

# Horizontal plane sound source localization and auditory enhancement

Jan R. Smith<sup>a,\*</sup>, Wesley R. Lombard<sup>a</sup> and Moses N. Shaba<sup>a</sup>

<sup>a</sup>*Ergonomics Technologies, PO Box 6264, Pretoria, 0001, South Africa*

**Abstract.** Accurate sound source localization has advantages for the performance of work by humans. The ability to accurately localize sound sources contributes to perception, decision making and task performance. Two studies were conducted to investigate the prevalence of accurate sound source localization and the enhancement that spatially separated sound source locations can have on speech perception. The first study was conducted to characterize the ability to detect the location of horizontal plane sound sources. A sample of 117 participants with the hearing capacity within the normal limits participated in the study. The results indicated that sound sources located towards the front of the participant were identified more frequently than those sound sources located towards the rear positions. Based on the results found in the first study, a second study was conducted to assess performance within a listening task. Three different spatial configurations were used to assess if similar trends in performance translated to sound sources through headphones. Fifteen research participants performed a Coordinated Response Measure (CRM) task requiring the identification of a speech phrase and its associated information for a diotic configuration and two different spatial sound source configurations. Performance measured for the diotic configuration was significantly ( $p \leq 0.05$ ) less than for the two spatial configurations. The current studies indicate distinct advantages of utilizing localized sound sources to present auditory signal and speech to listeners.

Keywords: diotic, binaural, CRM

## 1. Introduction

Listeners, in military and industrial environments are often required to monitor and respond to auditory stimuli originating from sound sources in their immediate vicinity. Accurate sound source localization contributes to the development of individual situation awareness [1], a primary construct for improved decision making and effective task performance. The inability to accurately differentiate between sound sources can lead to increases in cognitive workload and decreases in effectiveness of task performance. Sound source localization in humans is typically described as a coordinate point with characteristic azimuth, elevation and distance relative to the listeners head position in space [2]. Horizontal plane localization refers to the azimuth and distance components of the sound source location coordinates when the elevation of the source is equal to the head height. The cues that are typically used to predict the sound

source distance include the sound intensity, the frequency content and the decay characteristics of the sound [1]. Of these cues, sound intensity is the primary cue for distance estimation [2]. There is strong evidence that when a sound source is relatively distant from the listener, the auditory cues that allow a listener to localize the sound are largely independent of distance [3, 7]. Beyond a distance of 1 meter the interaural level differences at the ears are not contributing to the estimation of distance to the sound source [3, 6]. Therefore, if the distance to the sound source is not a protocol requirement and the influences of the distance to the perception of sound are to be minimized, the sound source location beyond the 1 meter limit provides specific experimental benefits. Since distance of the sound source beyond 1 meter does not contribute to the estimation of sound source localization, this leaves azimuth determination as the remaining component of horizontal plane localization. Interaural time differences and interaural intensity differences are the primary cues for azimuth determi-

---

\* Corresponding author. E-mail: janryno@ergotech.co.za

nation [2]. An important contributing cue may come from the listener's head movement, since the resulting binaural perceptual differences increase the accuracy with which the localization of a static sound source can be estimated [19]. The objective of the first study was therefore to characterize the frequency of accurate sound source localization for sound source stimuli of long duration from a fixed distance beyond 1 meter at any azimuth around the listener.

Within real-world environments, the timing and intensity of the sound source reaching the ear enable the listener to take advantage of binaural cue differences that occur when competing sound sources are presented from different locations. Human listeners are able to utilize binaural cue differences to separate the speech message when competing talkers are spatially separated. This so-called "cocktail party effect" allows a listener to perform much better when they are listening to multiple sound sources at the same time.

Within standard military multi-talker communication systems the speech signals are generally electronically mixed into a single signal that is presented diotically to listeners over headphones [5, 21]. Research supports the notion that the efficiency of multi-talker communications can be improved by audio displays that spatially separate competing talkers [5, 12] and therefore mimic the real world "cocktail party effect".

Based on the knowledge gained from the first study, a second methodology was developed to assess the impact of a two dimensional horizontal plane auditory display configuration on a listener's performance. Various spatial configurations were assessed for a communication task to study the impact of spatial configuration on performance. The study aimed to measure performance differences on a coordinated response measure (CRM) task between a diotic audio configuration and two spatial configurations while using headphones.

