The Influence of Graphical User Interface Design on Critical Listening Skills

Josh Mycroft, Joshua D. Reiss, Tony Stockman
Centre for Digital Music,
Queen Mary's, University of London.
j.b.mycroft@qmul.ac.uk

ABSTRACT

Current Digital Audio Workstations include increasingly complex visual interfaces which have been criticised for focusing user's attention on visual rather than aural modalities. This study aims to investigate whether visual interface complexity has an influence on critical listening skills. Participants with experience mixing audio on computers were given critical listening tests while manipulating Graphical User interfaces of varying complexity. Results from the study suggest that interfaces requiring the use of a scroll bar have a significant negative effect on critical listening reaction times. We conclude that the use of scrolling interfaces, by requiring users to hold information in working memory, can interfere with simultaneous critical listening tasks. These results have implications for the design of Digital Audio Workstations especially when using small displays.

1. BACKGROUND

In current Digital Audio Workstation (DAW) design, unlimited track counts, multiple effects plug-ins and a large number of conceptual additions have resulted in increasingly complex interfaces [1]. It has been suggested that this increased interface complexity risks focusing user's attention on the visual display to the cost of aural engagement [2], with many DAW users opting to turn off the VDU at times during mixing [3].

This paper highlights some of the perceptual and creative implications of mixing using screen based interfaces then proceeds to report the findings from a study designed to quantify the influence of Graphical User Interfaces (GUI) design on aural acuity. Participants with experience mixing audio on computers were given critical listening tests while manipulating GUIs of varying complexity. The results were analysed to see whether the visual presentation style influenced the critical listening skills typical of those required in audio mixing workflows.

2. INTRODUCTION

The increasing visual complexity of current DAWs has potential consequences for the successful mixing of audio. In creative terms, the need to navigate through several windows risks inhibiting the engagement and 'flow' of the mixing process. For example, they may impede the user's ability to make requisite adjustments such as pan, level and effects changes [4]. Furthermore, the interface may compromise the realisation of creative ideas, which due to their fleeting nature are 'lost' when the user has to negotiate a badly implemented GUI. [5].

In perceptual terms, the large amount of information on the screen and the navigation required to access it across multiple windows can place high cognitive load on short-term and working memory [6] and overload the limited capacities of the visual mechanism [7]. The large amount of visual detail within the interface may also bias the perception of auditory information in favour of visual information [8]. For example, Macdonald and Lavie [9] found that when test subjects 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 (79% in the high-visual-load condition, significantly more than in the low-load condition). In a similar study Dehais et al [10] found a link between complexities of the GUI and reduced aural awareness. In flight simulations 57 % of trained pilots failed to notice auditory alarms under high visual load conditions. The authors suggest that visual information processing interfered with concurrent appraisal of auditory alarms, thereby inducing 'Inattentional Deafness' [9]. In order to ameliorate the effect of visual overload when using these GUIs, they suggest a temporary simplification of the user interface (Cognitive Countermeasures) to redress this problem [11].

Given the complex visual presentation of many contemporary DAWs (with scrolling and window switching a major part of the interface navigation) and the increased use of small screen displays for music and audio mixing (such as Cubasis, Auria, Nanostudio and FL Studio Mobile) it may prove insightful to quantify how GUI complexity influences the speed and accuracy of critical listening tasks typical of audio mixing workflows. In so doing it is hoped that heuristics may be realised that acknowledge the perceptual limitations of the user, decrease cognitive load and minimise the extraneous complexity of the interface encroaching on the intrinsic complexity of the user's main task [12].

3. STUDY DESIGN

3.1 Participants

There were eighteen participants recruited (eight from the Centre for Digital Music, Queen Mary, University of London and ten from second year ‘A level’ Music Technology Students at City and Islington College, London). All participants were experienced using DAWs. All gave informed consent to participate in the study. The study was conducted in accordance with the guidelines of the University. The Ethics Committee of Queen Mary, University of London, approved the details of the study.

3.2 Procedure

Participants were played an excerpt of a mix of eight audio tracks which 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 panned over the duration of the excerpt (two minutes) till 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 response button (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 completing this critical listening task, the participants were asked to 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. This was done using four interfaces (figures 1-4):

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 one: This consisted of a play button, the three response buttons and the source and target equalisers.

Interface two: 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 three: 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 patch so they could acquaint themselves with manipulating the equaliser.

