International Journal of Industrial Ergonomics 39 (2009) 246–254

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

journal homepage: www.elsevier.com/locate/ergon

Using the Callsign Acquisition Test (CAT) to investigate the impact of background noise, gender, and bone vibrator location on the intelligibility of bone-conducted speech

B. Osafo-Yeboah [a], X. Jiang [a,*], M. McBride [b], D. Mountjoy [a], E. Park [a]

[a] Department of Industrial and Systems Engineering, North Carolina Agricultural and Technical State University, 1601 East Market Street, Greensboro, NC 27411, USA

[b] Department of Human Factors and Systems, Embry-Riddle Aeronautical University, 600 South Clyde Morris Boulevard, Daytona Beach, FL 32114, USA

ARTICLE INFO

Article history:
Received 5 January 2008
Received in revised form 23 May 2008
Accepted 15 July 2008
Available online 11 September 2008

Keywords:
Speech intelligibility
Bone conduction
Callsign Acquisition Test (CAT)
Background noise
Gender

ABSTRACT

Traditionally, combat operations have relied on air-conducted radio communication to receive and transmit information. Recently, however, there has been the need to find alternatives because headphones used in air conduction are bulky and cover the ears of the listener, thus reducing the listener's awareness of his/her surroundings. Bone-conducted radio communication, however, uses lightweight and inconspicuous transducers which allow radio communication without compromising the listener's awareness of his/her surroundings. This research investigated the intelligibility of bone-conducted speech in white, pink, babble and quiet background environments at the condyle and mastoid locations on the head using the Callsign Acquisition Test (CAT). Data were collected and analyzed from 20 normal hearing participants (10 males and 10 females) between the ages of 19 and 31 years. Significant interaction effect between gender and background noise was found from the results. Post-hoc analysis showed that for both males and females, background noise had a significant impact on speech intelligibility. In babble background, there was a significant difference in speech intelligibility between the male and female listeners (males performed better than females). However, no significant effects were found for the other type of background noises. The results also indicated that there was no statistically significant difference in intelligibility scores between the condyle and mastoid locations.

Relevance to industry: This study investigated the impact of background noise, gender of listener, and location of bone vibrator on the intelligibility of bone-conducted speech. Findings of this study revealed the impact of different background environments and listener's gender on speech intelligibility and will assist in the development of improved bone conduction devices in the future.

1. Introduction

The soldiers' survival on the battlefield depends to a large extent, on their ability to hear and understand communicated speech clearly. In other words, soldiers need to be able to hear and understand radio messages clearly, receive verbal orders and communicate with other members of their infantry (Rao and Letowski, 2003). The success of any tactical military operation depends heavily on efficient and timely radio communication. Delivering the right information to soldiers at the right time and in a form that is clear can mean the difference between life and death for these brave soldiers.

Consider this scenario where a field commander and his/her troops come under surprise enemy fire. The field commander asks for air support from command and control to neutralize the enemy fire. He is asked to provide the location of the suspected enemy but instead furnishes command and control of his own location because of low speech intelligibility. A short time later, the commander and his/her troops are under intense air bombardment. If the communication from command and control had been intelligible enough, the commander would have provided the location of the enemy instead of his/her own. In light of the above, it can be argued that unintelligible speech is worthless, but low intelligible speech can be worse. It is also worth mentioning that, a speech unit that is loud enough to be heard or audible is not necessarily intelligible. This is because though the speech may be audible, it may also be overly reverberant or distorted in some way, making it unintelligible and therefore useless (Letowski, 2002).

The purpose of this study was to use the CAT to investigate the impact of background noise, gender of listener, and vibrator location on the intelligibility of bone-conducted speech.

* Corresponding author. Tel.: +1 336 334 7780x522; fax: +1 336 334 7729.
E-mail address: xjiang@ncat.edu (X. Jiang).

0169-8141/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ergon.2008.07.003

1.1. Air and bone conduction

Generally, there are two alternate means by which radio communication can be transmitted: air conduction and bone conduction. Air conduction is the natural process by which sound waves are collected by the outer ear and transmitted through the ear canal to the inner ear for processing. Bone conduction, on the other hand, is the mechanical process by which sounds are transmitted from the cranial bones of the head to the inner ear without passing through the ear canal (Studebaker, 1962).

Traditionally, air conduction is more commonly used in industry; however, bone conduction radio communication offers an attractive means for infantry to communicate on the battlefield because it enables auditory signals to be transmitted and received without compromising the soldier's awareness of his or her surroundings. With bone conduction, soldiers can receive and transmit radio communication while maintaining full awareness of the noisy, dangerous and often unpredictable environment of military warfare, and yet maintain portability and privacy. Bone conduction interfaces are also lighter and smaller than earphones and can easily be integrated into military headgear. Developing efficient and reliable bone conduction interface for use in the battlefield will lead to improved situation awareness and enhance the intelligibility of communicated speech.

In bone conduction, the sound waves can be transmitted through vibrations from the skull of the talker to a contact microphone and from a bone vibrator to the skull of the listener (McBride et al., 2005). There are several ways through which hearing by bone conduction can be achieved. The most common method is to place a bone vibrator on the skin covering the skull or by surgically attaching the bone vibrator to the skull such as in bone anchored hearing aids. Hearing by bone conduction can also occur naturally when exposed to loud background or ambient noises which stimulate the cranial bones or when listening to one's own voice (Gripper et al., 2007).

1.2. The impact of gender on speech intelligibility

Although there have been some studies to investigate the impact of gender on speech intelligibility, there is no unanimous consensus on what impact gender really has on intelligibility. Stevens et al. (2005) reported that gender of speaker and the quality of signal affect intelligibility, and that generally the male voice is more intelligible than the female voice. Kelic and Ogut (2004) concluded that female speech is significantly more difficult to discriminate than male speech in normal hearing subjects when noise is present and Cerrato (1995) reported that female speakers tend to be more intelligible than male speakers. From the listener's perspective, Ellis et al. (1996) found no significant difference between male and female listeners' intelligibility scores; however, their overall impression of the intelligibility of the speaker tended to differ. While female listeners indicated that they found male voice more intelligible, male listeners indicated that they found the female voice more intelligible. Wilding and Cook (2000) reported that while males showed no differentiable ability to recognize male over female voices, females showed an enhanced ability to recognize female voices as well as being more accurate overall.

