Stereophonic Listening and Speech Intelligibility against Voice Babble*

IRWIN POLLACK AND J. M. PICKETT
Operational Applications Laboratory, Air Force Cambridge Research Center, Bolling Air Force Base 25, D. C.
(Received September 24, 1957)

The utilization of a particular form of stereophonic information was studied with respect to the intelligibility of a single speech source heard in the presence of a voice babble of other speakers. The reception of monosyllabic words, presented agsinst a babble of 1, 2, 4, or 7 talkers, was compared under two listening conditions: (1) In the *stereophonic* listening condition, one set of background talkers was presented to one earphone, another set of background talkers was presented to the other earphone, and the test words were presented binaurally, in phase; (2) In the *control* listening condition, only a single set of background talkers, and the test words, were presented to a single ear. The stereophonic advantage, for 50% word intelligibility, ranged from 12 db with 1 background voice per channel to 5.5 db with 7 background voices per channel.

IT has long been felt that directional information of a talker may improve the intelligibility of his speech heard against the voice babble of other talkers.[1] This effect has sometimes been referred to as "the cocktail party effect."†

The present study examines the role of binaural directional information in listening to speech against a background of voice babble. We shall examine a situation which appears extreme in direct person-to-person communication, but is entirely feasible in an interphone system utilizing binaural information.[2]

PROCEDURE

Groups of 1, 2, 4, or 7 talkers, simultaneously reading newspaper text at a rapid rate, were recorded upon a single channel of a two-channel tape recorder. Other groups of 1, 2, 4, or 7 talkers were recorded upon the other channel. The newspaper material served as the "background" against which the relevant messages were presented.

The relevant messages were the Harvard PB monosyllabic word lists, consisting of 1000 words, read by a single recorded speaker.

The relevant messages and the background materials were presented under two conditions: In the *stereophonic* condition, one set of background talkers was presented to one earphone; the other set of background talkers was presented to the other earphone; and the relevant messages were presented stereophonically, in phase, to both earphones. The *control* condition was identical with the stereophonic condition except that one earphone was disconnected. Thus, the relevant messages were presented against only a single set of background talkers over a single earphone in the control condition. It may be noted that this comparison is conservative because, in the control monaural listening condition, the total number of different background talkers was reduced by a factor of two. It may be noted that presentation of the monaural control condition over two earphones, rather than one, does not substantially modify the intelligibility scores. That is, the relevant factor which distinguished between the binaural stereophonic condition and the monaural control condition was the stereophonic presentation of the desired speaker and not the monaural presentation of the control condition.

In the experimental tests, the speech level of the relevant messages was held constant and the level of the background materials was varied over a wide range of conditions.

RESULTS

The results of the intelligibility tests are presented in Figs. 1–4. The ordinate is the average percentage of words correctly reproduced by the testing crew; the abscissa is the speech-to-background $(S/B)_0$ ratio (the subscript *o* indicating that the background was measured in terms of the *over-all* VU meter reading for the group of talkers). The sectional entry indicates the number of background talkers employed. The squares are associated with binaural-stereophonic listening condition, and the circles with the monaural-control listening condition. Over a wide range of speech-to-background ratios, the binaural stereophonic listening condition is consistently superior.

The two conditions are compared in Fig. 5 in terms of the $(S/B)_0$ ratio at which 50% intelligibility was obtained as a function of the number of background voices. In addition, the difference between the respective $(S/B)_0$ ratios, i.e., the difference between the two solid curves, is presented as the dotted curve in Fig. 5 with the right-hand ordinate. The average difference at 50% intelligibility in the S/B ratios of the two conditions ranges from about 12 db with a single extraneous background speaker (per channel) to about 5.5 db with seven extraneous background speakers (per channel).

The two conditions are compared further in Fig. 6 in terms of the average word intelligibility obtained at a $(S/B)_0$ ratio of 0 db (the subscript 0 again indicating that the background level is the *over-all* level of the combined set of talkers). The average difference in

* This is Air Force Cambridge Research Center Technical Note 57-3, ASTIA Document No. AD 110081, in support of Air Research and Development Command Project 7681.

[1] Investigation of the effect of competing sound sources has been the subject of much recent experimental work. The bibliography in this area is extensive. The reader is invited to examine the work on multichannel listening of Broadbent, Cherry, Egan, and Webster and their co-workers, as well as work on binaural phase by Hirsh, Jefferies, Koch, and Licklider, as well as the masking of speech by speech by G. A. Miller. In addition, the reader is invited to examine a very important recent paper on the mechanism of binaural fusion in speech [see B. M. Sayers and E. C. Cherry, J. Acoust. Soc. Am. **29**, 973–987 (1957)].

