AM 70-6

AUDITORY PROCESSING FOR SPEECH INTELLIGIBILITY IMPROVEMENT

Jerry V. Tobias, Ph.D.

Approved by

J. R. Dille, M.D.
Chief, Civil Aeromedical Institute

Released by

P. V. Siegel, M.D.
Federal Air Surgeon

April 1970

Department of Transportation
FEDERAL AVIATION ADMINISTRATION
Office of Aviation Medicine

AUDITORY PROCESSING FOR SPEECH INTELLIGIBILITY IMPROVEMENT*

I. The Problem

Everyone is familiar with environments that are too noisy to permit speech signals to retain much intelligibility. In some of these environments, it is rare for safety to be affected by the interference with speech; the commonest of these is probably the cocktail party.[7 13 15]

Where safety *is* concerned, there are a number of standard solutions. They call for an increase in the signal intensity at the ears, or for a decrease in the intensity of the noise. However, in most light aircraft, and in many heavy aircraft, boosting the signal level and attenuating the noise is not the sort of solution that can be readily applied.

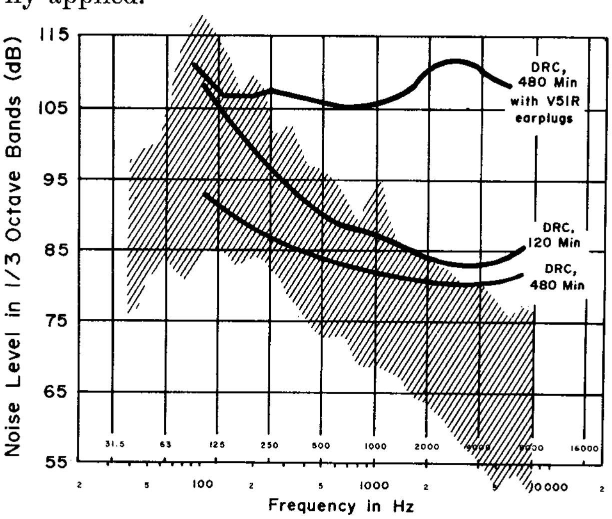

FIGURE 1. Range of noise levels measured in a number of types of single-engine airplanes.[19] Most of the popular planes now being sold in the United States are included. The superimposed DRC (damage-risk criterion) curves show the limits of noise exposure per day[12] that can be endured without creating permanent hearing loss.

The problem is a real one. Almost all the noises that have been measured in airplanes, no matter what sorts of engines are used, are adequate to produce hearing losses after a time. Such noise levels inevitably must interfere to some degree with the intelligibility of speech. With the exception of helicopter pilots, who are invariably exposed to extremely high sound pressures, the pilots of light planes are most commonly exposed to the highest noise levels (Figures 1 and 2). However, the difference between the values for light planes and those for large planes is generally only a few decibels, and the spectra are similar, even though the noise sources differ from type to type.[19 20] On the other hand, the pilots of heavier planes tend to fly more, and so get longer exposures to noise. A few planes are as much as 10 or 15 dB lower in cockpit-noise level, but for the most part, the curves in Figures 1 and 2 are representative. (The superimposed damage-risk criterion, or

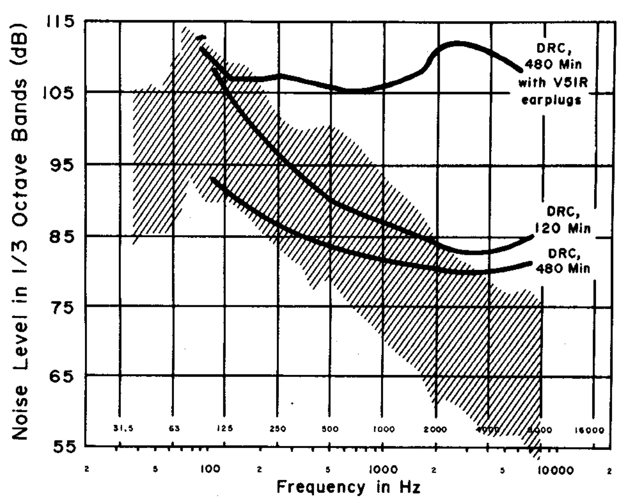

FIGURE 2. Range of noise levels measured in a number of types of twin-engine airplanes.[20] Most of the popular planes now being sold in the United States are included. The superimposed DRC (damage-risk criterion) curves show the limits of noise exposure per day[12] that can be endured without creating permanent hearing loss.

