Auditory Stream Disruption in Human Computer Interaction

Ronan O'Dea
Interactive Systems Research Group
Limerick Institute of Technology
ronan.odea@lit.ie

Rokaia Jedir
Interactive Systems Research Group
Limerick Institute of Technology
rokaia.jedir@lit.ie

Flaithri Neff
Interactive Systems Research Group
Limerick Institute of Technology
flaithri.neff@lit.ie

This paper presents results of the first phase of a study evaluating disruptive auditory features during a visual memory task in a HCI context. The disruption of primary tasks by secondary information streams is an important consideration in the design of multimodal interfaces. Although no statistical difference was exhibited in terms of task performance between the various auditory streams, a TLX questionnaire suggests that auditory streams comprising complex rhythmic features are perceived by users as being notably disruptive.

Multimodal Interfaces; Sonification; Attention Mechanisms; Working Memory; Changing State Hypothesis.

1. INTRODUCTION

Mobile, wearable and IoT devices now dominate the consumer computer market. This has resulted in interface designers contending with significantly reduced screen real-estate and needing to utilise other sensory modalities to present information to the user, such as auditory cues and haptic feedback. Such interfaces call for careful consideration of the idiosyncrasies inherent in the human perceptual system. Of particular importance is the possibility of introducing crossmodal interference in interfaces that concurrently present information both visually and aurally. Appropriate auditory cue design can avoid excessive disruption during a primary visual task while allowing for peripheral awareness of secondary information. Furthermore, user engagement and interpretation of data streams can be facilitated by sonification. Sonification is the use of non-speech audio to convey data.

As the capabilities of small-screen devices increase with technological advances, sonification may be utilised to relay complex system states and multivariate data (Hussain, Kang, & Lee, 2014; Lemmelä et al., 2008). However, if sonified streams are to effectively relay information to the user without negatively impacting on visual information streams, it is necessary to underpin the complex interplay between sensory streams in human attention mechanisms and working memory.

To this end, the authors present findings from a study investigating the potential interference elicited by two features of a secondary auditory stream on a concurrent visual working-memory task. The features investigated are rhythmic variations in the presentation of sine tones and the addition of reverberant spatial cues.

2. BACKGROUND

2.1. Design Considerations for the Presentation of Sonified Data.

Sonification is a prospective method for relaying data when using visually restrictive devices such as wearable, mobile and IoT devices. (Lemmelä et al., 2008) The integration of sonification in multimodal interface design requires consideration beyond simply choosing which sonification technique to use. Various techniques have been developed and reviewed in literature (Gaver, 1986; Blattner, Sumikawa, & Greenberg, 1989; Gavin, Jedir, & Neff, 2016). These techniques are generally categorised as auditory icons, earcons, parameter mapping, spearcons and hybrids (Neuhoff, 2011).

Depending on the task involved, consideration needs to be given to complex constituents intrinsic to how auditory information is processed in the human auditory system (Hussein et al., 2009). These include attention mechanisms; how sonic dimensions such as timbre are processed; the constraints associated with cognitive load; and how visual and auditory modalities interact on a perceptual level. These factors are important in scenarios requiring the user to maintain a certain level of task performance. Therefore, the inclusion of sonified data that functions to augment visualised information needs to be examined from these perspectives.

2.2. Working Memory Constraints and Attention Mechanisms

Working Memory (WM) is an established multicomponent model of human memory, proposed and refined through continuous research by Alan D. Baddeley and colleagues (Baddeley & Hitch, 1974). The model itself consists of three subsystems (Salamé & Baddeley, 1982; 1986; 1987; Baddeley, 2015). These are termed the phonological loop; the visuospatial sketchpad; and the episodic buffer. The three subsystems engage congruently to serve WM's 'central executive', which controls all subsystems and defines links to long-term memory. The study described by the authors relates directly to the function of the phonological loop, which determines a user's capacity to encode to-be-remembered (TBR) items in short-term memory (STM). The phonological loop provides a short-term store which is reinforced by a deliberate articulatory rehearsal process to account for a capacity limit. When an individual is presented with a visual-recall task, two processes are immediately engaged, allowing the order of presentation of visual items to be reinforced as they deliberately attempt to memorise them.

