

The Design of Audio Mixing Software Displays to Support Critical Listening

Joshua Mycroft

Submitted in partial fulfilment of the requirements of the Degree of
Doctor of Philosophy

School of Electronic Engineering and Computer Science Queen Mary,
University of London

July 2017

Statement of Originality

I, Joshua Mycroft confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below, and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date:

Abstract

The mixing desk metaphor found in Digital Audio Workstations (DAW) is built upon a specialised and technical knowledge of signal flow and audio engineering. However, since their inception the DAW has gained a far wider and less technically specialised user-base. Furthermore, the limited screen space of laptop and tablet computers, combined with potentially limitless tracks in current DAWs has resulted in the need for complex interface navigation during mixing which may inhibit a fluid and intuitive approach to mixing.

The research outlined in this thesis explores novel designs for Graphical User Interfaces (GUIs) for mixing, which acknowledge the changing role of the user, the limited space of tablet and mobile computers screens and the limitations of human perception during cross modal activities (aural and visual). The author designs and conducts several experiments using non-expert participants drawn from several music technology courses, to assess and quantify the extent to which current DAW designs might influence mixing workflow, aiming our research especially at beginner and non-expert users.

The results of our studies suggest that GUIs which load visual working memory, or force the user to mentally integrate visual information across the interface, can reduce the ability to hear subtle simultaneous changes to the audio. We use the analysis of these experiments to propose novel GUI designs that are better suited to human cross-modal perceptual limitations and which take into account the specific challenges and opportunities afforded by screen-based audio mixers. By so doing, we aim to support the user in achieving a more fluid and focused interaction while mixing, where the visual feedback supports and enhances the primary goal of attending to and modifying the audio content of the mix. In turn, it is hoped this will facilitate the artistic and creative approaches required by music computer users.

Table of Contents

Abstract	3
Table of Contents	4
List of Figures	9
List of Tables	11
List of Abbreviations	12
1. Chapter 113	
1.1 Overview of Audio Mixing.....	13
1.2 Motivation	14
1.3 Research Questions.....	17
1.4 Thesis Structure	17
1.5 Publications	19
1.5.1 Peer-reviewed Conference Papers.....	19
1.5.2 Conference Posters and Demos	20
2. Chapter 2	
2.1 Development of Multi-Track Mixing.....	21
2.1.1 Development of Mixing Desks.....	22
2.1.2 Transfer from Hardware to Software	23
2.2 The CS Metaphor in DAWs.....	26
2.2.1 Affordances.....	26
2.2.2 Visual Feedback.....	28
2.2.3 Interface Navigation.....	31
2.2.3.1 Window Switching.....	32
2.2.3.2 Zooming.....	32
2.2.3.3 Scrolling.....	34
2.2.3.4 Overviews.....	35
2.3 State of the Art	37
2.3.1 Non-technical Interfaces.....	37
2.3.2 Interaction Techniques.....	39
2.3.2.1 Gestural Interaction.....	39
2.3.2.2 Tangible Interaction	40

2.3.2.3 Multi-touch Interaction.....	40
2.3.3 Stage Metaphor Mixers	42
2.3.3.1 Limitations of the SM Design.....	43
2.3.4 Evaluation of Novel Designs	44
2.4 Summary	45

3. Chapter 3

3.1 Thesis Experiments.....	47
3.1.1 Rational for Methodological Framework.....	47
3.1.2 Summary of Experiments	50
3.2 Experimental Design	51
3.2.1 Experimental Considerations.....	51
3.2.2 Experimental Approach.....	53
3.2.3 Quantitative Data Analysis	53
3.2.4 Qualitative Data Analysis.....	54
3.2.4.1 Questionnaires.....	54
3.2.4.2 Interviews.....	54
3.3 Participants.....	55
3.3.1 Recruitment and Ethical Considerations.....	55
3.4 Specific Experiment Details	56
3.4.1 Experiment One.....	56
3.4.1.1 Experimental Methods.....	56
3.4.1.2 Experiment Tasks	56
3.4.1.3 Data Analysis Methods	56
3.4.2 Experiment Two	57
3.4.2.1 Experimental Methods.....	57
3.4.2.2 Experiment Tasks	57
3.4.2.3 Data Analysis Methods	57
3.4.3 Experiment Three.....	58
3.4.3.1 Experimental Methods.....	58
3.4.3.2 Experiment Tasks	58
3.4.3.3 Data Analysis Methods	58
3.4.4 Experiment Four.....	58
3.4.4.1 Experimental Methods.....	58
3.4.4.2 Experimental Tasks	59
3.4.4.3 Data Analysis Methods	59

3.4.5 Experiment Five.....	60
3.4.5.1 Experimental Methods.....	60
3.4.5.2 Experimental Tasks	60
3.4.5.3 Data Analysis Methods	61
3.4.6 Experiment Six	61
3.4.6.1 Experimental methods	61
3.4.6.2 Experimental Tasks	61
3.4.6.3 Data Analysis Methods	62
3.4.7 Experiment Seven.....	62
3.4.7.1 Experimental methods	62
3.4.7.2 Experimental Tasks	62
3.4.7.3 Data Analysis Methods	63
3.5 Summary	63
4. Chapter 4	
4.1 Study Design	65
4.1.1 Participants.....	65
4.1.2 Procedure	66
4.2 Analysis and Results.....	67
4.3 Discussion	68
4.4 Conclusion.....	69
4.5. Progression to Subsequent Study	70
5. Chapter 5	
5.1. Study Design	73
5.1.1 Participants.....	73
5.1.2 Procedure	73
5.2. Analysis	76
5.3. Results	77
5.4 Discussion.....	78
5.5 Conclusion.....	80
5.6 Summary	81
5.7 Progression to Subsequent Study	81
6. Chapter 6	
6.1 Study Design	85
6.1.1 Participants.....	85

6.1.2 Listening Task.....	85
6.1.3 Visual Task.....	86
6.2 Analysis and Results.....	87
6.3 Discussion.....	88
6.4 Summary.....	91
6.5 Progression to Subsequent Study.....	92
7. Chapter 7	
7.1. Study Design.....	94
7.1.1 Participants.....	94
7.1.2 Visual Task.....	95
7.1.3 Listening Task.....	96
7.2 Analysis and Results.....	96
7.3 Summary.....	100
7.4 Progression to subsequent study.....	101
8. Chapter 8	
8.1 Study Design.....	103
8.1.1 Participants.....	103
8.1.2 Visual Task.....	104
8.1.3 Listening Task.....	107
8.1.4 Study Procedure.....	107
8.2 Analysis and Results.....	108
8.2.1 Visual Task Analysis.....	108
8.2.2 Aural Task Analysis.....	109
8.2.3 Survey Results.....	109
8.3 Discussion.....	114
8.4 Summary.....	115
8.5 Progression to subsequent study.....	115
9. Chapter 9	
9.1 Study A: Representing an Additional Mix Parameter.....	117
9.1.1 Participants.....	117
9.1.2 Study Design.....	118
9.1.3 Procedure.....	121
9.2 Analysis and Results.....	121
9.3 Study B: Adding a Further Mix Parameter.....	123

9.3.1 Participants.....	123
9.3.2 Study Design.....	124
9.3.3 Procedure	124
9.4 Analysis and Results.....	125
9.5 Discussion	127
9.6 Summary	128
9.7 Progression to Subsequent Study	129

10. Chapter 10

10.1 Study Design.....	131
10.1.1 Participants.....	131
10.1.2 Interface Designs	131
10.2 Procedure	134
10.2.1 Pre-test Screening	134
10.2.2 Test Procedure	135
10.2.3 Mixing Tasks.....	137
10.3 Analysis and Results.....	138
10.3.1 Speed to Complete Tasks	139
10.3.2 Amount of Correctly Completed Tasks.....	140
10.3.3 Accuracy of Mixing Tasks	141
10.3.4 Analysis of Participant Comments.....	143
10.3.4.1 CS Mixer.....	143
10.3.4.2 Hybrid.....	143
10.3.4.3 Stage Mixer	144
10.4 Discussion	145
10.5 Conclusion.....	147
10.6 Summary	148

11. Chapter 11

11.1 Research Summary.....	151
11.2 Summary of Heuristics Arising from this Research	156
11.3 How Research Questions are Answered	162
11.4 Contributions.....	164
11.5 Directions for Future Research	165
11.6 Application of research to other aspects of audio mixing.....	169
11.6.1 Live Sound mixing;.....	169
11.6.2 multi-channel and surround mixing;.....	170

11.6.2 Frequency masking;.....	171
11.6.3 Dealing with different genres.....	174
Bibliography.....	175

List of Figures

Introduction

Figure 1:1: Mixing desk channel strips.....	15
Figure 1:2: Cubase 9 mixer window.....	16

Background

Figure 2:1: Reaper DAW mixer screen.	23
Figure 2:2: Reason rack view.	27
Figure 2:3: Cubase 9 mixer window.....	28
Figure 2:4: Logic Pro 9.	33
Figure 2:5: Overview and detail in effects page in Ableton.	36
Figure 2:6: Overview and detail of arrange page in Ableton	36
Figure 2:7: Pro Tools Universe Window..	36
Figure 2:8: Serato’s use of coloured waveforms	37
Figure 2:9: WaveMachine Labs Auria.....	41
Figure 2:10: Gibson’s Stage Metaphor	43
Figure 3:1: The Pro Tools channel EQ gain scale.....	49

Experiments and Analyses

Figure 4:1: Pro Tools equaliser, version 5.1.	66
Figure 4:2: Pro Tools equaliser, version 6.9 onwards	66
Figure 4:3: Study One experimental interfaces	68
Figure 4:4: Mean errors and p-values for study two interfaces.	70
Figure 5:1: Experimental Control interface, study two	75
Figure 5:2: Experimental interface 1, study two.....	75
Figure 5:3: Experimental interface 2, study two	75
Figure 5:4: Experimental interface 3, study two	76
Figure 5:5: Times for completion per interfaces for guitar file.	79
Figure 5:6: Times for completion per interface for string file.	80

Figure 5:7: Times for completion per interface for tambourine file.	80
Figure 6:1: The Pro Tools Channel EQ, use of colour.	84
Figure 6:2: The use of colour to represent frequency in Serato DJ Software.	84
Figure 6:3: Experimental interface designs used in study three.	86
Figure 6:4: Mean time to correctly identify audio changes per interface type.	88
Figure 6:5: Mean number of channels compared per interface design.	89
Figure 7:1: Experimental interface designs used in study four.	97
Figure 7:2: Number of correctly answered visual questions per interface type.	98
Figure 7:3: Number of correctly answered panning positions per interface type.	100
Figure 8:1: Experimental interafce designs used in study five.	105
Figure 8:2: Mixer designs with DQ functionality.	106
Figure 8:3: Visual searches successfully completed.	110
Figure 8:4: Questionnaire; results from the question 1.	112
Figure 8:5: Questionnaire; results from the question 2.	112
Figure 8:6: Questionnaire; results from the question 3.	113
Figure 8:7: Questionnaire; results from the question 4.	113
Figure 9:1: Experimental interfaces used in study six, part a.	119
Figure 9:2: Colour gradients used in the interface designs.	121
Figure 9:3: Experimental interfaces used in study six, part b.	126
Figure 9:4: The visual search time for interface designs.	127
Figure 10:1: Experimental interfaces designs used in study seven.	132
Figure 10:2: Size differences between channels.	133
Figure 10:3: Highlighting instruments in the display.	135
Figure 10:4: Time taken to correctly complete mixing tasks per interface.	140
Figure 10:5: Analysis of the accuracy of panning the tracks in question 4.	142
Figure 10:6: Analysis of the accuracy of matching the bass amounts.	142
Figure 10:7: Number of participants to identify the LP filter per interface type.	146

Summary of Contributions, Discussion and Final Conclusions

Figure 11:1. Microsoft Word right click functionality.	158
Figure 11:2. Logic X editing menu.	159
Figure 11:3. Interactive map of crime in London.	167
Figure 11:4. Zoopla website.	168
Figure 11:5. DQ filters to show or hide tracks in a channel strip mixer.	168
Figure 11:6. Surround sound panning in Pro Tools and Logic Pro.	172
Figure 11:7. Representing frequency bands in a novel user unterface object.	173

Figure 11:8. Relative size of frequency bands in a novel user interface object173

List of Tables

Studies and Analyses

Table 2:1: Comparison of features in commercial DAWs (as of May 2018).	25
Table 2:2: Comparison of navigation management in commercial DAWs.....	30
Table 4:1: Population Comparisons.....	65
Table 5:1: The <i>P</i> values for time difference between interface types.....	77
Table 5:2: Time difference for task completion between Control and interfaces.....	78
Table 6:1:Percentage of channels matched incorrectly per UI type.	88
Table 7:1: Questions asked for each occurrence of the three interface designs.....	96
Table 7:2: The mean and standard deviation for visual search questions	98
Table 7:3: The mean and standard deviation for audio panning.....	98
Table 8:1: Visual search questions asked per interface design.....	108
Table 8:2: The Percentage of correctly identified audio file attenuations per interface type.....	111
Table 8:3: Results of the Z-test for dependent groups analysis.....	111
Table 9:1: Error rates (%) for each design.....	122
Table 9:2: Error rates for both design at different value differences.. ..	127
Table 10:1: The list of instruments used in the multi-track recording.....	136
Table 10:2: Mixing tasks given during the experiment.	138
Table 10:3: Mixing tasks. Percentage of participants correctly completing the mixing task (per interface type and parameter amount).	141
Table 10:4: Percentage of participants who successfully identified the LP filter per interface type and parameter amount.....	141
Table 10:5: Percentage of participants who successfully completed both tasks (correct mixing and hearing LP filter).....	141

List of Abbreviations

CI	Confidence Intervals
CLT	Cognitive Load Theory
CS	Channel Strip
DAW	Digital Audio Workstation
dB	Decibels
DQ	Dynamic Query
EQ	Equaliser/Equalisation
GIS	Geographical Information Systems
GUI	Graphical User Interface
HCI	Human Computer Interaction
Hz	Hertz
LP	Low Pass
O+D	Overview and Detail
P	Probability
SM	Stage Metaphor
STM	Short Term Memory
TUI	Tangible User Interface
UI	User Interface
WM	Working Memory
SD	Standard Deviation
2D	Two dimensional
3D	Three dimensional

Chapter 1

Introduction.

1.1 Overview of Audio Mixing

Audio mixing, at the highest level, can be understood as engaging with the various sounds within a performance or composition as a blended entity, a meta-performance consisting of all the individual parts (Moylan, 1992). The act of the mixing engineer is to engage in shaping this level of perspective and establishing qualities and relationships between them (Moylan, 2017). The mix engineer effectively delivers the song by ‘shaping its dimensions to match, complement or enhance the character of the song, its message and expression’ (ibid, p.50). This process is goal-oriented, with the aim being a desired change in the emotional effect of a sound (Sauer et al, 2013). The primary elements of these modifications are timbral balance; the distribution and density of pitch/ frequency information and the combination of all the pitch densities of all the sounds within the mix, volume; the adjustments of dynamics between the various sounds so that all loudness levels coalesce into a single sensation of loudness, and panning; the placement of sounds between the speaker to make sense of a recording’s spatiality (Moylan, 2017., Walther-Hansen, 2017).

In order to undertake these modifications, a mix engineer will use an audio mixing console. While the actual implementations vary between different manufacturers and designs, the common element of the audio mixing console is the channel strip: vertical strips which group together processing and routing options. Each of the various channel strips (or channels) on a mixer has identical controls to one another and employs pan pots (rotary dials) and faders to allow access to key mixing parameters. These typically include equalisation (boosting or cutting the amount of treble, bass or mid-range frequencies), level (the volume of the sound), pan position (the relative position between the two speakers) and effects, such as reverb, chorus, delay etc. (these are set using the auxiliary dials, which send a user-defined amount of the instrument or sound to an effect, thereby allowing the user to add as much or as little as required).

In mixing, each sound or instrument within the performance or recording is referred to as a track (e.g. bass track, vocal track, snare drum track). These tracks can be assigned to a channel, allowing subsequent adjustments by the mix engineer to shape, refine and

enhance the overall blend between the sounds and instruments. (figure1.1). Tracks can be assigned to channels on a one-to-one basis, or multiple tracks can be assigned to the same channel, allowing groups of instruments, e.g. multiple backing vocals, to be adjusted together, rather than having to change the characteristics of each one separately. The mixing console therefore provides the mix engineer with the necessary tools to control and adjust the characteristics of the various audio signals generated by musical instruments, microphones, and the like, within the performance or recording.

Since the 1990s, due to increases in computer processing speed, it has become increasingly common to undertake the entire mixing process using virtual audio mixers, rather than a physical mixing desk. These digital audio mixers, referred to as Digital Audio Workstations (DAWs), allow producers to undertake mixing tasks using a computer screen. DAW based mixing now accounts for most mix productions; it is estimated that by 2007, between 70 and 80 percent of all pop music (and nearly 100 percent of all Hip-Hop, R&B, and Dance music) was mixed using DAWs, without any recourse to physical mixing desks (Milner, 2009, p.338). However, despite this rapid and fundamental change in the interface from the physical to the virtual, the visual representation used in DAWs continues to be based heavily on the real-world counterpart of the physical mixing desk. Indeed, DAWs such as Pro Tools, Logic and Cubase have transferred the CS metaphor to Graphical User Interfaces (GUI). Though DAWs have extended the metaphor to incorporate ‘more virtual sliders, knobs and waveform and track displays’ (Golkhe et al, 2010, p.1), fundamentally, they continue to employ skeuomorphic GUIs modelled on analogue mixer designs (figure 1.2).

1.2 Motivation

While the use of the physical mixing desk metaphor in DAWs is familiar and provides transferable knowledge from the hardware, there may be a risk that the structural metaphors and affordances do not always translate well into virtual representations (Mihnkern 1997). For example, while a dial on the physical desk suggests and supports the perceived affordance of turning, this action is not supported when using a mouse, potentially leading to errors and breaks in the user’s engagement with the task (McGrenere & Ho 2000). Faders are a further example; their inclusion on physical mixing consoles were an ergonomic development to allow engineers to control several



Figure 1:1: Mixing desk channel strips. Each channel typically contains controls for (top to bottom), microphone input level, equalisation, effects (auxiliaries), panning position and volume level. Common desk sizes include 8, 16, 24 and 32 channels.

faders at one time. However, as Htalky et al (2009) highlight, there may be a mismatch with their digital representations in terms of limited screen size and resolution, with virtual faders in many DAWs being ill-suited to refined adjustments (ibid). Furthermore, when using a mouse, only one channel can be changed at one time, failing to reproduce the performative nature that faders support on a physical desk.

There are also perceptual issues to consider; while physical mixing desks are often limited to 24 channels, DAWs offer limitless tracks for an audio mix, giving rise to situations where there may be too many channels to fit onto one screen. The subsequent screen management required to access the mixer across multiple windows can place high cognitive load on short-term and working memory (Shneiderman and Bederson, 2005) and overload the limited capacity of the visual mechanism (Rensink, 2012).



Figure 1:2: Cubase 9 mixer window. The interface employs a skeuomorphic GUI modelled on analogue mixer designs.

Moreover, the need to search through several windows of mix information may inhibit the engagement and ‘flow’ of the mixing process and impede the user’s ability to quickly respond to the programme material (Szalva, 2009, Gelineck et al 2013a), leading to a situation where some users find it “impossible to navigate those interfaces [Logic, Pro Tools] while also trying to be artistic” (Crane, 2010, p.12). This situation is further compounded when using tablet and laptop computers to mix, where screen space is at a premium.

Thirdly, since the late 1990s, there has been a shift in the demographic of DAW users from traditional studio environments to home studio-based recording and production (Leyshon 2009). This change may require interface designers to broaden the definition of a DAW user beyond a functional professional, performing in a commercial environment. Indeed, some users of music production software may have never used a physical mixing desk (Battino and Richards 2005). For such users, an alternative metaphor, which can be used alongside the more segmented mixing desk metaphor, may help maintain the user’s creative engagement with the mixing process and provide different contexts of use (Gelineck et al 2013b).

By acknowledging the perceptual limits of the user, the limited screen-sizes of laptop and tablet computers, the large track and effects count of contemporary mixes and the changing profile of users, the work described in this thesis aims to develop heuristics that inform the design of audio mixing interfaces. It focuses on the main functions and tools within audio mixing including level setting, panning, equalisation and effects and

is aimed especially at non-expert users who may lack the resources or technical skills to handle a mixing console or professional DAW.

However, due to the continuing ubiquity of the CS design, the heuristics and research reported here are not intended to replace the CS metaphor per se, but rather to provide an additional conception of the mixing interface to be used in conjunction with it

1.3 Research Questions

The main research area investigated in this thesis is concerned with how best to present the mixing tools and information within DAW mixers so that the sonic content of the mix is the user's primary focus and visual information is used to support and enhance this primary function. It should be noted, however, that we acknowledged that visual feedback of a mix is a valuable resource, and our research is focussed, not on removing or dismissing it, but rather on finding ways of designing it which support and enhance the audio mixing process. Toward this end three specific questions are addressed:

1. To what extent does visual perceptual load interfere with or support the processing of auditory information?
2. Does the specific design and layout of the visual information have an impact on aural acuity or speed of visual tasks?
3. How can visual mix information display be managed more effectively?

1.4 Thesis Structure

Chapter 1 Introduction. We begin by outlining the objectives and motivations of the research and identifying the key research questions that we will investigate in the course of the studies presented in this thesis.

Chapter 2 Background: In this chapter, we review four key areas fundamental to conducting research on audio mixing interfaces. Firstly, we briefly discuss the development of the audio mixing console in general, and the CS design in particular. Secondly, we consider the translation of the CS strip metaphor to DAWs. Thirdly, we discuss perceptual implications of screen-based mixing, specifically focussing on cross-modal perception. The chapter ends with a discussion of interface navigation, both in

general and in relation to mixing and includes a review of scrolling, window switching, zooming, Overview + Detail and Dynamic Query filters.

Chapter 3 Strategy to Investigate Research Questions: Here, we outline the thesis research strategy discussing the design of the experiments, the rationale for the research methodology and data analysis, and the profile and recruitment of the participants.

Chapter 4 Visual Feedback and Critical Listening: We examine in detail the first research question, namely the extent to which visual perceptual load impedes or supports processing of auditory information. We present the results from our first study which investigates to what extent the amount of visual information in a GUI affects aural acuity.

Chapter 5 Visual Load in Mixing GUIs: In this chapter, we continue to examine the first research question, and we consider how navigating the interface to find visual mix information affects visual search times and simultaneous critical listening.

Chapter 6 User Interface Object Design: Here, we examine the second research question; to what extent does the design of the visual information have an impact on concurrent aural acuity and visual search? To do this, we examine the design of User Interface (UI) objects. This takes the form of a study that examines the UI objects commonly found in CS mixers (dials and faders, numbers and colours) to see whether different levels of visual perceptual demand inherent in these designs influences the simultaneous perception of auditory stimuli.

Chapter 7 Managing Visual Search in the GUI: We present findings from our fourth study, which investigates how the GUI can be presented to manage visual search more effectively. This includes the use of overviews to manage the amount of visual feedback in the interface. The study explores whether, by removing interface navigation, we affect visual search times and critical listening skills.

Chapter 8 Filtering the GUI: We incorporate Dynamic Query filters (sliders, buttons and other filters) into the GUI. By so doing, we aim to quantify the extent to which filtering of the GUI may facilitate improvement in search times and/ or concurrent critical listening. We also assess whether their inclusion may distract users from the mixing workflow.

Chapter 9 Representing Multivariate Mix Data: In this chapter, we consider how in other areas of data visualisation (such as medial visualisations, geo-spatial and cartographic displays), multiple parameters are often integrated into one graphical object to leverage perceptual skills and help users discern and interpret patterns within data. We compare interfaces which integrate UI objects into one graphical object against interfaces which have a one-to-one mapping of controls. This is then assessed in terms of accuracy of visual search and interpretation of visual information in the GUI.

Chapter 10 Mixing Interface Prototype: Here we present our final study. We build on the results from the previous experiments, and design a mixing interface incorporating several heuristics from previous studies, including the design of the UI objects, the design of the GUI and the management of the visual information within it. We test this against a channel strip mixer, and assess several mixing workflow tasks using both designs. We use this experiment to test, develop and refine our heuristics.

Chapter 11 Summary of Contributions, Discussion and Final Conclusions: In this chapter, we state the achievements of the thesis and outline the contributions made. We summarise the findings from the studies and present a list of heuristics for the design of mixing GUIs. Finally, we consider future directions, and how we the research undertaken in this PhD could be further developed.

1.5 Publications

Portions of the research from this thesis were published and presented at international conferences:

1.5.1 Peer-reviewed Conference Papers

- Mycroft, J., Stockman, T., and Reiss, J.D. (2018). ‘A Prototype Mixer to Improve Cross-Modal Attention During Audio Mixing’ Accepted for presentation at *Audio Mostly Conference* 2018, Wrexham, U.K.
- Mycroft, J., Stockman, T., and Reiss, J.D. (2016). ‘Visually Representing and Interpreting Multivariate Data for Audio Mixing’ In: *Proceedings of Sound and Music Computing*, August 2016, Hamburg, Germany.
- Mycroft, J., Stockman, T., and Reiss, J.D. (2016). ‘Visual Information Search in Digital Audio Workstations’. In: *Proceedings of Audio Engineering Society*, June 2016. Paris, France.

- Mycroft, J., Stockman, T., and Reiss, J.D. (2015). 'Audio Mixing Displays: The Influence of Overviews on Information Search and Critical Listening'. In: *Proceedings of Computer Music Multidisciplinary Research*, June 2015, Plymouth, U.K.
- Mycroft, J., Reiss J.D., and Stockman, T. (2015). 'The Effect of Differing User Interface Presentation Styles on Audio Mixing'. In: *Proceedings of the International Conference on the Multimodal Experience of Music*, March 2015. Sheffield, U.K.
- Mycroft, J., Reiss J.D., and Stockman, T. (2013) 'The Influence of Graphical User Interface Design on Critical Listening Skills'. In: *Proceedings of the Sound and Music Computing*, June 2013. Stockholm, Sweden.
- Mycroft, J and Paterson, P. (2011). 'Activity flow in music equalization: the cognitive and creative implications of interface design'. In: *Proceedings of the Audio Engineering Society Convention*, May 2011, London, U.K.
-

1.5.2 Conference Posters and Demos

Portions of the research from this thesis were presented as posters and demonstrations at the following conferences:

- DMRN +6. "Supporting fluid interaction in audio mixing software", J. Mycroft, T. Stockman, J.D. Reiss and A. J. Wall. December 2011.
- DMRN+7. "Toward simplifying the mixing work-space: an outline of possible interface designs", Josh Mycroft, Tony Stockman and Josh Reiss. December 2012.

Chapter 2

Background.

In section 2.1 we discuss the development of the Channel Strip (CS) mixer. We explore this in the context of historical and technological developments. We also briefly discuss the adoption of the CS design in software mixers and the reasons for the design being used in the virtual realm. In section 2.2 we discuss the practical implications of mixing with the CS metaphor in DAWs by highlighting areas where it may impact on workflow. This is discussed in terms of perceived affordances, screen display sizes, screen navigation and perceptual and attentional limitations of users. In section 2.3 we briefly review the state-of-the-art in audio mixing GUIs, and introduce some alternative metaphors to the CS paradigm. We discuss the advantages and disadvantages of the designs in terms of conceptual models, user knowledge, mixing workflow, and interface design.

2.1 Development of Multi-Track Mixing

In the 1940s, when recording with tape was still in its infancy and many studios recorded directly to acetate disc, all the audio sources were recorded simultaneously onto one track. Since there was just a single track, there was no conceivable way for the producer to adjust the individual levels of the recorded instruments after the initial recording. Direct recording relied on microphone placement, equalisation, acoustics and mixing before recording (Channan 1995). If the mix wasn't satisfactory, or if a musician made a mistake, the music had to be performed again until the desired balance or performance was achieved.

In the early 1950s, this limitation was overcome when American composer and technician Les Paul commissioned the Ampex Corporation to build a custom tape recorder in which eight separate audio tracks could be simultaneously recorded onto one-inch audio tape. Its introduction created the possibility for an incremental approach to mixing. The producer could now hear parts of the production in isolation, repeating and correcting for technical, musical, or creative reasons until satisfied. The development of multitrack recording was a major progression in music recording and can be understood as a critical step in the history of music production. With its

inception, instead of trying to capture a recording at the time of performance, it could be enhanced and changed after the event. “An hour after the session or even the next day, you could sit down and readjust the balance of the bass or the guitar on a particular recording to create a whole new different mix” (Cunningham, 1998, p.48).

Multitrack tape machines soon developed beyond eight-tracks, with sixteen-track recorders available by the late 1960s and twenty-four track tape machines following soon after. Since the 1990s, multitrack tape has largely been replaced by hard-disc recording (though analogue tape is still favoured by some studios). This has reduced the limitations of the channel count that tape imposed, and modern DAWs offer the potential for unlimited tracks in multichannel recordings, the consequences of which are discussed in more detail below.

2.1.1 Development of Mixing Desks

The move to multitrack recording signalled a fundamental change in the production process where the recording could be separated from the mixing process for the first time. The introduction of multitrack led to the development of a plethora of new mixing techniques and equipment including delays, reverberation units, equalisers, filters, compressors and limiters (Channan 1995). Soon, recording studios started to realise the need to develop mixing desks to fully realise the potential of multitrack mixing (Cunningham 1998). While it should be acknowledged that mixing desks had been in use prior to multitrack recording, they were relatively primitive, with early studios like Sun Studios in Memphis often using broadcast equipment or rudimentary rotary mixers, which offered limited functionality (Cogan 2003).

The development of mixing desks to meet the possibilities of multitrack mixing can be attributed in a large part to the work of Bill Putman and his company Universal Audio (who have been credited with both the layout of the CS and much of its functionality). One of their key innovations involved integrating a preamp, equalizer, three program outputs, and a delay send into one module (Robjohns 2001). Though this was still relatively basic by modern standards, it set the precedent for the channel strip-based mixer as we know it today. Another key figure in the development of consoles was Tom Dowd, an engineer and producer at Atlantic Records (Massy 2010). Finding the rotary dials used on early studio mixers difficult to use, and seeking a more tactile and performative interaction whilst mixing, Dowd pioneered the inclusion of faders to the

console design. As Dowd himself said of their adoption in the CS mixer; “Finally, I could play the faders like you would a piano” (Simons, 2004, p.53).

The CS mixer went on to be developed throughout the 1970s by a plethora of companies such as Soundcraft, API and SSL and, though variations exist from model to model and manufacturer to manufacturer, the CS design remains the de facto standard for the analogue mixing console to this day, allowing producers control over levels, pan, filtering, equalisation and routing of effects and signal processors for multichannel mixing.



Figure 2:1: Reaper DAW mixer screen. Its design is closely modelled on an analogue mixing desk. However, unlike a physical desk, it offers potentially unlimited channels.

2.1.2 Transfer from Hardware to Software

Whereas the early decades of multitrack recording were dominated by physical mixing desks, since the 1990s the mixing process has largely moved to DAWs, especially in the burgeoning home studio sector, with many producers never having used a physical mixing desk (Massey, 2009). However, despite this fundamental change in both technology and users, most DAW interfaces remain visually the same as their physical counterparts (Bell et al 2015). The designs of various commercial DAWs vary slightly from one to another, with different strengths and weaknesses (table 2.1). However, a commonality between them all is that for each incoming channel, a channel strip is used

with several (virtual) dials, switches and a fader, each controlling a specific parameter in a one-to-one mapping.

The adoption of skeuomorphic design as the principle paradigm in DAWs is in part due to their emergence in the early 1990s when analogue mixing desks were still the dominant method for audio mixing workflow. The familiarity of the metaphor may be another reason for its longevity, as it can help users more acquainted with hardware feel comfortable mixing with computers (Battino & Richards 2005, p.204). Indeed, the machine-centric metaphor of the mixing desk design allows mapping of knowledge from the source domain of physical mixing desks to the target domain of the DAW interface (Duigan et al 2004).

It should be stressed that the channel strip metaphor is not without its benefits. The vertical layout of dials and faders allows a strong physical interaction, permitting engineers to control multiple channels simultaneously (Bell et al 2015) and use the mixing desk almost like an instrument (Simons, 2004, p.53). Indeed, artists such as the Beatles (Massey 2007, p. 112) and the Beastie Boys (Brown, 2009, p.45) have remarked on the performative nature of audio mixing on a physical console.

The channel strip can also help conceptualise the signal flow of audio through the console. In this metaphor, each track enters the console at the top of the channel strip, where it is subjected to adjustment (equalisation, panning, volume) or sent (bussed) to an effect unit, using auxiliary sends. The signal flow metaphor is the primary focus of the majority of textbooks on music production and forms the basis of much educational material on mixing workflow, and many mix engineers have learned to operate equipment modelled on this logic (Walther-Hansen 2017).

Furthermore, as graphical representations of the analogue channel strip layout are found in most of today's DAWs, its use can minimise the learning curve of moving from one platform to another (e.g. Logic to Pro Tools), at least for the fundamental aspects of the mixing process. Moreover, by implementing the analogue in DAWs, there are benefits in terms of allowing music producers to make a smooth transition from the physical to the virtual domain and visa-versa, which is an important consideration, as many large commercial recording studios, and many live performance venues still use physical channel strip audio mixing desks.

DAW	Logic (10.4)	Pro Tools (12.8.3)	Cubase (9.5)	Ableton (10.0.1)
Compatibility	Mac	Mac/PC	Mac/PC	Mac/PC
Maximum audio tracks	Unlimited	Ultimate: 256 Standard: 128 First: 16	Pro: Unlimited Artist: 64 Elements: 48	Suite: Unlimited Standard: Unlimited Intro: 16
Maximum software/MIDI tracks	Unlimited	Ultimate: 512 Standard: 512 First: 16	Pro: Unlimited Artist: 128 Elements: 64	Suite: Unlimited Standard: Unlimited Intro: 16
Included Audio effects	69	Ultimate: 60 Standard: 60 First: 20	Pro: 90 Artist: 70 Elements: 40	Suite: 41 Standard: 34 Intro: 21
Maximum sends per channel	8	10	8	Suite: 12 Standard: 12 Intro: 2
Maximum effects per channel	15	8	8	Suite: 12 Standard: 12 Intro: 2
Key Strength/Weaknesses	Detailed focus/ support for MIDI editing Limited workflow/ support for audio editing	Multiple shortcuts and interaction support for audio editing MIDI editing more limited than other DAWs	Focus on both compositional and engineering workflow. Metering not as advanced as Pro Tools.	Advanced support for live performance using clip view. Well integrated touch-based hardware interface (Push) Mouse-driven Automation control is limited.

Table 2.1: Comparison of features in commercial DAWs (as of May 2018).

The above may go some way to explain why use of the CS metaphor in software shows little sign of changing in the mainstream market with the launch of the latest versions of Pro Tools, Cubase and Logic all continuing to use the CS metaphor in their interfaces, and companies such as Slate Audio offering full-size mixing desks based around virtual CS mixers.

2.2 The CS Metaphor in DAWs

While, as mentioned above, the CS metaphor is the most commonly used design for software mixers, its ubiquity does not necessarily mean it is the best design. Users of applications usually acquire interface skills through exposure and repetition, meaning that interface designs which are common will often be accepted as the most natural by users, even though they may not represent the best possible interaction (Harrower and Sheesley 2005). In the case of DAWs, there is a danger that modelling of the GUI on real world counterparts risks keeping the disadvantages of physical hardware without utilising the advantages that software interfaces offer (Duigan et al, 2004., Battino & Richards 2005). An example of this can be seen in Propellerhead's Reason, where the metaphor of the studio has been taken through to the software in the need to 'wire' together the various outboard processors (figure 2.2). Though this provides a visually striking interface, in this case emulating the physical devices onscreen, it can compromise usability (Battino and Richards, 2005, p.200) and result in users being forced to work in a way directed by the limits of the physical technology rather than the potential of the computer (Harrower and Sheesley 2005).

2.2.1 Affordances

When mixing on screens, the structural metaphors and affordances from real world counterparts allow transfer of knowledge, but they do not always translate practically into the virtual representations, resulting in the metaphor breaking down (Mihnkern 1997). For example, while a dial on the physical desk suggests the perceived affordance of turning, this action is not supported when using a mouse, leading to errors and breaks in the user's engagement with the task (McGrenere & Ho 2000). Indeed, even at the launch of Cubase VST in 1996, some commentators were critical of the consequences of the modelling of physical dials in the virtual CS:

"The built-in effects are designed to look and operate as much as possible like their hardware counterparts. This is a friendly metaphor, but it may be unnecessarily limiting. The parameters are edited with a 'knob' using the mouse. The knob has fine resolution, but it's still tricky to 'nail' specific values" (Aikin, 1997, p.90).

Furthermore, when using dials to adjust pan position in CS mixers, the user must look at the position of the pan dials for every channel to get a sense of the lateral position of a source, which may impede the ability to globally visualise the panning of sound sources between the speakers (Gelineck et al 2013a). For example, the channel that is physically located on the left-most side of the mixer may in fact be panned hard right, creating a dissonance for the user (Ratcliffe 2014a). This is further compounded when using DAWs, as not all the channels may be visible on one screen, necessitating visual search across multiple pages to ascertain the panning of individual channels.

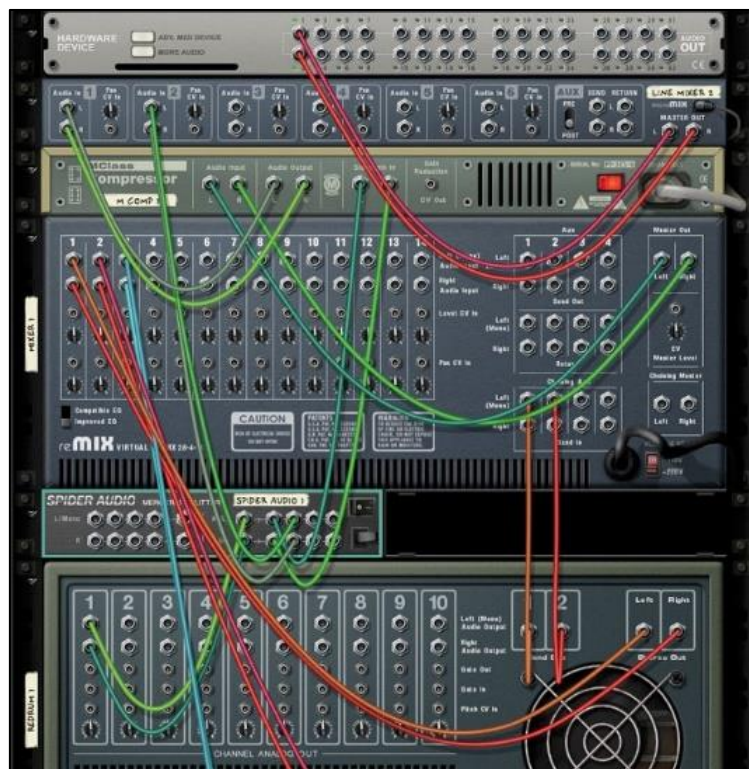


Figure 2:2: Reason rack view. The user is required to wire together the virtual rack modules. Though this gives the user flexibility, it can compromise usability

Faders are another fundamental part of physical mixing desks, allowing a tactile and ergonomic interaction with the mixing process (Simons, 2004). However, their adoption in DAW interfaces may be flawed. Htalky et al (2009) highlight a mismatch between

the faders used on physical mixing desks and their digital representations in terms of limited screen size, with virtual faders in many DAWs being of a size and resolution that are ill-suited to refined adjustments. Furthermore, when using a mouse, only one fader can be controlled at one time, disallowing the original performative premise on which faders were introduced to the mixing console.

2.2.2 Visual Feedback

Beyond the meters, physical mixing desks have very little dynamic visual feedback (Moorefield, 2005). However, in current DAW designs, conceptual additions have extended the complexity and visual feedback of the channel strip mixer (Golkhe et al 210) (figure 2.3). Changes to sound will often be reflected as changes to the waveform, and the tools employed to make those changes will also have a visual correlate. Furthermore, due to the potential for unlimited channels, the limited visual feedback found in a 24-channel physical desk can be far exceeded when using a virtual mixer. Subsequently, DAWs can make the mixing process not only an aural, but also an increasingly visual undertaking. Indeed, the use of visual feedback in DAWs and the subsequent interface complexity has been criticised for the way it focuses attention on the visual display directly to the cost of aural engagement (Crane, 2010). While commercial DAWs do address the issues raised above this is limited and varies between DAW manufacturers (table 2.2).



Figure 2:3: Cubase 9 mixer window. While still based on the CS metaphor, DAW mixers such as Cubase have extended the metaphor. For example, in Cubase 9, the mixer includes equalisation curves and waveform displays (at top of screen).

In literature beyond audio mixing, the possible negative effect of visual feedback on aural acuity has been widely discussed, a phenomenon which may have implications for the effectiveness of screen-based mixing. For example, several studies have suggested that visual stimuli can strongly modulate perceptions of auditory events and inhibit auditory perception (e.g. Colavita 1974., Sinnet et al 2007., Soto-Faraco & Spence 2010). The Colavita effect (Colavita 1974) demonstrated that the presentation of simultaneous visual stimuli during bi-modal (vision and audition) trials could ‘extinguish’ a participant’s perception of the auditory signal on a significant amount of trials (ibid). Similarly, the McGurk effect (McGurk and MacDonald 1976) has shown visual information (lip movements) can bias the perception of an auditory message (the words spoken) in favour of the visual information in a significant amount of trials. Saldana and Rosenblum (1994) and Shutz and Lipscomb (2007) demonstrated the same visual biasing effect on auditory perception on the ability of the trained musicians to correctly observe either plucking or bowing movements of a string instrument when conflicting visual information was simultaneously presented. Studies by Macdonald and Lavie (2011) found that when participants made either a low or high-load visual discrimination concerning a cross shape (respectively, a discrimination of line colour or of line length with a subtle length difference) the participant’s ability to notice the presence of a simultaneously presented brief pure tone was significantly reduced (87% of the low-load group reported hearing the tone, as opposed to 56% in the high-load condition).

This level of visual perceptual load is often found in DAWs where the user may have to attend to the mix while making very fine visual judgments or modifications (such as changing a parametric equalizer or observing the threshold movements of a compressor).

However, while the above shows evidence of the susceptibility of hearing to visual capture, evidence also supports the notion of independence of attentional resources for vision and audition (e.g. Triesman and Davies., 1973; Alais et al., 2006; Santangelo et al., 2010). For example, Larsen et al. (2003) compared subjects’ accuracy for identification of two concurrent stimuli (a visual and spoken letter) relative to performance in a single-task. They found that the proportion of correct responses was almost the same for all experimental conditions. Similarly, when Alais et al (2006) measured the discrimination thresholds for visual contrast and auditory pitch, they found that visual thresholds were unaffected by concurrent pitch discrimination of chords (and vice versa). Indeed, studies have suggested that the use of visual

information can be beneficial to auditory perception. When a speaker's lips and facial

	Logic	Pro Tools	Cubase	Ableton
Workflow views	Uses 'Arrange' view for editing and channel strip for mixing	Uses 'edit' view for arrangement and channel strip for mixing	Uses 'Arrange' view for editing and channel strip for mixing	Uses 'Arrangement' view for editing, 'Session' view for mixing and 'Clip' view for live performance.
Managing mixer window display	Channel strip width can be resized as standard or narrow	Channel strip width can be resized as standard or narrow	Channel strip width can be resized as standard or narrow	Channel strip width fully resizable. Channels strip height is resizable (extends the height and detail of the track meters).
Managing mixer channel display	Channels can be shown/hidden per channel type (Audio, MIDI, Auxiliary track etc.). Mixer can display folder tracks	Channels can be shown/ hidden per track. Channels can be shown/ hidden per auxiliary send	Channel strip elements configurable (fader, sends, inserts etc). Mixer can display folder tracks (several channels controlled by one fader)	Channel strip elements configurable (fader, sends, inserts etc.) Channels can be grouped (Arrangement view only).
Global view support	No overview of arrangement. Overview of Equaliser on channel strips is supported.	Overview of arrangement using 'Universe Window'	No Overview of arrangement.	Overview and Detail for arrangement and for the effects used on individual channels.