1.3. Speech intelligibility

Speech intelligibility is the percentage of speech units that can be correctly identified by a listener over a given communication system in a given acoustic environment or the degree to which speech can be understood within a given acoustic environment (Letowski et al., 2001). Put differently, it is the degree to which a speaker's intended message is recovered by a listener. Reduced speech intelligibility severely compromises communication and social interaction for affected individuals. Several factors influence speech intelligibility. These include background noise, distortions, reverberations and frequency among others; however, background noise is usually the most important factor to consider. The effectiveness of a communication system or of the ability of people to communicate in noisy environments can be measured by performing speech intelligibility testing.

Speech intelligibility can be measured directly or indirectly. In a direct measurement, a number of talkers will speak words or sentences and a number of listeners will indicate what they hear. However, tape recordings of the speakers are often used rather than live speech, so that different communication systems can be compared with exactly the same speech material (Lower, 1997). In indirect testing, either a speech or a special test signal is broadcast over the system, and the received signal is picked up by a microphone and analyzed to produce the signal and degradation components. A ratio of useful signal to noise signal is computed.

Traditional speech intelligibility tests include the Modified Rhyme Test (MRT), Diagnostic Rhyme Test (DRT), Northwestern University Test number 6 (NU-6), Diagnostic Medial Consonant Test (DMCT) and Central Institute for the Deaf Test (W-22) among others. However, these tests have been criticized for having poor validity in military settings (Blue, 2002). One likely reason for poor performance on these tests in military applications is because most of the tests were developed primarily for clinical diagnostic testing. To address this shortfall, the Callsign Acquisition Test (CAT) was developed.

1.4. Callsign Acquisition Test (CAT)

In response to the criticism of MRT and the other clinical speech tests that have been described above, the Callsign Acquisition Test (CAT) was developed by the Auditory Research Team (ART) of the Army Research Laboratory Human Research and Engineering Directorate (ARL–HRED) specifically for military applications. It consists of 18 two-syllable military code words and seven one-syllable digits. In all, there are 126 items in CAT. The CAT words were derived from NATO and the International Civil Aviation Organization standard word list. However, only the two-syllable words in the list were selected for CAT. The digits are the one-syllable digits from one to eight, except seven which is two-syllable. The CAT uses military code words familiar to soldiers and therefore has greater appeal among military personnel (Blue et al., 2004). Soldiers are therefore more likely to respond correctly when these speech materials are presented. For this reason, the CAT was used as the speech testing material in this research.

2. Methodology

In this research, the condyle (the bony protrusion in front of the ear) and mastoid (the protrusion of the temporal bone behind the ear at the base of the skull) were selected as the bone vibrator placement locations. The selection of the condyle and mastoid for this research was based on the previous study by the authors (Osafo-Yeboah et al., 2006) who investigated the most favorable locations on the head for optimum bone vibrator placement and found the condyle and mastoid as the most favorable locations. Another study conducted by McBride et al. (2005) investigated the sensitivity at 11 bone vibrator locations on the head using pure-tone signals and found the condyle and mastoid were the most favorable and second most favorable locations, respectively, for bone vibrator placement.

Another reason is that both the condyle and the mastoid are close to the ear, and therefore, it is possible for the ear to collect residual sound waves that might leak from the bone vibrator that is

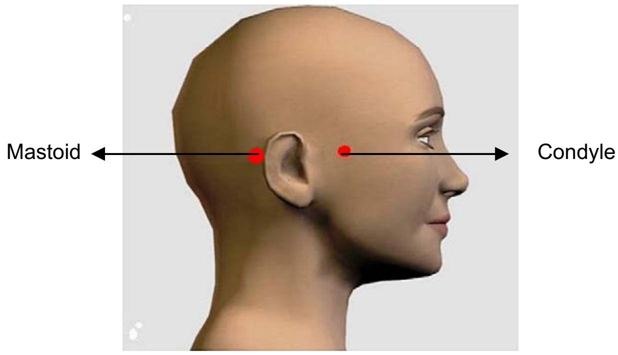

Fig. 1. Condyle and mastoid locations used for the current study.

placed at these two locations. Similarly, both the condyle and the mastoid are located conveniently close to where the headgear rests, thus a bone vibrator could quite easily be incorporated into headgear. The two locations selected for the current research are shown in Fig. 1.

2.1. Participants

Twenty undergraduate and graduate students (10 males and 10 females) enrolled at North Carolina Agricultural and Technical State University (NCATSU) were recruited to participate in the study. Male participants were between the ages 20 and 31 years old, inclusively (mean age = 22.7 and standard deviation = 3.2) and female participants were between the ages 19 and 31 years old, inclusively (mean age = 22.3 and standard deviation = 3.3). In all, there were 10 African American females, nine African American males and one Asian male. The lack of diversity in the participant pool reflects the fact that NCATSU is a Historically Black University. All participants in this study were required to have a normal hearing defined by a threshold better than or equal to 20 dB hearing level (HL) at audiometric frequencies of 250 Hz to 8 kHz (ANSI S3.6, 1996). For each individual, the difference between the left and right ear thresholds had to be 10 dB or less to ensure hearing symmetry. The audiometric screening was performed by the investigator. The screening involved standardized clinical equipment and procedures and complied with the ANSI S3.1 (1997) requirements for audiometric testing under earphones. This was carried out with a Fonix Hearing Evaluator FA-12 Digital Audiometer in a sound treated booth. Participants who passed the audiometric screening were invited to participate in the study.

2.2. Stimulus material

CAT with 60 test items was presented with three background noises (white, pink, and babble). The test was also conducted in quiet (without any background noise) to serve as the basis for comparison. Table 1 shows the 60 callsigns that were randomly presented to listeners. Gripper (2006) reported that this list has the capability of providing the same predictive power as the full CAT with good test–retest reliability.

Three noise files (white, pink and babble), each with overall intensity of 89 dB SPL, were mixed with CAT items using the Sound Forge 8.0® software. The CAT items were presented at 80 dB SPL to produce a signal-to-noise ratio (SNR) of −9 dB. In order to ensure accurate measurement of the sound intensity produced by the bone vibrator, an artificial mastoid was used together with a sound level meter to measure the intensity of the sound output of the bone vibrator.

