ICSV14

Cairns • Australia

9-12 July, 2007

14th International Congress on Sound & Vibration

BINAURAL AND MONAURAL HELMET SYSTEMS: NOISE CONSIDERATIONS

Tomasz Letowski and Larry Ferguson

Human Research and Engineering Directorate, U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5425, USA
trl2@arl.army.mil

Abstract

The main function of the combat helmet is to provide ballistic protection against small arms fire and fragmentation. However, the helmet also serves as a mounting platform for several systems enhancing the soldier's situational awareness. The two main such systems are the helmet mounted (visual) display (HMD) and the radio (audio) communication interface. Audio interfaces may be built into the helmet or mounted directly on the head under the helmet shell. In both cases the main challenge in designing an effective audio interface is to provide radio communications and protect the soldier's ears from high noise levels while permitting adequate auditory awareness. One solution is to mount small loudspeakers in the helmet shell and to provide stand-alone level-dependent hearing protection. This study addresses the issues of monaural and binaural speech reception using helmet-mounted transducers. Monaural and binaural audio interfaces built into the same type of helmet (PASGT) were compared with regard to speech recognition in noise using the Modified Rhyme Test. Several noise conditions differing in noise source location and noise level were investigated. Ten subjects participated in the study. The collected data reveal a 2-6 dB improvement in effective signal-to-noise ratio by changing from the monaural to the binaural version of the helmet and provide support for the energy model of speech intelligibility summation across left and right ear channels. A signal-to-noise ratio of about 0 dB or better is needed to provide the 91% binaural speech intelligibility required by MIL STD 1472D.

1. INTRODUCTION

A soldier's cognitive performance is affected by the complexity of the mission requirements, knowledge of the battlefield, and numerous physical and physiological factors such as fatigue, temperature, and noise. Communication networks and wearable computers combined with easily accessible information displays such as helmet mounted (visual) displays (HMD) can increase the soldier's battlefield awareness and enhance decision-making processes. However, the practical benefits that this increase in information offers are dependent on the soldier's ability to capture, understand, prioritize, and act upon it. Complex and multithreaded information that is made available to the soldier simultaneously through one or few communication channels may actually hamper rather than increase the soldier's effectiveness if it is not properly managed. There are concerns that the soldier's visual sense is overloaded with various types of information that should instead be allocated to the auditory and tactile channels.

The auditory channel is commonly regarded in soldier system designs as a networking channel that facilitates radio communication across vertical and horizontal operational structures. Radio signals are received by either earmuff-based earphones (earphones) or insert earphones (earplugs) and delivered to one ear (monaural reception) or both ears (binaural reception). A number of binaural audio displays have been successfully used in airplane cockpits, submarines, and ground vehicles. Most of these displays deliver the same signal to both ears (diotic presentation) although a growing number of displays deliver different signals to each ear (dichotic presentation). Dichotic presentation is well suited for multi-channel communication and permits sound spatialization in which various talkers (or other sound sources) are heard as located at different positions in three-dimensional space. Displays utilizing dichotic presentation are called *spatial displays* or *3D audio displays* and require a signal processing unit that restores the natural head-related transfer function (HRTF) of the wearer.

Binaural audio displays as described above are fully satisfactory for mounted solder operations where the auditory channel's primary function is communication. However, it should be noted that the primary function of the auditory sense is to provide information about changes taking place in the surrounding environment. This is of critical importance to dismounted soldiers. In this case, binaural earphones that deliver signals directly to the ears while providing some hearing protection against hazardous external noises are not an acceptable solution. The soldier needs to not only hear the radio messages but also to detect, identify, and localize potential sound sources in the local environment and monitor their behaviour. Moreover, the addition of a multi-channel auditory information display may require too much of the soldier's attention, thus reducing situational awareness and possibly interfering with the completion of other critical tasks. Therefore, the use of auditory information displays by dismounted soldiers should only be considered when it does not degrade the situational awareness of the soldier.

