



Convention Paper

Presented at the 152nd Convention
2022 May, In-Person and Online

This paper was peer-reviewed as a complete manuscript for presentation at this Convention. This paper is available in the AES E-Library, <http://www.aes.org/e-lib>. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Effect of spatial audio processing on attention redirection and speech intelligibility in multi-talker communication

Eric Grownney¹, Agnieszka Roginska²

¹ Otto Engineering Inc, 2 East Main St, Carpentersville IL 60110

² New York University, 35 West 4th St. New York NY USA

Correspondence should be addressed to Eric Grownney (Eric.Grownney@ottoexcellence.com) and Agnieszka Roginska (roginska@nyu.edu)

ABSTRACT

Most current communication systems for mission critical applications use devices with relatively simple audio presentation to the user, employing single channel communication. An increasing number of applications are requiring users to monitor and listen to multiple talkers concurrently. This paper describes results of an experiment investigating the differences in performance in a transcription task when listeners were listening to multi-talker communication using two different Head-Related Transfer Functions (HRTF) data sets in comparison to monaural listening. The goal of this research is two-fold: 1) to assess the impact on attention redirection and the accuracy of the NATO phonetic word transcription based on the pre-selected HRTFs used, as compared to a diotic presentation, as the number of talkers increases; 2) to investigate whether there is a difference in performance between younger and older listeners. Results indicate a significant improvement when using spatial audio, in the accuracy of the transcription and number of missed characters, especially for participants in the older age group.

1 Introduction

Maintaining situational awareness can be a life-saving ability in mission-critical application. Often, soldiers are engaged in multi-talker communications scenarios where they are forced to devote cognitive resources to understanding critical messages. Spatial separation of audio can reduce cognitive load and help soldiers maintain situational awareness. Past research has indicated benefits of using 3D audio for speech communication with improved intelligibility and recognition [1][2][6][7]. The goal of this research is to quantify the impact of different Head-Related Transfer Function (HRTF) datasets on listener performance in a task involving attention redirection to a set target phrase followed by NATO phonetic

words (e.g. “alpha”, “bravo”, etc) which were to be transcribed. This research has two primary goals. First, to assess the impact on attention redirection and the accuracy of the NATO phonetic word transcription based on the two different HRTF datasets used and compare this to the baseline performance of the diotic presentation. The focus of this research is on the accuracy of the number of characters correctly transcribed and the number of missed characters, as the number of talkers increases from one to four. Second, to investigate whether there is a difference in performance between younger and older listeners.

2 Background

2.1 3D audio

In the natural listening environment, the human auditory system uses these cues to localize and externalize sound sources. When listening in the free field, we are able to discern what the sound source is, where it is coming from and, when multiple sound sources are presented concurrently, we are able to attend to a specific source through the benefits of the “cocktail party effect”. The “cocktail party effect” was first described by Cherry [3] as the ability of the brain to focus on a specific auditory stimulus in the presence of other competing stimuli, as when a person is in a noisy room at a cocktail party. Stream segregation occurs because of several factors; however, the strongest contributor to stream segregation is the distinct spatial location of sound sources, and binaural audio. The cocktail party effect has been studied extensively and evaluated in multi-talker situations [4].

The localization accuracy of a sound source and externalization highly depend on the HRTFs used to render the signals. These directionally dependent filters are unique to an individual and best results are obtained with individually-measured HRTFs [12]. However, individual measurements can be impractical, time-consuming, and expensive to measure. Alternatively, HRTF datasets measured on another individual or a dummy head can be used. But the perceptual results of the quality of the spatial presentation can vary from person to person and finding a good non-individualized HRTF dataset that would accommodate most listeners is challenging. Much work has been done to investigate ways to find a good non-individualized HRTF dataset to enhance performance. In order to do so, the application must be considered to determine what factors are important in evaluating a good performing dataset. For speech and communication, one of the most critical factors are speech intelligibility, in addition to localization and localization accuracy.

2.2 Spatial audio in speech communication

Traditional radio communication presents the listener with a diotic signal, where both ears are presented with the same signal. Diotic listening leads to the

perception of a sound inside the head and, when multiple talkers are presented concurrently, can make it very difficult to attend to any one of the talkers. Research has shown that informational masking occurs and intelligibility drops significantly when speech signals are presented concurrently with speech interferers [10]. In military communications, the inability to segregate multiple streams of speech can result in missing critical information, and potential loss of life.