Table 2:2: Comparison of navigation management in commercial DAWs.

movement are visible, an improvement in performance has been observed equivalent to increasing signal to noise ratio by up to 15 dB (Ross et al 2007). Furthermore, a study by Marozeau et al (2010) showed that visual representation of a melody line had a positive effect on participants' ability to segregate the melody from a background of interleaved distractor notes. Similar research by Nakagaway et al (1999) has also shown that visual information can help in improving auditory stream separation.

As visual and aural stimuli are so tightly interwoven with DAWs, it is important that the influence of these two modes of perception are understood in terms of the influence visual display has on aural acuity. Understanding and quantifying the effect of visual load on concurrent critical listening tasks is therefore critical to informing the design process for mixing GUIs.

2.2.3 Interface Navigation

Physical mixing desks allow the visual mix information to be viewed simultaneously, as all channels are laid out in front of the user. However, in common with many computer applications, DAW users need to interact with more information than can be conveniently displayed on a single screen (Cockburn and Gutwin 2009). With increasing track counts and use of effects, there are times when the applications cannot be run on screens that adequately contain the workspace data. This problem is compounded by the fact that many DAWs have complex interface controls that consume considerable screen space and include large amounts of visual data (toolbars, tool pallets, status bars etc.). This is especially acute when the programmes are run on mobile and tablet computers, which even at their best, display a highly-reduced amount of visual information and are considerably slower to navigate than desktop versions (Gutwin & Fedak 2004).

To allow the user to access mix information beyond the confines of the screen, DAW GUIs use navigation techniques to display different parts of the visual data, including floating windows, nested menus, zooming and scrolling. However, because music is experienced in real time, there is a short window for making changes to mixes, therefore the navigational controls have to be fast and easy to access, otherwise they may obstruct the speed of workflow (Gelineck et al 2013b). Indeed, Csikszentmihlyi (1991), Bederson (2003), Mistry & Agrawal (2004) and Shneiderman & Bederson (2005) relate the concept

of 'flow' specifically to interface design. Shneiderman and Bederson (2005) maintain that well-designed interfaces can encourage user engagement and support task completion. Among the suggestions that they put forward to elicit flow are that interfaces should reduce load on short-term and working memory (WM) (Baddeley 2003), reduce excessive window navigation and use macros and command key shortcuts.

Below we discuss the principle navigation methods, using examples from commercially available DAWs, and we relate them to the mixing process to highlight potential issues of implementation.

2.2.3.1 Window Switching

DAWs allow multiple views of the mix space, which can be set by the user. For example, all effects (such as reverb, delay etc.) can be viewed as floating windows in addition to the main mix/ edit windows. The use of floating windows can be of benefit when mixing. As comparisons of data sets can be made via eye movements (Plumlee and Ware 2006 p.13), floating windows can minimise search times (Baldando et al, 2000). Multiple windows can also facilitate the partitioning of complex data, creating manageable chunks that can provide insight into the interaction among different dimensions (ibid). However, though floating window layering can be useful, it can also potentially reduce fast intuitive interaction (Gelineck et al, 2013b). For instance, additional views can incur a cost in terms of context switching from one screen set to another (Plumlee and Ware 2006). Additionally, setting up, resizing, repositioning and changing the size of floating windows can all potentially reduce the speed of interaction and can incur considerable cost in terms of navigation, set-up and access (Convertino et al, 2003). Layering of information through floating windows can also tax available display space requirements and obscure other (useful) screen information. An example of this can be seen in Apple's Logic Pro, where the floating windows for effects and inserts cannot be resized below a set level thereby obscuring large parts of the available screen space (fig 2.4). Floating windows can also contribute to screen clutter (Rosenhaltz et al 2005), obscure more relevant visual information, and potentially overload the users' visual and attentional perceptual bandwidth (Rensink 2012).

2.2.3.2 Zooming

DAWs require frequent scale changes to search for details and make comparisons, such

as seeing waveforms at different level, both fine (sample level) and coarse (level of a musical phrase). In most DAWs, this is achieved through zooming, or scalable view changes (Golkhe et al, 2010). For example, in Logic and Pro Tools, zooming on individual channels is achieved through the use of horizontal and vertical sliders.



Figure 2:4: Logic Pro 9. Two floating windows cannot be resized below default minimum, thereby taking up much of the available screen space.

There are, however, perceptual costs to zooming that need to be understood as they may influence the efficiency of screen-based mixing. For instance, relating the overview zoom level to the detail zoom level requires rapid changes in orientation. Cockburn and Savage (2003) found that these abrupt transitions between discrete zooming levels caused significant disorientation as participants reoriented themselves and assimilated the new view with each zoom change. Furthermore, during zooming navigation, the user is required to hold the visual information in their limited short-term memory (STM) while different scale views are revealed through the zoom functionality (Plumlee and Ware, 2006).

This problem is compounded by the ‘temporal separation’ between views and therefore becomes more critical as the levels of zooming increase. In a comparison of zooming interfaces to floating windows, Plumlee and Ware, (2002 and 2006) found that when using zooming navigation, user’s STM was taxed due to the fact they had to hold everything outside their current view in their memory, whereas multiple windows reduced the load on STM by allowing direct visual comparison. They concluded that zooming interfaces were unsuitable for complex comparison tasks (which are often

found in DAW workflow, such as comparing multiple channels settings). Lastly, zooming interfaces represent data at a highly scaled down level when fully zoomed out. This can make it hard to perceive and quantify fine levels of detail in the data (Bederson 2003). For this reason, zooming navigation works best when the objects can be recognised at zoomed out levels. Within the CS metaphor, fine details in dials and faders may prove hard to perceive meaningfully at high zoom levels, as they become limited in screen size and resolution, meaning that their ability to be analysed and changed may become compromised (Htalky et al 2009).

2.2.3.3 Scrolling

A common feature across all DAWs is scrolling, which is implemented due to the fact the information “overflows” off the screen out of immediate view (Sanchez and Wiley, 2009). As such a common navigation method, it is necessary to consider the perceptual and workflow implication of using scrolling while mixing.

Scrolling (or panning) is a fundamental technique for freely moving around two-dimensional continuous space, allowing the user to move to different locations. However, GUIs still have some fundamental limitations when using scrolling navigation (Igarashi & Hinckley 2000). For example, it is difficult to browse a large data space efficiently as the user must move back and forth between the GUI and the scroll bar. This can increase the operational time and may cause significant attentional overheads (Lam 2008). Indeed, studies on the use of screen presentation have highlighted higher levels of mental load when the use of a scroll bar is included in the interface (Piolat et al, 1997). This was especially true of users with a limited WM capacity, who could be predicted to show the largest decrease in comprehension caused by scrolling, as they were less able to maintain both the surface representation of the data while also processing the relationship between the scrolled content (Sanchez and Wiley, 2009). Scrolling can also incur split attention deficits in trying to relate different views to each other (Feinberg and Murphy 2000) as the user must remember one information element while searching for the other, leading to a high WM load.

Another factor to consider is the disorientation caused by scrolling, which may compound the problems of STM. Sanchez and Wiley (2009) found disorientation an issue with scrolling interfaces since they lack a static ‘place on a page’ (p. 731). In a more recent study, Sanchez and Branaghan (2011) found that by simply rotating small screen device displays by ninety degrees, and thus minimising the need to scroll, reasoning was

significantly improved.

2.2.3.4 Overviews

In order to avoid disorientation during navigation, strong local-global orientation cues are needed to reinforce a sense of place and increase efficiency of use (Harrower and Sheesley 2005). Indeed, under complex navigation, user orientation becomes a key issue (Gahegan 1999) and providing well-designed global views of the data is considered a fundamental criterion for successfully navigating the information space (Shneiderman 1996; Card et al 1999).

Overview + Detail (O+D) interfaces (Plaisant et al. 1995), address these issues by allowing the simultaneous display of both an overview and detailed view of an information space, each in a distinct presentation space. Users interact with the views separately, although actions in one are often immediately reflected in the other, which allows users to maintain spatial awareness during navigation tasks. For comprehending the content of an interface (e.g. reading texts) they have been found to be particularly effective; studies by Cockburn (2009) have shown that navigation is faster with O+D interfaces since users can navigate in both the overview and detail view. Furthermore, the contextual information provided by the overview has been found to help users maintain orientation and make decisions about future navigation actions (Nekrasovski et al 2006). By using an O+D display, the user is not only helped in understanding individual data points, but also in understanding the relationships between several data points (Duigan et al 2010). Furthermore, there is some evidence that at global-view scales, overviews may also continue to support interaction techniques such as targeting, steering, scrolling, selection, and dragging (Gutwin and Fedak 2004) thereby potentially reducing the need for multiple changes in zoom level.

Within DAW interfaces the use of overviews has been adopted to a limited extent. For instance, Ableton uses overview and detail in both the arrange views and the effects and inserts view (figs 2.5 and 2.6). Pro Tools (version 8 onwards), includes the 'Universe Window', an overview of the edit page (fig 2.7), while Logic displays an overview of each channels EQ curve, and Serato makes uses coloured waveforms to identify individual drum parts (kicks, snares etc.) even at high zoom levels (figure 2.8). However, the CS mixer may not be well suited to overviews, as users must search the detail of every UI object (such as dials) to find relevant parameters, and at high zoom levels, the requisite detail may not be visible to represent or adjust this information in a useable way (Gelineck et al 2013a).



Figure 2:5: Overview and detail in effects page in Ableton (Overview is in bottom right hand corner).



Figure 2:6: Overview and detail of arrange page in Ableton (Overview is in top left of screen)



Figure 2:7: Pro Tools Universe Window. This provides an overview of the arrangement (top of screen) to give context to smaller sections (main screen).



Figure 2:8: Serato uses coloured waveforms so that users can identify individual drum hits even at high zoom levels (darker colours are mapped to lower frequencies).

2.3 State of the Art

In this section, we review the current State of the Art. We provide an overview of audio mixing GUI developments, and briefly discuss interaction techniques to control the mixing software. We consider the advantages and disadvantages of the designs in terms of conceptual models, user knowledge, information presentation and usability. We conclude the section by reviewing the evaluation of novel designs.

2.3.1 Non-technical Interfaces

It has been argued that the choice and implementation of a metaphor used in software design can have a far-reaching psychological influence upon the user (Norman 1986). For example, research by Yamamoto et al (2005) suggests that the interface design can strongly influence a user's cognitive process, 'encouraging, discouraging, permitting or prohibiting a user from taking a certain course of action' (ibid, p. 533), while Bertelsen et al (2008) suggest a fixed metaphor can constrict models and create narrow conceptions of the domain object. Moreover, Carrascal and Jorda (2011) suggest that the aesthetic features of an interface can affect both its usability and perceived performance, while Sabin and Pardo (2009) and Sabin et al (2011) believe that the complexity of music production GUI tools can be a significant bottleneck in the creative process. Indeed, the intricacies of the conventional DAW designs can result in 'musicians without technical expertise spending a great deal of time stumbling through a large range of parameter settings' (ibid, p.435), which can disrupt the creative process and direct attention away from the music itself. These designs can be confusing to novice users, who face 'what seems like inexplicably-motivated design decisions' (Stowell and Mclean 2011, p.1) and feel 'overwhelmed 'by the CS controls and

subsequently ignore much of their functionality' (Carrascal and Jorda 2011, p.102). Due to the influence of the CS design, and its impact on usability, several researchers have begun to examine simplified, non-technical GUIs, which focus on the most central features for mixing.

With 'Mixploration' (Cartwright et al 2014), the authors put forward a simplified GUI intended to encourage a more exploratory approach to mixing. By dividing the mixing process into two stages- exploration and then refinement- the software helps users create high quality rough mixes based on an evaluation of what they hear. The authors contrast their visually simple GUI with CS designs, which they criticise for prioritising the minute detail of adjusting single faders and dials over global goals of achieving an acceptable overall mix. Related to this, Mecklenburg and Loviscach (2006) have designed a GUI which supports non-expert users in music equalisation tasks and minimises distraction caused by an overly technical interface. In their design, the GUI is dramatically simplified, containing only subjective terms ('warm', 'present' 'boomy') and a dot on a graph, which the user positions over the appropriate term. Similarly, Pardo et al (2012), motivated by their belief that DAW equalisers do not adequately meet the non-expert's conceptual and creative goals, have created a GUI in which the user manipulates the equaliser (EQ) in terms of a descriptive language (e.g. "warm", "bright" or "tinny"). Though their research is focussed on EQs, they conclude that intuitive non-technical interface designs could have benefits in mixing workflows beyond equalisation and may be successfully applied to other areas of music production, such as reverberation effects (Seetharaman and Pardo, 2016).

However, while a less technically demanding metaphor may allow non-experts quicker access to the mix parameters, there is a risk that they can also reduce the level of functionality. For example, in conventional DAW EQs, the user can make very specific and refined adjustments to the frequency spectrum. In EQ designs that use natural language and descriptive terms, user satisfaction with the results can be varied (Pardo et al 2012) and precise control is limited. Indeed, reducing the amount of detail within the interface does not necessarily present a clear-cut solution. Norman (1986) suggests that deciding on the amount of detail to include in an interface can result in an inevitable trade-off; while detailed interfaces can increase navigation time, they can be useful in making various choices and functionality explicit. Furthermore, there is a danger that interfaces with very limited functionality for precise refinement can start to feel limited, even to novice users (Gelineck and Korsgaard 2014). However, despite these concerns,

non-technical interfaces may help non-expert users gain confidence and knowledge of the mixing process (Dewey and Wakefield 2014) and provide a less daunting visual experience than the banks of dials and faders found on a mixing desk (Selfridge and Reiss 2011).

2.3.2 Interaction Techniques

2.3.2.1 Gestural Interaction

As discussed above, in developing new designs for the GUI, methods of interaction beyond the dials and faders of analogue mixing or the mouse driven control of desktop DAWs need to be considered. One solution is to allow the user to step away from the mixing desk, and undertake the process remotely via gestures. Among the potential advantages identified in this approach are the removal of visually obtrusive UI elements such as buttons and dials (Balin and Loviscach 2011) and a greater sense of immersion in the mixing process. Several implementations of gestural mixing designs have been proposed. These include Selfridge and Reiss (2011), who used a Wii controller to mix audio tracks, Ratcliffe (2014b), who implemented a gestural mixing interface using Leap Motion and the Microsoft Kinect, Lech and Kostek (2013), who designed a sound mixing interface in which gestures are recognized via camera stream processing, and Law and Hou (2015) who implemented a four-channel mixing interface in which wearable technology controlled the process.

While these approaches reportedly helped increase the sense of immersion and involvement in the mixing process, (Lech and Kostek 2013) and were met positively by users, there were also negative aspects. For example, fine adjustments were not as easy with gestures as with a mouse and in certain cases it was difficult to set the parameters to exact amounts (Selfridge and Reiss 2011). Furthermore, problems were identified with the hardware used, with Lech and Kostek (2013) concluding that controllers, infrared sensors, and accelerometers do not provide sufficient ergonomics to be adopted for the subtleties of audio mixing. Indeed, in interfaces that allowed detailed manipulation, such as changes to equalisation, users found it difficult to accurately control these with gesture, with much of the mixing time being taken up trying to master the movements needed to control the subtleties of parameter setting (Selfridge and Reiss 2011).

2.3.2.2 Tangible Interaction

A further criticism of gestural mixing is the lack of haptic feedback (Law and Hou 2015). Indeed, research has emphasised the importance of tangible controls (Ratcliffe 2014a, Gelineck and Overholt, 2015) to enhance users' performance by providing tactile feedback, considered by many to be an essential part of playing music and producing sound (Law and Hou 2015). For example, Gelineck, et al (2013a) have developed a mixing interface that lets users control effect parameters of up to 24 channels by manipulating tangible user interface (TUI) blocks over a multi-touch surface. Similarly, Diamante (2007) emphasises the importance of tangibles in mixing interfaces, with his design using a pen as the input device, thereby allowing a better sense of tactile feedback. Likewise, Lopes et al (2011), in an attempt to provide a more natural interaction method, experimented with tangibles to control parameters of DJ mixing, including loop speed, cross fading and turntable motion.

While tangibles helped users with proprioception and provided physical feedback, in some experiments their use resulted in limitations. For instance, Gelineck et al (2013a) found the tracking system too slow to respond to lifting and placing of the physical objects controlling the interface. Lopes et al (2011) found that some participants considered the TUIs a distraction, with many expressing 'increasing discomfort with the control objects' (p.369). Relatedly, Gelineck et al (2013a) found that the TUIs tended to occlude important information on the screen, and in some cases, distracted users from listening to the music and fully engaging with the task of mixing.

2.3.2.3 Multi-touch Interaction

The advent of multi-touch technology has 'raised general awareness of the possibilities presented by direct manipulation of the visual information' (Dewey and Wakefield 2014, p.1). Multi-touch surfaces may provide more precise control than tangible or gestural mixing and, unlike mouse-based mixing, allow multiple parameters to be adjusted at one time (e.g. both the pan and volume of a channel). Furthermore, an evaluation of multi-touch mixing by Carrascal and Jorda (2011) suggested that compared to an analogue mixer, the multi-touch interface was easy to learn for novices, more time-efficient for mixing tasks and generally preferred.

In keeping with these advantages, multi-touch technology has been widely adopted by commercial DAW manufacturers. Steinberg's Cuabasis takes the desk-top metaphor and functionality and translates it almost intact to a multi-touch platform, allowing

direct manipulation of faders and dials from the screen. Similarly, WaveMachine Lab's Auria transfers the existing CS design directly to a touch-screen application (figure 2.9). There have also been uses of multi-touch technology which extend and adapt the CS metaphor. For example, the Cubert mixer (Liebman et al 2010) takes the channel strip and modified its functionality to provide a more flexible instantiation of it, using it as a space to allow quick access to several controls (such as effects) within the horizontal layout.



Figure 2:9: WaveMachine Labs Auria. This software transfers the desktop CS metaphor directly to touchscreen platform.

However, multi-touch mixing interfaces have experienced limitations. The size of multi-touch surfaces may not be ideal for large multi-track projects and may become a constraint on the development of the software interfaces (Law and Hou 2015). Furthermore, when using small screens (such as tablet computers), widgets and on-screen tools may be too small to apply fingers to (Gelineck and Korsgaard 2014), consequently disallowing the amount of detail required for more complex mixing (ibid). They also found the touch control itself to be an issue, with exit errors (lifting fingers) making it difficult for users to make very specific parameter adjustments. However, despite this, the authors were positive about the technology, and identify several potential areas in which multi-touch based mixing interfaces might provide benefits, especially when used on small-screen devices. These include mobility, demo mixes for novices or those who lack the resources or skills for a console or desktop DAW, and quick and easy on-stage monitor mixes (the individual mix each musician hears, usually different from that played through the main speakers to the audience) (Gelineck and Korsgaard 2014).

2.3.3 Stage Metaphor Mixers

Common to several of the implementations in gestural, tangible and multi-touch designs discussed above (Dewey and Wakefield 2016.; Law and Hou, 2015., Gelineck, and Overholt 2015., Ratcliffe, 2014., Gelineck and Korsgaard, 2014., Gelineck et al 2013a.; Gelineck et al 2013b., Lech and Kostek, 2013., Carrascal and Jorda, 2011., Diamante 2007) is the concept of the stage metaphor (SM) mixer. In this design, the mix is broken down into three planes; width, depth and height. This creates a virtual space within which producers can place elements front-to-back, side-to-side and top-to-bottom (Moylan, 2007). As the mix is broken down into three planes it allows a strong visual correlate; e.g. things which need to be panned left are put to the left of the screen, enabling the user to visualise the absolute and relative distribution between audio channels (Dewey and Wakefield 2014). This is in contrast to the CS metaphor, which forces the users to ‘scrutinise each channel’s pan knob position’ to assemble a mental image of the stereo positioning (Dewey and Wakefield 2016, p.2). Indeed, in evaluations of panning in a CS and stage mixer, Gelineck and Korsgaard (2014) found panning was utilised more fully in the latter, especially with novices, many of whom who did not undertake any panning at all when using the CS mixer. The enhanced spatial representation of channels within the SM mixer may also allow better use of global views as it facilitates the user to quickly see patterns within the mix, which may be of benefit in avoiding common mix errors such as masking, bunching of elements within a certain pan position (Case, 2007) and ensuring any outliers (in terms of volume, pan etc.) can be easily attended and selected (Wolfe et al 2000).

Gelineck et al (2013b) suggest that conceptually, the SM is much closer to the mental model of how the user may conceive the mix, and subsequently it may help reduce the gulf between evaluation and execution (Norman 1988) by narrowing the gap between the mental model of the user's goals and the physical action of the controls. In an evaluation of an SM mixer, Gelineck and Overholt (2015) found that participants, especially non-experts, preferred the intuitive aspect of the SM mixer, while Carrascal and Jordà (2011) suggest it helps users exploring new ways of approaching the task and supports them in undertaking a more creative exploratory approach to their mixes. Indeed, the SM has been used as a way of visualising mixing concepts (Dewey and Wakefield 2014.; Hodgson, 2011.; Dockwray and Moore 2010.; Moylan 2007) for music production education. For example, Gibson (1997) uses the concept of the SM mixer expressly to analyse mixes in terms of equalisation, panning, levels and reverberation and ensure that no elements are audibly hidden (fig 2.10).



Figure 2:10: Though never realised as a working mixer, Dave Gibson uses the concept of the Stage Metaphor to analyse mixes and visualise ways of placing elements within a multi-channel mix. From Gibson (1997)

2.3.3.1 Limitations of the SM Design

Despite positive findings with the SM metaphor, there are issues with its usability. For example, because channels are randomly distributed (unlike a CS mixer where they are in a fixed order), the GUI may become difficult to search, especially with larger multichannel mixes (Gelineck et al 2013b). Furthermore, channels with similar pan positions and level may overlap each other on the display leading to problems finding and adjusting specific channels. In response, a number of solutions have been proposed to address these issues. Carrascal and Jorda (2011) suggest layering, with different ‘stages’ for different functionality, so that the main mix is one stage, the reverb of the channels is another etc. In this way, the amount of visual information on the screen at one time is minimised. Gelineck and Korsgaard (2014) also propose a form of layering, allowing users to toggle between a ‘limited’ or ‘full visual’ mode (ibid, p.5.) the former hiding effects and focusing solely on volume and panning. However, as there is a short window for making changes to the mix, and controls need to be quickly available for the user, (Gelineck et al 2013b.; Liebman et al 2010) access to functionality through layering may slow down the mixing process and obstruct users from making quick responses to the mix. While this concern is recognised, it is not investigated in the literature, and no comparison is made of layering interfaces compared to ones in which the visual information is visible at all times.

Another proposed solution to cluttered SM GUIs is filtering the interface to only show details relevant to the user. For example, Gelineck and Uhrenholt (2016) experiment with dynamic query (DQ) sliders to improve legibility of the interface. These are User Interface (UI) objects (sliders, buttons and other filters) that facilitate rapid exploration of interfaces by real time visual display of query formulation and results (Li & North 2003). While the authors felt DQ filters showed promise, with the majority of participants responding positively to their inclusion, they voice their concerns that the visualisations may potentially draw attention away from the mixing process and slow down mixing. However, this is not assessed in their study.

2.3.4 Evaluation of Novel Designs

Evaluating mixing interfaces can be a difficult undertaking. As Carrascal and Jorda (2011) remark, though mixing has an inherent technical component, there is no precise good or bad way of carrying it out (*ibid*, p.102) and in many ways it is a subjective undertaking. Evaluating the success of a mixing interface can also provide contradictory results, as preferences can be biased due to resistance to change when presented with designs which differ from established GUIs (Dewey and Wakefield 2014). For these reasons, structured HCI approaches need to be applied to the field (Stowell and McLean 2011), and quantitative as well as qualitative measures of usability need to be considered.

While Dewey and Wakefield (2014) apply HCI techniques to the development and evaluation of a novel music equaliser, and Lopes et al (2011) record and analyse the time taken to complete tasks, a review of the literature reveals several evaluations of novel interfaces which fall short on quantitative measures of usability. For example, Cartwright et al (2014), while comparing their novel mixer to a CS design, base their comparison only on subjective evaluations, such as asking participants which interface facilitated more satisfying mixes. Similarly, Gelineck and Korsgaard (2014) use grounded theory to collect subjective opinions of their software design, while, Liebman et al (2010) also use comments from the participants rather than quantitative metrics as a basis for their evaluation. Gelineck and Uhrenholt (2016) omit metrics of timing, accuracy or comparison to a traditional CS interface when making conclusions about the effectiveness of DQ filters, relying rather on subjective user feedback. In other studies, a lack of comparison with current DAW designs makes it hard to draw comparisons or establish a reference for any improvements (e.g. Liebman et al 2010.; Diamante 2007.; Gelineck and Uhrenholt 2016). While Carrascal and Jorda (2011) do

compare their interface with a traditional digital mixer, they choose a physical rather than virtual mixer, and the authors themselves recognise that a comparison with a software CS mixer would have been a valuable undertaking (ibid, p.103). Similarly, Gelineck et al (2015) compare their virtual mixer with a channel strip mixer, a decision they acknowledge may not be suitable.

Lastly, some of the literature uses limited tasks for evaluation. While Gelineck and Overholt (2015) did consider quantitative analysis (2015), the mixing task given to participants was basic, requiring only three channels to be balanced, and only involving panning and volume setting. The authors admit that lack of statistically significant differences found in the study may be attributable to the simplicity of the task. Likewise, Carrascal and Jorda (2011) and Law and Hou (2015) used only a four-channel mix when evaluating their interfaces, while Dewey and Wakefield (2016) and Gelineck et al (2015) limited their evaluation tasks to panning and volume, and did not address the need for greater mixing functionality (such as equalisation and effects).

2.4 Summary

In this chapter, we reviewed the development of multichannel mixing, the adoption of the paradigm in audio mixing software interfaces, and investigated the practical implications of mixing using this design, both in software and hardware. In the first section of this chapter, we reminded ourselves of the development of multitrack recording and related this to the CS mixer and the functionality that it allows. This historical and technical knowledge was provided as a basis to understand how the mixing workflow is mediated through CS mixers; knowledge which we draw upon in subsequent chapters.

In the second section of this chapter, we investigated the practical implications of mixing using DAWs. We explored the emphasis DAWs place on visual feedback and how this may influence auditory attention. We also examined the navigation required to access visual information and related it to perceptual limitations and workflow. With a limited analysis of DAW navigation to draw on in published research, we looked at more general areas of HCI to understand the perceptual implications of interface navigation. However, to make it germane to this thesis, we related this to specific DAWs including Logic, Pro Tools, Cubase, Ableton, Reason and Serato.

In the last section, we looked briefly at the state-of-the-art, and considered different interaction techniques for mixing, and different metaphors for representing and controlling the mix elements. We discussed the SM metaphor, which is common to several different interaction techniques, and compared this to the CS metaphor in terms of conceptual models, information design and mixing workflow. Finally, we commented on the limitations found in some evaluations of novel mixing interfaces.

Chapter 3

Strategy and Research Methods.

In this chapter, we outline the strategy to investigate the research questions identified in chapter one (section 1.3). In section 3.1, we provide a broad summary of the experiments described in the thesis. We illustrate how one experiment follows on from the other, set out our strategy and provide an overall picture of how the various experiments form a coherent body of work to address the main research questions. In section 3.2, we continue to provide a broad focus, discussing the general rationale for the experimental tasks included in the studies, and providing an overview of our experimental approach. In section 3.3, we present a summary of the data analysis methods used within the experiments, looking at our general approach in terms of quantitative and qualitative analysis. In section 3.4, we discuss the participants used in the study, and provide a rationale for their inclusion, including our recruitment procedure and sampling approach. In Section 3.5, we begin to provide more detailed information, focusing on each experiment in turn and discussing the experimental methods, experimental tasks and data analysis used for each one. The chapter concludes with a summary of the key issues discussed.

3.1 Thesis Experiments

3.1.1 Rational for Methodological Framework

To serve as a starting point for our studies, we consider the on-going discussion, found both anecdotally and in academic literature, as to whether when mixing, the method of interaction with the audio changes the end results. For example, producers such as Adrian Sherwood and Frank Fillipetti (Fillipetti, 2018) mix with little recourse to quantitative information, such as dB levels, specific frequencies, or millisecond settings of effects and compressors. Rather they use the mixing desk in a proprioceptive way, akin more to a performance, where mixing involves the hand-eye relationship of moving dials and faders (Owsinski 2006). Indeed, some producers believes that the visual feedback in DAWs can be a hindrance to aural engagement with the mix; by focussing on the quantitative information in DAW GUIs while mixing users don't listen as critically as in a situation in which there is nothing visual to be drawn to (Valeriotte, 2016).

A similar argument, which again serves as a starting point for our experimental designs, is the concern that the visual feedback per se will lead decisions and approaches when mixing. For example, Don Hahn comments that producers don't listen when confronted by visual feedback, rather they just "start adjusting faders and grabbing the EQs" (Owsinski 2006, p.169). Similarly, in terms of equalisation, there is tendency for users to be too driven by the visual feedback, setting EQ curves for instruments by sight, rather than considering the sonics of each one in turn, and the affect the EQ curves have on the sound (Massey 2000). In this way, the graphical information takes precedence over critical listening.

Relatedly, our methodology is informed by acknowledging the possibility that there may be a correlation between the level of visual feedback and the perceived impact of auditory changes. An example can be seen when users believe that a track is too quiet, based not on the perception of what they hear, but rather on the physical size of the graphic representation of the waveform on the screen (Battino and Richards, 2005 p.169). This tendency can also be seen in equalisation, where it has been suggested users are tempted by the visual display to 'draw' a good-looking EQ curve, instead of basing the EQ curve on listening (Larsson, 2014, p.3). This problem is compounded by some DAW EQs, where the scale of the gain axis means that minor EQ changes of 1 or 2 dB will appear very small visually, tempting the user to increase the amount of gain to make it visually, rather than sonically meaningful (fig 3.1). This can also be seen at the other extreme, where users may refrain from making large gain adjustments as it appears visually too significant.

Finally, our methodology is informed by the consideration of psychological aspects of the interfaces. Because, for example, equalising is a task guided by subjective preferences, the "technical" visuals could influence, and possibly hinder or modify the process (Larsson 2014). Does the segmented, quantitative design of Channel Strip elements inform the workflow? Is it the case that the graphics and visual information design itself will affect how the users undertake mixing workflow?



Figure 3:1: The Pro Tools channel EQ. The gain scale of +/- 18dB can potentially make small gain changes (in the region of 1 or 2 dB) seem visually insignificant. While some EQs allow rescaling of the gain axis (e.g. to show it as +/- 3dB) this is not possible here.

Using these broad discussions as the starting point for our approach, we aim to address and quantify whether the above have a bearing on the critical listening decisions of the user. This informs our studies in general, and our first study in particular, and sets the course for the further studies described in this thesis. We also wish to lend ecological viability to our studies, and we therefore ensure the studies involve participants undertaking actual mixing tasks with differing interaction methods. Therefore, in all the studies described in this thesis, we design and trial different interfaces, with varying amounts of visual information, and varying methods for showing mix parameters. We do this in order that we are able to directly compare the outcomes from differing approaches in terms of both quantitative and qualitative aspects of the mixing workflow.

We therefore begin study one by setting up a critical listening test, using equalisers with varying amounts of visual feedback, to ascertain if, and to what extent, a visually based approach may differ to an aural one. Taking a pragmatic point of view, we then use the results from this study to inform the design of subsequent studies, and thereby refine and extend our approach to answering the research questions as we proceed through the experimental chapters.

3.1.2 Summary of Experiments

The experiments described below are designed to address the main research question, namely how best to present the mixing tools and information within DAWs so that the sonic content of the mix is the user's primary focus, and visual information is used to support and enhance this primary function. Within this area of research, three specific questions are also addressed: To what extent does visual perceptual load interfere with or support the processing of auditory information? Does the design and layout of the visual information have an impact on aural acuity or speed of visual tasks? How can visual mix information display be managed more effectively? In the aim of addressing these questions we have designed the following seven experiments:

- **Experiment 1:** This experiment is a purposefully broad, high-level investigation into bi-modal attention while mixing to ascertain to what extent visual representations of mixing tools affects critical listening. The knowledge gained will provide a starting point for further studies, set future directions for research, and crucially, begin to address our first research question; to what extent does visual perceptual load interfere with or support the processing of auditory information?
- **Experiment 2:** Building on the results from experiment one, we conduct an experiment in which the GUI is more akin to DAWs. Firstly, as scrolling navigation is commonly employed in DAW GUI design, we include this functionality. Secondly, we increase the number of channels to that encountered in a simple mixing session, and thirdly, we include interfaces with moving distractors, in the form of peak meters and equaliser spectrum analysers. By so doing, we continue to address our first research question and further investigate whether varying amounts of visual perceptual load affects simultaneous critical listening tasks.
- **Experiment 3:** Here we assess UI elements commonly used in CS mixers design (faders, dials, numbers and colours) to see whether different levels of visual perceptual demand inherent in these designs influence the simultaneous perception of auditory stimuli and visual search. In this experiment, we also address the second research question; does the design and layout of the UI influence critical listening skills, and do certain designs improve or detract from the ability to hear subtle changes to the audio content and find mix information while navigating the GUI?
- **Experiment 4:** In this experiment, we continue to address our second research question. Specifically, we test interfaces incorporating overviews of the mix information to assess whether by removing interface navigation, we affect visual search times and critical listening skills. The experiment is informed by the results

of the previous two chapters, and the theoretical and practical implementations of overviews within other fields, such as instructional design and cartographic displays, which suggest that they may help avoid the negative effects of disparate data in the GUI, and orient users in the search environment.

- **Experiment 5:** For this study, we test the inclusion of dynamic query (DQ) filters within the GUI. By so doing, we aim to quantify the extent to which filtering of the GUI may facilitate improvement in search times and/ or concurrent critical listening. We also assess whether their inclusion may distract users from the mixing workflow. We use both SM and CS mixer designs for the overview. The experiment begins to address our third research question; how can visual mix information display be managed more effectively?
- **Experiment 6:** This experiment continues to address our third research question and builds on the results from the previous studies. We consider how in other areas of data visualisation (such as medial visualisations, geo-spatial and cartographic displays), multiple parameters are often integrated into one graphical object to leverage perceptual skills and help users discern and interpret patterns within the data. We therefore compare interfaces which integrate UI objects into one graphical object (using an SM mixer) against interfaces which have a one-to-one mapping of controls. By so doing, we aim to assess and quantify which designs are most effective, how much detail can be encoded and what impact such designs have on visual search times and visual search accuracy.
- **Experiment 7:** In this, our final study, we build on the results from the previous experiments, and design a mixing interface incorporating several findings from previous studies, including the design of the UI objects, the design of the GUI and the management of the visual information within it. We test this against a CS mixer, and assess several mixing workflow tasks using both designs. We use this experiment to test, develop and refine our heuristics, which we present in chapter 11, discussing limitations, future work and future directions.

3.2 Experimental Design

3.2.1 Experimental Considerations

In the studies reported in the thesis, we have implemented several experimental design considerations. These have been included to ensure that both quantitative as well as qualitative metrics of evaluation are used. These include the following:

- Ensure that traditional interfaces are evaluated alongside novel interfaces: All the experiments have included traditional UI objects (faders, dials etc.) alongside any novel designs being evaluated as this provides a benchmark against which to measure the effectiveness of novel designs and establishes a context for the results (Lopes et al 2011.; Carrascal and Jorda 2011.; Dewey and Wakefield 2014).
- Elements of the test should be randomised to ensure unbiased evaluation: With all the studies, the order in which various GUI designs were tested, the order in which the tasks were presented and the order in which participants undertook the studies were randomised. This ensures the influence of listening fatigue, improvements occurring due to continued use of the interface, and the impact of continued exercise of critical listening are accounted for.
- Test subjects should be allowed to practice using each interface: Training was provided for all the interfaces used in the thesis studies. This was not time-limited and participants were asked to confirm they had had enough time to overcome any issues of interface learnability and familiarity. This was especially important as most participants were more familiar with CS mixers than the novel interface designs presented. Research by Reiss (2016) suggests that auditory tasks in experiments provide improved statistical significance when training and practice are part of the experimental design.
- Task completion time should be recorded to quantitatively evaluate effectiveness and efficiency: Wanderly and Orio (2002), propose that when evaluating audio production interfaces, tasks should be timed to provide a quantitative metric of usability. The experimental interfaces all included a timer function (hidden from the participants). This was used not only in recording completion time for the whole task, but also for smaller, sub tasks, such as timing how long it takes for a participant to find a specified audio channel. Timing the tasks also allows clear comparison of differing GUI designs (Dewey and Wakefield 2016)
- Tasks should consist of discrete, simple tasks for evaluation: A comprehensive review of HCI usability methodologies conducted by Wanderly and Orio (2002) reveals the importance of testing within well-defined contexts, using discrete, simple musical tasks for evaluation. Dewey and Wakefield (2014), recommend that when evaluating new user interfaces for mixing, complex actions should be broken

down into a series of simple tasks, as it can aid the design process by clarifying the steps needed by a user to achieve their goals. In line with these findings, all the studies reported herein break down the potentially complex and open-ended process of mixing into defined subtasks and actions (Crystal and Ellington 2004).

3.2.2 Experimental Approach

Both Repeated Measures and Independent Measures test designs have been used in the studies reported herein. Repeated Measures tests were used when the participants were required to use all the experimental interfaces being assessed. This helped to reduce the variability between participants (Field & Hole 2003), such as when the participants came from different populations or had varying levels of experience. When using Repeated Measures, the order in which interfaces were presented was alternated for each participant to ensure the influence of listening fatigue, continued use of the interface, and continued exercise of critical listening were accounted for.

Independent Measures were used when participants were required to use only one of the experimental interfaces. When running Independent Measures tests, all subjects were taken from the same population and had an equivalent amount of exposure to mixing using DAWs. This was done to compensate for potential variations in participants. Furthermore, in the Independent Measures tests, random assignment was used to assign participants to the different condition groups.

3.2.3 Quantitative Data Analysis

In analysing the quantitative data, we have used descriptive statistics, graphical analysis and inferential statistics. Descriptive statistics, such as average, mode and standard deviation have been used to describe the basic features of the data in the studies. Descriptive statistics have been used to examine the frequency and distribution of each variable and identify any outliers in the data.

Hypothesis testing has been used to produce conclusions regarding any observed difference: either that the difference is statistically significant or that it is statistically insignificant. Confidence intervals (CI) have been used to show what effects are likely to exist in the general population with similar levels of mixing experience, allowing us

to make inferences from the sample data. In line with convention, CIs were set at 95% in all the studies, unless otherwise stated. Paired and unpaired *t*-tests have been used to compare data from the different conditions (different interface designs). Unpaired *t*-tests have been used to examine the results of the data when different participants were testing different interface designs (Independent Measures). Z-tests for proportions have been used to test the results of interface evaluations and ascertain whether populations differed significantly between interface designs.

3.2.4 Qualitative Data Analysis

While traditional research in usability lends itself well to task-based evaluation (Jeng 2005.; Dubey and Rana 2010) usability extends beyond traditional definitions of goals to more complex interactions (Beauregard and Corriveau 2007). This is especially relevant with software for music mixing, as this is a creative, as well as technical undertaking. For this reason, qualitative analysis has been used to assess participants' preferences, feelings and acceptance of the software interfaces being used. Specifically, the following methods were used:

3.2.4.1 Questionnaires

Questionnaires were used to collect the participants' opinions regarding the interfaces used and to gather demographic and experiential information. Questionnaires were also used to collect usability satisfaction levels and quantify responses to several criteria including mental demand, temporal demand, performance, effort and engagement levels.

3.2.4.2 Interviews

In all the studies, opportunities were afforded to the participants to expand on the responses given in the questionnaires in the form of a semi-structured interview. These provided an opportunity to better understand users' perceptions (Flick, 2009) and helped reveal any opinions and preferences not covered within the confines of the questionnaires. For example, after all studies, participants were asked individually (by the author) if they had any further comments on the interfaces in terms of personal preferences, effectiveness of task completion, aesthetic considerations, interface design or learnability. Interviews were voluntary and not all participants contributed further comments.

3.3 Participants

3.3.1 Recruitment and Ethical Considerations.

Participants with experience of audio mixing using DAW interfaces have been recruited for the studies, with a strong emphasis on beginners, students and non-experts. We recruited participants from several music technology courses, namely, City and Islington College, Camden School for Girls, the College of North East London and the Centre for Digital Music at Queen Mary University of London. Where staff were used in addition to pupils, these were non- music technology staff within the colleges and schools, who self-classified themselves as beginners, giving them a broadly similar mixing background to the students recruited. Non-probability sampling was used throughout these studies to recruit participants. This is a method where participants are chosen according to predetermined criteria (Wilson & MacLean, 2011, p. 163). In the case of these studies, the criteria were that participants should have experience of audio mixing in general and DAW use in particular, and not be an expert or professional producer. This recruitment strategy aligned with our Purposeful Sampling process, due to the participants' interest, experience and background in audio mixing. For all the studies, we aimed for as many participants as possible for the studies, setting eight as a minimum target. However, for reasons of economy and pragmatism, such as response rate from volunteers, some experiments had smaller sample sizes than others. We conducted analysis on the data from the sample sizes recruited, and in all cases stopped appealing for and recruiting further volunteers once the data provided significant results.

For all the user studies in this thesis, we received approval from the Research Ethics Committee at Queen Mary University of London. Approval was also received from the Curriculum Managers and Heads of Department at City and Islington College, Camden School for Girls and the College of North East London. Participants were given detailed information regarding the studies. All participants gave informed consent to take part in the studies and they were informed that they could leave the study without giving any reasons and without any consequences for them.

3.4 Specific Experiment Details

3.4.1 Experiment One

3.4.1.1 Experimental Methods

As the participants who took part in this experiment came from different populations (students at the College of North East London, and students and researchers at the Centre for Digital Music at Queen Mary University of London), this experiment used a Repeated Measures design to reduce issues of Sensitivity (Field & Hole 2003), reduce the variability between participants and minimise the variation in scores between conditions due to non-experimental factors (ibid).

3.4.1.2 Experiment Tasks

Participants were required to equalise six individual audio files to match as closely as possible a pre-equalised reference version of the same file, using a two-band equaliser. Participants undertook the equalisation tasks using two interfaces, one which displayed visual information for the equalisation curve, the other which had this visual information blacked-out. Three audio files were presented for each interface, with their order alternated for each participant.

3.4.1.3 Data Analysis Methods

This experiment used a paired *t*-test to compare data from the different conditions (hidden and visible interfaces). This was chosen as we tested different interfaces on the same participants using Repeated Measures designs. As the paired *t*-test reduces inter-subject variability (because it makes comparisons within the same subject), it is useful for testing any change or difference in means between the two experimental conditions.