* Some of the material in this paper was presented at an AGARD/ASMP Symposium on Aeromedical Aspects of Radio Communication at Brooks AFB, Texas, in May 1969; some parts were presented at the 40th Annual Scientific Meeting of the Aerospace Medical Association.

DRC, curves show the limits of daily noise exposure, under the noted conditions, beyond which a hearing loss can ultimately be expected.)

Most pilots of light aircraft seem to prefer loudspeakers to earphones for various reasons. Air-transport pilots also seem to like the idea of having nothing over their ears. Yet the pilot wearing a full headset has two advantages: one is that he is partially insulated from the ambient noise by his earphone cushions; the other is that the sounds he most needs to hear are transduced near his eardrums, thus increasing his signal-to-noise ratio (S/N) without either driving his receiver so hard that the speech is distorted or having to have particularly powerful (and therefore heavy) amplifiers such as would be necessary to operate loudspeakers.

Since cockpit noise does interfere with speech recognition (takeoff is an especially difficult time), and since further increases in receiver volume are already impractical because of distortion, some other answers are needed for the pilot or crew member who prefers loudspeakers to headsets. Certainly education is an obvious approach, but many fliers will insist that headsets are uncomfortable, and that earplugs, which would protect the wearer's hearing and, in noisy surroundings, improve speech intelligibility,[4 11] are even more uncomfortable. Loudspeakers are thus likely to be with us for a long time.

II. Some Solutions

The most obvious approach to the cockpit intelligibility problem is one that has already been pushed nearly to its limit. It is the brute-force technique of driving the transducers at very high levels. At high enough levels, the procedure is self-defeating, for it forces the speakers beyond their limits of linearity, and adds distortion to the signals, thus deteriorating rather than assisting recognition. Even if the very high intensities could be transduced without distortion, though, the sound intensity at a listener's ears would become great enough to produce pain. So, under the most difficult listening conditions, several things may happen: there will not be enough power available to drive the loudspeakers, the signal level will remain too low to override the noise, enough distortion will be added to the signal to make it unintelligible, or the sound intensity will be so great as to be unacceptable.

Other kinds of solutions have been offered, but each has several inherent problems. A particularly appealing approach is to increase the fidelity and frequency range of the transmitted and received signals. Tone quality is improved, the redundancy of the signal is increased (making it easier to get enough information to understand what was said), and, all in all, speech transmitted over a high-fidelity system[1] or over a wide-range system with a high-frequency emphasis[5] is better. Such signals seem to overcome the noise's deleterious effects. However, the constraints imposed by the limited radio-frequency spectrum available for flight communications mean that the common practice of limiting the band to about a 3000–Hz width will likely be with us forever, and so make this solution unusable.

An architectural approach would be to reduce the noise by adding acoustical insulation to the bulkheads. However, particularly in light aircraft, this answer is not a useful one because the effective characteristic of such insulation is mass. Mass, of course, cuts down on the payload.

Improved engine muffling would also improve the S/N, although propeller and wind noise would remain unchanged. Too, this change would not be especially meaningful in aircraft with engines far to the rear of the cockpit.

Theoretically, there are ways to modify the noise waveform in a cockpit by adding other sounds.[2] The techniques turn out to be impractical, however, when the receiver (in this case, a human head) has finite dimensions or can move.

