"© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works."

Investigating cognitive workload in concurrent speech-based information communication

Muhammad Abu ul Fazal[a], Sam Ferguson[a,*], Zafar Saeed[b]

[a]*School of Computer Science, Faculty of Engineering & IT, University of Technology, Sydney*
[b]*Faculty of Information Technology, University of Central Punjab, Lahore, Pakistan*

Abstract

Users are capable of noticing, listening, and comprehending concurrent information simultaneously, but in conventional speech-based interaction methods, systems communicate information sequentially to the users. This mismatch implies that the sequential approach may be under-utilising human perception capabilities and restricting users to seek information to sub-optimal levels. This paper reports on an experiment that investigates the cognitive workload experienced by the users when listening to a variety of combinations of information types in concurrent formats. Fifteen different combinations of concurrent information streams were investigated, and the subjective listening workload for each of the combination was measured using NASA-TLX. The results showed that the perceived workload index score varies in all concurrent combinations. The workload index score depends on the types and the amount of information presented to users. The perceived workload index score in concurrent listening remained the highest in Monolog with Interview (three concurrent talkers) combination, medium in Monolog with News Headlines (two talkers where one is intermittent) combination, and the lowest in Monolog with Music (one talker and a concurrent music stream) combination. Users descriptive feedback remained aligned with the NASA-TLX-based results. It is expected that the results of this experiment will contribute to helping digital content creators and interaction designers to communicate information more efficiently to users.

Keywords: Voice-based Interaction, Concurrent Audio, Speech-based Information

*Sam Ferguson is the corresponding author.

Email addresses: `Muhammad.AbuulFazal@uts.edu.com` (Muhammad Abu ul Fazal), `Samuel.Ferguson@uts.edu.au` (Sam Ferguson), `zafar.saeed@ucp.edu.pk` (Zafar Saeed)

1. Introduction

In an auditory scene, users are capable of focusing their attention on speech-based information streams of their choosing when they receive competing speech-based information in parallel. The well-known example highlighting this phenomenon is the cocktail party problem (Bee and Micheyl, 2008) where a person receives multiple voice streams concurrently and manages to pay attention to a particular stream using the selection and attention abilities by prioritising the interest (Cherry and Taylor, 1954). In contemporary implementations of voice-based interaction, the question arises: Are current information designs optimally utilising human auditory capabilities for voice-based interaction?

Visual interfaces, the most common method of information communication, can provide multiple streams and sources of information in parallel to a user by using a variety of methods, such as numeric displays, text, graphical representations, and even computer user interface elements, one of which is the use of overlays (Ware, 2012; Neil, 2009). Similar concepts theoretically may be adopted in speech-based interaction interfaces for communicating multiple information concurrently, because the human auditory system is capable of performing filtering of the sounds received and can allow users to ignore extraneous noise and concentrate on relevant information (Bregman, 1994; Dix, 2003). For example, a possible increase in information communication could theoretically be provided by broadcasting two voice streams concurrently, one as a primary stream representing the main information, and the other stream, as an assistant that provides additional information based on the context and behaviour (Sato et al., 2011). However, designing such concurrent information streams can be a real challenge that would decide whether such communication method is helpful to the users, or distracts users in interacting with the system.

There are many other applications of concurrent information communication (Guerreiro, 2016). For many applications, such as concurrent speech synthesisers, interactive voice response systems (IVRs), and obtaining audio or video information within large corpora, listening to two concurrent streams and gaining a gist from multiple information streams at the same time could be of great utility. For example, a user might listen to various live talk shows that focus on different topics. That user might be interested in listening to more than one live program at the same time, such as a talk show discussing politics while listening to a program that discusses music. A few other activities among the wider population that motivate to explore concurrent communication are: a) Students engaged in study may have multiple screens at hand. While studying, the student might have their laptop that they are working on, their phone in arms reach and a television/radio/stream playing in the background. b) Parents who are obliged to have a child's program playing on a large television screen, while they have their programming on a secondary (possibly less audible) screen. The parent is likely to be attempting to pay attention to both streams to ensure that appropriate content is playing on the television screen for the child while being entertained by their programming choice. c) Video game players may have an instructional video streaming on one screen while they are gaming on a second screen.

Besides the less critical applications of concurrent speech-based information communication, many other critical real-life domains may benefit from a systematic understanding of concurrent information communication designs. Professionals who engage in listening to multiple talkers simultaneously, such as air traffic controllers, watchstanding sailors, or physicians working in an emergency ward, who generally balance competing priorities, duties, and tasks by listening and interacting with multiple sources simultaneously (Walter et al., 2017) may benefit from concurrent designs. In the medical industry, research is already heading where auditory displays enable the head-up monitoring of the patient during theatre operations (Sanderson, 2006). Similarly, possibilities of non-speech concurrent communication have also been explored in the context of flight-decks (Towers, 2016). Concurrent speech-based information communication in such critical fields would require careful considerations and research.

Exploring concurrent information communication is generally important as it may enable users to perform roles and tasks efficiently, and therefore research interest in this topic has recently been increasing. There is large body of research in the multimodal computing community on modality combinations that allow predictions which modality combinations are favourable in certain conditions and which are not (Munteanu et al., 2016, 2017; Cohen and Oviatt, 2017). However, providing information optimally using multiple channels within the same speech-based interaction mode requires further investigation. As discussed in the section-2, researchers have delineated many parameters that could improve concurrent information communication to users. The research also has shown that concurrent information communication creates a high amount of cognitive challenge for users when listening to multiple information streams (Fazal et al., 2018a, 2020a; Xia et al., 2015). In this regard, to the best of authors' knowledge, what has not been explored is which types of information streams can be concurrently listened to without creating excessive cognitive load. This research explores concurrent speech-based information communication where various types of information streams (including musical streams) are combined to comprehensively investigate the cognitive workload encountered when listening to concurrent information streams.

2. Background

Many researchers (Schmandt and Mullins, 1995; Mullins, 1996; Parente, 2008; Guerreiro and Goncalves, 2016; Ikei et al., 2006; Werner et al., 2015; Hinde, 2016; Towers, 2016; Fazal and Shuaib Karim, 2017; Fazal et al., 2018b; Feltham and Loke, 2017; Deroche and Culling, 2013; Iyer et al., 2013; Aydelott et al., 2015; Brungart and Simpson, 2005) have worked on introducing concurrent communication through auditory display, and have reported remarkable performance by participants listening to two simultaneous voice streams, showing that a listener can process secondary information present in the voice stream that is not the immediate focus. For instance, *AudioStreamer* by Schmandt and Mullins (1995) is one of the first speech interfaces that endeavoured to use people's ability to attend the desired stream from the competing streams selectively. In this system, three concurrent speech-based streams were presented, applying spatial manipulation to each. The system was designed to track head movement to identify user's interest in a stream of the competing streams. Mullins (1996) stated that *AudioStreamer* users were cognitively overwhelmed by three channels of concurrent speech. To overcome this, Mullins suggested introducing five-second onset asynchronies between the streams. Schmandt (1998) introduced *Audio Hallway* as his second speech interface exploiting the concurrent speech-based presentation that allowed the browsing of vast compilations of audio files. Parente (2008) developed a speech interface prototype, called Clique, where users instead of interacting with the underlying graphical interfaces, listened to and interacted solely with the display. For improved television experiences, Hinde (2016) explored how auditory displays can offer an alternative method that depends on users' desire to being able to attend screen-based information. The results showed that offering sound-based secondary content from a smartphone after removing the speech from the television program was the best auditory approach. There are many prototypes addressing user's interaction with the system are introduced by the researchers to communicate speech-based information concurrently.

As the concurrent information communication may also aid in critical domains, the researchers in the U.S. Naval Research Laboratory (NRL) for improving the Navy watch standing operations conducted studies (Brock et al., 2008, 2011) aimed at developing a set of comparative measures of attention and comprehension in a variety of multi-talkers information contexts involving concurrent and serial speech communications. Similarly, for improving a pilot's situational awareness for the changing state of systems information, Towers (2016) supported the use of spatial auditory displays within flight decks. The results of the studies supported the use of concurrent spatial sonifications as it helped users to spend more head-up time to an out of flight deck visual search task and fly the aircraft more precisely.

The speech-based interaction also benefits visually impaired persons for interacting with the system as visually impaired users mostly rely on their auditory system to receive information. Guerreiro and Goncalves (2016) carried out research on blind and sighted users, and conducted experiments to determine the information scanning abilities of the sighted and the visually impaired person from the concurrent speech. Guerreiro and Goncalves leveraged the concept of cocktail party problem. Guerreiro and Goncalves (2016) found that the spatial difference in sources is the best cue in concurrent speech. The study established that sighted and the visually impaired users have similar abilities to scan the information from the concurrent speech (Guerreiro and Goncalves, 2016). Two concurrent information streams were more useful in understanding and identifying the content. The study showed that the use of three speech sources depends on the task intelligibility demands and listener capabilities. In another study by Guerreiro (2013), it was found that the concurrent speech with slightly higher playback-rate enables a significantly quicker scanning for relevant content. The authors of this paper based on their study involving blind and sighted users proposed *Vinfomize* framework (Fazal et al., 2019) that may help in developing systems to communicate multiple voice-based information to the users subject to users' contextual and perceptual needs and limitations.