The four interfaces and panning file types were arranged in a randomised order and presented to the participants. The time it took to respond to the panned file was recorded for each interface, though this information was not visible to the participants and they were not told they were being timed.

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 skills would result in participants taking longer to hear the panning (which becomes easier to identify at extremes).

Figure 1: Control interface only displays response buttons.

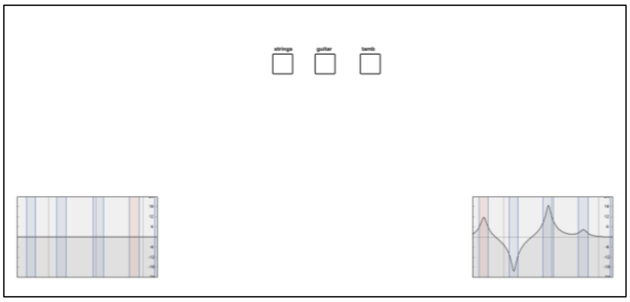

Figure 2: Interface 1 includes the addition of source and target equalisers.

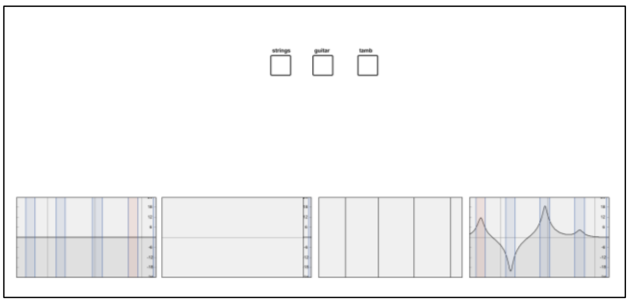

Figure 3. Interface 2 includes the addition of moving meters between the source and target equalisers.

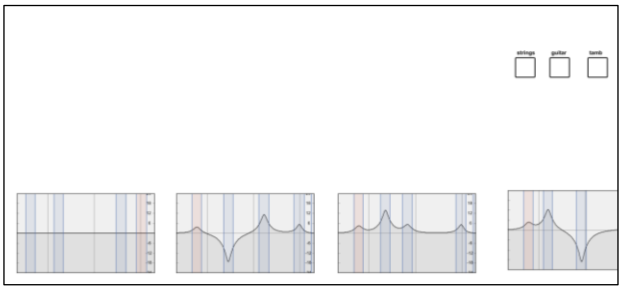

Figure 4. Interface 3 includes the addition of several equalisers, requiring scrolling.

4. ANALYSIS

Of the eighteen participants recruited, four 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 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 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 (see table one). A dependent *t*-test was then conducted between the control interface and the independent variable interfaces. The dependent *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).

5. RESULTS

While Interfaces one and two 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 dependent *t*-tests being greater than 0.05 (p>0.05). See table two.

However there were significantly slower response times for all three instruments in interface three (requiring scrolling) compared to the control interface. The dependent *t*-test consistently generated *P* values less than 0.05, thereby rejecting the null hypothesis at the 95% confidence level.

The time difference between the Control and the interfaces was also calculated to discern how the interface affected the ability to complete the task. The analysis (table three) shows that interface 3 (at 95% confidence level) has a range for the true population mean that is greater than the control across all three file types.

The analysis also reveals that overall the Control provided the fastest response for the majority of 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 (figure 5).

File		Interface type			
		Control	**1**	**2**	**3**
GUITAR	**Mean**	36.3	46.1	45.33	52.3
	SD	14.5	15.17	22.9	15.38
	CI (90%)	±6.88 29.41 To 43.19	±7.2 38.9 to 53.3	±10.87 34.46 to 56.2	±7.3 43.36 to 57.96
STRINGS	**Mean**	37.3	44.58	49.41	50.66
	SD	15.7	18.39	15.47	15.38
	CI (90%)	±7.45 29.85 to 44.75	±8.73 35.85 to 53.31	±7.35 42.06 to 56.76	±7.3 43.36 To 57.96
TAMB	**Mean**	49.0	51.9	53.83	66.41
	SD	16.94	19.49	18.68	21.78
	CI (90%)	±8.04 40.96 to 57.04	±9.25 42.65 to 61.15	±8.87 44.96 to 62.7	±10.34 56.07 to 76.75

Table 1. Mean time for task completion using the different interfaces.

