REVIEW ARTICLE

# DUAL AUDIO: EFFECTS OF VOLUME RANGE ON DICHOTIC LISTENING VIA SPATIALLY SEPARATED SPEAKERS.

Steven J. KERR[1], Carmen SIAU[1]

[1]*Motorola Solutions, Human Factors Research, Penang, Malaysia.*

**ABSTRACT**

Professionals who utilize communication devices may have to listen and monitor multiple audio streams whether from different Land Mobile Radio (LMR) or public broadband (LTE) channels whilst carrying out their day to day tasks. Understanding how humans cope with multiple audio streams is therefore important to help engineers design solutions that will allow professionals to both hear and understand vital information from numerous sources. This paper describes a usability study of a spatialized dual speaker system (one speaker on each shoulder) whereby 16 participants from age 23 to 54 had to listen to 2 audio streams played simultaneously, a primary audio stream on one speaker to be concentrated on to understand content and a secondary stream played on the other speaker to be scanned for keywords. Participants were asked true or false questions on the primary stream to gauge level of information understood whilst being scored on their ability to recognize when selected keywords were spoken in the secondary stream. These audio streams were played at different volume ranges and deltas to understand if differences in volumes across a range of volumes can assist users in this task. The results has shown that there does not appear to be a specific volume range or delta that assists users in listening to content from 2 spatialized audio sources and that individual capabilities are more likely to be an important factor. Devices that have capability of broadcasting multiple channels therefore need the ability for each channel's volume to be controlled individually and not be auto changed, as each user will have their own personal preferences.

***Keywords**: Dichotic Listening, Multi-Stream Audio, Divided Attention*

## INTRODUCTION

Public safety and other professionals utilizing communication devices are having to deal with ever more streams of audible information as technology allows for both Land Mobile Radio (LMR) and LTE communications to be handled at the same time. Adding to the mix of the audio from other sources such as virtual assistants and professionals could find themselves overloaded with audio. Understanding how humans deal with multiple audio streams and how it affects cognitive load and attention is therefore imperative to help engineers in designing solutions that will allow professionals to both hear and understand vital information from numerous sources.

Previous research carried out on monitoring multiple auditory channels has involved the use of spatial audio to improve speech intelligibility and reduce workload for operators listening to multiple radios or intercom channels (Simpson et al, 2004). Research has also looked at Signal to Noise Ratios (SNR: the ratio of levels of target and interfering sounds) to understand how people can listen and obtain information from an audio stream by effectively blocking out content from other audio sources. However, those other audio sources may also be of interest and people will want where possible to listen to more than one channel effectively.

This paper describes a usability study of a spatialized audio system whereby a user can listen to 2 audio streams simultaneously. Different volume levels and SNRs were used to understand whether these have any affect on a person's ability to listen to 2 audio streams at once with the aim of understanding whether auto adjustments to incoming communication would help professionals obtain information and carry out their tasks more effectively.

## BACKGROUND

### Speech Intelligibility / SNR

Hearing is a passive process, merely detecting sounds around us, listening is an active process, it requires a conscious desire to determine the meaning of what we hear. Good hearing is the foundation of good listening. Good listeners push irrelevant stimuli to the background (Listenwell, 2017).

In many cases when there are competing audio streams, the intent is for the user to concentrate on only one (usually an audio information source) as the other competing audio is just noise. In those cases, the main audio information cannot just be heard over the noise, it has to be understood. *Intelligibility* is a measureable aspect of electronic voice transmission systems that indicates the degree that human listeners will be able to understand the voice messages transmitted through them (Moran, 2013). Measurement is usually expressed as a percentage of a message that is understood correctly (Moran, 2013). Related to intelligibility is Signal-to-Noise Ratio (SNR). 'SNR compares the sound level output from the speaker to ambient noise in the room. It is the obscuring of voice due to background noise, the higher the ratio, the greater the intelligibility. Humans can tolerate a significant amount of background noise. However, once intelligibility begins to diminish, it diminishes rapidly' (Moran, 2013). Elements such as pitch and harmonic structure in a talker's voice can help separate speech from general background noise, but if the competing audio source is also speech then it will be more difficult to separate (Aiken, 2004).

Individuals with normal hearing can usually still understand speech with an SNR of -6 to -10 dB (Agnew, 1999), though to have better success rates at understanding information it is better to have higher SNR. Most prior research on speech intelligibility metrics was carried out due to military communication requirements and emergency mass notification systems. Normal calculations for alarm professionals for audibility is to maintain 15 dBA over ambient (Moran, 2013). Higher than +15 dB results in reduced returns in terms of improving intelligibility, though increasing the number of speakers can usually help intelligibility better than just increasing speaker wattage which can lead to distortion (Agnew, 1999). Other research looking at optimizing a person's ability to focus on one audio stream over another have included Brungart and Simpson (2005), who in a study with an interfering talker found that a +12dB SNR was enough for nearly perfect user performance whilst Schmandt and Mullins (1995) created a spatialized audio interface that utilized a +10dB SNR on the primary audio source as the user turned their head towards it. Therefore it has been shown that a +10 to +15 dB SNR range can improve intelligibility when concentrating on one audio source over another. However there are circumstances where a person may be trying to actively listen to two or more audio sources at once, e.g. monitoring separate audio channels.

