Audio Engineering Society
Convention Paper 9566
Presented at the 140th Convention
2016 June 4–7 Paris, France

Novel designs for the audio mixing interface based on data visualisation first principles

Christopher Dewey[1] & Dr. Jonathan P. Wakefield[2]

[1] University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK
c.dewey@hud.ac.uk

[2] University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK
j.p.wakefield@hud.ac.uk

ABSTRACT

Given the shortcomings of current audio mixing interfaces (AMIs) this study focuses on the development of alternative AMIs based on data visualisation first principles. The elementary perceptual tasks defined by Cleveland informed the design process. Two design ideas were considered for pan: using the elementary perceptual tasks 'scale' to display pan on either a single or multiple horizontal lines. Four design ideas were considered for level: using 'length', 'area', 'saturation' or 'scalable icon' for visualisation. Each level idea was prototyped with each pan idea, totalling eight novel interfaces. Seven subjects undertook a usability evaluation, replicating a 16 channel reference mix with each interface. Results showed that 'scalable icons', especially on multiple horizontal lines appear to show potential.

1. INTRODUCTION

This paper focuses on the development and evaluation of novel designs for the audio mixing interface (AMI). Initially it considers current AMI paradigms and summarizes current thinking in the literature before outlining data visualisation first principles. A usability evaluation is then conducted on a range of novel AMIs, which are based on these principles, and the results are discussed. The aim of this paper is to explore potential fundamental AMI paradigms for further consideration in future more sophisticated AMI designs.

2. AUDIO MIXING INTERFACE PARADIGMS

Originally the layout of the AMI was dictated by its underlying analogue electronic components leading to a one-to-one mapping of controls. Interestingly, since the 1970's most AMIs have continued to conform to this layout despite evolving from mainly analogue to mainly digital and software solutions. This AMI design is referred to as the Channel Strip Paradigm (CSP) (see Figure 3 for a simple example). Recently, researchers have questioned whether this commercially established paradigm really meets the needs of the user [1] and have proposed alternative designs based on psychoacoustic principles that correlate with sound localization in humans [2]. These proposed AMIs conform to the stage paradigm.

The basic concept behind this paradigm is that each audio channel is graphically represented on a stage by an icon/node. The position of each icon/node on the stage represents its level and pan. In contrast to the CSP, the stage paradigm adopts a ‘depth mixing’ approach [3] with regard to channel level with the icons/nodes closest to the user having the highest level. Although very few commercial embodiments of this paradigm exist [4, 5], it has been suggested as a possible alternative to the CSP in the academic literature given its psychoacoustic associations. Ratcliffe [2] helps define this paradigm further by distinguishing those solutions that feature a three dimensional stage and those that feature a two dimensional stage.

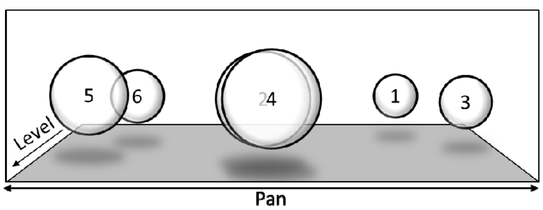

Figure 1 The three-dimensional stage paradigm

The three-dimensional stage paradigm proposed by Gibson [6] was the first attempt to present an alternative to the CSP and features a virtual cuboid stage with individual audio channels represented as coloured spheres as shown in Figure 1.

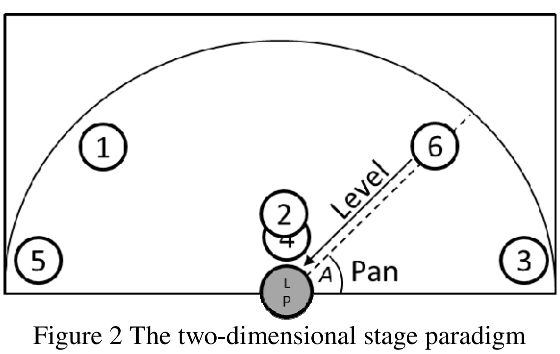

Figure 2 The two-dimensional stage paradigm

The two-dimensional stage paradigm has been considered in numerous studies [7, 8, 9, 10] and is shown in Figure 2. In contrast to the three-dimensional stage paradigm, the two-dimensional stage paradigm features a listening point aligned centrally at the bottom of the stage. The relative distance of each circle from this listening position relates to the channel’s level with those closer to the listening point being louder than those further away. The relative angle of each circle from the listening point defines the channel’s pan position.