Various techniques have been proposed for combining the networking function of the auditory channel with the natural situational awareness function of the sense of hearing in one operating system. The most common solution is to use a single communication earplug and allowing the other ear to hear the environment. A modern civilian variant of this solution is the use of clip-on-ear cellular phone headsets. This monaural solution may be acceptable in some situations, but it greatly reduces the localization capability of the wearer and is detrimental to the soldier's spatial orientation. Another solution is to use two communication earplugs or earphones with external microphones. These types of monaural or binaural *talk-through* systems have a clear advantage over the monaural communication earphone but still affect sound localization and are high maintenance devices. It is also possible to combine the natural situational awareness provided by unocluded ears with the networking function provided by the bone conduction channel. Such systems are very promising but are still in the development stage for military applications.

One audio communication system proposed for the dismounted soldier that facilitates situational awareness uses small loudspeakers built into the helmet shell of the *Personnel Armor System for Ground Troops* (PASGT). The system may have one small loudspeaker located on the right side of the helmet or two loudspeakers for binaural reception. Such a system leaves the ears unoccluded yet allows reception of radio messages by radiating them towards the ears of the wearer. The operational limitations of the proposed system are numerous, but it is not clear what the potential communication advantages and limitations of such monaural and binaural helmet interfaces are. Therefore, the scope of the present study was to compare the effectiveness of these monaural and binaural helmet systems in transmitting speech messages under selected noisy operational conditions including ipsilateral, contralateral, and binaural modes of listening.

2. METHOD

The study consisted of two experiments labeled Experiment 1 and Experiment 2 that were conducted using similar test procedures but under different test conditions.

2.1 Experiment 1

The objective of Experiment 1 was to compare the effectiveness of one-loudspeaker (monaural) and two-loudspeaker (binaural) helmet systems in transmitting speech messages in the presence of a single source of noise. The goal of the study was to determine the speech-to-noise ratio required to provide 75% speech intelligibility under select listening conditions.

2.1.1 Helmet

The test helmet was a modified PASGT helmet with a pair of loudspeakers mounted on the left and right side of the helmet between the chinstrap screws and the bottom edge of the helmet. The PASGT helmet is a U.S. infantry combat helmet made of Kevlar and resembling the German Stahlhelm in shape. The loudspeakers were 1” Calrad MS drivers with 8Ù impedance. The loudspeakers were matched for frequency response and impedance. A set of two switches mounted at the back of the helmet was used to switch the individual loudspeakers on and off.

2.1.2 Listeners

Ten volunteers participated in the study. The participants were 19 to 42 years old with a mean age of 28 years. All participants were native English language speakers, had a pure tone hearing threshold better than or equal to 25 dB HL at audiometric frequencies from 250 through 4000 Hz [1], otoscopically normal ears, and no recent history of otologic pathology.

2.1.3 Speech stimuli and noise

The speech stimuli were 300 monosyllabic words comprising the Modified Rhyme Test (MRT) [2]. Each word consists of a single vowel surrounded by two consonants (CVC). The words are arranged in 50 sets of 6 words and include all the phonemes of the American English language. The target words were spoken by a male speaker and presented in the carrier phrase “Mark the ______ now”. The peak RMS levels of the target words were normalized to be within ±1 dB.

The noise masker used in Experiment 1 was a live recording of a Bradley Fighting Vehicle (BVF) travelling at a constant speed of approximately 50 km/h on unpaved road. The noise was recorded using a B&K 4133 microphone mounted to the exterior of the moving vehicle and a Sony TCD-D7 portable DAT recorder.

2.1.4 Instrumentation

The study was conducted in a large audiometric test room (IAC Inc.) complying with the ANSI S3.1-1999 (R2004) [3] standard for sound field listening. The listener was seated in the center of the room and wore a test helmet. Two matched Infinity 150 Studio Monitors located at both sides of the listener were used to reproduce the masking noise. The left and right loudspeakers were used individually to simulate a noise source at the left (NL) or, respectively, right (NR) side of the listener. The noise was presented at an average level of 72 dB (A) as measured at the listener location while unoccupied. The instantaneous noise level varied from 69 dB (A) to 77 dB (A) while generally staying within the 71-73 dB (A) range.