## 2. Methodology

Ethics approval was obtained to perform the studies. All participants were required to volunteer, provide informed consent and have normal hearing. The participants underwent audiometric screening to assess their magnitude of hearing loss.

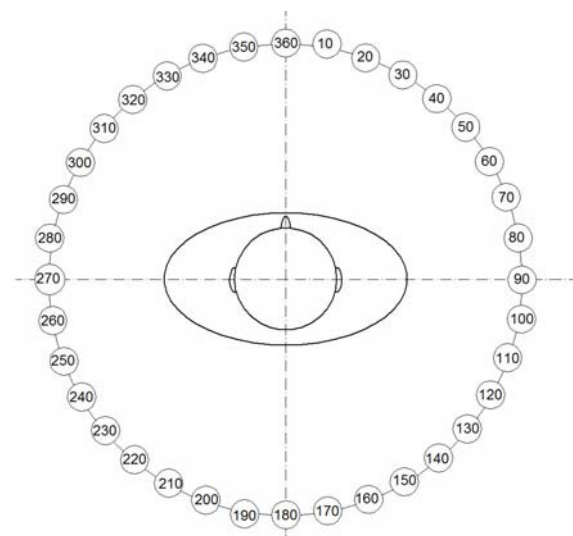


Fig. 1: Subject positioning and loudspeaker locations.

Only participants with a hearing capacity within the normal limits were accepted as research participants to perform the sound source localization and the CRM tasks. For the purpose of these studies normal hearing was defined as having a pure-tone average hearing loss  $\leq 25$  dB.

### 2.1. Sound source localization

The sound source localization study was conducted in a rectangular 10 meter long x 8 meter wide x 6 meter high anechoic chamber. The participants were tested individually while seated in the center of a 4 meter radius circle in the anechoic chamber.

Each participant sat upright at the center of the circle facing the 360 degree mark with the back aligned with the 180 degree mark. The participant's ear position was aligned with the 270 degree mark to the left and 90 degree mark to the right as shown in Figure 1.

#### 2.1.1. Stimuli and equipment

A broadband white noise generator device was custom developed to present the sound to the participants. The sound source was presented through any one of 36 high resolution ultra linear reference studio loudspeakers (Behringer B2031P) with a response frequency range of 55 Hz to 21 kHz. The loudspeakers were equidistantly positioned on a 4 meter radius in a circular pattern around the participant, separated center to center by 10 degrees azimuth. The locations of loudspeakers were covered with cloth to hide their exact positions and to prevent visual detection

of the target sound source. The azimuth labeled markers were placed at every 5 degree position relative to the center where the participant was seated. The loudspeakers were placed on loudspeaker stands which were set at 900 millimeter above the floor. The top level of speaker from the floor was 1300 millimeter. The participant sat on a 495 millimeter high stool. The height of the loudspeakers was not adjusted for the individual's seated ear level.

### 2.1.2. Procedure

The presentation of the sound was driven by computer software which randomly presented white noise through any one of the loudspeakers at a short burst (duration of one second) for fast localization and for a period of two seconds for slow localization at a fixed intensity sound pressure level of 33 dB(A). The sound pressure level was not changed for the test conditions throughout the evaluation period. The sequence of sound presentation to each loudspeaker was repeated randomly and was different for fast and slow localisation. The participants were allowed to move their heads at any time after the stimulus was presented. The participants were allowed free head movement and upper body rotation when pointing at the perceived location of the sound source and had to return to the original position before the next stimulus was transmitted. The participants responded by using a laser pointer to indicate the perceived sound source. An observer recorded the position of the indicated sound sources to the closest azimuth label.

The data per participant were categorized into two groups for each of the comparisons; for example, observations to the front of the participant was compared to observations to the rear, left rear observations to right rear observations and left front observations to right front observations for both fast and slow localizations. The differences in the mean values of the percentage of accurate responses for the sound source comparisons were evaluated using the dependent t-test procedure at a significance level of  $p \leq 0.05$ .

