Audio Engineering Society
Convention Paper

Presented at the 142nd Convention
2017 May 20–23 Berlin, Germany

Formal usability evaluation of audio track widget graphical representation for two-dimensional stage audio mixing interface

Christopher Dewey[1] and Jonathan P. Wakefield[2]

[1,2] *University of Huddersfield, Queensgate, Huddersfield, West Yorkshire, UK HD1 3DH*

Correspondence should be addressed to Christopher Dewey (c.dewey@hud.ac.uk)

ABSTRACT

The two-dimensional stage paradigm (2DSP) has been suggested as an alternative audio mixing interface (AMI). This study seeks to refine the 2DSP by formally evaluating graphical track visualisation styles. Track visualisations considered were text only, circles containing text, individually coloured circles containing text, circles colour coded by instrument type with text, icons with text superimposed, circles with RMS related dynamic opacity and a traditional AMI. The usability evaluation focused on track selection efficiency and included user visualisation preference for this micro-task. Test subjects were instructed to click five randomly selected tracks for a six, sixteen and thirty-two track mix for each visualisation. The results indicate text only visualisation is best for efficiency however test subjects preferred icons and traditional AMI.

1 Introduction

The layout of the audio mixing interface (AMI) emerged in the late 1950s when engineers replaced the large three inch dials on the broadcasting consoles they were using for mixing music with slide-wires (i.e linear faders) [1]. The layout of the AMI has remained largely unchanged since this time with channels presented to the user as repeated vertical strips of controls that feature faders, knobs and buttons to manipulate and blend the constituent audio channels. This implementation-centric layout is termed the channel strip paradigm (CSP).

Recently researchers have proposed alternative designs based on psychoacoustic principles that correlate with sound localisation in humans [2-8]. Several of these studies [3-7] present the two dimensional stage paradigm (2DSP) as an alternative to this established AMI. The 2DSP adopts a depth mixing approach with tracks represented as circular widgets on a stage [6]. A widget's stage position relates to its pan position and perceived level, with the x-axis defining channel pan position and the y-axis defining the channels perceived level relative to the front of the stage, termed the listening position.

The 2DSP represents a significant improvement over the CSP in enabling the user to visualise the absolute and relative spatial distribution of audio channels. Although the paradigm has been adopted in only a few commercial AMIs [9, 10], it has received promising feedback from test participants in several studies [3-7, 11].

Unfortunately the 2DSP can become cluttered in when mixing sessions with higher track counts [6]. This is because tracks with similar pan positions and levels overlap on the display. Furthermore, the movable nature of the widgets, when compared with the static CSP display, presents another potential barrier to track selection. The effect of intuitive track identification/selection on User Experience (UX) was observed in prior work with users having to visually search the Graphical User Interface (GUI) for tracks when using some of the AMIs tested [11]. This visual search increased track selection time, and reduced UX, leading the authors to form the opinion that intuitive track selection is a key AMI user requirement. In light of this observation the aim of this paper is to investigate the effect of track widget visualisation style on track selection.

2 Background

2.1 Track widget visualisation

To the best of the authors knowledge, no study has yet been conducted that explicitly investigates the effect that track widget visualisation style has on track selection for the 2DSP.

Two studies have recently been conducted that explore 2DSP track widget visualisation with differing motivations [6, 7]. Gelineck & Uhrenholt [6] considered variations of the widgets to provide a visual representation of information “at a glance” for channel activity, monitoring of levels and monitoring of frequency content. In each variation the individually coloured circular widgets were replaced with an alternative, augmented visualisation style. In the variation that considered track activity tracks below a certain level were dimmed down on the display. Three variations were considered for the monitoring of levels including the mapping of real-time audio levels to circle diameter, circle colour brightness and by displaying a wavy line around each circle. For the monitoring of channel frequency content, three implementations were considered. The first mapped channel spectral centroid to circle colour brightness. The two other variations tested featured a line displayed around each circle which was “induced with noise” with the amount of noise added increasing with spectral brightness.

An informal evaluation approach was adopted which involved six professional engineers using each interface variation to mix multiple audio channels with no fixed mix task defined. The results of the evaluation indicate that the enhanced visualisations were not used directly for identifying differences between channels but performed a more supporting role, complementing the audio. The test subjects favoured dimming to indicate channel activity which proved very useful for quickly seeing which tracks were currently active. While the subjects found the noise lines displayed around each circle useful for indicating spectral content the visualisations were generally deemed too subtle. Gelineck & Uhrenholt advise that rapidly changing brightness and size should be used with care as they distract the user from the auditory task. Furthermore using circular widgets with large diameters increases the potential for clutter.