When analysing the data, the frequency modifications for both frequency bands were extracted from the equalisation curves set by each of the twenty-four participants under each condition (hidden and visible). The data was analysed and the mean error across frequency bands were calculated. The distribution of errors across frequency selection and gain adjustment were plotted as a frequency spectrum for each condition. The time taken for the completion of each file's equalisation was also calculated for each condition.

3.4.2 Experiment Two

3.4.2.1 Experimental Methods

As the participants who took part in this experiment came from different populations (music technology students at City and Islington College and Camden School for Girls), this experiment used a Repeated Measures design.

3.4.2.2 Experiment Tasks

Participants were played an excerpt of an eight-channel mix. They were asked to listen to three specified instruments from the mix to ascertain which of these instruments was being gradually panned during the extract. At the same time as undertaking this listening task, they were asked to visually match the frequency curves of a four-band equaliser (the target) with a pre-equalised four band equaliser (the source) using four interface designs with varying levels of visual feedback and interface navigation. These consisted of the control interface, which only displayed response buttons (allowing participants to select which instrument had panned), an interface which included source and target equalisers, another which included the addition of meters, and a fourth which required scrolling to view the source and target. The excerpt was played twelve times in total, during which each of the specified instruments was panned three times. The four interfaces and panning file types were arranged in a randomised order. The time it took to successfully compare the channels was timed.

3.4.2.3 Data Analysis Methods

The time taken to correctly identify panning was compared between the four interface types. As all three of the specified instruments were panned in each of the interface types it was possible to directly compare the response times for each instrument across interface types. The mean time and standard deviation was calculated for the response times of all the interfaces and file types. A paired *t*-test was then conducted between the control interface and the independent variable interfaces (at 95% CI). The paired *t*-test generated a *p* value, where values of 0.05 or less reject the null hypothesis (that the interfaces design does not have any effect on critical listening skills).

3.4.3 Experiment Three

3.4.3.1 Experimental Methods

The participants who took part in the experiment came from different populations (staff and students at City and Islington College, London), therefore a Repeated Measures design was used.

3.4.3.2 Experiment Tasks

The participants were required to listen to a two minute, eight-channel mix and identify which of the three specified instruments was gradually decreased in volume over the course of the excerpt. At the same time, participants were presented four 16 channel mixers. Each mixer used different UI object designs to represent the mix parameters. These consisted of dials, faders, numbers and colours. All the interface designs required scrolling to view the requisite information. Participants were asked to look at channel one and compare the subsequent 15 channels to ascertain if they were the same or different (by clicking on a 'same/ different' button below channels 2-16) while listening to the audio. The order of the interfaces and the attenuated instrument was randomised for each participant. The time taken to compare the channels was timed.

3.4.3.3 Data Analysis Methods

The time taken to correctly identify the attenuated audio in each interface design was analysed for each participant. From this the mean and standard deviation was calculated and used to generate confidence levels (at 95%) showing the range of the true population. Secondly, the amount of UI objects compared for each of the four interface designs were calculated. The amount of correctly matched channels was used to generate a mean and standard deviation to produce confidence levels, again at 95%. As all three of the specified instruments were attenuated in each of the interface types, it was possible to directly compare the response times and channel matching for each instrument across each interface type.

3.4.4 Experiment Four

3.4.4.1 Experimental Methods

As the participants who took part in this were recruited from different populations (students and staff at City and Islington College, London), this experiment used a

Repeated Measures design. A Questionnaire was given at end of the study to ascertain which of the interfaces individual participants felt provided the best results and experience.

3.4.4.2 Experimental Tasks

Three versions of a 24-channel mixer showing volume and pan-position were presented to the participants. The designs consisted of a channel strip (CS) mixer with all 24 channels shown on a single page, a stage metaphor (SM) mixer presented on one page, and a CS mixer requiring scrolling navigation to view all 24 channels. For each of the interface designs, participants were asked to answer four questions about the visual information and select the correct answer from a drop-down menu above each interface. While doing this, participants were played a twelve-channel audio mix. Each time the excerpt was played, four of the instruments within the mix were randomly panned either left, right or centre. Participants were asked to select the correct pan position of two of the instruments (chosen at random) from a drop-down menu with the categories; 'left, centre, right or couldn't tell' after the audio had finished. Participants were presented four occurrences of each interface type (with the order randomised for each participant). The mix was played twelve times for each participant (corresponding to the twelve visual interfaces).

On completion of the study, each participant was asked to comment on how they perceived their performance in each interface type in terms of their success in locating the visual information and correctly detecting the audio panning.

3.4.4.3 Data Analysis Methods

The amount of correctly answered visual and auditory questions (and the time taken to answer) were recorded and analysed for each participant per interface type. From this, the mean and standard deviation was calculated for the three interface types. The mean and standard deviation generated confidence levels (at 95%) showing the range of the true population per interface type. A comparison of time and accuracy by question type was also analysed to quantify whether the various interface designs supported visual search better than others.

The results of the questionnaire were analysed to quantify the proportion of participants who perceived their performance as being better while using each interface design. This was done by collating the responses and creating bar charts of preferences to rank interfaces.

3.4.5 Experiment Five

3.4.5.1 Experimental Methods

As the participants who took part in this experiment were from different populations (students and staff at City and Islington College, London), this experiment used a Repeated Measures design. Following the study, participants were given a questionnaire to rate their experience of using the various interfaces. The questions were designed to assess their levels of comfort, perceptions of task completion and their success using the different designs.

3.4.5.2 Experimental Tasks

Participants were presented three interface designs of a 24-channel mixer showing each channel's volume and pan-position. For each design, a version with and without dynamic query filters (DQ) was included (creating six interfaces in total) so that the influence of DQ filters could be analysed for each design. The designs consisted of a CS design with all 24 channels shown on a single page, a CS mixer where scrolling navigation was required to view all 24 channels and a SM mixer presented on one page without the need to navigate. For each interface, a series of questions about the visual display was included on the screen. Participants were required to search for the relevant information and select the correct answer from a drop-down menu. Each question was asked in each interface design with the question order randomised for each participant. At the same time, participants were played a twelve-channel audio mix. Each time the excerpt was played three of the instruments within the mix were randomly attenuated by 6dB. As soon as the excerpt had finished playing, the interface which the participants were using was automatically closed, and they were asked to select which instrument had been attenuated from a drop-down menu.

Following the experiment, participants were given an adapted NASA Task Load Index questionnaire (Hart and Staveland 1988). This was used as a subjective assessment tool to rate their perceived workload and interface effectiveness. Questions asked included;

- Which interface did you feel most comfortable (least stressed/ rushed) using?
- Overall how much did having the DQ sliders help in each interface design?
- Which interface do you think helped you do the listening task best?
- Which interface do you think helped you do the visual task best?

3.4.5.3 Data Analysis Methods

The amount of correctly identified visual searches was analysed for each participant per interface type. From this, the mean and standard deviation were calculated for the participants' responses in the six interface types. These were used to generate CI, at 95%, showing the range of the true population per interface type. The amount of correctly identified file attenuations were analysed for each of the thirteen participants. This was used to calculate the percentage of correct answers per interface type. A z-test for proportions-dependent groups was used to determine if the percentages of correct answers from the six interfaces were significantly different from one another.

The responses to the questionnaire were analysed to rate the interfaces according to user preferences. This was done by entering the responses into a spread sheet, and calculating how many people selected each response.

3.4.6 Experiment Six

3.4.6.1 Experimental methods

Participants comprised students on a two-year music technology course at City and Islington College, London. Due to the range of experience between the participants (1st and 2nd year students), a Repeated Measures test design was used.

3.4.6.2 Experimental Tasks

Participants were presented five, eight-channel mixers; a CS design and four SM mixers, showing each channel's volume, pan, reverb and delay amounts. For the CS design, faders were used for volume, while dials were used for the pan position and the reverb. For the SM, x and y positions were used for the pan and volume, while five designs were used to represent the reverb: size, transparency, saturation (single colour) and hue (multiple colours). The effects' ranges (1-100) were divided into increments of five, ten and twenty values. For each of the five mixer designs, a target was included in the eight channels and placed within a border. For each design three screens were created; one with effects differences between the target and other channels set at +/- 5 (increments of 5), one with differences set at +/-10 (increments of 10) and one with differences at set +/-20 (increments of 20). This created a total of fifteen screens for the study. For each screen participants were asked to identify which of the other channels on the mixer had the same value as a target channel by clicking on the corresponding channel. The

screen order was randomized for each participant and they were presented one after the other.

3.4.6.3 Data Analysis Methods

The amount of errors (incorrectly identified channels) was calculated for each participant in each of the fifteen interfaces at the three increment levels between target and other channels. From this the total number of errors made on each screen by all participants could be calculated in each interface design. The error rates found between the different mixers were analysed using a z-test for proportions dependent groups at 95% CI to test if difference between designs was significant.

3.4.7 Experiment Seven

3.4.7.1 Experimental methods

Since all the participants who took part in this experiment had comparable experience of DAW mixing, and equivalent training in audio mixing (all were first year students on a music technology A level course), an independent measures test was used. Participants were assigned to each of the three groups using random assignment.

3.4.7.2 Experimental Tasks

Participants were presented one of three eight-channel mixer designs, showing each channel's volume, pan position, reverb amount, treble and bass levels. The designs consisted of a CS mixer, an SM mixer using multivariate UI designs and DQ filters, and a hybrid design, incorporating all the features of the stage metaphor design with the inclusion of dials to show parameter values. Participants were required to undertake six specific mixing tasks, including changing volumes, equalising, panning, and muting specified channels. The mixing tasks required varying amounts of visual search. Two questions required analysis of one parameter (e.g. volume), two required analysis of two parameters (e.g. bass amount and pan position) and two required analysis of three parameters (e.g. volume, pan and reverb amounts). The order of these questions was randomised. At the same time as completing these mixing tasks, one of the audio channels within the mix had a Low Pass filter (reduction of treble frequencies) applied to it for a short duration. Participants were required to identify which instrument had been affected by selecting the correct instrument from a list provided on the screen.

3.4.7.3 Data Analysis Methods

The times taken to correctly complete the mixing tasks were analysed for each interface design. From this, the mean time and standard deviation were calculated. This was used to provide CI (at 95%). This allowed us to both ascertain which interface was quickest, and quantify the influence of increased visual search on task completion. The amount of correctly completed mixing tasks and correctly identified audio changes was also analysed for each interface as a percentage. This data was used for a z-test for proportions to see if the difference was significant between interface designs in terms of visual search complexity and aural acuity. Finally, any participants' comments or feedback given following the tasks was recorded. These responses were then analysed to collate any common themes, and identify patterns and trends that may have been voiced.

3.5 Summary

In this chapter, we have described our strategy and provided an overall picture of how the various experiments described herein form a coherent body of work to address the main research questions. We have provided a rationale for the experimental tasks, experimental approach, participant recruitment and analysis of the data accrued. Finally, in Section 3.4 we put this into context, by describing in finer detail the experimental approaches, tasks and analysis methods used in each of the seven studies. In the following chapters (four to ten) we go on to detail each experiment further and continue to relate them to both our research strategy and research questions outlined in chapter one.

Chapter 4

Visual Feedback and Critical Listening.

For this first experiment we have purposefully designed a broad, high-level investigation into bi-modal attention while mixing. This is done in the anticipation that the knowledge gained will provide a starting point for further studies, set future directions for interface design and crucially, begin to address our first research question; namely to what extent does visual perceptual load interfere with or support the processing of auditory information?

When mixing using analogue mixing desks, engineers will often adjust the equalisation without looking at details of the control. Instead, they engage in a proprioceptive action, in which they listen while moving the mixing desk dials and faders (Valeriotte 2016). However, when applying equalisation using DAWs, the visual feedback cannot be so easily ignored (Bell and Ratcliffe 2015). Since there is no tactile response (as turning a dial on a physical mixer) users must engage with the pictorial representation of the equaliser, requiring an increased level of visual engagement (ibid). As discussed in section 2.2.2, there is contrasting evidence regarding the influence of increased visual attention on aural modalities. Some researchers maintain visual stimuli can strongly modulate perceptions of auditory events and inhibit auditory perception and aural acuity (e.g. Colavita 1974., Sinnet et al 2007., Soto-Faraco & Spence 2010) while others posit the notion of independence of attentional resources for vision and audition (Triesman and Davies., 1973., Alais et al., 2006., Santangelo et al 2010). Though there is little published research on the effect of visual load on mixing tasks, a related study by Dehais et al (2012) found a link between complexities of aeroplane cockpit GUIs and reduced aural awareness. In flight simulations, 57 % of the participants failed to notice auditory alarms under high visual load conditions. They suggest that visual information processing interfered with concurrent appraisal of auditory alarms, thereby inducing 'Inattentional Deafness' (Macdonald & Lavie 2011). In order to ameliorate the effect of visual overload they suggest a temporary simplification of the user interface, a process they referred to as Cognitive Countermeasures (Dehais et al 2012). Similarly, mix engineers will often turn off the screen whilst listening to the mix in order to minimise potential visual distractions (Battino and Richards 2005). Indeed, Lech and Kostek

(2013) state that ‘visualizing audio parameter values can affect the decision process during sound mixing’ (ibid, p. 311).

Quantifying the extent to which visual feedback affects mixing decisions may therefore be of benefit to the research questions outlined in this thesis. This is especially the case given the tendency for DAW GUIs to become more visually complex with each subsequent version (figures 4.1 and 4.2). In this chapter, therefore, we present our first experiment; a GUI design study which examines single channel equalisation tasks with two levels of visual feedback- one condition in which the equaliser is visible, and the other in which the equaliser is hidden (blacked-out).

4.1 Study Design

4.1.1 Participants

There were 16 students recruited in total. Nine (9) Participants were recruited from Music Technology students at the College of Haringey, Enfield and North East London (population 1) and seven (7) students and researchers from the Centre for Digital Music at Queen Mary University, London (population 2). Participants were asked to give their age, (the means were 28.6 for population 1, and 31.9 for population 2) and classify their prior experience of music equalisation as beginner, intermediate or expert. The data from the participants show that the age and experience from both populations were broadly similar, with both groups self-classifying themselves as Novices (population 1: 62%, population 2: 56.25%) or Intermediate (population 1: 37.5%, population 2: 43.75%). As the participants came from different populations, we used a Repeated Measures design for the study, with all participants using all the experimental interfaces. This was chosen as we wished to control for factors that may cause variability between the various participants as they were drawn from two separate populations.

Population	Mean age	% Expert	% Intermediate	% Novice
1	28.6	0	37.5	62.5
2	31.9	0	43.75	56.25

Table 4:1: Population Comparisons. Both populations are broadly similar in terms of experience using DAWs for mixing



Figure 4:1: Pro Tools equaliser, version 5.1. Except for the sliders, there is limited visual feedback in the GUI design.



Figure 4:2: Pro Tools equaliser, version 6.9 onwards. The GUI is more complex, with colour coded dials and numeric information in addition to a graphical representation of the frequency curve.

4.1.2 Procedure

Participants were required to equalise six separate audio files to match as closely as possible a pre-equalised reference version of the same file. These consisted of two full mixes, two drum loops and two solo vocals. Participants undertook the equalisation using a pictorial equaliser interface designed using Max/MSP (Cycling '74), a paradigm

which is commonly utilised in DAWs. Participants were presented with two interfaces. Interface 'a' displayed visual information for the equalisation curve including gain and frequency, and Interface 'b' had this visual information blacked-out (figure 4.3). Both the equalisers were dual band with a frequency range of 0-20,000 Hz, a gain boost/reduction of +24 dB and a fixed bandwidth.

Participants were instructed to match the reference files as closely as they could. They could listen to the reference files as many times as they wished, and were not put under any time constraints (though timing data was captured, this was hidden from the participants and they were not told they were being timed). Three audio files were presented using interface a, and three were presented using interface b. The order in which either interface was presented first was alternated for each participant and the six audio files were rotated to ensure the influence of listening fatigue, continued use of the interface and improvement were accounted for. Participants were provided with training in how to use both equalisers and only moved on to the task when they felt confident in using them.

Each participant began with a different audio file, and each audio file was used in both the hidden and visible interfaces. Furthermore, the audio files were evenly distributed between interface a and interface b, so for each interface the participants were required to equalise a full mix, a solo vocal and a solo drum loop.

4.2 Analysis and Results

In the analysis for the study we used a Type 1 Anova analysis to compare the data from the different conditions (hidden and visible interfaces). As there were multiple occurrences of these (the different mixes within the hidden and visible conditions) the Type 1 Anova was chosen to assess the equivalence of mean levels of participant's equalisation results across the six interface designs, showing them as pair-wise comparisons between the hidden and visible interfaces for each of the equalised audio files.

The frequency modifications for both bands were extracted from the equalisation curves set by each of the sixteen participants under each condition (hidden and visible). The data was analysed and the mean error across frequency bands was calculated. Regardless of condition, the differences between the mean errors appeared marginal. A

Type 1 Anova analysis confirmed this impression (figure 4.4). It showed no significant difference between the mean error for either condition (visible/ hidden interface) with p values for both conditions being above chance (>0.05)

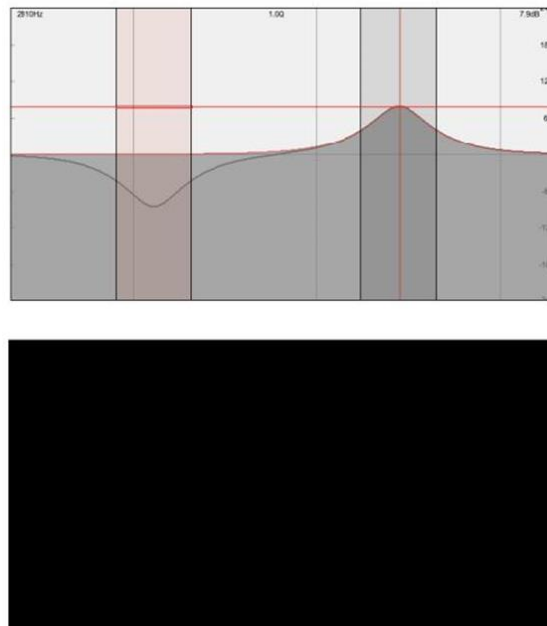


Figure 4.3: Top, Interface 'a'. Bottom, Interface 'b', which had the screen blacked out. Participants used a mouse to adjust the two bands (in terms of frequency and gain). For the blacked-out version, the cursor turned into a cross when the mouse was on a band.

The time taken for the completion of each file was also calculated for each condition. Again, the plotted data shows the mean time between either condition being marginal, and once again a Type 1 ANOVA analysis confirmed that there was no significant effect in terms of visible or hidden interface. The results show that there is no statistically significant difference between the two approaches, and that in this study, having an interface with quantitative visual feedback neither increases the accuracy or speed of the equalisation process, nor reduces aural acuity or time taken.

4.3 Discussion

The fact that there was no reduction in accuracy of audition when monitoring the visual

feedback is in keeping with research suggesting monitoring visual and auditory streams simultaneously does not incur any perceptual or attentional cost (e.g. Larsen et al. 2003., Alais et al. 2006., Arrighi et al. 2011). However, literature specific to DAWs suggests that the heavy use of complex visual detail can overload a user's visual bandwidth (Htalky et al 2009).

The failure to demonstrate cross-modal perceptual load may be due to our study simply not using sufficient manipulation of perceptual load. Indeed, in the use of DAWs, the user is expected to take in a large amount of visual information, while simultaneously scrolling and window switching. Furthermore, by not loading the visual bandwidth in a manner akin to DAW mixing, the participants may have had the ability to time-share, so that visual and auditory monitoring effectively became a single task (Johnson & Proctor 2004). Indeed, Molly et al (2015) suggest that experiments which don't show an effect of visual load on concurrent aural acuity may be due to the fact they do not require consistently focused attention to the visual task, and offer the opportunity for task switching to the auditory stimuli.

Another consideration from the current study is that the visual stimulus was never absent. Though in the 'hidden' interface trial the equalisation curve and numerical feedback were hidden, the position of the mouse was still clear within the equalisation rectangle, with the cursor turning into a cross when the mouse was positioned over a frequency band. This leaves open the question as to how much of the participants' focus was on the auditory information and how much was on the visual feedback still present in the interface. Furthermore, there was a proprioceptive modality taking place during both conditions involving the up/down, left/right movement of the mouse. The presence of this third modality can again make it difficult to assess accurately the separate aural/ visual reaction to the trial. Lastly, as raised by Molly et al (2015), the participants may have been able to switch between the sonic and visual content.

4.4 Conclusion

In conclusion, this study has found that for simple single channel equalisation tasks, the use of a visual representation of frequency changes does not minimise the attentional resources given to the auditory response, though in the case of the current study this may be due to lack of strong enough perceptual load on either the visual or auditory modalities. Furthermore, the use of a visual representation did not improve accuracy or

speed, suggesting that visual feedback per se is not necessarily an aid to mixing.

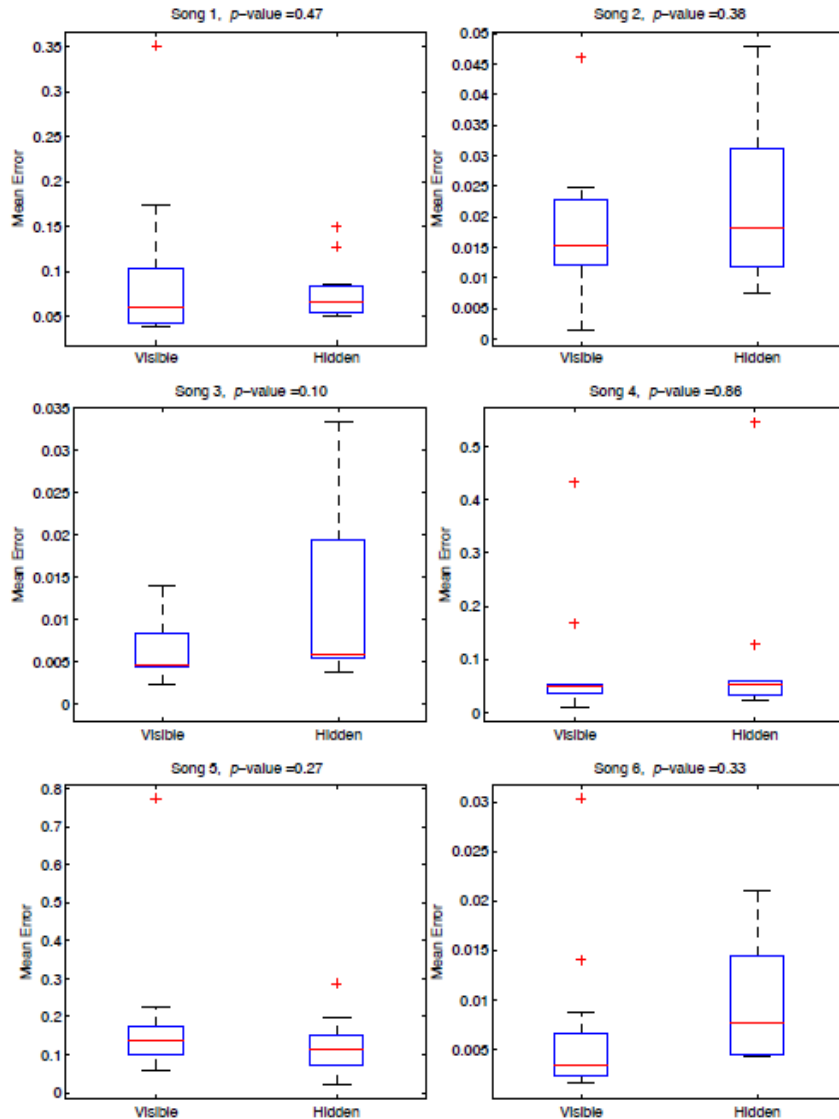


Figure 4:4: Mean errors and p-values for visible and hidden interfaces.

4.5. Progression to Subsequent Study

In our next study we wish to better understand the reasons for our null result. Specifically, we wish to clarify whether it is due to human perception per se, or whether the interface used by the participants may have been a contributing factor. Indeed, even on moderately small mixes, a typical DAW such as Logic or Cubase will present a large amount of visual information, far surpassing the level of visual load found in the current study's interface designs. This visual feedback includes multiple channel strip UI elements such as dials and faders, and moving meters, including channel volume and

frequency spectrum analysers. In order to replicate the workflow encountered when using such software displays, it may be necessary to introduce more visual load, thereby making the experience more comparable with actual use of a DAW. Accordingly, our next chapter introduces a study where mixing interfaces with increased amounts of channels, visual feedback and interface navigation are assessed. By so doing we hope to better clarify whether these factors play a part in participant responses to audio changes during interface interaction.

Chapter 5

Visual Load in Mixing GUIs.

The results outlined in the previous chapter suggested that visual load does not necessarily impact either negatively or positively on simple audio mixing workflow. However, we concluded that this may be because the interfaces used in the previous study did not accurately reflect the amount of information that a DAW user would typically interact with. MacDonald and Lavie (2011) found evidence that ‘Inattentional deafness’ was ‘clearly influenced by the level of visual perceptual load in the task’ (p. 1785), and the subjective experience of noticing a sound depends on the visual perceptual load in the task being undertaken (ibid p.1786). Furthermore, Molly et al (2015) suggest that experiments which don’t show an effect of visual load on concurrent aural acuity may be due to the fact they do not require consistently focused attention to the visual task, and offer the opportunity for task switching to the auditory stimuli- a feature we believe may have been present in our previous study.

In DAWs, there are commonly several channels in a mix, with dynamic visual feedback, such as peak and VU meters on each one, which were lacking from our previous experiment. We must also consider the influence GUI navigation on mixing workflow- another element that was not included in the previous experiment. In creative terms, the need to navigate through several windows risks inhibiting the engagement and ‘flow’ of the mixing process. It may impede the user’s ability to quickly respond to the programme material and make requisite adjustments such as pan, level and other effects changes (Szalva, 2009, p.10) and may compromise the realisation of creative ideas, which due to their fleeting nature are ‘lost’ when the user must navigate the GUI (Tano et al 2012). For example, end-users have commented on the excessive amounts of navigation required to adjust channel controls on tablet DAWs and the distraction this causes (Janney 2012) while others have voiced concerns over the disparity between the ever-increasing amount of information displayed and the lack of quick and useful access to parts of the screen space (Golkhe et al 2010). As navigation in general, and scrolling navigation in particular are commonly employed in DAW GUI design, quantifying the extent to it may affect concurrent critical listening forms a key stage in developing heuristics for more efficient screen-based mixing. This is especially germane given the

increased use of DAWs designed specifically small screen mixing (such as Cubasis, Auria, Nanostudio and FL studio mobile).

In this chapter, therefore, we seek to address the gap between the visual load presented in chapter one and the visual load found during DAW mixing. We compare interfaces with and without scrolling navigation to access mix information, we increase the number of channels from two to eight and we include interfaces with moving distractors (Sanabria et al 2005) in the form of peak meters and equaliser spectrum analysers, both features of DAWs.

5.1. Study Design

5.1.1 Participants

There were sixteen participants recruited (six from Music technology students at the Camden School for Girls, and ten from second year 'A level' Music Technology Students at City and Islington College, London). All participants had at least one year's experience mixing on DAWs and one year undertaking a formal Music Technology course (Edexcel Music Technology GCE A2 level). Although all participants were undertaking the same Music Technology course and were broadly the same age, due to the recruitment of participants from two different educational establishments, we used a repeated measures design for the experiment (where all participants used all software interfaces). This was done in order to control any non-experimental factors arising from the different populations. None of the participants had taken part in the previous study.

5.1.2 Procedure

Participants were played an excerpt of an eight-channel mix that they monitored on headphones. They were asked to listen to specified instruments from the mix (strings, guitar and tambourine) to ascertain which of these instruments was being panned (changing the apparent position of the sound between the headphone speakers). All files began panned centrally (pan position 0) and one of the three specified files was gradually panned over the duration of the excerpt (two minutes) until it was panned hard left or right (pan position -60 or +60). The participants were asked to respond to the panning by pressing one of three panning response buttons (labelled strings, guitar or tambourine) as a timed response task. The excerpt was played twelve times in total,

during which each of the specified instruments was panned three times. At the same time as undertaking this listening task they were presented four interfaces (figures 5.1-5.4) displayed on a 10” by 5.8” screen (these dimensions are directly comparable with commercially available tablets such as the Apple iPad, Galaxy Nexus and Microsoft Surface). The participants were asked to visually match the frequency curves of a four-band equaliser (the target) with a pre-equalised four band equaliser (the source) so that the target and source frequency curves were as visually close as possible. The equaliser used was a four-band parametric (a design very commonly used in DAWs) created using the Max/MSP ‘filtergraph’ object. The interfaces comprised the following visual information:

- **Control interface:** This consisted of a play button and three response buttons labelled guitar, strings and tambourine. There was no source or target equaliser, and the participants were not required to complete any interface manipulation task during the excerpt other than selecting a response button.
- **Interface 1:** This consisted of a play button, the three response buttons and the source and target equalisers.
- **Interface 2:** This consisted of a play button, the three response buttons, a source and target equaliser and three moving meters (a gain meter, a phase meter and a frequency analyser) placed between the source and target.
- **Interface 3:** This consisted of a play button, the three response buttons, the source and target equaliser as well as five additional equalisers placed between them. Due to the additional equalisers the source and target equalisers did not fit on the same screen and participants were required to scroll between them.

Participants were asked to begin matching the source and target as soon as they pressed the play button, but were informed they could stop at any point at which they clarified which instrument was panning, even if they had not completed matching the target equaliser curve to the source curve. Prior to the study, participants were given a test screen so they could accustom themselves with manipulating the equaliser. The order in which the interfaces were presented was alternated for each participant and the audio files were rotated to ensure the influence of listening fatigue, improvement occurring due to continued use of the interface and improvement due to continued exercise of critical listening were accounted for.

The time it took to respond to the panned file was recorded for each interface. Due to

the increased aural acuity required to hear small panning amounts and the potential distraction of visual feedback, it was hypothesised that interfaces which impact negatively on critical listening would result in participants taking longer to hear the panning (which becomes easier to identify at extremes).



Figure 5:1: Control interface only displays response buttons, labelled ‘strings’, ‘guitar’ and ‘tamb’ (tambourine).



Figure 5:2: Interface 1 includes source and target equalisers, no scrolling is required

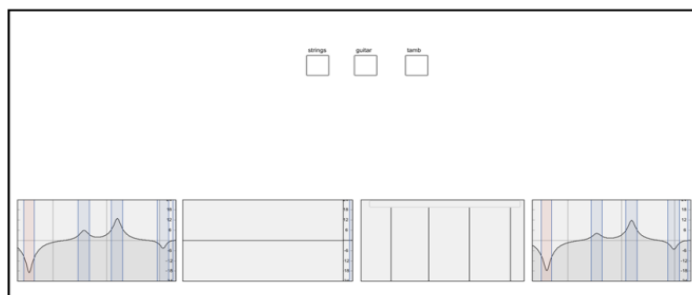


Figure 5:3: Interface 2 includes the addition of meters; again, no scrolling is required

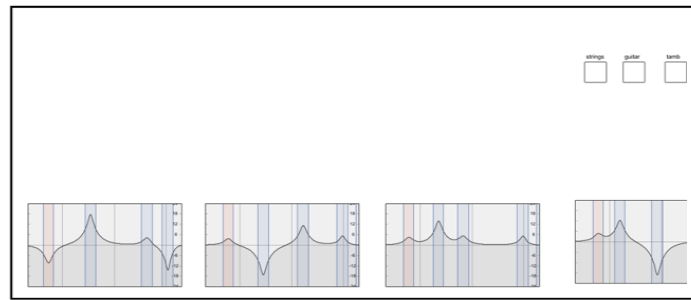


Figure 5:4: Interface 3 includes the additional EQs. Scrolling is required to see all the equalisers.

5.2. Analysis

As the analysis compared each experimental design interface against a control (Control Interface) we used a paired t -test to directly compare the response times for each instrument across each interface type. This enabled us to compare results between the control interface and the independent variable interfaces using Confidence Intervals (at 95%) to see if the results from the participants was representative of the larger population. Our Analysis method for this chapter also allowed us to generate a p value, where values of 0.05 or less reject the null hypothesis (that the interfaces design does not have any effect on critical listening skills). This allowed us to quantify whether differences in results were due to chance or the design of the GUI.

Of the sixteen participants recruited, two were discounted due to incorrectly identifying some of the panning instruments, one was discounted due to an inability to clearly hear the panning instruments within the mix, and a further participant was discounted for failing to attempt matching the source and target equalisers. Of the twelve remaining participants (eight from City and Islington College and four from Camden School for Girls), the time taken to correctly identify panning was compared between the four interface types. As all three of the specified instruments (tambourine, guitar and strings) were panned the same amount of times in each of the interface types, it was possible to directly compare the response times for each instrument across interface types. The mean time and standard deviation (SD) was calculated for the response times of all the interfaces and file types. A paired t -test was then conducted between the control interface and the independent variable interfaces at 95% confidence level (CI). The paired t -test generated a p value, where values of 0.05 or less reject the null hypothesis (that the interface's design does not have any effect on critical listening skills). This

allows us to see if any differences in results between the two conditions occurred by chance, thus suggesting whether the interface design may or may not be an important factor.

5.3. Results

While Interfaces two (non-scrolling) and three (moving distractors) had slower response times across all three of the specified instruments compared to the control, none of these were statistically significant, with p values from the paired t -tests being greater than 0.05 (table 5.1). However, there were significantly slower response times for all three instruments in interface three (requiring scrolling) compared to the Control interface. The paired t test consistently generated p values less than 0.05, thereby rejecting the null hypothesis at the 95% confidence level (only 5% of the time would the statistical process produce a finding this extreme if the null hypothesis were true). The time difference between the Control and the interfaces was also calculated to discern how the interface affected the speed to complete the task (figures 5.5-5.7). The analysis (table 5.2) shows that interface 4 (at 95% CI) had a range for the true population mean that is greater than the Control across all the three file types. The analysis also revealed that overall, the Control provided the fastest response for most participants on all file types (overall being the quickest interface 58 % of the time), while interface 3 provided the quickest response only 4% of the time.

	Control to interface 1	Control to interface 2	Control to interface 3
Guitar file	$P=0.120$	$P=0.261$	$P=0.033$
Strings file	$P=0.308$	$P=0.070$	$P=0.047$
Tamb file	$P=0.701$	$P=0.514$	$P=0.040$

Table 5:1: The P values for time difference between Control and interface types.

Values < 0.005 are significant.

File type	Interface type	Mean (sec)	S.D. (sec)	Confidence Interval
Guitar	Interface 1	9.83	18.33	-0.54 to 20.2
	Interface 2	9	20.88	-2.81 to 20.81
	Interface 3	16	19.12	5.18 to 26.82
String	Interface 1	7.25	17.03	-2.39 to 16.89
	Interface 2	12.08	15.90	3.08 to 21.08
	Interface 3	13.33	10.59	7.34 to 19.32
Tambourine	Interface 1	3.58	19.58	-7.5 to 14.66
	Interface 2	4.83	22.68	-8 to 17.66
	Interface 3	17.41	22.65	4.59 to 30.23

Table 5:2: The time difference for task completion between Control and interfaces.

5.4 Discussion

In accordance with the findings from the experiment in chapter four, the analysis of the data suggests that increased visual load by itself does not have a statistically significant effect on reaction time to the critical listening test, though it is worth noting that the Control interface (where there was no visual feedback) had the quickest reaction time across all the files. However, introducing scrolling navigation to the interface design did have a significant negative effect on critical listening reaction times.

Literature on the use of scrolling navigation in GUIs may supply an explanation for this reduction in aural acuity when using the navigation interfaces in our study. For example, research by Piolat et al (1997) and Feinberg and Murphy (2000) have shown significantly higher levels of mental load when the use of scroll bars are included in an interface, since users must keep one information element active in Working Memory (WM), while searching for the other. Related research by Harms et al (2015) supports this finding. In a study where scrolling on small screen devices was compared to other navigation techniques (tabs, menus and collapsible field) it performed significantly worse than the other navigation techniques in terms of users' ability to remember GUI content during navigation.

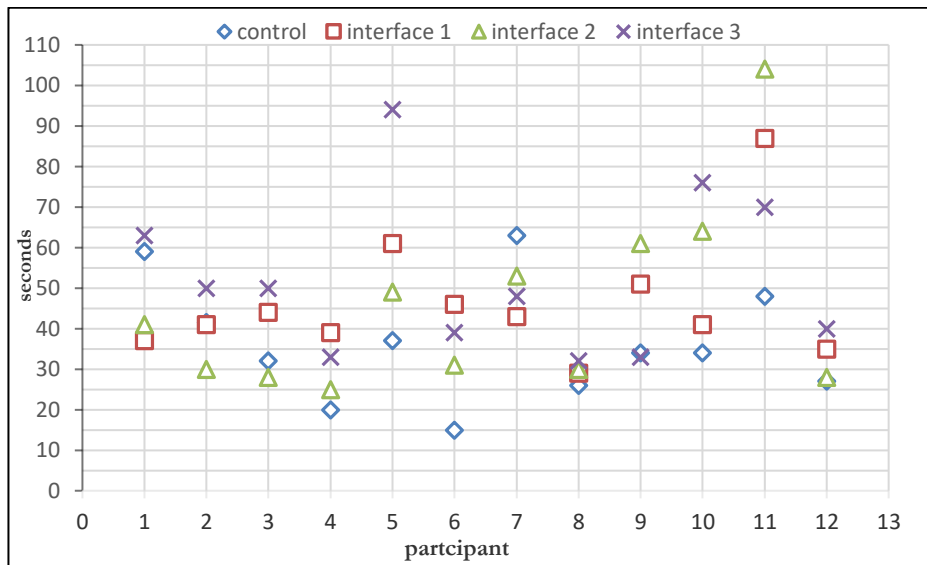


Figure 5:5: Times for completion per interfaces for guitar file.

Relatedly, literature suggests that there is a link between WM capacity and critical listening. For example, Alho et al (1992) and Berti & Schroger (2003) have found that when users continuously maintain visual information in WM, attentional resources cannot be effectively shared between them. In a study by Otten et al (2000), participants were given a dual modality task (vision and audition) under two conditions, one of which required the use of short term visual WM (hard condition) and one without (easy condition). Neural imaging showed greater activity for processing sounds under the easy visual load than when visual WM was engaged. Likewise, Klemen et al (2010) conducted a study where WM load was varied in an auditory/ visual dual modality task. The experiment showed a significant suppression of secondary stimuli processing under high WM load, leading the authors to conclude that WM load engages limited attentional resources and ‘directly’ reduces the processing of task- irrelevant stimuli (ibid, p.445). Moreover, Janata et al (2002) found that attentive listening to multi-channel music employs neural circuits underlying ‘multiple forms of working memory, attention, semantic processing, target detection, and motor imagery’ (p. 9). They concluded that attentive listening to music appears to be enabled by areas that serve general functions rather than by music specific areas, and as such must be shared in terms of available attentional resources.

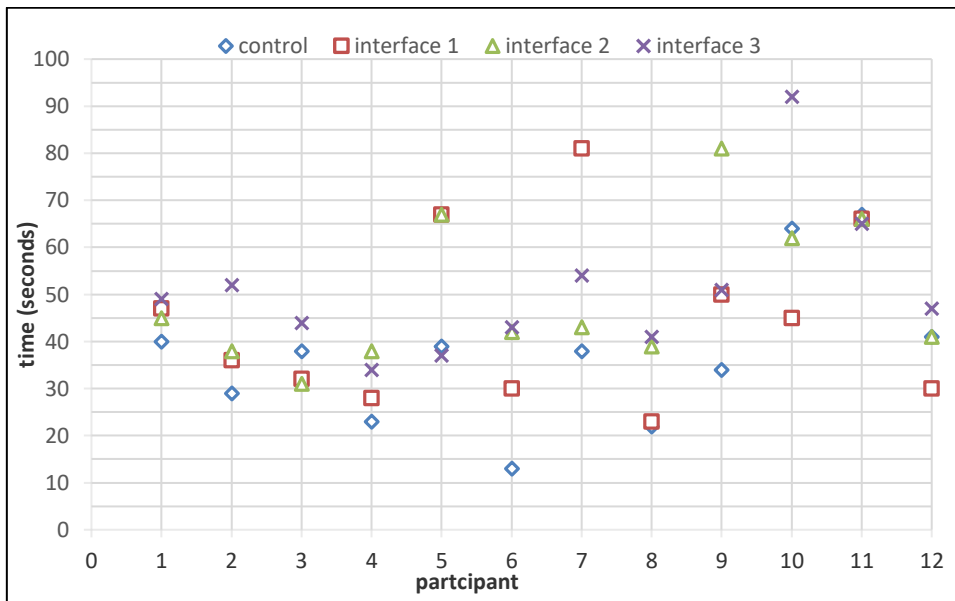


Figure 5:6: Times for completion per interface for string file.

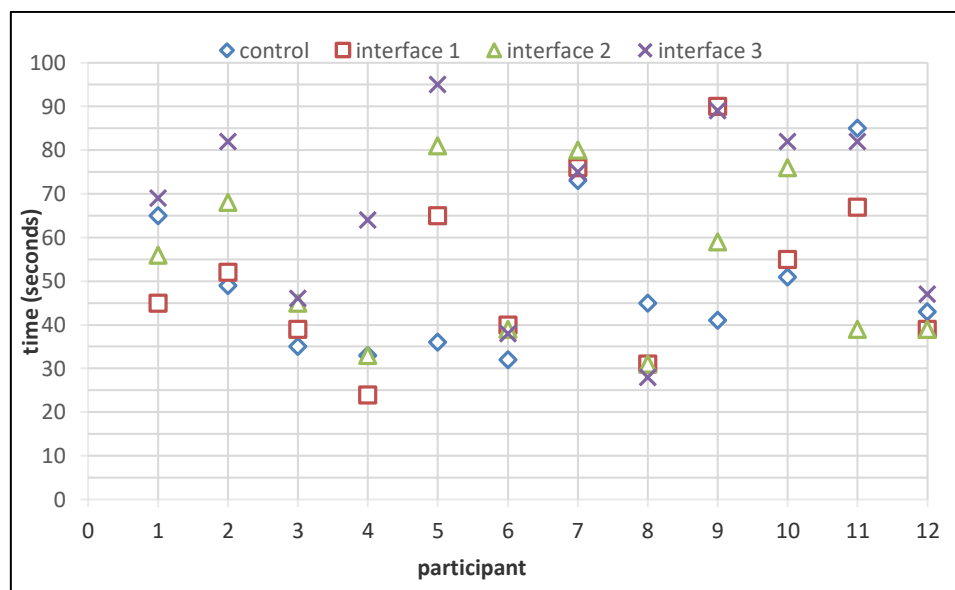


Figure 5:7: Times for completion per interface for tambourine file.

5.5 Conclusion

The findings from our study results, and the literature mentioned above, leads us to conclude that the inclusion of scrolling navigation led to an increase in WM load as participants had to remember visual information (in our case the equaliser curve) during

navigation. This in turn may have consumed available attentional capacity, reducing ability to process the subtle audio changes in the mix. This result points toward a susceptibility of auditory processing to cross modal WM load manipulation (Sorquist et al, 2010). If this is the case, it may be of direct importance to our future DAW heuristics, especially when using small screen devices, as limited screen space necessitates the use of visual WM, while different views of the mix (such as effects, volume level, track automation etc.) are revealed through the navigation functionality.