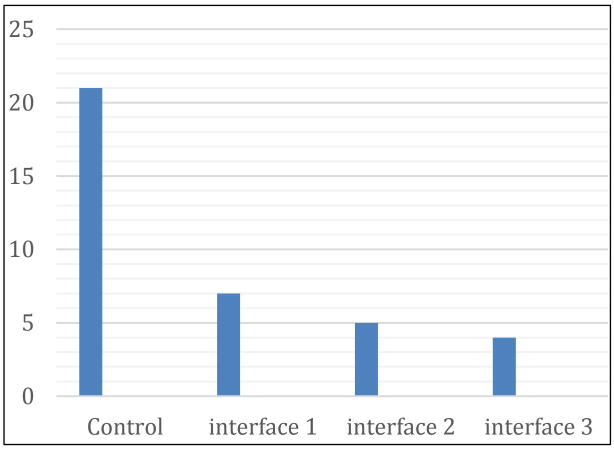

Figure 5. Occurrences of interface types being fastest across all participants and file types.

File Type	*P* values at 95% Confidence Intervals		
	Control to interface 1	**Control to interface 2**	**Control to interface 3**
Guitar	P = 0.120	P = 0.261	P= 0.033
Strings	P = 0.308	P= 0.070	P = 0.047
Tamb	P = 0.701	P = 0.514	P= 0.040

Table 2. The *P* values for time difference between Control and interface type.

6. DISCUSSION

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, though it is interesting to note that the control interface had the quickest reaction time across all the files. This result confirms a previous study by the authors [13] and aligns with research which postulates the independence of attentional resources for vision and audition [14, 15, 16].

However, as noted, introducing a scrolling interface has a significant effect on participant's critical listening reaction time. This may be due in part to the ergonomic issues of having to access information 'off the page', and future work will explore the influence of improving interface ergonomics on mixing workflow (see below). However, the negative effect on critical listening skills invoked by scrolling may be compounded by further cognitive issues, which require consideration. For example, Janata et al [17] 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' (page 9). Thus, attentive listening to music appears to be enabled by areas that serve general functions rather than by "music specific" areas. In this way the use of working memory and attention to process the visual task may consume most of attentional capacity, leaving little or none remaining for processing other modalities [18]. This notion is further supported by Tano et al [5] who consider the fragility of Short Term Memory (STM) as being at odds with complex Graphical User Interfaces, especially in creative support software (ibid). They suggest that software built for creativity support (in their case Design software) should be designed with the 'fragility' of STM as a corner stone of the design process.

Another factor to consider is the disorientation caused by scrolling, which may compound the problems of STM. Sanchez and Wiley [19] found disorientation an issue with scrolling interfaces since they lack a static 'place on a page' [19, p.731]. The context switching between the two views may result in users becoming disoriented or lost during reading. In a more recent study, Sanchez and Branaghan [20] found that by simply rotating small screen device displays by ninety degrees, and thus minimising the need to scroll, reasoning was significantly improved.

File type	Interface type	Mean	S.D.	Confidence Interval (95%).
Guitar	Interface 1	9.83	18.33	±10.37 -0.54 to 20.2
	Interface 2	9	20.88	±11.81 -2.81 to 20.81
	Interface 3	16	19.12	±10.82 5.18 to 26.82
String	Interface 1	7.25	17.03	±9.64 -2.39 to 16.89
	Interface 2	12.08	15.90	±9 3.08 to 21.08
	Interface 3	13.33	10.59	±5.99 7.34 to 19.32
Tamb	Interface 1	3.58	19.58	±11.08 -7.5 to 14.66
	Interface 2	4.83	22.68	±12.83 -8 to 17.66
	Interface 3	17.41	22.65	±12.82 4.59 to 30.23

Table 3. The time difference for task completion between Control and interface types.

Being aware of the cognitive and perceptual factors of GUIs may contribute to the optimal use of DAWs, especially when limited display area is a factor. In so doing it is hoped that the users will be better able to engage in "high-level planning, integrative thinking, and problem solving" rather than being sidelined by the interface itself [12, p.3].