There is still another way to insulate the aircraft's occupants against noise—that is by supplying passengers and crew with ear protectors. It is difficult to convince someone that a headset is better for him because it covers his ears, especially when he has chosen to have loudspeakers installed in his plane in the first place. Although the statement is true, it is often difficult to convince people that earplugs (either alone or under a headset) are an optimum solution both to the noise and to the S/N problems (because earplugs' patterns of selective attenuation are more effective on noise than on speech[4 11]). Certainly earplugs are not used enough to serve as the *only* answer to the intelligibility-improvement question.

So, although the ideal condition could be reached by insulating cockpits, muffling engines, widening broadcast bandwidths, and having everyone wear earplugs and use headsets, none of these situations is likely to come about in the foreseeable future for more than a tiny minority of the world's flying personnel.

Approach. One more approach is available. It is based in a family of auditory phenomena that can produce an improvement in the apparent S/N without actually changing either the speech or the noise intensity. Derived from studies of binaural hearing under earphones, the method applies data on "masking-level differences" (MLDs)[3,9] to loudspeaker-listening situations.[6,8,10]

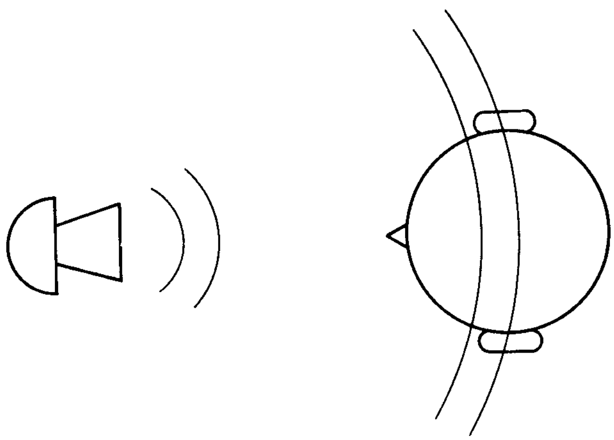

FIGURE 3. Wave pattern at a listener's ears from a single source of sound in the median plane. The sounds at the two ears are essentially identical.

When several signals mix before they reach a listener's ear, no matter how diverse a combination of sources might be involved, they form a single waveform that includes all the information from all the signals. In order for an observer to separate one signal from the others, he has somehow to figure out which fractions of this single, continuous wave belong to that desired signal. With one-eared listening, the cues available to him for this task are limited: he can pick segments that are similar to what he expects to hear, he can try to form a unitary auditory image from those parts whose dynamic characteristics are similar, or he can try to form an image from those parts that share common fundamental frequencies. The task is extremely complex because the signal mixtures are extremely complex, and each signal is almost constantly changing. This signal-selection problem is generally beyond the capability of any form of mechanical or electronic analysis now available, but the human analyzer does very well at isolating wanted information from the mass of sounds that reach it simultaneously.

Yet, a listener who has *two* ears that can be differentially stimulated does much, much better. By a neural equivalent of adding and subtracting the two waves in various ways,[3] he can first equalize and then cancel the unwanted parts of the waveforms. Although the procedure is imperfect, as much because of variations in the neural coding as anything else, the binaural system is much better than any one-channeled system can be.

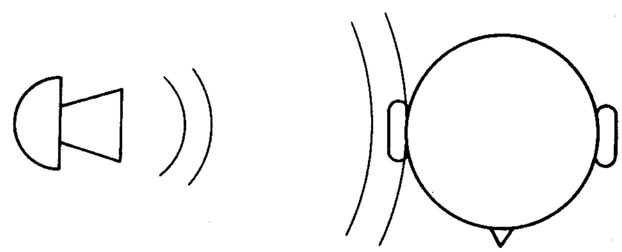

FIGURE 4. Wave pattern at a listener's ears from a single source of sound at the side. The sound at the more distant ear is less intense, and each part of the waveform arrives slightly delayed compared to the near ear.