Mycroft et al considered the potential of harnessing Dynamic Query (DQ) filters to improve visual search tasks and critical listening tasks when using both the CSP and 2DSP [7]. This study built on prior work which indicated that visually more complex interfaces negatively affect a user’s ability to perform critical listening tasks [12,13]. DQ filters are commonly used to ameliorate clutter in web design, and in Mycroft’s implementation enabled the user to query pan position, level and individual channels. Mycroft’s implementation of the 2DSP featured a grey stage with numbered dark blue spheres for track widget visualisations. In the 2DSP the DQ filter was displayed as numbers. Selecting numbers on the x-axis queried pan position, y-axis queries level and numbers at the top of the screen allowed individual channel selection.

Mycroft et al adopted a two-stranded approach to evaluate each interface by considering a visual search task and a critical listening task. The results of this investigation revealed that the 2DSP with DQ filters enabled the subjects to perform the visual search tasks and critical listening task significantly more accurately when compared against the CSP and 2DSP with no DQ filters. Furthermore the 2DSP with DQ filters was most favoured by the test subjects overall.

These two studies indicate there is the potential when designing new AMIs that we risk overburdening the user with visual information which detracts from the primary auditory task. In order to explore this problem further it is first necessary to consider cognitive load with regard to mixing multi-track audio.

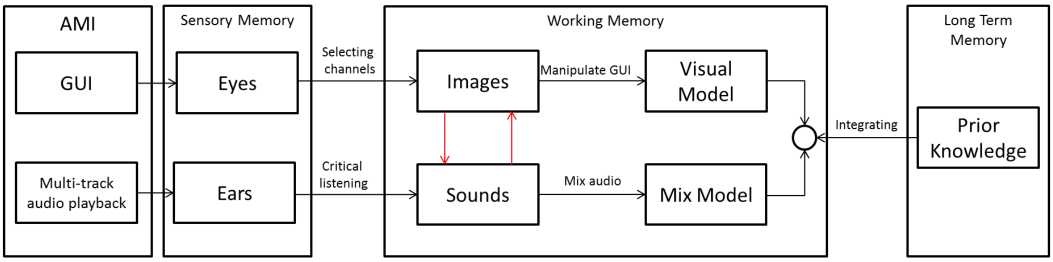

Figure 1: An overview of cognitive theory in the context of audio mixing (after Mayer & Moreno [20]).

2.2 Cognitive Load

In the context of user interfaces Cognitive Load (CL) theory [14] can be considered as the amount of mental resources required by the user to successfully operate the system to achieve their goals [15]. In this context the term goal refers "to a mental representation of an intention to accomplish a task" [16]. CL is both intrinsic and extraneous [17]. Intrinsic CL refers to the inherent level of difficulty of the task. Extraneous CL is generated by the manner in which the information is presented to us. Ideally designers should seek to reduce the extraneous CL when designing new interfaces.

Humans have a limited working (short-term) memory which is determined by our cognitive architecture [18]. This working memory is used for all conscious activities, is the only memory that we can consciously monitor and is limited to around seven items/elements of information [19]. Working memory primarily uses two separate information processing channels: auditory/verbal and visual [20]. Long Term Memory (LTM) in contrast, represents a repository for more permanent knowledge/skill and is filtered through the working memory as required.

Figure 1 provides an overview of cognitive theory in the context of audio mixing. The two rows represent the two information processing channels. The five columns represent the modes of knowledge representation and the arrows represent cognitive processing. In the working memory column the two red arrows that connect Images and Sounds illustrate the innate interplay between the two cognitive channels. When using a visually demanding interface there is the potential for a user's cognitive system to become overloaded if the processing demands placed by the interface exceed the processing capacity of the cognitive system. This cognitive overload results in a "split-attention affect", meaning the user can only focus on one information processing channel at a time [17]. In the context of the audio mixing interface, if the visual channel is overloaded, the user cannot focus on the auditory task of critical listening.

Whitenton (2013) presents three strategies for reducing CL in GUIs [15]. Firstly, designers should seek to remove visual clutter from the GUI. Secondly, GUIs should be built on existing mental models of systems they have used. Thirdly, designers should "offload tasks" by implementing designs that do not require the user to read extraneous text or remember information. This includes using pictures/icons and reordering information (for example the implementation of DQ filters in [7]). These strategies have been used to create a range of differing channel widget style visualisations for scrutiny via a specific track selection task.