For providing guidelines to designers to build concurrent speech interfaces, the authors of this paper, reported on an experiment that aimed to test different speech-based designs for concurrent information communication (Fazal et al., 2018b, 2020a,b). In the experiment, a standardised Discourse Comprehension Test (DCT) and custom IELTS test was used. Users were asked to listen to the concurrent streams and then answer the questions from concurrent stimuli. The questions were arranged to determine the comprehension depth by comparing comprehension performance across several different formats of questions (main/detailed, implied/stated). Two audio streams from two types of content were played concurrently to 34 users, in both continuous or intermittent form, with the manipulation of a variety of spatial configurations (i.e. Diotic, Diotic-Monotic, and Dichotic). In total, 12 concurrent speech-based design configurations were tested with each user. The results showed that the concurrent speech-based information designs involving intermittent form and the spatial difference in information streams produce comprehensibility equal to the level achieved in sequential information communication. However, many users reported high cognitive load in concurrent speech-based information communication. The task-based experiment appeared extensive and boring to the users, and cognitive feedback by the users remained task-specific (memorising content). Fifteen users out of 31 said they would 'sometimes' prefer concurrent speech-based communication over sequential. Vazquez Alvarez and Brewster (2010) used a divided-attention task and conducted an experiment where an audio menu and continuous podcast competed for attention. In the experiment, the impact of the cognitive load was assessed using the NASA-TLX subjective assessment tool. The results showed that users' ability to attend two concurrent streams enhances by spatial audio, and also the divided attention creates cognitive load and impacts the overall performance significantly.

This paper extends exploring concurrent speech-based information communication and comprehensively investigates the cognitive workload experienced when listening to a variety of combinations of information types in concurrent formats.

3. Aims & Motivation

3.1. Aims

This experiment aims to comprehensively analyse the cognitive load by subjectively measuring workload that users endure while listening to two different audio streams concurrently. We sought to obtain data to satisfy the following questions: a) Does the cognitive workload remain similar in each concurrent combination? b) Irrespective of combinations, which information type(s) are preferred most by users when presented in concurrent combinations? c) Do users presented with different combinations show differences in preference and likely frequency of use? d) Does an intermittent form of communication in one of two streams create lower cognitive workload in speech-based information communication when compared to the two continuous concurrent streams?

3.2. Motivation

The motivation behind conducting this experiment is to determine the viability of communicating concurrent information within a scenario likely to be encountered by prospective users. Also, some combinations in this study dichotically deliver speech-based information streams in one ear and songs or non-vocal music in the other ear concurrently with an expectation that such presentation of information will smooth a path to deliver information efficiently. Additionally, for some forms of information, e.g. headlines, tweets, or RSS feeds, an intermittent design (4.2.1) may be a natural choice, and may also help with lowering cognitive load. This study will attempt to ascertain whether the intermittent form decreases cognitive load. Knowledge from these studies may help to inform for designing auditory overlays that are similar to visual overlays used in many visual user interface approaches as mentioned in the introduction, section 1.

4. Method

The method adopted for this experiment is outlined below.

4.1. Participants

After receiving institutional Human Research Ethics Committee approval for the research protocol, user participation campaigns were launched. The participants were selected based on two criteria: 1) not having a significant hearing impairment, and 2) having competent English language skills, as the listening experiment's content was in the English language. The users, selected for participation, were offered gift cards worth 30 AU$ each. In total, 40 participants, 20 female, and 20 male took part in the experiment. The mean age of the participants was 23 with the standard deviation of 6.

Table 1: Combinations of Different Types of Information Streams in Concurrent Stimuli

		Monolog	**Interview**	**Commentary**	**News**	**Songs**	**Music**
Source (BBC)	Right Ear	Left Ear					
Lip Service	**Monolog**	-	-	-	-	-	-
Celebrity Inter.	**Interview**	✓	-	-	-	-	-
5 Live Radio	**Commentary**	✓	✓	-	-	-	-
World News	**News**	✓	✓	✓	-	-	-
Radio-1	**Songs**	✓	✓	✓	✓	-	-
Viceroy Movie	**Music**	✓	✓	✓	✓	✓	-

4.2. Design

4.2.1. Concurrent Condition

In this condition, two concurrent information streams were communicated in a dichotic form to users. We created a series of stimuli for concurrent communication where each stimulus was created by combining the two different types of information streams. One stream was presented in the right ear, and the other stream was presented in the left ear, using the panning feature available in an open-source software Audacity (Audacity). The previous studies (Fazal et al., 2018b, 2020a,b; Guerreiro and Goncalves, 2016) showed that the spatial difference between the competing streams is one of the best cues to communicate information concurrently. Therefore, we designed to present information dichotically in this experiment.

The concurrently presented combinations included two of the following sources: Monolog (Documentary), Interview (Dialog), Commentary (Football), News Headlines, Song (Vocal), Music (Non-Vocal), representing a range of common types of auditory content a user of an auditory presentation system might most likely encounter (e.g. a mobile phone user or computer user) in their routine interaction. Each type of information stream was combined with the other types of information streams once, as described in Table 1 and illustrated in Figure 1.

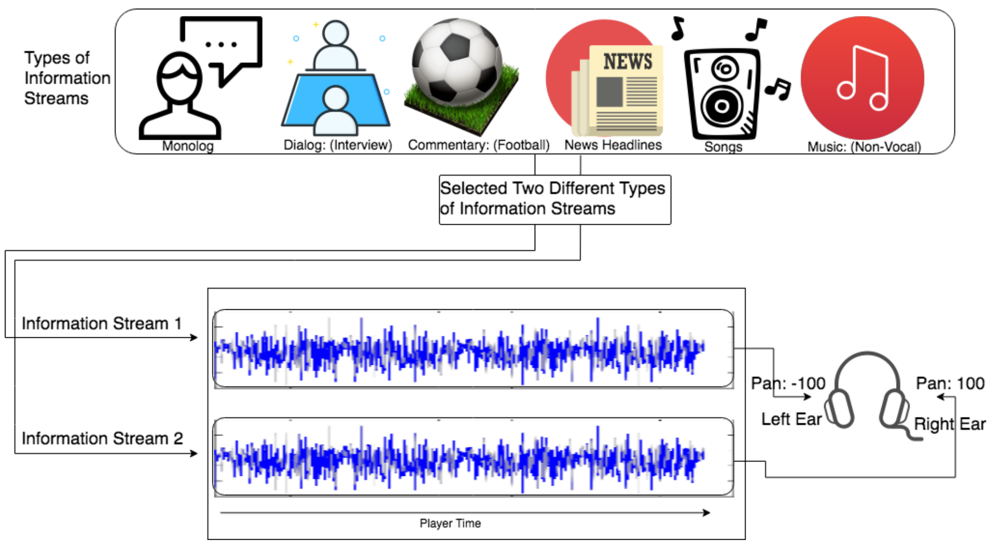

Figure 1: Concurrent Stimulus Design (Continuous)

Intermittent. In all the concurrent combinations, the information streams were presented continuously, except for the combinations where an information stream was combined with a news headlines information stream. Following the intermittent design introduced in our previous study (Fazal et al., 2018b), we manipulated the news headlines stream and transformed it into an intermittent form from the continuous information presentation. For this, the news headlines bulletin was broken into temporal segments, and after each news headline, a silent interval of 15 seconds was added. We involved this intermittent form because our previous studies showed that the users' comprehension was the best in this form among the several other concurrent information communication formats. Therefore, the information streams combined with the news headlines information stream was of an intermittent concurrent design type. The rest of the concurrent combinations were based on continuous information design.

4.2.2. Baseline Condition

The reported experiment mainly focused on investigating and comparing concurrent combinations. The baseline condition was mainly included to have a yardstick (baseline) to perform the comparison between the concurrent combinations. In the absence of a baseline, the one to one comparison between 15 concurrent combinations would have made the analysis and reporting complex.

In this condition, no concurrency was involved. A type of information stream was randomly selected from six information steam types and sequentially communicated to the users in a conventional regular form. The user experience benchmark gained from this baseline condition was used to compare the user experience between concurrent conditions.

4.3. Material

In the previous experiment by these authors (Fazal et al., 2018b, 2020a), users reported that concurrent communication is a fairly interesting and intriguing method that carries the potential to improve the multi-tasking perspective of life, subject to better implementation and considerations of design. They mentioned that concurrent approach could be useful in contexts where information is not critical – for instance, listening about match commentary, news reports, stock markets, and listening to music or sounds, rather than densely layered narratives etc. This feedback, coupled with our deliberations provided us with the foundation to investigate various concurrent combinations.

For this, six types of information streams were selected that included: Monolog (documentary), Dialog (interview), Commentary (football), News Headlines, Songs (vocal) and Music (instrumental / non-vocal). For each information type, BBC online channels were searched to find high-quality samples of information presentations. For each information type, six samples of a maximum of 2 minutes duration were selected, except for the commentary information type. For commentary, the first twelve minutes of the sports match was broken in 6 equal (in duration) files. Following the selection method, there were 36 files in total, each with a duration of 2 minutes.

Based on the ecological choices, the *monolog* streams were selected from the BBC program *Lip Service* wherein each selected documentary, a woman discussed a trait of her life.

Interviews (*dialog*) were selected from the BBC's program *BBC Celebrity Interview* where a male host interviewed a male celebrity. The sports *commentary* was from the BBC 5 Live Radio and was recorded in the male voice covering a football match between Napoli and Manchester City. The *news headlines* spanned six different dates and were selected from the BBC World News. Three news headlines were in the female voice and three in the male voice. The *songs* were selected from the BBC Radio-1 Channel where three of the singers were female and three male. For the *music*, the background music of the Hollywood movie *Viceroy* composed by the Academy Award winner AR Rahman was selected.

4.4. Stimuli Presentation

Since each type of information stream was combined with the rest of information stream types, users were presented with all 15 different concurrent combinations. Besides the concurrent combinations, a baseline stimulus was also presented. Hence, each user was presented with the 16 different stimuli (15 Concurrent, 1 Sequential).

In total, 576 stimuli combinations were created, including the stimuli representing the baseline condition in order to remove the combinational effect. A user was presented with 15 stimuli, each representing one combination, in addition to the baseline stimulus. The randomisation removed the combinational effect, making sure that users were not repeatedly provided with the same information stream (content) twice. The order of presenting combinations was random to remove the ordering effect.

Sixteen sample stimuli out of 576 representing each concurrent combination and baseline stimulus generated for this study are made publicly accessible at Figshare (Fazal and Ferguson, 2021b).