Prior research has shown that separating competing talkers at different locations by using binaural audio provides many benefits to the listener compared to monaural or diotic presentation. Such benefits include improved talker recognition, speech intelligibility (especially in the presence of noise), and soldiers in a subjective experiment indicated reduced effort when listening to multiple talkers [5][6][11].

Speech source separation for improved recognition, intelligibility and reduced fatigue has been the subject of research for mission-critical applications for a number of years. In 2001, Brungart and Simpson [2] observed a 25-35% improvement in overall performance during a multi-talker task with same-sex subjects in a warfighter-type situation. The biggest advantage, they reported, was in better intelligibility. Abouchacra et al. [1], describe the binaural helmet as equipped with a binaural and active noise reduction (ANR) system that helps with localizing multiple sound sources in noisy environments. Results indicate that this system aided soldiers with completing missions quicker, and they were more efficient at completing the tasks [1]. Spatial audio has also been investigated and has shown improved performance in aviation with regards to improved situational awareness, target acquisition, threat avoidance, and cockpit communications [7]. In 2019, it was announced that the US Air Force would be working to incorporate 3D audio into cockpits because 3D audio “helps reduce the pilot's workload, improve situational awareness and survivability; reduce hearing loss and fatigue; and improve speech intelligibility”[7]. Fighter planes are also looking to 3D audio for “increased situational awareness and reduced workload and fatigue”.

3 Methodology

A subjective experiment was conducted where subjects were asked to listen to audio samples that contained from one to four concurrent talkers, and perform a transcription task for each sample. The goal was to examine the impact of spatial audio presentation on transcription accuracy, especially in the presence of other interferers.

3.1 HRTF Pre-Selection

In a pilot study, 21 participants were presented with multiple talker combinations of the same spatial placement as the study described in this paper. Participants were asked to evaluate speech signals processed using 7 different binaural presentations; 1 diotic and 6 non-individualized HRTF datasets taken from CIPIC, LISTEN, and SADIE II databases. Subjects evaluated speech stimuli based on 5 criteria: Attention redirection (ability to select one speaker from many), Intelligibility (understand one talker in the presence of many), Externalization (audio sound inside or outside of the head), Spatial Distinction (how specific is a given talker's perceived direction), and Sound Quality. Results showed that there were two HRTF datasets that most participants preferred in all multiple channel combinations. These two most preferred sets of filters were selected and used in the current study.

3.2 Participants

A total of 35 participants were subjects in this study, 30 males and 5 females, ranging in age from 20 to 64 years old. All had self-reported normal hearing. The subjects were untrained listeners with no prior experience in audio evaluation experiments. Participants were excluded from the study if they self-reported dyslexia or related reading/writing challenges or were unable to complete the test fully.

Given the findings of previous studies like Pearson, et al. [13], analysis was performed by dividing the subjects into two age groups to determine whether there was any impact of age, and age-related hearing loss, on results. Age group 1 contained 15 subjects below the age of 40. Age Group 2 consisted of 20 subjects 40-64 years old. There was an imbalanced gender representation in the available participants

that did not permit further meaningful comparisons between males and females in this study.

3.3 Apparatus

Subjects were seated in a quiet office-type environment while testing was conducted. Spatial audio was presented to subjects over closed back headphones (Anker SoundcoreVortex, wired mode). A laptop computer running a custom developed MATLAB script was used to spatially process audio stored in WAV files (see Stimuli section) and play the audio through the internal sound card (HP ZBook 14u G6 with Intel® SST) with all Windows audio enhancements deactivated. For most mission critical radio applications, this represents the high end of available audio quality.

The MATLAB script provided the prompts for the subject to proceed from trial to trial at their own pace. For each trial, the subject listened for target alpha-numeric digits (see Stimuli section) and wrote them down on a structured worksheet to simulate a real-world transcription task. For the convenience of data collection, the subject was then asked to type each transcription answer into a text field provided by the application to capture their response. All responses were reviewed with the subject after the test was complete to assure that errors unrelated to the spatial listening task did not skew results (e.g. typing "o" for an answer when the written answer showed they had correctly captured "0" – zero, etc.).