2.3. Equipment

The equipment used for this study included a Gateway Desktop Computer with a CD ROM drive, Sound Forge 8.0® Software, proprietary CAT software developed at the U.S. Army Research Laboratory to administer CAT items, Dell Desktop Monitor and Keyboard inside an acoustically treated sound chamber (Acoustic Systems 143 MC), step attenuator (Kay Elementrics 839 Attenuator), Radioear B-71 bone vibrator, a pair of Telephonics TDH-39 earphones, digital force gauge, headband, and calibration equipment (Larson Davis Precision Sound Level Meter Kit containing a(n) 824 Sound Level Meter, AMC 493 Artificial Mastoid, and Precision Acoustic Calibrator).

2.4. Experimental design

A $2 \times 2 \times 4$ factorial design was used in this study. Three independent variables, background noise, gender and location, were investigated. Background noise had four levels (white noise, pink noise, babble noise and quiet), location had two levels (condyle and mastoid) and gender had two levels (male and female). The dependent variable was listener's word recognition score, expressed as percentage of words, numbers, and word–number combinations (total callsign) identified correctly. Each of the 20 participants (10 males and 10 females) went through each of the eight treatment conditions. The experiment was completely randomized to remove carry-over effect. The CAT items were randomly presented each time a participant took the test. The following hypotheses will be tested in this study; (1) that the location of bone vibrator has no impact on speech intelligibility, (2) that the gender of listener has no impact on speech intelligibility, (3) that background noise has no impact on speech intelligibility, and (4) that there is no interaction effects between gender, location and background noise.

2.5. Procedure

All participants were briefed on the purpose of the research followed by verbal instructions and a chance for them to ask questions. All volunteers in the study were required to have normal hearing. Volunteers who passed the hearing evaluation were asked to read and sign consent forms in order to participate in the study. They were seated at the listener's station in a sound treated booth in front of a Dell Desktop Computer (that displays the CAT items) with keyboard for data input.

A bone vibrator was placed at the listener's condyle or mastoid and fastened in place with a headband (Fig. 2). A digital force gauge

Table 1
Sixty CAT items used in this research.

Alpha 1	Charlie 4	Hotel 1	Kilo 5	Lima 8	Papa 2	Tango 2	Victor 6	Whiskey 8	Zulu 2
Alpha 2	Charlie 5	Hotel 3	Kilo 6	Oscar 2	Papa 3	Victor 1	Victor 8	Yankee 3	Zulu 3
Bravo 2	Charlie 6	Hotel 8	Kilo 8	Oscar 3	Papa 4	Victor 2	Whiskey 1	Yankee 4	Zulu 4
Bravo 3	Charlie 8	Kilo 1	Lima 3	Oscar 4	Quebec 2	Victor 3	Whiskey 2	Yankee 5	Zulu 5
Bravo 8	Echo 2	Kilo 3	Lima 4	Oscar 5	Quebec 3	Victor 4	Whiskey 4	Yankee 8	Zulu 6
Charlie 3	Echo 8	Kilo 4	Lima 5	Papa 1	Tango 1	Victor 5	Whiskey 6	Zulu 1	Zulu 8

Fig. 2. A participant taking test with bone vibrator attached to condyle.

was used to measure the force exerted by the headband to ensure that adequate, but not unnecessarily high and uncomfortable force was applied to the head of the listener. For this experiment, a static force of 3.5–3.9 N (to ensure consistency) was used in accordance with audiology related literature, which recommends that the static force applied by bone vibrator to the human head must fall within the range 2.5 N (minimum required for stable position) and 5.9 N (level of discomfort) (ANSI S3.5, 1997).

Participants were asked to listen to incoming CAT items (displayed on the monitor in front of them) and identify them by pressing the appropriate keys on their keyboard. For example, if a listener heard the callsign "Quebec 3", the correct response would be to press the "Q" key, followed by the "3" key and then the "Enter" key.

Listeners were instructed to make their best guess if they were unsure of what they heard, because there was no repetition of signals and the next signal was not presented until a response to the current signal was given. A screen shot of what listeners saw on their computer screen while taking the test is shown in Fig. 3.

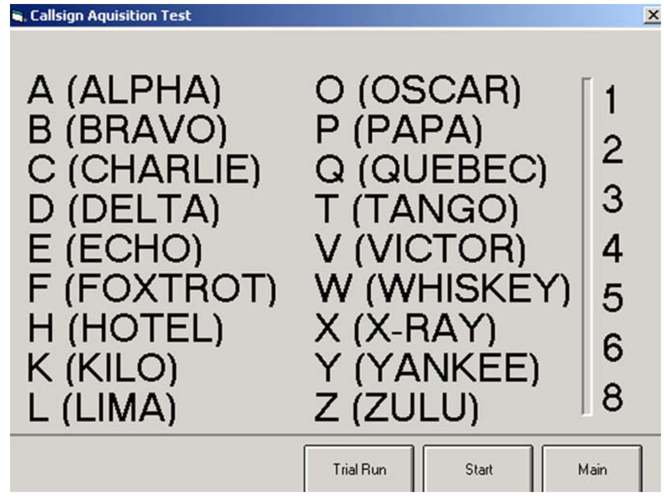

Fig. 3. Screen shot of CAT items on participant's computer.

At the start of the experiment, the experimenter clearly explained the procedure to participants and showed them how to enter their responses by running a trial. At the end of the trial, all questions were answered and the actual experiment commenced. Listeners' responses during the experiment were stored in a file and subsequently imported into an Excel spreadsheet for analysis.

2.6. Data collection

Both objective and subjective data were collected in this study. In order to compile a subjective assessment of participants, pre-test and post-test questionnaires were used. Each participant who passed the hearing evaluation was given a pre-test questionnaire to complete. This was used to collect biographical data on participants. Also, after completing the test, each participant was debriefed, thanked and given a post-test questionnaire to complete.