3 Usability Evaluation

3.1 Channel Widget Visualisation Styles

Six different graphical channel widget visualisation styles were considered in this investigation. Figures 2-8 present example screenshots of each style. The first style was the simplest and represented channels with only text (Figure 2) and the second represented channels as white circles with a black border containing text (Figure 3). These two interfaces represented simplified representations of the 2DSP, adopting Whitenton's guidance of avoiding visual clutter. The third style represented channels as individually coloured circles containing text (Figure 4). This visualisation style has been used by Gelineck [5]. The fourth style considered was inspired by a common user approach to colour

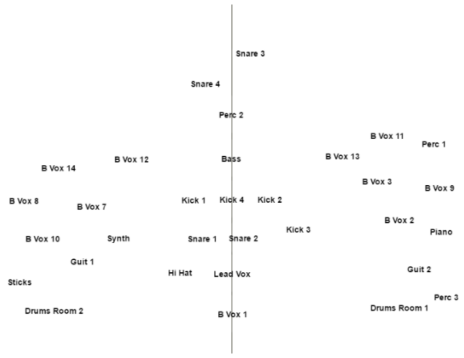

Figure 2: Interface 1: text only

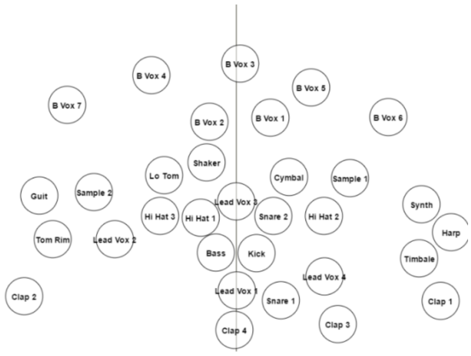

Figure 3: Interface 2: black and white circles

Figure 4: Interface 3: individually coloured circles

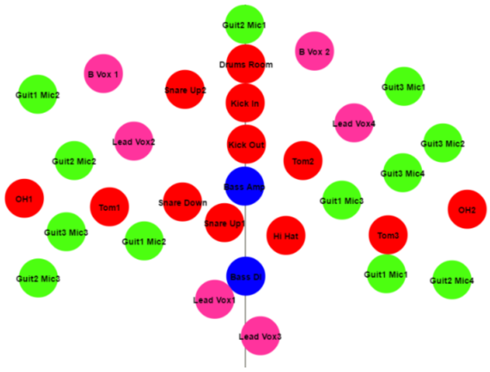

Figure 5: Interface 4: circles coloured by group

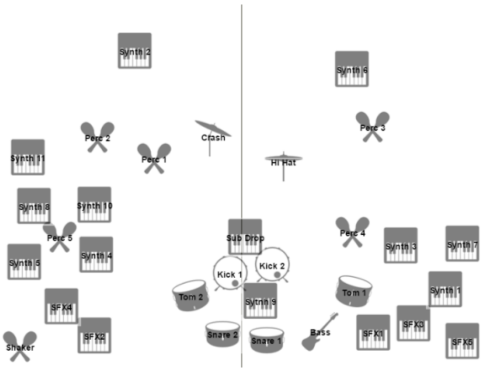

Figure 6: Interface 5: icons

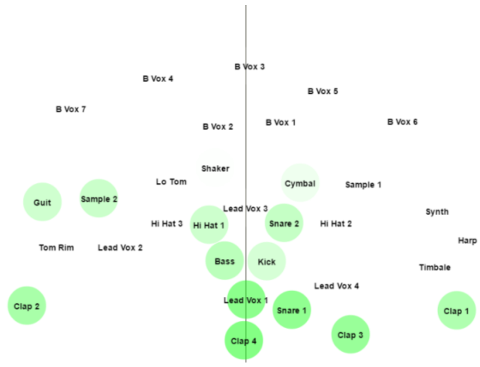

Figure 7: Interface 6: dynamic circles

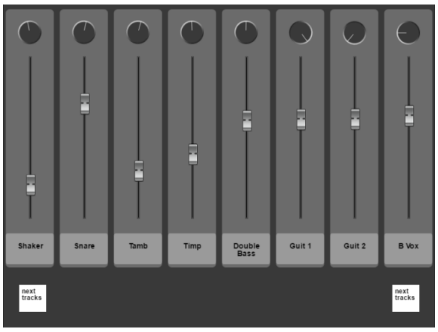

Figure 8: Interface 7: channel strip benchmark

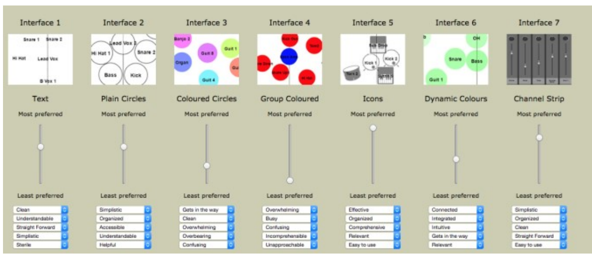

Figure 9: Evaluation interface

coding channels in the Digital Audio Workstation (DAW) depending on their instrument group and instead of colouring circles individually circles were coloured according to their instrument group (e.g. all drum track widgets coloured blue etc.) (Figure 5). This design was included in an attempt to build on a user's existing mental model as per Whitenton's second recommendation. In the fifth style the channels were represented by skeuomorphic icons of the channel's instrument with text superimposed (Figure 6). This style was included as it was favoured by participants in an earlier test [11] and conformed to Whitenton's third recommendation to use pictures/icons. The sixth style was more dynamic in nature and featured green circles, containing text, which changed opacity in real time relative to the channel's RMS level (Figure 7). This visualisation style is similar to Gelineck's implementation where inactive tracks were dimmed down. One further interface that replicated the DAW channel strip paradigm (CSP) was also developed to act as a benchmark during testing (Figure 8). This interface was similar to the implementation used by Mycroft et al. [7] and featured buttons for scrolling between tracks. Channels on this benchmark interface were ordered according to instrument type (i.e. drums, bass guitar, guitars, vocals, keyboards).

A consistent method of displaying channel name was adopted in all 2DSP interfaces. A single line of text was drawn over the centre of each widget to denote the channel name. The main 2DSP interface used with each visualisation style remained consistent and deliberately simplistic featuring a white background with a single line to represent the pan centre and black rectangular border to illustrate the bounds of the stage.

3.2 Evaluation Approach