† Cf. I. Pollack and J. M. Pickett, J. Acoust. Soc. Am. **29**, 1262(A) (1957), or The New Yorker, November 16, 1957, p. 203.

[2] The writers wish to credit Dr. Maurice Rappaport of the Stanford Research Institute whose research proposal for an interphone system utilizing stereophonic information paralleled our own plans for evaluation of stereophonic effects upon speech intelligibility.

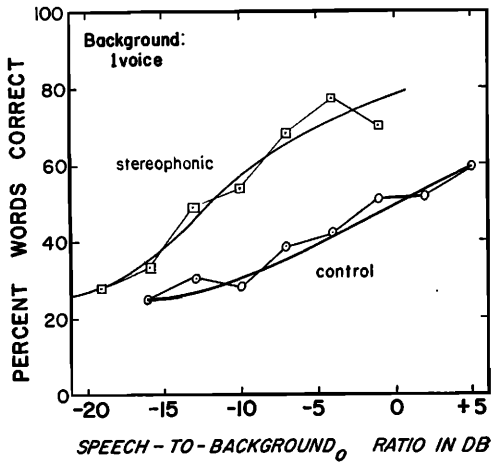

FIG. 1.

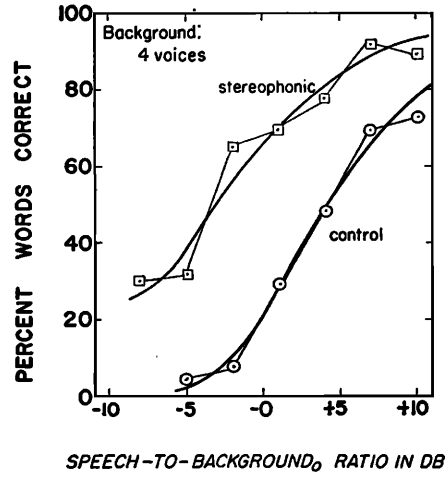

FIG. 3.

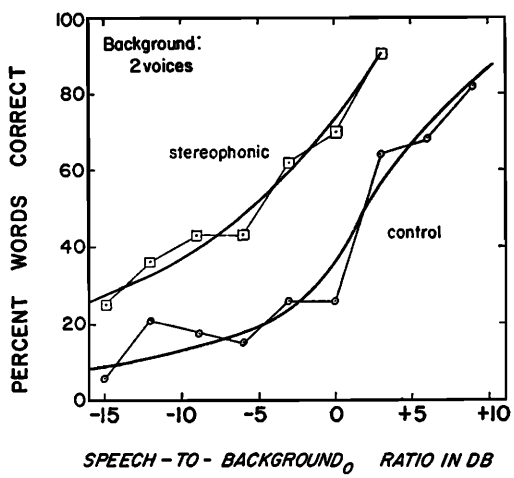

FIG. 2.

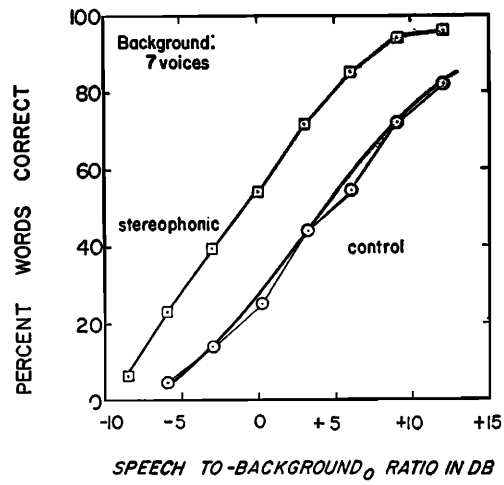

FIG. 4.

FIGS. 1–4. The stereophonic effect for isolated words read against a babble of newspaper text. The ordinate is the average percentage of words correctly reproduced by the listening crew. The abscissa is the speech-to-background $(S/B)_0$ ratio in decibels, in which the extraneous background is expressed in terms of the *over-all* level of the combined group of talkers. The number of extraneous background talkers is listed on the separate graphs. The results associated with the binaural stereophonic listening condition are presented as squares; the results associated with the monaural control condition are presented as circles. Each point represents 200 observations—4 independent tests of 10 words presented to a testing crew of 5 listeners.