Auditory attention can be involuntarily elicited using intelligent auditory stimulus design (Jones, Miles, & Page, 1990; Jones & Macken, 1993). This shift in attention can disrupt the performance of visually demanding cognitive tasks (Alho et al., 1997; Escera et al., 1998; McCloy et al., 2017). The Irrelevant Sound Effect (ISE) is a theory associated with eliciting auditory attention, which warrants consideration in sonification design. The ISE is associated with a user's retention of visually presented information being disrupted by inconsequential background sound present in the environment (Jones & Macken, 1993). Researchers have consistently demonstrated that WM processes are reliably disrupted by irrelevant sound (Jones, Miles, & Page, 1990; Morris & Jones, 1990a; Morris & Jones, 1990b; Jones, Madden, & Miles, 1992; Jones & Macken, 1993; Parmentier & Beaman, 2015; Macken, Mosdell, & Jones, 1999). Furthermore, research by Salamé and Baddeley spawned interest in links to verbal STM mechanisms with the assumption that confusion occurs between phonologically similar material in both the auditory and visual domains (Salamé & Baddeley, 1982; Baddeley, 2015).

The authors in this paper assert that the identification of specific sonic properties that may be responsible for WM disruption is paramount in sonification design for multimodal interfaces, as it allows for the elimination of stimuli that draw on primary attention mechanisms during data presentation. The principle theory expressing this process is the Changing-State Hypothesis (Jones, Madden, & Miles, 1992).

2.3. The Changing-State Hypothesis

The Changing-State Hypothesis (CSH) examines the disruption of WM tasks by auditory stimuli. The CSH is based on the theory that dynamic elements within auditory stimuli are the main cause of WM disruption (Jones, Miles, & Page, 1990; Morris & Jones, 1990a; Morris & Jones, 1990b; Jones, Madden, & Miles, 1992; Jones & Macken, 1993). The theory suggests that this disruption is due to a conflict in memory whereby two similar processes occur, each involving seriation and each sharing a representational space (Jones & Macken, 1993). Parmentier and Beaman (2015) were the first to test the CSH as it relates to rhythmic change. The disruption of a serial recall task was examined using speech content with two types of variance: changes in temporal regularity; and changes in speech content. Findings consistently demonstrated that variations of the content, but not the rhythm of the irrelevant speech, reliably disrupted performance (Parmentier & Beaman, 2015).

In 1992, Jones suggested a new theory, 'interference at the attention level', which implied that the passing of sonic information to STM is reliant on auditory characteristics. Jones' model implies that a perceptual filter operates in response to the simultaneous presentation of visual and auditory stimuli, and that it fails to portray exact processes occurring in memory. Jones suggests in this model that speech content has privileged access to STM, based on the differences in disruption caused by non-speech signals in comparison to speech (Jones, Madden, & Miles, 1992). The effect is observable in a number of studies where disruption is compared between stream formats, with the effect being most prominently exhibited when contrasting speech content to broadband noise (Tremblay, Macken, & Jones, 2001).

This study examines the CSH as it relates to rhythmic and reverberant characteristics of non-speech auditory streams, as well as fixed versus dynamic reverberation in sine-tone auditory streams. Reverberation can be described as the time-domain response of a space to audio-spectral excitation, and imposes influence on other sonic dimensions, such as rhythm. A potential benefit of using reverberation in sonification applications is its potential to lessen the disruptive effects of seriation between sequential auditory objects while adding a sense of depth or source distance. However, excessive use of reverberation can obstruct the delivery of information (Lee, Shinn-Cunningham, 2008; Shinn-Cunningham 2000a; 2000b; 2004). Perham et al. (2007), examined the influence of reverberation levels applied to recordings of office background noise, which were presented as a distractive auditory stream during a serial recall task. Their findings indicated that increasing levels of

reverberation lessened the disruptive effect of a changing-state auditory stream but did not differ significantly to non-reverb office noise in terms of recall task performance. In a similar approach, the study presented in this paper applies reverberation to a sine-tone auditory stream to eliminate potential compounding influences found in complex sound sources.