Ratcliffe [2] argues that the whist the one-to-one mapping of parameters in the CSP offers precise control over many mix parameters, this paradigm offers no direct way to visualise the stereo distribution of audio channels as the user must scrutinise each channel’s pan knob position to assemble a mental image. Furthermore a channel to the left of the console may well be panned to the right potentially causing cognitive confusion. This assertion is reinforced by Mycroft *et al* [11] who argue that this visual task places an undue cognitive load on the user, detracting from their performance of the auditory tasks.

Both stage paradigms represent a significant improvement over CSP in enabling the user to visualise the absolute and relative spatial distribution between audio channels. Unfortunately these visualisations can become cluttered in real-world scenarios. Gelineck [10] remarks that because mix engineers are usually treating many channels of audio in any one mix the stage paradigm quickly becomes cluttered and potentially difficult to use. This is because channels with similar pan positions and level will overlap each other on the display as illustrated in Figures 1 and 2. This represents a deficiency with this paradigm.

As an alternative to existing AMIs, this paper explores whether there is potential to develop better AMIs based on data visualisation first principles.

3. DATA VISUALISATION FIRST PRINCIPLES

Shneiderman [12] asserts that interface researchers and designers are increasingly using data visualisations to display dynamic information because visual displays take advantage of the users’ cognitive ability to “scan, recognize, and recall images and ... detect changes in colour, size, shape, movement or texture”.

Green [13] advocates the use of Bertin’s ‘Image Theory' [14] to help produce informed graphical visualisations. Bertin asserts that data visualisation is a joint function of computer graphics and perception. Green attests that

'Image Theory' is the only coherent perceptual theory in the vision literature and that it closely parallels recent theories in human vision.

Bertin defines an image as the fundamental perceptual unit of any visualisation with each image consisting of two parts termed components and invariants. A component is the concept conveyed to the user and an invariant links these components together. Ideally, one image should be presented to the user for simplicity. Users extract information from data visualisations by firstly externally identifying what is being represented. The user then internally identifies how the components are mapped before perceiving what is being displayed, a process termed abstraction [15].

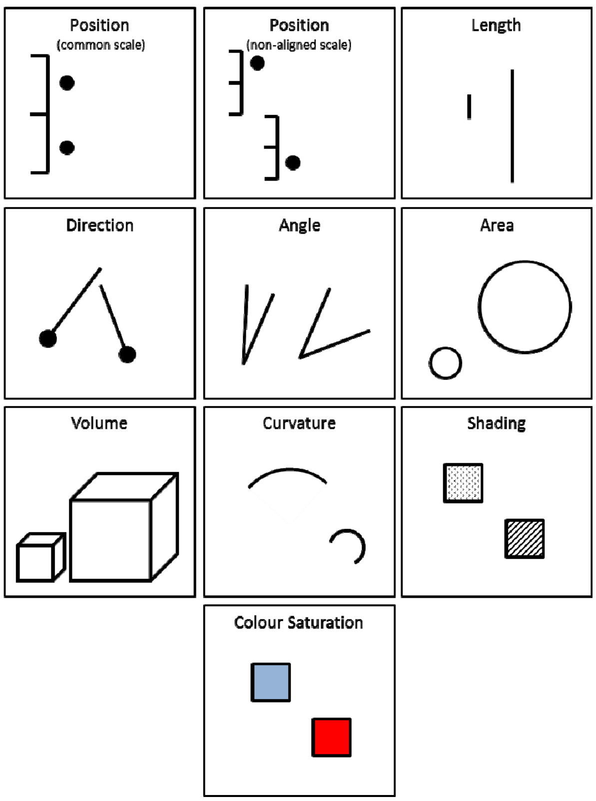

Figure 3 Elementary Perceptual Tasks (after Cleveland and McGill [16])

Bertin advises that three visual variables are optimal for a user to extract information from an image. These visual variables are classified as either planar or retinal. Planar variables exist as spatial dimensions, i.e. height, width and depth and retinal variables include size, colour, shape, orientation or texture. Importantly, each image must consist of both planar and retinal variables.

