHANKWEILER, D. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, **48**, 578-594.
- TREISMAN, A. M. Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 242-248.
- TREISMAN, A. M. Shifting attention between the ears. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 157-167.
- TREISMAN, A. M., & GEFFEN, G. Selective attention: Perception or response? *Quarterly Journal of Experimental Psychology*, 1967, **19**, 1-17.
- UNDERWOOD, G., & MORAY, N. Shadowing and monitoring for selective attention. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 284-295.
- WARREN, R. M. Identification times for phonemic components of graded complexity and for spelling of speech. *Perception & Psychophysics*, 1971, **8**, 345-349.
- WARREN, R. M., OBUSEK, C. J., FARMER, R. M., & WARREN, R. P. Auditory sequence: Confusion of patterns other than speech or music. *Science*, 1970, **164**, 586-587.
- WINGFIELD, A., & KLEIN, J. F. Syntactic structure and acoustic pattern in speech perception. *Perception & Psychophysics*, 1971, **9**, 23-25.
- WINGFIELD, A., & BYRNES, D. L. Decay of information in short-term memory. *Science*, 1972, **176**, 690-692.
- ZURIF, E. B., & SAIT, P. E. The role of syntax in dichotic listening. *Neuropsychologia*, 1970, **8**, 239-244.

(Received for publication December 7, 1974;  
revision accepted June 15, 1975.)