The MRT speech signals were played from an IBM PC computer through a Crown D-75 amplifier. The helmet switches were set to deliver the signal to either the right ear (SR) or to both ears (BL). The signal levels and word selection were controlled using a Tucker-Davis Technologies (TDT) System II [4] and custom-written software. In a single trial the listener was presented with the target word in the carrier phrase and responded using a computer mouse by selecting one of the six possible alternatives displayed on the computer screen.

2.1.5 Procedure

Four test conditions were investigated in the study: SR-NR (right loudspeaker on; noise source on the right), SB-NR (both loudspeakers on; noise source on the right); SB-NL (both loudspeakers on; noise source on the left), and SR-NL (right loudspeaker on; noise source on the left). The right ear location of the monaural loudspeaker is consistent with reports showing slight prevalence of left ear hearing loss in male population and the fact that the rifle shooters generally show more hearing loss in the ear opposite the shoulder to which the rifle stock is held, that is, right-handed shooters generally have more hearing loss in the left ear and the left-handed shooters generally have more hearing loss in the right ear.

For each of the four test conditions a single run of the MRT was administered as a 50-item 6-alternative forced choice test. The test was run as an adaptive procedure as developed by Zera [5]. In this procedure the word presentation level is adjusted up or down depending on the listener's response in order to achieve a preset percent of correct responses at the end of the test. A correct response results in a decrease in signal level while an improper response results in an increase. The level adjustment is controlled by a maximum likelihood estimation procedure described by Green [6]. After the first N_1 trials of each run, the procedure converges upon a specific speech-to-noise (SN) ratio corresponding to a preset level of correct responses. An additional set of N_2 trials is run at that constant level to provide more accurate data. The numbers of trials run in the present study were N_1=25 and N_2=25 and the converging level was 75%.

2.2 Experiment 2

The primary objective of Experiment 2 was to determine the effect of noise level on speech intelligibility delivered through a binaural version of the helmet operating in the diotic mode. The secondary objective of this experiment was to assess binaural advantage in speech intelligibility when speech signal is presented in surrounding background noise.

2.2.1 Helmet

The same helmet used in Experiment 1 was used in Experiment 2.

2.2.2 Listeners

The same ten volunteers who participated in Experiment 1 participated in Experiment 2.

2.2.3 Speech stimuli and noise

The same speech stimuli and masking noise used in Experiment 1 were used in Experiment 2.

2.2.4 Instrumentation

The study was conducted in the same large audiometric test room as Experiment 1. The listener was seated in the center of the room and wore a test helmet. Two Infinity 150 Studio Monitors located at both sides of the listener were used to reproduce the masking noise. Both loudspeakers were used in parallel (NB) to simulate surrounding vehicle noise. The noise was presented at three average noise levels of 65 dB (A), 75 dB (A), and 85 dB (A) as measured at the listener location while unoccupied.

The MRT speech signals were played from an IBM PC computer and delivered to the helmet assembly through a Crown D-75 amplifier. The helmet switches were set to deliver the signal to either the right ear (SR) or to both ears (BL). The average speech level was set at 65 dB (A) for all target words. Listeners' responses were collected in the same way as in Experiment 1.

2.2.5 Procedure

Four test conditions were investigated in the study: a binaural condition SB-NB (both loudspeakers on; both noise sources on) at 0, -6, and -12 dB SN ratio and a monaural condition SR-NB (right loudspeaker on; both noise sources on) at -6 dB SN ratio. The monaural condition was added to assess an effectiveness of binaural summation under given listening environment. For each of these conditions a single run of the MRT was administered as a 50-item 6-alternative forced choice test. The test was run as a fixed procedure at a specific SN ratio.