4.5. Measures

In the previous experiment by these authors (Fazal et al., 2018b, 2020a), users also reported that the use of concurrent systems depends on an individual's mental capabilities and differing preferences regarding information processing. This inferred that controls that allow a user to configure the precise format for how to process concurrent information should be given to the users so that they may set up the listening session according to their preferences. This user feedback, coupled with previous research findings provided a basis to design an experimental methodology where results do not get influenced by an intense task (e.g. memorising content). The task was kept simple where users were rationally restricted to the listening task only. This arrangement helped to assess users' interest and cognitive load subjectively where users could freely employ any information processing and attention (e.g. selective, divided) approach as they adopt in real-life scenarios or ecological setting. This design approach also mimicked the cock-tail party effect that provided listeners with the ability to both segregate different stimuli into different streams, and subsequently decide which streams are most pertinent to them for gaining information. Hence, information processing control remained with the user in the experimental design.

Users were asked to listen to the provided combinations and share their experience regarding each combination. The duration of each stimulus was 2 minutes. After listening to each stimulus, users were presented with a questionnaire to share their experience. The experience was primarily obtained using the NASA-TLX subjective, multidimensional assessment tool (Hart and Stavenland, 1988; NASA, 2018b), in addition to descriptive feedback. NASA-TLX is a standardised instrument that rates perceived workload. Besides being cited in over 4400 research studies, this tool has been used in many research studies investigating concurrent communication (McGookin and Brewster, 2004; Parente, 2008; Vazquez-Alvarez et al., 2015; Hinde, 2016; Truschin et al., 2014; Vazquez Alvarez and Brewster, 2010; Towers, 2016; Vazquez-Alvarez et al., 2014; Moffat and Reiss, 2018). The NASA-TLX test has two parts. In the first part, the total workload is measured using the following NASA (2018a) subjective sub-scales, which are rated for each combination within a 100-points range with 5-point steps: 1) "Mental Demand - How mentally demanding was the listening task in a combination? 2) Physical Demand - How physically demanding was the listening task in a combination? 3) Temporal Demand - How hurried or rushed was the pace of the listening task in a combination? 4) Overall Performance - How successful were you in obtaining the relevant information from a combination? 5) Effort - How hard did you have to work to accomplish your level of performance in a combination? 6) Frustration Level - How insecure, discouraged, irritated, stressed, and annoyed were you while listening to a combination?"

The second part of the NASA-TLX procedure created an individual weight of the above mentioned six sub-scales by asking the subjects to compare the dimensions in a pairwise manner based on their perceived importance. Following the standard practice (Hart and Stavenland, 1988; NASA, 2018b), each weight was multiplied by the scale score for each dimension and then divided by 15 to get a workload score from 0 to 100, the overall workload load index.

In order to gain more information about the user experience, we added two additional questions / sub-scales, which, however, were not used to calculate the workload index score: 7. Like (Preference) - How much did you like the combination? 8. Frequent - How frequently will you be using the combination?

4.6. Apparatus

In order to minimise participation time, a usable web-based system using PHP, MySQL, JQuery, HTML5, CSS, and Bootstrap was developed to play the stimuli. Sixteen HTML audio players were designed to play each stimulus design that was presented in sequential web pages. Users were only able to move to the next stimulus when they submitted their NASA-TLX form response for the current stimulus. Users responses were directly recorded in a MySQL database.

The tests were conducted in a quiet purpose-built room in the Creativity and Cognition Studios (CCS) of the University of Technology, Sydney. Three identical Apple iMac computers, having 2.7GHz quad-core Intel Core i5 processor, 8GB RAM, installed with Yosemite 10.10.5 OS were arranged in the studio. To listen to the audio stimuli Beyerdynamic's DT770 250 OHM headphones were used that were connected to the headphone jack of the computer. Users were provided with control of the gain of the system, in order to find a comfortable listening level. Since three computers were used in the studio, up to three participants were engaged in the study simultaneously.

4.7. General Procedure

The users were verbally briefed on the study protocol before the start of the study. Instructions were presented on a screen after registration. Before starting the study, users entered their demographic profile information that included, name, email, age, gender, primary language, qualification, profession, country, mood and hearing/visual impairment status (all profile fields were optional).

In addition to measuring the workload index using NASA-TLX, users were also asked an open-ended question and provided a text area at the end of the experiment to descriptively write their feedback and thoughts regarding their experience from listening to the concurrent information communication. The dataset generated during this study is made publicly available at Figshare (Fazal and Ferguson, 2021a).

5. Results

There were four approaches we undertook to analyse the NASA-TLX-based results: a) the baseline condition was analysed to set the benchmark. b) determined the subjective workload index for each combination and compared with the baseline condition. c) the impact of each information stream type on the user's experience when combined with the rest of the information stream types was explored. d) determined the workload index and other experiential observations of the information stream types with respect to their presentation in the left ear and the right ear. These analysis steps are discussed in the subsequent subsections.

5.1. Baseline Condition

The baseline mean scores for each rating scale was calculated first. The mean rating for the mental demand in baseline condition was 36.75, whereas, for physical demand, temporal demand, effort, frustration and performance was 28.25, 30.00, 38.00, 24.50, 15.88 respectively. Using these subjective subscale ratings, combined with the weighting measure of the NASA-TLX, the calculated mean index score for baseline listening task appeared 30.72. Similarly, regarding the frequent scale, that is asking users what would be the frequency of listening to the baseline condition, the mean rating was 61.37. For the baseline condition, the mean preference rating was 71.12. These ratings set a benchmark to draw comparisons between concurrent combinations.

5.2. Concurrent Combinations

The results of each combination type have been compared with the baseline condition, which provides an indirect comparison between the concurrent combinations. The mean scores differences with the baseline condition are individually illustrated in Figure 2 for all the combination designs. Besides illustrating the results in Figure 2, we discuss the concurrent combinations with the ANOVA results descriptively in the following subsections.

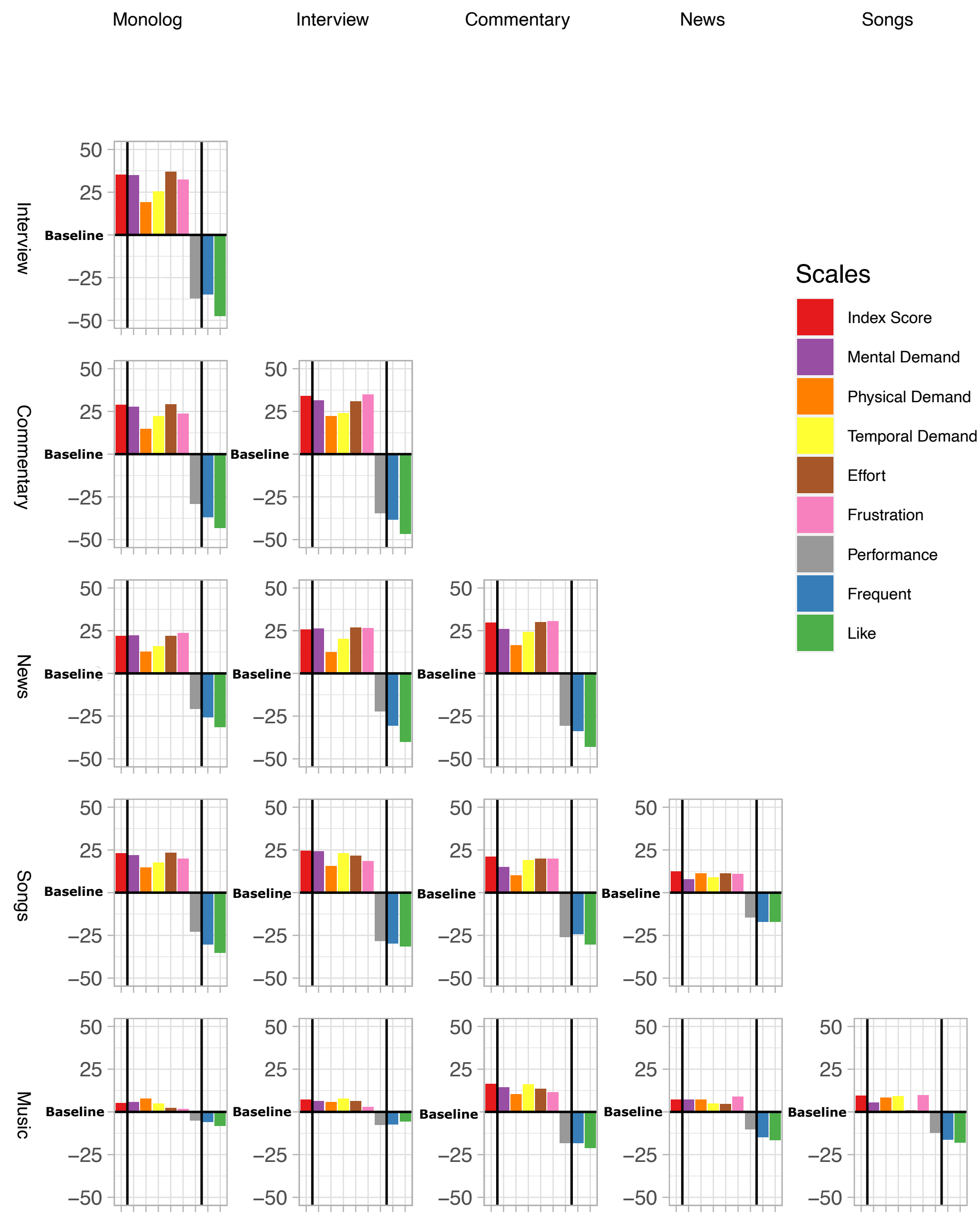

Figure 2: Users' Experience in each Combination in comparison to the baseline condition. Given the reverse nature of NASA-TLX performance scale, the value for this scale was subtracted with 100 to represent the bar in the same sense that a reader will employ to interpret the other scales.