3.4 Stimuli

The audio presented to the subject in this experiment ranged from 1 to 4 simultaneous audio samples. The audio samples consisted of one target phrase and zero to three Background Interference (BI) phrases. target phrases were composed of a preamble portion of 3-5 words and then a list of 7-10 alpha-numeric digits presented as NATO Phonetic words. For example, "Grid Location for Target: Bravo, Zero, Eight, Six, India, Niner, Four, Tree, Tango, Seven".

BI phrases were created specifically not to contain NATO Phonetic words but were instead phrases and words commonly used in military radio communications. BI phrases were constructed to be as long or longer than target phrases in duration,

usually ~10 seconds. For example: “Negative, did not copy, not sure which structure you are referring to, please repeat”.

In addition to target phrases, training phrases were created having a similar structure of preamble and 7-10 alpha-numeric digits spoken in NATO Phonetic format.

Source material for all phrases was scripted and recorded from 4 male talkers (native English speakers) in a sound isolation room at 48 kHz. Prior to use, all captured samples were de-noised, resampled to 16 kHz, bandlimited to 300 Hz - 3600 Hz, and level normalized to -20 LUFS as measured by MATLAB in accordance with EBU R 128 and ITU-R BS.1770-4 standards. Subjects were given additional control of the playback audio level (see Figure 1). The bandwidth restriction is representative of military and first-responder radio communication channels, the area of interest for this study. From these samples, 16 unique training phrases, 24 unique target phrases and 40 BI phrases were created and stored as WAV files.

The spatial audio processing for this experiment was done in a MATLAB script that conducted the evaluation (see Procedure section). As described in the HRTF Pre-Selection section, two HRTFs were selected for use. From each HRTF database, the right and left channel filter pairs for four azimuth positions on the horizontal plane symmetrically distributed along the inter-aural axis (left/right) were selected. The minimum angular separation was 60 degrees. The HRTFs were resampled at 16kHz. No headphone equalization was applied.

3.5 Procedure

Prior to testing, subjects were trained on recognizing and transcribing NATO Phonetic words. Using a developed MATLAB app, training was conducted by diotically presenting to the subject training phrase audio samples over headphones, one at a time. The preamble portion of each phrase was shown in a text field of the MATLAB app prior to each training trial. Each subject was able to play the audio sample and was asked to transcribe the alpha-numeric digits portion of the phrase. The app tracked the accuracy of the result against the target string and reported back

to the subject whether or not the answer was correct. The subject needed to have 3 consecutive answers that were 100% correct in order to proceed to the main test.

Once training was complete, the Task Test itself was conducted. The test was run by another MATLAB app developed for presenting audio over headphones and capturing results. The app showed the preamble of the target phrase in a text box for which the subject needed to listen. The next portion of that phrase was the alpha-numeric digits that the subject needed to transcribe. The 24 unique target phrases were randomly presented to the subject for each trial. There was no repetition of the target phrases over the course of the test.

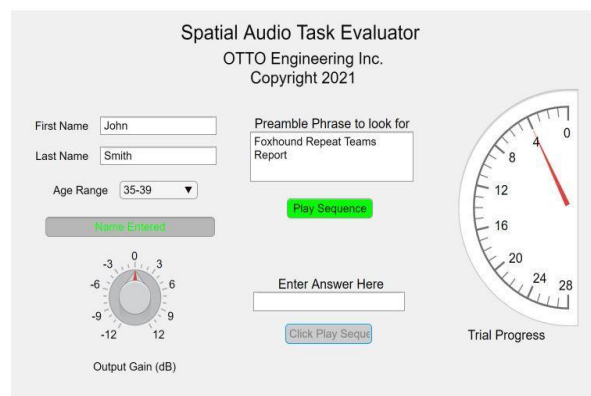


Figure 1 Application used by subjects for the procedure of the experiment, and to record the responses.

For each trial, a random number of BI phrases, from 0 to 3, were also presented at the same time. The BI phrases were randomly assigned for each trial from the available pool of BI phrases such that each target and BI phrase came from each of the 4 male talkers (no talker was ever used more than once within the same trial). The target and BI phrases had a unique and randomly assigned start delay between 0 and 300 ms.