### Multistream Listening

Auditory scene analysis is where listeners use information perceived by their auditory system to form a mental representation of the world around them (Schmandt and Mullins, 1995). A number of acoustic cues allows the separation of a mix of sounds into distinct sources: these include location, harmonics and frequency, continuity, volume, and correlation to visual events (Schmandt and Mullins, 1995).

Schuett and Walker (2013) found that if information was presented slowly then participants could attend to 5 or 6 streams, but at faster rates then they could only manage 2 or 3. When listening to faster data streams then people could attend to 2 sources. This would indicate that people could cope with listening to more than one audio stream, but how much information can they understand and obtain from those different sources?

**The Cocktail Party Effect**

Understanding speech in multiple-talker situations and the difficulties associated with this is often referred to as the “cocktail-party problem” (or “cocktail party effect”). The term has been used in various contexts and sometimes refers to different conditions, with initial focus being the ability of listeners to select target speech while ignoring other sounds in conditions where signals were either mixed or presented to separate ears. An important finding for “cocktail-party” conditions is that the auditory system benefits considerably from the fact that we have two ears. “The head provides an acoustic shadow, which can favor one ear, depending on the location of the talkers. In addition, the differences between the signals entering the two ears enable us to partially ‘unmask’ interfering sounds, effectively providing an increase of the signal to noise ratio of up to 4 dB” (Bronkhorst, 2015).

Vasquez-Alvarez & Brewster (2010) has stated that the Cocktail Party effect provides evidence that humans can, in fact, monitor several audio streams simultaneously, selectively focusing attention on any one and placing the rest in the background. A spatial representation of the auditory display provides orientational information that aids segregation and attention switching between the audio streams to maintain intelligibility when auditory information is being used.

**Divided Attention and Spatial Hearing**

Multiple Resource Theory (MRT) is a theory of multiple task performance in high work-load environments (Wickens, 2002). Closely tied is the Theory of Working Memory which maintains that short term working memory consist of multiple independent processors e.g. one for visual and one for auditory information (Oviatt, 2006). Therefore, an important element of MRT relates to whether there is conflict of resources in performing cognitively related tasks (Wickens, 2002). Listening to two audio streams at once clearly provides a conflict and people’s attention is likely to be switching between them.

Binaural processing of audible cues has been shown to be a significant factor in increasing intelligibility. People with normal hearing can usually take advantage of the difference in spatial location that often occurs between a desired speaker and undesired noise. If spatial separation is present, the normal functioning of the central auditory system allows suppression of undesired background noise and the ability to concentrate on desired speech. Binaural listening has been shown to offer people the ability for greater distinction between voices due to the separation of the sounds in space and can help improve SNR by at least 2 to 3 dB. Thus, one simple way of improving speech intelligibility is to use a binaural fitting in appropriate situations (Agnew, 1999).

When information is presented dichotically, selective attention tasks involving identification of one of two sound sources improves by 2 to 5 dB (Yost et al, 1996). Humans can leverage dichotic presentation or simultaneous spatial presentation of multiple audio streams, if the information intake required is not too high (Ranjan et al, 2006). When audio streams are perceived at different locations (spatialized) then selective attention tasks can be performed better than if the same audio streams are perceived to be at the same location. Divided attention tasks can also perform better under spatialized audio conditions though not as pronounced as selective attention tasks Shinn-Cunningham and Ihlefeld (2004). Simpson Et. al (2004) has also found that spatial minimization techniques were perceived as more usable for selective-attention tasks and not so effective in divided-attention tasks.

Vazquez-Alvarez & Brewster (2010, 2011) carried out experiments using both a divided-attention and selective attention task where a continuous podcast (divided) or classical music piece (selective) and an audio menu compete for attention. Four conditions were tested across the tasks, 1) *baseline*, podcast was interrupted when participant carried out the audio menu tasks with both audio sources 1m in front of participant, 2) *concurrent*, podcast played whilst participant carrying out audio menu task, 3) *user-activated spatial*

*minimization*, podcast was located 1m in front of the listener and moved 90° to the right hand-side only when the participant was engaged in the audio menu tasks. The volume level of the podcast was attenuated by approximately -10 dB, 4) *fixed spatial minimization*, the podcast was kept at 90°and 2m away from the listener whilst the audio menu remained directly in front of listener for entire duration of task, whilst the audio streams were presented continuously.
Results showed that the divided-attention task in simultaneous conditions, increased the users' cognitive load. In these conditions one could expect a performance drop in terms of recall (70 to 50%) and a 10% increase in task time.