### 2.2. Coordinated Response Measure (CRM) task

The methodology of the second study was developed based on the findings of localization studies where sound sources are presented at a distance greater than one meter from the head. The findings indicated that sound sources located in front of the participant (arc of 180 degrees) produced a signifi-

cantly higher accuracy in localization of the sound sources than compared to those presented behind the participant. The aim therefore was to assess if similar trends in performance translated to sound sources recorded further than 1 meter from the listener when presented through headphones. Nineteen research subjects participated in the study. Four males were used as human speakers in the recording of the CRM task. Fifteen listeners participated in the performance measurement protocol of the pre-recorded CRM task. The sample consisted of 11 males and 4 females within an age range of 20 years to 26 years.

#### 2.2.1. Stimuli and equipment

Moore's [20] CRM task, which was developed to provide greater operational validity for military communication tasks as compared to standard speech intelligibility tests, was used as the performance measure. In the CRM task, the listener hears one or more simultaneous phrases of the form "Ready, (Call Sign), go to (color) (number) now" with one of eight call signs ("Baron", "Ringo", "Eagle", "Arrow", "Hopper", "Tiger", and "Laker"), one of four colors (red, blue, green, and white), and one of eight numbers (1 – 8). The listener's task was to then listen for the target sentence containing the call sign "Baron" and respond by identifying the color and number combination which follows the target phrase.

For the CRM recordings, microphones were mounted inside a dummy head which was fitted with KEMAR manikin pinnae and placed in the center of an anechoic chamber.

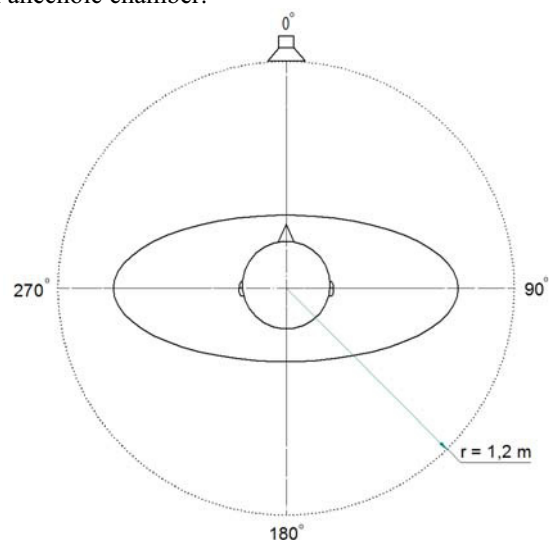


Fig. 2: Control condition.

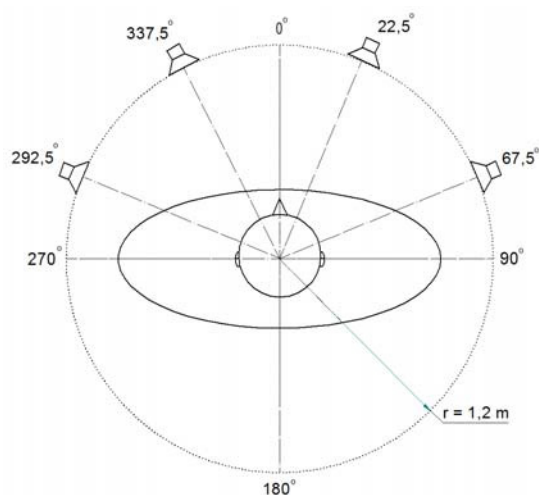


Fig. 3: Four-front condition.

Individually, each human speaker stood on a predefined marker 1.2 meter away from the dummy head and read the CRM task for each particular condition. The diotic condition formed the control condition.

Each human speaker was positioned in front of dummy head at 0 degrees as shown in Figure 2. In the Four-Front condition, the human speakers were positioned at equidistant points in front of the dummy head as shown in Figure 3, and for the front-back condition, the human speakers were positioned at equidistant points on the left and right of the dummy head as shown in Figure 4. For the control test condition each participant individually read the CRM list while standing at the 0 degrees mark as indicated in Figure 2. The two spatial configuration conditions consisted of one condition with four sources equidistantly spaced in front of the dummy head as indicated in Figure 3 and one configuration with two to the front and two to the rear of the dummy head (Figure 4). Individually, each human speaker read a CRM list at a predefined marker for each spatial condition.