3. RESULTS AND DISCUSSION

The results of Experiment 1 and Experiment 2 are shown in Figures 1 and 2, respectively. The data presented in Figure 1 were normalized so that each listener achieved a 75% correct response rate under the SR-NR condition. Data normalization was made using the cumulative normal function relating MRT score to SN ratio [2][6].

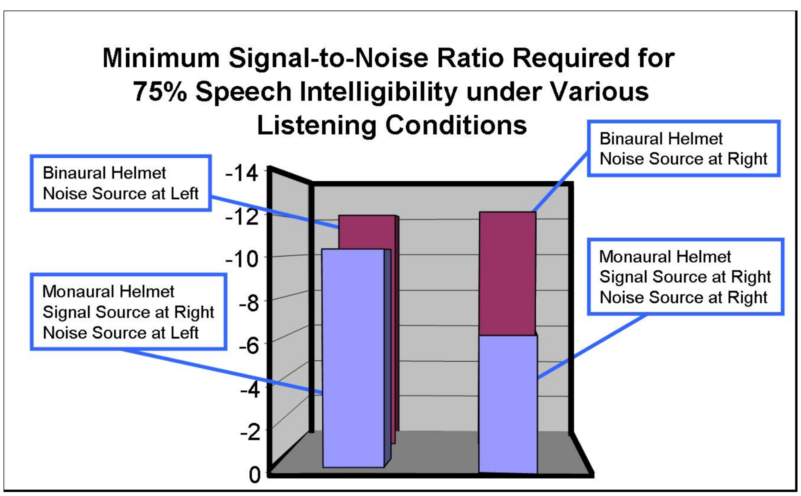

Figure 1. Average (mean) signal-to-noise ratios for the four test conditions.

The SN ratio needed to obtain 75% speech intelligibility for the SR-NR condition was -6.0 dB. The SN ratios for other test conditions were -10.6 dB (SD=2.3 dB) for SR-NL condition, -12.1 dB (SD=2.6 dB) for SB-NL condition, and -12.3 dB (SD=1.7 dB) for SB-NR condition. The subsequent differences between SN ratios for SR-NL, SB-NL, and SB-NR conditions and the SN ratio for the SR-NR condition were - 4.6, -6.1, and - 6.3 dB, respectively. These differences correspond respectively to approximately 14%, 19%, and 20% increases in word recognition scores keeping the SN ratio in each of the conditions constant at the -6.0 dB level.

T-tests with Bonferroni correction for multiple tests showed that the differences in the SN ratios between SR-NR and the other three test conditions were statistically significant at the 0.001 level ($t=6.273$ or higher; $df=9$). The differences between SR-NL and the two binaural conditions were not statistically significant ($p>0.05$).

The data presented in Figure 1 show a clear advantage for the two-loudspeaker helmet assembly design over the single-loudspeaker design. The binaural improvement in the listeners' performance was quite substantial and equal to a 6-20% increase in speech intelligibility. Alternatively, the same 75% speech recognition score could be obtained at a 2-6 dB lower speech level. This, in turn, increases the security of the communication, reduces the electrical current drained from the battery supplying the system and extends battery life.

It is well known that binaural listening has an advantage over monaural listening in both signal detection and loudness perception tasks. This applies to both listening in quiet and listening in noise. For example, Pollack [7] reported that for normally hearing listeners, binaural thresholds of audibility were 2.5 dB lower for a 1 kHz tone and 2.0 dB lower for white noise than the corresponding monaural thresholds. The binaural level of sound can be 7-10 dB lower than the monaural level and still be perceived as being equally loud [8]. Even more importantly, the most comfortable listening level (MCLL) for binaural listening to speech seems to be 4-6 dB lower than the respective MCLL for monaural listening [9]. All these effects result from the binaural summation of the physiological excitations evoked in both ears of the listener.