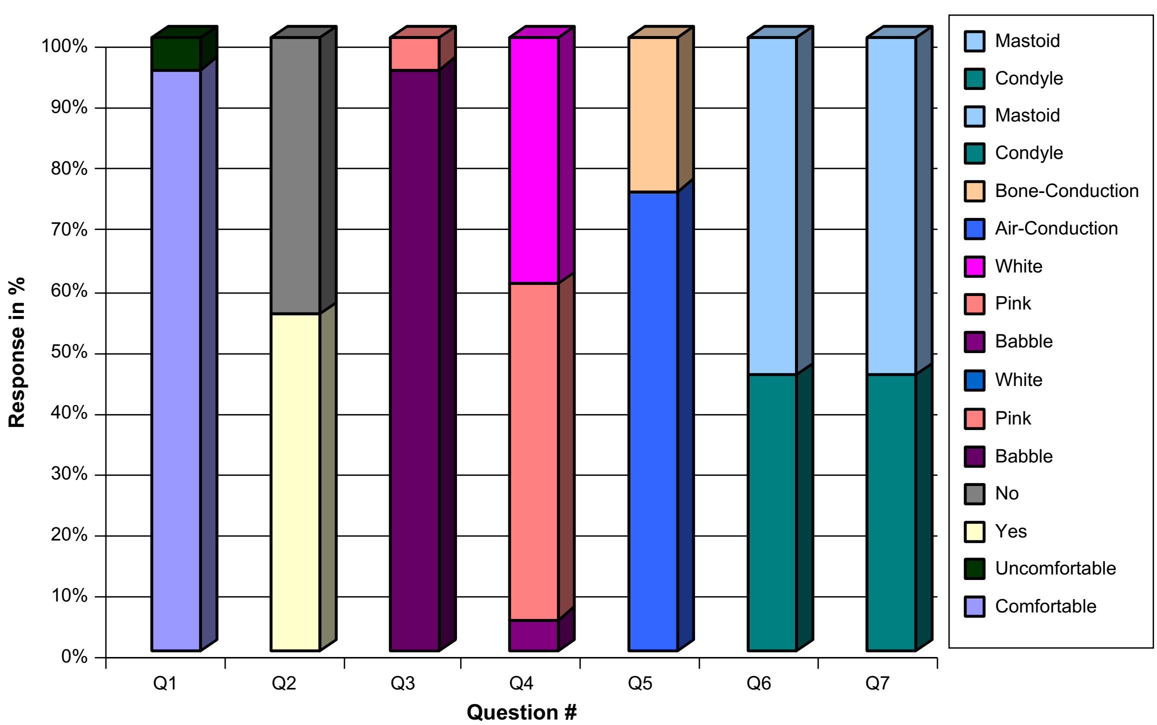

Fig. 4. Graphical presentation of subjective assessment data.

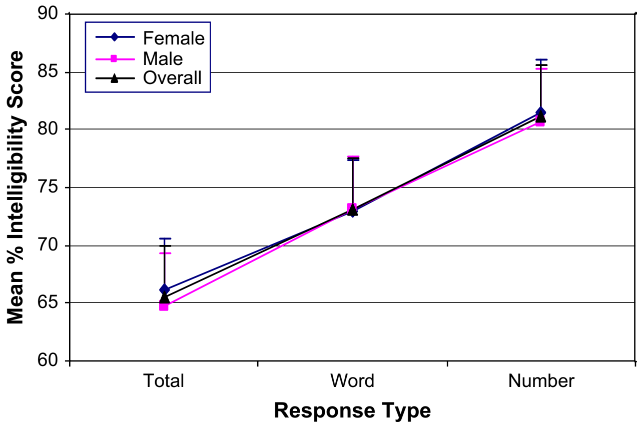

Fig. 5. Mean % intelligibility scores across gender for each response type.

This was used to compile a subjective assessment of participants. The objective data collected from the experiment were the percentage of words, numbers, and word–number combinations correctly identified by participants.

3. Results

3.1. Subjective data

Data from the post-test questionnaire were collected from all 20 participants and analyzed. Fig. 4 provides a brief graphical overview of the subjective assessment results. The columns shown in Questions 1–7 of Fig. 4 represent the percentages of responses to the subjective assessment questions obtained from participants. A detailed description of the subjective assessment follows.

Of the 20 participants tested, 95% felt comfortable using the bone vibrator and 5% felt the bone vibrator was uncomfortable to use (Fig. 4, Column Q1). All male participants felt comfortable listening with the bone vibrator and 90% of female participants felt comfortable wearing the bone vibrator. Majority of female participants (70%) said they would be comfortable wearing the bone vibrator for long-term use (at least 4 h of use), while a smaller percentage of male participants (40%) responded that they would be comfortable for such long-term use of the bone vibrator. Overall, 55% of the participants said they would be comfortable using the bone vibrator long-term (Fig. 4, Column Q2). In terms of background noise, 95% of the participants felt that the signals were most difficult to hear in babble while 5% felt that the signals were most difficult to hear in pink noise (Fig. 4, Column Q3). For clarity of signals, 55% of the participants felt that the signals were most clear in pink noise, 40% felt that the signals were most clear in white noise and 5% felt that the signals were most clear in babble (Fig. 4, Column Q4).

Lastly, the majority of females (70%) felt more comfortable with the bone vibrator placed on their condyle and 30% preferred the bone vibrator at the mastoid location. However, only 20% of male participants felt more comfortable with the bone vibrator placed at the condyle location. The majority of male participants preferred the bone vibrator placed at the mastoid location. Overall, 55% of the participants felt more comfortable when they wore the bone vibrator at the mastoid location while 45% of the participants felt that the bone vibrator was more comfortable to wear at the condyle location (Fig. 4, Column Q6). Similarly, the majority of female participants found signals more intelligible when the bone vibrator was placed at the condyle location whereas 80% of male participants felt that the signals were more intelligible when the bone vibrator was placed at the mastoid location (Fig. 4, Column Q7).