Building on the work of Bertin, Cleveland & McGill [16] identify ten 'elementary perceptual tasks' (see Figure 3) which closely relate to these visual variables. Cleveland suggests that we perform multiple elementary perceptual tasks when abstracting information from any visualisation or image and that these tasks can be ordered in terms of accuracy as shown in Table 1.

Elementary Perceptual Tasks	Order (in terms of accuracy)
Position along a common scale	1
Position along a non-aligned scale	2
Length, direction, angle	3
Area	4
Volume, curvature	5
Shading, colour saturation	6

Table 1 Elementary Perceptual Tasks ordered in terms of accuracy by Cleveland and McGill [16]

We have used these first principles to prototype a range of novel interfaces to control and display the core functionality of an AMI, namely level and pan.

4. INTERFACES

Following initial paper prototyping, two design ideas were selected for further consideration to visualise pan. In the first idea pan position was presented as a single horizontal line with all channels placed on the line according to their pan position. This design idea was based on the 'common scale' elementary perceptual task. The second idea presented each audio channel's pan position on individual vertically stacked horizontal lines (i.e. one pan line for each channel). This design idea was based on the 'non-aligned scale' elementary perceptual task. Three elementary perceptual tasks ('length', 'area' and 'colour saturation') were selected for further investigation to visualise level. In the design that represented level as 'length' each channel was represented by a vertical line with a height proportional to the channel's level. In the designs that represented level as 'area' and 'colour saturation' each channel was represented by a coloured circle with each circle's radius used to represent channel level in the former and

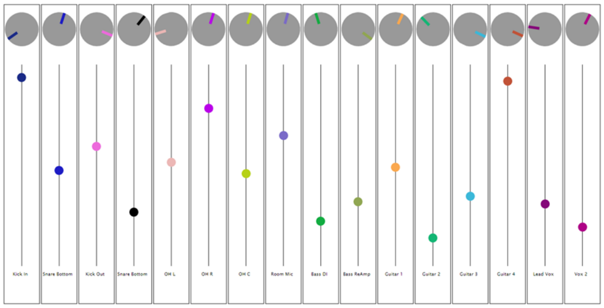

Figure 4 Interface 1: Channel Strip Paradigm

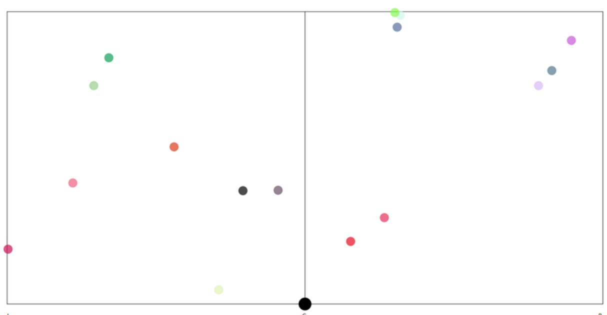

Figure 5 Interface 2: 2D Stage Paradigm

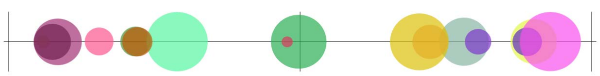

Figure 6 Interface 3: Pan 'common scale', level 'area'