#### 2.2.2. Procedure

The test procedure occurred within an audiometric booth to minimize extraneous noise factors. Participants, fitted with stereo headphones (Sennheiser HD 558), were given a test trial to familiarize themselves with the procedures of listening and then recording the response on the data sheet. The order of condi-

tions was permuted throughout the testing procedure.

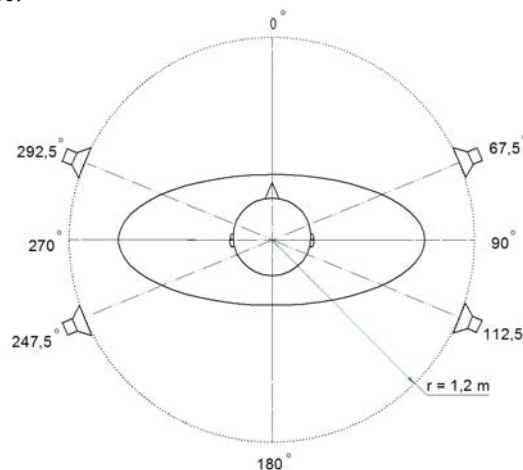


Fig. 4: Front-back condition.

The participants completed all three conditions one after the other with a 30 second break in between each condition.

Again the data for each test condition per participant were categorized into two groups; this time correct responses of each of the spatial configurations were compared to the diotic condition. The differences in the mean values of the percentage of correct responses for the CRM task comparisons were evaluated using the dependent t-test procedure at a significance level of  $p \leq 0.05$ .

### 3. Results

#### 3.1. Sound source localization results

One hundred and seventeen (117) unpaid participants with normal hearing and of ages between 23 and 52 years participated in this study. The sample consisted of 36 females, and 81 males. Data analysis on each participant's responses was performed to determine the frequency with which accurate estimates of the azimuth of the sound source location could be made. The results presented in Figure 5 show the percentage of the frequency of participant responses that were within 5 degrees (i.e. range from -5 degrees to +5 degrees) from the actual sound source location.

The results indicate that sound sources located towards the front (from 270 degrees to 90 degrees, (see Figure 1)) of the participant were identified correctly

more frequently than those sound sources presented towards the rear of the participants. There was significant statistical differences ( $p < 0.05$ ) between the means of the responses for both fast and slow signal presentations to the front compared to the rear sound source presentations. The sound sources for an inclusive 140 degrees in front of the participants (from 290 degrees to 70 degree) when compared to an inclusive angle of 180 degrees in front of the participant indicated no significant statistical difference ( $p < 0.05$ ). Similarly, there were no significant statistical differences for the identical inclusive angles to the rear of the participants. The comparison of the results of detection of sound sources coming from the front left quadrant and front right indicated no significant statistical difference ( $p < 0.05$ ) for both fast and slow localization. Similarly there was no significant statistical difference between the rear right and rear left quadrant for both fast and slow localization.

The deviation in correct responses at 190 degrees for fast localization cannot be fully explained. The deviation found in the identification of sounds coming from the rear left quadrant can be partially attributed to the fact that the participants had better hearing capacity in the right ear than the left ear. Analysis of the participant's audiometric data indicated significant statistical difference ( $p < 0.05$ ) between the left ear and right ear. The mean value for the participant's hearing loss of the left ear was higher than the loss for the right ear by 4 dB.

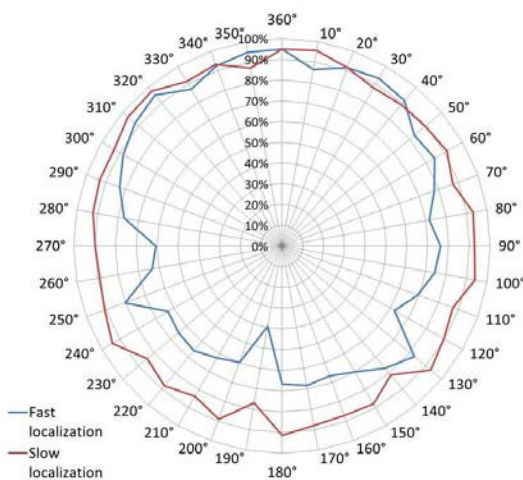


Fig 5: Accurate responses [% accurate responses].