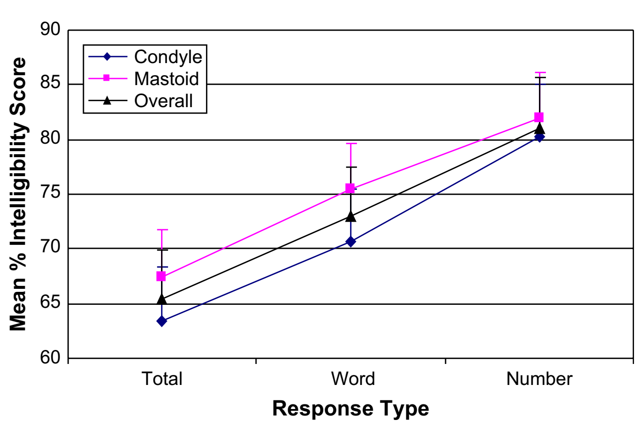

Fig. 6. Mean % intelligibility scores across location for each response type.

3.2. Objective data

Descriptive statistics for the three dependent variables obtained from the experiments are summarized below. The results showed that female participants scored slightly higher than their male counterparts in quiet, pink and white noise backgrounds; however, in babble/multi-talker noise, the percentage correct scores for male participants were slightly higher than those of the female participants. Considering all three independent variables, participants' scored highest on numbers correctly identified, followed by words correctly identified and word–number combination correctly identified. This finding did not come as a surprise as several previous studies by Rao and Letowski (2003), Osafo-Yeboah et al. (2006), and Gripper et al. (2007) who all found that numbers are more intelligible than both words and complete callsigns.

Participants' intelligibility scores in quiet environment were as expected, very high, however, this was not the case when background noises were introduced into the listening environment. Of the three background noises investigated, intelligibility scores were highest in pink noise (mean = 81.42%), followed by white noise (mean = 66.99%) and least in babble (mean = 44.21%). This was expected because babble is a speech noise and, therefore, has linguistic content similar to the target signal (callsigns). Babble, thus, had a greater ability to mask the target speech, and as a result tended to have a greater impact on speech intelligibility than the nonlinguistic (pink and white) background noise. Figs. 5 and 6 provide graphical representations of the intelligibility scores across gender and location, respectively.

3.3. Hypotheses testing and inferential statistics

Prior to the use of any parametric statistical analysis, a normality check was performed with the Shapiro–Wilk's test on each of the 16 combinations of data sets to ensure that no obvious normality violations exist. The Shapiro–Wilk's test statistic for numbers, words and total callsigns was compared to 0.05 significance level ($\alpha = 0.05$), to determine whether or not the data sets were normally distributed or otherwise. Normal probability plots of percentage correct number, word and number–word combination (total

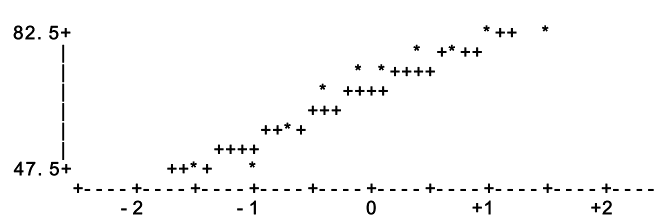

Fig. 7. Normal probability plot of callsign number for male–mastoid–babble data set.

Table 2
ANOVA results for 'callsign number'.

	SS	df	MS	F	p
Location	126.02	1	126.02	0.42	0.5187
Gender	28.90	1	28.90	0.09	0.7639
Noise	35,303.72	3	11,767.90	121.17	<0.0001
Location × gender	32.40	1	32.40	0.34	0.5586
Gender × noise	1014.25	3	338.08	3.59	0.0154
Location × noise	325.62	3	108.54	1.15	0.3306
Location × gender × noise	49.25	3	16.41	0.17	0.9137

callsign) were plotted for each data set, to expose obvious normality violations.

For example, the data set with gender = male, location = mastoid, and background noise = babble, the Shapiro–Wilk's (W) test statistic for CAT-number was: $W = 0.8848$, $p = 0.1482$, the test is not significant and, therefore, normality is not violated. The normal probability plot is shown in Fig. 7. Finally, a two-way ANOVA was used to analyze the results for each dependent variable as shown in the sections below. All statistical analyses were conducted using the SAS software.

Tables 2–4 show the summary of main effects and interaction effects of the statistical analysis. As seen in Tables 2 below, there was no significant main effect for location ($F_{1,159} = 0.42$, $p = 0.5187$) or gender of listener ($F_{1,159} = 0.09$, $p = 0.7639$). However, there was a significant main effect for background noise ($F_{3,159} = 121.17$, $p < 0.0001$). Similarly, there was no significant interaction effect between location of vibrator and gender of listener ($F_{1,159} = 0.34$, $p = 0.5586$) or location and background noise ($F_{3,159} = 1.15$, $p = 0.3306$); however, there was a significant interaction effect between gender of listener and background noise ($F_{3,159} = 3.59$, $p = 0.0154$). Detailed description of the analysis is provided in the following subsections.

3.3.1. Impact of bone vibrator location on speech intelligibility

Results from two-way ANOVA show that there is no statistically significant difference in the intelligibility scores obtained for the condyle and the mastoid locations. For numbers correctly identified, the means were 82.02% and 80.20%, respectively, for the mastoid and condyle with $F_{1,159} = 0.42$, $p = 0.5187$. Words correctly identified had means of 75.48% and 70.58%, respectively, for the mastoid and condyle location ($F_{1,159} = 1.61$, $p = 0.2059$). For total callsign, the means were 67.56% and 63.33%, respectively, for the mastoid and condyle ($F_{1,159} = 0.90$, $p = 0.3447$). From the above analysis, there is no enough statistical evidence that location of bone vibrator placement on the head impacts speech intelligibility and the null hypothesis cannot be rejected. Fig. 8 shows the mean percentage intelligibility scores at the condyle and mastoid locations across different background environments.

3.3.2. Impact of listener's gender on speech intelligibility

The ANOVA results show that there is no statistically significant difference in the intelligibility scores of male and female participants. For callsign number correctly identified, the mean intelligibility scores were 80.68% and 81.53%, respectively, for male and female participants ($F_{1,159} = 0.09$, $p = 0.7639$). For callsign words correctly identified, mean scores were 73.16% and 72.91%, respectively, for male and female participants ($F_{1,159} = 0.00$, $p = 0.9487$); and for total callsigns correctly identified, mean scores were 64.80% and 66.10% for male and female participants, respectively ($F_{1,159} = 0.08$, $p = 0.7715$). From the analyses above, there is no enough statistical evidence that gender of the listener has an impact on speech intelligibility scores. Fig. 9 shows the mean intelligibility scores of male and female participants across different background environments.