A formal approach to usability evaluation was adopted to assess the relative merits of each interface. This approach has been successfully adopted by the authors to evaluate interfaces in prior work [11]. Usability is defined in the ISO 9241-11:1998 as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [21]. Wanderley advises conducting such tests using a focussed musical micro-task in a defined context [22]. For this reason the task of instructing test subjects to select channels using each track widget style while the associated audio tracks were playing was used to replicate a real-world scenario.

Track selection time was used as a measure of efficiency. An interface that enabled users to quickly identify/select target channels was deemed a primary user concern in prior work [11]. This assertion was reinforced by Calrec engineer, Henry Bourne in a recent panel [23]. Furthermore, in a study that considered differing methods of CL measurement, using selection time produced consistently reliable results [24]. Effectiveness was measured in terms of accuracy i.e. did the subjects select the correct target widget. Satisfaction was measured using a preference score. The Microsoft Desirablity Toolkit was used as a second subjective measure of satisfaction. Following Nielsen [25] the range of keywords was reduced from 150 to 25 to simplify the process of selecting keywords.

The evaluation software and candidate interfaces were developed in HTML5/JavaScript and harnessed a range of open source libraries for abstracting control of the Web Audio API (Tone.js) [26] and the display/control of visual elements via the canvas element (Create.js) [27].

The mixes were created in a 2DSP interface using short loops (less than 15 seconds in length) from twenty-one different multi-track sessions sourced from the Open Multitrack Testbed [28]. In order to correctly map the level of each track to a perceived distance all audio tracks were normalised to -23 LUFS. The sessions were then mixed to a standard deemed acceptable using the 2DSP in a bid to emulate the "in-the-mix" context. It is noteworthy to mention that in creating the sessions that contained 32 tracks, the potential for widgets to overlay was common place. In order to prevent any one widget being obscured, and therefore harder to select, the channel's mix positions were adjusted away from the authors preferred position e.g. a visually competing snare bottom track was panned slightly to the left of a snare top track.

3.3 Test structure

Eighteen test subjects took part in the formal evaluation. Each subject had at least one year's experience of music production. The subjects' ages ranged from 19 to 55. Four of the subjects were female and the rest male. The evaluation was conducted using a desktop computer with the web browser maximised to full screen. The audio playback level was pre-set and the subjects used their own headphones as a means of monitoring.

Each evaluation began with a training session which involved instructing each subject on how to participate in the evaluation alongside providing an introduction to the 2DSP. Any queries were addressed prior to commencing the evaluation.

The evaluation interface included a large box for displaying the candidate interface and a smaller text box positioned to the right of the screen to instruct the test subject to select a specific channel by clicking on it. This design consideration was employed to minimise the inherent addition of CL brought about by periodically adding/removing text instructions from the display during the test.