## RESEARCH FOCUS

Public safety or other professionals using communication devices may wish to monitor or deal with more than one audio source. Wearable speakers on the shoulder or normal stereo headphones allow different audio sources to be spatialized for the benefits discussed in the previous section. The background literature also noted how a 10 to 15 dB increase in volume on a primary audio source over a secondary audio source can help people focus on the primary audio source better whilst still monitoring a secondary source, though it does not look at a comparison of volume differences across a range of volumes. Vazquez-Alvarez & Brewster (2010, 2011) used a 10dB differential in their spatial audio interface study and whilst they found that for a divided attention task performance dropped significantly, they also had not tested the effect of a range of volumes.

For communication system designers it is of interest if volume differences could help users understand content from different channels as they could make automatic changes to audio stream volumes depending on what was deemed to be a primary source and thus allow the user to maintain a hands free approach to using the system which would allow them to focus on their task. The authors were therefore interested in the following research questions.

1. To understand human capability in listening to two audio streams simultaneously under different audio conditions.

2. Does a difference in volume across a range of volumes (volume spectrum) help users in understanding content while two audio streams play simultaneously?

*Study Outline*
This research study was carried out over a couple of months from research planning, usability study preparation, participants' recruitment, usability testing, and lastly, data consolidation. Usability testing was carried out with 16 internal Motorolan participants (8 male, 8 female) from ages 23 to 54 such that effects of aging on multi stream listening could also be gauged. No other special criteria was required.

## USABILITY STUDY

### Hardware Setup

The test equipment was set up in collaboration with Motorola Solutions engineering team with the main objective was to be able to control two audio stream outputs at the same time.

The experimental setup (see Figure 1) included:

- Control station (for facilitator): Two laptops – each connected by wire to a separate radio (transmitter radio); both laptops were equipped with Windows Media Player.
- Audio output (for participants): A pair of external speakers was placed below the ear on each side with each speaker connected via wire to a receiver radio. (Volume on both sides of radios is adjustable.)

Audio files were played from the control station laptops and transmitted via the connected radios to the worn output speakers. File output from the laptops is set at the same audio level with speaker output loudness levels (what the participant hears) set locally via toggle at the external speakers.

**Audio Streams**

Audio stream samples were acquired from a free English language learning website called the Breaking News English. This website provides more than 2404 English lessons, and in each lesson, news listening files are provided in both British and North American English which can be downloaded in mp3 format. Eighteen (or nine pairs of) news tracks were chosen randomly from the websites with a balanced of British and North American English tracks; each track length is within one to two minutes. These audio tracks were chosen as they provided near continuous verbal output on topics that participants could at least relate to and are narrated in a similar news broadcast format. The eighteen chosen tracks were sent to the engineering acoustics team for 'normalization' to ensure that all audio files have comparable baseline loudness.