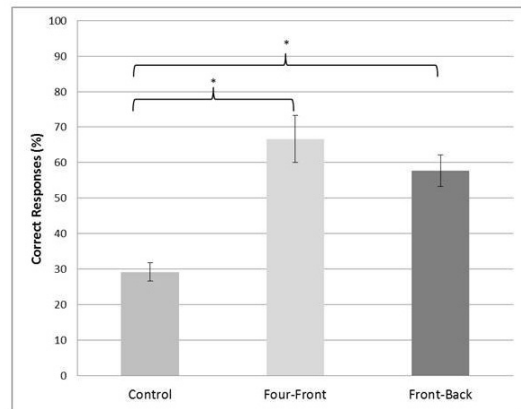


Fig. 6: Percentage of correct color and number identifications on the CRM task. (\* denotes a significant difference,  $p \leq 0.05$ )

It would therefore be expected that more frequent correct responses would be recorded for sound sources towards the right rather than the left.

### 3.2. CRM task results

The results for the CRM task are presented in Figure 6. The figure indicates the percentage of correct identifications of the call sign "Baron" and the color and number combination which followed. The current study indicated a 37.5% and 28.6% improvement being noted for the four-front and front-back speech source conditions respectively when compared to the control condition.

## 4. Discussion

### 4.1. Sound source localization

A high number of participants correctly identified targets of sound sources in the front of the participants for both fast and slow localization, indicating that more participants were able to accurately locate the sound source in front of them rather than behind them. This is in line with the published findings of other researchers [10, 15, 16]. The frequency of detecting sound sources correctly was reduced for sounds presented to the rear of the participants. Less correct responses were also prevalent for the rear left quadrant. The accuracy measure used for correct response identification was well within the accuracy capability reported for humans and contributed, at

least in part, to the high frequency of correct responses.

Humans can detect sound sources accurately to within 4 degrees for broadband sound presented directly ahead at ear level [9]. In the current study, correct responses were noted for indications within a 5 degree range to each side of the actual sound source. The high percentage of correct responses can also be attributed to the long duration of the stimulus. Other studies used sound bursts of 100 milliseconds to 500 milliseconds [11, 14, 15, 18], while the current study used 1 second and 2 second exposures. This allowed the participants to utilize localization cues from head movement in addition to the contribution of long binaural exposure to broad spectrum sound. The total number of participants utilized in the study provided confidence in the reported results, but did not explain the noticeable deviation found for fast localization in the left rear quadrant.

#### 4.2. CRM task

The results of the CRM task indicated significant improvement over the diotic presentation of the speech input. Ericson et al. [13] reported similar findings in the assessment of headphone-based multi-talker communication using the CRM speech intelligibility test. Previous research assessing spatialized speech signals in multi-talker headphone communications reported a 30% to 40% increase in performance on the CRM task when spatial auditory techniques were used [17, 21]. Similar enhancements in performance were observed for the current study.

Assessment of spatialized multi-talker communication systems reported greater improvements on the CRM task when the “talkers” were located in frontal azimuth between 0 degree and 180 degree [13] as indicated within Figure 3. The configuration presented in Figure 4 has previously been noted to result in ‘Front-Back’ confusion which occurs when the sounds are perceived at the reversed position across the frontal plane [8]. The consequence of such a phenomenon is that the listener will perceive an overlap of “talkers” at certain positions and therefore result in a decrease in the effectiveness of the spatial separation. This phenomenon might potentially explain the slight difference in percentage of correct responses between the two spatially separated auditory configurations.

The current study on auditory performance enhancement using the CRM task showed the advantages of utilizing localized sound sources to present

auditory signal and speech to listeners. The findings of this study are in agreement with the concurrence that substantial performance benefits are achieved by using spatial auditory displays [4].

Future work is identified to compare the real world performance improvement of military communication equipment using binaural auditory signals with sound source location configurations. The practical application of the laboratory results to the user environment could prove valuable, not only for the enhancement of perceptual cues, but also to the overall improvement of communication efficiency.