The binaural advantage in speech intelligibility observed in this study can be attributed to binaural (diotic) summation, acoustic head shadow, and binaural squelch. Acoustic head shadow is the shielding effect of the listener's head on the ear farther away from the sound source, which results in less masking of the speech signal in this ear. Binaural squelch is the speech recognition improvement resulting from the difference in binaural summation between coherent (speech) and incoherent (noise) signals. Davis, Haggard, & Bell [10] studied diotic summation for sentence recognition tests and reported a 5-9% improvement at a -7.4 dB SN ratio (monaural score of 70%). This leaves 50% or more of the total improvement as being due to the other two phenomena. It can also be expected that at lower intelligibility levels the binaural advantage may be even greater. For example, Davis and Haggard [11] reported a speech intelligibility increase as large as 20% due to binaural summation of speech with monaural intelligibility of about 50%.

The 75% binaural speech intelligibility observed in Experiment 1 required about a -12.2 dB SN ratio for a laterally located noise source. Monaural speech intelligibilities of 75% for the left and right ear for the same noise source location required -10.6 dB and -6 .0 dB SN ratios, respectively. The sum of the speech energy delivered simultaneously at these two SN ratios would result in about a -12.0 dB SN ratio that matches quite well with the binaural SN ratio of -12.2 dB reported in the study. This agreement provides support for the energy summation hypothesis of the binaural advantage.

The results of Experiment 2 presented in Figure 2 show progressive rapid decrease in binaural (diotic) speech intelligibility in surrounding background noise (SB-NB) as the SN ratio changes from 0 dB to -12 dB. The average speech intelligibility scores were 98.2% in quiet, 91.2% at SN = 0 dB, 77.0% at SN = -6 dB, and 50.4% at SN = -12 dB. All these differences were statistically significant [F(3,24)=201.3; p<0.0001]. The military standard MIL-STD-1472D [12] requirement of 91% speech intelligibility for the MRT is met by this system at a SN ration of 0 dB or better. The minimum acceptable by the standard speech intelligibility of 75.0% is achieved at the -6 dB SN ratio.

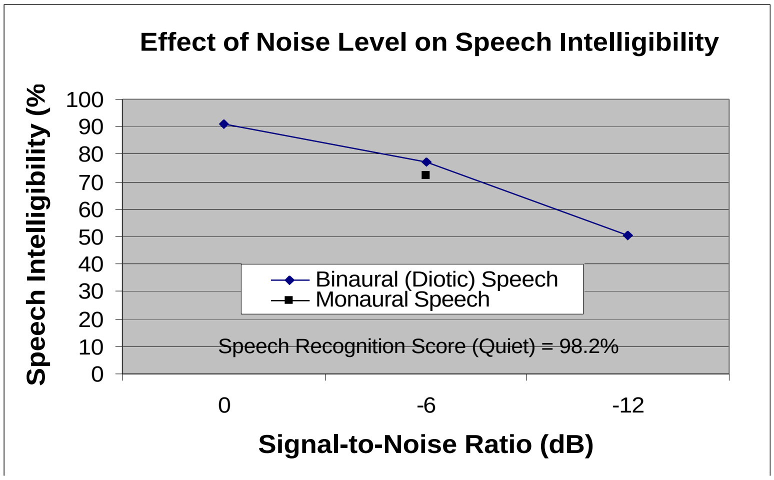

Figure 2. Effect of signal-to-noise (SN) ratio on speech recognition score in surrounding background noise.

The speech intelligibility score measured in Experiment 2 for monaural speech (SR-NB condition) at SN = -6 dB was 72.2%. This value is 4.8% lower than speech intelligibility measured for SB-NB condition at the same SN ratio [F(1,9)=36.0; p=0.0002] and is comparable to the 6.0% binaural advantage observed in Experiment 1 for the contralateral location of the directional sound source. Similar advantage of about 4.0-6.0% was reported for binaural fitting over monaural fitting in bilaterally-impaired listeners [13] and for binaural listening in reverberation [14]. Recall, however, that the binaural advantage depends on the positions of the speech and noise sources in relation to the listener and under unfavourable listening conditions the amount of the advantage can be as large as 20% or more [15][16].