The differences between signals at the two ears are primarily a function of the azimuth of the source. A source from straight ahead (or anywhere in the median plane) reaches both ears in the same form (Figure 3). Any variations that occur in the wave that goes to the right ear also occur symmetrically in the wave that goes to the left. A sound off to the side (Figure 4) is more intense at the nearer ear, and each part of the wave gets there sooner, which gives the auditory nervous system both an *interaural intensity difference* and an *interaural time difference* to manipulate during the equalization and cancellation processes. Sounds whose sources lie in the same direction arrive with similar interaural disparities, and so are harder to discriminate from each other than are sounds whose sources are spatially separate. The masking effect, then, of a given amount of noise changes as its source moves relative to the wanted signal's source. This change in masking effect with change in binaural stimulation, measured in decibels, is called the masking-level difference or MLD.

Solution. Tests of binaural masking phenomena typically are done under earphones so that artificial disparities between ears can be easily inserted into the signals. Also, with earphones, it is possible to make signal disparities that are quite unlike anything found in the real world. One interesting finding of these earphone MLD experiments is that the less "natural" the interaural disparity, the more effective the brain seems to be in the signal-discrimination task. For example, a theoretical optimum in improvement in apparent S/N is approached by sending the signal to only one ear, while the noise goes to both. In earphone listening, this optimum can be reached.* Of course, with loudspeakers, it is not practical to stimulate only one ear.

A large MLD can also be produced under earphones when the phase of the signal (or the noise) is reversed in one phone while the phase of the noise (or the signal) is not. This unusual situation cannot occur in the real world with normal signals except for a few simple, tonal sounds. Listeners report hearing the phase-reversed sound as if its source were literally inside the skull, and the audible image is thus difficult to confuse with anything whose source is outside. A simple arrangement in which loudspeakers are used as the sources[6 10 16] will approximate the intracranial sensation, and this study reports the investigation of speech-intelligibility improvement during such loudspeaker stimulation.

Experiment. It is clear that only the signal (usually speech) can be usefully manipulated in the cockpit environment; cockpit noise is simply not amenable to the kinds of phase shifts necessary. Only the speech signal can be treated, but in order to see what constitutes a useful treatment, it is necessary to examine the comparable auditory situation under earphones. The critical features of the phase-flipped signal turn out to be the constant interaural phase and the fact that, with earphone-delivered sounds, head movements create no interaural changes. When two loudspeakers driven from a common transmitter are placed symmetrically on either side of the listener's median plane (Figure 5), the sound

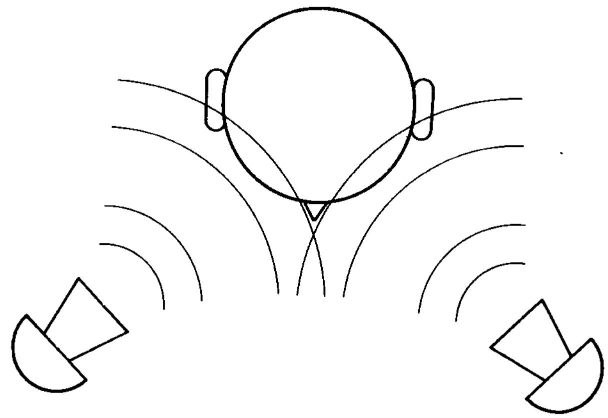

FIGURE 5. Wave patterns at a listener's ears from two sources, symmetrically arranged about his head. Some information from each source reaches both ears, although, as in Figure 4, the far ear (for each source) receives an attenuated and delayed version.

delivered to the two ears is essentially identical; interaural differences are zero except as head movements make small changes. The effect is the one strived for in setting up a home stereophonic listening system; when the two speakers receive the same information, the sound appears to be coming from directly between them. When the phase of one of the speakers is reversed, though, the signals at the two ears become distinctly different. Each ear receives the signal from the nearer speaker a little sooner and a little louder than the signal from the speaker on the other side. The contralateral signal is phase shifted, and so partially cancels the ipsilateral signal, creating an effect very like the one that comes from earphone stimulation with one phone's waveform inverted. Additionally, head movements are not very effective in changing this percept, so the auditory system functions much as it does during similar stimulation with a headset.