#### References

- [1] A.A. Scharine, T.R. Letowski and J.B. Sampson, Auditory situation awareness in urban operations, *Journal of Military and Strategic Studies*, 2009, 11 (4), 1-24.
- [2] A.M. Ahmad, K.M. Stanny and H. Fouad, Theoretical foundations for integrating sound in interactive interfaces: Identifying temporal and spatial information conveyance principles, *Theoretical Issues in Ergonomics Science*, 2009, 10(2), 161-186.
- [3] B. Shinn-Cunningham, Creating three dimensions in auditory displays, in *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality*, in the Proceedings of the Human Computer Interaction International Conference 2001, New Orleans, 5-10 August, M.J. Smith, G. Salvendy, D. Harris and R.J. Koubek, eds., N.J. Erlbaum, pp. 604-608.
- [4] D.S. Brungart and B.D. Simpson, Improving multitalker communication with advanced audio displays, in *New Directions for Improving Audio Effectiveness*, Meeting Proceedings RTO-MP-HFM-123, 2005, 30-1 – 30-18.
- [5] D.S. Brungart and B.D. Simpson, Optimizing the spatial configuration of a seven-talker speech display. Proceedings of the 2003 International Conference on Auditory Display. Boston USA, July 6-9, 2003, pp. 188-191.
- [6] D.S. Brungart and W.M. Rabinowitz, Auditory localization of nearby sources. Head-related transfer functions, *Journal of the Acoustical Society of America*, 1999, 106(3), 1465-1479.
- [7] D.S. Brungart, N.I. Durlach and W.M. Rabinowitz, Auditory localization of nearby sources. II. Localization of a broadband source, *Journal of the Acoustical Society of America*, 1999, 106(4), 1956-1968.
- [8] G. Lorho, J. Marila, and J. Hiipakka, Feasibility of multiple non-speech sounds presented using headphones, Proceedings of the 2001 International Conference on Auditory Display, Finland, July 29 – 1 August, 2001, pp. 01-32 – 01-37.
- [9] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*, London, Massachusetts Institute of Technology, 1997.
- [10] J.C. Makous and J.C. Middlebrooks, Two-dimensional sound localization by human listeners, *Journal of the Acoustical Society of America*, 1990, 87(5), 2188-2200.
- [11] L.E. Humes, S.K. Allen and F.H. Bess, Horizontal sound localization skills of unilaterally hearing-impaired children, *Audiology*, 1980, 19, 508-518.
- [12] M.A. Ericson and R.L. McKinley, The intelligibility of multiple talkers separated spatially in noise, United States

- Air Force Research Laboratory Report, Human Effectiveness Directorate Crew System Interface Division, AFRL-HE-WP-SR-2001-0009, 2001.
- [13] M.A. Ericson, D.S. Brungart, and B.D. Simpson, Factors that influence intelligibility in multi-talker speech displays, *The International Journal of Aviation Psychology*, 2004, 14(3), 313–334.
- [14] M.M. Van Wanrooij and A.J. Van Opstal, Sound Localization under perturbed binaural hearing, *Journal of Neurophysiology*, 2007, 97, 715–726.
- [15] N. Merat and J.A. Groeger, Working memory and auditory localization: Demand for central resources impairs performance, *The Quarterly Journal of Experimental Psychology*, 2003, 56A(3), 531–549.
- [16] R. Martin, K. McAnally, T. Watt and P. Flanagan, The effect of spectral variation on sound localization, *Defence Science and Technology Organisation, DSTO-RR-0308*, 2006.
- [17] R.S. Bolia, W.T. Nelson, M.A. Ericson and B.D. Simpson, A speech corpus for multi-talker communications research. *Journal of Acoustical Society of America*, 2000, 107(2),: 1056–1066.
- [18] S.M. Abel and C. Tikuisis, Sound localization with monocular vision, *Applied Acoustics*, 2005, 66, 923–944.
- [19] S.P.A. Parker, G. Eberle, R.L. Martin and K.I. McAnally, Construction of 3-D audio systems: Background, research and general requirements, *Defence Science and Technology Organisation, DSTO-TR-2184*, 2008.
- [20] T. Moore, Voice communication jamming research, *Advisory Group for Aerospace Research and Development (AGARD). Conference Proceedings No. 311, Annual Communication in Aviation, CP311, 2-1-2-6*, 1981.
- [21] W.T. Nelson, R.S. Bolia, M.A. Ericson, and R.L. McKinley, Monitoring the simultaneous presentation of specialized speech signals in a virtual acoustic environment, *Proceedings of the 1998 IMAGE Conference*, 1998, pp. 159–166.