Each trial randomly assigned the direction from which the target phrase and the 0-3 BI phrases were presented. Also randomized was which of three HRTF conditions would be used for the spatial audio treatment – Diotic (none), HRTF1, or HRTF2. Each

permutation of HRTF and 0-3 BI phrases was repeated twice for a total of 24 trials.

To help participants acclimate to the task, 4 dummy trials were presented at the beginning of the experiment, which were not counted toward the results.

4 Results and Analysis

Analysis of the results of the 35 subjects was performed to determine the trends of the performance of stimuli presented diotically, with HRTF 1, and with HRTF 2. The results reported in this paper will focus on general trends observed. A repeated-measures ANOVA was performed using SPSS v.28.0 to determine statistically significant results, impact of the spatial processing conditions, the number of concurrent talkers, and the difference in the results as a function of the age group of subjects.

In both the training and the actual task test portions of this study, “correctness” of an answer was based on the accuracy of each NATO phonetic character as follows:

Correct: characters accurately captured by the subject in the correct sequence.

Incorrect: wrong characters that neighbored correct characters as defined above.

Missed: absent characters in one of three ways: (1) subject unknowingly skipped characters in a captured sequence, (2) subject knowingly skipped one or more character(s) for which they did not/could not guess, or (3) the subject recorded a blank answer. In all cases, each of the intended characters absent in an answer counted towards a missed character total.

4.1 Correct characters

Results of the experiment showing the average percentage of correctly transcribed characters in the NATO alpha-numeric strings are shown in Table 1 for all subjects, in Table 2 for subjects in Age Group 1, and in Table 3 for subjects in Age Group 2. Results are presented graphically in Figure 2 for all subjects (top), Age Group 1 (middle) and Age Group 2 (bottom). The results show the average for all subjects, and their performance as a function of the spatial audio processing used (diotic, HRTF 1, or

HRTF 2), and the number of talkers that were presented concurrently.

# Talkers	<i>Diotic</i>	<i>HRTF 1</i>	<i>HRTF 2</i>
1	99%	100%	100%
2	90%	97%	96%
3	41%	81%	83%
4	21%	68%	65%

Table 1 Average percentage of correctly transcribed characters for all subjects

# Talkers	<i>Diotic</i>	<i>HRTF 1</i>	<i>HRTF 2</i>
1	100%	99%	100%
2	94%	98%	98%
3	45%	86%	94%
4	28%	83%	74%

Table 2 Average percentage of correctly transcribed characters for Age Group 1

# Talkers	<i>Diotic</i>	<i>HRTF 1</i>	<i>HRTF 2</i>
1	99%	100%	100%
2	87%	97%	95%
3	37%	77%	78%
4	17%	56%	59%

Table 3 Average percentage of correctly transcribed characters for Age Group 2

As can be seen in Table 1, for the one (1) talker condition, subjects’ identification of characters is nearly perfect for the diotic (no spatial audio), HRTF 1 and HRTF 2 conditions, for all subjects. When the number of talkers increases to two, the performance drops for all subjects, with a 90% accuracy for diotic, 97% for HRTF 1, and 96% for HRTF 2. With three talkers, the performance drops even further. However, we see a much more significant decrease of accuracy in the diotic condition than either of the HRTFs – 41% for diotic, 81% and 83% for HRTF 1 and HRTF 2 respectively. In the 4-talker condition, HRTF 1 and HRTF 2 showed over 3 times the accuracy of the diotic condition (21% diotic, 68% HRTF 1, 65% HRTF 2). A repeated-measures statistical analysis shows statistically significant

results ($p < 0.001$) of whether spatial audio was off (Diotic condition) or on (HRTF & HRTF 2). Results also show a statistically significant effect of the number of talkers ($p < 0.001$), and interaction between the spatial processing and the number of talkers ($p < 0.001$). There was no statistical significance between HRTF 1 and HRTF 2.

There is a statistically significant effect ($p < 0.03$) of the age group on accuracy results, and an interaction between the number of talkers and age group ($p < 0.01$). Results for Age Group 1 (under 40 years old) show that the one and two-talker conditions are nearly perfect for all three spatial audio conditions. For the 3-talker condition there is a significant decrease in accuracy in the diotic condition (45%), but the HRTF1 and HRTF2 conditions show a much lower decrease (86% for HRTF1, and 94% for HRTF2). Similarly, in the 4-talker condition, the diotic presentation drops to 28% accuracy, while HRTF1 maintains an accuracy of 83% and HRTF2 drops to 74%.