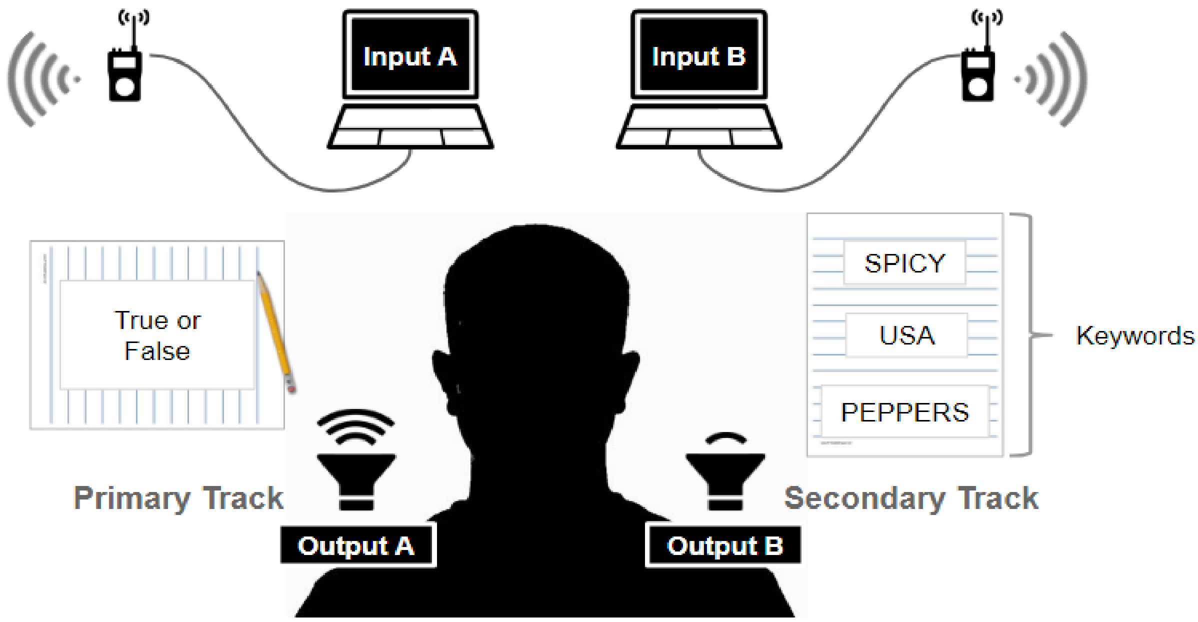


**Fig. 1.** Usability Study set up.

**Testing Outline**

The main intention of this usability study was to be able to provide a dual audio experience to the participants to simulate scenarios where a user would want to listen to a primary audio source and understand the content therein whilst monitoring a second audio source for something of interest (keywords). In each testing, all participants were required to listen to two audio streams at the same time. Every pair of tracks was configured with one track at a higher volume than the other. Tracks with the higher volume were grouped as the primary audio stream and tracks with the lower volume were the secondary audio stream. During testing, the participant had to listen to and remember content from the primary stream whilst scanning for keywords in the secondary stream.

Volume output was controlled by the receiver radio with an adjustable volume level of 1 to 10. In order to provide a comfortable dual audio experience, the team went through a few rounds of volume adjustments. The team found out that any volume below level 3 was too low to catch; volumes above level 8 were too loud to be comfortable. An optimal range of volume was set between Volume 3 (83.5 dB) to Volume 8 (100.6 dB) for tracks to be played to participants. The dual audio streams were then configured based on set volume deltas of Δ1, Δ 2, Δ3 and Δ4 (Table 1). Note though that during test, 4 participants who were comfortable with louder noises had one Δ3 primary track played at Vol 9 to gauge possible effects of the louder volume.

***Table 1 - Audio sets sample with differing volume***

| Volume deltas | Example of audio sets | |
|---|---|---|
| | Primary/ content stream | Secondary/ keywords stream |
| Δ1 | Vol 4 | Vol 3 |
| | Vol 8 | Vol 7 |
| Δ2 | Vol 5 | Vol 3 |
| | Vol 8 | Vol 6 |
| Δ3 | Vol 6 | Vol 3 |
| | Vol 8 | Vol 5 |
| Δ4 | Vol 7 | Vol 3 |
| | Vol 8 | Vol 4 |

**Test Questions**

As previously mentioned, audio tracks with higher volume level were the primary '*content*' stream. Participants were required to listen to and understand the context of the stream. At the end of each testing, participants had to answer five "True or False" questions based on the last listened track (Table 2 gives examples). Whilst mainly concentrating on the primary track, participants were required to scan keywords on the secondary tracks. Before starting each test, participants were given a piece of paper which contain 4 words. Within the 4 words given, only 3 words are correctly adopted from the news stream. All words chosen from the stream are noun or adjectives or combination of both.

***Table 2 - Sample True/False Questions***

| Track | Question Sample |
|---|---|
| Primary Track Article: Ants use the Sun to get from A to B | a) The article says ants are impressive creatures on Earth. |
| | b) The study was conducted in a desert in Spain. |
| | c) The article says we know almost everything about ants. |
| | d) The professor said ants have fairly large brains. |
| | e) Understanding ants will help us to build robot systems. |

**Experimental Procedure**

*Subjects*

Sixteen participants were recruited internally from Motorola Solutions in Penang, Malaysia. Participants were classed into three age group categories: first group (between age 20 - 29), second group (between age 30 - 39) and, third group (age 40 and above).


**Fig. 2.** Speaker Setup

During the recruitment, all volunteers were requested to go through a hearing age test. The hearing age test was conducted through an IOS application - Hearing Test & Ear Age Test. Testing was carried out in a quiet indoor environment, using the same type of earpiece and same application specified volume. From the overall hearing test result, all participants were reported with normal hearing between 3 -18 kilohertz (within the hearing age of 20 - 55). All users except one recorded a hearing age that was younger than their actual age, the exception still falling within their age group bracket.

*Procedure*

There were 16 participants in this study. Volunteers did not have to require any special skill or advance preparation in order to partake this testing. To ensure a fair a comparison as possible, the following controls were implemented:

1. Every testing was carried out in the same simulated environment.
2. The order of each task was randomized for every participant.
3. Standardized test questions were provided according to each audio sample.
4. Primary audio streams were alternately assigned between left and right ears for different users.

**Fig. 3.** Usability Test Session


**Fig.4.** Undertaking Content and Keyword Tasks.

Each test session took roughly 60 minutes. During each test, all participants had to perform three sessions, 15 minutes each. In each session, participants were required to listen to two audio streams simultaneously; each track played to a separate ear. Participants were required to pay more attention to the primary track (at higher volume) as they have to concentrate on the audio content. At the same time, participants had to listen out for keywords on the secondary track. After each test, participants are requested to answer "True or False" questions based on the primary tracks and identify keywords on secondary tracks. Each participant completed 9 pairs of tracks across the 3 sessions. The order of the tasks was randomly assigned.