To discuss the results, we categorised combinations into two types: 1) Speech-based information combinations, 2) Music-based (vocal and instrumental) experience combinations. In the speech-based information combinations type, the combinations having both the streams from speech-based information types (i.e. monolog, interview, commentary, news headlines) were categorised, whereas, in music-based experience combinations, the combinations having one stream either from the song or music types were categorised.

5.2.1. Speech-based information combinations

In monolog with interview combination, the mean values of the mental demand, physical demand, temporal demand, effort, frustration, and performance were 71.75, 47.25, 55.50, 75.00, 57.00, 53.13 respectively, whereas, the index score for this combination was 66.05. Similarly, regarding the frequent and preference scales of this combination, the mean ratings was 26.62, and 23.75 respectively. To statistically compare the monolog with interview combination ratings with the baseline condition, we used two-way *analysis of variance*-(ANOVA) test (Copenhaver and Holland, 1988). Two-way ANOVA test showed that the presentation type (monolog with interview — baseline) has a significant impact on user response, $F(1, 702) = 85.793, p < 0.01$. Also, the interaction between the presentation and rating scales, $F(8, 702) = 42.62, p < 0.01$, had a significant impact on user's response. Since the interaction between the presentation and rating scales was significant, we performed the *Post hoc* Tukey HSD analysis (Miller, 198; Yandell, 1997) to compare monolog with interview combination rating with the baseline mean rating. The *Post hoc* Tukey HSD test on the ANOVA results of the interaction between the presentation types and rating scales showed significant differences ($p < 0.05$) in all the scales. Table 2 shows the statistical difference ($p - values$) between this (monolog with interview) combination and the baseline condition concerning each rating scale.

In two other Speech-based continuous combinations, i.e. monolog with commentary and interview with commentary, similar statistical results appeared as seen in monolog with interview combination, see Table 2. The results showed that users' experience in these combination was similar to the monolog with commentary combination.

As mentioned in the methodology section, the news headlines in all combinations were

Table 2: *Post hoc* Tukey HSD Analysis: ($p - values$) comparing mean ratings of speech-based concurrent combinations scales with the relevant baseline scales: ($Signif.codes :<= 0.001 = ***, <= 0.01 = **, <= 0.05 = *$)

Combination	Index	Mental	Physical	Temporal	Effort	Frustration	Performance	Frequency	Like
Monolog w. Interview	***	***	*	***	***	***	***	***	***
Monolog w. Commentary	***	***	0.30	**	***	**	***	***	***
Interview w. Commentary	***	***	**	**	***	***	***	***	***
Monolog w. News	**	**	0.60	0.19	**	**	*	***	***
Interview w. News	***	***	0.66	*	***	***	**	***	***
Commentary w. News	***	***	0.13	**	***	***	***	***	***

presented intermittently with other continuous streams. In monolog with news headlines, the mean index score for this listening task was 52.62. The ANOVA test on this combination showed that the presentation type (monolog with news headlines — baseline) has a significant impact on user response, $F(1, 702) = 26.812, p < 0.01$. Also, the interaction between the presentation and rating scales, $F(8, 702) = 16.931, p < 0.01$, had a significant impact on user's response. In the extended analysis using *Post hoc* Tukey HSD test performing comparison between this combination and the baseline condition, the significant difference appeared in index score ($p < 0.01$). Similarly, because of the significant difference in index score and other rating subscales ($p < 0.05$), users significantly preferred ($p < 0.001$) the baseline condition for frequent use and preference over monolog with news headlines combination. In two other Speech-based intermittent combinations involving news headlines, i.e. interview with news headlines and commentary with news headline, similar statistical results appeared as seen in monolog with news headlines, see Table 2.

5.2.2. Music-based experience combinations

In addition to two concurrent speech-based information streams combinations, song (vocal) and instrumental (non-vocal) music streams in different combinations were included on the assumption that it would enhance user experience.

In the monolog with song combination, a song was presented with a monolog stream and the user experience was evaluated. In this combination, the mean index score for the listening task was 53.73. The two-way ANOVA test comparing this combination with the baseline condition showed the significant difference in users response concerning presentation type

Table 3: *Post hoc* Tukey HSD Analysis: ($p - values$) comparing mean ratings of music-based concurrent combinations combinations scales with the relevant baseline scales: ($Signif.codes$:$<= 0.001 = ***, <= 0.01 = **, <= 0.05 = *$)

Combination	Index	Mental	Physical	Temporal	Effort	Frustration	Performance	Frequency	Like
Monolog w. Song	**	**	0.30	0.08	**	*	**	***	***
Interview w. Song	**	**	0.22	**	**	0.05	***	***	***
Commentary w. Song	**	0.31	0.91	*	*	*	***	**	***
News w. Song	0.67	0.99	0.82	0.97	0.83	0.85	0.40	0.14	0.14
Monolog w. Music	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99
Interview w. Music	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
Commentary w. Music	0.23	0.46	0.92	0.24	0.56	0.83	0.09	0.09	*
News w. Music	1.00	1.00	1.00	1.00	1.00	0.98	0.94	0.41	0.20
Song w. Music	0.97	1.00	0.99	0.98	1.00	0.96	0.78	0.27	0.13

(monolog with song — baseline), $F(1, 702) = 24.57, p < 0.01$. Also, the interaction between the presentation and rating scales had significant impact on the user's response, $F(8, 702) = 20.488, p < 0.01$. In the extended analysis using *Post hoc* Tukey HSD test performing the comparison between this combination and the baseline condition, the significant difference ($p < 0.01$) appeared in index score. Also, frequent, and like scales were significantly different than the baseline condition ($p < 0.001$), see Table 3.

In two more combinations involving song, (that is, interview with song and commentary with song), users' responses were similar to the monolog with song combination. Statistical results, mentioned in Table 3, show that the users rating was significantly different for index scale, and they preferred the baseline condition over these combinations. significantly

In the news headline with song design, intermittent news headlines were combined with song. The results showed that the mean index score for listening task was 43.29. The two-way ANOVA test comparing this combination with the baseline condition showed the significant difference in the users' responses concerning presentation type (news headlines with song — baseline), $F(1, 702) = 7.101, p < 0.008$. The interaction between the presentation and rating scales had significant impact on the user's response, $F(8, 702) = 5.39, p < 0.01$. In *Post hoc* Tukey HSD test performing the comparison between this combination and the baseline condition, no significant difference ($p > 0.05$) appeared in all the rating scales.

In the monolog with music combination, results show that users have a greater interest in this combination than the combinations previously discussed. In this combination, the mean index score for the listening task was 36.01. The two-way ANOVA test comparing this combination with the baseline condition shows neither a significant difference in presentation type (monolog with song — baseline), $F(1, 702) = 1.461, p < 0.227$, nor in the interaction between the presentation and rating scales, $F(8, 702) = 1.08, p < 0.375$. Since no significant impact appeared in the interaction between the presentation and rating scales, in the extended analysis using *Post hoc* Tukey HSD test performing the comparison between this combination and the baseline condition, no significant difference ($p = 1$) appeared in all the rating scales.

In the other four combinations involving music (instrumental — non-vocal), that is interview with music, commentary with music, news headlines with music and song with music, similar statistical results appeared as seen in monolog with music combination discussed above, see Table 3. There was one exception, as a significant difference appeared in like scale in commentary with music combination. This statistical analysis shows that the user experience was similar to the baseline condition in each concurrent combination involving music.

5.3. Information Streams Impact in Concurrent Communication

Besides discussing each combination, we also carried out an overall comparison regarding each information type to determine its impact when presented concurrently with rest of the information types. In other words, we determined the viability of each information type to be concurrently presented with other types of information streams. Figure 3 shows the difference for each information type with the baseline condition. The statistical comparisons of each combination with the baseline condition using *Post hoc* Tukey HSD test are mentioned in Table 4.

In this analysis, we start with the monolog, and follow the same pattern adopted previously, calculating the overall mean values for index score based on the subjective subscales. The ratings for the other two scales that include frequent and like are also compared. The results showed that the index score for listening to a concurrent combinations that had a stream type of monolog was 53.62. For this information type, the two-way ANOVA showed that the stream type (monolog — baseline) had a significant impact on users' response