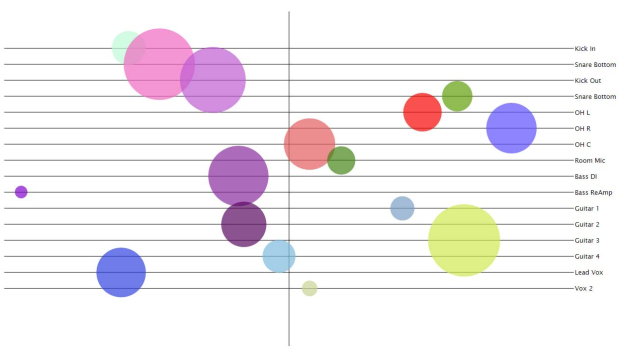

Figure 7 Interface 4: Pan 'non-aligned scale', level 'area'

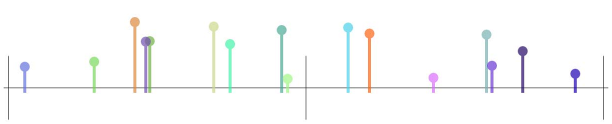

Figure 8 Interface 5: Pan 'common scale', level 'length'

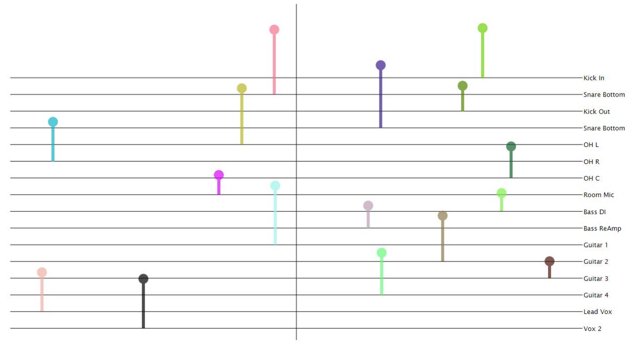

Figure 9 Interface 6: Pan 'non-aligned scale', level 'length'

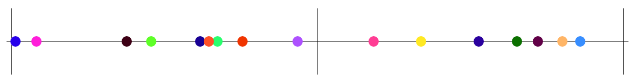

Figure 10 Interface 7: Pan 'common scale', level 'colour saturation'

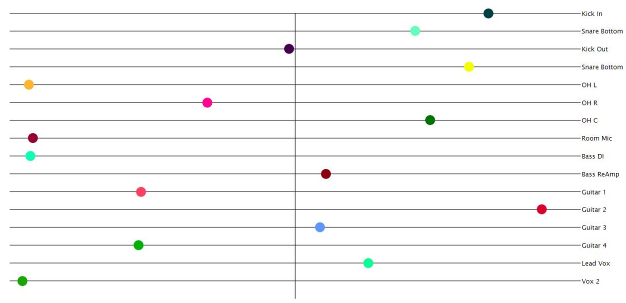

Figure 11 Interface 8: Pan 'non-aligned scale', level 'colour saturation'

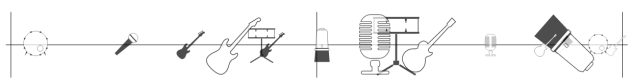

Figure 12 Interface 9: Pan 'common scale' level 'icon size'

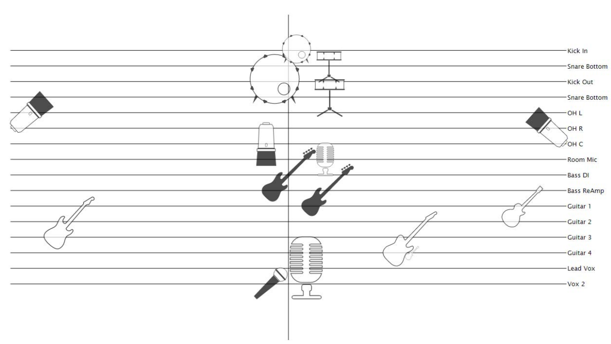

Figure 13 Interface 10: Pan 'non-aligned scale', level 'icon size'

colour intensity used to represent level in the latter. One further design was selected which displayed each channel as a graphical icon of the corresponding instrument with icon size proportional to channel level. This was inspired by Bertin's Image Theory coupled with the increasing use of icons in digital mixing consoles [18].