## RESULTS

This section summarizes and discusses the results obtained from the dual audio usability testing. Results were obtained based on participants' performance on both test tasks - primary (content) and secondary (keywords) tracks. There were 16 participants for this study and altogether there were 144 pairs of tracks tested.

*Gender and Ear Differences*

Figure 5 shows the average score for each content (C) and keyword (K) track from the 9 pairs of tracks played simultaneously to 16 participants, 8 Males and 8 Females users. Looking at the split between male and female participant scores, it can be observed that there is no real difference in performance due to gender. Overall, females have similar scores on both content and keywords category, with scores of 48.61% (content) and 42.24% (keywords). However, males score slightly better than females at content category with an overall score of 51.11% but scored lower at keywords category (38.36%). General differences between left and right ear participants were also assessed. Right ear participants were consistent in both content and keyword scores though left ear participants appeared to score slightly better at content but performed more poorly on the keyword task (Table 3).

***Table 3 - Left/Right Ear Scoring***

| Left Ear | | Right Ear | |
|---|---|---|---|
| Content | Keywords | Content | Keywords |
| 193/360 =**53.6%** | 82/232 =**35.3%** | 166/360 =**46%** | 108/232 =**46.55%** |

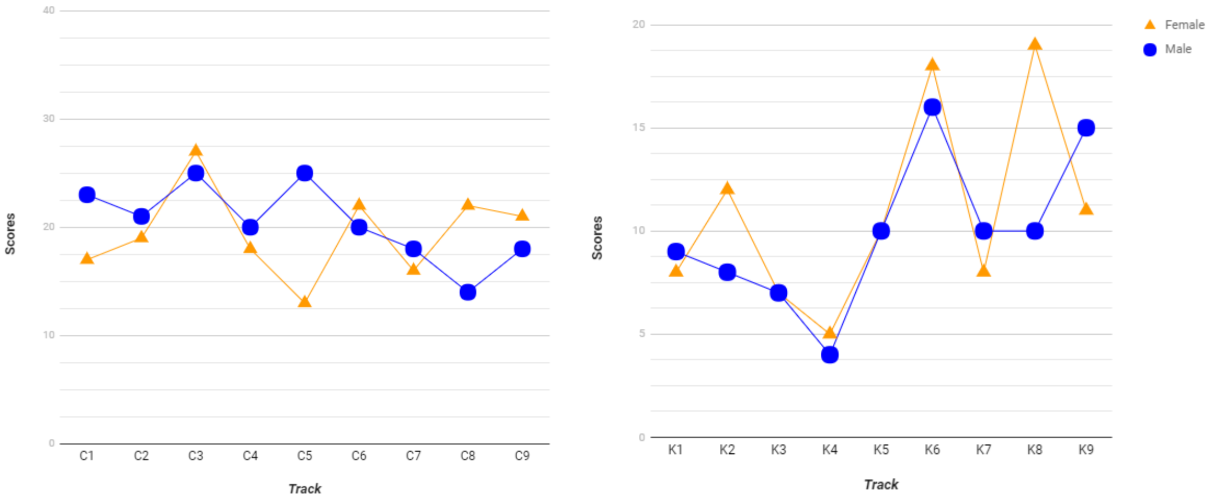


**Fig. 5.** Female (▲) Vs Male (●) on Content (left) and keywords (right) categories.

*Age Group Comparison*

Referring to Figure 6, when looking at performance across age groups, it can be seen that there is little difference in results overall, though there's a notable difference in the over 40s category for listening to keywords.

At content category, participants Group C (40 and above) scored roughly 30% higher compared to Group A (age 20-29) and Group B (age 30-39). Differences between Group A (40.7%) and Group B (44.8%) on content category performance are minimal. On the other hand, at keywords category, Group C participants scored the lowest among the age groups, which is 14.6%. In this category, Group A participants scored 51.7%, and Group B scored 46%.

In general, Group A participants scored lowest (40.7%) at content (primary) category, but scored the highest in keywords (secondary) category. Group B participants had similar scores for both categories, which were 44.8% (content) and 46% (keywords).