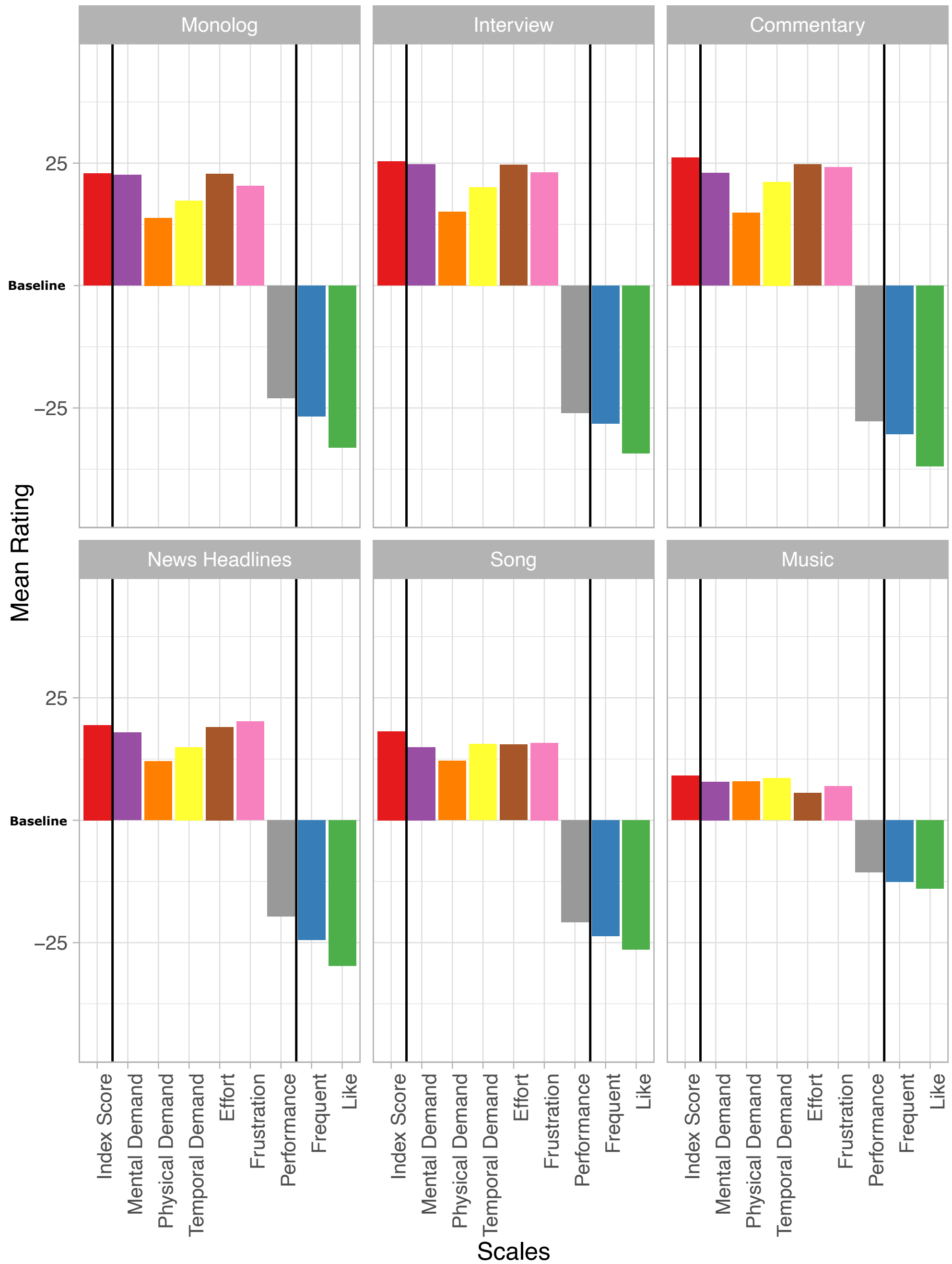

Figure 3: Users' experience regarding each information type when presented with rest of the information types, shown in comparison to the baseline condition

Table 4: *Post hoc* Tukey HSD Analysis: ($p - values$) comparing mean ratings of each stream type with the relevant baseline scales: ($Signif.codes :<= 0.001 = ***, <= 0.01 = **, <= 0.05 = *$)

Stream Type	Index	Mental	Physical	Temporal	Effort	Frustration	Performance	Frequency	Like
Monolog	***	***	**	***	***	***	***	***	***
News	***	***	*	**	***	***	***	***	***
Song	***	***	*	***	***	***	***	***	***
Music	0.30	0.58	0.57	0.41	0.96	0.78	0.09	*	**
Interview	***	***	***	***	***	***	***	***	***
Commentary	***	***	***	***	***	***	***	***	***

$F(1, 2502) = 75.84, p < 0.01$. Also, the interaction between the presentation and rating scales, $F(8, 2502) = 50.069, p < 0.01$, had a significant impact on the users' response. Moreover, the extended analysis using *Post hoc* Tukey HSD test showed significant difference ($p < 0.05$) in all sub-scales. Table 4 shows the statistical difference ($p - values$) between this stream type and the baseline condition. The results for news headlines, song, interview, and commentary, as shown in Table 4, appeared similar as seen in monolog type of stream. In all these types of streams, the index score and other sub-scales difference were significant compared to the baseline condition.

For music type, the same analysis pattern that was adopted for monolog type was followed. The extended analysis using *Post hoc* Tukey HSD test showed no significant differences ($p < 0.05$) in all the scales except frequent use and preference, as shown in Table 4. This shows that the user experience was favoured in this information type when compared to the rest of the information types in concurrent form.

5.4. Impact of Presentation in Left — Right Ears

In this study, the first type of information, that is monolog streams, was always played in the left ear for all relevant concurrent combinations. Similarly, the last information stream type, that is music, was set to play in the right ear, always. However, the rest of the information streams, in some combinations were presented in the left ear, and in some combinations in the right ear.

The interview stream was once presented in the right ear, and four times it was presented in the left ear. The commentary stream was presented twice in the right ear, and the remaining three times it was presented in the left ear. Regarding news headlines, it was presented in the left ear twice, and three times in the right ear. Finally, the song was presented once in the left, and the remaining four times to the right ear of the users in concurrent combinations. The composition is also reflected in Table 1 as mentioned in the method section above.

This composition enabled us to further extended our analysis to see the users' response with reference to presenting information in different ears, and to determine whether one ear had an advantage over the other ear in terms of enhancing the user experience. For this, we compared users response regarding each information stream concerning its concurrent presentation in different ears. The analysis revealed an interesting pattern and showed that users reported lower workload index score for each of the information streams presented in the left ear, and similarly, rated higher for frequent and like scales for left ear presentation. Figure 4-(a) and 4-(b) show the pattern.

Following the same statistical pattern, a three-way ANOVA was performed which included the interactions between the three independent variables to determine the significant impact of all the independent variables on the user's response. The statistical results show that the ear presentation was significant $F(1, 7128) = 78.012, p < 0.01$ in impacting the user response. However, the three-way interaction between the information type, scale type, and the ear presentation variables had a non-significant impact $F(24, 7128) = 0.852, p < 0.671$ on user response, and therefore, we did not perform the *Post hoc* Tukey HSD analysis on ANOVA results.

6. Discussion

In our analysis, we calculated the perceived workload index score for each of the combinations. The study showed that the perceived workload index was significantly high in all concurrent combinations comparing to the baseline condition. The workload index varied in different concurrent combinations. Figure 5-(a) shows an ascending order of the combinations with respect to the perceived index score. The statistical tests, as discussed in section 5, showed significant difference ($p < 0.05$) between the concurrent combinations and the baseline condition.

Similar to the results found for perceived workload index, user responses in preference

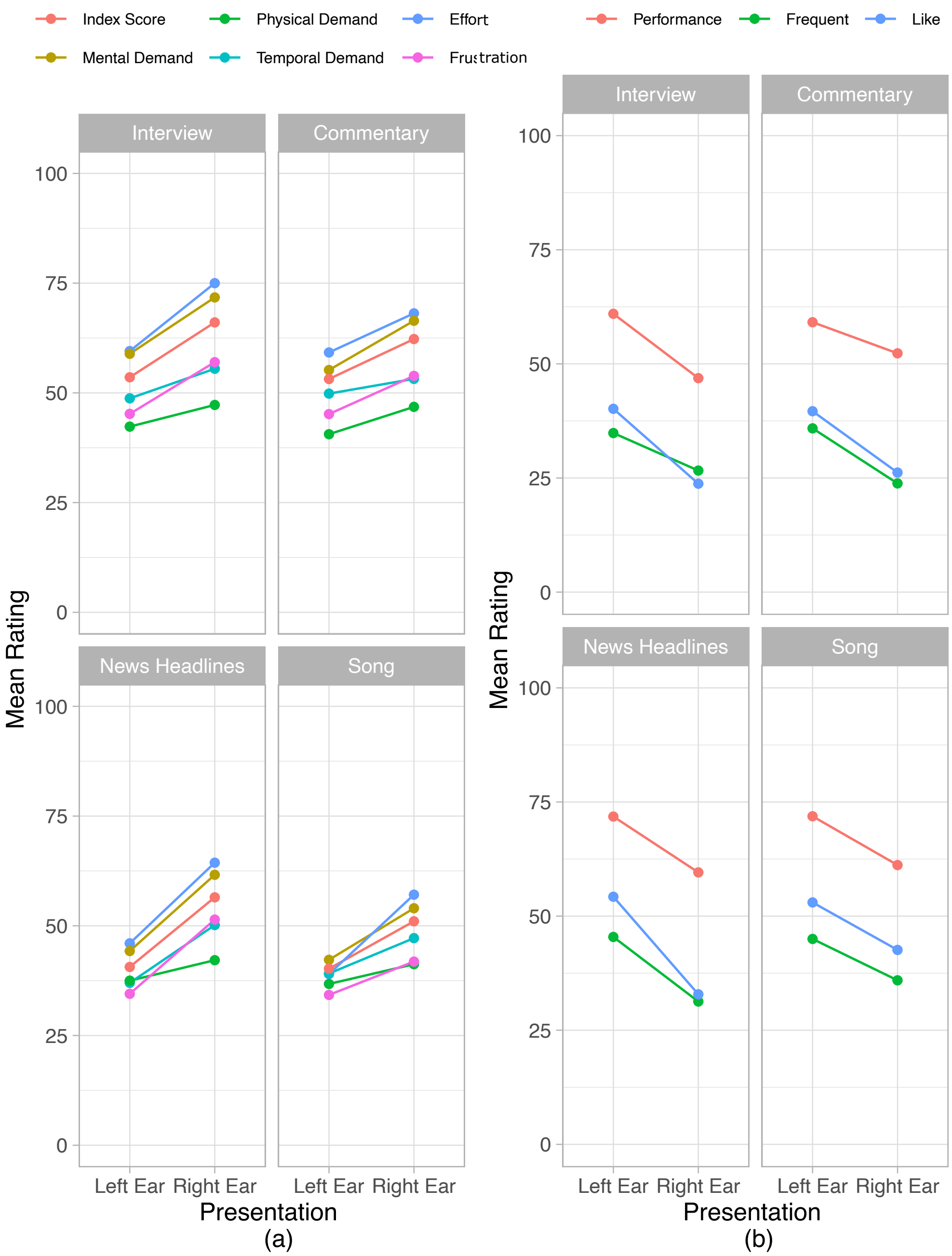

Figure 4: Impact on Users' Experience ('Stress' in (a) and 'Acceptance' in (b)): in each of the four information types with reference to the ear presentation.