When the masking effect is tested under earphones, the ability to hear speech in noise is shown to improve under phase-reversed condi-

* This situation is *not* duplicated by the pilot who wears only one phone from his headset; he would need to have similar noise at the two ears, and that is not possible unless he keeps both ears covered by the earphones and cushions. He is not using the earphone cushion to attenuate and filter the noise to his open ear, so a simple subtraction or cancellation process still leaves a great deal of noise. However, if he did have *both* phones on, but only one of them operative, he would be able to hear better in the noise.

tions by the same amount that would have been produced by increasing the speech intensity by a factor of three or four.[14 17 18 21] In difficult, noisy listening conditions, the associated increase in the intelligibility of a given message might be 30 or 40%. For messages from a limited set (such as air-traffic-controller transmissions), the change could be even more.

Our tests used an incomplete circular array of nine loudspeakers, 30° apart, one at each clock position from 8 through 4 (Figure 6). The

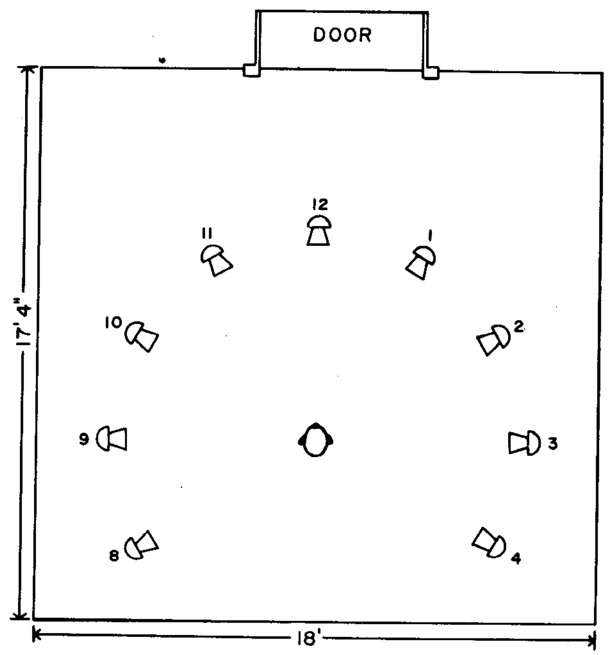

FIGURE 6. Experimental arrangement of listener and nine loudspeakers.

rearward positions were not included because the acoustic paths followed by sounds emanating from them would lead signals to the ears in exactly the same ways that would occur for speakers that were ipsilaterally symmetrical—for instance, 1 and 5 o'clock speakers have the same effect, as do 8 and 10 o'clock speakers. Pairs of speakers were tested for the intelligibility-level difference (ILD) between in-phase and out-of-phase presentations of signals that were immersed in noise. The noise was wideband white noise. The signals were 75 tape-recorded, 120-word passages that subjects were trained to repeat in a quiet voice while the material was being presented. A highly directional microphone system carried the responses to another room where an observer noted correctly and incorrectly spoken words on a prepared copy of the passage. The observer also heard the recorded passage without noise, so it was simple to keep track of where in the text a subject might be. The last 100 words of each passage were scored for the number of correct responses. Conditions were assigned randomly to the passages.

Subjects were young men with normal hearing, and a total of 40 were tested in the various parts of the work. The major data reported here are based on tests of 25 subjects.

The determination of signal levels for speech is nearly impossible to standardize, so numbers representing signal-to-noise ratios are not as meaningful as they might be. However, once a calibration value has been accepted for the speech signal, changes in level can be compared to each other. In the case of these experiments, a 20-dB range of speech intensity levels was used, leading to S/Ns of +5, 0, −5, −10, and −15 dB. Four of these levels were adequate to cover the useful intelligibility range for any subject, although for some the four were +5, 0, −5, and −10 dB, and for others they were 0, −5, −10, and −15 dB. Each level was tested in each phase condition, and each of those combinations was tested with each pair of loudspeakers. Replications were made, and the whole series was randomized for each subject.