Both pan design ideas were realized in HTML5 and JavaScript for each level design; totalling eight candidate interfaces (see Figures 6-13). Two further interfaces were realised for benchmarking during testing. The first visualised the traditional CSP (see Figure 4) and the second visualised the two-dimensional stage paradigm (see Figure 5). A consistent drag-and-drop interaction style was adopted throughout the interfaces. The Web MIDI API was used to link the interfaces to an Ableton Live session containing sixteen audio channels. It was felt to be important to develop interfaces with a realistic number of audio channels from the outset given the previously identified shortcomings of the stage paradigm. In the 'common scale' interfaces the track name was displayed beneath the interface when the cursor was placed over the interactive visual elements (mouse over events). In the 'non-aligned scale' interfaces the track names were permanently displayed to the right of the interface.

5. USABILITY TESTING

Seven undergraduate music technology students, who regularly produce music, took part in a preliminary usability evaluation of the candidate and benchmark interfaces (a total of ten tests). Each test involved the subjects being asked to reproduce an 8 bar reference mix using one of the interfaces. This mix task only involved setting levels and panning for the 16 channels of the mix. The mix was a typical band with the following channels; kick, snare, hi-hat, hi-tom, lo-tom, overhead left, overhead right, bass guitar, four channels of guitar, lead vocal and three channels of backing vocals. Interface order was randomized per subject with the same audio material used in each test. The subjects used their own headphones to provide a means of monitoring with which they were familiar. Each test was supervised and any interesting observations/comments recorded. Prior to testing the subjects were allowed to practice each interface and any queries were addressed. Each test started with the reference mix playing and the pan position and level of all audio channels randomised. The test involved the subjects interacting with the interface using the computer's mouse to position each channel to match the reference mix. A button was provided to enable subjects to toggle between reference mix and their own mix. When the subjects were satisfied they pressed another button to end the test.

In accordance with ISO 9241 efficiency, effectiveness and satisfaction were measured to evaluate the usability of each AMI under test [19]. Satisfaction was measured by asking subjects to score each interface using a screen that featured a slider for each interface under test with a scale from 0 (least preferred) to 1 (most preferred). Subjects were also asked to select five keywords to best describe their experience of each interface in accordance with the Microsoft Desirability Toolkit [20].Task completion time was considered as a measure of efficiency. Effectiveness was considered in terms of error by comparing final channel settings to the reference mix channel settings.

6. RESULTS

6.1. Satisfaction

6.1.1. Preference Scores

Figure 14 shows that interface 10, pan 'non-aligned scale', level 'icon size' appears to be the most preferred although this result is only statistically significant against interfaces 3, 7 and 8. Interface 2, 2D Stage Paradigm, appears to be a close second with this again being statistically better than interfaces 3, 7 and 8. The results appear to suggest that interfaces 2 and 10 may be better than interface 1 which is the traditional channel strip paradigm interface. Given that all the subjects tested have several years of experience using the benchmark interface 1, and the preference bias this may lead to, this probably indicates that the 2D Stage Paradigm is actually a better AMI and helps to confirm Gelineck's assertions [10].

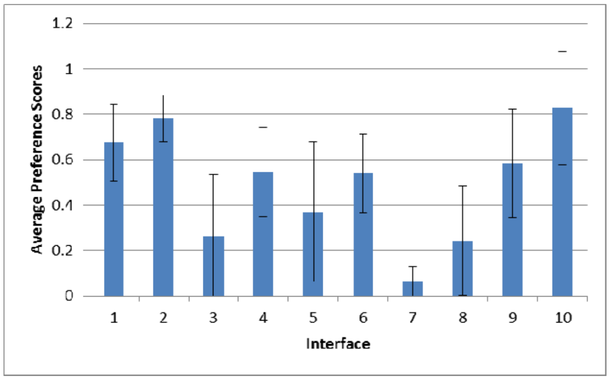

Figure 14 Average preference scores for all interfaces

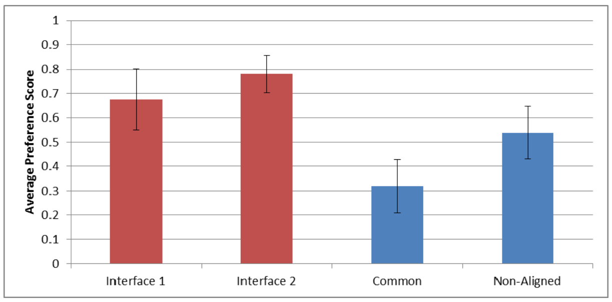

Figure 15 Average preference scores for common and non-aligned scale pan visualisation styles