3. USER STUDY DESIGN

The study was designed to allow participants to navigate fully through the presented conditions with no time limits applied. Participants were required to recall a sequence of seven visually-displayed digits (between 1 and 9), presented in random order. With the exception of the first condition, once the serial display was complete, auditory stimuli were presented via headphones during the serial rehearsal phase. No training or trial phase was implemented in order to eliminate potential conditioning through exposure or practice.

Each participant was given three attempts per condition. A list of instructions was also presented to the participants in order to reinforce the process. The participants were first presented with a quiet condition (no auditory stimulus during the serial rehearsal phase), followed by four irrelevant auditory streams presented in random order for the remaining conditions (see table 1).

Table 1: Auditory Conditions Presented:

Silence	Simple Rhythm	Complex rhythm	Simple Rhythm (Fixed Reverb)	Simple Rhythm (Changing Reverb)

3.1. Methodology
Two potential distractors were examined: rhythmic change in the auditory content, and reverberant timbral change. The effects of rhythmic change were examined using a stream of sine-tone pulses, presented in a sequence with random temporal intervals, creating a random and irregular rhythmic stream. To examine the effects of this irregular pattern, a simple stream of repeated sine-tone pulses in precisely equal temporal intervals was also presented to allow for a direct comparison. All sine-tone pulses were presented at a frequency of 391Hz, with an identical amplitude envelope applied to all items. To examine the potential effects of reverberant timbral change, a fixed, digitally synthesised impulse-response reverb with a 750ms reverberation time was applied to the simple, repetitive rhythmic sine-tone stream. For the final condition, a randomised reverberation time between 150ms and 2000ms was dynamically applied to the

aforementioned stream in real-time to investigate the effects of changing-state reverb.

Participants were specifically instructed to ignore the auditory stimuli to the best of their abilities while completing the recall tasks. A typical trial began with participants clicking on a 'start' toggle, and 7 digits visually presented 2 seconds (s) later. Each digit was presented for 1s. An additional 1s gap was presented after the last digit followed by the appearance of a submission box prompting the user to recall the sequence. Following 3 trial iterations, the participants were presented with a software replication of a Task Load Index (TLX) Questionnaire designed by NASA (Hart, 1986). The TLX is used to examine various categories of cognitive load and to gather perceptual feedback from participants. The categories examined were mental demand; physical demand; temporal demand; performance; effort; and frustration. These categories were summated to provide a total rating. In total, 15 trial iterations (3 per auditory condition) and 5 TLX questionnaires were completed by each participant.

52 undergraduate students voluntarily participated in the user study. The age of participants ranged from 18 to 49 years old. 82.35% of participants were male and 17.65% were female. The application used to present the stimuli, as well as the follow-on questionnaire, was developed using the Max 7™ visual programming language on a 2012 Apple® Macbook Pro® running macOs® Sierra. The study took place in a controlled lab environment. Participants were seated at individual workstations in the lab. The auditory stimuli were rendered using Csound, an audio DSL (domain-specific language). Reverb was applied to auditory stimuli using Space Designer® in Apple Logic Pro X®.

4. RESULTS

The disruptive effects on serial recall of the auditory conditions outlined in table 1 were compared. Silence; sine tone pulses presented in a repetitive rhythmic pattern; sine tone pulses presented arrhythmically; sine-tone pulses (repetitive rhythmic pattern) with a fixed reverb setting applied; and finally, sine-tone pulses (repetitive rhythmic pattern) with randomised changes in the reverb time across items. Serial-recall data was scored depending on the serial order entered by the participant. The user-study application was programmed to output a percentage error calculation, which compared each display list of TBR items (randomly generated) to the user's response that followed.