III. Results

The results show the two-speaker transmission system to be nearly as effective in increasing speech intelligibility as an earphone system is (see Table I). Asymmetrical pairs of speakers are not as good as symmetrical ones and are not included in the Table. The differences between symmetrical speaker pairs (1 and 11 o'clock, 2 and 10, and so on) are statistically insignificant when tested at each S/N, but the narrower angles seem to be slightly better (speakers 1 and 11 are slightly better than 2 and 10, which in turn are a little better than 3 and 9). This systematic change suggests that a larger sample, or use of a signal with less inherent variability, would lead to significant differences.

Average responses for 25 subjects show that, for speech signals so thoroughly immersed in noise as to be nearly unintelligible with the signals in phase at both speakers, changing the signal phase at one speaker increases the intelligibility by amounts averaging from 26 to 36%, depending on the condition. For the least noise used in these tests, the mean improvement was still approximately 9 to 13%. (Note that, as the intelligibility of the original masked speech grows higher, there is less room available for improvement, but that, as the masking effect of the noise increases, the phase-reversal offers greater and greater assistance. That is, the greatest help occurs in the situation where the need is greatest.) In each case, the result is comparable to the range of results found for tests done with earphone listening, a very useful, if unexpected, outcome.

TABLE I.—Summary of data; averages of 25 subjects. Scores (percentage of correct word identifications) for four signal-to-noise ratios are shown for three symmetrical loudspeaker placements and for two phase conditions. Differences between phase conditions are tabulated in the last column.

Conditions		*Scores*		
S/N	*Loudspeaker pair*	*Phase normal*	*Phase reversed*	*Difference*
	1 and 11	72.0%	84.2%	12.2%
+5dB	2 and 10	71.6%	79.7%	8.1%
	3 and 9	74.5%	79.6%	5.1%
	1 and 11	58.5%	73.7%	15.2%
0dB	2 and 10	57.0%	66.6%	9.6%
	3 and 9	60.5%	61.0%	.5%
	1 and 11	35.6%	60.5%	24.9%
−5dB	2 and 10	31.7%	48.4%	16.7%
	3 and 9	33.0%	48.9%	15.9%
	1 and 11	13.0%	47.6%	34.6%
−10dB	2 and 10	13.1%	26.3%	13.2%
	3 and 9	11.0%	18.8%	7.8%

IV. Conclusions

From these data, it appears that one solution to the problem of losing the meaning of messages transmitted to fliers who do not use any ear covering, especially during critical times, is to install *pairs* of loudspeakers, symmetrically, approximately equidistant from the pilot's head, toward the front or (because of the greater availability of space, and because the symmetry of the acoustic paths permits it) toward the rear, with one speaker's leads wired in reverse to the other's, so that phase inversion is automatic. The use of two speakers also allows a higher receiver-gain setting without overdriving the loudspeakers to produce the distortion that can destroy the advantage of the increased S/N, and both can be operated from the same amplifier, so the added weight is only for the second speaker.*

Although this approach is not as universally beneficial and satisfactory as using a headset in the same way, or as using earplugs (either alone or under a headset) would be, it can help to improve speech reception for aviators who prefer to leave their ears uncovered. The technique makes no appreciable change in the ability to hear and understand the highly intelligible transmissions received during most of a flight. Only during times when the noise level is especially high can an improvement in S/N be at all meaningful. But of course, it is during just those times that an improvement in speech reception can be most important. The cost, both in money and in weight, is extremely low, and the payoff can occasionally be quite high. Therefore, it is suggested that serious consideration be given to the advisability of putting two loudspeakers in aircraft cockpits in one of the described configurations.