Table 3
ANOVA results for 'callsign word'.

	SS	df	MS	F	p
Location	902.50	1	902.50	0.90	0.3447
Gender	2.50	1	2.50	0.00	0.9487
Noise	70,899.67	3	23,633.22	152.95	<0.0001
Location × gender	11.02	1	11.02	0.07	0.7864
Gender × noise	1176.30	3	392.10	2.62	0.0530
Location × noise	337.70	3	112.56	0.75	0.5225
Location × gender × noise	136.87	3	45.62	0.31	0.8217

Table 4
ANOVA results for 'total callsign'

	SS	df	MS	F	p
Location	664.22	1	664.22	0.90	0.3447
Gender	67.60	1	67.60	0.08	0.7715
Noise	98,174.60	3	32,724.86	181.51	<0.0001
Location × gender	4.22	1	4.22	0.02	0.8768
Gender × noise	1622.20	3	540.73	3.09	0.0292
Location × noise	397.07	3	132.35	0.76	0.5207
Location × gender × noise	154.47	3	51.49	0.29	0.8296

3.3.3. Impact of background noise on speech intelligibility

Results from ANOVA performed on the data showed that there is a significant difference in speech intelligibility scores across the three background noises investigated. For callsign number correctly identified, the means were 88.97%, 75.42%, 60.20% and 99.85% for pink, white, babble, and quiet, respectively ($F_{3,159} = 121.17$, $p < 0.0001$). The statistical analyses indicate that there is a significant difference in speech recognition scores of callsign numbers. For callsign words correctly identified, the mean scores were 81.45%, 69.10%, 41.85% and 99.75%, respectively, for pink, white, babble and quiet backgrounds ($F_{3,159} = 152.95$, $p < 0.0001$). For number–word combination (total) callsign, the mean scores were 73.90%, 56.45%, 31.80% and 99.65%, respectively, for pink, white, babble and quiet ($F_{3,159} = 181.51$, $p < 0.0001$). Also, a post-hoc Tukey test revealed that all the mean intelligibility scores were significantly different from each other.

From above, there is evidence to conclude that background noise does indeed have a significant impact on speech intelligibility as stated in the alternate hypothesis. Fig. 10 graphically illustrates the change in intelligibility scores for number, word and total callsign across background environments investigated.

3.3.4. Gender of listener and background noise interaction effect

The ANOVA results show that for callsign number, there is a significant interaction between gender of the listener and background noise ($F_{3,159} = 3.59$, $p = 0.0154$); however, there is no

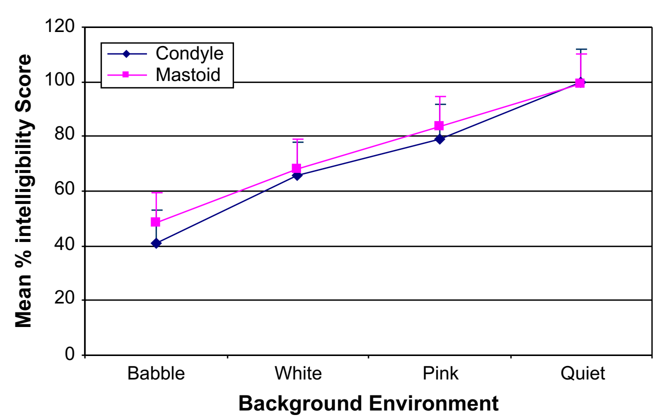

Fig. 8. Mean % intelligibility scores at condyle and mastoid locations across different background environments.

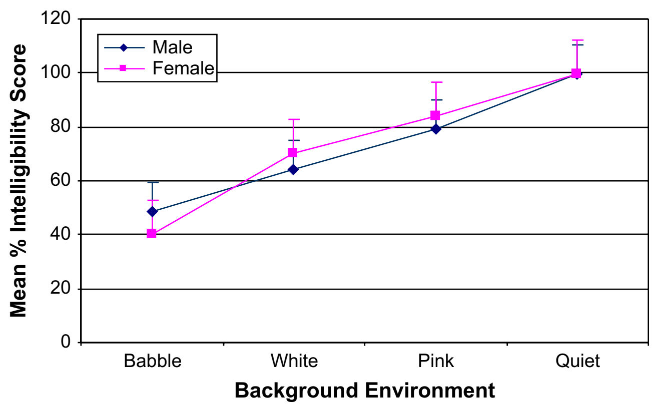

Fig. 9. Intelligibility scores of male and female participants across different background environments.

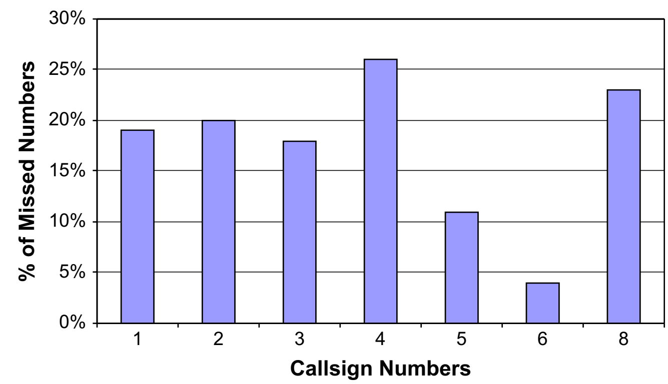

Fig. 11. Percentage of misses for each callsign number.

significant interaction between gender and background noise for callsign words ($F_{3,\ 159} = 2.62$, $p = 0.050$). Similarly, the interaction between gender and background noise is significant for total callsign ($F_{3,\ 159} = 3.09$, $p = 0.0292$).

To further investigate the interaction effect of gender and background noise, a post-hoc slicing was performed using SAS software. One variable was kept at a fixed level so that the effect of the other variable could be investigated. The results indicate that (1) background noise has a significant impact on speech intelligibility regardless of gender ($F_{3,\ 159} = 94.98$, $p < 0.0001$); and (2) when the test was performed in babble, gender had a significant impact on the intelligibility of callsign words (males performed better than females, $F_{1,\ 159} = 4.32$, $p = 0.0394$); however, there was no significant difference in intelligibility between males and females for the other background conditions.