A mean error rate was derived from the percentage error rate for each condition, incorporating the percentage error from three iterations. Inter-quartile functions were applied to eliminate outliers in terms of serial recall performance. No outliers were

identified. Further analysis and calculation of the confidence interval (CI=95%) demonstrated no statistically significant difference in mean (M) error rates across presented conditions (see figure 1).

TLX rating data was then analysed to determine perceived cognitive load demand. Inter-quartile functions were calculated to eliminate outliers (see table 3). The physical demand category was the prominent source of outliers.
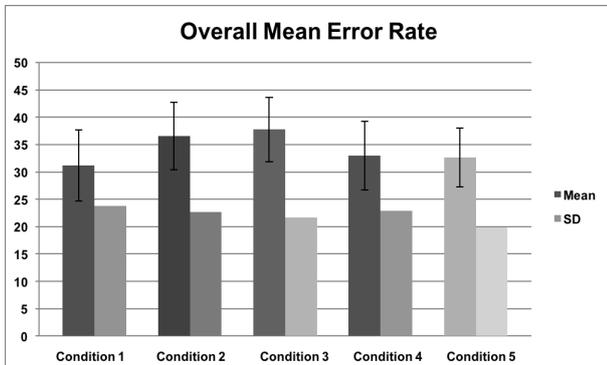

Fig 1: *Mean % Error Rates for conditions presented.*

Table 3: *Number of TLX Result Outliers Removed*

Condition	1	2	3	4	5
No. of Outliers	8	4	13	9	9

Further analysis of individual conditions across all TLX categories demonstrated statistical differences between condition 1 and 3. This indicates increased perceived cognitive load demand amongst users when comparing silence to sine tones presented in a rhythmically complex manner. The categories whereby statistical differences occurred were: mental demand, effort, frustration and total rating (see table 4).

Table 4: *Statistical Difference between Conditions 1 & 3.*

TLX Categories	Condition 1 (Silence)	Condition 3 (Complex Rhythm)
Mental Demand	M=65.31 95% CI [59.09,71.53]	M=80.65 95% CI [76.01,86.29]
Effort	M=59.1 95% CI [52.91,65.29]	M=72.44 95% CI [67.25,77.63]
Frustration	M=45.30 95% CI [38.55,52.05]	M=65.00 95% CI [57.69,72.31]
Total	M=58.75 95% CI [54.88,62.62]	M=70.85 95% CI [66.01,75.69]

Additionally, a statistical difference was evident between condition 1 (silence) and condition 4 (simple rhythm, fixed reverb). Frustration ratings for condition 1, M=45.30, CI 95% [38.55,52.05]; in comparison to frustration ratings for condition 4, M=59.9, CI 95% [52.77, 67.03].

5. DISCUSSION

The user study presented examines the potential disruptive effects of constant variation in rhythmic and reverberant characteristics in auditory streams on the performance of a cognitive task involving both WM and auditory attention mechanisms. Analyses support a null hypothesis over the hypothesis that irregular rhythms would disrupt serial recall to a greater extent than regular rhythms, and that dynamic reverb would disrupt performance to a greater extent than fixed reverb. Combined analysis from all conditions revealed that none of the auditory streams presented were statistically more disruptive to task performance compared to the silent condition. However, TLX data exhibit interesting results in relation to the perceived difficulty of the task across each auditory condition. In particular, differences were observed when comparing condition 1 (silence) with auditory condition 3 (arrhythmic sine tone pulse pattern). This indicates a difference in the perceived disruption caused by these two conditions, with an increased cognitive load under arrhythmic conditions (see table 4).

Additionally, in the frustration category there was a statistical difference between condition 1 and 4 (silence vs. simple rhythm with fixed reverb). No differences were found in task performance relating to both the fixed and changing-state reverberant conditions. Interestingly, the fixed reverb stream was perceived to add to the frustration in completing the task. This opposes the view that reverberant streams reduce the elicited perceived distraction.

Based on the results of the user study, the use of both rhythmic and reverberant dimensions in multimodal interfaces does not negatively impact on actual task performance when using simple auditory streams. However, TLX results indicate that users perceive a higher cognitive load, especially in relation to arrhythmic sonification, which has the potential to impact on the user experience.