Age Group 2 (40-64 years old) results for accuracy in the one- and two-talker conditions show a relatively good performance. It is in the three- and four-talker condition that we see interesting results. Here, there is a much larger drop in accuracy in the diotic condition than we saw in the younger age group. In the 4-talker condition, Age Group 2 sees a significant benefit of HRTF1 (56% accuracy) or HRTF2 (59% accuracy), in comparison to the 17% accuracy in the diotic condition.

4.2 Missed Characters

The results of the number of missed characters are shown in Table 4 for all subjects, in Table 5 for Age Group 1, and in Table 6 for Age Group 2. Results are presented graphically in Figure 3 for all subjects (top), Age Group 1 (middle) and Age Group 2 (bottom). The number of missed characters in the one-talker case is less than 1% in all three spatial audio conditions (0% for HRTF 2). When the number of talkers increases to two, the number of missed characters increases to 8% for the diotic condition, and increases to 2% for both HRTFs. With 3 talkers, the number of missed characters increases to 52% for diotic, but we only see an increase in missed characters to 17% (HRTF 1) and 15% (HRTF 2). In

the 4-talker condition, the number of missed characters is more than double for the diotic presentation (72%) than HRTF 1 (28%) or HRTF 2 (30%).

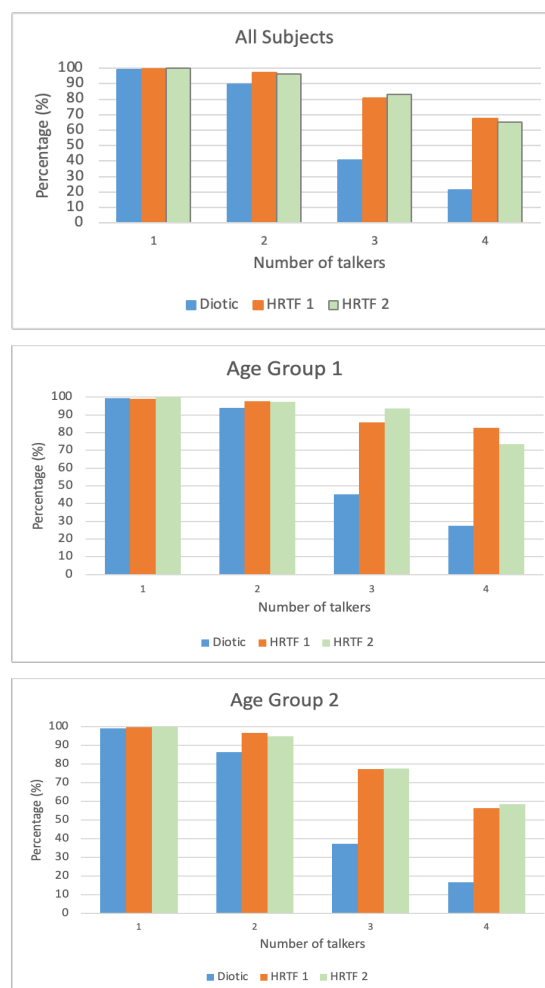


Figure 2 Average percentage of correctly identified characters for the diotic presentation (blue), HRTF 1 (orange), and HRTF 2 (green) stimuli. Results are presented for all subjects.

The number of missed characters in Age Group 1 for the one- and two-talker conditions is very low. We see a large increase of missed characters in the diotic

# Talkers	<i>Diotic</i>	<i>HRTF 1</i>	<i>HRTF 2</i>
1	0.5%	0.5%	0%
2	8%	2%	2%
3	52%	17%	15%
4	72%	28%	30%

Table 4 Average percentage of missed characters for all subjects

# Talkers	<i>Diotic</i>	<i>HRTF 1</i>	<i>HRTF 2</i>
1	0%	1%	0%
2	5%	2%	2%
3	51%	12%	4%
4	66%	13%	8%

Table 5 Average percentage of missed characters for Age Group 1

# Talkers	<i>Diotic</i>	<i>HRTF 1</i>	<i>HRTF 2</i>
1	1%	0%	0%
2	10%	3%	3%
3	53%	21%	22%
4	77%	39%	37%

Table 6 Average percentage of missed characters for Age Group 2

condition for 3 talkers (51%), in comparison to 12% and 4% missed in the HRTF1 and HRTF2, respectively. In the 4-talker condition the diotic presentation results in 66% missed characters, but only 13% and 8% missed for HRTF1 and HRTF 2.