5.6 Summary

We began the chapter by reviewing the findings from our previous experiment, and considered how the amount of visual feedback therein may have fallen short of that experienced when a mixing on a DAW. We went on to discuss how this may have influenced the study results, and identified specific aspects of the GUI to add to our experimental interfaces.

In section 5.1 we outlined our study design, including details of the interface designs and participant profile. In section 5.2. we gave details of our analysis methods. In section 5.3 we summarised our results and discovered that the introduction of scrolling navigation had a significant detrimental effect on results. In section 5.4 we discussed the findings and considered how, by necessitating users to remember visual information, scrolling navigation may have loaded WM, consequently reducing the attentional resources available for auditory processing. We concluded the chapter by identifying the need to test this further to ascertain if the use of different User Interface designs may influence the level to which this takes place.

5.7 Progression to Subsequent Study

Having found that interface navigation appears to be a contributing factor to critical listening acuity, we wish to explore this further in the next study. We wish to investigate whether the reduction in auditory attention that we found in the current study was due specifically to the design of the UI objects (faders and dials) or whether it is due to navigation of the interface per se. Therefore, in the next study, we design scrolling mixing interfaces using both novel (colours and numbers) and traditional DAW UI elements (faders and dials). Using these designs, we compare critical listening tasks during interface navigation, to better establish the cause of the results found in the

current study. By so doing we hope to assess whether certain UI designs may tax WM more than others under navigation and subsequently reduce aural acuity.

Chapter 6

User Interface Object Design.

The results of the experiment outlined in chapter 5 suggested a correlation between WM load and reduced auditory processing, as found when participants had to search for and remember visual information from a parametric equaliser while scrolling. However, we should consider how much of this was due to navigation per se, and how much due to the visual complexity of the particular User Interface (UI) object tested in the study, and the load it put on WM. DAWs use a variety of UI objects in the GUI, such as faders, dials and numbers. As these all contain different levels of pictorial detail, they may vary in how much they impact on WM and visual load. For example, quantitative information in dials (a common part of DAW GUI design) can be difficult to interpret due to the fact the human eye has difficulty estimating area and comparing angles (Chawla & Whitman, 2011) specifically underestimating acute angles and overestimating obtuse angles (Robbins, 2005, p. 49). Faders, however, though also requiring visual comparison, are perceptually easier to analyse, as the human eye can compare the two-dimensional positions of objects (such as the ends of bars) or their lengths more easily and precisely than angles (Few, 2007).

Albeit in a more limited form, colours are also used within audio software design to represent mix data. For example, in Pro Tools, colours are applied to the different bands of the parametric equaliser, with lighter colours representing higher frequencies (fig 6.1) while in Serato, coarse colour coding is used to show the frequency of drum sounds (fig 6.2), allowing quick visual reference of snares and hi-hats etc. within the waveform itself. In terms of visual WM, colours have been found to play a significant role in enhancing memory performance (Wichmann et al 2002) with several studies suggesting that they can increase recall rates (Pan 2012., Smilek et al 2002) and elicit faster reaction times (MacKay et al 2005). Indeed, in a study by Pan (2012) participants were asked to identify whether the colour or the shape of two objects that were presented were the same. In the first experiment, the colours of the two objects were the same but the shapes were different, while in the second experiment this was reversed. The result showed that response times were faster in identifying the differences in colours compared to differences in the shapes of the objects in both experimental conditions.

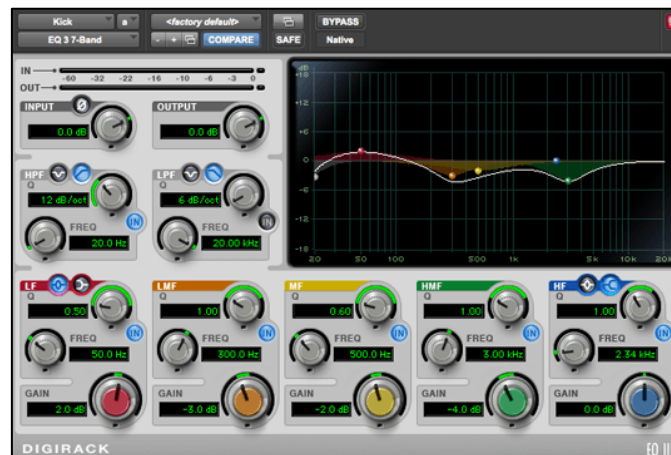


Figure 6:1: The Pro Tools Channel EQ. The five bands of the EQ are colour coded, with dark red being the lowest frequency, and light blue representing the High frequency band. Numbers are used to provide more detailed secondary source of information.



Figure 6:2: The use of colour to represent frequency in Serato DJ Software. Darker colours represent lower frequencies.

In this chapter, therefore, we assess UI elements commonly used in CS design (faders, dials and numbers) and compare them with the less detailed pictorial information found in colours. We do this to see if any design is more perceptually robust during navigation and whether different levels of visual perceptual demand inherent in the different designs, influence the simultaneous perception of auditory stimuli (Molloy et al 2015). By so doing we hope not only to quantify the extent to which the design of the UI elements impact on WM load, but ascertain if the process of remembering the mix information during navigation is responsible for reduced critical listening, or if the UI

object design has a part to play. Furthermore, we continue to address the second research questions; does the design of the UI influence critical listening skills, and do certain designs improve or detract from the ability to hear subtle changes to the audio content while navigating a CS mixer?

6.1 Study Design

6.1.1 Participants

Nine participants (seven male and two female, aged 18-42), all with a similar level of prior mixing experience, were recruited from staff and Music Technology students at City and Islington College, London. The students were all studying for the 1st year of the Edexcel Music Technology GCE A Level Course. The staff recruited were not Music or Music Technology specialists, but rather staff from other subjects with an interest in audio mixing. All participants were asked to classify themselves from the categories beginner, intermediate or expert. All participants identified themselves as beginners. In order to further control any non-experimental variability between the participants (especially given the use of students and staff) we used a Repeated Measures design, in which all participants used all the experimental interface designs. None of the participants had taken part in the previous studies.

6.1.2 Listening Task

The participants were required to listen to a two minute, eight-channel mix. During this task, they were asked to identify which of three specified instruments (guitar, snare or shaker) was decreased in volume over the course of the excerpt (as the audio diminished from full volume at the start to become inaudible at the end it became easier to hear the attenuation further into the excerpt). As the investigations of interface heuristics detailed throughout this thesis are aimed at beginners and non-experts, attenuation was chosen for this study as it is a non-technical aspect of the mix, and does not require any specialised knowledge of audio effects to discern or recognise.

The excerpt was played twelve times in total, during which each of the specified instruments was attenuated four times (with the order randomised for each participant). The participants were asked to identify which instrument was being attenuated as soon as they heard it. At the same time as undertaking this listening task they were presented

with one of four visual interfaces displayed on a 10” by 5.8” screen.

6.1.3 Visual Task

A group of 16 channels were created in Max/MSP. Each channel had four parameters with a range of 16 values (1-16). The design of the 16 channels was represented by four different UI designs (fig 6.3); numbers, dials, faders and colours (the 16 hues used for the colours were created using an online colour ramp creator). Due to the number of channels, scrolling was required to view the channels in all four designs. Participants were asked to look at channel one and compare the subsequent fifteen channels to ascertain if they were the same or different (by clicking on a ‘same/ different’ button below channels 2-16) while listening to the audio. This task was chosen as it is a common part of mixing workflow, where for example, a user may wish to compare two, non-adjacent channels to compare their frequency balance, effects setting, etc. and make requisite adjustments to one or both channels.

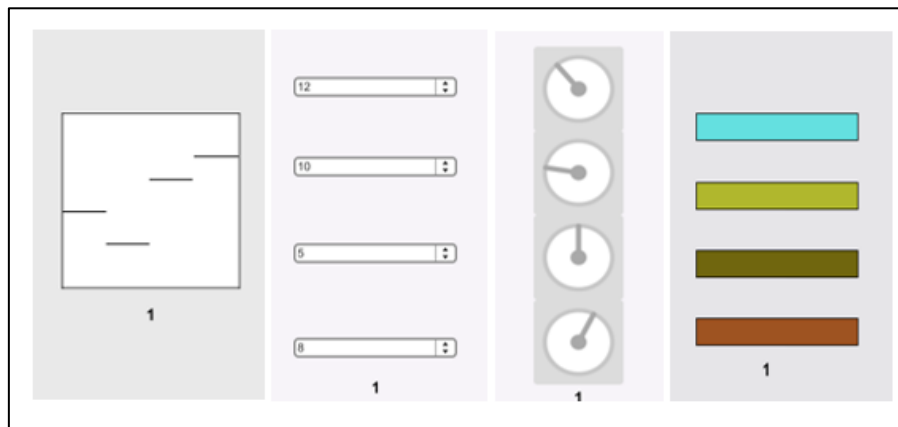


Figure 6:3: The four interface designs used for each of the sixteen interface channels.

There were sixteen channels of each design (the figure here shows channel 1 of each design). These included faders, numbers, dials and colours. Each channel consisted of four parameters, each with a range of 16 values.

The participants were presented twelve interfaces (three occurrences of the four interface designs) with the order and parameter values randomised for each participant. Participants were asked to begin comparing the channels as soon as they began the audio. They were told to press the appropriate key on the QWERTY keyboard as soon as they heard the track attenuation (which would stop the audio) and proceed directly onto the next interface. The time taken to hear the audio changes and the number and

accuracy of the channels compared was recorded. We hypothesised that designs which resulted in longer and less accurate reaction times to the subtle audio changes may use more WM and attentional resources, as less capacity is available for auditory processing.

6.2 Analysis and Results

As the analysis compared results of each experimental design interface, we used a paired *t*-test. This was chosen as it allowed us to directly compare the result times for participants across each interface type (error rate, completion rate and response time). Using a paired *t*-test analysis also allowed us to generate a *p* value, where values of 0.05 or less rejected the null hypothesis (that the interfaces design does not have any effect on critical listening skills). Furthermore, as all three of the specified instruments were attenuated in each of the interface types, using this analysis method allowed us to directly compare the response times and channel matching accuracy for each instrument across each interface type per participant.

The time taken to correctly identify the attenuated audio in each interface design was analysed for each participant. From this, the mean and standard deviation (SD) was calculated and used to generate confidence levels (CI) at 95% showing the range of the true population. Secondly, the amount of UI objects compared for each of the four interface designs was calculated. Any of the channels that were incorrectly matched were discounted from the analysis (the error rate of each UI design is shown in table 6.1). The amount of correctly matched channels was used to generate a mean and SD to produce CIs at 95%. As all three of the specified instruments were attenuated in each of the interface types it was possible to directly compare the response times and channel matching for each instrument across each interface type.

The analysis of the reaction times to the audio attenuation shows that there was no significant time difference between the four interfaces designs (figure 6.4) suggesting that none of the interface designs diverted attention from the auditory task more than any other. However, the analysis for the number of channels successfully compared reveals that participants could complete significantly more channels with UI designs using colours, faders and numbers compared to dials (figure 6.5).

	Colours	Dials	Faders	Numbers
Total Channels matched.	170	84	141	174
Channels matched incorrectly.	15	17	8	0
Error percentage.	8.8%	20.2%	5.6%	0%

Table 6.1: The percentage of channels matched incorrectly per UI type. Dials have a significantly higher error rate than the other UI object designs.

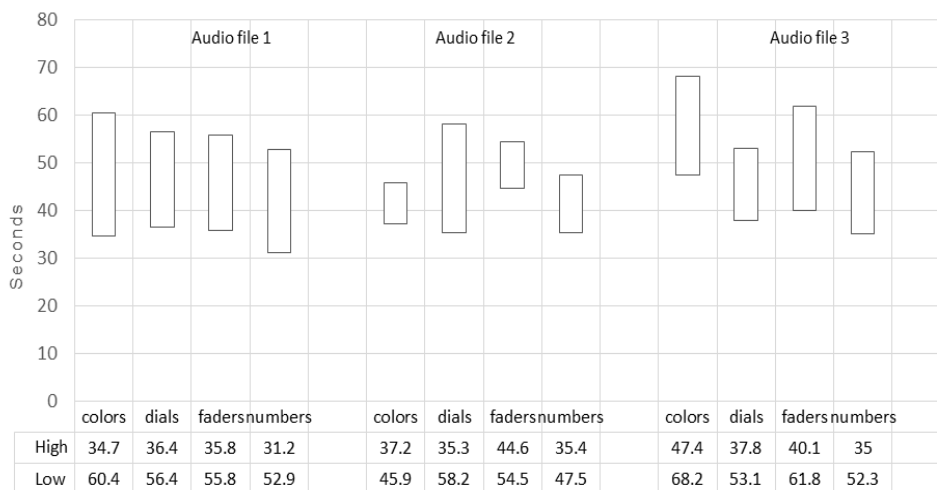


Figure 6.4: The mean time taken to correctly identify the changes to the audio using the four different interface types. The analysis (at 95% CI level) shows there is no significant difference in the time taken to hear the changes when using different UI designs.

6.3 Discussion

In terms of the listening test, as mentioned above, the analysis did not reveal any significant difference in the ability to hear audio changes across the various UI designs, even though they varied in visual detail. This may be because the level of perceptual difficulty between the designs was not great enough.

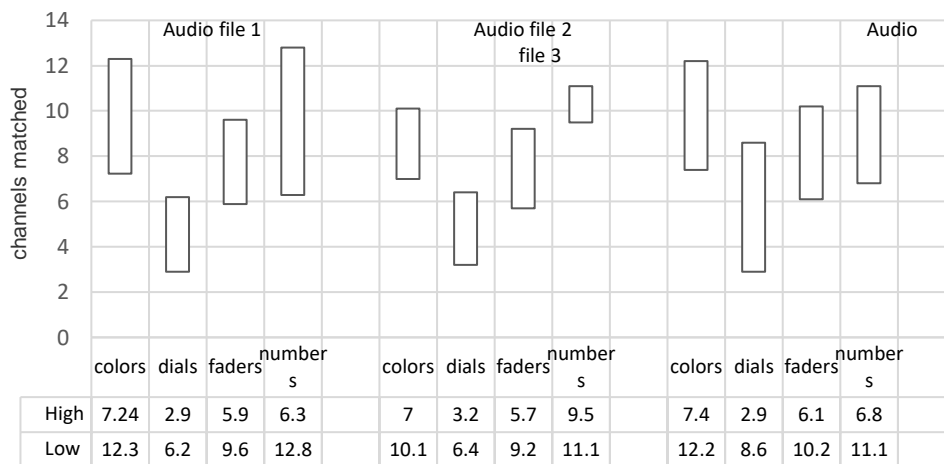


Figure 6:5: The mean number of channels compared using the four interface designs. The analysis (at 95% CI level) shows that participants could successfully compare significantly more channels when the UI was presented as colours and numbers as opposed to dials.

However, given the difference in visual complexity between colour and a dial, this seems doubtful. An explanation may be found in literature from Cognitive Load Theory (CLT) (Sweller 1988). CLT is an area of research concerned with the optimal design of information for instructional design, originally in terms of paper based materials, but increasingly focused on complex electronic learning environments (Van Merriënboer & Ayres, 2005). In CLT, ineffective design is seen to occur if learners are required to mentally integrate any information between two disparate sources (for example diagram and text). Users must devote substantial cognitive resources to remembering and mentally integrating the various sources of information (Chandler & Sweller 1992), which in turn is understood to have a direct bearing on WM load (Hollender et al 2010). In order to minimise this, it is recommended that disparate sources of information in the GUI should be physically integrated where possible (Chandler and Sweller 1992), and information should be displayed so that users don't have to remember it from one screen to the other (Hollender et al 2010). Indeed, Ferguson et al (2005) found that when visually representing several audio features simultaneously, this is best achieved when these are shown as different attributes of the same object, due to the limited nature of human attentional bandwidth.

As we discussed in chapter 5, increased visual WM load appears to have a direct bearing on concurrent aural acuity (Alho et al 1992., Otten 2000., Berti & Schroger 2003., Janata

et al 2002., Klemen et al 2010) as it limits the capacity for secondary stimuli processing. Subsequently, the process of representational holding (Mayer & Moreno 2003) - remembering and integrating the information while scrolling across pages - may have meant that auditory attention was affected equally in all conditions with no UI design produced significantly better or worse results than any other. This finding further confirms the work outlined in Chapters 4 and 5, and suggests that visual WM load during interface navigation occurs irrespective of the pictorial detail of the UI object design. In order to further validate this explanation, we continue to investigate and test this in the next chapter, when GUI designs using either recognition or recall are compared to assess if there any differences in critical listening and auditory attention.

In terms of the visual task, the analysis revealed that dials produced significantly less channel matching than the other UI designs. This may be due to the difficulty in interpreting quantitative information in dials (Chawla & Whitman, 2011) specifically underestimating acute angles and overestimating obtuse angles (Robbins, 2005, p. 49). For this reason, it is also difficult to compare dial positions that are not in close proximity to each other (ibid), which may explain why it fared poorly under navigation and had the highest percentage of errors (table 6.1). This supposition is in line with Guided Search Theory (Wolfe 2007), which suggests that efficient and rapid preattentive processes which guide visual attention are influenced by salience factors, such as the target-distractor similarity. In difficult visual search tasks, it has been found that reaction time is directly influenced by these salience factors (Zenon et al 2008). This may further explain the inefficiency found in visually searching dials in our experiment. As the position of the dials became more similar between various channels, the search became harder. Indeed, while it is possible to resolve orientation differences in the order of one degree (Olzak & Thomas, 1986), it requires approximately 15 degrees of difference to reliably attract attention (Foster & Ward, 1991b; Moraglia, 1989).

In terms of colours, these performed well in the test, producing the highest number of channels correctly compared in two of the three tests. Given the somewhat limited implementation of colours to convey ordinal data in DAWs, this may be a potential design paradigm to pursue. However, there are both perceptual and physiological caveats that need to be considered when using colours to represent mix information. Colour discrimination can be compromised by a variety of factors, such as the lighting conditions, display position, display quality, and viewing angle (Yeh et al, 2013) while colour vision deficiencies (such as, “colour blindness”) affect approximately nine

percent of the population (Galitz 1997). Colours also need to be selected carefully to ensure that they are sufficiently different and easily discriminable from each other. The more colours that are used the closer in hue each colour will be and the harder it will be to discriminate between them (Smith and Mosier, 1986). While sixteen colours were discriminated in this study, they still had an error rate of 8.8 % (table 6.1). This error rate may increase as the colour palette extends to represent more values, and the distractors become more heterogeneous. Moreover, the fact that colours can help in memorising information is due to their greater ability to capture attention than other variables (Dzulkifli & Mustafar 2013). In this way, and accepting a dual modality view of attention, the increased visual attention recruited by colours in our study may have negatively impacted on concurrent auditory attention, negating the benefit of reduced pictorial complexity.

Lastly, numbers performed well in the study, with a significantly higher search rate than dials (figure 6.5). However, while numbers show parameter values precisely and allow users to easily compare two values (as required in the study) they may not be as effective in showing patterns, trends, or exceptions among parameters values, or allowing the user to compare whole sets of numbers (Few 2004). Furthermore, numbers may not be suited to high zoom levels where they become hard to read, therefore making them ill-suited to overviews of the mix space.

6.4 Summary

We began the chapter by reviewing the findings from our second study. We considered the influence of visual load and WM on concurrent critical listening in terms of UI design. We questioned whether results found in the study may have been due to the UI design tested and went on to consider how the complexity and perceptual demands of other UI objects may influence auditory attention.

In section 6.1 we tested this in our experiment, where we quantified various UI object designs on visual search times, search accuracy and critical listening skills while navigating an interface. In section 6.2. we analysed the results and discovered that while differing UI designs produced quicker and more accurate visual search, there was no significant difference between the tested designs on aural acuity. In section 6.3 we discussed the findings and considered the possibility that the act of integrating visual information across the GUI and holding the UI objects in WM, may have reduced

auditory attention equally under all conditions, regardless of visual detail. We concluded the chapter by identifying the need to test this further to ascertain if reduction in aural acuity during navigation can be ameliorated by the design of the GUIs.

6.5 Progression to Subsequent Study

In the current study we further clarified our findings from study 2 (chapter 5). Specifically, we found that WM appears to be taxed equally by both novel (colour, numbers) and traditional (faders, dials) UI elements, with all designs resulting in a broadly similar reduction in aural acuity while navigating the mixing interface. This result suggests that interface navigation per se may be responsible for reduced aural acuity, rather than the design of the UI elements themselves. In the next study we therefore investigate interface designs which eschew scrolling navigation completely. To do this, we design mixing interfaces using overviews of the mix information (thereby negating the need to scroll) and compare these with designs which require scrolling to access mix information. Using these designs, we quantify participant's speed and accuracy for both visual search and concurrent listening tasks and assess whether removing navigation affects the results.

Chapter 7

Managing Visual Search in Mixing GUIs.

In the previous chapter, we discussed the influence of the User Interface (UI) object design on the speed and accuracy of visual search time and critical listening. The results of the study revealed that, while the complexity of the design of UI objects does not necessarily have a significant bearing on critical listening, remembering and integrating visual information presented in the UI objects does appear to negatively affect concurrent auditory attention. We suggested that the process of holding the visual information in visual Working Memory (WM) resulted in cognitive overload, as users devoted substantial resources to remembering and mentally integrating the various sources of information (Chandler & Sweller 1992).

Within other fields, such as Instructional Design, research suggest that overviews of the information space may help avoid the negative effects of disparate data in the Graphical User Interface (GUI) (Thüring et al 1995). Overviews can minimise navigation and support recognition rather than recall, subsequently avoiding the representational holding which may contribute to cognitive overload and reduced WM capacity (Hollender et al 2010). Furthermore, by providing navigational cues, overviews may be useful for orienting users in the search environment (Jul and Furnas 1998), which may be of benefit when mixing with DAWs. For example, in a scenario where one track in a large multi-track mix is distorting, the user must search through a potentially large number of UI objects, requiring navigation through multiple pages to find and correct the offending channel. If this is time critical, such as during live mixing, the temporal separation between hearing the distortion and locating its channel can potentially cause them to lose their sense of place (Rodgers, 2001). In such a scenario, the broad focus provided by overviews may ameliorate the tendency to become lost while searching the GUI (Grayden et al 2015., Jul and Furnas 1998). Finally, by presenting the data as a whole, overviews may allow the user to discover which sections of an interface is densely or sparsely populated so they can see ‘where there are clusters, exceptions, gaps and outliers’ (Card et al 1999, p.239). This may be useful in mixing, where elements interact sonically, and one channels settings can impact on another (Dewey and Wakefield 2015).

Providing a holistic overview may therefore allow the user to build a more cohesive mental image of the mix (Ratcliffe 2014).

However, while overviews of mix information may support orientation and recall, there are potential drawbacks due to reductions in UI object resolution and screen size (Büring et al 2006). For example, as we have seen, quantitative information in dials can be difficult to interpret (Chawla & Whitman, 2011) and reducing their detail further may compound this problem. The same issue may be experienced with numeric information, with legibility becoming increasingly difficult as size is reduced. Furthermore, in an overview, faders may become minimised to heights to as little as 100 pixels (Htalky et al 2009), leading to a corresponding loss in usable resolution.

This study therefore tests the efficacy of overviews to convey mix data, comparing them with scrolling navigation interfaces and assessing their influence on the speed and accuracy of visual search and concurrent critical listening. In order to address the perceptual and display issues highlighted above regarding the reduced size of CS UI objects, the study includes both a CS overview using dials and faders, and a Stage Metaphor (SM) mixer, in which volume and pan position are represented by a single circular graphical object- with its X/Y position representing pan position and volume respectively.

7.1. Study Design

7.1.1 Participants

Nine participants were recruited from staff (2) and Music Technology students (7) at City and Islington College, London (7 male, 2 female and aged 18-43). The students were all studying on the Edexcel Music Technology GCE A Level Course and were all at the same stage of their studies. The staff recruited were not Music or Music Technology specialists, but rather staff from other subjects with an interest in audio mixing. All participants were asked to classify themselves from the categories beginner, intermediate or expert. All participants identified themselves as beginners. In order to further control any non-experimental variability between the participants (especially given the use of students and staff) we used a Repeated Measures design, in which all participants used all the experimental interface designs. None of the participants had taken part in the previous studies.

7.1.2 Visual Task

Three versions of a 24-channel interface showing volume and pan-position were designed using Max/MSP. The designs consisted of a CS with all 24 channels shown on a single page without the need to navigate (fig 7.1, top), a design using an SM mixer presented on one page without the need to navigate (fig 7.1. middle) and a CS mixer requiring scrolling navigation to view all 24 channels (fig 7.1, bottom).

For each of the interface designs, participants were asked to answer four questions about the visual information and select the correct answer from a drop-down menu above each interface (fig 7.1). The questions (table 7.1) were designed to test quick visual referencing (i.e. which channels are panned to extremes, whether the mix has more channels above or below the centre volume) as well as more specific visual referencing questions (i.e. how many channels have volume or pan set between certain values; what is the panning/volume positions of specific named channels). Participants were presented four occurrences of each interface type (with the order randomised for each participant) making twelve screens in total.

1	2	3	4
Overall is the mix panned more to the left, or right?	How many channels have pan set to 2?	How many channels have pan below 3?	What is the volume of channel 8?
How many channels have volume on 6 or above?	How many channels have volume on 8?	How many channels have volume on 4 and below?	Are there more channels with volume above or below half way?
How many channels have pan and volume set to half way or above?	What is the pan position of channel 12?	Which has highest volume out of channels 2 12 and 24?	What is difference in pan position between channels 3 and 18?
How many channels have volume set between 2 and 6?	How many channels have pan set between 7 and 15?	How many channels have the volume set between 4 and 7?	How many channels have pan set between 3 and 6?

Table 7.1: Questions asked for each occurrence of the three interface designs. Each group of four questions was asked on each interface design, with the order randomised for each participant.

7.1.3 Listening Task

While the participants were undertaking the visual task, they were played a twelve-channel audio mix (duration 44 seconds). Each time the excerpt was played, four of the instruments within the mix (namely vocal, snare, flute and tambourine) were randomly panned either left, right or centre. As soon as the excerpt had finished the mixer screen was automatically closed and participants were asked to select the correct pan position of two of the instruments (chosen at random) from a drop-down menu with the categories; 'left, centre, right or couldn't tell' (this last option was included to try and minimise the occurrence of participants guessing the answer if they were unsure).

7.2 Analysis and Results

Inferential statistics were used to analyse participant results, as we wished to assess whether the differences found between the various interfaces being tested were due to chance or interface design. We therefore analysed the amount of correctly answered visual and auditory questions (and the time taken to answer) for each participant per interface type. From this, the mean and standard deviation was calculated for the three interface types. The mean and standard deviation generated Confidence Intervals (at 95%). This analysis method allowed us to quantify any overlap between results and see the range of the true population per interface type, with non-overlapping confidence levels suggesting a statistically significant difference between GUI designs. A comparison of time and accuracy by question type was also analysed to quantify whether any of the various interface designs supported visual search better than others.

The amount of correctly answered visual and auditory questions (and the time taken to answer) were recorded and analysed for each participant per interface type. From this, the mean and standard deviation (SD) was calculated for the three interface types (tables 7.1 and 7.2). The mean and standard deviation generated confidence levels (CI) at 95%, showing the range of the true population per interface type (figures 7.2 and 7.3).

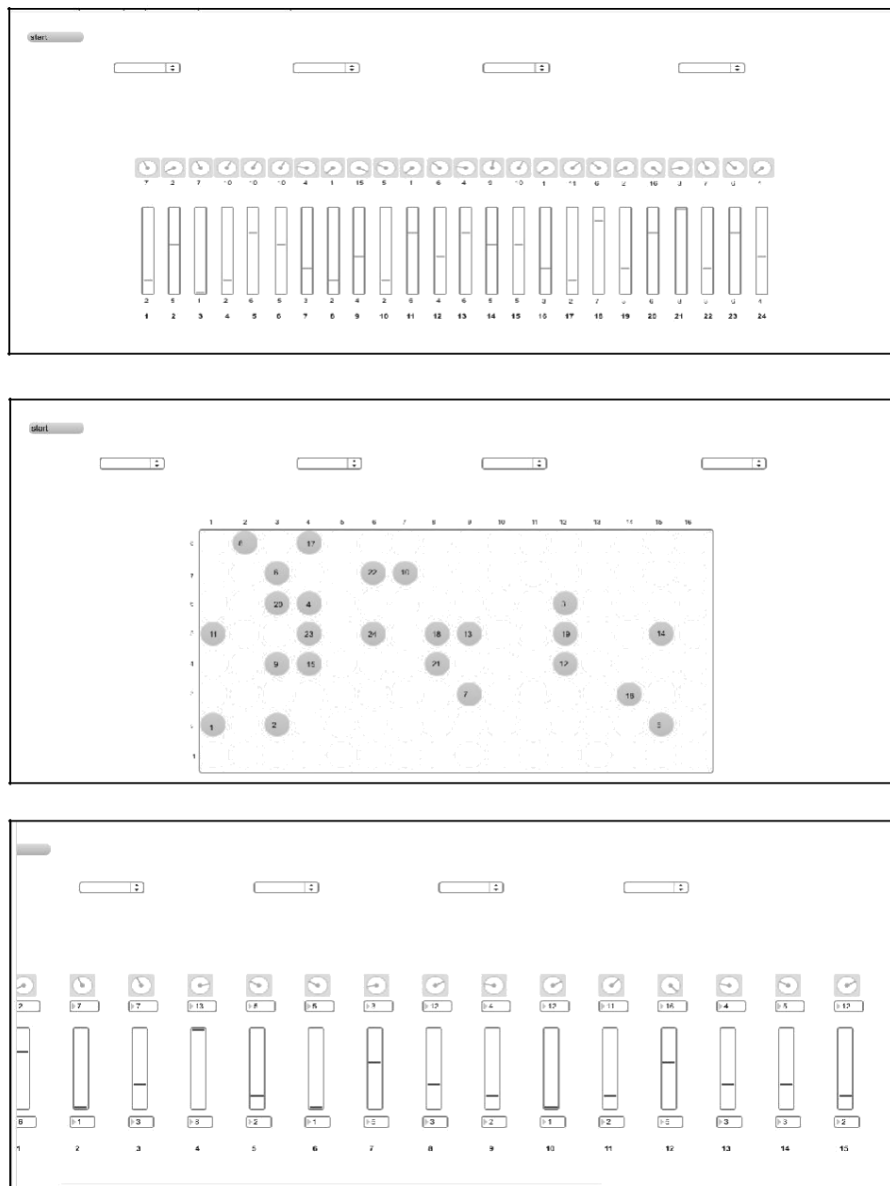


Figure 7.1: Top; the 24 channels as a mixer overview, all channels are on one page without the need for navigation. Middle; the 24 channels as an SM overview, the left right position represents pan position while the up down position represents volume, channel numbers appear in the circles. Bottom; The 24 channels as a traditional scrolling design requiring navigation to view the channels.

In terms of the visual search task, there were significant differences between the three interface designs. The scrolling interface had the lowest amount of questions correctly answered, significantly less in fact than both the CS and SM overviews (fig 7.2). While the amount of correctly answered visual search questions in the SM overview was not significantly greater than the CS overview, it was an increase.

Interface Type	Mean	Standard Deviation
CS Overview	7.66	0.86
SM Overview	8.66	1.73
Scrolling	5.6	1.22

Table 7:2: The mean and standard deviation for amount of visual search questions correctly identified per interface type.

Interface Type	Mean	Standard Deviation
CS Overview	5.11	1.76
SM Overview	5.77	1.09
Scrolling	3.88	1.16

Table 7:3: The mean and standard deviation of amount of audio panning positions correctly identified per interface type.

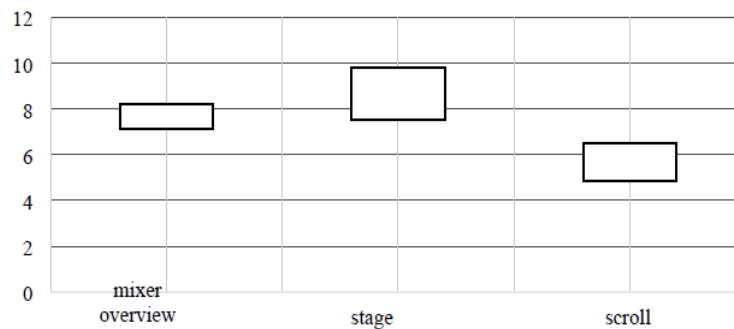


Figure 7:2: Confidence intervals (95%) for the number of correctly answered visual questions per interface type. The CS and SM overviews show a significant increase compared to the scrolling design. The highest amount of correctly answered visual search questions occurred in the stage overview.

The improved visual search results found in the SM design may be attributable to the mapping of mix attributes to visual properties (e.g. position), which leveraged perceptual skills to discern and interpret patterns within the data (Heer and Shneiderman 2012). Indeed, the SM design provided the quickest times when it came to recognising patterns within the mix (e.g. whether more channels were panned left than right, whether more channels had volume below rather than above half way) which tallies with literature which suggests that overviews allow the user to effectively

comprehend the relationships between data and discern global patterns more readily (Shneiderman 1998). This is in contrast to the CS design, where users must ‘scrutinise each channel’s setting to assemble a mental image of the mix’ (Dewey and Wakefield 2016, p.2).

However, the SM mixer showed less efficient results where participants were required to find information about a specific channel. This may be attributable to the random distribution of the channels. Unlike the CS mixers, channels were not numerically ordered, so when asked about a specific channel, the participants had to search through the GUI to find it. This problem has been identified by other researchers (e.g. Gelineck et al 2013) who acknowledged that unlike the CS metaphor, the SM’s channels random distribution in the virtual stage creates potential difficulties for information search. In Chapter 8, we address this issue, by applying strategies found with data visualisation to visually filter the display of mix information.

The analysis of the listening task per interface type reveals that the SM design provided significantly higher amounts of correctly identified audio panning than the scrolling interface (fig 7.2). Though not significant, the CS overview also provided higher amounts of correctly identified panning relative to the scrolling interface. This finding is in line with the results from our previous studies, which suggest that simplifying visual search, specifically by reducing interface navigation, may result in improvements in concurrent auditory processing. In terms of the difference between the two overview designs, the SM overview resulted in an increase in correctly answered listening questions, though this was not statistically significant (fig 7.2). This result further supports the notion presented in CLT literature (e.g. Chandler & Sweller 1991) that visually integrating elements in the UI (in our case volume and panning) may reduce representational holding (Mayer & Moreno 2003) and subsequently reduce cognitive load and visual WM. In line with our findings from the previous studies, the results further suggest a link between a reduction in visual WM and increased auditory attention. In the case of the present study, the removal of navigation, the integration of UI objects and the mapping of mix attributes to visual properties (e.g. position), contributed to improvements in critical listening during visual search.

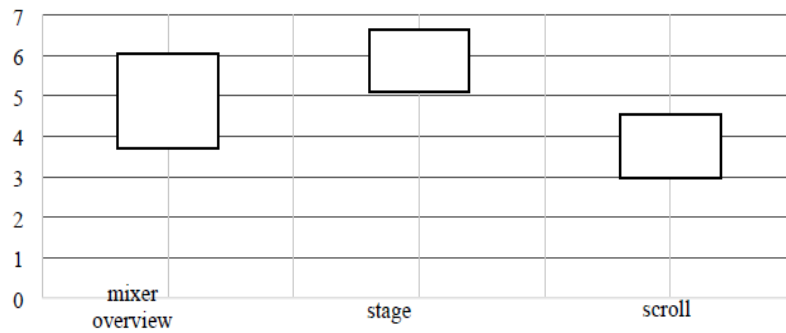


Figure 7:3: Confidence intervals (95%) for the number of correctly identified panning positions per interface type. While there is an overlap between the scrolling interface and the CS overview, the SM overview shows a significant increase compared to the scrolling design.

7.3 Summary

In this chapter, we considered the positive aspects of using mixer overviews; gaining a broad focus of the mix space, minimising navigation, orientating the user and allowing the analysis of patterns in the data. We also considered potential drawbacks, including lack of resolution and reduction in the GUI size and detail.

In section 7.1 we gave details of our experiment to quantify the extent to which overviews impacted on mixing workflow, using a traditional scrolling interface, a CS overview and a SM metaphor overview, and we related these to mixing, identifying potential influences on workflow.

In section 7.2 we discussed our results, which suggested that overviews provided quicker visual search than a navigation based interface, and resulted in improved aural acuity, especially with the SM mixer. We discussed this result and suggested it may be due to the mapping of mix attributes to visual properties (e.g. position), which leveraged perceptual skills to discern and interpret patterns within the data (Heer and Shneiderman 2012). We also suggested that by integrating the UI elements into a single object in the stage metaphor overview, we may have reduced cognitive load and WM, leading to increased auditory attention. Lastly, we highlighted the potential difficulties with random distribution of channels when using a SM overview, and the need to explore ways to organise and filter the information further.

7.4 Progression to subsequent study

In this study we clarified the results from studies 2 and 3; we found that using overviews of the mix (thereby removing the need to navigate the interface) provided quicker visual search than a navigation-based interface, and resulted in improved aural acuity, especially with the SM overview mixer. However, while the results were encouraging, the SM overview mixer lead to difficulties with the random distribution of channels. Furthermore, as mixes become more complicated and channels increase, there is a danger that an overview mixer will become congested with information, as unlike scrolling designs the user may have to interact with more information than can be conveniently displayed on a single screen (Cockburn and Gutwin 2009). In the next study, we therefore investigate the use of filtering in overview designs to show/ hide mix information per user requirements. To do this, we design overview mixers (using both CS and SM designs) incorporating Dynamic Query (DQ) filters and compare these with overviews without filters. By so doing, we hope to quantify and assess how managing the amount of visual feedback in an overview affects both visual search speed and critical listening accuracy and whether such designs can ameliorate potential difficulties with mixer overviews.

Chapter 8

Filtering Data in Audio Mixing Interfaces.

In the previous chapter, we found that overviews of mix data allowed faster search for visual information compared to a navigation interface, and provided improvements in simultaneous critical listening tasks. We concluded that this may be due to the reduced visual search and visual WM compared to scrolling. In the SM design, where the results were the most improved, we discussed how the visual integration of separate UI objects representing pan and volume into one UI object in may have further minimised cognitive load and representational holding. However, though the broad focus of the overview proved useful, we found the SM's random ordering of channels was less efficient for certain tasks, due to the non-sequential distribution of the channels.

In his discussion on exploring information in Graphical User Interfaces (GUIs), Shneiderman (1998) states the importance of designing interfaces to display information in an orderly and user-controlled way. While he acknowledges that there are several approaches to achieving this aim, he asserts they should be underpinned by a basic information seeking principle; overview first, zoom and filter, then details-on-demand (zooming in this context refers to focussing on an area of an object, rather than zooming in navigational terms). Within other domains, such as maps and websites, we can see an application of this heuristic in the implementation of dynamic query (DQ) filters. These are User Interface (UI) objects (sliders, buttons and other filters) that facilitate rapid exploration of interfaces by real time visual display of query formulation and results (Li & North 2003). Ordinal, quantitative, and temporal data can be filtered using a standard slider (for a single threshold value) or a range slider (for specifying multiple values) to allow rapid and reversible exploration of visual data. By incrementally adjusting a DQ filter, users can rapidly explore and filter the information while continuously viewing the changing results.

In application, DQ filters have been shown to improve the usability of overviews. For example, research by Ahlberg and Shneiderman (1994) demonstrated performance improvements and high levels of user satisfaction when implementing DQ widgets into an interface. By allowing rapid, incremental and reversible changes to query parameters,

often simply by dragging a slider, users could explore and gain feedback from displays in a few tenths of a second, a time frame which fulfils the desire for visual tools to support interface search at rates which resonate with the pace of human thought (Heer and Shneiderman 2012). Furthermore, DQ filters may help users to filter the GUI, and find visual information without being distracted by excessive amount of ‘ink’ on the screen (Tufte 1983).

While there is little reported use of DQ filters with DAWs, a study of their implementation by Gelineck and Uhrenholt (2016) explored their use in an audio mixing context. While their study was, by their own admission, an informal exploratory evaluation, they concluded that DQ filters benefited users in understanding attributes of mix information ‘at a glance’ (ibid, p.3) such as whether a channel was active or inactive. However, the authors concluded that there was a risk that DQ filters might draw attention away from the mixing task, by ‘stealing focus’ (ibid, p.5) thereby causing the extraneous complexity of the interface to encroach on the intrinsic complexity of the user’s main task (Oviat 2006).

For this study, we aim to quantify the extent to which filtering of overviews may facilitate improvement in visual search time and/ or concurrent critical listening. We also wish to assess whether their inclusion distracts users from the mixing workflow. We use quantitative as well as qualitative metrics of success, and directly compare results from mixers with and without the inclusion of DQ filters. The experiment continues to address the finding from the previous chapter, and refines our second research question; does the design and layout of the visual information have an impact on aural acuity or speed of visual tasks?

8.1 Study Design

8.1.1 Participants

The thirteen participants selected for this study comprised staff (3) and students (10) on a two-year music technology course at City and Islington College, London. The students were all studying for the 1st year of the Edexcel Music Technology GCE A Level Course and were at the same point in their studies. The staff recruited were not Music or Music Technology specialists, but rather staff from other subjects with an

interest in audio mixing. All participants were asked to classify themselves from the categories beginner, intermediate or expert. All participants identified themselves as beginners and all participants indicated that they had prior experience of mixing using DAWs (namely Logic and Pro Tools). Participants were 10 males, 3 females. None of the participants had taken part in our previous studies. Despite the broadly similar experiential level of the participants, in order to further control any non-experimental variability (especially given the use of students and staff) we used a Repeated Measures design, in which all participants used all the experimental interface designs. None of the participants had taken part in the previous studies.

8.1.2 Visual Task

Three interface designs of a 24-channel mixer showing volume and pan-position were designed using Max/MSP. For all interface designs the pan and volume had a range of 12 values. For each of the three designs a version with and without DQ filters was included (creating six interfaces in total) so that their influence could be analysed for each design. The designs consisted of a CS design with all 24 channels shown on a single page without the need to navigate (figure 8.1a), a CS mixer where scrolling navigation is required to view all 24 channels (figure 8.1b) and a SM mixer (figure 8.1.c).

In the case of the DQ versions, the DQ filters allowed the users to query the pan, volume and position of individual channels (figure 8.2). In the case of the SM mixer, pan position was queried by selecting the numbers on the x-axis, and volume queried by selecting numbers on the y-axis, individual channels were highlighted by clicking the numbers at the top of the screen. For the channel strip mixer designs, pan was queried using the horizontal sliders, volume queried using the vertical sliders, and individual channels selected by clicking on the channels strip numbers. Once selected, the relevant channels were highlighted in the mixer displays.

For each interface, a series of six questions about the visual display was included on the screen (see table 8.1). When one question was answered the next would appear. The questions were designed to test visual referencing of volume and panning of a mix. Panning and volume levels were chosen as they are fundamental elements of mixing workflow (Owsinski 2006), and being aware of each sound's position in terms of these attributes are fundamental in giving each sound space within the arrangement, preventing sounds from 'fighting for attention' (ibid, p.11), and achieving clarity.