5. SUMMARY

A binaural version of the PASGT helmet offers relatively effective speech communication in quiet and in moderate noise levels. It provides 6-20% higher speech intelligibility than the monaural version when compared at the same SN ratio. The actual amount of the binaural benefit depends on the location of the noise source and the type of noise. Unfortunately, the PASGT helmet itself disrupts auditory localization, and this is one of the reasons that it is being replaced by the Advanced Combat Helmet (ACH) which has less extensive head coverage around the ears. It should also be remembered that communication through loudspeakers mounted in the helmet shell is highly insecure and can easily be overheard by others or captured by acoustic or laser microphones. Therefore, such a technical solution has only limited military and security applications even for helmet platforms that do not negatively affect the wearer's localization capabilities. There are, however, numerous civilian applications where such communication systems could be successfully used (e.g., construction, forest services, rescue squads). Combined with appropriate helmet form and a suitable microphone they can provide hands-free multi-channel communication without compromising the wearer's auditory situational awareness.

REFERENCES

[1] American National Standards Institute, *Specifications for audiometers*, ANSI S3.6-1996, ANSI, New York, 1996.

[2] A.S. House, C.E. Williams, H.M.L. Hecker, and K.D. Kryter, “Articulation-testing methods: Consonantal differentiation with a closed-response set”, *Journal of the Acoustical Society of America* **37**, 158-166 (1965).

[3] American National Standards Institute, *Maximum permissible ambient noise levels for audiometric test rooms*, ANSI S3.1-1999 (R2003), ANSI, New York, 2003.

[4] T. Tucker, *Tucker-Davis Technologies System II Manual,* Tucker-Davies Technologies, Gainesville, Florida, 1992.

[5] J. Zera, “Speech intelligibility measured by adaptive maximum-likelihood procedure”, *Speech Communications* **42**, 313-328 (2004).

[6] D.M. Green, ”A maximum-likelihood method for estimating thresholds in a yes-no task”, *Journal of the Acoustical Society of America* **93**, 2096-2105 (1993).

[7] I. Pollack, “Monaural and binaural threshold sensitivity for tones and white noise”, *Journal of the Acoustical Society of America* **20**, 52-57. (1948).

[8] H. Stephenson, J.M. Higson, and M.P. Haggard, “Binaural hearing in adults with histories of otitis media in childhood”, *Audiology* **34**, 113-123 (1995).

[9] T. Letowski and D. Emanuel, “Most comfortable listening levels for listening through earphones and in a sound field”, *Proceedings of the 15th International Congress on Acoustics* (15th ICA), vol. 3, pp. 269-272. 26-30 June 1995. Trondheim, Norway.

[10] A. C. Davis, M.P. Haggard, and I. Bell, “Magnitude of diotic summation in speech-in-noise tasks: performance region and appropriate baseline”, *British Journal of Audiology* **24**, 11-16 (1990).

[11] A.C. Davis and M.P. Haggard, “Some implications of audiological measures in the population for binaural aiding strategies”, *Scandinavian Audiology,* Suppl. **15** (1982).

[12] Department of Defense, *Human Engineering Design Criteria for Military Systems, Equipment, and Facilities,* MIL-STD-1472D, Washington, DC, 1992.

[13] S. Köbler, “Bilateral hearing aids for bilaterally hearing-impaired persons – always the best choice?”, Ph.D. Dissertation, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden, 2007.

[14] A.K. Nabelek and P.K. Robinson, “Monaural and binaural speech perception in reverberation for listeners of various ages”, *Journal of the Acoustical Society of America* **71**, 1242-1248. (1982).

[15] W.E. Kock, “Binaural localization and masking”, *Journal of the Auditory Society of America* **22**, 801-804 (1950).

[16] M. Rao, J. Dreyer, A. Scharine, and P. Henry, “A model for predicting intelligibility of binaurally perceived speech”, *ARL Technical Report* **xxxx**, Army Research Laboratory, Adelphi MD, 2007 (in print).