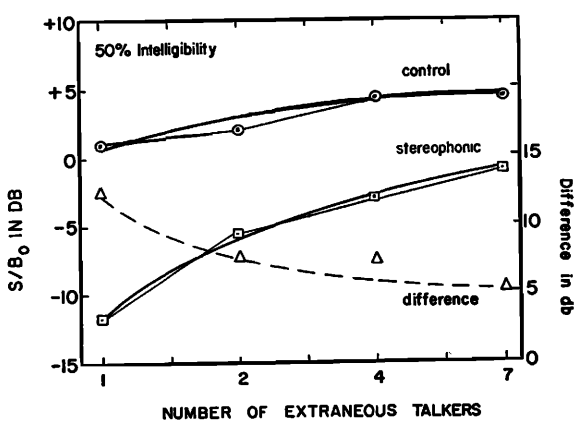

FIG. 5. Comparison between the $(S/B)_0$ ratios necessary at 50% intelligibility for the binaural-stereophonic listening condition and the monaural-control listening condition. Points obtained by horizontal interpolation of Figs. 1–4 at 50% words correct. The dotted curve is plotted against the right ordinate and shows the decibel advantage of stereophonic listening.

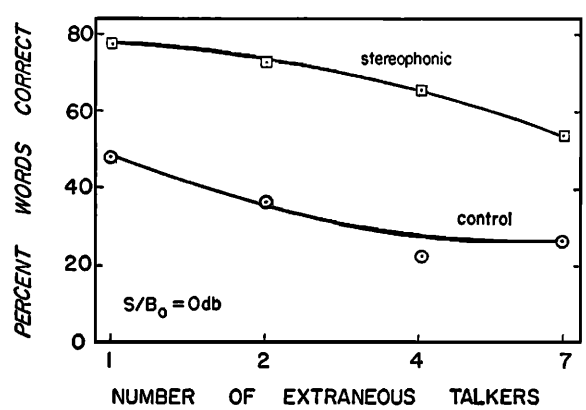

FIG. 6. Comparison between the average intelligibility scores at an *over-all* $(S/B)_0$ ratio of 0 db for the binaural stereophonic listening condition and for the monaural control listening condition. Points obtained by vertical interpolation of Figs. 1–4 at a $(S/B)_0$ ratio of 0 db.

word intelligibility is about 30% at the $(S/B)_0$ ratio of 0 db.

The most direct evaluation of the stereophonic effect is presented in Fig. 7. Figure 7 compares the two conditions with the assumption that the desired speaker, and each of the background speakers, were speaking at the same level. The abscissa is the speech-to-background $(S/B)_a$ ratio (the subscript a indicating that the background level was *adjusted* to that of a single talker). In this comparison, the over-all background level was adjusted on the assumption of simple power addition among the background talkers. For example, for a single background speaker, the curves of Fig. 1 were intersected at a $(S/B)_0$ ratio of 0 db; the curves of Fig. 2 were intersected as a $(S/B)_0$ ratio of -3 db, etc.

The consistent superiority of the binaural stereophonic condition is maintained. However, even in the stereophonic condition, word intelligibility is nearly zero in the presence of seven background talkers (per channel), if all are speaking at the same level at the same time.[3]

[3] One moral of the result is that, if you wish to communicate at a cocktail party, make sure your listener has fewer than seven people speaking to each ear, or make sure your communications have less information than Harvard PB monosyllabic words.

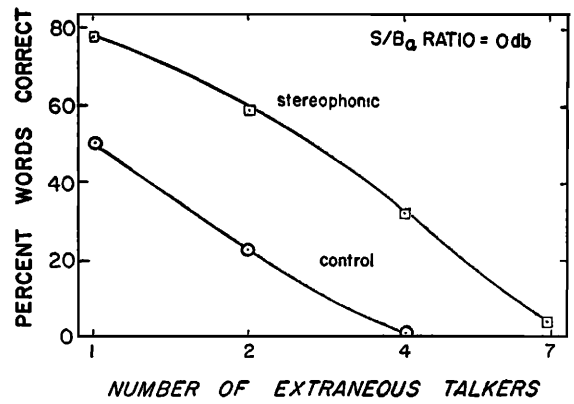

FIG. 7. Comparison between the average intelligibility scores at an *adjusted* $(S/B)_a$ ratio of 0 db for the binaural-stereophonic listening condition and for the monaural-control listening condition. Points obtained by vertical interpolation of Figs. 1–4 with the assumption of equal voice level per background talker summating by power addition, i.e., interpolation at $(S/B)_0$ ratios of 0 db, -3 db, -6 db, and -8.5 db, respectively.

In summary, large gains in word intelligibility above a background of speech babble may be obtained with stereophonic listening as compared with nonstereophonic listening.