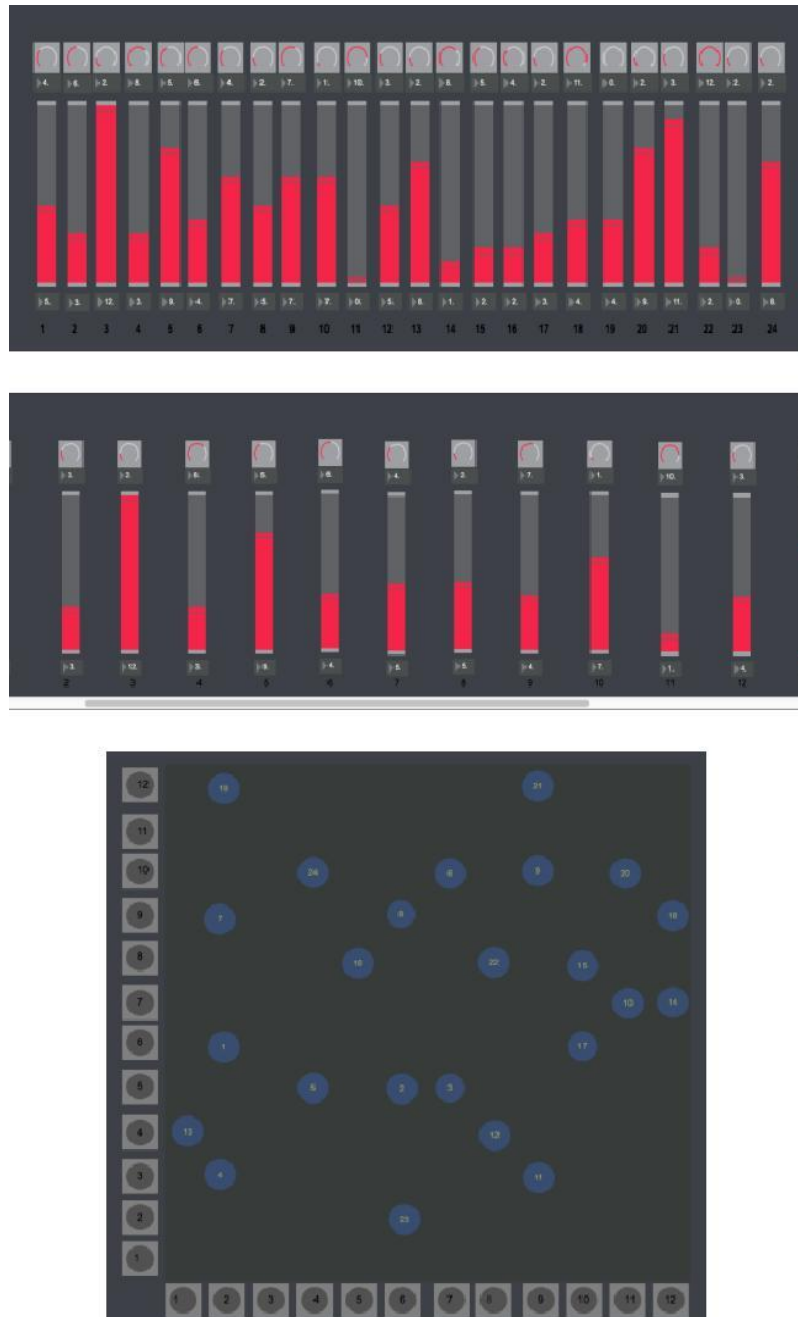


Figure 8:1 a) top; the mixer design with no scrolling. b) middle; the mixer design requiring scrolling navigation to view all the 24 channels. c) bottom; the stage mixer design, the numbered circles represent channels, the x- axis represents panning and the y-axis represents volume.

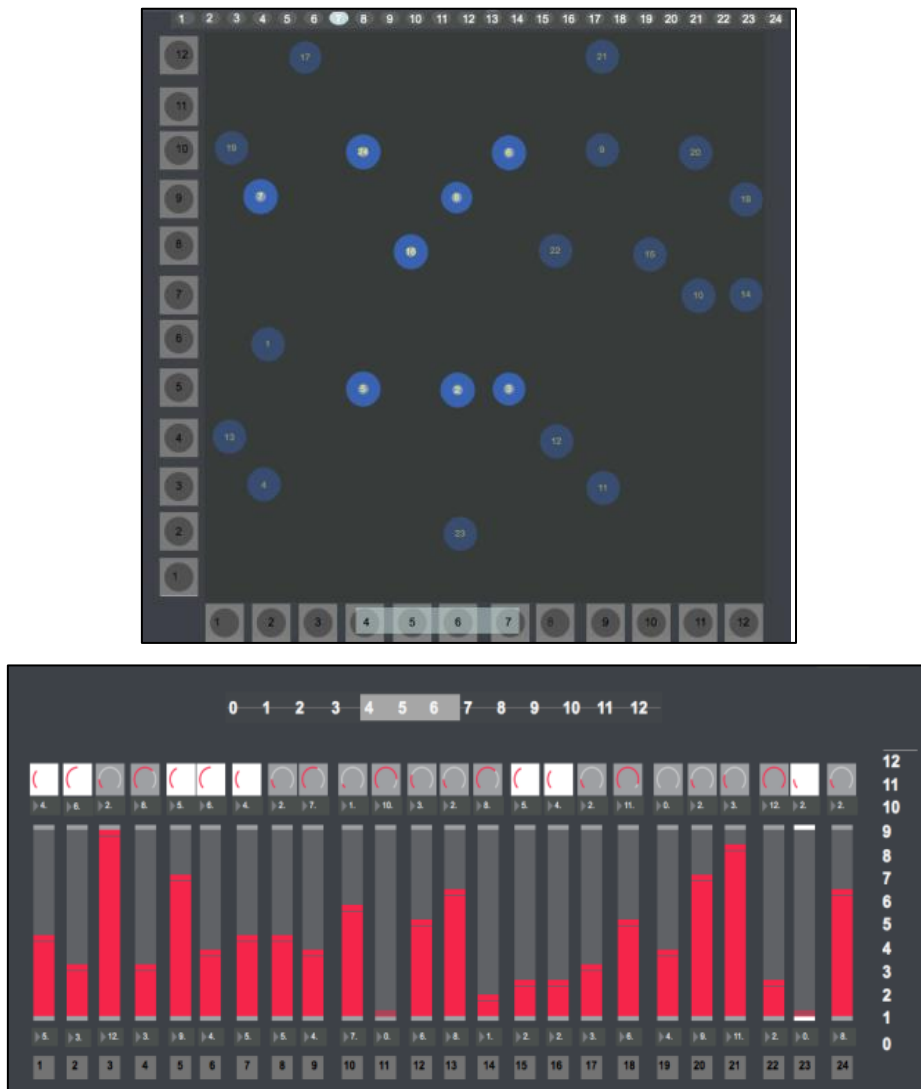


Figure 8:2: The stage mixer design and channel strip mixer designs with DQ functionality. The selected range highlights the relevant channels. In the stage mixer (top), channels panned between 4 and 7 are selected, as is the individual channel 7. In the channel strip mixer, channels panned between 4 and 6 are selected.

Participants were asked to answer as many questions as they could in the 45 seconds that the excerpt played, and as soon as the audio had finished the interface was automatically closed. Each question was asked in each interface design with the question order randomised for each participant. A maximum of six questions was asked as we considered this to be as many as users could reasonably be expected to answer in the time available.

8.1.3 Listening Task

The listening task was designed to assess whether DQ filters, by reducing visual search, allowed greater cognitive resources to be given to the aural modalities, thereby increasing aural acuity. The participants were played a twelve-channel audio mix (duration 45 seconds, created using Apple Loops from Logic Pro 9 and imported as 16 bit/ 44.1 KHz audio files into Max/MSP) at the same time as undertaking the visual search tasks. Each time the excerpt was played three of the instruments within the mix (namely backing vocal, snare and tambourine) were randomly attenuated by 6dB. This gain increment was chosen as it is considered an easily discernible reduction in volume (Everest 1998).

The instrument attenuated in each trial was pseudo-randomised with the condition that each instrument was turned down twice for each participant (so that a direct comparison could be made between the interface designs). The point in the excerpt at which the attenuation was applied was also randomised for each participant. As soon as the excerpt had finished playing, the interface which the participants were using was automatically closed and they were asked to select which instrument had been attenuated from a drop-down menu with the categories; backing vocals, snare, tambourine or couldn't tell (this last option was included to avoid participants guessing the answer if they were unsure).

8.1.4 Study Procedure

Before the study began, participants were given an opportunity to use the software and familiarise themselves with all six interface designs. Participants were also given a screening test to see if they could hear the attenuation of the specified instruments (this was done without any concurrent visual task). Participants who could not identify the attenuation would not have their results included in the study. Participants were asked to rate how easily they could hear the attenuation on a five point Likert scale (very easy, easy, hard, very hard, couldn't hear). All participants chose either very easy or easy for all three instruments, suggesting that discerning audio attenuation at -6dB was within their capabilities when there was no simultaneous visual task to conduct. Immediately after the test a survey was given to evaluate the participant's subjective views on task completion using the various interface designs.

8.2 Analysis and Results

As with the previous study, we used inferential statistics to analyse the results. These took the form of Confidence Intervals (CI) and a χ -test for proportion dependent groups. The CI (at 95%) analysis allowed us to quantify any overlap between results from the experimental interfaces to see the range of the true population per interface type, with non-overlapping CIs suggesting a statistically significant difference between GUI designs. The χ -test (at 95%) was chosen so that we could quantify the significance of the percentage of correct answers per interface type. Using A z-test for proportions-dependent groups allowed us to determine if the percentage of correct responses generated by the participants when using each of the six interfaces were significantly different from one another. The data for the thirteen participants was analysed for three main criteria; the amount of correctly answered visual questions, the amount of correctly identified file attenuations and an evaluation of the post-study survey.

8.2.1 Visual Task Analysis

The amount of correctly identified visual searches was analysed for each participant per interface type. From this, the mean and standard deviation (SD) were calculated for the participants' responses in the six interface types. These were used to generate Confidence Intervals (CI) at 95%, showing the range of the true population per interface type (figure 8.3). The analysis revealed that participants were able to correctly identify significantly more visual information with the DQ version of the interfaces, with no overlap between the CIs.

Q.1	Which channel is loudest, 3, 13 or 23?
Q.2	How many channels are panned between 2 and 4?
Q.3	How many channels have volume between 11 and 12?
Q.4	Is the volume of channel ten between 1 and 3?
Q.5	What is the loudest channel panned between 1 and 3?
Q.6	What is the difference in volume between channels 3 and 7?

Table 8:1: Visual search questions asked per interface design. Panning and volume were chosen as they are fundamental attributes of a mix, and correct placement in terms of these attributes are essential to creating a clear mix.

8.2.2 Aural Task Analysis

The amount of correctly identified file attenuations were analysed for each of the thirteen participants. This was used to calculate the percentage of correct answers per interface type (table 8.2). A test for proportions-dependent groups was used to determine if the percentages of correct answers from the six interfaces were significantly different from one another (table 8.3).

The analysis, at 95% CI, showed that the mixer DQ, SM DQ design, and the SM interface had a higher amount of correctly identified audio attenuations than the scrolling interface. Furthermore, the SM DQ interface produced significantly more correct answers than the mixer interface and the scrolling DQ interface (table 8.2) making it the most effective design in allowing the participants to discern the audio changes.

8.2.3 Survey Results

Following the study, participants were asked to rate their experience of using the various interfaces. The questions were designed to test their levels of comfort and their perceptions of task completion and success using the different designs with and without DQ filters. The questions asked were as follows:

- Which interface did you feel most comfortable (least stressed/ rushed) using?
- Overall how much did having the sliders help in each interface design?
- Which interface do you think helped you do the listening task best?
- Which interface do you think helped you do the visual task best?

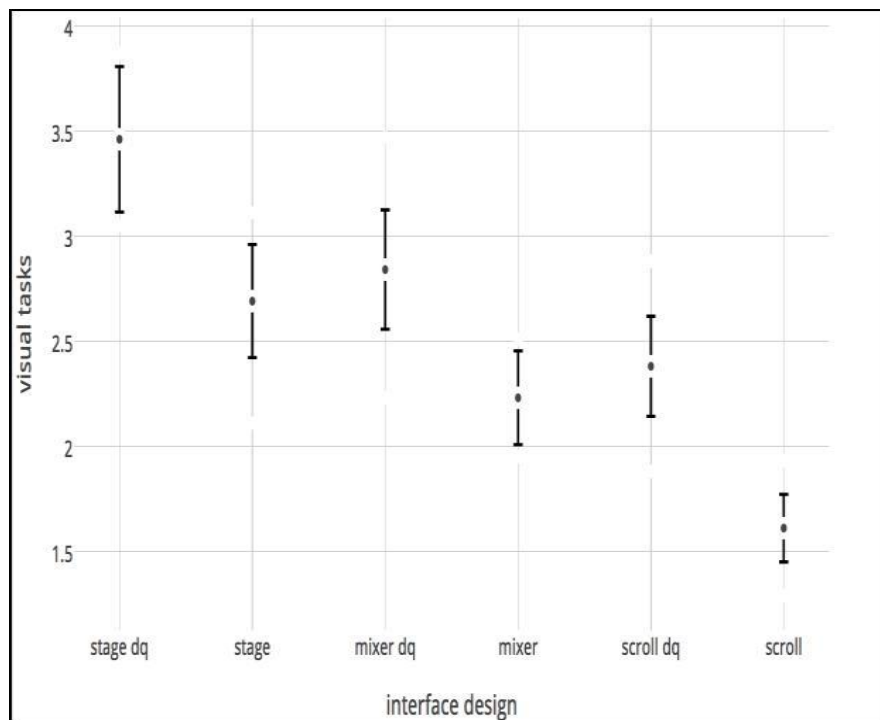


Figure 8:3: Visual searches successfully completed; Confidence Intervals at 95%.

There is an increase in the amount of visual questions answered with the DQ versions of all the interface designs. Except for the mixer DQ design, the stage DQ interface yields a significantly greater amount of correctly identified visual information than any of the other interfaces used in this study.

Analysis of the survey shows that the SM design was rated favourably on all measures. This is especially notable given its novelty to the majority of participants. Indeed, the SM interface had the highest number of respondents rating it as the interface they felt most comfortable using (figure 8.4). Furthermore, when asked which interface they thought had helped them to successfully complete both the visual task and listening task, the majority of participants named the SM DQ design (see figures 8.6 and 8.7 for precise figures).

Stage DQ	Stage	Mixer DQ	Mixer	Scroll DQ	Scroll
76.9	38.4	38.4	23	30.7	7.6

Table 8:2: The Percentage of correctly identified audio file attenuations per interface type.

Stage DQ		No	No	Yes	Yes	Yes
Stage	No		No	No	No	Yes
Mixer DQ	No	No		No	No	Yes
Mixer	Yes	No	No		No	No
Scroll DQ	Yes	No	No	No		No
Scroll	Yes	Yes	Yes	No	No	
	Stage DQ	Stage	Mixer DQ	Mixer	Scroll DQ	Scroll

Table 8:3: Results of the Z-test for dependent groups analysis at 95% CI. ‘Yes’ indicates that there was a significant difference between the interfaces. The Stage DQ design had a significantly higher amount of correctly identified audio attenuations than the mixer, scrolling DQ and scrolling interfaces.

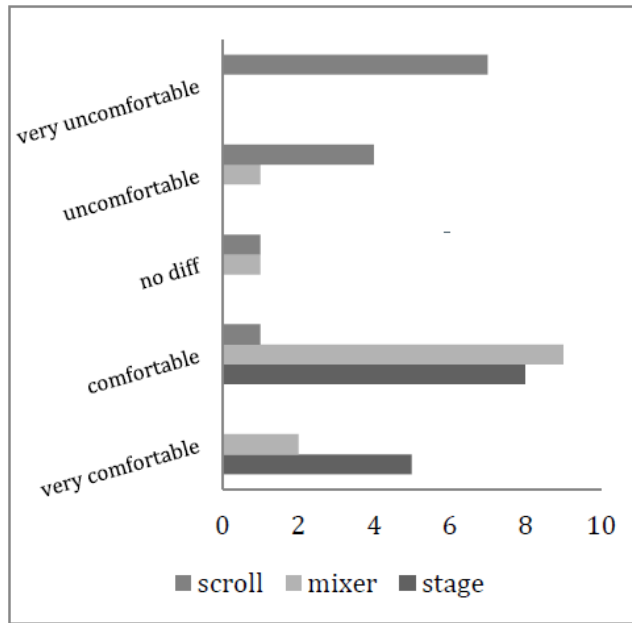


Figure 8:4: Results from the question ‘Which interface did you feel most comfortable using?’ The SM and CS mixers fare more favourably than the scrolling interface.

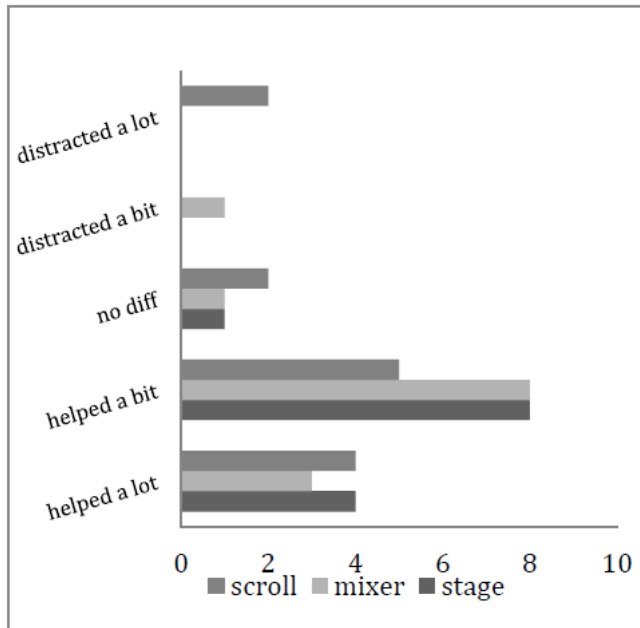


Figure 8:5: Results from the question ‘Overall how much did having the sliders help?’

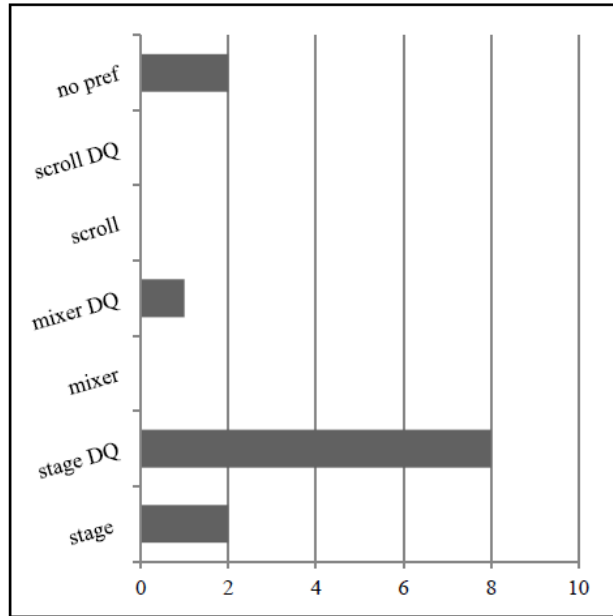


Figure 8:6: Results from the question ‘Which interface do you think helped you do the listening task best?’ The SM DQ design was perceived as the most effective.

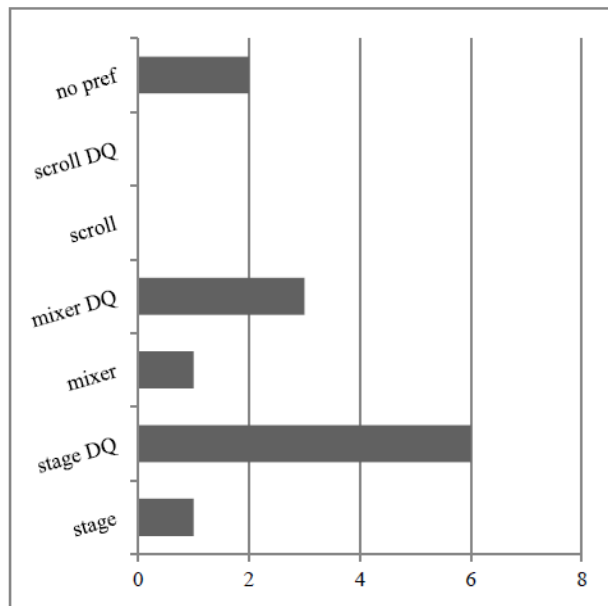


Figure 8:7: Results from the question ‘Which interface do you think helped you do the visual task best?’ The SM DQ and mixer DQ designs are perceived as being most effective. No respondents chose either of the scrolling designs.

8.3 Discussion

Improvements were found in all instances where DQ filters were added to the GUI. One explanation for this may be due to the relation between the user's perceptual abilities and the amount of visual feedback in GUI. In chapters 4 and 5 we found that visual complexity of the GUI per se does not affect simultaneous critical listening (users could ignore meters and visual distractors on the screen when they were not immediately relevant to task completion). However, it appears from the results of this study that searching more complex GUIs (when the visual feedback is directly related to the task) does appear to present an obstacle. In the 'Load theory' of conscious perception (Lavie 1994, 2005), it is suggested that the brain is constantly engaged in an encoding process of selecting which stimuli to attend to. Additionally, it is engaged in cognitive retrieval of that information, and execution of actions upon it (Begault et al 2016). Therefore, in our study, reducing the amount of extraneous detail in the UI may have reduced cognitive load and created spare capacity to attend to and process the subtle attenuation of the audio. Indeed, in an experiment on inattentional deafness, Raveh and Lavie (2015) found suppression of auditory stimuli, (even when the auditory stimulus was highly expected) was significantly reduced when the complexity of a visual target search was simplified. This suggests that it is not only the manipulation of working memory (WM) that results in inattentional deafness (MacDonald and Lavie 2011), but it may also be the manipulation of visual search complexity. Simplifying the interface during bi-modal tasks may therefore help prevent users from neglecting events on other (relevant) channels (Wickens & Alexander 2009).

In the introduction to this chapter, we considered whether the inclusion of DQ filters might distract users from efficient use of the interfaces. As mentioned above, the analysis of the quantitative data shows that this was not the case. In fact, DQ filters increased visual search and critical listening results in all designs in which they were included. Furthermore, the DQ filters appear to help address the random distribution of channels we experienced in our previous study, by allowing them to be examined according to user requirements, for instance, highlighting an individual channel to reduce time searching through extraneous channel information. However, it is important to note, more visually arresting visualisations may cause distraction to the user (Gelineck and Uhrenholt 2016) and the design of DQ filters should carefully consider the amount of visual attention they attract and the interaction required to access and modify them.

Finally, on a more subjective level, the inclusion of sliders was received favourably by the participants, with 84% of the responses rating them as helpful compared to non-DQ designs. This finding is in line with Gelineck and Uhrenholt's study on DQ filters (2016), in which participants found the idea of DQ filters an interesting and potentially useful addition to the interface design.

8.4 Summary

In this chapter, we continued our investigation of overviews of mix information. We briefly discussed literature on overviews, and recommendations that filtering the visual data may be beneficial in comprehending the data on several metrics, including search time, quantity of visual information and perceptual limits of users.

In section 8.1 we outlined the detail of our experiment quantifying the outcomes of interfaces incorporating filtering on search times and auditory acuity. The results of the study showed improvements on all measures when interfaces incorporated filters, especially so when the SM overview was used with DQ filters.

In section 8.3 we discussed the results of the study, and concluded that the layout of the visual information does indeed have an impact on aural acuity or speed of visual tasks. In line with Shneiderman's (1998) 'mantra' of 'overview first, details on demand', the use of query-respondent overviews does seem to allow quicker access to mix information and increase the speed and accuracy of concurrent visual search. We suggested that this may be due to DQ filters reducing extraneous detail in the GUI and minimising cognitive capture by the visual information. We concluded that aural acuity during simultaneous visual tasks can be improved, not only by reducing WM, but also by simplifying the visual search process, which in our case was achieved by filtering out (irrelevant) information.

8.5 Progression to subsequent study

The results from this study showed improvements on all measures when interfaces incorporated filters, this was especially marked with the SM overview design. However, we are aware that our SM mixer only displayed two channel parameters, namely pan position and volume. In order to improve the usability of an SM overview while mixing, further parameters may need to be included. In the next study we therefore refine the

design of the SM overview and assess how, in addition to showing a channel's volume and panning, we can add effects (reverb and delay). To do this we design SM overviews using varying UI elements to represent these parameters (including UI size, colour, saturation and transparency) and compare these with a CS mixer using dials and faders. We go on to assess the speed and accuracy of visual search in an eight-channel mix and quantify how the different designs affect the results.

Chapter 9

Representing Multivariate Mix Data.

In chapter seven, we concluded that the SM design supported integration not only of the interface as a whole, but also of each channel, as dual elements (faders and dials) were integrated into one User Interface (UI) object. We discussed how this relates to the notion presented in Cognitive Load Theory (CLT) literature (e.g. Chandler & Sweller 1992) that by integrating elements in Graphical User Interfaces (GUIs), users no longer have to engage in Representational Holding (Mayer & Moreno 2003), which can subsequently decrease cognitive load. Indeed, within the field of data visualisation, multiple parameters are often integrated into one graphical object, by assigning data to specific visual variables such as position, size, shape, hue, saturation, texture, opacity and dynamics (Borgo et al 2012). Such designs, generally referred to as multivariate data objects, or glyphs, have been shown to reduce screen clutter, help support the interpretation of data and enhance visual analysis by allowing both inter and intra-record relationships to be more easily detected (Ward 2008).

While the implementation of multivariate objects is limited in audio GUIs, Ferguson et

al (2005) concluded that when visually displaying musical parameters (in their case loudness, noise and harmonic content) these are optimally understood if they are represented as different attributes of *one* object, as this takes into account the limitations of human attentional bandwidth. Furthermore, research by Dewey and Wakefield (2016) has shown that the use of icon based mixers can not only reduce cognitive load but also increase immersion in the mixing task. However, due to the limits of human visual perception, there are constraints on the design of multivariate data objects (Cleveland 1993). For example, while colours can be easily interpreted when displayed at reduced sizes (Stone 2006), they are liable to certain caveats, beyond the obvious considerations of colour blindness. For example, in chapter 6, using sixteen different colours resulted in an error rate of 8.8 %. This may increase as the colour palette extends to represent more values and the colours used become more similar. Similarly, if using size to represent mix parameters, one needs to consider at which point the differences become too similar to efficiently convey ordinal data. Furthermore, some studies suggest that visually representing several streams of information at the same time can increase cognitive processing load (Gudur et al 2009, Gelineck & Overholt 2015). If this is the case, multivariate designs may become counterproductive, as the cognitive load involved in analysing and interpreting the information conveyed starts to impact negatively on task completion.

The experiment described in this chapter therefore evaluates the efficacy of multivariate data designs to visually represent mix parameters and assesses how fine a range of values can be represented using different designs, relating these to human visual perception. We use an SM mixer for evaluation, comparing the visual search times and accuracy to a CS mixer. It is important to note that the experiments outlined in this chapter are preliminary, in the sense that they focus exclusively on visual aspects of information search. In the following chapter, we take the findings from these studies and incorporate them into mixing GUI designs using audio to assess and quantify their influence on concurrent audio analysis tasks.

9.1 Study A: Representing an Additional Mix Parameter

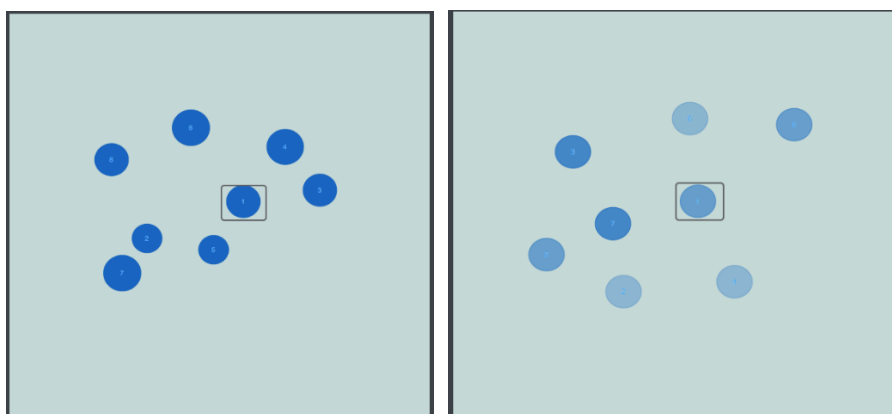
9.1.1 Participants

The Participants comprised first and second year students on a two-year music technology course at City and Islington College, London. All participants had at least

one year's experience mixing on Logic Pro (with a minimum of five hours a week exposure to DAWs and mixing). Sixteen participants were selected (10 male, 6 female, aged 17-19). All students were studying on the Edexcel GCE A Level Music Technology course. Due to the differences between the participants (1st and 2nd year students) we used a Repeated Measures design, in which all participants used all the experimental interface designs. None of the participants had taken part in the previous studies.

9.1.2 Study Design

Five eight-channel mixers; a CS design and four SM mixers (figures 9.1, a-e) were designed using Max/MSP showing each channel's volume, pan and reverb amount (reverb is a commonly used audio effect used to simulate real acoustic space, giving sounds a sense of ambience in the mix). As the visual representation and interpretation of the mix data was the object of the investigation, no audio was used. Each mixer design was a visual representation only. The term reverb was used solely to contextualise the visual tasks and place the additional parameter within an audio mixing framework.



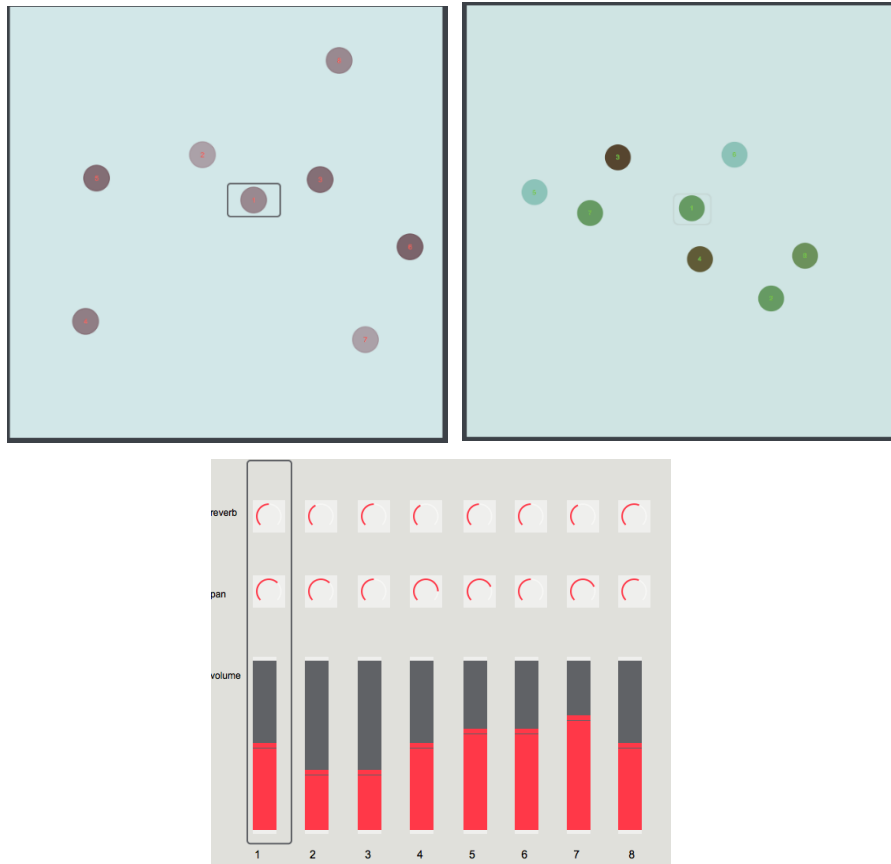


Figure 9:1: Screens for study A, clockwise: (a) size, (b) transparency (c)colours (d) channel strip (e) saturation.

For the CS design, faders were used for volume, while dials were used for the pan position and the reverb. For the SM designs, x and y positions were used for the pan and volume, while four designs were used to represent the reverb using visual variables commonly employed in glyphs (Borgo et al 2012). These included size, transparency, saturation (single colour) and hue (multiple colours). Rate of flashing (dynamics) was not used due to concerns that this might trigger seizures among people with photosensitive epilepsy. Shading was discounted due to the difficulty of interpretation at the high zoom levels required to analyse an overview, and shape was not included since it is chiefly a categorical data set and the current study examines representing ordinal data (such as the amount of reverb on a channel etc.).

The objective of the study was to ascertain how subtle a difference could be visually perceived between channels with different reverb amounts and how fine a range of values could be represented using each design. In order to do this, the reverb's range (1-100) was divided into increments of five, ten and twenty values and assigned to each

design. With increments of five there were twenty reverb values (100 divided by 5), for increments of ten there were 10 reverb values, and for increments of twenty, five reverb values. Increments of less than five were not included due to perceptual issues; colour schemes divided into multiple steps become increasingly hard to differentiate, with the values represented becoming difficult to distinguish (Harrower and Sheesley 2005). Furthermore, some displays will not accurately display small colour differences due to varying visual display characteristics (ibid).

To represent reverb values using colour and saturation, twenty gradients were created for both saturation and colour (fig 9.2). For increments of 5 (where the reverb was divided into twenty values) a separate gradient colour/hue was assigned to each reverb value. So, for example, reverb value 8 was represented by gradient colour/hue 8. For increments of ten, alternate gradients were used (ten gradients for ten reverb values). In this case, rather than assigning each of the twenty gradients to reverb, every other gradient was used, e.g. gradient 1 represented the lowest reverb value (1), gradient 3 represented reverb value 2, and so on. For increments of twenty, every fifth gradient was used (one for each of the five reverb values). Here gradient 1 represented reverb value 1, gradient 5 represented reverb value 2 etc. In all cases, darker colours were used to represent less reverb. For size, the difference between the minimum and maximum circle diameter was divided into twenty sizes. To represent increments of five reverb values, all twenty circle sizes were used, with each one representing a separate reverb value. For increments of ten, every other circle size was used, and for increments of twenty, every fifth circle size was used for each reverb value. Finally, the same method was used for transparency; the most and least transparent settings were divided into 5, 10 and 20 differences and assigned to reverb amounts with the most transparent settings representing the most reverb.

For each of the five mixer designs (channel strip, size, colour, saturation and transparency) a target was included in the eight channels and placed within a border (fig 9.1). For each design, three screens were created; one with reverb differences between the target and other channels set at +/- 5 (increments of 5), one with differences set at +/-10 (increments of 10) and one with differences set at +/-20 (increments of 20). This created a total of fifteen screens for the study.



Figure 9:2: Colour gradients used in the studies: Top; single colour saturation (less saturated colours were mapped to greater reverb amounts). Bottom; colours (lighter colours were mapped to more reverb).

9.1.3 Procedure

Each participant was presented with each mixer design at the three increment differences between target and other channels. This meant that, for example, on the screens showing increments of 5, if the target reverb value were set to 50, the other channels would all be 45 or 55 except for one other channel that was also set to the target's value. For each screen, participants were asked to identify which of the other channels on the mixer had the same reverb value as the target channel by clicking on the corresponding channel. The screen order was randomised for each participant and they were presented one after the other. The mapping of the designs to reverb amount (e.g. larger circle size to more reverb) was explained to each participant and they were given time to familiarise themselves with the different interface designs using practice screens. Participants were asked if they suffered from any known form of colour blindness prior to the test (no respondents reported this).

9.2 Analysis and Results

As with the previous study we used Confidence Intervals (CI) and a χ -test for proportion dependent groups to analyse the data. We chose this approach as the CI (at 95%) analysis allowed us to quantify any overlap between participants' results from the experimental interfaces to see the range of the true population per interface type, with non-overlapping CIs suggesting a statistically significant difference between GUI designs. The χ -test (at 95%) was chosen so that we could quantify the significance of

the percentage of correct answers per interface type, again to determine if the various interfaces tested in the study were significantly different from one another in terms of percentages of correct answers. We kept a broadly similar analysis method to the previous two studies to maintain a uniformity of analysis methods across the experimental chapters.

The amount of errors (incorrectly identified channels) was calculated for each participant in each of the fifteen interfaces. From this, the total number of errors made on each screen by all participants could be calculated (table 9.1). The results show that within all designs the error rates increased as the visual differences between the target and other channels' reverb values became smaller. However, the most errors for all differences were found in the dials and transparency designs. Size, colour and saturation resulted in fewer errors even at smaller differences (where the target and other channels were visually similar).

In order to test the significance of the error rates found between the different mixers, the data was analysed using a z-test for proportions dependent groups at 95% Confidence Intervals (CI). The results of the analysis show that the difference between the dials and transparency compared to the other designs was significant for increments of 5 and 10 per cent differences. However, the analysis showed no significant difference in accuracy between size, colours and saturation (though size had the least errors).

Increments between target and other channels' reverb amounts.	5	10	20
Dial	68	50	18.7
Colour	25	18.7	12
Saturation	25	18.7	6.2
Size	18.7	6.2	6.2
Transparency	68	65	31.2

Table 9:1: Error rates (%) for each design at different value differences between target and other channels. Correctly identifying similarity between the channels was

worst for the dial and transparency designs at all increment differences. Size proved the least error prone, with saturation and colour being generally evenly matched.

After completing the study, the participants were asked if they agreed with the mappings (i.e. lighter colours to represent more reverb). This was done by responding 'yes' or 'no'. There was also space to make additional comments regarding the mapping if required. This question was not addressing whether they had understood the mapping (this was explained clearly at the start and practice time was given to familiarise themselves with the designs) but rather if it fitted their conceptual map of reverb mapping. For the colour mapping, seven of the participants responded that that they felt it should have been mapped the other way around, e.g. darker colours represented more reverb. This issue did not occur with size, where all participants were agreed with "bigger is more" metaphor. This was also less of a problem with the saturation of the single colour where less saturated was more readily understood as representing more reverb (only two participants responded that they disagreed with this mapping). While all agreed with the transparency mapping, a number of the participants made additional comments suggesting that they found the transparency design very difficult (the numbers below refer to the participants' order in the study test):

P3: 'the transparency was too subtle'

P7: 'This was so hard'

P9: 'Hardest one by far'

P1: 'The differences are so similar; it's difficult to tell them apart'

P11: 'They looked the same'.

9.3 Study B: Adding a Further Mix Parameter

9.3.1 Participants

Participants were comprised of staff and students on a two-year music technology course at City and Islington College, London. All participants classified themselves as beginners, and all had some previous experience mixing audio using DAWs. For study b, twelve participants were selected (7 male, 5 female, aged 17-35). Separate participants were used for studies one and two to avoid the risk of possible learning effects.

9.3.2 Study Design

This study was designed to evaluate the efficacy of adding two mix parameters (reverb and delay) in addition to panning and volume. Again, this was done using both CS and SM designs. As with study a, no audio was used, as the aim of the study was to focus on and evaluate the efficacy of visual representation and interpretation. The terms reverb and delay were used to place the visual tasks within an audio mixing context, rather than specifically assessing these audio effects.

The choice of visual designs for the study was based on the results from study a. As outlined in section 9.2, size had performed best, while colour and saturation had both been equally successful. Transparency, however, had shown a significantly higher error rate (table 9.1), a result which corresponds with research suggesting that colour and size are the dominant visual factors and are most efficiently interpreted (Borgo et al 2012). For this reason, transparency was discounted for study b. Lastly, between colour and saturation, the latter was taken forward since it is a colour-blind safe design and due to the fact that multiple colours had resulted in disagreement from users over mapping. Again, a CS design using faders and dials was included so that a direct comparison could be made between designs. For the SM design, x-axis and y-axis were linked to pan and volume while reverb was linked to size and delay linked to saturation. As with study a, the reverb and delay parameters were given values of 100 steps, and the mixers represented these in increments of 20, 10 and 5 divisions.

9.3.3 Procedure

Participants were presented with both designs of an eight-channel mixer (figure 9.3) and were asked to identify a particular channel in relation to the target channel (surrounded by a border). For example, they were asked which channel was panned left of the target, of a higher volume than target, with the same amount of reverb and less delay than target? These tasks were chosen as they required the simultaneous analysis of all four visual channels (x and y position, size and saturation).

There were 18 screens in total. Nine SM screens and nine CS screens. Both designs included three screens with 5% differences between the target and other channels' delay and reverb settings, three with 10% difference, and three with 20% differences. So, for example, if the target had a setting of 50 on reverb and 75 on delay, the 5% difference would mean the other tracks were set to reverb being either 45 or 55 and delay of 70 or

80, with the exception of one other channel which was assigned the same reverb and delay settings as the target. As with study a, participants were asked to identify which one other channels had the same setting as the target channel by clicking on it with the mouse.

The order in which the mixers were presented was randomised for each participant. The reverb and delay values of the other seven channels were randomised for each participant (within variations of 5, 10 or 20 increments). The channel(s) chosen and the time taken to choose them were recorded for each participant, though this was not visible to them. Participants were given time to familiarise themselves with the mixer designs using practice screens before beginning the evaluation.

9.4 Analysis and Results

The amount of errors (incorrectly identified channels) were calculated for each participant in all eighteen screens. From this, for each mixer design, the error rate could be calculated for each of the three increment differences between the target and other channels (table 9.2). These results were analysed using a z-test for proportions (at 95% CI) to see if the error rate between designs at each increment difference was significant.

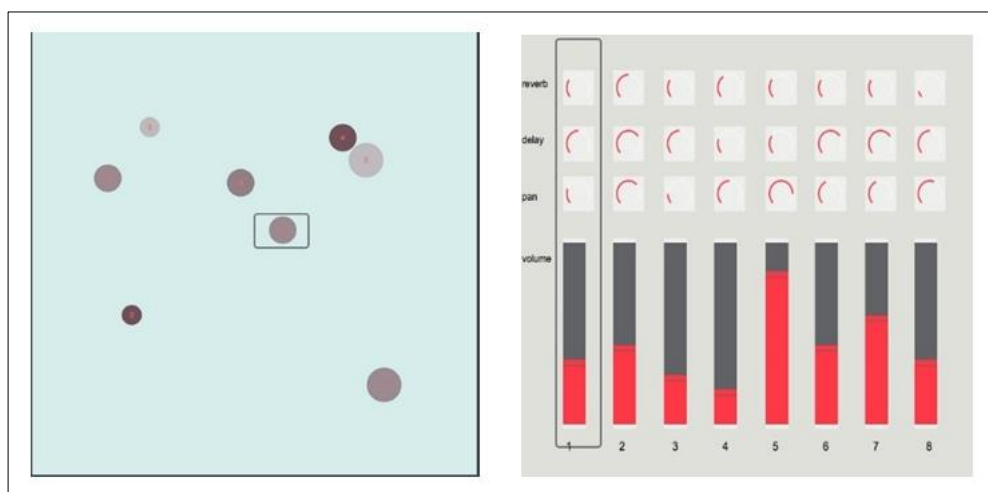


Figure 9:3: Left (a), the stage metaphor mixer; x and y positions show pan and volume, saturation of red colour shows delay amount and size shows reverb amount. Right (b) channel strip mixer; faders show volume, dials show pan, reverb and delay

The analysis shows that at 5% increments, there was a difference in error rate of 25 % between the SM and CS designs, providing a p value of 0.0131, suggesting that this is a significant difference at $p < 0.05$. For mixers where increments between channels were at 10%, the analysis shows an error difference of 27.8 percent, giving a p value of 0.003, which again is significant at $p < 0.05$. However, at increments of 20%, the analysis provided a p value of 0.3925, suggesting that the differences in error between the two GUI designs was not significant. This may be due to the orientation differences between dials being large enough at this level to reliably attract attention (Ozlak and Thomas 1986), and minimise any difficulties in estimating the quantitative information they represented.