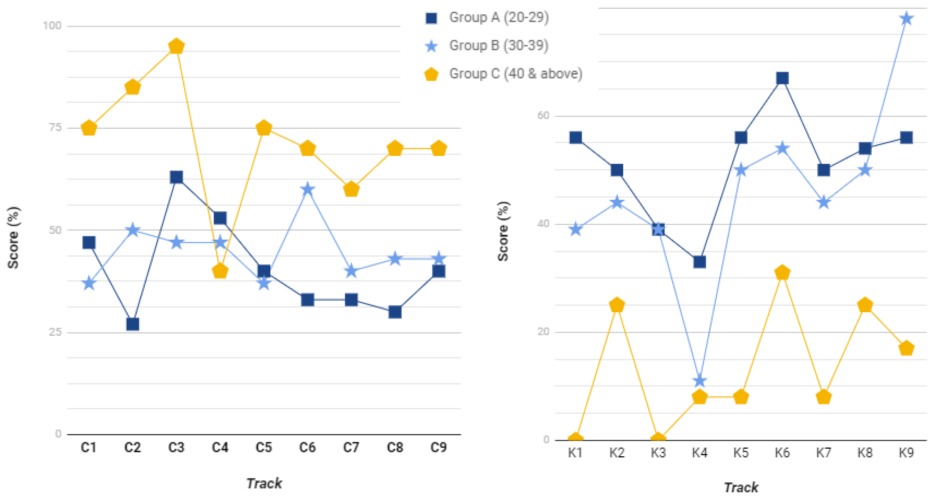


**Fig. 6.** Graph above show the *comparison between age groups on both content (left) and keywords (right) category*


**Fig.7.** *Volume ranges for each occasion that participants scored well on both content and keywords category.*

*Optimal Scoring Volume Ranges*

Here we assessed at what audio levels participants managed to score well in both content and keyword categories as this will give the best indication as to where specific volume ranges may assist in being able to listen to 2 audio streams at same time. If participants are able to listen to two audio streams at same time well then they would have to have good performance in both the content and keyword scores. This was to filter out the results where someone performing very well on one category such as content was likely to do so at the expense of the other category. Whilst overall performances were not very good, for this section it was decided to incorporate better than average scores in both Content (scoring 3 or above) and Keywords (scoring 2 or above). Figure 7, shows the volume ranges for all tasks where participants scored well on both content and keywords. It shows that volume deltas can be anything from 1, 2, 3 or 4 and it spread across many volumes (between 3 to 9). No volume was lower than level 3 because if it's too low then becomes difficult to hear or distinguish from other audio streams; and upper volume levels were only once higher than level 8, because it can be too loud and it will drown out other stream. A clear observation can be made from this result, that there's no best range or volume setting for dual audio listening.

*Audio Delta Effects*

Table 4 represents the average scores among audio deltas, such as, Δ1 (3.4dB difference), Δ2 (6.8dB difference), Δ3 (10.3dB difference), and Δ4 (13.7dB difference). Table 4 shows the scored results on different audio deltas for both content and keywords category. Female participants performed better at the content category at 1 audio level difference, which is 20% higher than keywords category. However, as audio deltas increases from Δ2 to Δ4, the results of the female group showed no obvious score differences for both content and keywords categories. On the other hand, at Δ1, males scored better at keywords category of 52% compared to content category (46%). As audio differences increase, results showed male participants started to perform better at content category with the percentage of 47% (Δ2), 57% (Δ3) and then increase to 68% (Δ4). However, results also showing that, at audio differences Δ2, 3 and 4, keywords category scoring lower than content category. Overall, combining both genders' results, there is no clear pattern to scoring with no one audio delta appearing to perform better and in fact in general most people do not perform well on the tasks.

*Table 4 - Audio Delta Scoring*

| Audio Delta | Female | | | | Male | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Content | | Keywords | | Content | | Keywords | | Content | Keywords |
| Δ1 | 56/100 | **56%** | 24/67 | **35.8%** | 48/105 | **46%** | 29/56 | **52%** | **104/205 = 51%** | **53/123 = 43%** |
| Δ2 | 56/120 | **47%** | 35/68 | **51.5%** | 56/120 | **47%** | 29/78 | **37%** | **112/240 = 47%** | **64/146 = 44%** |
| Δ3 | 53/115 | **46%** | 33/75 | **44%** | 63/110 | **57%** | 22/58 | **38%** | **86/225 = 38%** | **55/133 = 41.3%** |
| Δ4 | 9/20 | **45%** | 6/13 | **46%** | 17/25 | **68%** | 9/17 | **53%** | **26/45 = 58%** | **15/30 = 50%** |