4. Discussion

The results from this research showed that among the three dependent variables (callsign numbers, callsign words and complete callsign) investigated, numbers were consistently the most intelligible irrespective of the background environment, gender of the listener or bone vibrator location being tested, while word–number combination was the least intelligible. Among callsign numbers, the number 6 was the most intelligible (missed only 3.85% of the time); this was followed by 5 and 3, respectively, with 11.25% and 18.12% of misses. On the other hand, the number 4 was the least intelligible number with a 25.83% miss rate, followed by 8 with a 22.68% miss rate and 2 with a 19.81% miss rate. Fig. 11 provides a histogram of the percentages of missed callsign numbers.

In terms of callsign words, Hotel was the least intelligible with a miss rate of 44.37%, Echo and Quebec followed with miss rates of 39.68% and 37.81%, respectively; while the most intelligible callsign words were Zulu, Whiskey and Oscar, respectively, with 11.33%, 19.37% and 19.83% miss rates. One of the reasons for better speech intelligibility scores on callsign numbers may be familiarity. Since participants are usually more familiar with numbers from their everyday interactions, it is reasonable that they are able to recognize numbers with a lot more ease than callsign words with which they are less familiar. Fig. 12 is a histogram of callsign words and their percentage of misses. For complete callsign, Hotel 3 was the least intelligible with a miss rate of 54.37%; this was closely followed by Charlie 3 and Quebec 3 with miss rates of 53.75% and 51.25%, respectively.

On the other hand, the most intelligible complete callsign was Zulu 3 which had a miss rate of 9.37% followed by Yankee 3 and Zulu 2, each with 18.75% miss rate. Fig. 13 shows a histogram of complete callsigns and their percentages of misses.

Results show that among the independent variables investigated, background noise had the most impact on speech intelligibility and participants performed worst in babble. This could be attributed to the fact that there was greater similarity between the frequency and the vocal track characteristics of the voices that made up babble and the callsigns (target speech). Babble was, therefore, more effective in masking the callsigns than pink noise and white noise.

In general, intelligibility scores in pink noise were about twice the scores in babble, while the scores in white noise were about one and a half times the scores in babble. Just as their performance indicated, the majority of participants (95%) felt the signals were more difficult to understand in babble noise. Similarly, the majority of respondents (55%) felt signals were most clear in pink noise.

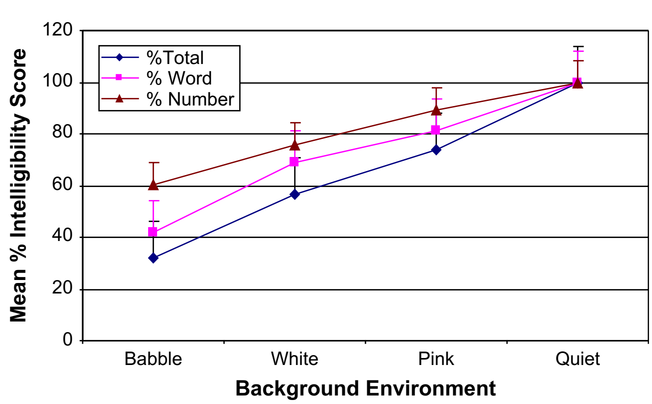

Fig. 10. Mean % intelligibility scores at condyle and mastoid locations across different background environments.

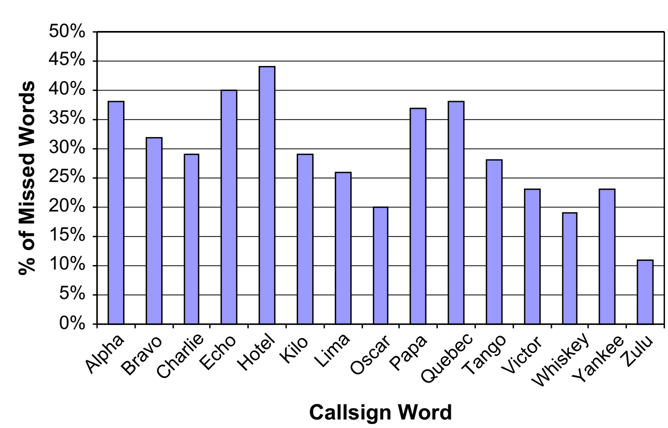

Fig. 12. Percentage of misses for each callsign word.

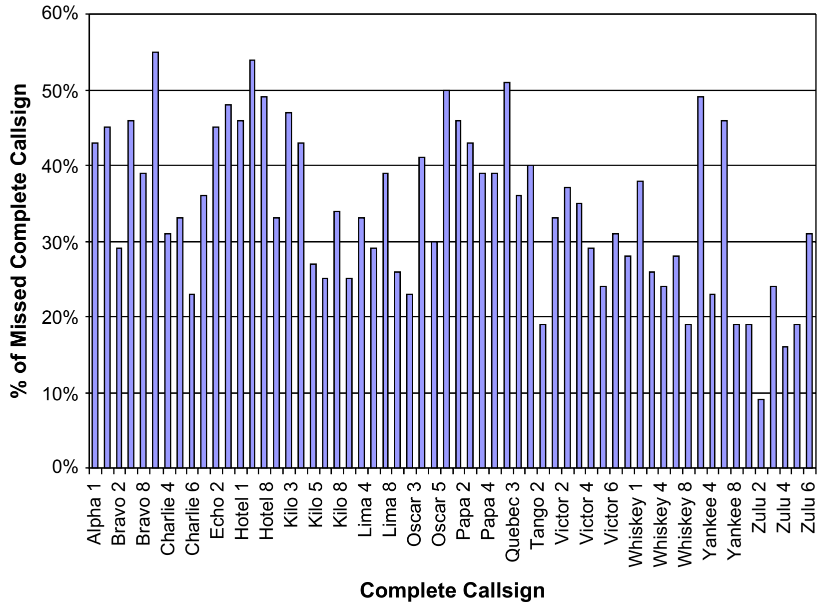

Fig. 13. Percentages of missed complete callsigns.