The results of Age Group 2 show a worse performance of missed characters. The number of missed characters for the one-talker condition is very low. In the 2-talker case we see an increase to 10% in the diotic condition, and only to 3% in both HRTFs. In the 3-talker condition there is a sharp increase to 53% missed characters in the diotic case, both HRTFs are at 21-22%. In the 4-talker condition the diotic presentation results in 77% missed characters, and 39% and 37% missed for HRTF1 and HRTF 2, respectively.

The statistical analysis of this study shows that the correct characters and missed characters variables are highly correlated. The statistically significant effects for the missed characters are the same as for the correct characters, as presented above.

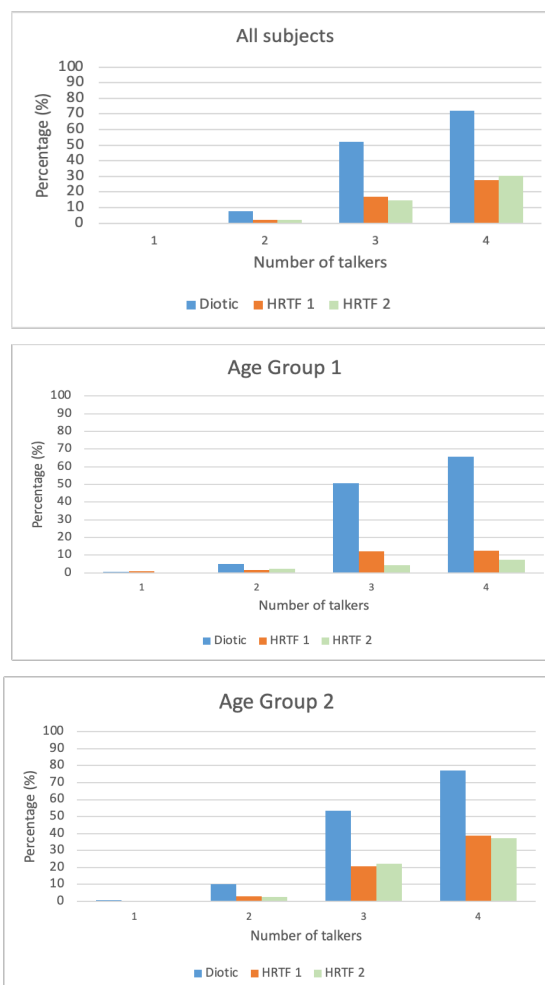


Figure 3 Average percentage of missed characters for the diotic (blue), HRTF 1 (orange), and HRTF 2 (green) stimuli. Results are presented as an average of subjects.

4.3 Results trends

To better see the trends and effects of the spatial audio processing used and impact of the age group, Figure 4 contains the boxplots of the correct characters and missed characters separated by age groups and spatial audio filters. In this figure, each row represents a spatial audio condition (diotic - top row; HRTF 1 - middle row; and HRTF 2 - bottom row), and each column represents subjects separated by ages (all

subjects – left column; Age Group 1 – middle column; Age Group 2 – right column).

The effect of the spatial audio filter can be seen on a couple of levels. First, we see that there is no effect of the spatial audio processing on performance in the 1-talker condition. In the 2-talker condition, we see the performance get worse for the diotic condition, and it is a little worse for Age Group 2. The 3-talker condition sees a significant difference in performance. The diotic presentation shows a significant drop in performance, while HRTF 1 and 2 maintains good results. This is even more pronounced in the 4-talker case. Second, the effect of the age

group can be seen. In the diotic presentation, the age group shows a difference in performance starting with the 2-talker condition, and the gap in performance gets worse in the 3- and 4-talker case. Meanwhile, the HRTF 1 and HRTF 2 results show that performance is very similar for 1- and 2- talkers, regardless of age group. In the 3-talker scenario, we see there is a noticeable difference between the two age groups. Age Group 1 still maintains very good performance in the 3-talker condition for both HRTFs, while Age Group 2 gets worse. In the 4-talker case, we again see the much better results for Age Group 1 than Age Group 2, for both HRTFs.