6. CONCLUSION

No significant difference in participant performance of the serial recall task was found. TLX results indicate that participant cognitive load demand was influenced by the irregular rhythmic presentation of sine-tones in the following categories: mental demand, effort, frustration and total rating. Furthermore, the addition of reverb was perceived to add to frustration in comparison to silence. A future study will incorporate harmonically complex tones to evaluate potential changes to error rate results and perceived cognitive load.

7. REFERENCES

Alho, K., Escera, C., Díaz, R., Yago, E. and Serra, J.M. (1997) Effects of involuntary auditory attention on visual task performance and brain activity. *Neuroreport*, Vol. 8*(15), pp. 3233-3237.

Baddeley, A. (2015) Working Memory. *Memory. A. Baddeley, Michael W. Eysenck and Mickael C. Anderson (Eds.), Memory*, Ch. 3, pp. 67-98.

Baddeley, A. D. and Hitch, G. (1974) Working memory. In *Psychology of learning and motivation Vol. 8*, pp. 47-89.

Beaman, C. P. and Holt, N. J. (2007) Reverberant auditory environments: the effects of multiple echoes on distraction by 'irrelevant' speech. *Applied Cognitive Psychology, Vol. 21*(8), pp. 1077-1090.

Blattner, M. M., Sumikawa, D. A. and Greenberg, R.M. (1989) Earcons and icons: Their structure and common design principles. *Human–Computer Interaction, Vol. 4*(1), pp. 11-44.

Bregman. A. S. (1990) Schema-Based Segregation and Integration. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, pp. 411 - 415.

Escera, C., Alho, K., Winkler, I. and Näätänen, R. (1998) Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of cognitive neuroscience, Vol. 10*(5), pp. 590-604.

Gaver, W. W. (1986) Auditory icons: Using sound in computer interfaces. *Human computer interaction*, *Vol. 2(2)*, pp. 167-177.

Gavin, M., Jedir, R. and Neff, F. (2016) Sonification playback rates during matching tasks of visualised and sonified EEG data. *Proceedings of the International Audio Systems Symposium, University of York.*

Hanley, J. R. and Broadbent, C. (1987) The effect of unattended speech on serial recall following auditory presentation. In the *British Journal of Psychology, Vol. 78(3)*, pp. 287-297.

Hart, S. G. (1986) NASA Task Load Index (TLX). Vol. 1.0; *Paper and pencil package.*

Hedrick, M. S. and Younger, M.S. (2007) Perceptual weighting of stop consonant cues by normal and impaired listeners in reverberation versus noise. *Journal of Speech, Language, and Hearing Research, Vol. 50*(2), pp. 254-269.

Helfer, K. S. (1994) Binaural cues and consonant perception in reverberation and noise. *Journal of Speech, Language, and Hearing Research, Vol. 37*(2), pp. 429-438.

Hussain, S., Kang, B. H. and Lee, S. (2014) A wearable device-based personalized big data analysis model. Proceedings of the *International Conference on Ubiquitous Computing and Ambient Intelligence*, pp. 236-242.

Hussein, K., Tilevich, E., Bukvic, I. I. and Kim, S., (2009) Sonification design guidelines to enhance program comprehension. In *Program Comprehension, Proceedings of the 17th ICPC,* pp. 120-129.

Jones, D., Madden, C. and Miles, C. (1992) Privileged access by irrelevant speech to short-term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology*, *Vol. 44*(4), pp. 645-669.

Jones, D. M. and Macken, W. J. (1993) Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *Vol. 19*(2), pp. 369.

Jones, D. M., Miles, C. and Page, J. (1990) Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, *Vol. 4*(2), pp. 89-108.

Lee, A. K. and Shinn-Cunningham, B.G. (2008) Effects of reverberant spatial cues on attention-dependent object formation. *Journal of the Association for Research in Otolaryngology*, *Vol. 9 (*1), pp. 150-160.