In terms of gender, female participants performed slightly better than their male counterparts in pink and white noise environments; however, in the babble environment, the performance of male participants was slightly better than those of the female participants. These differences were, however, not statistically significant. This finding seems to agree with Nixon et al. (1996) who reported that although the intelligibility of male and female speech is approximately equal under ordinary noise conditions, the intelligibility of the female speech was significantly lower at high noise levels. In answering the pre-test questionnaire, about 40% of male participants indicated that they occasionally listen to loud music in their cars and at home; however, none of the female participants indicated listening to loud music on such a regular basis. The author believes this could be one of the possible reasons why male participants performed slightly better than their female participants in babble noise since they have more practice segregating speech from noise.

With regard to location of the bone vibrator, results suggest that the intelligibility scores at the mastoid location were slightly higher than those at the condyle location; however, this difference was not statistically significant. This finding came as a surprise because a previous study by the author found the intelligibility scores at the condyle to be slightly higher than those at the mastoid. However, in both studies, the difference in intelligibility scores was not found to be statistically significant. Analysis of the subjective post-test questionnaire shows that 55% of participants (80% of males and 30% of females) felt more comfortable when they wore the bone vibrator at the mastoid location. This perceived mindset that it was more comfortable to wear the bone vibrator at the mastoid location may have indeed contributed to the slightly better performance.

5. Conclusion

The current research investigated the impact of background noise, gender of listener and location of bone vibrator on speech intelligibility. The results show that background noise has a significant impact on speech intelligibility. Among the three background noises investigated, babble had the most significant impact on speech intelligibility due to its ability to mask the target signal, while pink noise had the least impact. The results also show that gender of the listener does not have significant impact on speech intelligibility. This suggests that any successful bone conduction vibrator developed in the future could be used effectively by both male and female listeners in a comparable manner.

Finally, the results show that there is no statistically significant difference in intelligibility scores between the condyle and mastoid locations. Expanding this research to include multiple bone vibrator locations in future research will be a logical follow-up to the present study. The study also shows that gender of the listener and background noise have a significant interaction effect on speech intelligibility. Probing the effect of interaction showed that background noise had a significant impact on speech intelligibility regardless of the gender of listener; however, gender of the listener had a significant impact on speech intelligibility only when the test was carried out in babble noise. Male participants performed slightly better than female participants in this environment; however, the reverse was true when the test was performed in white noise or pink noise.

Acknowledgments

The authors would like to thank Dr. Misty Gripper from Wright State University for her expertise. The authors would like to thank Dr. Elmar Schmeisser from the US Army Research Office for his assistance of this research. The research reported in this paper was performed in connection with Grant number W911NF-05-1-0518 with the U.S. Army Research Office. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Office.

References

ANSI S3.5, 1997. Methods for the Calculation of the Speech Intelligibility Index. American National Standards Institute, New York.
ANSI S3.1, 1997. Criteria for Permissible Ambient Noise During Audiometric Testing. American National Standards Institute, New York.
ANSI S3.6, 1996. American National Standard Specification for Audiometers. American National Standards of the Acoustical Society of America, New York.
Blue, M., 2002. Speech Intelligibility of the Callsign Acquisition Test in a Quiet Environment. Master's Thesis, North Carolina A&T State University.
Cerrato, L., 1995. Using a Multitalker Database to Investigate the Speaker-Related Characteristics of Speech Intelligibility. NAL. Annual Report.

Ellis, L., Reynolds, L., Fucci, D., Benjamin, B., 1996. Effects of gender on listeners' judgments of speech intelligibility. Perception and Motor Skills 83 (3), 771–775.

Gripper, M.A., 2006. Evaluations of the Callsign Acquisition Test CAT in Acoustic Environments. Doctorial Dissertation, North Carolina A&T State University.

Gripper, M., McBribe, M., Osafo-Yeboah, B., Jiang, X., 2007. Using the Callsign Acquisition Test (CAT) to compare the speech intelligibility of air versus bone conduction. International Journal of Industrial Ergonomics 37 (7), 631–641.

Kelic, M.A., Ogut, F., 2004. The effect of the speaker gender on speech intelligibility in normal-hearing subjects with simulated high frequency hearing loss. Rev Laryngol Otol Rhinol (Bord). 125 (1), 35–38.

Letowski, T., 2002. Performance Intensity Function for the Callsign Acquisition Test (CAT). Unpublished Research Protocol. Aberdeen Proving Grounds, MD.

Letowski, T., Karsh, R., Vause, N., Shilling, R., Ballas, J., Brungart, D., McKinley, R., 2001. Human Factors Military Lexicon: Auditory Displays. U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Grounds, MD (unpublished technical report).

Lower, M., 1997. Communications and speech intelligibility. Available from: http://www.isvr.co.uk/spe_/hearcomint.htm (accessed 22.11.06).

McBride, M., Letowski, T.R., Tran, P.K., 2005. Bone Conduction Head Sensitivity Mapping: Bone Vibrator. Army Research Laboratory, Aberdeen Proving Ground.

Nixon, C., Anderson, T., Morris, L., McCavitt, A., McKinley, R., Yeager, D., McDaniel, M., 1996. Female voice communications in high level aircraft cockpit noises – part 1: spectra, levels, and microphones. Journal of Speech and Hearing Research 39 (6), 1159–1170.

Osafo-Yeboah, B., Gripper, M., McBride, M., Jiang, X., 2006. Effect of Bone Conduction Vibrator Placement on Speech Intelligibility Using the Callsign Acquisition Test – A Pilot Study. In: Proceedings of the 2006 IEMS Conference, Coco Beach, FL, pp. 234–240.

Rao, M., Letowski, T., 2003. Speech Intelligibility of the Callsign Acquisition Test (CAT) for Army Communication Systems. In: Paper Presented at the 114th Convention of the Audio Engineering Society.

Stevens, C., Lees, N., Vonwiller, J., Burnham, D., 2005. On-line experimental methods to evaluate text-to-speech (TTS) synthesis: effects of voice gender and signal quality on intelligibility, naturalness and preference. Computer Speech and Language 19, 129–146.

Studebaker, G.A., 1962. Placement of vibrator in bone-conduction testing. Journal of Speech and Hearing Research 5 (4), 321–331.

Wilding, J., Cook, S., 2000. Sex differences and individual consistency in voice identification. Perceptual and Motor Skills 91 (21), 535–538.