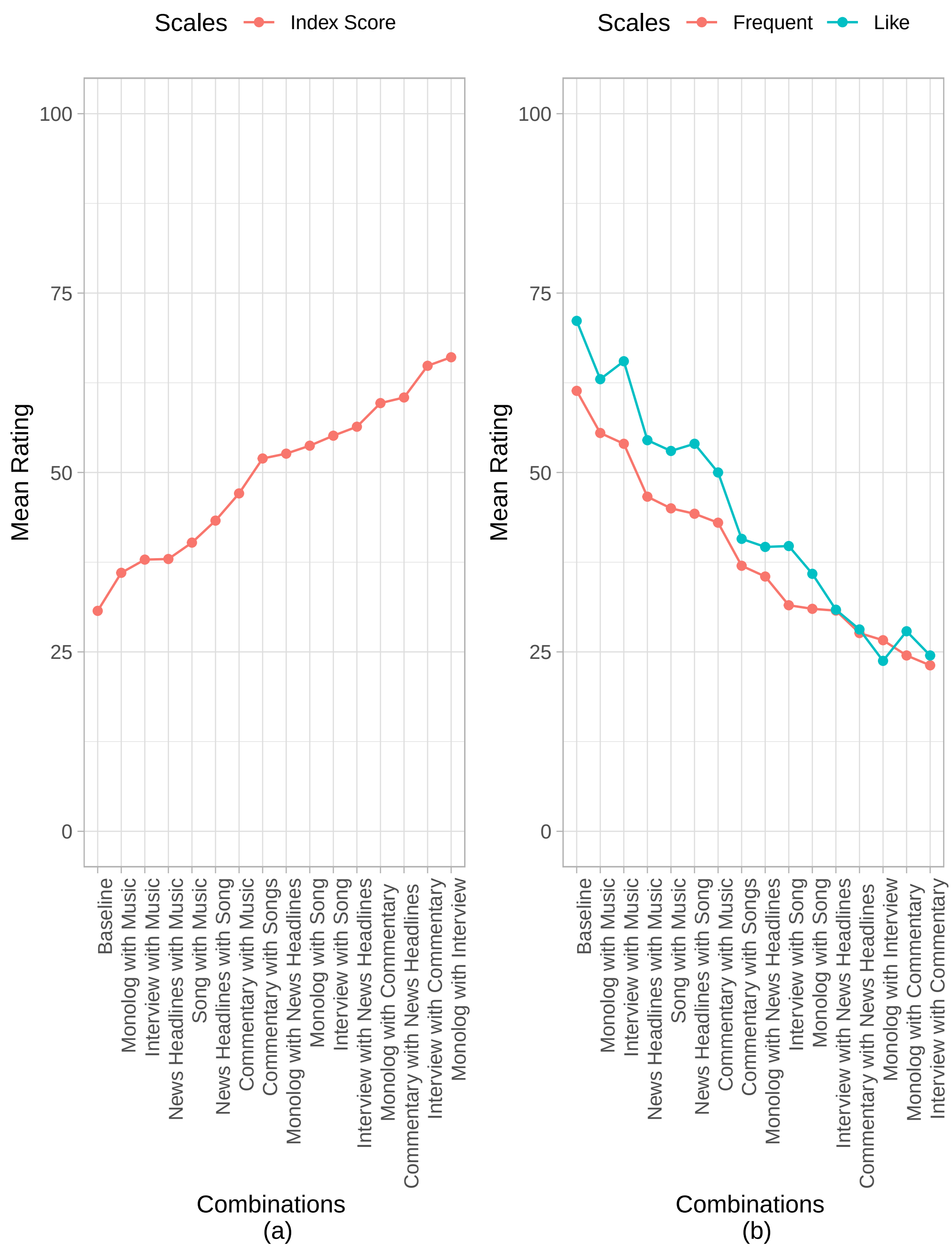

Figure 5: An order of combinations with reference to perceived Workload Index Score shown in (a), and Ratings for Frequency of Use and Preference (Like) shown in (b).

and frequently using different combinations were significantly different. Figure 5-(b) shows an order of the combinations with reference to frequent use. In the statistical analysis for many of the combinations, users' ratings regarding frequent and like scales was significantly less than the baseline condition. In other words, preference and frequent use remained varying in many concurrent combinations, shown in Figure 5-(b).

The illustration in Figure 5-(b), shows a positive correlation between frequent and preference (like) scales. The analysis also shows negative correlation between perceived workload index score and frequent & like scales. This negative correlation relationship between the index score and the frequent and like scales is illustrated in Figure 6-(a). This relationship shows that an increase in the perceived workload index for listening task means the relevant concurrent combination would less likely be preferred for frequent use by the users.

The pattern in Figure 5-(b) also suggests that the perceived workload for a listening task in concurrent combination is dependent on the type of information as well as the amount of information presented to the users. From the order of the combinations appearing in Figure 5-(b), the combinations created with music were preferred the most within concurrent combinations, followed by song-based concurrent combinations. This shows, as the music and songs usually do not require focused attention to process the information stream, there is apparently less cognitive load as users rated them high for frequent use. Similarly, in news headlines, the controlled and limited amount of information was being provided intermittently to the users in chunks, therefore, users selected it the third highest choice to hear them in all concurrent combinations. Similarly, the concurrent combinations created with monolog, interview, and commentary were continuously delivering a high amount of voice-based information, therefore, were rated low for frequent use by the users. This pattern shows that the high amount of information delivery requiring greater attention and cognitive processing from the users to comprehend information makes it less acceptable for the users. Here multiple resource theory can also provide guidance as to how careful redesign of concurrent combinations can restore performance to the residual capacity region (i.e. the demand is less than the capacity of resources available) (Wickens, 2008).

The extended analysis discussing the viability of information type in concurrent combination also validates the above observation that the perceived workload for a listening task

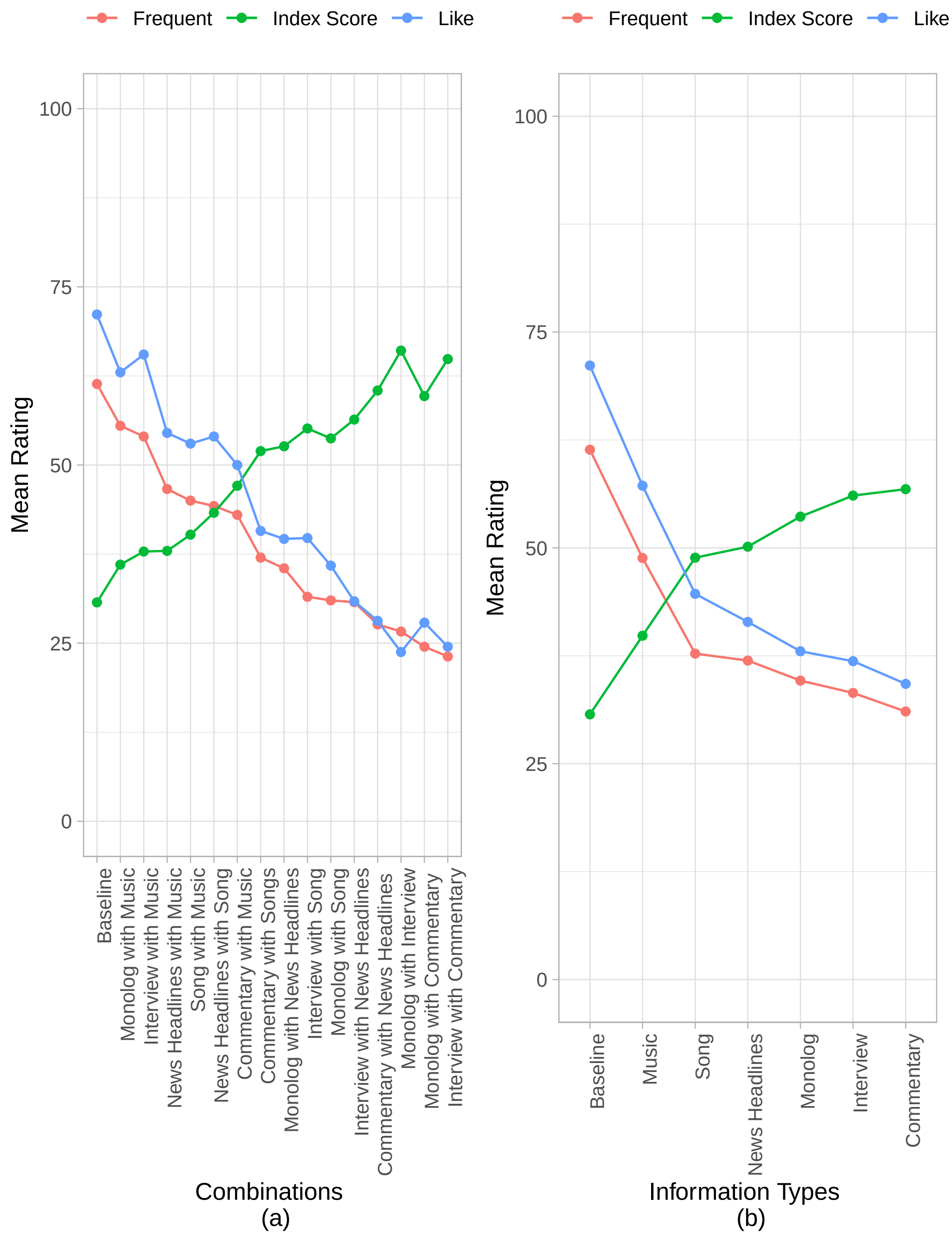

Figure 6: Order of Combinations with reference to their listening task index score, ratings for frequent and preference reported by the users, shown in (a), and order of Information Types with reference to their potential for being a part of concurrent communication, shown in (b).

in concurrent combination is dependent on the type of information as well as the amount of information presented to the users. The order shown in Figure 6-(b) validates the same as users rated music the highest regarding frequent use when listened in concurrent combination, followed by song as the second.