The time taken to identify the correct channel was also analysed for each participant in both mixer designs at the different increment levels. From this the mean time and SD were calculated. This was used to generate CIs at 95%. The analysis revealed that in both designs, search times decreased as differences between channels became greater. It also showed that there were significant time differences in identifying the correct channels between the CS and SM designs, with the former taking longer at all increment levels (fig 9.4).

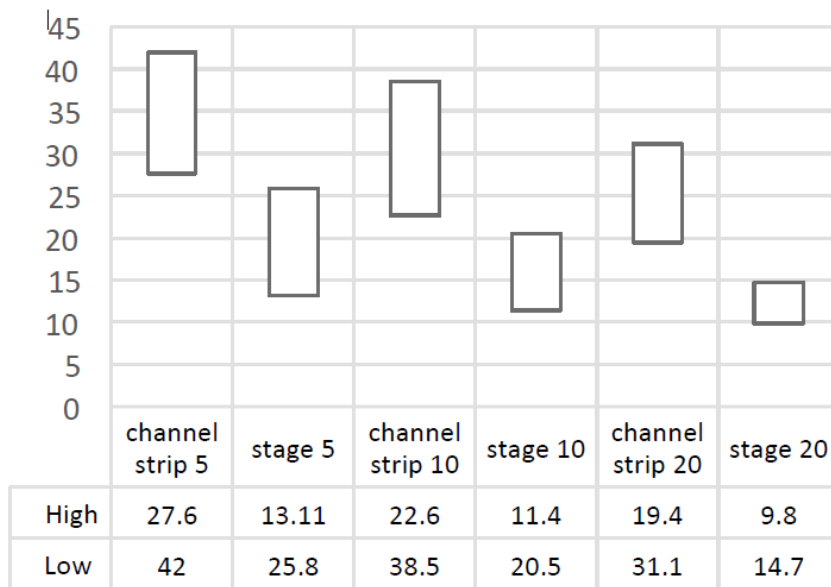


Figure 9:4: The visual search time (seconds) was significantly quicker in the stage metaphor design. However, in both designs, search times decreased as differences between channels became greater.

Increments	5	10	20
Channel strip	36.1%	33.3%	11.1%
Stage metaphor	11.1%	5.5%	5.5%
Difference in error rate	25%	27.85%	5.6%

Table 9:2: Error rates for both design at different value differences. 36 screens were shown for increment differences per interface type (3 screens over 12 participants). The analysis of the difference in error rate shows that the SM design was significantly more accurate at increments of 5 and 10 per cent.

9.5 Discussion

The results of the two studies reported in this chapter suggest that mapping mix attributes to a single multivariate object can result in improvements in visual search time and accuracy compared to 1-2-1 mapping, without any subsequent increase in error rates. The multivariate designs allowed users to find four separate mix parameters (pan,

volume, reverb and delay) more rapidly within one UI object than the four UI objects required in the CS design. As expected, this confirms our hypothesis that visually integrating parameters within UI objects helps to minimise representation holding, and simplifies visual search, both of which we have suggested contributes to increased processing of auditory information (sections 6.3, 7.2 and 8.3).

However, the results also raise awareness that the design of the visual channels used to encode additional mix parameters must be perceptually suitable, and cannot be assigned in an arbitrary manner. Multiple colours caused confusion over mapping, while transparency became difficult to interpret at reduced values. However, while not all visual channels used in the studies were equally effective, there may still be uses for them. For example, transparency may be useful for showing coarser values, such as muted and unmuted channels or indicating occlusion in mixes where channels visually overlap (Harrower & Sheesley 2005). Multiple colours, while prone to mapping confusion, may be suitable to more ordinal tasks such as identifying which channels are grouped together, such as vocals, drums, percussion instruments etc. (Ronan et al 2015). Furthermore, the relative novelty of the colour mappings in this study may be a factor in confusion, and prolonged use may lead to a greater acceptance. (Borgo et al 2012, p.2).

Lastly, it is important that differences between channels are kept within visual perceptual limits (for example making sure UI object sizes or colours are not too similar). This supposition is in line with Guided Search Theory (Wolfe 2006), which suggests that the preattentive processes which guide visual attention are influenced by salience factors, such as the target-distractor similarity, with reaction time directly influenced by these salience factors (Zenon et al 2008). Therefore, as visual differences between channels become more heterogenous, it may become harder to perceive differences. Indeed, error rates increased as visual differences between channels decreased (table 9.2).

9.6 Summary

In this chapter, we continued our investigation of multivariate UI objects, which we began in Chapter 7. We briefly covered some background on multivariate data objects, and discussed how they may reduce screen clutter, help support the interpretation of data, enhance visual analysis and improve the perceptual limits of users.

In section 9.1 we outlined the detail of our first experiment, quantifying the outcomes on visual search of interfaces incorporating three parameters into one UI object. The results, in section 9.2, showed improvements on all measures when interfaces incorporated parameters into a single UI object, compared to a one-to-one mapping, as long as the mappings used were perceptually relevant.

In section 9.3 we described our second study, in which further mix parameters were incorporated into a single UI object, using the results from study a to inform the mapping designs. In section 9.4, we discussed the results of the studies and concluded that at differences between channels of 5% and 10%, the speed and accuracy of visual search is significantly improved using multivariate UI designs.

In 9.5, we discussed our findings from both studies. We suggested that while multivariate objects can help convey information efficiently, they need to be designed in a way which is perceptually relevant, and cannot be assigned in an arbitrary manner. For example, we found multiple colours caused confusion over mapping, while transparency became difficult to interpret at reduced values. In the next chapter, we address the lack of audio in these studies, and incorporate an eight-channel audio mix into the experimental designs.

9.7 Progression to Subsequent Study

In this study, we found that incorporating several mix parameters into a single UI object using an SM mixer resulted in a quicker and more accurate search for specified mix parameters compared to CS mixers using a 1-2-1 mapping. However, our SM overview only showed limited elements for each channel, namely volume, pan and effects (reverb and delay). In the next study we wish to assess how we can display further channel parameters in an SM overview. To do this, we design an eight-channel SM mixer incorporating volume, pan, EQ and effects for each channel. To address the concerns with a congested interface, discussed in study 5, we add filtering to the interface to manage visual feedback and allow display of the mix parameters to be tailored per user reequipments. We compare this to a CS mixer using a 1-2-1 mapping with the same number of channels and channel parameters. We quantify the time and accuracy of participants undertaking several mixing tasks using the various designs and go on to

compare and assess the results. Furthermore, as the next study incorporates all the designs from the previous experiments (mix overviews, information filtering and multivariate UI elements) it acts as a prototype mixer and allows us to test and evaluate our combined design decisions detailed within our previous empirical chapters.

Chapter 10

Mixing Interface Prototypes.

The experiment outlined in the previous chapter showed that for visual search of mix attributes (such as reverb and delay), a GUI that combined several parameters into one graphical object was faster and more accurate than a design using 1-2-1 mapping. However, the previous study was focussed exclusively on visual search and it did not include any audio component. In this study, therefore, we address this issue by adding audio to the mixers, using the findings from the previous studies to inform the designs. We undertake several mixing tasks with the novel interfaces, and compare these with a CS mixer to analyse their efficacy for audio mixing workflows. We use the results to test, develop and refine our heuristics, which we present in chapter 11.

10.1 Study Design

10.1.1 Participants

Twenty-four participants (aged 16-17) took part in this study. All the participants were drawn from the same population, namely first year students on the Edexcel GCE Music Technology A Level at City and Islington College, London. All the students were from the same year of the course. As the students were recruited from the same population, the experiment used an Independent Measures design, requiring participants to use only one of the three experimental interfaces (8 in each group). To compensate for potential variations between participants, we ensured they all had comparable experience of DAW mixing, and equivalent training in audio mixing. This was done by recruiting participants from the same year of the course and comparing the experience of participants, none of whom had previous formal training in audio mixing prior to enrolling at the college. None of the participants had taken part in previous studies.

10.1.2 Interface Designs

Three eight-channel mixers were designed using Max/MSP. These comprised a CS mixer, an SM mixer, and a hybrid design mixer (combining functionality from both the CS and stage mixer). All designs showed each channel's volume, pan, reverb, treble and bass controls (figures 10.1).

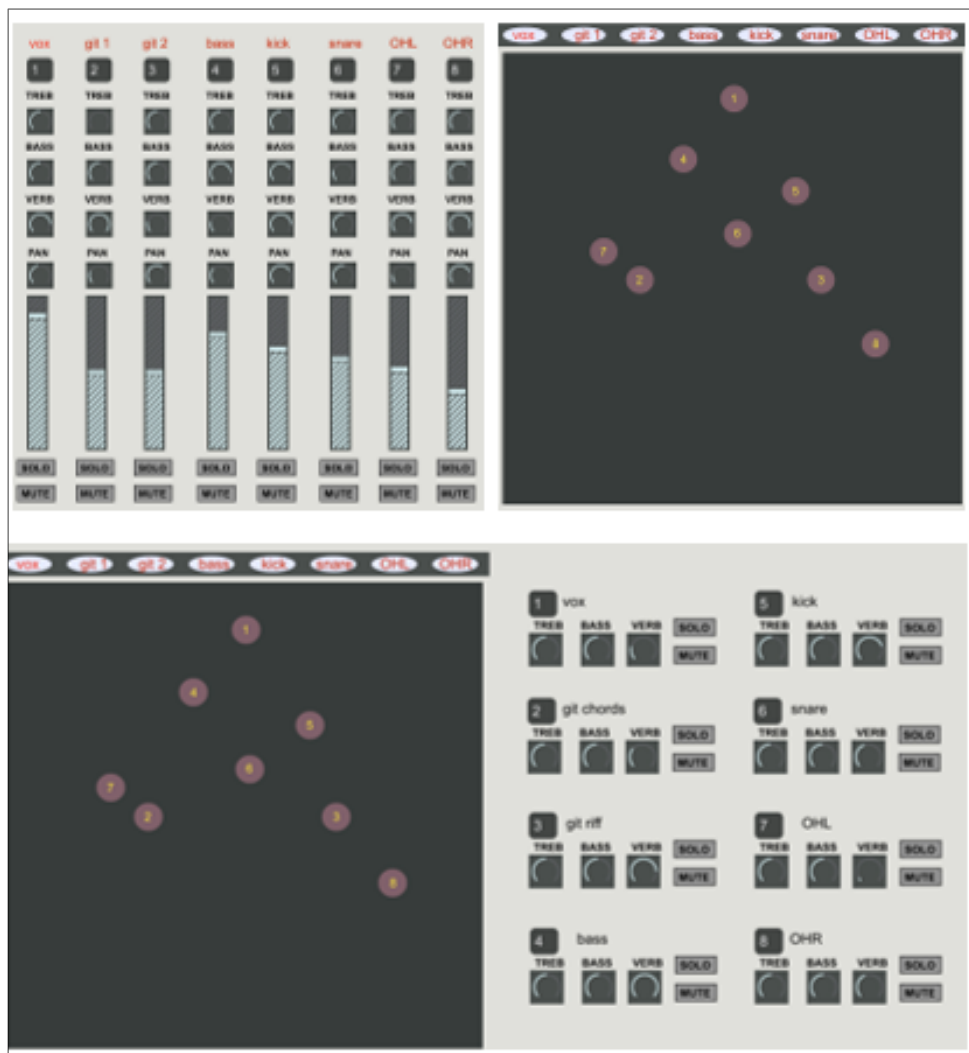


Figure 10.1: The three mixer designs used in the study. Top left, CS mixer. Top right, SM mixer. Bottom, hybrid mixer (combining the SM functionality with dials for treble, bass and reverb controls).

For the CS design, faders were used to adjust volume, while dials were used for the pan position, treble, bass and reverb amounts. For the SM and hybrid designs, each channel was represented as a circle (using Max/MSP's nodes object). Each channel's x and y position was used to adjust pan and volume respectively, while the relative size of each channel's circle was used to represent and control the reverb, bass and treble (figure 10.2). Clicking and dragging up or down on the nodes increased or decreased the circle size and the corresponding parameter value respectively. The choice of size to represent and modify frequency and effect amounts was in response to the previous experiment, which showed this to be the most easily identifiable visual channel for showing mix parameter differences (compared to transparency, colour or saturation).

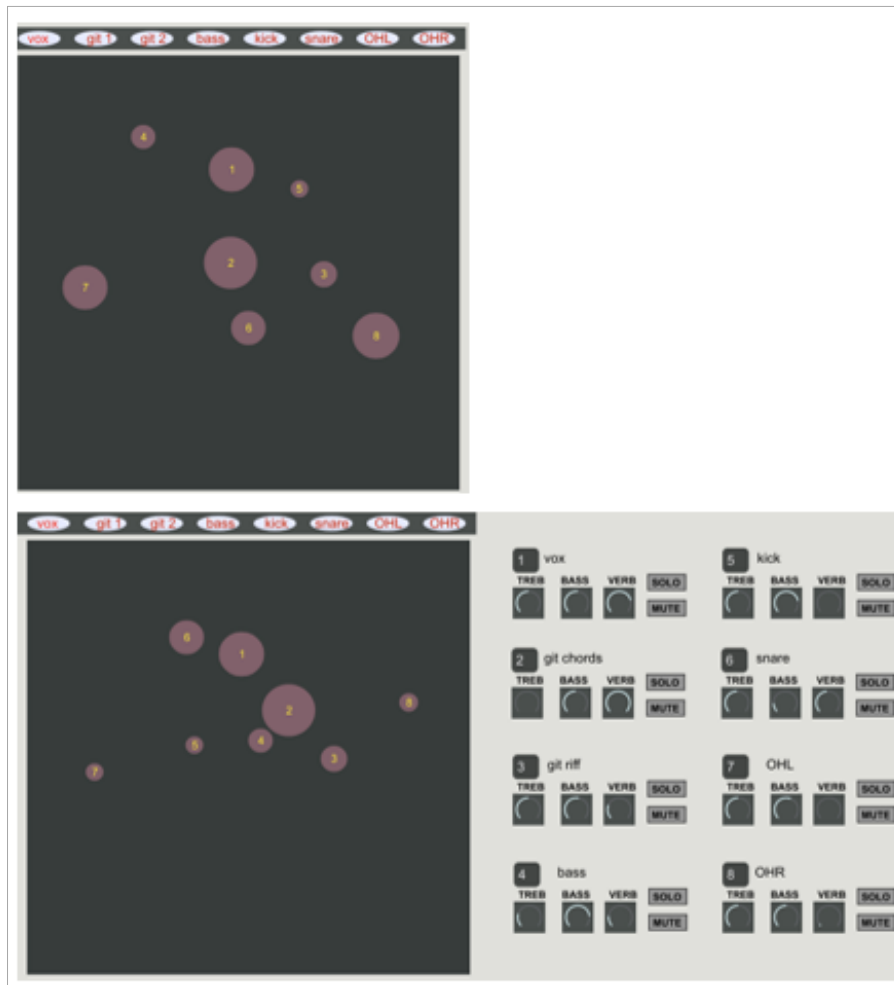


Figure 10.2: Size differences between channels for SM mixer (top) and hybrid (bottom). This allows users to see values of either treble, bass or reverb. Only one parameter (circle) can be viewed at one time. In the Hybrid design, values are also displayed as dials.

As node size was used for reverb, treble and bass, each parameter was viewed separately by pressing modifier keys. Pressing ‘r’ displayed the channel’s reverb amounts, pressing ‘t’ displayed treble, and pressing ‘b’ displayed the bass. When this was done, the pan position and volume of the channels remained constant, with only the circle size changing accordingly (figure 10.2). As soon as the modifier key was released, channels returned to the default view, in which all channel circles were the same size, regardless of parameter values. The decision to assign all three parameters to size was included in response to two concerns. Firstly, our previous study had suggested that filtering the amount of information in the interface decreased visual search times (without any reduction in concurrent critical listening response). Secondly, it addressed concerns that

when using an SM design, the legibility of the GUI may become compromised as multiple channel parameters are displayed simultaneously.

For all interfaces, the EQ used the MAX/MSP filtergraph parametric EQ. The treble control had a centre frequency of 5000 Hz, with a fixed bandwidth of 1.33 octaves and a boost and attenuation range of +18 dB and -18 dB respectively. The bass band had a centre frequency of 125 Hz, with a fixed bandwidth of 1.33 octaves and a boost and attenuation range of +18 dB and -18 dB respectively. The reverb used the Max/MSP 'reverb2' object, and the mixers controlled the wet/ dry level (wet refers to reverberant sound, dry refers to lack of reverb), with a range of 0 to 100%.

Finally, in the SM and hybrid designs, a list of names of the tracks could be clicked (e.g., bass, vocal etc.) to highlight the appropriate channel (fig 10.3). This was done to address the random distribution of channels, and potential problems of searching through the interface to find a target channel. We did not include this functionality in the CS design, where the channels are in a fixed numerical position, left to right, at all times and were labelled for ease of identification.

The decision to include a hybrid design was informed by three considerations. Firstly, we wished to investigate whether there might be disorientation effects in jumping from one view to another (as found in the SM design when using the modifier keys). As the hybrid mixer included dials as well as circles for the reverb, bass and treble values, it provided a secondary, constantly visible representation for these parameters. Secondly, we wished to assess whether layering of mix attributes (where only one parameter can be seen at one time) might slow down the mixing process (Liebman et al 2010). Finally, we sought to address whether including a secondary source of information may influence working memory (WM) or affect limits of visual bandwidth.

10.2 Procedure

10.2.1 Pre-test Screening

Three eight-channel practice mixes were created using royalty-free audio recordings. Before the experiment began, each participant was individually played each mix of the tracks (without any mixer or visual feedback). During this time, three separate instruments (vocal, high-hat, and electric guitar) each had a Low Pass (LP) filter (-18 dB

cut, 3000 Hz centre frequency, bandwidth of 2 octaves) applied for three seconds (one instrument per mix). As soon as they heard it, participants were asked to identify which instrument had the LP filter applied. Any participants who were not able to identify the instrument would have their results removed from the study. In the event, all participants answered these screening questions correctly, suggesting that without any visual stimuli, it is possible for the participants to clearly discern and identify this level of frequency attenuation within an audio mix.

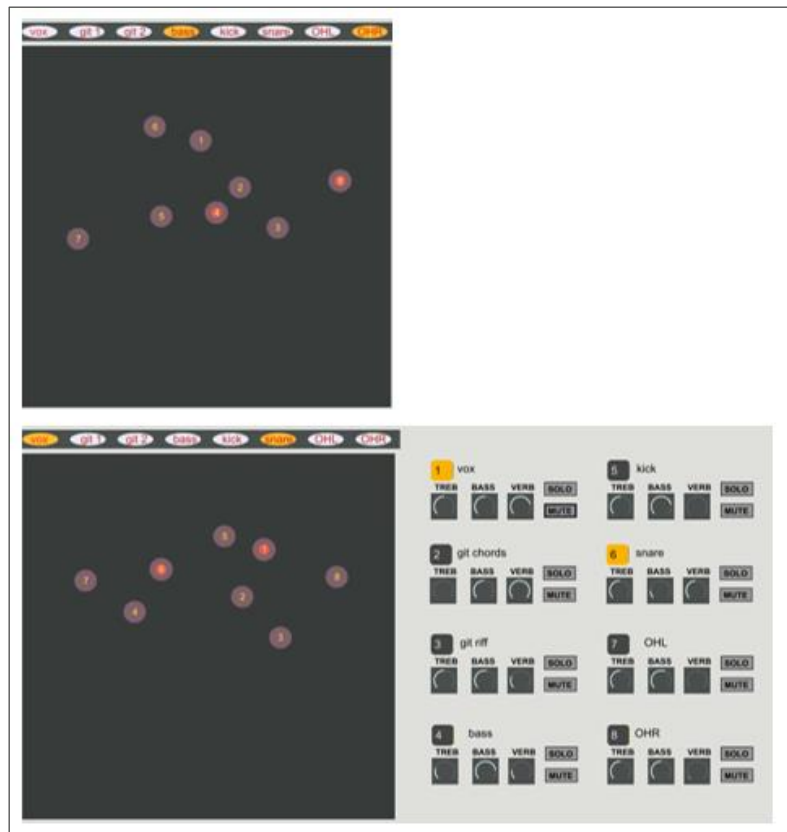


Figure 10:3: Selecting the instruments from the list at the top of the screen highlights the relevant channel(s). This functionality was not added to the CS design, as the channel order remains constant.

10.2.2 Test Procedure

For the test, an eight-channel mix (duration 2 minutes 50 seconds) was created using royalty free samples (see table 10.1 for list of instruments), which was used in all interface designs. Each participant was shown the interface design that they would be using. This was either the CS, SM or hybrid design. Random assignment was used to

allocate participants to the different interface types, with eight participants using each mixer type. The controls and functionality of the interface design were explained, and the participants were given time to practice using the mixer with a separate eight-channel mix. This was not time-bound, and participants were informed that they could spend as much time as they liked building their familiarity and confidence with the interfaces. Once they were happy they were told to begin the experiment.

Pressing a 'ready' button on the screen revealed the first of six mixing tasks, written as text on the screen (table 10.2). The tasks included in the study were chosen as they deal with the fundamental elements common to all good mixes (Moylean 2007). These comprise balance: the volume level between musical elements, frequency range: the correct balance of frequencies in the mix, panorama: correct placement of sounds in the stereo field, and dimension: creating depth and ambience through use of reverb (Owsinski, 2006, p.10).

Track	Description
Kick	Mono acoustic kick drum
Snare	Mono acoustic snare drum
Over-head L	Mono over-head drum kit recording- panned left
Over-head R	Mono over-head drum kit recording- panned right
Bass	Mono electric bass
Guitar	Mono electric guitar chords
Guitar 2	Mono arpeggiated guitar riff
Vocal	Mono male vocal

Table 10.1: The list of instruments used in the multi-track recording given to the participants. Eight tracks are typical of a small studio or Live Sound mix.

Once the participants had read the mixing task, and acknowledged that they understood

what was required, they were told to press the ‘start’ button. Once pressed, the mixer appeared on the screen, the audio of the eight-channel mix started, and the participants began undertaking the required mixing task. The channels began in the same position for all the mixers used. This was an unmixed position, with all sound panned left to right in channel order, with all channels set at the same volume. This meant that there was no overlap between channels in the stage mixer designs. During the mixing process, the LP filter was applied for three seconds to one of the three specified instruments (vocal, high-hat, or electric guitar riff) within a randomised period of 2-12 seconds of the participant interacting with the interface controls (e.g. moving a dial, clicking on a channel etc.). This was done to ensure that the visual and auditory tasks were completed simultaneously. Once the mixing task was complete, the participants were asked to press a ‘finish’ button. This saved their mix and completion time, and revealed a screen asking them to select which of the three specified instruments had the LP filter applied. This list included a ‘couldn’t tell’ option to discourage the participants from guessing. As soon as they had entered their response, the mix reset and the instructions for the next mixing task was presented on the screen. This procedure was repeated for all six mixing tasks.

10.2.3 Mixing Tasks

The six mixing tasks presented to the participants ranged in the level of difficulty of visual search (table 10.2). Tasks 1 and 2 required users to visually search for one User Interface (UI) object to complete the mixing procedure (e.g. the position of the bass dial/ circle size). Tasks 3 and 4 required visual search for two UI objects, while tasks 5 and 6 required participants to search for 3 UI objects. Participants were not asked to mute or EQ the target channels (those which had the LP filter applied to them) in any of the tasks. This was done to ensure that these tracks were always audible and remained constant in frequency balance, thereby allowing users to hear any frequency attenuation. The order of the mixing tasks was randomised so that the difficulty was not progressive and improvements due to learning and practice were minimised. After the experiment, each participant was asked individually if they had any additional thoughts or comments on using the interfaces. These were written down at the time, and later analysed for common trends or themes.

Question number	Mixing task presented to participants	Number of UI objects	UI Object type
1	Match the bass of channel 7 to channel 8	1	Bass
2	Mute all channels with volume below the bass.	1	Volume
3	Remove reverb on the channel panned furthest left and the channel panned furthest right.	2	Reverb / panning
4	Pan tracks with most bass to same position as channel 3	2	Panning / bass
5	Mute channels panned left of channel 4 which have more reverb, but less treble.	3	Panning/ reverb / treble
6	Mute any channels which have volume below the snare, more bass than the snare and more reverb than the snare.	3	Volume / bass / reverb

Table 10:2: Mixing tasks given during the experiment. The number of User Interface objects that need to be checked to complete the mixing tasks vary between and 1 and 3, with two questions for each

10.3 Analysis and Results

As with the previous studies, and in order to maintain consistency across the experimental chapters, we used Confidence Intervals (CI) and χ -tests for proportion dependent groups to analyse the data. We chose this approach as the CI (at 95%) analysis allowed us to quantify any overlap between participants' results from the experimental interfaces to see the range of the true population per interface type, with non-overlapping CIs suggesting a statistically significant difference between GUI designs. The χ -test (at 95%) was chosen so that we could quantify the significance of

the percentage of correct answers per interface type, again to determine if the various interfaces tested in the study were significantly different from one another in terms of percentages of correct answers, and whether that difference could have occurred by chance. Finally, any participants' comments or feedback given following the tasks was recorded (these were voluntary and not all participants commented). These responses were then analysed to collate any common themes and identify patterns and trends that may have been voiced using a rudimentary form of Thematic Analysis.

The time taken to correctly perform the mixing tasks, and the time taken to discern the LP filters were analysed for each participant. From this, the mean time and SD were calculated per interface type. This was used to provide CI, at 95%, to ascertain if correctly completing the mixing tasks or hearing the frequency attenuation was faster on any of the interface designs. The amount of correctly completed mixing tasks (adjusting the correct parameters on the correct channels) and correctly discerned LP filter were also recorded and analysed for each participant. This data was then subjected to a z-test for proportions, to ascertain if there were any significant differences between interface designs. Finally, participants' comments were analysed to collate any common themes and identify any patterns or trends that may have emerged.

10.3.1 Speed to Complete Tasks

The analysis of task completion time (fig 10.4), shows that mixing tasks, which required analysis of one UI object, did not result in any significant time difference between the designs. At two parameters, however, the CS was significantly slower than the hybrid and SM design. At three parameters, the amount of correct answers from participants using the CS design was so small that it resulted in a margin of error too large to create a meaningful CI figure. For the hybrid and SM designs, the speed of completion remained constant between one and two UI objects, becoming significantly slower when three UI objects were involved. However, even with three UI objects, they still resulted in faster mean completion times than two UI objects in the CS mixer.

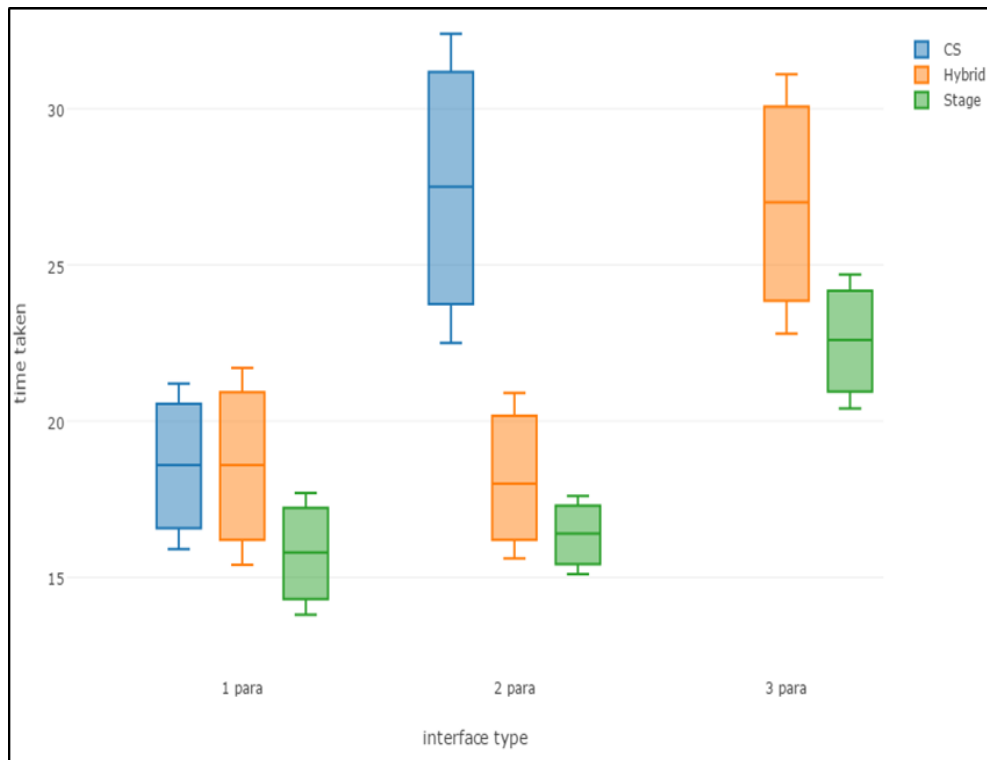


Figure 10:4: Confidence Intervals for time taken to correctly complete mixing tasks (in seconds, y axis) by interface type (x axis) and UI object amount. When more than one UI object needed to be checked to complete the mixing task, the mean time was significantly worse for the CS compared to the other designs

10.3.2 Amount of Correctly Completed Tasks

The χ -test for proportions analysis (table 10.3) shows that in terms of task completion, the difference between interface designs was not significant when one or two UI objects had to be searched for. However, when searching for and analysing three UI objects, a significantly greater number of participants correctly completed the tasks with the hybrid and SM mixers. Analysis of correctly identifying the channel with the LP filter shows a similar trend (table 10.4). When two or three UI objects had to be found and analysed, the percentage of participants who successfully identified the frequency attenuation increased significantly with the SM and hybrid designs, compared to the CS.

Finally, the z-test analysis for the percentage of users per interface type who completed the mixing task and the listening task (table 10.5), showed that results were significantly improved with the SM and hybrid mixers compared to the CS design in all the mixing tasks involving more than one UI object.

parameters	CS	Hybrid	Stage	Significant?
1	87.5 (7)	100 (8)	100 (8)	N
2	87.5 (7)	87.5 (7)	87.5 (7)	N
3	50 (4)	75 (6)	87.5 (7)	Y

Table 10:3: Mixing tasks. Percentage of participants correctly completing the mixing task (per interface type and parameter amount) with the significance of difference between CS and hybrid/ SM designs.

parameters	CS	Hybrid	Stage	Significant?
1	87.5 (7)	100 (8)	100 (8)	N
2	50 (4)	100 (8)	87.5 (7)	Y
3	25 (2)	87.5 (7)	87.5 (7)	Y

Table 10:4: LP filter. Percentage of participants who successfully identified the LP filter per interface type and parameter amount, with significance of difference between CS and hybrid/ SM designs

parameters	CS	Hybrid	Stage	Significant?
1	87.5 (7)	100 (8)	100 (8)	N
2	37.5 (3)	87.5 (7)	87.5 (7)	Y
3	12.5 (1)	75 (6)	87.5 (7)	Y

Table 10:5: Both tasks. Percentage of participants who successfully completed both tasks (correct mixing and hearing LP filter), per interface type and parameter amount, with significance of difference between CS and hybrid/ SM designs.

10.3.3 Accuracy of Mixing Tasks

The analysis reveals conflicting results regarding accuracy when manipulating channels using dials compared to circle size. In question 4, the accuracy of setting the pan position was worse when using the CS dials, with a mean error value above both other designs (figure 10. 5). However, when using the dials to set the bass amount (question 1), the difference in accuracy is not as marked, with the mean error rate broadly similar across the three mixers (figure 10.6).

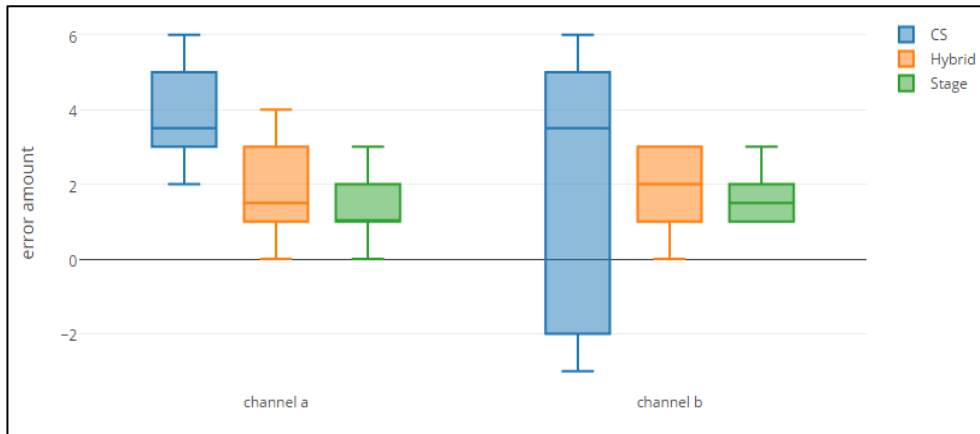


Figure 10:5: Analysis of the accuracy of panning the tracks in question 4. ‘0’ represents a perfectly matched panning position to the target channel. While not significantly different, panning was less accurate when using the dials compared to circle size

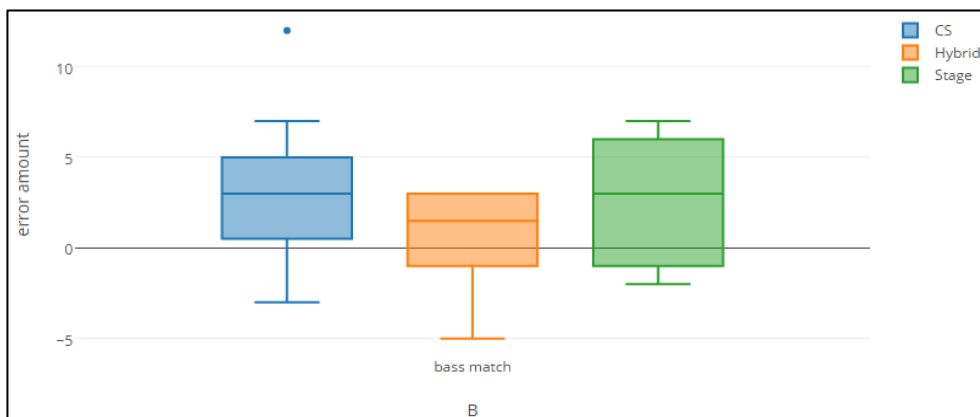


Figure 10:6: Analysis of the accuracy of matching the bass amounts in question 1. The accuracy appears to be broadly even among the three designs

This may be attributable to the fact that the bass matching between tracks could not be completed visually, as it was different on all three channels (matching the bass of channel 7 to 8 could only be done by attentive listening). Therefore, setting the dials/circles to the same positions would not have been effective. With question 4, however, the same dial position/ node size represented the same panning position, consequently, visual matching would be effective in setting the panning correctly (the amounts of bass could be checked visually, and the requisite tracks visually set to the same pan position as channel 3). This result suggests that visual cues were an important part of completing

question 4. This supposition is further confirmed by participant comments, many of whom stated that visual referencing of the mixer parameters was a key part of their mixing strategy. We discuss this further in section 10.4 below.

10.3.4 Analysis of Participant Comments

Following the tasks, all participants were asked individually if they wished to make any additional comments regarding their experience of using the software. Not all participants chose to make additional comments. Those comments given have been systematically analysed for common themes using thematic analysis methods (Cassell and Symon 2004) by analysing all responses for recurring answers. When referring to specific participants throughout this section, the letter refers to the mixer type, and numbers refer to the participant who used the design being discussed (e.g. HP1, Hybrid mixer, first participant).

10.3.4.1 CS Mixer

The use of dials was mentioned by participants in relation to several aspects of the task. These consisted of interaction. CSP3: “I found it quite fiddly to change the dials, getting them just right with the mouse”. Use of dials in the visual search was also raised: CSP3: “It was confusing with all the dials, finding their values was hard”. CSP8: “It was hard to concentrate on the mixer dials and listen out for the E.Q. at the same time”. CSP6: “It got quite stressful trying to find the right tracks and listen out too”

Search strategy was mentioned by two participants, specifically relating to the fact that visual referencing was used to address critical listening tasks. CSP6: “I found some channels by looking through settings on the mixer”. CSP3: “I could answer some of the questions by looking at the mixers tracks”. Familiarity was also raised. This was anticipated given the design. CSP2: “It was pretty much a standard mixer”, CSP6: “It reminded me of a small portable hardware mixer”.

10.3.4.2 Hybrid

Display of mix information was commented upon by participants. Firstly, while most participants didn't use the dials on the hybrid interface while mixing (only one of the participants was observed to make any channel adjustments with the dials) two respondents commented that it was good to have them. HP4: “The dials were useful

for checking what I'd done". HP3: "I didn't use the dials, but I quite liked them being shown". The use of the SM design to display overviews of mix information was also commented on. HP3: "I liked that I could check the volume of channels easily. It was clear to see just by looking, it helped that I could highlight tracks". HP1: "It's easy to comprehend- you can look at everything's overview". Interaction was commented on, specifically the use of the modifier keys to find information. HP1: "The [modifier] keys sped up checking the different parts of the mix" [the participant explained that this was in comparison to his experience using a CS mixer (Logic) encountered during the course]. HP5: "I liked the short-cut keys, they helped me compare the settings". The use of dragging to change the values was also raised. HP6: "It would be nice to have finer control over changing size, maybe pressing a key to get it to move more gradually". HP1: "dragging with the mouse is good, just takes a bit of getting used to". The design of the interface was also mentioned, one response positive, one less so; HP4: "it's an interesting design, good idea". HP6: "I generally prefer a normal one [virtual CS mixer], it's what I'm used to".

10.3.4.3 Stage Mixer

Simplicity of the interface design was a theme raised by three of the participants. SMP4: "It's much better than a traditional DAW for beginners". SMP8: "It's easy to get the hang". SMP7: "it's ok for getting going, but I'd like a normal mixer too for more complex mixing". Display of mix information was mentioned by six participants. Four made positive comments. They remarked that they found it easy to see patterns quickly, such as relative volume and panning between the channels. SMP8: "I really liked the circle size- it was clear to see the differences between tracks quickly, just with a look". SMP2: "I could gauge the whole mix by sight, and find what I needed by lighting up [highlighting] the instruments, that was quite nice". SMP1: "the presentation was clear, you could compare and see what was going on". SMP4: "I could see each channel's panning compared to the others quickly. I find that a pain on DAWs". However, two respondents mentioned the lack of precise ordinal information on the interface being a problem: SMP4: "It would be nice to have had dials or numbers too, for making sure the settings were 100% accurate". SMP7: "Having the dB figures somewhere on the screen would be helpful." Interaction was also mentioned. As with the hybrid design, use of modifier keys was addressed by participants. Two comments were positive; SMP1: "Toggling was useful, I could check straight away what each instrument had on it". SMP8: "Going from one view to another with the keys saved having to keep clicking on the screen". However, another participant stated that she would have liked to have

had the parameter which she was adjusting made clearer. SMP7: “I found it confusing switching the views; it would be good to have a ‘reverb’ box on the screen if I press ‘r’, so I know what I’m changing”.

10.4 Discussion

In all three GUI designs, the ability to correctly notice the LP filter reduced as the number of UI objects to be searched and analysed increased. This reduction was greatest when using the CS mixer (figure 10.7), with a 62.5% reduction in the ability to hear the LP filter when searching through three UI objects compared to one. In comparison, the SM and hybrid designs resulted in a less marked correlation between visual search complexity and aural acuity. In the hybrid design, there was a reduction of 25% in the number of participants who heard the LP filter when searching through three UI objects compared to one. For the SM design, this was further reduced to 12.5%. In fact, with the SM design, almost as many participants successfully heard the LP filter when searching through three UI objects, as participants using the CS design did when searching one.

The fact that the impact of increased visual search was less marked in the SM and hybrid designs may have been due not only to the design of the GUI, but also to the modifier key functionality included in the novel interface designs. As mentioned above, an observed strategy among participants using the novel designs was a toggling between views, indeed this functionality was commented upon by participants using both the SM and hybrid designs. For example, participants, were seen to rapidly check conditions of channels (such as the amount of bass and reverb) to ascertain which channels needed to be modified. This was especially marked in the tasks involving three UI objects. As comparisons of data sets can be made most efficiently via eye movements, (Plumlee and Ware 2006), the quick visual comparison may have helped minimise the load on visual WM and consequently reduced the search times (Baldando et al, 2000). Conversely, the intricacies of the CS mixer may have contributed to the slower and less accurate visual search. Indeed, it’s design may have caused participants to engage in inefficient search, subsequently directing attention away from the audio itself (Sabin and Pardo 2009., Dewey and Wakefield 2016).

While one participant commented that she found the use of the modifier keys confusing, concerns over disorientation effects caused by modifier keys in the SM

design seem not to have been a factor. In fact, the converse appears to be true, with the ability to rapidly change view appearing to be an advantage, rather than an impediment. This conclusion is further confirmed by the differences in the results between the stage and hybrid designs. The secondary, stable source of information provided by the dials in the hybrid design lead to a slight reduction in task speed and ability to discern the LP filter. This may have simply been due to users double checking the result with both sources of information (nodes and dials). However, the effects of Information Redundancy (e.g. Sweller, Ayres, & Kalyuga, 2011) present in the hybrid interface and the subsequent increase in UI objects may have resulted in a level of visual feedback surpassing that which can be efficiently processed in WM (Baddeley, 1998). Nevertheless, the differences were minimal, and inclusion of dials may be useful on a more subjective level as a confidence-builder in an otherwise unfamiliar interface design. This was noted in responses from participants too, two of whom mentioned that having further sources of visual feedback (dials or ordinal information) would have benefited their experience of using the novel designs (SMP4 and SMP7).

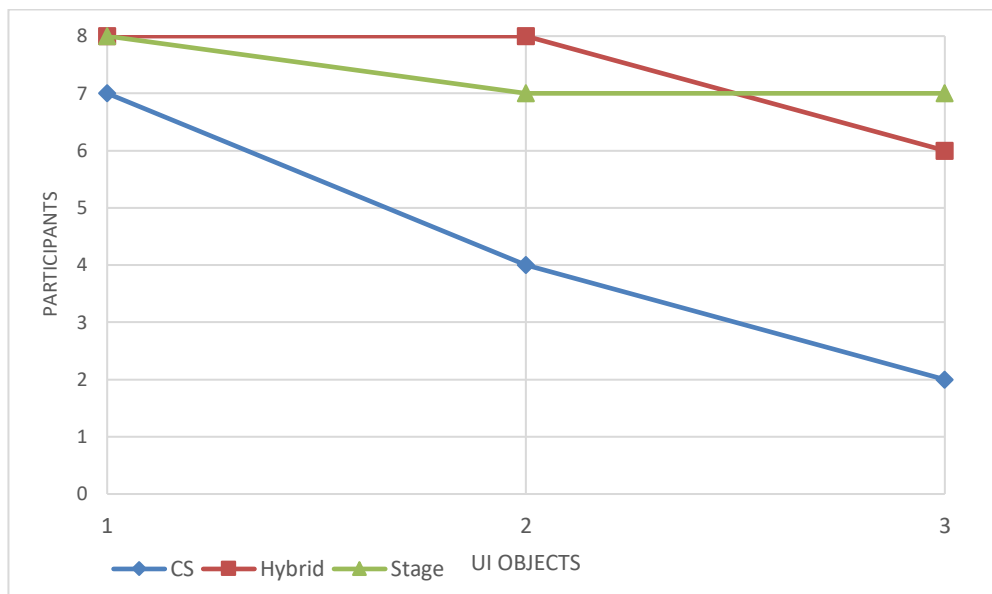


Figure 10:7: The reduction in number of participants to correctly identify the LP filter per interface type and UI objects.