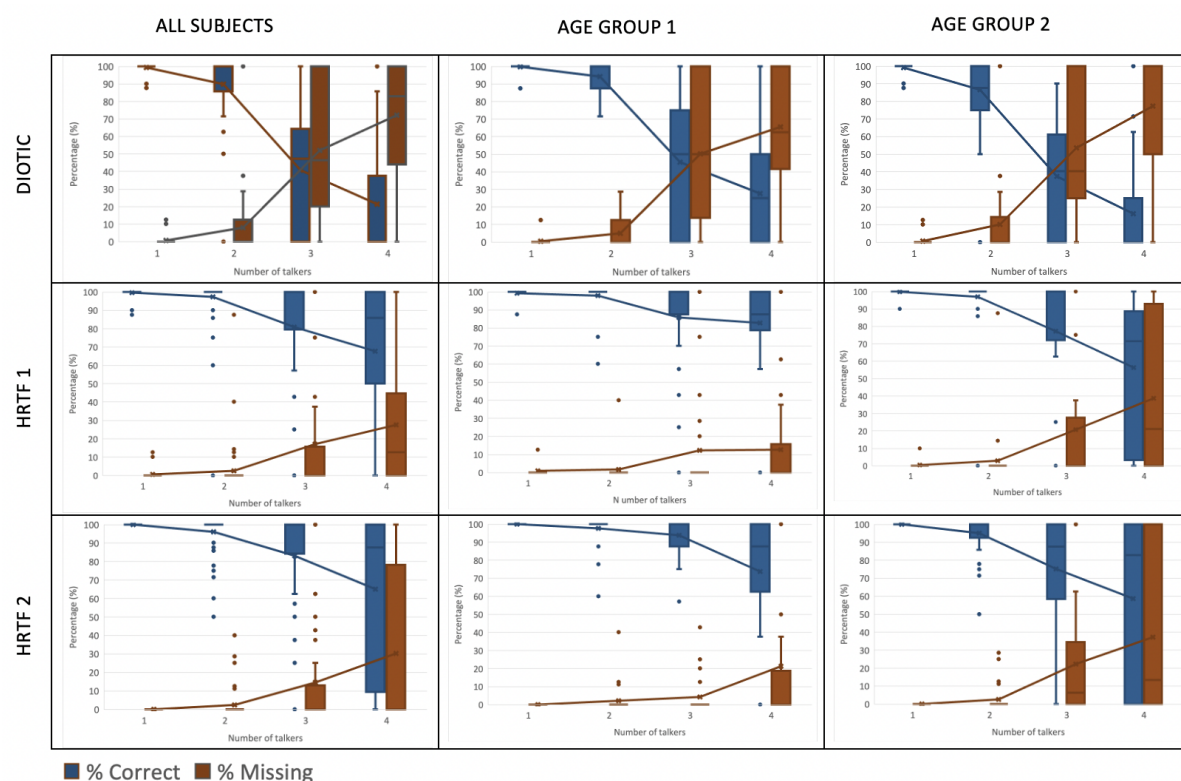


Figure 4. Boxplots of the correct characters (blue) and missed characters (orange) for each spatial audio condition – diotic (top row), HRTF 1 (middle row) and HRTF 2 (bottom row). Results are presented for each subject group: all subjects (left column), Age Group 1 (middle column), and Age Group 2 (right column). Lines represent the mean.

5 Discussion and Conclusions

The goal of the reported study was to determine the effect of spatial audio processing in a task where listeners are asked to transcribe NATO phonetic words during a multi-talker scenario, with up to four talkers. Results show a statistically significant effect of the spatial audio processing, the number of talkers, and an effect of age of the participants.

Overall, processing concurrent talkers for 3D audio presentation by using HRTFs increases the accuracy of correctly transcribed characters and reduces the number of missed characters as the number of talkers increases. These results are in line with prior studies done with speech processed using non-individualized HRTFs [12]. In the study presented here, results are similar for both HRTFs used, in comparison to the diotic presentation. No statistically significant difference was found between the two HRTFs.