For each of the seven widget visualisation styles the subjects were instructed to click a randomly selected channel for each mix five times. Given the concerns raised regarding scalability of the 2DSP, three different mixes were considered for each visualisation. The first mix consisted of 6 tracks, the second 16 tracks and the third 32 tracks.

The order in which the interfaces were presented to the user and the audio material used in each test were randomised. Navigational prompts were used to step the subjects through the tests with opportunities provided for the subjects to pause if they so desired between tests. On average the test took around 15 minutes to complete.

Once all tests had been completed, each subject was presented with a user preference evaluation interface as shown in Figure 9. This interface featured a vertical strip for the evaluation of each interface including a picture of the interface, a vertical slider for preference rating and five keyword selection menus.

4 Results

4.1 Efficiency

Unsurprisingly track selection time increased with track count. Average track selection times (AST) for the 6 track scenario were the most consistent across all interfaces with subjects taking 2 seconds (+/- 0.3 seconds) to select target tracks. ASTs for the 16 track scenario show a similar relationship with subjects taking 3.5 seconds on average (+/- 0.3 seconds) to select target tracks with Interfaces 3 and 4 being slowest (ASTs of 3.8 seconds). The ASTs for the 32 track scenario was 5.3 seconds. In contrast to the 6 and 16 track scenarios, ASTs for the 32 track scenario fluctuated to a greater extent across the interfaces. Interfaces 2 and 3 have the slowest ASTs (6.1 and 6.4 seconds respectively) and interfaces 1 and 4 have the fastest ASTs (4.3 seconds and 4.6 seconds respectively).

Individual subject track selection time fluctuated in all tests that considered the 32 track scenario. This prompted a deeper scrutiny of the data. Conditional formatting of all subject track selection times was used to create-heat maps for each interface. Figure 10 shows an example plot for Interface 2. One column is provided for each test subject and the rows show the track selection times for the 5 tasks considered in each scenario. First 5 rows are for 6 track mix, second 5 rows are for 16 track mix and last 5 rows are for 32 track mix. Faster selection times are coloured green and slower selection times are coloured red.

The heat maps reveal this fluctuation to be most evident for Interfaces 2 and 3. This fluctuation was observed during testing, with subjects noticeably having to visually scan for a track for some time on one task and other times quickly and easily identifying a track.

	1203	2485	3500	1969	1845	1657	3090	2170	2132	1371	2632	1938	2663	1513	2449	2718	3655	1944
	1711	1631	1039	1911	1743	1480	2391	1120	1838	1495	1206	1007	1743	1388	2446	1894	2943	1485
	1775	2088	1023	1551	1455	1679	1614	1383	975	1182	1167	1174	1756	1903	3070	1127	2398	1134
	1001	1351	1919	2663	1255	1848	1351	895	1574	1463	976	1086	1415	1552	3622	1095	2135	1424
6	1363	1152	1095	1383	1911	1679	1494	1616	2336	1007	2248	936	1926	1911	4072	1014	1695	1774
	1105	4054	4437	6086	2239	2244	5102	2343	3153	1997	4177	2277	2914	3512	6565	2483	10585	4270
	2406	2629	2672	4327	3158	2271	5358	3089	6806	4965	3167	4191	3806	4451	2016	3333	3552	3246
	1697	4344	2087	2288	2071	1479	1888	3895	3201	1399	6966	7222	1231	1671	3517	1743	3303	2029
	1518	1967	3592	8208	3391	2535	2063	4063	2125	4007	2918	2222	2991	9775	2503	4541	4703	2327
16	1799	2271	1722	2703	1999	2742	6215	2032	7720	1534	3272	1950	4349	3183	3447	4575	4246	8981
	2409	4940	4509	13073	2530	25398	6479	16123	1704	2646	9176	16211	9510	8416	2562	7833	4717	6257
	2959	2422	2205	5118	3503	5160	1649	20614	6861	8222	3270	2968	3487	9135	3136	4100	4367	1991
	8278	5856	7631	5422	3903	11542	3046	15463	1969	6319	5198	7420	9846	1495	4519	2246	6294	4756
	5326	2326	3094	9792	2479	2913	2432	14910	7047	7293	16535	2078	3232	2118	15085	1415	11862	5703
32	9374	4318	16159	8462	1678	2136	4078	4687	13395	24118	1549	3535	6398	2580	3782	2971	2192	10105