From the information type providing speech-based information (non-music/song), users rated news headlines the highest for frequent use when listening to concurrent combinations. Rating news headlines the highest support our previous studies (Fazal et al., 2018a, 2020a) that show the intermittent design with the spatial difference in sources is the best form to communicate multiple information concurrently. In our previous study, in the intermittent form of concurrent communication, users comprehended the content equal to the amount that they comprehended in the baseline sequential presentation. In this study, users reported that intermittent form of communication creates the least perceived workload index in speech-based information communication. From the speech-based information types, after the news headlines, users rated monolog the second highest information stream regarding frequent use. In monolog, one speaker presented information, whereas in interview (dialog) two speakers were involved in presenting information. Comparing these two types of information streams, users rated monolog higher than the interview. This shows that the number of talkers in concurrent streams also affects users perceived workload index. A stream with one talker is rated higher in terms of frequent use than the stream involving two talkers.

We used commentary in combination designs on the assumption that users usually do not pay in-depth attention to commentary types of information streams. Users mostly remain interested in gaining gist from the match that they can get on different points, for example, the commentator becomes louder and passionate indicating that some interesting event is happening in the field. Such cues may help the users to divert their attention immediately towards commentary with increased focus, else, pay attention towards the other competing streams. It was expected that combinations involving commentary would be liked profoundly by the users. Contrarily, users rated commentary the least. There could be many factors, such as there being background noise generated from the spectators in the commentary stream or the speaking speed of the commentator being fast in order to keep up with the pace of the game. The higher speaking rate/pace might also have created a difference.

Some researchers, (Aydelott et al., 2012; Westerhausen and Hugdahl, 2008; Hugdahl, 2016), showed that the right ear may provide some advantages in competing voice streams because in dichotic listening, the signal that reaches to right ear gets direct access to the left posterior temporal lobe for speech processing in the brain, whereas the signal presented to left ear enters to the wrong hemisphere that then gets transferred across the corpus callosum for processing (Westerhausen and Hugdahl, 2008; Pollmann et al., 2002). The REA offers better reporting of the content from the voice stream presented to the right ear compared to the other competing voice presented in the left ear of the listeners in dichotic listening (Hugdahl, 2016). However, in our experiment, the detailed analysis showed that in concurrent communication, users preferred an information type more when presented in the left ear as compared to the right ear presentation. In all the information types, the analysis showed that users reported lower workload index score for each of the information stream when it was presented in the left ear, and rated higher for frequent and preference scales compared to right ear presentation, and vice-versa.

It appeared that the monolog presented with the music achieved almost the same user experience as users enjoyed in the baseline condition. The user experience almost remained the same in baseline condition and combinations with music. This shows that presenting music with an information stream does not create a significant difference in user experience as compared to the baseline condition. Some users rated combinations with music higher than the baseline condition in terms of frequent use and preference. It sets another direction for investigation to test the impact of music presented in one ear while comprehending content from other voice-based streams presented in the other ear. Many people do tasks with music playing in the background. Dichotic listening could be tested where a speech-based information stream is provided in one ear and the music stream in another ear. A study based on the same design pattern that was adopted in a previous study of the researchers (Fazal et al., 2018a) could be used to compare the content comprehension in such concurrent communication with the baseline sequential communication.

6.1. Users Descriptive Feedback

As mentioned in section 4, at the end of the experiment, users were also requested to share feedback descriptively regarding their concurrent listening experience. Users feedback were received using a text area and stored in the database. Users feedback remained mostly aligned with the results we received from the NASA-TLX measure.

Most of the users found the experiment very interesting and also listening to the concurrent information streams. Users mentioned that it was a unique experience of listening to two different sounds and interpreting them simultaneously. They said that it was amusing to see and realise how we can understand what we focus on even with many distractions around. A user mentioned that *"It was sometimes good to try and prioritise one track at any given moment. This was not necessarily a bad thing, but it was challenging at times"*.

Six users mentioned that they are not used to listening to the multiple information concurrently and can only focus on one of both information streams simultaneously. They did not like the idea of listening to concurrent information. A user mentioned that *"It can be very frustrating to listen to two streams simultaneously. I personally always preferred the moments during which only one stream was playing. It was easier to understand and less frustrating"*.

Regarding users' experience, reactions, and expectations provided descriptively, the following sub-sections concisely discuss the points reported by users based on their experience. These points may provide several hints and apparently open avenues for human-computer interaction researchers to explore the directions of communicating multiple information concurrently.

Combinations Type — Users in their feedback pointed out that their listening experience depended on the types of information streams combined to form a combination. Users mentioned that listening to some of the concurrent combinations went pretty well, and others were frustrating and contrasting. As seen in NASA-TLX-base results they generally found a combination of music and a personal account more enjoyable and thus easier to listen to. In contrast, a user mentioned *"the experience depended on the different audios [combinations] displayed. For example, audios [combinations] like the interviews or personal experiences were hard to understand with music or the news. Because you want to be focused on the interview, it is really frustrating to be disturbed. Nevertheless, it was quite ok to listen to some music while hearing a caster commenting a match"*. Most users liked the combinations that included a music stream.

Attention and Selection — A few users shared their experience about the attention and selection of information streams when listening to the concurrent information communication. Users mentioned that their attention mostly was asymmetric; often biased towards the storytelling streams, not the music or sports-commentary. A user mentioned that *"It can be very frustrating to listen to two streams at the same time. But it was sometimes good to try and prioritise one track at any given moment"*. Like cocktail party effect, users applied attention and selection abilities to select information stream at given time.

Comprehension Challenge — Some users mentioned though it was interesting to listen to concurrent streams, they had a challenge to fully comprehend the concurrent information stream. A user mentioned that *"I think the most frustrating part was not fully comprehending what I was focusing on. That made me feel annoyed at myself"*.Comprehending concurrent information streams was cognitively demanding. The same is reflected in NASA-TLX measurement analysis.

Fatigued vs. Training — Some users commented that the experiment was lengthy, and it was hard for them to focus towards the end. For example, a user suggested that the number of combinations to listen to should have been ten. Some users, on the other hand, reported that at the beginning, it was more difficult to focus on concurrent streams than at the end, as they got used to focusing on both information streams at the same time. This sets to an interesting study to comprehensively analyse whether users get training from the previous listening experiences and perform better in the subsequent listening tasks.

Type of Music — Many users mentioned they liked especially the parts where there was epic music and a nice, dramatic story at the same time. Some of the users mentioned that fast-paced, beat-heavy music was frustrating and distracting when played simultaneously with speech. They could not focus when there was a strong beat at the back. However, for some users, the combinations were enjoyable, which had speech and slow music. A user mentioned that *"some of the combinations were really easy [to] comprehend, mostly the ones with some music on the other side but without lyrics"*.

Subject Matter — A few users mentioned that their response varied according to their interest in the subject matter. A user mentioned that *"It was interesting to hear multiple different sound combinations because I started to find the narrative stories interesting and* *compelling after a minute or so"*. This shows that content matter can benefit the concurrent communication if it carries the listener's interest.

Intermittent Information Stream — Some users reported that they generally felt disturbed when music or a discourse interrupted by the intermittent information stream (News Headlines). *"It was frustrating at times because you would try to listen to the song or the interview, and then something would interrupt and distract your attention away from it"*. One user mentioned that *"Shifting attention to sudden information stream caused them to forget about the other stream"*. For some users, the intermittent information presentation created issues with focus and attention.

Overall, users feedback mostly remained aligned with the findings we received using the NASA-TLX measure. Users reported about the high cognitive workload as seen in the NASA-TLX. Users told their interest remained dependent on the type of information presented and the amount of information presented. However, some users in their descriptive feedback didn't like the intermittent presentation of information.

7. Conclusion

Despite experiencing a high workload index in listening to the concurrent information, the majority of users showed interest in such communication method, which carries the potential to improve the multi-tasking perspective of life, subject to better implementation and considerations of design. Our study showed that cognitive workload varies in concurrent communication. The workload index score depends on the types and the amount of information presented to users. Our study showed many insightful findings of concurrent communication. Achieving the better implementation of such an efficient system requires further investigations to design concurrent information stimulus carefully. For example, 1) comprehensively analysing whether users get training from the previous listening experiences and consequently perform better in the subsequent listening tasks, and 2) testing the impact of music while comprehending the content from the competing speech-based stream presented in another ear are a few to mention.

Acknowledgment

This research was supported by the School of Computer Science, Faculty of Engineering and IT, University of Technology Sydney, Australia.

References

Audacity, . Audacity. https://www.audacityteam.org/. [Online; accessed 23-Dec-2018].

Aydelott, J., Baer-Henney, D., Trzaskowski, M., Leech, R., Dick, F., 2012. Sentence comprehension in competing speech: Dichotic sentence-word priming reveals hemispheric differences in auditory semantic processing. Language and Cognitive Processes 27, 1108–1144. doi:10.1080/01690965.2011.589735.

Aydelott, J., Jamaluddin, Z., Nixon Pearce, S., 2015. Semantic processing of unattended speech in dichotic listening. The Journal of the Acoustical Society of America 138, 964–975. doi:10.1121/1.4927410.

Bee, M.A., Micheyl, C., 2008. The cocktail party problem: what is it? how can it be solved? and why should animal behaviorists study it? Journal of comparative psychology 122, 235.

Bregman, A.S., 1994. Auditory scene analysis: The perceptual organization of sound. MIT press.

Brock, D., McClimens, B., Trafton, J.G., McCurry, M., Perzanowski, D., 2008. Evaluating listeners' attention to and comprehension of spatialized concurrent and serial talkers at normal and a synthetically faster rate of speech, in: Proceedings of the 14th International Conference on Auditory Display (ICAD), Georgia Institute of Technology. pp. 1–8.

Brock, D., Wasylyshyn, C., Mcclimens, B., Perzanowski, D., 2011. Facilitating the watchstander's voice communications task in future navy operations, in: MILCOM 2011 Military Communications Conference, pp. 2222–2226. doi:10.1109/MILCOM.2011.6127692.

Brungart, D.S., Simpson, B.D., 2005. Optimizing the spatial configuration of a seven-talker speech display. ACM Trans. Appl. Percept. 2, 430–436.