There are no benefits of using spatial audio when a single talker is present. There is also no detriment in using spatial audio with a single talker. In the two-talker case there is a mild benefit of HRTF processing. But the true benefit of 3D audio begins when three or more talkers are presented concurrently. This is the point where speech becomes mostly unintelligible in the diotic presentation while, with 3D audio, performance is much better. It is important to note that we didn't see any evidence that would suggest that HRTF processing has any negative impact on intelligibility, for any number of talkers.

Lastly, there appears to be a pronounced effect of age on results, and the benefits of 3D audio are different for Age Group 1 (younger group) and Age Group 2 (older group). The younger group didn't see any benefit of spatial audio with two talkers. It is with three talkers that the benefits are significant, where for the 3- and 4-talker conditions accuracy remains 74-85%, in comparison to 28% with no spatial audio. For the older age group, the benefits start with the 2-talker scenario, and are significant in the 3- and 4-talker conditions. In this age group, not having 3D audio is detrimental to performance.

In conclusion, the benefits of 3D audio in a multi-talker communication scenario to understand and transcribe NATO phonetic characters include more

correctly identified characters, fewer missed characters, fewer completely missed character strings, and fewer incorrectly heard and/or transcribed characters. We acknowledge the limitation of this study to be focused on male talkers. Future work will extend the range of talkers to female and male, and will include a more balanced male/female representation in the participant pool.

References

- [1] Abouchacra, K. S., Breitenbach, J., Mermagen, T., & Letowski, T. (2001). Binaural Helmet: Improving Speech Recognition in Noise with Spatialized Sound. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(4), 584-594.
- [2] Brungart, D.S. , & Simpson, B.D. (2005). Improving multitalker speech communication with advanced audio displays. In *New Directions for Improving Audio Effectiveness* (30.1 - 30.18). Meeting Proceedings RTO-MP-HFM-123, Paper 30 . Neuillysur-Seine, France, RTO.
- [3] Cherry, E. C. (1953). "Some Experiments on the Recognition of Speech, with One and with Two Ears" . *The Journal of the Acoustical Society of America*. 25 (5): 975–79.
- [4] Crispin, K., & Ehrenberg, T. (1995). Evaluation of the cocktail-party effect for multiple speech stimuli within a spatial auditory display. *Journal of The Audio Engineering Society*, 43, 932-941.
- [5] Doll, T. (1995). Spatial and spectral release from masking in three dimensional auditory displays. *Human Factors*, 37, 341–355.
- [6] Ericson, M. A., & McKinley, R. L. (1997). The intelligibility of multiple talkers separated spatially in noise. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 701–724). Mahwah, NJ: Lawrence Erlbaum.
- [7] Keller, J. (2019, October 24) "Air Force chooses Terma to provide cockpit 3-D audio for pilot situational awareness on A-10 combat aircraft", *Military & Aerospace Electronics*, Retrieved on July 17 2021

<https://www.militaryaerospace.com/communications/article/14069281/3d-audio-combat-aircraft-situational-awareness>

- [8] Simpson, B., Brungart, D., Gilkey, R., and McKinley, R. (2012). Spatial Audio Displays for Improving Safety and Enhancing Situation Awareness in General Aviation Environments.
- [9] Song, H., & Zhang, S. (2020). Perceptual Characteristics of Chinese Speech Intelligibility in Noise Environment. *Scientific Programming*, 2020, 1–10.
- [10] Summers, J.S., Roberts, B. (2020) Informational masking of speech by acoustically similar intelligible and unintelligible interferers, *Journal of the Acoustical Society of America* 147(113).
- [11] Vause, N., Abouchakra, K., Letowski, T., Resta, E. (2001). Spatial (3-D) audio displays improve speech communication in the command and control vehicle (C2V). *Proceedings of the 22nd Army Science Conference, Baltimore (MD): OASA*, pp. 834-841.
- [12] Wenzel, E. M., Arruda, M., Kistler, D. J. , Wightman, F. L. (1993). "Localization using non-individualized head-related transfer functions", *The Journal of the Acoustical Society of America*. 94 (1): 111-123.
- [13] Pearson JD, Morrell CH, Gordon-Salant S, Brant LJ, Metter EJ, Klein LL, Fozard JL. (1995) Gender differences in a longitudinal study of age-associated hearing loss. *The Journal of the Acoustical Society of America*. 97(2): 1196-1205.