Figure 10: Example heat-map for Interface 2

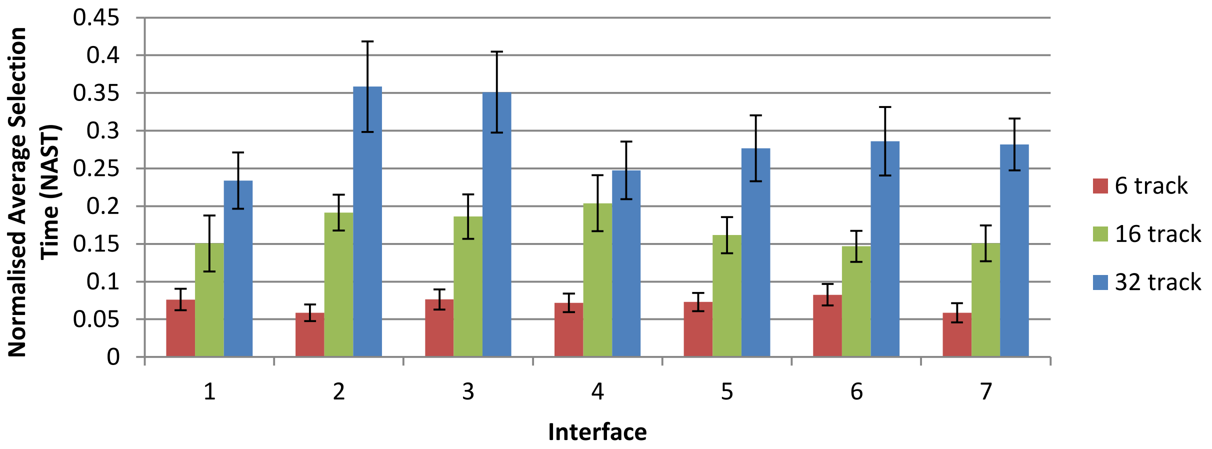

Figure 11: Normalised Average Track Selection Times (NAST) for all interfaces and scenarios considered.

To take account of differences between the subjects the AST results were normalised using the formula:

$$y_{(i)} = \frac{x_{(i)} - x_{(\min)}}{x_{(max)} - x_{(\min)}} \quad (1)$$

Where $y_{(i)}$ is the task's normalised track selection time, $x_{(i)}$ is the task's raw track selection time, $x_{(min)}$ is the test subject's fastest track selection time and $x_{(max)}$ is the test subjects slowest track selection time.

Figure 11 presents the Normalised Average Selection Time (NAST) for each interface and track count scenario. The NAST results echo the AST results with Interfaces 1 and 4 appearing to significantly reduce NAST when compared with interfaces 2 and 3 in the 32 track scenario. For 32 tracks, Interfaces 1 and 4 have a better mean than the benchmark interface (Interface 7) with Interface 1 appearing best, although the confidence intervals overlap so this can't be said with any statistical confidence. Interfaces 5 and 6 perform similarly to the benchmark interface.

For the 16 track mix, Interfaces 1 and 6 had similar NASTs to the benchmark interface, although given the confidence interval overlap nothing can be claimed with any statistical confidence.

For the 6 track mix, Interface 2 appears to perform comparably with the benchmark interface, although again nothing can be claimed with any statistical confidence.

Interestingly, whilst the results for Interface 4 do show an increase for NAST going from 16 tracks to 32 tracks, this is the only instance where we cannot be confident that the increase in track count has led to an increase in NAST. It may be the case that this interface might perform better with higher track counts but this will need further investigation.

4.2 Effectiveness

All the interfaces tested proved equally effective with over 99% of target channels correctly selected by the subjects. This is not surprising given the simplicity to the task considered in this evaluation.

4.3 Satisfaction

Figures 12-18 present word-clouds for the interfaces evaluated. There appears to be a strong correlation with the keywords selected by the subjects to describe each interface and the average preference score (shown in Figure 19). The keywords selected for Interfaces 3's word-cloud (which had the slowest AST/NAST) appears to mirror its reduced efficiency with subjects finding the use of colour *overwhelming*, *busy*, *confusing* and *gets-in-the-way*. Contrastingly, the keywords selected for the Interface 1, which scored well in terms of efficiency are very negative, i.e *dull*, *disconnected*, *simplistic*. Interface 2's word-cloud features a similarly disparaging selection of keywords. There appears to be a mixed opinion between the subjects with regard to Interface 6's keywords with some subjects