7. FUTURE WORK

Future studies will explore the design and use of scrolling interfaces against modifications or alternatives that reduce STM load and disorientation. As noted in section 6, the problems of access caused by scrolling may contribute to disrupting the mixing workflow. To measure this influence, future studies will use alternative scrolling designs (such as vertical scrolling) which support the use of the scroll wheel. Additionally Overview + Detail designs will be evaluated to quantify to what extent this may reduce any disorientation caused by scrolling [19]. User definable displays will be trialed to reduce the amount of information on screen, thereby reducing the need to scroll. Future studies will also explore other interface objects frequently found in DAWs, such as dials and faders, so that a broader range of interface elements can be investigated. By so doing it is hoped further refinements can be made toward possible design heuristics for interfaces which allow monitoring of multiple sources of visual information while simultaneously supporting critical listening.

8. REFERENCES

[1] K. Golkhe, M, Hlatky, S. Heise, D. Black, J. Loviscach. “Track Displays in DAW Software: Beyond Waveform Views”. In: *Proc. Audio Engineering Society,* London, 2010.

[2] L. Crane, L. “This is your Brain Creating and Recording Music”. *Tape Op*, No.74, p.12, 2010

[3] N. Porter. “Mixing With your Eyes Closed” http://audio.tutsplus.com/tutorials/mixing-mastering/quick-tip-mixing-with-your-eyes-closed. Accessed 5/7/2012

[4] W. Szalva. Behind the Gear. *Tape Op Magazine*, No.73, pages 10-11, 2009.

[5] S. Tano, S. Yamamoto, M. Dzulkhiflee, J. Ichino, T. Hashiyama, M. Iwata. (2012). “Three Design Principles Learned through Developing a Series of 3D Sketch Systems: ‘Memory Capacity’, ‘Cognitive Mode’, and ‘Life-size and Operability’” *IEEE International Conference on Systems, Man, and Cybernetics.* COEX, Seoul, Korea, 2012,

[6] B. Schneiderman, & B. Bederson. “Maintaining Concentration to Achieve Task Completion”. *Proceedings DUX,* 2005.

[7] R. Rensink. “The Management of Human Attention in Visual Displays”. *In Human Attention in Digital Environments*. Edited by Claudia Roda. Cambridge University Press, 2012.

[8] M. Schutz, S. Lipscomb. "Hearing gestures, seeing music: Vision influences perceived tone duration" *Perception 36(6) 888 – 897*, 2007.

[9] J. Macdonald, N. Lavie. “Visual perceptual load induces inattentional deafness.” *Atten Percept Psychophys.* 73(6): 1780–1789, 2011

[10] F. Dehais, M. Causse, N. Régis,E. Menant, P. Labedan, F. Vachon, S. Tremblay. “Missing Critical Auditory Alarms in Aeronautics: Evidence for Inattentional Deafness”? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56: 1639, 2012.

[11] F. Dehais, C. Tessier, L. Chaudron. “GHOST: experimenting conflicts countermeasures in the pilot’s activity.” *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) 18, 163-168,* 2003.

[12] S.L. Oviatt. “Human-centered design meets cognitive load theory: Designing interfaces that help people think”. *Proceedings of the Conference on ACM Multimedia* New York, 871–880, 2006

[13] J. Mycroft, J.D. Reiss, T. Stockman. “The Influence of Visual Feedback on the Speed and Accuracy of Music Equalisation Tasks”. *Unpublished study,* 2012.

[14] A. Treisman, A. Davies. “Dividing attention to ear and eye”. In S. Kornblul (Ed.*), Attention and Performance IV (pp. 101- 117).* New York: Academic Press., 1973

[15] D. Alais, D. Burr. “The ventriloquist effect results from near optimal bimodal integration”. *Current Biology 14*: 257- 62, 2004.

[16] V. Santangelo, C. Spence. “Crossmodal attentional capture in an unspeeded simultaneity judgment task”. *Visual Cognition,* 16, 155–165, 2010.

[17] P. Janata, B. Tillmann, J. Bharucha. “Listening to polyphonic music recruits domain-general attention and working memory circuits”. *Cogn Affect Behav Neurosci* 2:121–140, 2002

[18] Lavie, N.“Perceptual load as a major determinant of the locus of selection in visual attention.” *Percept. Psychophys.* 56, 183–197, 1994.

[19] C. Sanchez, J. Wiley.“To scroll or not to scroll: scrolling, working memory capacity and comprehending complex text”. *Human Factors,* 51(5), 730−738, 2009.

[20] A. Sanchez, R. Branaghan.“Turning to learn: Screen orientation and reasoning with small devices”. *Computers in Human Behavior*. 27 793–797, 2011.