10.5 Conclusion

In this study, we added audio to our GUIs using designs developed in chapter 9, and carried out mixing tasks addressing fundamental aspects of successful audio mixing workflow (Moylan 2007, Owsinski 2006). In line with the previous study, the results of this experiment suggest that by presenting channels as multivariate data objects, users can more accurately make visual comparisons of mix parameters compared to 1-2-1 mapping of the CS layout, resulting in a faster more accurate visual search and improved critical listening. Furthermore, and somewhat unexpectedly, the use of modifier keys to show mix parameters (such as effects and frequency) not only minimised screen clutter, but by supporting rapid toggling between parameters, allowed quick visual referencing between channels. We suggest this attribute of the functionality may have reduced WM load and allowed resources to be better shared between visual and aural modalities, resulting in faster and more accurate mixing, and better awareness of the LP filter being applied to the audio. The ability to highlight channels in the display, was only mentioned by two of the participants [HP3 and SMP2], though this may have been due to the fact that only two questions explicitly specified reference to instrument names, rather than track number(s). However, in chapter 8 (section 8.2), we did find significant improvements when channels were highlighted in the display, for both SM and CS mixer designs.

Providing a secondary, stable source of information in the hybrid design (dials as well as node size) did not seem to offer any improvements in speed or accuracy of the tasks, but neither did it significantly worsen them. Providing dials, faders or ordinal information (such as dB, Hz etc.) may help users gain confidence and make fine adjustments, though in future designs these may be better displayed in response to user requirements, rather than being constantly visible. The fact that searching one UI object in the CS design had very similar results to the novel designs suggests that the CS design may benefit from being simplified during visual search. Indeed, user comments suggest that visual checking of dial positions was a strategy employed by some participants. Comments also suggest that participants using the CS design found the information presentation to impact negatively on their aural acuity. Being able to show only relevant UI objects in a CS design in response to user queries, (such as channels with reverb, or channels within a set range of pan positions) may improve task completion time and accuracy, allowing better balance between visual and aural modalities. Though Pro Tools (version 10 upwards) has a 'Bus Interrogation' function (accessed through a drop down menu), in which only tracks using a specified effect are displayed, this is limited

to auxiliary effects. Filtering the CS interface may prove useful in a broader range of mixing tasks such as inserts, volume, panning and frequency modifications. Furthermore, Pro Tool's Bus Interrogation feature disallows display of multiple effects and has an undo function limited to one level, meaning that querying single effects one after the other can result in all other channels being hidden, with subsequent interruptions to mixing workflow.

Finally, the comments from the participants suggests that the mixing process involved visual referencing, as well as critical listening, in all three interfaces. This reliance on visual referencing is not confined to this study. Indeed, researchers have commented on the way that mixing engineers approach mixing visually as much as aurally (Battino and Richards, 2005). Furthermore, audio mixing guides (e.g. Computer Music, May 2014) regularly give guidance on how best to use visual metering and analysis tools to assist the mixing process, citing them as a useful way to deal with poor monitoring or room acoustics. Given this reliance on visual referencing in mixing, minimising the complexity of visual search and making the visual feedback perceptually more appropriate may benefit users. In line with this, the GUI presentation used in the novel designs, such as showing parameter amount by size or colour amounts, filtering the amount of visual data, or highlighting certain channels may be transferable to the CS mixer, potentially allowing faster visual referencing of the mix. We discuss this further in chapter 11.

10.6 Summary

In this chapter, we presented our final study in which we added audio to novel designs (hybrid and SM mixers) and compared these to a CS mixer. We designed these GUIs to test and refine our heuristics and improved ecological validity by applying them to an audio mixing context. Our designs incorporated design elements in response to findings from the previous studies. These included multivariate data objects, dynamic query access, UI objects design and GUI layout.

In section 10.1 we outlined the detail of our GUI designs, the test procedure and study design. The results of the study (section 10.3) showed that, while visual search speed and critical listening accuracy declined in all instances as the visual complexity increased, this was less pronounced when using the novel designs compared to the CS design. The participants' comments were analysed to uncover any common themes or unexpected

responses. The analysis showed that, as anticipated from the results of our previous studies, visual referencing and reliance on visual feedback to carry out the mixing tasks was a common theme across all the GUI designs.

In section 10.4 we discussed the results of the study outlined in this chapter and concluded that the novel designs had improved aural response and visual search significantly. We especially noted the use of short cut keys to quickly compare information across the mix, and concluded that this may have reduced search times and WM load, thereby contributing the improvements shown. We discussed how a reliance on visual feedback is a phenomenon experienced in this study, and beyond it, and how this aspect of mixing workflow should be addressed in mixer GUI design.

Chapter 11

Summary of Contributions, Discussion and Final Conclusions.

In this thesis, we investigated the presentation of visual mix information, focusing on how the design of the GUI, UI objects and access to them affected concurrent aural acuity. To remind the reader, we were motivated to undertake this research by concerns that current DAW designs, with increasingly complex interfaces, have the potential to make the mixing process not only an aural, but also an increasingly visual undertaking (Ratcliffe 2014). We sought to quantify the extent to which this may impact on perceptions of auditory events, auditory perception and aural acuity while mixing using a DAW, as these are fundamental to successful audio mixing workflow. It should be added, however, that we acknowledged that the visual feedback of a mix is a very valuable resource, and our research has been focussed, not on removing or dismissing it, but rather on finding ways of designing it which support and enhance the audio mixing process.

As a starting point for our research, we investigated the dominant paradigm used in commercial DAWs, namely the Channel Strip (CS) metaphor. We examined the influence of the CS design elements (faders and dials and EQs) on aural acuity, and the navigation required to search through, analyse and modify them. We then related this to human perception, focusing on Working Memory (WM), attentional limits, and dual modality task sharing between vision and audition.

Using our findings, we examined alternative designs for the User Interface (UI) objects and novel ways of accessing their information, again relating and framing any design decisions within human perceptual limits and demands. We used the findings from the studies to design and test prototype interfaces, which we compared to CS mixers to assess and quantify any potential differences. To conclude this thesis, we propose heuristics developed from these findings, summarise our contributions to the field and discuss ideas and applications for future research directions.

11.1 Research Summary

We began our research by putting the current DAW in an historical and developmental context (chapter 2). We discussed the development of audio mixing technology and how it led to the development of the CS paradigm, a design that has been widely adopted in both physical and virtual mixing desks. In the second section of the chapter, we focussed specifically on the practical implications of mixing using the CS in DAWs. We summarised that DAWs can provide essential access to mix information, but can also place a heavy emphasis on visual feedback, and we questioned the extent to which this may influence concurrent auditory attention.

In chapter 4, we began our experimental studies. We introduced our first experiment, a broad, high-level investigation into bi-modal attention while mixing. Specifically, we investigated the extent to which the amount of visual feedback may help or hinder the speed and accuracy of frequency matching of two audio sources. The study found that for single channel equalisation tasks, the use of visual representation of frequency changes (using a design commonly found in DAW EQs) does not appear to minimise the attentional resources given to the auditory response, compared to a design with a blacked-out GUI. Neither however, did the use of the visual representation improve accuracy or speed. This result suggested that visual feedback per se was not necessarily an aid or impediment to aural acuity. We concluded that lack of a definitive answer from these results may have been due to a lack of perceptual load on either the visual or auditory modalities in the experimental design. We also considered that the recruitment of proprioceptive senses, while using the mouse to undertake the equalisation task, may have had a bearing on the result.

In Chapter 5, we addressed the issues found in the previous experiment and introduced more visual load to make the experience more comparable to actual use of a DAW. To remind the reader, even on moderately small mixes, the DAW will present a large amount of visual information, surpassing the level of visual load found in the interface design described in Chapter 4. To make the experience more akin to DAW mixing (especially using small screen displays), we included scrolling navigation, increased the track count to eight channels and included moving distractors, in the form of peak meters and equaliser spectrum analysers. The results of the study showed that, again, increased visual load by itself did not have a statistically significant effect on reaction time to concurrent critical listening. However, introducing scrolling navigation to the interface design (by far the most common navigational method used in DAWs) did

significantly impede the speed and accuracy of critical listening reaction times. We suggested that, when scrolling, participants had to remember parameter settings while navigating between the visual information. We discussed that the use of visual working memory (WM) during navigation may have loaded attentional resources and thereby diminished capacity to engage with the auditory stimuli. We concluded that this result suggested a susceptibility of auditory processing to visual WM load.

Following on from our findings, in Chapter 6 we investigated whether the design of individual User Interface (UI) objects may influence WM load. We considered whether certain designs were more perceptually robust than others during navigation, and if different levels of visual perceptual demand inherent in various designs might influence WM load during navigation, and therefore also influence aural acuity. We tested standard and novel UI object designs (dials, faders, numbers and colours) and quantified their influence on visual search times, search accuracy and concurrent critical listening skills while scrolling. To summarise our findings, the results suggested that there was no significant difference between the tested designs and the amount of correctly answered critical listening questions. We concluded that the act of integrating visual information across the GUI being navigated, and holding the UI objects in visual WM was experienced equally under all conditions, with no design showing an advantage in terms of increased auditory attentional capacity. However, the analysis showed that in terms of visual tasks, participants could complete significantly more accurate channel matching using colours, faders and numbers compared to dials. We concluded that this may be due to the difficulty in interpreting quantitative information in dials (Chawla & Whitman, 2011) with small differences between dials becoming increasingly difficult to discern (Foster & Ward, 1991., Moraglia, 1989). We concluded that this in turn reduced the speed and accuracy with which the study participants could use the visual mix information represented by dials. Conversely, we found that colours performed well in this regard. However, we cautioned that both perceptual and physiological caveats needed to be considered when using colours to represent mix information. For example, we must ensure that they are sufficiently different and easily discriminable from each other. The more colours that are used, the closer in hue each colour will be, making it harder to discriminate between them (Smith and Mosier, 1986).

In chapter 7, we focused further on the impact of interface navigation on critical listening. Building on our results from chapter 5 (which showed scrolling navigation to

reduce aural acuity) we discussed GUIs which displayed overviews of the mix information, and considered whether they may be of benefit, as they remove the need to navigate the visual information. We designed three interfaces showing the panning and volume of a 24- channel mix. These comprised a traditional CS scrolling interface, a CS mixer overview and a stage metaphor (SM) overview (using the x and y positions of circles for pan and volume respectively). The SM mixer was included to address the perceptual limitations which we found when using dials, as it allowed volume and panning to be represented without recourse to dial UI objects.

The results of the study suggested that both the overview designs provided quicker visual search and improved aural acuity compared to the scrolling navigation interface. This finding was especially strong with the SM design. We concluded that reducing navigation did improve task sharing between visual and aural modalities. We also raised the possibility, that in the SM design, by integrating the UI elements into a single object (both panning and volume were represented by one UI object), we may have reduced representational holding and visual WM, which we suggested may also have contributed to the improved results. However, we found that while using the SM design, participants experienced difficulties finding channels due to their random distribution in the virtual stage. Consequently, we raised the need to explore ways to organise and filter the information to improve search using this design.

In chapter 8, we considered in further detail the potential difficulties we found with random distribution of channels in the SM design. We discussed how, within other domains such as maps and websites, there is frequent implementation of Dynamic Query (DQ) filters to allow rapid and reversible exploration of visual data. We reviewed how DQ filters have been shown to improve the usability of overviews, by organising visual data, and discussed how this may be a potentially useful application within the mixing design. However, we also raised concerns (e.g. Gelineck et al 2016) that DQ filters in the GUI may distract users from the mixing workflow, and in fact contribute to visual load, rather than minimise it. To explore these factors, we designed mixing GUIs (using both CS and SM designs) with and without DQ filters and asked participants to undertake a series of concurrent aural and visual mixing tasks whilst using them.

The results of the study suggested that the layout of the visual information did indeed have an impact on both the accuracy of critical listening tasks and the speed of visual

search, with DQ filters resulting in improvements on both counts. We suggested that this may be due to DQ filters reducing extraneous details in the GUI and minimising cognitive capture by the visual information, thereby allowing attention to be shared more efficiently between visual and aural modalities. Furthermore, our qualitative analysis showed that participants were comfortable using DQ filters, and perceived them as helping in terms of mixing workflow and visual search.

In chapter 9, we considered further the one-to-one mapping inherent in the CS design. We discussed how the SM design supported integration of dual parameters (faders and dial values) into one UI object, and how this may have improved aural acuity by reducing WM load. We discussed how within some areas of data visualisation, multiple parameters are integrated into one graphical object, by assigning data to visual variables such as position, size, shape, hue, saturation, texture, opacity and dynamics (Borgo et al 2012). We reviewed literature which suggested that such designs have been shown to reduce screen clutter, help support the interpretation of data, and enhance visual analysis by allowing both inter and intra-record relationships to be more easily detected (Ward 2008). Furthermore, while the implementation of multivariate objects is limited in audio GUIs, we found literature, such as Ferguson et al (2005), which concluded that when visually displaying musical parameters, these are optimally understood if they are represented as different attributes of *one* object, as this takes into account the limitations of human attentional bandwidth. Furthermore, we cited research by Dewey et al (2016), which has shown that the use of icon based mixers (such as SM mixers) can not only reduce cognitive load but also increase immersion in the mixing task. We therefore designed GUIs which extended the integration of mix parameters. We tested designs that integrated volume, panning, reverb and delay within one UI object and compared these with CS mixers using a one-to-one mapping. The experiment evaluated the efficacy of different designs to visually represent mix parameters by assessing how fine a range of values could be represented using different designs. The results of the studies outlined in this chapter suggested that mapping mix attributes within a single UI object resulted in improvements for visual search time and accuracy compared to a one-to-one mapping. This result confirmed our hypothesis that visually integrating parameters within UI objects helps to minimise visual search times and increased search accuracy.

In chapter 10, we extended the findings from the previous chapter and applied the use of multivariate objects to an audio mixing context. We tested two novel eight-channel mixers (using an SM design) and a traditional CS mixer, all showing the volume,

panning, treble, bass and reverb amounts of an eight-channel mix. The novel designs included DQ filters to tailor the amount and type of visual feedback on screen depending upon user requirements, and overviews of the mix data to minimise the need for navigation. For one of the mixer designs (the hybrid design) we included dials (showing treble, bass and reverb amounts) in addition to the multivariate objects. This was done to assess whether a secondary representation of channel parameters affected the mixing workflow. We undertook several mixing tasks with the novel interface designs, and compared these with a CS mixer to analyse their efficacy for audio mixing workflows. Specifically, we assessed the participants' ability to hear a Low Pass (LP) filter while undertaking visual search tasks of varying complexity.

The results of the study showed that in all three GUI designs, critical listening accuracy declined as visual search complexity increased, suggesting that complex visual search has a direct bearing on concurrent aural acuity. However, this decline was greatest when using the CS mixer, with a 62.5 % reduction in the ability to hear the LP filter when searching through three UI objects compared to one UI object. In comparison, the hybrid design resulted in a reduction of 25% in the number of participants who heard the LP filter when searching through three UI objects compared to one UI object. For the SM design, this was further reduced to 12.5%. In fact, with this design, almost as many participants successfully heard the LP filter when searching through three UI objects, as participants using the CS design did when searching one UI object. We discussed why the effect of increased visual search was less marked in the SM and hybrid designs and suggested that this may have been due to the design of the UI, which integrated multiple parameters into one object thereby reducing representational holding and simplifying the amount of UI objects to be searched. We also addressed the possibility that the modifier key functionality included in the novel interface designs may have been responsible for the improvements. To remind the reader, participants were seen to rapidly check conditions of channels (such as the amount of bass and reverb) to ascertain which channels needed to be modified. This was especially marked in the tasks involving three UI objects, where we suggested that the quick visual comparison may have helped minimise the load on visual WM and consequently reduced search times (Baldando et al, 2009). We concluded that this allowed more attentional resources to be applied to the audio content of the mix. Conversely, the intricacies of the CS mixer may have contributed to the slower and less accurate visual search. We also considered that the secondary, stable source of information provided by the dials in the hybrid design led to a slight reduction in task speed, and ability to

discern the LP filter. We believed this may have simply been due to users double checking the result with both sources of information (nodes and dials). However, the effects of Information Redundancy (e.g. Sweller, Ayres, & Kalyuga, 2011) present in the hybrid interface and the subsequent increase in UI objects may have resulted in a level of visual feedback surpassing that which can be efficiently processed in WM (Baddeley, 1998). Nevertheless, the differences were minimal, and inclusion of dials may be useful on a more subjective level, as a confidence-builder in an otherwise unfamiliar interface design. This last point was also noted in responses from participants, four of whom mentioned that having further sources of visual feedback (dials or ordinal information) would have benefited their experience of using the novel designs.

11.2 Summary of Heuristics Arising from this Research

As a result of the research described in this thesis, and through analysis of the qualitative and quantitative study results, we have developed a number of heuristics for audio mixing GUI design. Although we relate these to the current CS design to give context to our findings, they are not specific to any particular design strategy or DAW, as these may change with technological advances, increases in computer processing speed and market developments. Rather, we provide high-level heuristics that form a more general approach to the design and implementation of visual mix information. For each one, we state the heuristic in general and then discuss how we propose a novel application to DAW designs, relating our decisions to the research detailed in this thesis.

1. Allow users to filter the information. The visibility of the system should provide appropriate feedback that is relevant to the task in hand (Nielsen 2001). Users should be able to control what is visible in the interface as extraneous information, which is not immediately relevant to the current mixing task(s), should be avoided as it can compete with the relevant units of information and diminish their relative visibility.

Proposed application to DAW design: Although certain DAWs allow the user to configure the mixer layout to reduce the amount of visual information (for example the Pro Tools Mix Window Display settings, Reapers' Theme and Appearance settings), these are not immediate and require navigation through nested menus to initiate. Furthermore, they cannot be set in terms of specific mix parameters such as frequency, panning or volume etc., but rather are used to configure global displays (such as whether inserts, effects etc. are displayed). We have found that search in visually complex GUIs

(where the visual information may be relevant to the search aims) increases cognitive load and representational holding, with negative effects on concurrent critical listening (sections 5.3, 7.2 and 8.3). We propose therefore that visual information can be quickly filtered to make it germane to user requirements in terms of all key mixing parameters including pan position, individual channel position, effects, frequency bands (e.g. bass, treble etc.) and dynamic control (e.g. compression). In our studies, we found that filtering led to reduced search times and improved concurrent auditory perception, compared to non- filtered interfaces (section 8.3). Though DQ filters were used, we propose any method of querying the visual information which allows rapid access to user defined thresholds and views, and keeps the time to access the filter controls to a minimum (see heuristic 2 below). We propose that this should be done by a set value (e.g. show/hide tracks with reverb), or by a threshold amount (e.g. show/hide with reverb mix above 50%). In our studies, we found this to significantly improve search times in both SM and CS designs (section 8.3).

How this Addresses Research Questions:

2. Minimise the time needed to access mix information. When using interfaces for creative work, there is a brief window to make adjustments before ideas are lost (Tano et al 2012). The need to navigate through several windows of mix information to find a relevant control may impede the user's ability to quickly respond to the programme material (Szalva, 2009). For example, with scrolling interfaces, it is difficult to browse a large data space efficiently as the user must move back and forth between the document and the scroll bar. The same issue may also be found with drop-down menus which require the user to leave the virtual mixer and interact with the menu tabs outside of the actual work space, potentially causing significant attentional overheads (Igarashi & Hinckley 2000) and interruptions to workflow. Access to additional parameters and mix information should therefore be implemented in a way that is as immediate as possible, and should be designed to reduce the need for the user to leave the composition space. Right click functionality to access key menu items may help in this regard. An example can be found in Microsoft Word (2007 and beyond), where access to tools can be initiated with a right click within the actual document space, as well as a drop-down menu (fig 11.1).

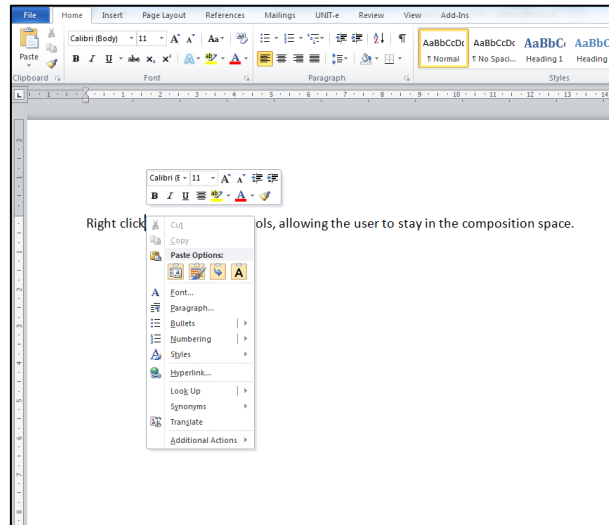


Figure 11:1. Microsoft Word (2007 and after). Users can access key functionality within the actual document space itself, thereby reducing the need to leave the composition space. <https://products.office.com/en-gb/word>.

Proposed Application to DAW design: In our studies (sections 10.1.2 and 10.4), we successfully applied shortcut modifier keys to rapidly access mix parameter information (such as delay, reverb, equalisation). Shortcut keys/ right-click editing views of the mixer space are currently limited in commercial DAWs. Though there is application of right click functionality in some DAWs, such as Logic X (figure 10.2), these are aimed at editing tasks, rather than mixing. We propose the inclusion of rapid, easily accessible commands (which may include, but are not limited to right click/ shortcut key functionality) to access views of key mixing parameters, such as effects, level, panning, and EQ. We have found that this can lead to faster and more efficient visual search with a subsequent improvement in the ability to focus on subtle changes in the audio content of the mix. Furthermore, on a subjective level they were met positively by users who welcomed their inclusion (section 10.3.4).

3. Support recognition rather than recall. Related to the heuristic above, designs should seek to minimise the user's visual WM load by making UI objects visible within the same screen, or allowing very rapid comparisons of information. The user should not have to remember information from one part of the screen to another. We have found that overloading visual WM (specifically by scrolling) can impede critical listening significantly in contrast to designs in which comparisons could be made immediately.

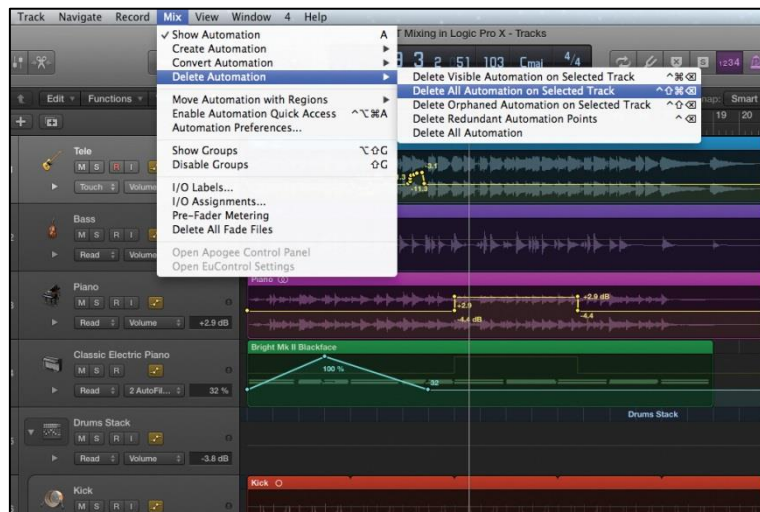


Figure 11:2. Logic X. Mixing functionality is often accessed from a menu at the top of the screen, outside of the actual mixing space. Though right click functionality is included for specific functionality (such as selecting items from the ‘toolbox’) it remains primarily for editing (cutting, pasting, truncating etc.)

Proposed Application to DAW designs. We propose that DAW mixers should support rapid comparison of parameters (effects, frequency, balance and blend etc.), so that users can quickly make comparisons between channels. We have found that this can increase the speed and efficiency of visual search and the accuracy of simultaneous critical listening tasks (sections 7.2, 8.2, 9.2, 9.4 and 10.3). In our studies we used modifier keys, DQ filters and overviews to achieve this aim (sections 7.1.2, 8.1.2, 9.1.2, 9.3.2 and 10.1.2). However, we propose that any method which allows immediate visual comparison (and which minimises the time to navigate the interface) may lead to a reduction in visual capture and visual WM load, thereby allowing attention to be shared more efficiently between vision and audition.

4. Use UI objects which are perceptually relevant to mixing. In terms of human perception, we have found literature suggesting that some of the UI objects found in the CS metaphor can have associated perceptual difficulties. For instance, dials can be difficult to interpret (Robbins 2005), especially where they are close in value, or located far apart from one another (Ozlak and Thomas 1986); fader resolution can be diminished to levels which may impede useful analysis (Htalky et al 2009) and numbers

are not effective in showing patterns, trends or exceptions among parameter values (Few 2004) especially at zoomed out levels.

Proposed Application to DAW design: We propose using alternative UI object designs to show mix parameter amounts. These may be used instead of, or in addition to the CS UI objects, depending on workflow (see heuristic 6). Specifically, we have found that the size of UI objects to represent values can successfully show differences between channels at increments of 5%, while supporting improved sharing of visual and auditory attention (section 9.4). Furthermore, we found it to be useful to exploit the widely accepted conceptual metaphor (bigger is more, smaller is less) among our participants (sections 9.2 and 10.3.4) thereby making it more appropriate to novice users unfamiliar with more technical designs. We found that saturation of colour also successfully matched our participants conception of value mappings; increased saturation being accepted as representing more of a parameter, and allowed successful comparison and modification of channel parameters amounts (e.g. reverb, frequency etc.) at 5% increments. Moreover, as it does not require the UI objects to vary in size, saturation can avoid occlusion of one channel by another, or UI objects becoming too small to be analysed or modified. Due to some disagreement among participants over mapping (section 9.2), we propose that colours be used for grouping channels (such as vocals, drums, guitars etc) or coding in terms of frequency bands. Related to this, we also propose that colour coding be user definable, as conceptions of whether a brighter colour represents more or less of a parameter value was not universally accepted in our studies and appears, as with many aspects of music, to be a largely subjective matter. Finally, we propose that transparency be used to indicate binary values, such as whether a channel is active or inactive, muted or playing, due to the perceptual difficulties in differentiating different levels of transparency, even at coarse amounts (section 9.2).

5. Support holistic understanding of the mix. The CS metaphor can lead to a segmented realisation of the mixer (Theberge 1997). Within a mix, elements interact sonically, and one channels setting can impact on another. For example, an increase in low frequencies on a bass guitar may create masking of a kick drum thereby diminishing its audibility (Dewey and Wakefield 2015). Providing a holistic overview supports rapid intra and inter record information to be more efficiently analysed and can allow the user to build a more cohesive mental image of the mix (Ratcliffe, 2014).

Proposed Application to DAW design: Related to heuristic 3, we propose that GUIs

allow overviews of the mix information. This is currently only supported in a limited way in DAWs. For example, the Pro Tools Universe Window only shows arrangement, while the Ableton effects overview can only be configured for individual channels. However, our findings suggest that CS UI objects (dials, faders, EQs) may lack the requisite detail at high over-view levels for efficient analysis (section 7.2). We therefore propose the use of visual variables such as position, colour, saturation and UI objects size to represent information in the mixer. Though these designs may not show parameter values as precisely as graduated dials or numeric information, we have found that they are better suited to showing patterns, trends or exceptions in a mixing GUI overview (sections 7.2, 8.2, 9.4 and 10.3.1-10.3.3). In our prototype design, we applied this to an SM mixer, though we suggest that this design method could be used with CS mixers to provide increased legibility at zoomed out views.

6. Allow flexibility in design of the GUI. While less technically demanding metaphors may allow non-experts quicker access to the mix parameters, there is a risk that they can also reduce the level of functionality (Pardo et al 2012) and limit precise control. Indeed, reducing the amount of detail within the interface does not necessarily present a clear-cut solution. Norman (1986) suggests that deciding on the amount of detail to include in an interface will result in an inevitable trade-off; while detailed interfaces can increase navigation time, they can be useful in making various choices and functionality explicit.

Proposed Application to DAW design: We propose that DAW GUIs should offer options to allow users to customise mappings, and access additional information. This should be implemented depending on user requirements to keep the GUI from becoming cluttered. For example, in some instances users may need precise numerical quantitative information to be displayed, such as when making fine adjustments to volume after a high quality rough mix has been achieved. In other scenarios, a broad approach may be required, such as trialling different mixes, or balancing musicians on stage for Live Sound, in which case detailed numerical feedback may be unnecessary. Comments from the participants in our studies highlighted the need for this (section 10.3.4.3), with several comments relating to the need for varying levels of information, such as dB levels, precise numerical information etc. depending on their workflow, experience and the progress in the mix. We also propose that visual channels such as colour or saturation be applied to the CS metaphor, as required by the user. In this way, for example, the mixer could show precise quantitative information, while also

displaying how much of a user selected parameter each one contains (for example, reverb, bass, delay etc.), potentially providing the ability to quickly see patterns and relations between channels, and increasing search speed.

11.3 How Research Questions are Answered.

1. To what extent does visual perceptual load interfere with or support the processing of auditory information?

Our research and prototypes have suggested a direct correlation between interface design and aural acuity. Our analyses have shown that minimising the complexity of Graphical User Interfaces (GUIs), especially where the visual feedback is related to the mixing task, decreases perceptual load, specifically visual Working Memory (WM), with a subsequent improvement in concurrent critical listening tasks (sections 5.3, 7.2 and 8.3, 10.3).

We have found that the use of perceptually relevant designs for the User Interface (UI) objects within a mixer (sections 7.2, 8.2, 9.2, 9.4 and 10.3) minimises the cognitive load required to hold information in visual WM and yields improvements in the speed and efficiency of visual search and simultaneous critical listening tasks (sections 7.1.2, 8.1.2, 9.1.2, 9.3.2 and 10.1.2). Specifically, by reducing the perceptual load required to analyse mixer UI objects (section 6.1.3, 6.1., 9.1) we found improvements in both time and accuracy when searching for and modifying channel parameters (e.g. reverb, frequency etc.) compared to a CS design, even at small increments in difference between mixer channels.

2. Does the specific design and layout of the visual information have an impact on aural acuity or speed of visual tasks?

In our studies we compared mixing interfaces using different layouts and designs for the mixer UI objects (sections 6.1.3, 7.1.2, 9.1, 9.3, 10.1). We found that UI objects which comprised non-CS elements (such as colour, size, saturation and channel x-y position) allowed faster and more accurate visual search than traditional UI elements (dials, faders EQ graphs). Our prototype mixer (10.1) showed that UI objects which utilised widely accepted conceptual metaphors (e.g. bigger is more, smaller is less) were

rated favourably among our participants (section 10.3.4) and resulted in more efficient mixing workflows compared to a CS design (sections 9.2 and 10.3.3). We also found that less perceptually demanding UI designs, such as the saturation of colour, successfully matched our participants conception of value mappings and resulted in increased accuracy when discerning and modifying small differences between channel parameters values compared to CS UI elements such as dials (sections 6.2, 9.2, 9.4).

We continued to address this question by investigating the efficacy of filtering the GUI. We found that interfaces which allowed the user to configure the design and layout of the GUI through the manipulation of Dynamic Query filters (sections 8.1, 10.1) lead to increased speed and accuracy for visual search, and improvements in discerning simultaneous small changes in the audio (sections 8.2, 10.3). Finally, when investigating overviews (using both a CS and SM design), our research suggested they allowed participants to quickly see patterns within in the mix (such as overall panning) as well as relations between channels, resulting in increased visual search speed and accuracy (sections 7.2, 8.2, 10.3).

3. How can visual mix information display be managed more effectively?

We explored this question by assessing the introduction of overviews into the GUI designs (sections 7.1., 8.1., and 9.1 and 10.1.) Our analyses suggested that using overviews (for both a Stage Metaphor and Channel Strip design), supported orientation and recall (sections 7.2, 8.2, 9.4 and 10.3.1-10.3.3) and helped avoid the negative effects of disparate data that we found when using scrolling CS designs (sections 7.2, 8.2).

We further addressed this question in chapters 8 and 10. Here we quantified the extent to which filtering the visual information in the GUI affected search times and concurrent auditory perception. Our analyses showed that using Dynamic Query filters to manage the amount of information shown on the screen, yielded improvements in speed and accuracy for both visual and aural tasks compared to non-filtered interfaces (section 8.3).

Finally, in section 10.2, we addressed this question by investigating the use key commands to manage the amount of visual feedback in the GUI. Our analysis (both qualitative and quantitative) suggested that this improved both simultaneous attention to auditory changes as well as the speed and accuracy of workflow for a range of mixing

tasks, and significantly reduced the time to find, access and modify the UI objects with the trialled mixers (section 10.3).

11.4 Contributions

In general terms, this thesis contributes knowledge concerning the human perception of, and responses to auditory information whilst undertaking visual search and analysis in a screen-based audio mixing context. This includes the following specific contributions.

1. **The provision of new quantitative information on the human perception of, and responses to aural stimuli during visual search.** This contribution bridges the gap between quantitative and theoretical/qualitative studies. For example, in chapter 10, we relate the use of DQ filters to task completion times, accuracy of information search and concurrent audio mixing tasks. This extends the subjective evaluations (e.g. Gelineck and Uhrenholt 2016) on filtering visual information in the GUI. Similarly, in chapters 7, 8 and 9, we have assessed the efficacy of SM mixers using metrics of time and accuracy as well as subjective measures, extending related qualitative evaluations (e.g. Gelineck and Korsgaard 2014, and Cartwright et al 2014).
2. **The development of heuristics for the design of mixing GUIs that acknowledge issues of human perceptual limits, working memory and cross-modal attention.** This contribution helps close the gap between general human perceptual and attentional research and domain specific application. For example, the work in this thesis addresses the use of visual load and information retrieval (e.g. Harms et al 2015., Feinberg and Murphy 2000., Piolat 1997), visual load and auditory processing (e.g. Molly et al 2015., Dehais et al 2012., MacDonald and Lavie 2011., Klemen et al 2010) and visual WM and interface navigation (e.g. Rensink 2012., Lam 2008., Mayer and Moreno 2003., Jul and Furnas 1998.) and relates them to aural and visual modality sharing during mixing. Specifically, we address the influence of cross-modal attention in terms of interface navigation (chapter 5), UI object design (chapters 6, 9 and 10) and the limitations of visual bandwidth (chapters 7 and 8).
3. **The provision of new quantitative information on the design of novel GUIs**

in relation to the existing paradigm. We have extended research which examines novel mixing GUIs in isolation from the existing software paradigm (e.g. Gelineck and Uhrenholt 2016., Law and Hou 2015., Ratcliffe 2014., Carrascal and Jorda 2011., Liebman et al 2010) and which focus on a limited range of mixing parameters (e.g. Law and Hou 2016., Dewey and Wakefield 2016). In all studies outlined in this thesis (chapters 4-10) we have used the CS metaphor mixer alongside novel designs, providing context and benchmarks and directly comparable data for novel designs compared to the CS metaphor in terms of visual search time, visual search accuracy, critical listening acuity and critical listening response time.

4. **The development of UI objects designs that acknowledge cross-modal human perceptual limits.** We have provided detailed quantitative information regarding the suitability of UI objects for audio mixing. While several authors have used novel UI objects or novel mappings of parameters (e.g. Gelineck and Uhrenholt 2016., Gelineck et al 2015., Cartwright et al 2014., Diamante 2007., Gibson 1997) these have not always been based on visualisation principles or human perceptual limits. We have extended this research therefore by assessing them in terms of cross-modal perceptual efficiency (chapters 6, 7, 9 and 10). Specifically, we have assessed mapping of visual information to size, hue, saturation, transparency, and position. We have provided data on the efficacy of each and compared and quantified the differences compared to CS UI objects in terms of simultaneous information search and critical listening (chapters 6, 7 and 9) and simultaneous visual search, parameter adjustment and mixing workflow (chapter 10).

11.5 Directions for Future Research

While our prototype designs (see chapter 10) assessed an eight-channel mix, we are aware that many mixes exceed this. Future research would need to address management of increased channel count. For example, when using the SM design, the UI objects used to represent channels may begin to obscure one another when the amount of tracks increases. While, as we have discussed, dynamic filtering may ameliorate this problem, further solutions may be required. Looking beyond the field of audio mixing, potential solutions may be found in the fields of Geographic Information Systems (GIS). Here too, designers are confronted by two dimensional (2D) representations of

data which can contain a very high density of visual information. Multiple data points need to be represented without compromising the legibility of the GUI. For example, the Crime map of London (fig 11.3) displays incidence of crimes as data point symbols on a 2D space in a similar way to channels presented as data points on the virtual stage in our prototype design. The information is represented at a variety of scales which can be accessed using a zoom wheel, or modifier keys. Users can start with an overview of an area's crime, which displays a single data point (circle containing numerical information) showing the combined amount of recorded crime incidents within the general area. Users can then click on the circle to explore the data in more detail, in the process zooming in and creating more specific data points. This increase in data circles per scale change allows the user to analyse subsets of data as individual information layers. In this way data never obscures the interface and its display is managed in relation to the zoom level chosen, a design methodology which has been broadly welcomed by users during its evaluation (Roth et al 2014).

A similar design can be seen in property search websites. Again, these use icons on 2D maps to show data. For example, the Zoopla website (fig 11.4) displays properties as icons superimposed over a map. Whenever there is a more than one property within a small area (such as a single street), rather than displaying multiple icons, which may obscure one another and the interface, the design allows users to reveal further data through right clicking, which reveals a list of properties within the area (figure 11.4, right). Again, this design allows a high-level view of data points in the map, whilst supporting analysis of densely populated areas. Furthermore, unlike the Crime Map, this design does not require zooming, or any navigation to access the subsets of data. If the user does zoom, the integrated icons are collapsed out, depending on the zoom level chosen. However, in this design it is not evident that an icon contains more than one data point, and perhaps a coarse coding, such as colour, could be used to differentiate single from multiple data icons.

Relating the design of these interactive maps to our heuristics (section 11.3), they may offer suitable design solutions. For instance, they support rapid and persistent filtering and remove the need for multiple or nested menus. However, implementation of such functionality in the SM designs would need to be undertaken carefully. More research is needed on how zooming may impact on concurrent auditory processing, a concern not investigated by designers of interactive maps. Specifically, future research would need to address whether zooming navigation impacts on critical listening in a similar



Figure 11:3. Interactive map of crime in London. Users can explore the data at a variety of different levels. As zoom level decreases, areas are merged into fewer icons to maintain the legibility of the GUI. <https://www.police.uk/city-of-london/cp/crime/> [Accessed May 25th, 2017].

way to scrolling. For example, Cockburn and Savage (2003) found that abrupt transitions between discrete zooming levels caused significant disorientation as participants reoriented themselves and assimilated the new view with each zoom change. Finding ways to minimise any disorientation and assessing its influence on critical listening would need to be carefully considered. Such research has not been dealt with in this thesis, as within commercial DAWs, scrolling is the dominant navigation method. Furthermore, successful implementation of multiple data layers in an SM design would involve multiple evaluation-and-revision stages to address use cases (Roth et al 2014), an undertaking that would have been beyond the scope of this thesis.

Finally, further work is needed to apply our heuristics to the CS metaphor. While our studies were focused on creating general design heuristics for novel mixing UI objects and GUIs, we believe our findings have potential value, not only in the creation of a novel design such as the SM metaphor, but also in application to the existing CS mixer.

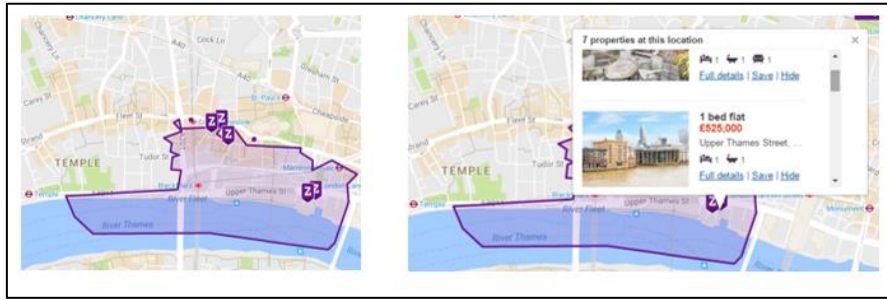


Figure 11:4. Zoopla website. Users can explore the data at a high zoom level. Right click functionality allows users to access subsets of data as an overview, without the need to zoom. <https://www.zoopla.co.uk/> [Accessed May 25th 2017].

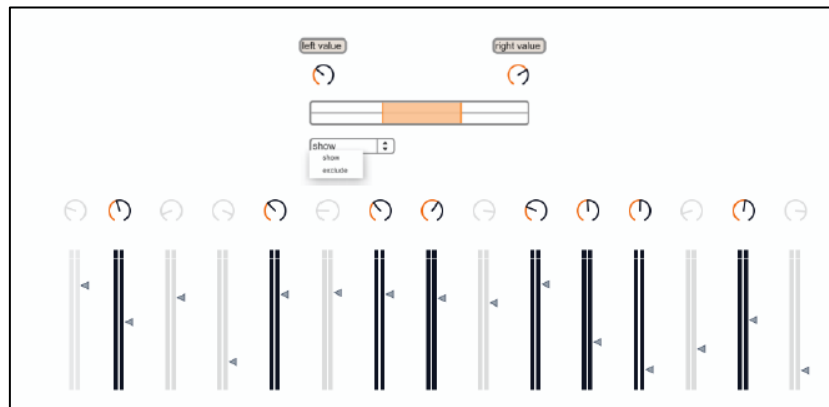


Figure 11:5. Authors' mock-up of a CS GUI which uses DQ filters to show or hide tracks (make transparent) depending on user requirements. In this figure, only channels panned within a specified range of values are shown.

For example, the CS UI objects (dials, faders, EQs) may lack the requisite detail at high over-view levels for efficient analysis. However, visual variables such as position, colour, saturation and UI objects size to represent information in the mixer could be combined with CS mixers to provide increased legibility at zoomed out views. For example, relative levels of colour saturation could be applied to the channels, as required by the user, allowing both display of precise quantitative information as well as coarser indications of parameter amounts. This may improve the virtual CS mixers ability to provide show patterns, and relations between channels, and identify specific channels while scrolling.

Likewise, while the use of DQ filters was applied to the SM mixer in our studies, we believe they may have potential in the CS metaphor. For example, hiding and showing

channel strips, or making them transparent in response to DQ filters (figure 11.5). This may make the CS mixer better suited to perceptual limits that dictate that only a few items are attended to at any one time (Rensink, 2012), and once again, it may improve visual search by reducing non-essential visual information. While this has not been covered in the present work, further trialling and assessment of these modifications to the CS mixer metaphor may prove useful in allowing the design to more accurately reflect the growing track and effects counts of contemporary audio mixes, the limited screen space of laptop and tablet computers and the human perceptual limits of the user.