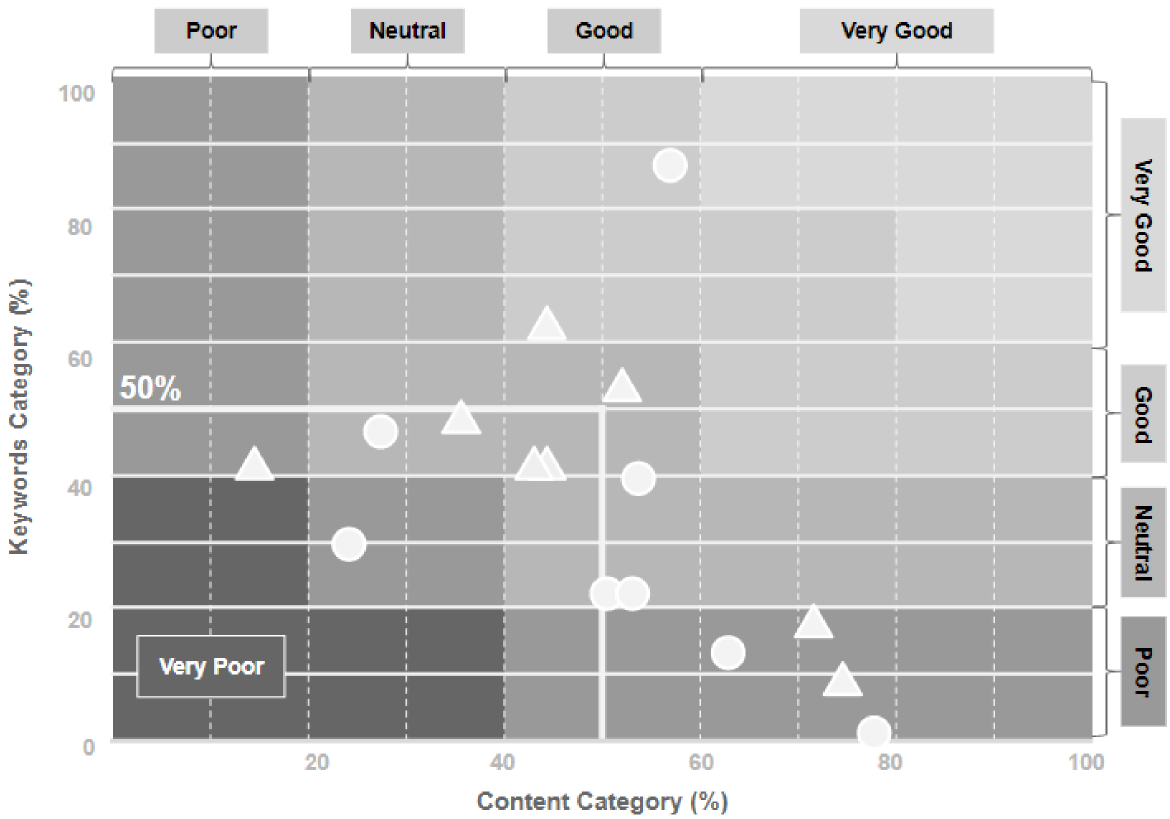


***Fig.8.*** *Combined Performance Overview. (▲-Female Participant; ● - Male Participant)*

*Participants' Comparison Result*

Figure 8 shows an overview of every participant's individual scores on both content (x-axis) and keywords (y-axis) categories. The graph is classed into 5 categories - Very Good, Good, Neutral, Poor, and Very Poor dependent on the percentage scores. Genders are represented as, ▲ (Females) and ● (Males). The average line was set at 50% on both axes. From the plotted graph, results show that apart from one participant who scored reasonably well on both categories, the majority were either around neutral or performed well at the content (e.g. 60 to 80%) whilst not scoring well on the keyword task (0 to 20%), showing that attention was either divided evenly between the tasks or focused more on the primary 'content' task. Some participants were poor in their capabilities to listen to two audio streams at once, scoring poorly on both categories.

## CONCLUSION

Results has shown that there are no specific volumes or volume deltas that assist users in performing better when *listening* to two audio streams simultaneously on 2 spatialized speakers. Dichotic listening ability has been said to be heavily influenced by your genes (National Institutes of Health, 2007) and results here suggested that individual capability may play a larger part in someone's ability to listen to multiple audio streams at the same time. Volume ranges and deltas that optimally assist a person in doing so appear to vary considerably. However audio streams which is less than 84dB (when other stream is higher) becomes difficult to hear whilst audio streams greater than 103dB becomes too distracting to hear the other audio stream.

Therefore devices with multiple audio channels will need the capability to control volumes independently, as generalized auto volume adjustments may interfere with a user's task unless personalized devices can be calibrated to individual preferences. Previous research has shown that if intent by user is to only concentrate on one audio source then increasing volume to be over 10dB higher than competing noise or sound is known to help user concentrate on that source though it could be argued that on a multi channel systems the other channel just be temporarily muted unless just listening for activity on another channel is satisfactory.

Another important factor would be on the specific information professional users would look to monitor on multiple channels and this relate to limitations of the study. Participants in this study were not familiar with the content of the audio in a way which professional day to day users might expect to listen out for commonly used terms and phrases. If one of the audio streams uses more structured phrases or commands then they might be able to scan a secondary audio stream more effectively, though this has to be determined in a follow up study. There is anecdotal evidence from end users that they can be trained to monitor multiple channels assuming there is no continuous information being streamed on each channel e.g. Comms Officers in Korea typically monitor multiple channels on up to 5 devices. In some cases users may not need to initially hear what is said on a specific channel, but just hearing that there is activity on a given channel is of interest and they can then maximize volume on that channel over the other channels if they need to hear the content. However if end users need to listen to multiple channels where the information could be classed as 'conversational' and vary from day to day then they will have to apply more attention to those streams and if this is required on more than one audio stream then their ability to do so may be down to their own skill.