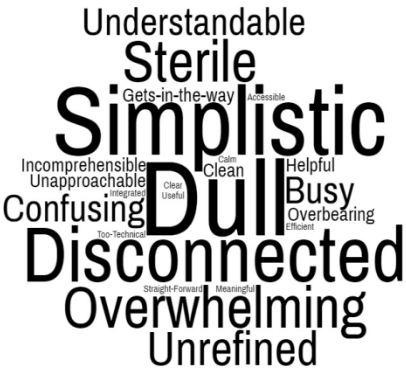

Figure 12: Word-cloud for Interface 1

Figure 13: Word-cloud for Interface 2

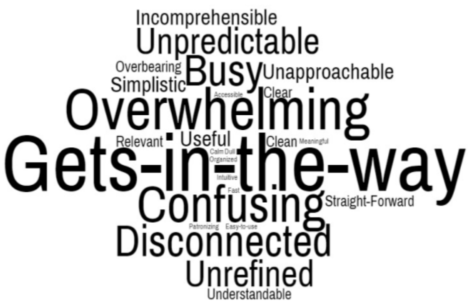

Figure 14: Word-cloud for Interface 3

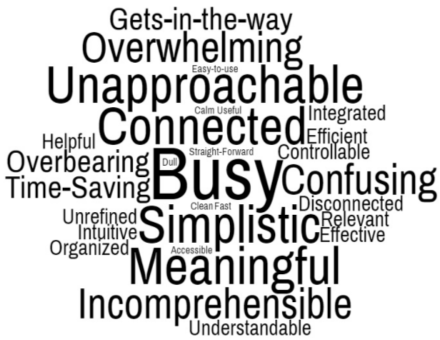

Figure 15: Word-cloud for Interface 4

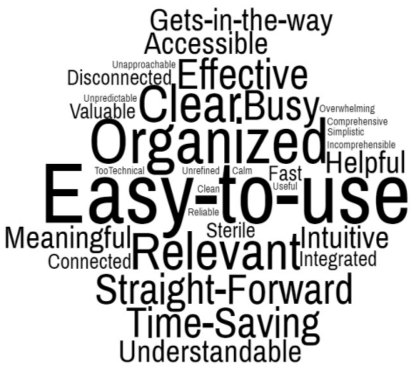

Figure 16: Word-cloud for Interface 5

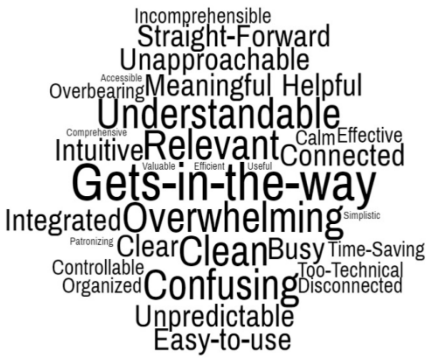

Figure 17: Word-cloud for Interface 6

Figure 18: Word-cloud for Interface 7

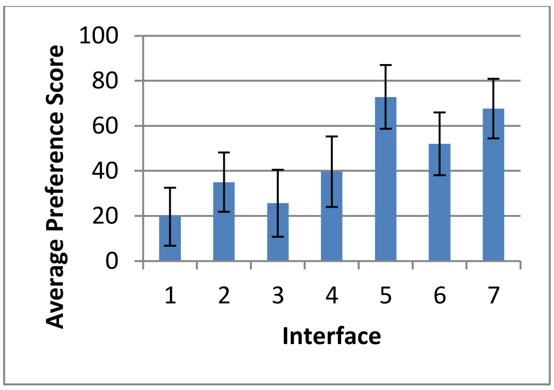

Figure 19: Average preference scores

selecting *clean*, *relevant* and *understandable* while others perceiving the dynamically changing display *overwhelming* and *gets-in-the-way*. Interfaces 5 and 7 both feature a predominance of positive keywords with subjects finding Interface 5 *easy–to-use*, *organised*, *clear* and Interface 7 *understandable*, *organised* and *clean*. Interfaces 5 and 7 have the highest average preference ranking scores overall with Interface 5 being favoured by the subjects significantly more than Interfaces 1, 2, 3 and 4. The familiarity of the bench mark Interface 7 arguably accounts for the subjects' apparent predisposition to this AMI. The satisfaction measures for Interface 5 indicate the subjects' preference for a 2DSP that features skeuomorphic representations of the channels over more simplified visual styles.

4.4 Discussion of results

The contrasting results for efficiency and satisfaction present conflicting results. Whilst the subjects appeared fastest selecting tracks using the text-only interface (Interface 1) in the 32 track scenario, they clearly preferred using a more complex widget visualisation style, namely icons (Interface 5) which is only fourth fastest in the 6 and 16 track scenarios, and third fastest in the 32 track scenario. One may conjecture that the preference for skeuomorphic icons on a metaphorical stage best supports the subjects' mental model of this AMI paradigm given its real world associations. The slower AST/NASTs for Interfaces 2 and 3 is particularly interesting because most published studies that consider the 2DSP implement widget visualisation styles that broadly conform to these two interfaces.