* Note that two loudspeakers driven from the same source, but wired in phase with each other, are as good as (but no better than) one loudspeaker except as their distortion is decreased because each can operate at slightly reduced power to give the same overall signal intensity. It is the phase reversal that makes the additional speaker an effective means of transmitting intelligence in exceptionally difficult listening conditions.

REFERENCES

1. Camp, R. T., Jr.: Personal communication, 1968.
2. Dolch, J. P.: Phase and intensity relationships in the interference of bone- and air-conducted sound. Supplementary Report #2, Signal Corps Contract #DA 36–039–SC–42562. State University of Iowa, Iowa City, Iowa, 1953.
3. Durlach, N. I.: Binaural signal detection: equalization and cancellation theory. In J. V. Tobias (Ed.) *Foundations of Modern Auditory Theory.* Academic Press, New York, 1971.
4. Froelich, G.: The effects of ear defenders on speech perception in the cabin of reciprocating engine and turboprop-powered transport aircraft Noratlas and Transall C–160. In *Aeromedical Aspects of Radio Communication and Flight Safety Symposium.* Advisory Group for Aerospace Research and Development (NATO), 1969.
5. Griffiths, J. D.: Improved analog voice communications. *Air Force Cambridge Research Laboratories—68–0227, Its Instrumentation Papers,* #141, 1968.
6. Hanson, R. L. and Kock, W. E.: Interesting effect produced by two loudspeakers under free space conditions. *Journal of the Acoustical Society of America,* 29:145, 1957.
7. Hardy, H. C.: Cocktail party acoustics. *Journal of the Acoustical Society of America,* 31:535, 1959.
8. Heffler, A. J. and Schultz, M. C.: Some implications of binaural signal selection for hearing aid evaluation. *Journal of Speech and Hearing Research,* 7:279–289, 1964.
9. Jeffress, L. A.: Binaural signal detection: vector theory. In J. V. Tobias (Ed.) *Foundations of Modern Auditory Theory.* Academic Press, New York, 1971.
10. Jeffress, L. A.: Method for the study of the localization of airborne sound. *Journal of the Acoustical Society of America,* 27:208, 1955.
11. Kryter, K. D.: Effects of ear protective devices on the intelligibility of speech in noise. *Journal of the Acoustical Society of America,* 18:413–417, 1946.
12. Kryter, K. D., Ward, W. D., Miller, J. D., and Eldredge, D. H.: Hazardous exposure to intermittent and steady-state noise. *Journal of the Acoustical Society of America,* 39:451–464, 1966.
13. Legget, R. F. and Northwood, T. D.: Noise surveys of cocktail parties. *Journal of the Acoustical Society of America,* 32:16–18, 1960.
14. Levitt, H. and Rabiner, L. R.: Binaural release from masking for speech and gain in intelligibility. *Journal of the Acoustical Society of America,* 42:601–608, 1967.
15. MacLean, W. R.: On the acoustics of cocktail parties. *Journal of the Acoustical Society of America,* 31:79–80, 1959.
16. Sandel, T. T., Teas, D. C., Feddersen, W. E., and Jeffress, L. A.: Localization of sound from single and paired sources. *Journal of the Acoustical Society of America,* 27:842–852, 1955.
17. Schubert, E. D.: Some preliminary experiments on binaural time delay and intelligibility. *Journal of the Acoustical Society of America,* 28:895–901, 1956.
18. Schubert, E. D. and Schultz, M. C.: Some aspects of binaural signal selection. *Journal of the Acoustical Society of America,* 34:844–849, 1962.
19. Tobias, J. V.: Cockpit noise intensity: fifteen single-engine light aircraft. *Aerospace Medicine,* 40:963–966, 1969.
20. Tobias, J. V.: Noise in light twin-engine aircraft. *Sound and Vibration,* 3:16–19, 1969.
21. Tobias, J. V. and Curtis, J. F.: Binaural masking of noise by noise. *Journal of the Acoustical Society of America,* 31:127, 1959.