Some real world users may not always need to initially hear content on a given audio channel but just listen out for activity on that channel and then turn its volume up if need be to hear it more clearly. An intelligent multi channel system could monitor for activity and alert the user in the most appropriate manner. The intelligent system could also be used to 'listen out' for keywords etc, especially if the user is busy on another channel and then either alert the user via tones or haptics (so as not to interrupt other channel) or even record that information and playback to the user at a more appropriate time.

Another limitation of the study was that both audio streams were continuous. This was done to ensure a fair comparison of different audio deltas, but in real world scenarios it is unlikely that audio would be so continuous with pauses allowing users to better hear information across sources.

Next steps for the authors include testing real users with familiar content, with varying combinations of unstructured 'conversational' audio vs. structured audio such as commands or commonly used instructions. Testing users in more realistic situations may have less overlap of audio and better chance to monitor 2 channels though will be challenging to standardize and offer fair comparisons between participants. Customer research of user workflows will need to understand the types of audio streams e.g. structured vs. unstructured content and more 'in the field' ability to handle more than 2 audio streams will also need to be investigated. Whilst no major differences were found between left and right ear users in this study, robust dominant ear testing will need to be considered and incorporated into future experiments. It is hoped that by further exploring multi audio stream listening that we can optimize solutions for professional users of communication systems.

**Acknowledgment**

Thanks to the participants and engineers from Motorola Solutions Penang in assisting with this study.

**REFERENCES**

Agnew, J. (1999). Challenges and some solutions for understanding speech in noise. *High Performance Hearing Solutions*, *3*, 4-9.

Aiken, S.J.(2004). Understanding the noise problem. Retrieved from Audiology Online: http://www.audiologyonline.com/articles/understandingthenoiseproblem1093.

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics*, *77*(5), 1465-1487.

Brungart, D. S., & Simpson, B. D. (2005). Improving multitalker speech communication with advanced audio displays. *In New Directions for Improving Audio Effectiveness*, *Neuilly-Sur-Seine, France* (30/1-30/18).

Lau, C.A. (2014, March). Dichotic listening training and CAPDOTS (CAPD Online Therapy System). Retrieved from: http://capdots.com/audiologists-and-speech-pathologists/the-science/scientific-papers/dichotic-listening-training-and-CAPDOTS

Listenwell. (2017). Retrieved from: www.listenwell.com/English/Articles/how_we_listen.html.

Moran, M. (2013). Designing for intelligibility vs audibility. Cooper Notification Solutions. Retrieved from: http://www.cooperindustries.com/content/dam/public/safety/notification/Resources/White_Papers/CN_AudibilitvsIntelligibilityWhitePaper.pdf

National Institutes of Health. (2007). Ability to listen to two things at once is largely inherited. Retrieved from: https://www.nih.gov/news-events/nih-research-matters/ability-listen-two-things-once-largely-inherited

Oviatt, S. (2006, October). Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proceedings of the 14th ACM International Conference on Multimedia, Santa Barbara, CA* (pp 871-880). New York, USA: ACM.

Ranjan, A., Balakrishnan, R., & Chignell, M.

(2006, April). Searching in audio: the utility of transcripts, dichotic presentation, and time-compression. In *Proceedings of Human Computer Interaction, Montreal, Canada* (pp. 721-730). New York, USA: ACM.

Schmandt, C., & Mullins, A. (1995, May). AudioStreamer: Exploiting simultaneity for listening. In *Conference Companion of Computer Human Interaction, Denver, CO* (pp. 218-219). New York, USA: ACM.

Schuett, J. H., & Walker, B. N. (2013, September). Measuring comprehension in sonification tasks that have multiple data streams. In *Proceedings of the 8th Audio Mostly Conference, Pitea, Sweden* (Article 11). New York, USA: ACM.

Shinn-Cunningham, B. G., & Ihlefeld, A. (2004, July). Selective and divided attention: Extracting information from simultaneous sound sources. In Barrass, S. & Vikers, P. (Eds.) *Proceedings of ICAD'04 –Tenth Meeting of the International Conference on Auidtory Display, Sydney, Australia*. Available online at www.icad.org/websiteV2.0/Conferences/ICAD2004/papers/shinn-cunningham_etal.pdf

Simpson, B. D., Bolia, R. S., & Draper, M. H. (2004). Spatial audio display concepts supporting situation awareness for operators of unmanned aerial vehicles. In Vincenzi, D. A., Mouloua, M. & Hancock, P. A. (Eds.) *Human Performance, Situation Awareness, and Automation: Current Research and Trends, I*, 61-65.

Vazquez Alvarez, Y., & Brewster, S. A. (2010, September). Designing spatial audio interfaces to support multiple audio streams. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, Lisboa, Portugal* (pp. 253-256). New York, USA: ACM.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

Yost, W. A., Dye, R. H., & Sheft, S. (1996). A simulated "cocktail party" with up to three sound sources. *Perception & Psychophysics*, *58*(7), 1026-1036.