The scrolling CSP interface performs well in terms of efficiency and satisfaction. This is despite the subjects having to interact with the interface to a greater extent in the 16 and 32 track scenarios (which feature 2 and 4 pages of channels respectively). The authors believe this may partly be attributed to the enhanced user experience provided by ordering channels by instrument group. This belief is certainly reflected in the keywords selected for Interface 7. Future work should consider variations of the 2DSP that employ hierarchical structures to ordering channels in the GUI.

5 Conclusions

A formal usability evaluation that considered channel widget visualisation style for the 2DSP was conducted. Differences between the interfaces in terms of track selection time were observed in the 32 track scenario with the visually simplistic text-only variation (Interface 1) proving to be the most efficient variation overall. This result supports the view that visually simplistic interfaces reduce extraneous CL. The variations which presented channel widgets as black and white (Interface 2) and individually coloured (Interface 3) circles appeared least efficient for 32 channels. These apparent efficiency gains are not matched with regards to satisfaction, with subjects favouring more complicated visualisations over their simplified variations with Interface 5 (icons) the most preferred variation. This poses a potential issue for AMI interface designers considering performance improvements as they must balance what the user wants with what the user needs.

References

[1] Bell, A., Hein, E., & Ratcliffe, J. (2015). Beyond Skeuomorphism: The Evolution of Music Production Software User Interface Metaphors. ARP, 9.

[2] Gibson, D. A. (2005). U.S. Patent No. 6,898,291. Washington, DC: U.S. Patent and Trademark Office.

[3] Diamante, V. (2007). Awol: Control surfaces and visualization for surround creation. Technical report, University of Southern California, Interactive Media Division.

[4] Carrascal, J. P., & Jordà, S. (2011). Multitouch Interface for Audio Mixing. In NIME (pp. 100-103).

[5] Gelineck, S., Büchert, M., & Andersen, J. (2013, April). Towards a more flexible and creative music mixing interface. In CHI'13 Extended Abstracts on Human Factors in Computing Systems (pp. 733-738). ACM.

[6] Gelineck, S., & Uhrenholt, A. K. (2016, September). Exploring Visualisation of Channel Activity, Levels and EQ for User.

[7] Mycroft, J., Stockman, T., & Reiss, J. D. (2015). Audio Mixing Displays: The Influence of Overviews on Information Search and Critical Listening. (CMMR).

[8] Ratcliffe, J. (2014). Hand motion-controlled audio mixing interface. NIME 2014

[9] Bourne, H., & Reiss, J. D. (2013, October). Evaluating iBall—An Intuitive Interface and Assistive Audio Mixing Algorithm for Live Football Events. In AES 135.

[10] DSpatial (2016). Reality VR. Retrieved from http://dspatial.com/product/reality-vr/

[11] Dewey, C., & Wakefield, J. (2016, May). Novel designs for the audio mixing interface based on data visualization first principles. In AES 140.

[12] Mycroft, J., Reiss, J. D., & Stockman, T. (2013). The influence of graphical user interface design on critical listening skills. (SMC), Stockholm.

[13] Mycroft, J., Reiss, J. D., & Stockman, T. (2015). The effect of differing user interface presentation styles on audio mixing'. Proceedings of ICMEM.

[14] Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. Learning and instruction, 4(4), 295-312.

[15] Whitenton, K. (2013). Minimize Cognitive Load to Maximise Usability. Nielsen Norman Group.

[16] Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. Cognitive science, 26(1), 39-83.

[17] Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. Cognition and instruction, 8(4), 293-332.

[18] Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning.

[19] Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological review, 63(2), 81.

[20] Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. Educational psychologist, 38(1), 43-52.

[21] ISO (1998) 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.

[22] Wanderley, M. M., & Orio, N. (2002). Evaluation of input devices for musical expression: Borrowing tools from hci. Computer Music Journal, 26(3), 62-76.

[23] AES 2nd Workshop on Intelligent Music Production (13th Sept. 2016).

[24] Haji, F. A., et al. (2015). Measuring cognitive load: performance, mental effort and simulation task complexity. Medical education, 49(8), 815-827.

[25] Meyer, K. (2016)Using the Microsoft Desirability Toolkit to Test Visual Appeal. Neilsen Norman Group.

[26] Tone.js (2017). https://github.com/Tonejs

[27] Create.js(2016). http://www.createjs.com/

[28] De Man, B., Mora-Mcginity, M., Fazekas, G., & Reiss, J. D. (2014, October). The open multitrack testbed. In AES 137.