11.6 Application of research to other aspects of audio mixing.

We feel our findings and heuristics may be applicable to areas beyond those detailed in the studies. These include live mixing, mixing across multiple genres, surround sound and multi-channel mixing, and revealing errors in mixes such as frequency masking. Although an in-depth investigation of each of these is beyond the scope of this thesis, we briefly discuss below how they may benefit from our research findings and how modification to GUIs based on our research may be applicable.

11.6.1 Live Sound mixing;

In recent years, the use of analogue mixing consoles for live performance applications have, to a large extent, been supplanted by digital mixers (Yeakel and Vallier 2002). With the majority of these, the user interfaces require the mix engineer to page through multiple layers of on-screen menus to locate the desired feature on the mixer. For example, the sound engineer at a live performance venue may notice problematic areas with the mix coming from the stage (e.g. the vocal is sounding sibilant). Using a typical digital mixer (such as the Yamaha 02R system) the sound engineer has to understand and recall which sub-mix the vocal mix is on, find and access the menu for the high frequency EQ (from a number of sub menus on that channel), and possibly access another layer or bank to find and alter the bandwidth of the vocal EQ so that the equalisation only attenuates the sibilant frequencies. This experience can lead to a slower and more frustrating workflow than operating a traditional analogue mixer where

controls are mapped physically on the mixer in a 1-2-1 configuration (Yeakel and Vallier 2002).

We believe that the interface designs and heuristics detailed in this thesis may help ameliorate some of the problems with such a segmented realisation of digital mixers. For example, providing overviews of a mix, using a stage metaphor design, may support rapid intra and inter record information to be more efficiently analysed, allowing the user to build a more cohesive mental image of the mix (Ratcliffe, 2014) and reducing the amount of interface navigation to access mix parameters (menus, window switching, scrolling etc) compared to channel strip or digital mixer designs. In our studies, we found that reducing the complexity involved in searching, accessing and manipulating the GUI lead to improved mixing workflow in terms of interface manipulation and concurrent auditory perception (section 8.3, 10.3); both essential attributes for live sound engineers, who must use mixing desks to rapidly respond to and fix problems coming from the stage in front of them.

11.6.2 multi-channel and surround mixing;

While the studies detailed herein have been solely focussed on stereo mixing, there may be transferable benefits to surround sound and multiple channel mixing (this most commonly consists of five speakers positioned around the listener, two behind, and three in front, with an additional low frequency speaker to provide extended bass response). While surround sound mixing shares many of the same fundamental requirements as a stereo mix in terms of effects, equalisation and dynamic control (Rumsey 2001), in terms of panning, surround sound mixing becomes more complex. In stereo mixing the audio can be panned to either the left or the right. Surround mixing however adds two new elements to the technique of panning. While audio can still be panned left to right, this is split into two dimensions: front left to right panning and rear left to right panning. In addition, the mix engineer has control over the panning between the front stereo speakers and rear stereo speakers.

While the surround sound panning in current DAWs such as Logic and Pro Tools offers a clear visualisation of the track's relative position between the speakers (figure 11.6), this is only on a track by track basis and is not carried through to the mix as a whole, compounding the problems we found with a stereo mix, where the user has to

‘scrutinise each channel’s pan knob position’ to assemble a mental image of the stereo positioning (Dewey and Wakefield 2016, p.2). We believe a stage metaphor design may be applicable to surround mixing, to clarify the position of all the channels within the surround sound space. Furthermore, visualisations of the all the mix channels within the 360° sound stage may allow users to see patterns within the mix, which may be of benefit in avoiding common surround sound mix errors such as crosstalk between pairs covering the same mix areas, such as Left-Centre, Centre-Right etc, (Theile 2000) and making explicit any areas of the sound stage where the phantom images may be unstable (Rumsey 2001) potentially helping to avoid holes in the soundstage (ibid) and improving the overall quality of the final mix.

11.6.2 Frequency masking;

Frequency masking is a phenomenon in which the perceived audibility of one sound is affected by the presence of another sound which has a similar spectral content (Dewey and Wakefield 2010). This is of concern in the mixing of multi-track recordings where many tracks may have similar spectral content and appear to compete for the listener’s attention, resulting in a lack of clarity and an undefined mix (ibid). Examples of frequency masking can be seen in the frequency relationship between the kick drum and bass or between vocals and lead guitar (Owsinski 2006).

We have discussed previously that it can be difficult to mentally integrate all the dials on a mixing desk, especially where scrolling is required to see them all (section 5.2). This can make it difficult to build up a mental image of the overall frequency content of a mix and see how various channels are interacting in terms of equalisation, as the EQ information is visually disparate within the GUI. Furthermore, as many home-studio speaker systems are not full range (they do not play back frequencies as low as 20Hz), it may become necessary to rely on visual feedback of frequency content, as the lowest frequencies of a mix may be inaudible through the speakers.

We propose that by filtering the interface, it may be possible to make the relative EQ content between channels more explicit. For example, using Dynamic Query filters, the user could specify that only tracks which have frequency content between a specified frequency range (using a Hz filter) and above a specified volume level (using a dB filter) should be shown. In this way, only channels matching the search criteria would be

visible. This method may be suited to both channel strip and stage metaphor mixers.

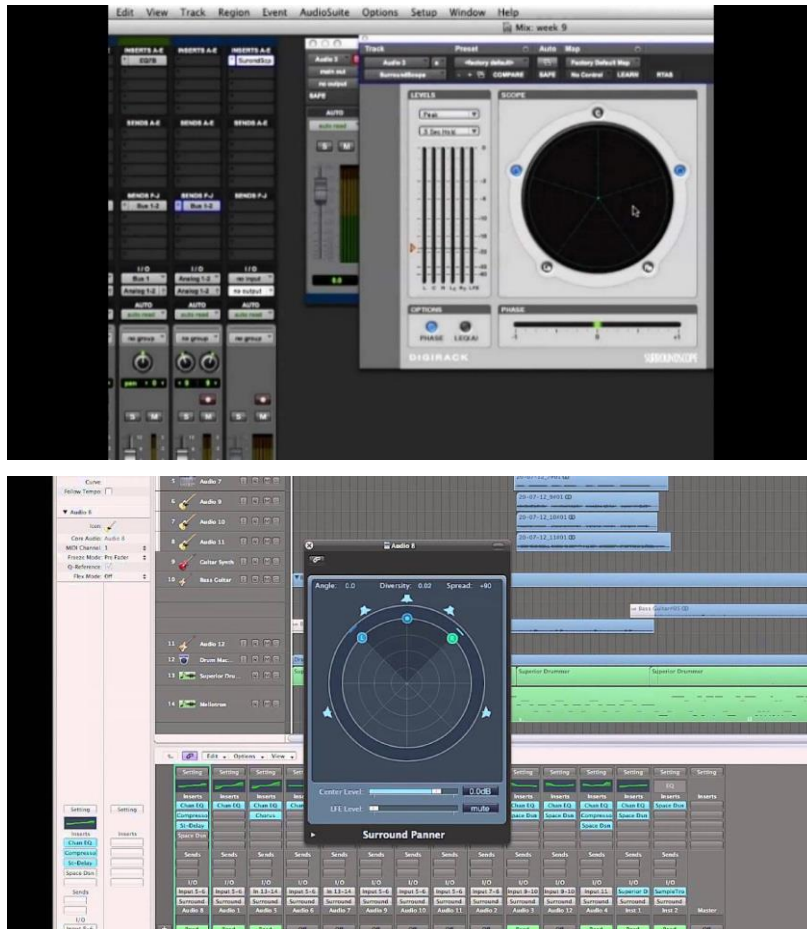


Figure 11:6. The surround sound panning in Pro Tools (top) and Logic Pro (bottom).

While the GUI shows the pan position of individual channels, this is not supported for the whole mix, making it potentially difficult to build a clear mental image of the relative position of all the mix channels.

The frequency content of channels could also be made explicit in a stage metaphor by utilising the shape of the channel. By dividing the circle into segments, each one representing a frequency band such as lows, low-mids, mids and highs, the relative size of the bands segment could give a quick visual indication of each channels frequency balance (fig. 11.7), making relationships of frequency content between channels easier to detect (fig 11.8). Finally, we believe that by using perceptually robust indicators of differing frequencies (such as colour or channel size in an SM mixer) users could gain a holistic overview of the relative frequency content of all channels. For example, dividing channels into colours based on which frequency band has the most energy, or allowing

users to visualise frequency bands in terms of channel size may be applicable. Employing these methods, the user could then go on to fix the offending channels using mixing techniques to reduce frequency masking such as panning, level setting, filtering and muting (Owsinski 2006, Dewey and Wakefield 2010).

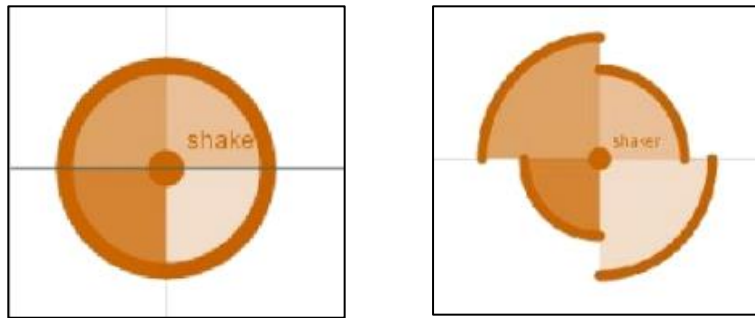


Figure 11:7. Dividing the circle representing a channel (in this case a shaker) into frequency segments. Each segment can be assigned a frequency range, in this case (anti-clockwise) lows, low-mids, mids and highs. In the picture on the right, the low and mid frequencies are reduced, while the low-mids and highs are boosted.



Figure 11:8. Frequency content of four channels in a mix. The relative size of each segment can give a visual indication of the gain of the frequency band, while the width of the segment indicates the bandwidth (which range of frequencies are contained within each band). This allows a quick visual overview of relative frequency profiles within and between the channels.

11.6.3 Dealing with different genres.

While most of the mixes within our studies have been of a Rock or Electronic Dance Music genre, we believe that the findings from the studies are applicable to a wider range of musical genres. As Owsinski says; every modern genre of music, including Rock, Pop, R&B, Rap, Country, New Age, Swing, Drum and Bass etc., contain the same common elements essential to a successful mix (Owsinski 2006, p.10). These include balance (volume levels), frequency range (equalisation), panning, depth (reverb), and dynamics (ibid, p.10). While the amount of mixing required within each of these elements may vary between genres (for instance Hip-hop, is often mixed with a narrower use of panning than Rock, and Reggae may require more use of low frequency equalisation than Jazz), these elements form commonalities between all styles of music. In this regard, the visualisation developed through the course of the studies described in the thesis, and the heuristics developed from them, may be relevant to a wide range of users working in a multitude of musical styles.

Bibliography

- Aikin, J. (1997) 'Keyboard Reports: Steinberg Cubase VST'. In: *Keyboard*. January, pp. 85–90.
- Ahlberg, C. and Shneiderman, B. (1994). Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In: *Proceedings of the Conference on Human Factors in Computing Systems*, Boston, MA, April. New York, ACM Press. pp. 313-317.
- Alais, D, Morrone, C, and Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society B-Biological Sciences*, 273 (1592), pp. 1339–1345.
- Alho, K., Woods, D.L., Algazi, A., Naatanen, R. (1992). Intermodal selective attention: effects of attentional load on processing auditory and visual stimuli in central space. *Electroencephalography & Clinical Neurophysiology* 82, pp.356–368.
- Arrighi, R., Lunardi, R., Burr, D., 2011. Vision and audition do not share attentional resources in sustained tasks. *Frontiers in Psychology* 2, p.56.
- Baddeley, A. (2003) Working memory and language: An overview. *Journal of Communication Disorders* 36, pp.189–208.
- Baldano, M.Q.W., Woodruff, A. and Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization. In: *Proceedings of Audio Visual Interfaces, Palermo, Italy*. New York, ACM Press. pp. 110–119.
- Balin, W and Loviscach, J. (2011) Gestures to Operate DAW Software In: *Proceedings of Audio Engineering Society* 130th Convention, London, UK, 13 – 16 May.
- Battino, D. & Richards, K. (2005). *The Art of Digital Music*. San Francisco, Backbeat Books,
- Bederson, B. (2004). Interfaces for Staying in the Flow. *Ubiquity Magazine*. Volume 2.
- Begault, D., Christopher, B., Zeamer, C., Anderson, M., Burns, K. (2016). *A Touchpad-*

- Based Method for Inducing Attentional Tunnelling*. NASA. Report number 219208.
- Bell, A., Hein, E. and Ratcliffe, J. (2015). Beyond Skeuomorphism: The Evolution of Music Production Software User Interface Metaphors. *Journal of the Art of Record Production*, 09.
- Beaugard, R. and Corriveau, P (2007). User Experience Quality: A Conceptual Framework for Goal Setting and Measurement. In: Duffy, V.G. (ed). *Digital Human Modelling*. Heidelberg, Germany. Springer, pp. 325-332.
- Bertelsen, O., Breinbjerg, M. and Pold, S. (2008). Tools-Material, Metaphor-Metonymy, Instrument (ness). In: *Proceedings of the Eighth Danish Human-Computer Interaction Research Symposium, November 2008, Aalborg, Denmark*. Denmark, Aalborg University, pp 5-9.
- Berti, S. & Schroger, E. (2003). Working memory controls involuntary attention switching: Evidence from an auditory distraction paradigm. *European Journal of Neuroscience*, 17, pp. 1119 –1122.
- Borgo, R., Kercher, J., Chung, D.H., Maguire, E., Laramee, R.S, Hauser H., Ward, M. Chen, M. (2012). “Glyph-based visualization: Foundations, design guidelines, techniques and applications”. In: *Eurographics State of the Art Reports*, pp. 39–63.
- Brown, J. (2009) *Rick Rubin: In the Studio*. Toronto, ECW Press.
- Card, S., Mackinlay, J. & Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. San Francisco, CA, Morgan Kaufmann Publishers.
- Carrascal, J.P. & Jorda, S. (2011). Multitouch Interface for Audio Mixing. In: *Proceedings of the International Conference on New Interfaces for Musical Expression, 30 May - 1 June 2011, Oslo, Norway*. University of Oslo, Norwegian Academy of Music.
- Cartwright, M, Pardo, Band. Reiss. J (2014). Mixploration: Rethinking the audio mixer interface. In: *Proceedings of the Conference on Intelligent User Interfaces, February 24-27, Haifa, Israel*. New York, ACM Press, pp. 365–370,
- Case, A. (2007). *Sound FX: Unlocking the Creative Potential of Recording Studio Effects*. Oxford, UK, Elsevier.

- Cassell, C and Symon, G (2004). *Essential Guide to Qualitative Methods in Organizational Research*. London, Sage Publications.
- Chandler, P. & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology* 62, pp. 233–246.
- Channan, M. (1995). *Repeated Takes: A Short History of Recording and Its Effects on Music*. London, Verso.
- Chawla. & Whitman, L.(2011). Finding the Best Display Type for Your Data. In: *Proceedings of the SAS global Forum, Las Vegas, Nevada, April4-7*. pp 1-7.
- Cleveland, W.S (1993). *Visualizing Data*. Hobart Press.
- Cockburn, A., & Gutwin, C. (2009). A predictive model of human performance with scrolling and hierarchical lists. *Human-Computer Interaction*, 24(3), 273-314.
- Cockburn, A., Karlson, A. and Bederson, B. (2009). A Review of Overview +Detail, Zooming, and Focus + Context interfaces. *ACM Computer Survey*., 41 (1):1.
- Cockburn, A. & Savage, J. (2003). Comparing Speed-Dependent Automatic Zooming with Traditional Scroll, Pan and Zoom Methods. In: *Proceedings of the 2003 British Computer Society Conference on Human-Computer Interaction*. Bath, England. pp.87-102.
- Cogan, J. (2003) 'Bill Putnam'. In: *Mix* October. Available at: http://mixonline.com/recording/interviews/audio_bill_putnam [Accessed: 7th November 2016].
- Colavita, F. (1974). Human sensory dominance. *Perception and Psychophysics* 16, pp. 409–4122.
- Convertino, G., Chen, J., Yost, B., Ryu, Y.-S., North, C (2003) Exploring context switching and cognition in dual-view coordinated visualizations. In: *Proceedings of the International Conference on Coordinated and Multiple Views in Exploratory Visualization, Los Alamitos, CA, USA*. pp. 57- 66.

- Crane, L. (2010). This is your Brain Creating and Recording Music. In: *Tape Op*, No.74, p.12.
- Crystal, A. & Ellington, B. (2004). Task analysis and human-computer interaction: approaches, techniques, and levels of analysis. In: *Proceedings of the Tenth Americas Conference on Information Systems, New York, New York, August 2004*. pp 1-9.
- Csikszentmihlyi, M. (1991). *Flow: The Psychology of Optimal Experience*. London, UK, Harper Collins.
- Cunningham, M. (1998). *Good vibrations: a history of record production*. London, U.K., Sanctuary Books.
- Dehais, F., Causse, M., Regis, N., Menant, E., Labedan, P., Vachon, F., Tremblay, S (2012). Missing Critical Auditory Alarms in Aeronautics: Evidence for Inattentive Deafness? In: *Proceedings of the Human Factors and Ergonomics Society*, Boston, MA, October 22-26. New York, Sage Publications. pp. 1639-1643.
- Dewey, C. and Wakefield, J. (2016). Audio Interfaces Should be Designed Based on Data Visualisation First Principles. In: *Proceedings of the 2nd AES Workshop on Intelligent Music Production*, London, UK.
- Dewey, C., and Wakefield, J. (2015). An Investigation into the Efficacy of Methods Commonly Employed by Mix Engineers to Reduce Frequency Masking in the Mixing of Multitrack Musical Recordings. In: *Proceedings of the Audio Engineering Society 138th Convention*.
- Dewey, C. and Wakefield, J. (2014). A guide to the design and evaluation of new user interfaces for the audio industry. In: *Proceedings of the Audio Engineering Society 136th European Convention*, Berlin, Germany, April 26-29. pp. 1-10
- Diamante, V. (2007). Awol: Control surfaces and visualization for surround creation. Technical report, University of Southern California, Interactive Media Division.
- Dockwray, R. and Moore, A. (2010). Configuring the sound-box. *Popular Music* 29 (2),

pp. 181-197.

Dubey, S. K. and Rana, A. (2010). Assessment of Usability Metrics for Object-Oriented Software Systems, *ACM SIGSOFT*, vol. 35, no. 6, pp. 1-4.

Duignan, M., Noble, J. and Biddle, R. (2010). Abstraction and Activity in Computer-Mediated Music Production. *Computer Music Journal*, Volume 34, Number 4, winter 2010, pp. 22-33.

Duignan, M., Noble, J., Biddle, R. (2004). Metaphors for Electronic Music Production in Reason and Live. In: *Proceedings of the 6th Asia-Pacific Conference on Computer-Human Interaction. Rotorua, New Zealand, June 29-July 2*. Berlin, Springer. pp.111-120.

Dzulkifli, M. and Mustafar, M. (2013). The Influence of Colour on Memory Performance: A Review. *Malaysian Journal of Medical Science*. 20(2), pp.3-9.

Everest, F.A. (1998). *Critical listening skills for audio professionals*. Boston, Thomson Course Technology.

Feinberg S. and Murphy, M. (2000). Applying cognitive load theory to the design of web-based instruction. In: *Proceedings of the IEEE professional communication society international professional communication conference*. Cambridge, MA. pp. 353-60.

Few, S (2007) Save the Pies for Dessert. Available at: http://www.perceptualedge.com/articles/visual_business_intelligence/save_the_pies_for_dessert.pdf. [Accessed May,4 2014].

Few, S. (2004) Show Me the Numbers: Designing Tables and Graphs to Enlighten. Oakland, USA: Analytics Press.

Field, A & Hole, G (2013). *How to report and design experiments*. London, Sage Publications.

Filipetti, Frank. (2016). Frank Filipetti Masterclass on Pop Mixing and Music Production [Online Video]. Available from: <https://youtube.com/watch?v=-aSSX0fIHr0>. [Accessed: 23 May 2018].

- Flick, U. (2009). *An Introduction to Qualitative Research*. London, Sage Publications.
- Foster, D. H., and Ward, P. A. (1991). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. In: *Proceedings of the Royal Society London: Series B*, 243, 75-81
- Gahegan, M., 1999. Four barriers to the development of effective exploratory visualization tools for the geosciences. *International Journal of Geographic Information Science* 13 (4), 289-309.
- Galitz, W.O. (1997). *Essential guide to user interface design: An introduction to GUI design principles and techniques*. New York, Wiley.
- Gelineck and Uhrenholt (2016). Towards Improving Overview and Metering through Visualisation and Dynamic Query Filters for User Interfaces Implementing the Stage Metaphor for Music Mixing. In: *Proceedings of the AES*, Los Angeles, CA.
- Gelineck, S. and Overholt, D. (2015). Haptic and Visual feedback in 3D Audio Mixing Interfaces. In: *Proceedings of the 9th Audio Mostly Conference*, Glasgow, UK. ACM, pp.14:1–14:6
- Gelineck et al (2015). Stage vs Channel Strip Metaphor-Comparing Performance when Adjusting Volume and Panning of a Single Channel in a Stereo Mix. In: *Proceedings of the International Conference on New Interfaces for Musical Expression*, Baton Rouges, LA, May 31-June 3. pp. 343-346.
- Gelineck, S. and Korsgaard, D (2014) Stage Metaphor Mixing on a Multi-Touch Tablet Device. *Audio Engineering Society Convention 137, October 9-12 Los Angeles*.
- Gelineck, S., Overholt, D., Buchert, M. and Andersen, J. (2013a) “Towards an Interface for Music Mixing based on Smart Tangibles and Multitouch.” in *Proceedings of the International Conference on New Interfaces for Musical Expression, May 27-30, Seoul, South Korea*. pp. 180–185
- Gelineck, S., Overholt, D., Buchert, M. and Andersen, J. (2013b) Towards a more Flexible and Creative Music Mixing Interface. In: *Proceedings of ACM SIGCHI Conference*

- on *Human Factors in Computing Systems, April 27 - May 02, Paris, France*. ACM, pp.733-738
- Gibson, D (1997). *The Art of Mixing*. Artistpro Publishing
- Ghirardelli, T and Scharine, A. (2009) Auditory-Visual Interactions in Helmet Mounted Displays; Sensation, Perception and Cognition Issues. *Auditory-Visual Interactions*. US Army Aeromedical Research Lab, pages 599–618
- Golkhe, K., Hlatky, M., Heise, S., Black, D. and Loviscach, J. (2010). Track Displays in DAW Software: Beyond Waveform Views. In: *Proceedings of the Audio Engineering Society, London, 2010*. Paper No. 8145.
- Gudur, R., Blackler, A., Popovic, V. and Mahar, D. P. (2009). Redundancy in interface design and its impact on intuitive use of a product in older users. In: *International Association of Societies of Design Research Conference, Seoul, Korea*, p. 209.
- Gutwin, C. and Fedak, C. (2004). Interacting with Big Interfaces on Small Screens: A Comparison of Fisheye, Zoom, and Panning Techniques. In: *GI '04: Proceedings of the 2004 conference on Graphics Interface, Ontario, Canada*. Canadian Human-Computer Communications, pp.145-152.
- Harms, J., Kratky, M., Wimmer, C., Kappel, K. and Grechenig, T. (2015). Navigation in Long Forms on Smartphones: Scrolling Worse than Tabs, Menus, and Collapsible Fieldsets. In: *Interact 2015*. Springer. pp 333-340
- Harrower, M. and Sheesley, B. (2005) 'Designing Better Map Interfaces: A Framework for Panning and Zooming', *Transactions in GIS*, 9(2), pp. 77–89.
- Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock P. A., and Meshkati, N. (eds), *Human Mental Workload*. Elsevier Science, pp. 139-183.
- Heer, J. and Shneiderman, B. (2012). Interactive Dynamics for Visual Analysis. *ACM Queue*, 10 (2), pp. 30–55.

- Hlatky, M., Gohlke, K., Black, D. and Loviscach, J (2009). Enhanced Control of On-Screen Faders with a Computer Mouse. In: *Proceedings of the Audio Engineering Society, Munich, Germany*.
- Hodgson, J. (2010). *Understanding Records: A Field Guide to Recording Practice*. New York, Continuum Books.
- Homer, M. (2009). Beyond the Studio: The impact of Home Recording Technologies on Music Creation and Consumption. Available at: <http://www.nobleworld.biz/images/Homer.pdf>. [Accessed 2/10/2011].
- Hollender, N., Hofmann, C., Deneke, M., and Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behaviour* 26(6), pp.1278–1288.
- Igarashi, T., & Hinckley, K. (2000). Speed-dependent automatic zooming for browsing large documents. In: *Proceedings of the 13th Annual Symposium on User Interface Software and Technology, San Diego, CA, November 5-8*, pp.139-148.
- Janata, P., Tillmann, B., Bharucha, J. (2002): Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive Affective Behaviour Neuroscience* 2. Springer, pp.121–140.
- Janney, E. (2012). *Review: Auria DAW for iPad*. Available at: <http://www.sonicscoop.com>. [Accessed:2/10/2013].
- Jeng, J (2005). Usability Assessment of Academic Digital Libraries: Effectiveness, Efficiency, Satisfaction, and Learnability. *LIBRI* 55. pp. 96–121.
- Johnson, A., & Proctor, R. (2004). *Attention; Theory and Practice*. London, Sage Publications.
- Jul, S., & Furnas, G. W. (1998). Critical Zones in Desert Fog: Aids to Multiscale Navigation. In: *Proceedings of User Interface and Software Technology, San Francisco, California*. ACM Press, pp. 97-106.

- Klemen, J., Buchel, C., Buhler, M., Menz, M.M. & Rose, M. (2010). Auditory working memory load impairs visual ventral stream processing: toward a unified model of attentional load. *Journal of Cognitive Neuroscience* 22, pp.437–446.
- Lam, H. (2008). A Framework of Interaction Costs in Information Visualization. *IEEE Transactions on Visualisation and Computer Graphics*, 14 (6), pp. 1149- 1156.
- Larrson, L. (2014). *Perceived Difference in Equalizing Workflow Between a Control Surface and a Mouse and Keyboard*. Bachelor Degree Thesis, Lulea University of Technology, Sweden.
- Larsen, A., McIlhagga, W., Baert, J., & Bundesen, C. (2003) Seeing or hearing? Perceptual independence, modality confusions, and crossmodal congruity effects with focused and divided attention. *Perception & Psychophysics* 65, pp. 568–574.
- Lavie, N. (2005). Distracted and confused? Selective attention under load. *TRENDS in Cognitive Sciences* 9(2), pp.75–82
- Lavie, N. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics* 56, pp. 183–197.
- Law, V and Hou, H. (2015). Audio Collage as an Instrument of Musical Expression: Combining Freehand and Tangible Controllers. In: *ICMC*, Sept 25-Oct 1, Texas. pp. 106-109.
- Leyshon, A. (2009). The software slump? Digital music, the democratisation of technology, and the decline of the recording studio sector within the musical economy. *Environment and Planning* 41, pp.1309–1331.
- Lech, M., & Kostek, B. (2013). Testing a novel gesture-based mixing interface. *journal of the audio engineering society*, 61(5), pp. 301-313.
- Li, Q., & North, C. (2003). Empirical comparison of dynamic queries sliders and brushing histograms. In: *Proceedings of the IEEE Symposium on Information Visualization*. IEEE Computer Society, pp. 147–153.

- Liebman, N., Nagara, M., Spiewla, J., and Zolkosky, E. (2010). Cuebert: A new mixing board concept for musical theatre. In: *Proceedings of the International Conference on New Interfaces for Musical Expression*, Sydney, Australia, June 15-18th. pp. 51-56.
- Lopes, P., Ferreira, A. and Madeiras-Pereira, J.A. (2011) Battle of the DJs: an HCI perspective of Traditional, Virtual, Hybrid and Multitouch DJing. In: *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, May 30 - June 1.
- Macdonald, J. S. P. and Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, perception & psychophysics* 73, pp. 1780-1789.
- MacKay, D. G. and Ahmetzanov, M. V. (2005). Motion, memory, and attention in the taboo stroop paradigm. *Psychological Science* 16 (1) pp.25–32.
- Marozeau J., Innes-Brown H., Grayden, D.B., Burkitt, A.N. and Blamey P.J. (2010) The effect of visual cues on auditory stream segregation in musicians and non-musicians. *PLoS One* 5: e11297.
- Massey, H. (2007). *Here, There and Everywhere: My Life Recording the Music of the Beatles*. New York, Gotham Books,
- Massey, H (2009). *Behind the Glass Volume 2*. San Francisco, BackBeat Books.
- Massy, S. (2010) ‘Gear Stories with Sylvia Massy: The Age of Customs’. In: *Mix* [Online] July. Available at: http://mixonline.com/recording/gear_stories/gear_stories_custom_consoles/ [Accessed: 5/11/2016].
- Mayer, R. E. and Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38 pp. 43-52.
- McGrenere, J. and Ho, W. (2000). Affordances: Clarifying and evolving a concept. In: *Proceedings of the Graphics Interface*, Quebec, May 15-17. Canadian Human-Computer Communications Society, pp. 179 – 186.

- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, pp.746-748.
- Mecklenburg, S., and Loviscach, J. (2006) Subject: controlling an equalizer through subjective terms. In: *Proceedings of Computer–Human Interaction EA '06 Extended Abstracts on Human Factors in Computing*, Montreal, Quebec, Canada. pp. 1109–1114
- Mihnkern, K. (1997). Visual interaction design: Beyond the interface metaphor. *SIGCHI Bulletin*, 29 (2), pp. 11-15.
- Milner, G. (2009). *Perfecting Sound Forever: An Aural History of Recorded Music*. New York, Faber and Faber.
- Mistry, P. and Agrawal, G. (2004). Functional Metaphoric Approach to be “in the flow” with computer interfaces. In: *Proceedings of the Indian Human Computer Interaction*, Bangalore, India. Retrieved August 11, 2012 from: <http://web.media.mit.edu/~pranav/research/stayingintheflow/functional-metaphoric-approach.pdf>.
- Molloy, K., Griffiths, T.D., Chait, M. and Lavie, N. (2015). Inattentional Deafness: Visual Load Leads to Time-Specific Suppression of Auditory Evoked Responses. *Journal of Neuroscience*. 9;35(49), pp.16046-54.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences* 7, pp. 134 –140.
- Moore, A. and Dockwray, R. (2010). “The Establishment of the Virtual Performance Space in Rock.” *Twentieth-Century Music* 5 (2): pp. 219–41.
- Moorefield, V. (2005). *The Musician as Producer*. Cambridge, MA, MIT Press.
- Moylan, W. (2017). ‘How to Listen, What to Hear’ in Hepworth-Sawer, R., and Hodgson, J. (eds.) *Mixing Music, Perspectives on Music Production*. New York, Focal Press, pp. 24-53.
- Moylan, W. (2007). *Understanding and Crafting the Mix: The Art of Recording*, 2nd Edition. Oxford, Focal Press.

- Moylan, W. (1992). *The Art of Recording: The Creative Resources of Music Production and Audio*. New York. Van Nostrand Reinhold.
- Nakagawa, Y., Okuno, H. and Kitano, H. (1999). Using vision to improve sound source separation. In: *Proceedings of 16th National Conference on Artificial Intelligence*. pp. 768–775.
- Neale, D.C. and Carroll, J.M. (1997). The role of metaphors in user interface design. In Helander, M. Landauer, T.K and Prabhu, P. (Eds.), *Handbook of Human-Computer Interaction*, Amsterdam, Elsevier Science. pp. 441-462.
- Nekrasovski, D., Bodnar, A., McGrenere, J., Munzner, T., and Guimbretiere, F. (2006). An Evaluation of Pan and Zoom and Rubber Sheet Navigation. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*.
- Norman, D. (1986). Cognitive Engineering. In D. N. Norman and S. W. Draper (Eds.), *User Centred System Design*. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 31-61.
- Norman, D (1988). *The Design of Everyday Things*. New York, Basic Books.
- Owsinski, B. (2006). *The Mixing Engineers Handbook*. Boston, MA, Thomson Learning Inc.
- Oviat, S.L. (2006). Human-centred design meets cognitive load theory: Designing interfaces that help people think. In: *Proceedings of the 14th annual ACM international conference on Multimedia*, Santa Barbara, CA. pp. 871–880.
- Otten, L.J., Alain, C. and Picton, T. (2000). Effects of visual attentional load on auditory processing. *NeuroReport* 11, pp. 875–80.
- Olzak, L.A., and Thomas, J.P. (1986). Seeing spatial patterns. In K.R. Boff, L. Kaufman, and J.P. Thomas, editors (1986). *Handbook of Perception and Human Performance, Volume I: Sensory Processes and Perception*. Wiley- Interscience: New York.
- Pan Y. (2012) Attentional capture by working memory contents. *Journal of Experimental*

- Psychology*. 64(2) pp.124–128.
- Pardo, B., Little, D. and Gergle, D (2012). Building a personalized audio equalizer interface with transfer learning and active learning. In: *Proceedings of the second international ACM workshop on Music information retrieval with user-centred and multimodal strategies*. ACM, pp.13-18.
- Piolat, A., Roussey, J. and Thunin, O. (1997). Effects of screen presentation on text reading and revising. *International Journal of Human Computer Studies*, 47, pp.565-589.
- Plaisant, C., Carr, D. and Shneiderman, B. (1995). Image browsers: Taxonomy, guidelines, and informal specifications. *IEEE Software*, 12 (2) pp. 21-32.
- Plumlee, M. D. and Ware C (2006). Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. *Computer-Human Interactions*, 13(2), pp.179– 209.
- Plumlee, M., and Ware, C. (2002). Modelling performance for zooming vs. multi-window interfaces based on visual working memory. In: *Proceedings of the ACM Working Conference on Advanced Visual Interface (AVI)*, 59–68.
- Ratcliffe, J. (2014a). Hand and finger motion-controlled audio mixing interface. In: *Proceedings of the international conference on new musical instruments for musical expression*, London, U.K., June 30- July 4. pp. 136-139.
- Ratcliffe, J. (2014b). MotionMix: A Gestural Audio Mixing Controller. In: *Proceedings of the 137th Audio Engineering Society Convention*. Los Angeles, California, October 9-12.
- Raveh, D, & Lavie, N (2015). Load-induced inattentive deafness. *Attentional Perception and Psychophysics* 77(2), pp.483-92.
- Rensink, R. (2012). The Management of Human Attention in Visual Displays. In: Roda, C (ed.), *Human Attention in Digital Environments*. Cambridge University Press. pp.63-92.
- Reiss, J.D. (2016). A meta-analysis of high resolution audio perceptual evaluation. *Journal of the Audio Engineering Society*, 64 (6).

- Robbins, N. (2005) *Creating More Effective Graphs*. Hoboken, New Jersey, Wiley.
- Robjohns, H. (2001) 'Universal Appeal: Universal Audio 1176LN Limiting Amplifier'.
In: *Sound on Sound* [Online] June. Available at:
<http://www.soundonsound.com/sos/jun01/articles/universal1176.htm>. [Accessed:
November 23, 2016].
- Rodgers, D., (2001). A Grammar for Zooming Interfaces: Using Interaction Design Strategies to Improve User's Navigation and Spatial Awareness. *Information Design Journal* 10, pp.250-57.
- Ronan, D., De Man, B., Gunes, H., and Reiss, J.D. (2015). The impact of subgrouping practices on the perception of multitrack mixes. In: *Proceedings of the 139th AES Convention*, New York, October 29–November 1.
- Rosenhaltz, R., Yuanzhen, L., Mansfield, J. and Jin, Z. (2005). Feature Congestion: A Display of Clutter. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, pp.761–770.
- Roth, R. E., Donohue, R. G., Sack, C. M., Wallace, T. R., & Buckingham, T. (2014). A process for keeping pace with evolving web mapping technologies. *Cartographic Perspectives* 78, 25-52.
- Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C. and Foxe, J.J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex* 17, pp.1147–1153.
- Rumsey, F. (2001). *Spatial Audio*. New York, Focal Press.
- Sabin, A.; Pardo, B. (2009). 2DEQ: An intuitive audio equalizer. In: *Proceedings of the ACM Conference on Creativity and Cognition*, Berkeley, CA, USA, 27–30 October. pp. 435–436.
- Sabin, A., Rafii, Z, and Pardo, B (2011). Weighting-Function-Based Rapid Mapping of Descriptors to Audio Processing Parameters. *Journal of the Audio Engineering Society*, 59 (6), pp. 419-430.

- Saldaña, H.M. and Rosenblum, L.D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *Journal of the Acoustical Society of America*, 95 (6) pp. 3658-3661.
- Sanabria, D., Soto-Faraco, S. and Spence, C. (2005). Assessing the effect of visual and tactile distractors on the perception of auditory apparent motion. *Experimental Brain Research*. 166 (3), pp.548–558.
- Sanchez, C. A. and Wiley, J. (2009). To scroll or not to scroll: scrolling, working memory capacity and comprehending complex text. *Human Factors*, 51(5), pp.730-738.
- Sanchez, A. and Branaghan, R. (2011). Turning to learn: Screen orientation and reasoning with small devices. *Computers in Human Behaviour*, 27, pp.793–797
- Santangelo, V., Fagioli, S., and Macaluso, E. (2010). The costs of monitoring simultaneously two sensory modalities decrease when dividing attention in space. *Neuroimage* 49, pp. 2717–2727.
- Sauer, C., Roth-Berghofer, T., Auricchio, N., Proctor, S. (2013). Recommending Audio Mixing Workflows. In: *Case-Based Reasoning Research and Development*, pp. 299–313. Springer.
- Seetharaman, p and Pardo, B (2016). Audioeasze: Crowdsourced Audio Production Tools, In: *Journal of the Audio Engineering Society*, 64 (9). pp. 683–695.
- Selfridge, R, and Reiss, J.D. (2011). Interactive Mixing Using Wii Controller. <http://www.elec.qmul.ac.uk/people/josh/documents/SelfridgeReiss-2011-AES130.pdf>. Accessed 10/10/11
- Shneiderman, B. (1996) The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, *Proc. VL*, pp. 226–242, 1996.
- Shneiderman, B. (1998) Designing the User Interface: Strategies for Effective Human-Computer Interaction. Addison-Wesley, Maryland.

- Shneiderman, B. and Bederson, B. (2005). Maintaining Concentration to Achieve Task Completion. In: *Proceedings DUX '05*.
- Shneiderman, B. (2000). Universal usability. *Commun. of the ACM*, 43, 5 (2000), 84-91.
- Schutz, M. and Lipscomb, S. (2007). "Hearing gestures, seeing music: Vision influences perceived tone duration" *Perception* 36(6) 888 – 897.
- Simons, D. (2004). *Studio Stories: How the Great New York Records Were Made*. San Francisco, CA: Backbeat Books.
- Sinnett, S., Spence, C. and Soto-Faraco, S. (2007) Visual dominance and attention: the Colavita Effect revisited. *Percept Psychophys* 69(5):673–686.
- Smilek, D., Dixon, M.J., Cudahy, C. and Merikle, P.M. (2002) Research Report: Synesthetic color experiences influence memory. *Psychol Sci.* 2002;13(6):548–552.
- Smith, S. L. and Mosier, J. N (1986). Guidelines for designing user interface software. Report ESD-TR-86-278. Bedford, MA: The MITRE Corporation.
- Soto-Faraco, S. and Spence, C. (2010). Assessing the role of attention in the audio-visual integration of speech. *Information Fusion*, 11(1), 4-11
- Sörqvist, P., Ljungberg, J. K. and Ljung, R. (2010). A sub-process view of working memory capacity: evidence from effects of speech on prose memory. *Memory* 18, 310–326.
- Szalva, W. (2009). Behind the Gear. *Tape Op Magazine*, No.73, pages 10-11.
- Stone, M (2006). "Choosing Colors for Data Visualization" http://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf. Accessed May 4, 2014.
- Stowell, D. and McLean, A. (2011). Live Music-Making: A Rich Open Task Requires a Rich Open Interface. <http://mcl.open.ac.uk/Workshop/7>. Accessed 1/08/11.

- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. In: *Cognitive Science* 12, 257-285.
- Szalva, W. (2009) Behind the Gear. *Tape Op Magazine*, No.73, pages 10-11.
- Tano, S., Yamamoto, S., Dzulkhiflee, M., Ichimon, J., Hachiman, T. and Iwata, M. (2012). Three Design Principles Learned through Developing a Series of 3D Sketch Systems: “Memory Capacity”, “Cognitive Mode”, and “Life-size and Operability” In: *IEEE International Conference on Systems, Man, and Cybernetics*. October 14-17, 2012, COEX, Seoul, Korea
- Theberge, P. (1997). *Any Sound You Can Imagine*. Wesleyan University Press
- Theile, G. (2000). Multichannel natural music recording based on psychoacoustic principles. In: *Proceedings of the AES Convention*, 19-22 Feb, Paris, France.
- Thüring, M., Hannemann, J., & Haake, J. M. (1995). Hypermedia and cognition: Designing for comprehension. *Communications of the ACM*, 38, 57- 66.
- Treisman, A. and Davies, A. (1973). Dividing attention to ear and eye. In S. Kornblum (Ed), *Attention and Performance IV* (pp. 101- 117). New York: Academic Press.
- Tufte, E. R. (1983) *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut.
- Van Merriënboer, J. and Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development* September 2005, Volume 53, Issue 3, pp 5–13.
- Valeriotte, J. (2016). The Case for Mixing with Your Eyes Closed. Available at: <http://therecordingrevolution.com/the-case-for-mixing-with-your-eyes-closed/>. [Accessed December 3rd 2016]
- Walther-Hansen, M. (2017). New and Old User Interface Metaphors in Music Production. In: *Journal of the Art of Record Production* (11) 2017.

- Wanderley, M. and Orio, N. (2002). Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal*, 26(3):62–76, 2002.
- Ward, M. (2008). Multivariate data glyphs: “Principles and Practice”. In: *Handbook of Data Visualisation*, 2008.
- Wichmann, F.A., Sharpe, L.T. and #, K.R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2002, Vol. 28, No. 3, 509–520.
- Wickens, C. and Alexander, A (2009). Attentional Tunnelling and Task Management in Synthetic Vision Displays. *The International Journal of Aviation Psychology*, 19(2), 182–199.
- Wilson, S. and MacLean, R. (2011) *Research Methods and Data Analysis for Psychology*. McGraw-Hill Education / Europe, Middle East & Africa.
- Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In *Integrated models of cognitive systems*, W. D. Gray, Ed. Oxford University Press, New York, NY, Chapter 8.
- Wolfe, J.M., Klempe, N. and Dahlen, K. (2000) Post-attentive Vision. *The Journal of Experimental Psychology: Human Perception and Performance*, 26(2): 693-716
- Yamamoto, Y. and Nakakoji, K. (2005). Interaction Design of Tools for Fostering Creativity in the Early Stages of Information Design. *International Journal of Human-Computer Science* 63, pp513-535).
- Yeakel, N., and Vallier, J. (2002). *Live Performance Audio Mixing System with Aimplified User Interface*. US7742609B2.
- Yeh, M., Jin Jo, Y., Donovan, C. and Gabree, S. (2013) Human Factors Considerations in the Design and Evaluation of Flight Deck Displays and Controls. Federal Aviation Administration Report.
- Zenon, A., Ben Hamed, S., Duhamel, J. R., & Olivier, E. (2008). Spatial and temporal dynamics of attentional guidance during inefficient visual search. *PLoS ONE*, 3.