Lemmelä, S., Vetek, A., Mäkelä, K. and Trendafilov, D. (2008) October. Designing and evaluating multimodal interaction for mobile contexts. In *Proceedings of the 10th international conference on Multimodal interfaces*, pp. 265-272.

Macken, W. J., Mosdell, N. and Jones, D. M. (1999) Explaining the irrelevant-sound effect: Temporal distinctiveness or changing state?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *Vol. 25*(3), pp. 810.

McCloy, D. R., Lau, B. K., Larson, E., Pratt, K. A. and Lee, A.K. (2017) Pupillometry shows the effort of auditory attention switching. *The Journal of the Acoustical Society of America*, *Vol. 41*(4), pp. 2440-245.

Moore, B. C. (2012) Auditory Pattern and Object Identification in *An Introduction to the Psychology of Hearing*, pp. 283-301.

Morris, N. and Jones, D. M. (1990) Habituation to irrelevant speech: Effects on a visual short-term memory task. *Perception & Psychophysics*, *Vol. 47*(3), pp. 291-297.

Morris, N. and Jones, D. M. (1990) Memory updating in working memory: The role of the central

executive. *British journal of psychology*, *Vol. 81*(2), pp. 111-121.

Neuhoff, J. G. ed. (2011) *The Sonification Handbook*. Logos Verlag, Berlin.

Parmentier, F. B. and Beaman, C. P. (2015) Contrasting effects of changing rhythm and content on auditory distraction in immediate memory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *Vol. 69*(1), pp. 28.

Parseihian, G., Katz, B. F. and Conan, S. (2012) Sound effect metaphors for near field distance sonification.

Perham, N., Banbury, S. and Jones, D.M. (2007) Do realistic reverberation levels reduce auditory distraction? In *Applied Cognitive Psychology*, *Vol. 21*(7), pp. 839-847.

Salame, P. and Baddeley, A. (1982) Disruption of short-term memory by unattended speech: Implications for the structure of working memory. In the *Journal of verbal learning and verbal behaviour*, *Vol. 21*(2), pp. 150-164.

Salamé, P. and Baddeley, A. (1986) Phonological factors in STM: Similarity and the unattended speech effect. In the *Bulletin of the Psychonomic Society*, *Vol. 24*(4), pp. 263-265.

Salamé, P. and Baddeley, A. (1987) Noise, unattended speech and short-term memory. *Ergonomics*, *Vol. 30*(8), pp. 1185-1194.

Shinn-Cunningham, B. G. (2000) Learning reverberation: Considerations for spatial auditory displays. In *Proceedings of the 2000 International Conference on Auditory Display*

Shinn-Cunningham, B. G. (2000) Distance cues for virtual auditory space. In *Proceedings of the IEEE-PCM, Vol. 2000*, pp. 227-230.

Shinn-Cunningham, B. G. (2002) Speech intelligibility, spatial unmasking, and realism in reverberant spatial auditory displays. In *Proceedings of International Conference of Auditory Display.*

Shinn-Cunningham, B. G. (2001) Creating three dimensions in virtual auditory displays. *Usability evaluation and interface design: Cognitive engineering, intelligent agents and virtual reality*, pp. 604-608.

Shinn-Cunningham, B. G. (2004) The perceptual consequences of creating a realistic, reverberant 3-D audio display. In the *Proceedings of the International Congress on Acoustics, Kyoto, Japan*, pp. 4-9.

Shinn-Cunningham, B. G., Santarelli, S. and Kopco, N. (1999) Distance perception of nearby sources in reverberant and anechoic listening conditions: Binaural vs. monaural cues. In *Assoc Res Otolaryn. Meeting*. Vol. 23.

Stothart, C., Mitchum, A. and Yehnert, C (2015) The attentional cost of receiving a cell phone notification. *Journal of experimental psychology: human perception and performance*, *Vol. 41*(4), pp. 893

Tremblay, S., Macken, W.J. and Jones, D. M. (2001) The impact of broadband noise on serial memory: Changes in band-pass frequency increase disruption. *Memory*, *Vol. 9*(4-6), pp. 323-331.