Cherry, E.C., Taylor, W.K., 1954. Some Further Experiments upon the Recognition of Speech, with One and with Two Ears. The Journal of the Acoustical Society of America 26, 554–559. doi:10.1121/1.1907373, arXiv:NIHMS150003.

Cohen, P., Oviatt, S., 2017. Designing speech, acoustic and multimodal interactions. Association for Computing Machinery (ACM), United States of America. volume 1. pp. 403–447.

Copenhaver, M.D., Holland, B.S., 1988. Multiple comparisons of simple effects in the two-way analysis of variance with fixed effects. Journal of Statistical Computation and Simulation , 1–15.

Deroche, M.L.D., Culling, J.F., 2013. Voice segregation by difference in fundamental frequency: Effect of masker type. The Journal of the Acoustical Society of America 134, EL465–EL470. doi:10.1121/1.4826152.

Dix, A., 2003. Human-computer interaction. Pearson. doi:10.4135/9781412982818.n29.

Fazal, M.A.u., Ferguson, S., 2021a. Investigating cognitive workload in concurrent speech-based information communication - csv. URL: https://bit.ly/3gJWvLs, doi:10.6084/m9.figshare.14743422.v1.

Fazal, M.A.u., Ferguson, S., 2021b. Investigating cognitive workload in concurrent speech-based information communication - sample files. URL: https://bit.ly/3cQtTiA, doi:10.6084/m9.figshare.14743545.v1.

Fazal, M.A.u., Ferguson, S., Johnston, A., 2018a. Investigating concurrent speech-based designs for information communication, in: Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion, ACM, New York, NY, USA. pp. 4:1–4:8. URL: http://doi.acm.org/10.1145/3243274.3243284, doi:10.1145/3243274.3243284.

Fazal, M.A.U., Ferguson, S., Johnston, A., 2020a. Evaluation of information comprehension in concurrent speech-based designs. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16, 1–19.

Fazal, M.A.u., Ferguson, S., Johnston, A., 2020b. Investigating efficient speech-based information communication: a comparison between the high-rate and the concurrent playback designs. Multimedia Systems 26, 621–630.

Fazal, M.A.u., Ferguson, S., Karim, M.S., Johnston, A., 2018b. Concurrent Voice-Based Multiple Information Communication: A Study Report of Profile-Based Users' Interaction, in: 145th Convention of the Audio Engineering Society, Audio Engineering Society.

Fazal, M.A.u., Ferguson, S., Karim, M.S., Johnston, A., 2019. Vinfomize: A Framework for Multiple Voice-based Information Communication, in: Proceedings of the 3rd International Conference on Information System and Data Mining, ACM - Accepted. pp. 1–7. Accepted - to be published.

Fazal, M.A.u., Shuaib Karim, M., 2017. Multiple information communication in voice-based interaction, in: Advances in Intelligent Systems and Computing. Springer, pp. 101–111. doi:10.1007/978-3-319-43982-2_9.

Feltham, F., Loke, L., 2017. Felt Sense through Auditory Display: A Design Case Study into Sound for Somatic Awareness while Walking, in: Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition, ACM. pp. 287–298.

Guerreiro, J., 2013. Using simultaneous audio sources to speed-up blind people's web scanning, in: Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility, ACM. pp. 1–2. doi:10.1145/2461121.2461154.

Guerreiro, J., 2016. ENHANCING BLIND PEOPLE'S INFORMATION SCANNING WITH CONCURRENT SPEECH. Ph.D. thesis. University of Lisbon.

Guerreiro, J., Goncalves, D., 2016. Scanning for digital content: How blind and sighted people perceive concurrent speech. ACM Transactions on Accessible Computing 8. doi:10.1109/CVPR.2016.105, arXiv:1512.01413.

Hart, S.G., Stavenland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, in: Hancock, P.A., Meshkati,

N. (Eds.), Human Mental Workload. Elsevier. chapter 7, pp. 139–183. URL: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000004342_1999205624.pdf.

Hinde, A.F., 2016. Concurrency in auditory displays for connected television. Ph.D. thesis. University of York.

Hugdahl, K., 2016. Dichotic listening and attention: the legacy of Phil Bryden. Laterality 21, 433–454. doi:10.1080/1357650X.2015.1066382.

Ikei, Y., Yamazaki, H., Hirota, K., Hirose, M., 2006. vCocktail: multiplexed-voice menu presentation method for wearable computers, in: Virtual Reality Conference, IEEE. pp. 183–190. doi:10.1109/VR.2006.141.

Iyer, N., Thompson, E.R., Simpson, B.D., Brungart, D., Summers, V., 2013. Exploring auditory gist: Comprehension of two dichotic, simultaneously presented stories, in: Proceedings of Meetings on Acoustics, Acoustical Society of America. pp. 050158–050158. doi:10.1121/1.4800507.

McGookin, D.K., Brewster, S.A., 2004. Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition. ACM Trans. Appl. Percept. 1, 130–155.

Miller, R.G., 198. Simultaneous Statistical Inference. Springer.

Moffat, D., Reiss, J.D., 2018. Perceptual evaluation of synthesized sound effects. ACM Trans. Appl. Percept. 15.

Mullins, A.T., 1996. Audiostreamer: Leveraging The Cocktail Party Effect for Efficient Listening. Ph.D. thesis. Massachusetts Institute of Technology.

Munteanu, C., Irani, P., Oviatt, S., Aylett, M., Penn, G., Pan, S., Sharma, N., Rudzicz, F., Gomez, R., Cowan, B., Nakamura, K., 2017. Designing speech, acoustic and multimodal interactions, in: Lampe, C., schraefel, m. (Eds.), CHI'17 Extended Abstracts - Proceedings of the 2017 ACM SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery (ACM), United States of America. pp. 601–608.

Munteanu, C., Irani, P., Oviatt, S., Aylett, M., Penn, G., Pan, S., Sharma, N., Rudzicz, F., Gomez, R., Nakamura, K., et al., 2016. Designing speech and multimodal interactions for mobile, wearable, and pervasive applications, in: Lampe, C. (Ed.), CHI'16 - San Jose, CA USA May 7-12, ACM, United States of America. pp. 3612–3619.

NASA, 2018a. NASA Task Load Index Sheet. https://humansystems.arc.nasa.gov/groups/TLX/downloads/TLXScale.pdf. [Online; accessed 22-Dec-2018].

NASA, 2018b. NASA-TLX. https://humansystems.arc.nasa.gov/groups/TLX/. [Online; accessed 22-Dec-2018].

Neil, T., 2009. Designing web interfaces: Principles and patterns for rich interactions. O'Reilly Media, Inc.

Parente, P., 2008. Clique: Perceptually based, task oriented auditory display for GUI applications. Ph.D. thesis. The University of North Carolina at Chapel Hill.

Pollmann, S., Maertens, M., Von Cramon, D.Y., Lepsien, J., Hugdahl, K., 2002. Dichotic listening in patients with splenial and nonsplenial callosal lesions. Neuropsychology 16, 56–64.

Sanderson, P., 2006. The multimodal world of medical monitoring displays. Applied Ergonomics 37, 501–512. doi:10.1016/j.apergo.2006.04.022.

Sato, D., Zhu, S., Kobayashi, M., Takagi, H., Asakawa, C., 2011. Sasayaki: augmented voice web browsing experience, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM. pp. 2769–2778.

Schmandt, C., 1998. Audio hallway: a virtual acoustic environment for browsing, in: Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology, ACM. pp. 163–170. doi:10.1145/288392.288597.

Schmandt, C., Mullins, A., 1995. AudioStreamer: Exploiting simultaneity for listening, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM. pp. 218–219. doi:10.1145/223355.223533.

Towers, J.A., 2016. Enabling the Effective Application of Spatial Auditory Displays in Modern Flight Decks. Ph.D. thesis. The University of Queensland.

Truschin, S., Schermann, M., Goswami, S., Krcmar, H., 2014. Designing interfaces for multiple-goal environments. ACM Transactions on Computer-Human Interaction 21, 1–24. doi:10.1145/2544066.

Vazquez-Alvarez, Y., Aylett, M.P., Brewster, S.A., von Jungenfeld, R., Virolainen, A., 2014. Multilevel auditory displays for mobile eyes-free location-based interaction, in: Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, ACM. pp. 1567–1572.

Vazquez-Alvarez, Y., Aylett, M.P., Brewster, S.A., Jungenfeld, R.V., Virolainen, A., 2015. Designing interactions with multilevel auditory displays in mobile audio-augmented reality. ACM Transactions on CHI 23, 1–30.

Vazquez Alvarez, Y., Brewster, S.A., 2010. 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, ACM. pp. 253–256.

Walter, S.R., Raban, M.Z., Dunsmuir, W.T., Douglas, H.E., Westbrook, J.I., 2017. Emergency doctors' strategies to manage competing workload demands in an interruptive environment: An observational workflow time study. Applied Ergonomics 58, 454–460. doi:10.1016/j.apergo.2016.07.020.

Ware, C., 2012. Information visualization: perception for design. Elsevier.

Werner, S., Hauck, C., Roome, N., Hoover, C., Choates, D., 2015. Can VoiceScapes assist in menu navigation?, in: Proceedings of the Human Factors and Ergonomics Society, pp. 1095–1099. doi:10.1177/1541931215591157.

Westerhausen, R., Hugdahl, K., 2008. The corpus callosum in dichotic listening studies of hemispheric asymmetry: A review of clinical and experimental evidence. Neuroscience and Biobehavioral Reviews 32, 1044–1054.

Wickens, C.D., 2008. Multiple resources and mental workload. Human factors 50, 449–455.

Xia, J., Nooraei, N., Kalluri, S., Edwards, B., 2015. Spatial release of cognitive load measured in a dual-task paradigm in normal-hearing and hearing-impaired listeners. The Journal of the Acoustical Society of America 137, 1888–1898.

Yandell, B.S., 1997. Practical Data Analysis for Designed Experiments. Chapman & Hall.