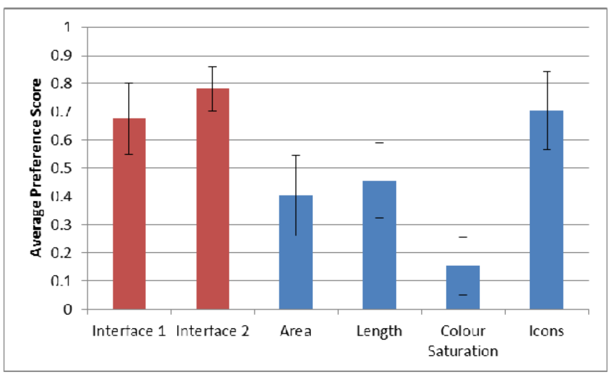

Figure 16 Average preference scores for each level visualisation style

In comparison with the 2D Stage Paradigm (interface 2) the combined results for 'common scale' and 'non-aligned scale' interfaces score statistically less favourably with subjects preferring the 'non-aligned scale' interfaces over their 'common scale' alternatives (see Figure 15).

Figure 16 combines the results for 'common scale' and 'non-aligned scale' interfaces and indicates that the 'icon size' interfaces (interfaces 9 and 10) appear to be preferred against the other new interfaces. They appear to have a similar level of preference to the benchmarks. Referring back to Figure 14, between the two 'icon size' interfaces, the 'non-aligned scale' appears to be better received than the 'common scale' alternative. This suggests that an interface that combines the 2D Stage Paradigm and icons could have potential.

6.1.2. Selected Keywords

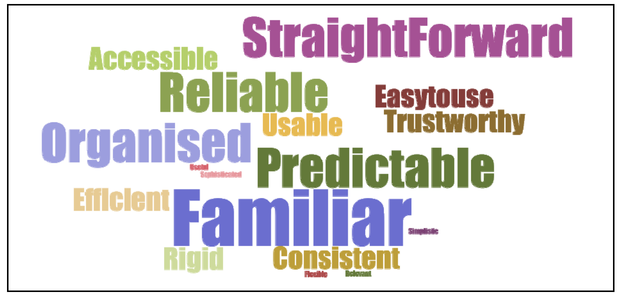

Figure 17 Word-cloud for interface 1

Figure 18 Word-cloud for interface 2

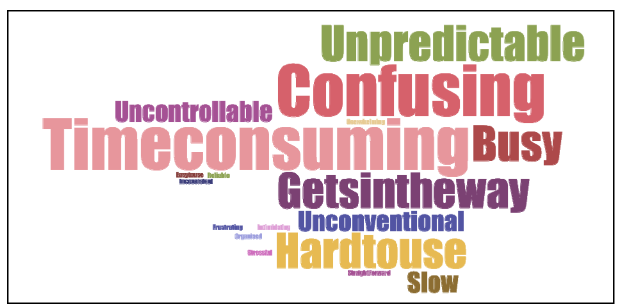

Figure 19 Word-cloud for interface 3

Figure 20 Word-cloud for interface 4

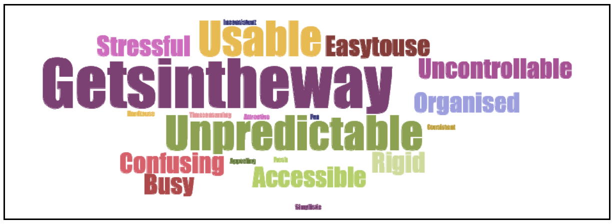

Figure 21 Word-cloud for interface 5

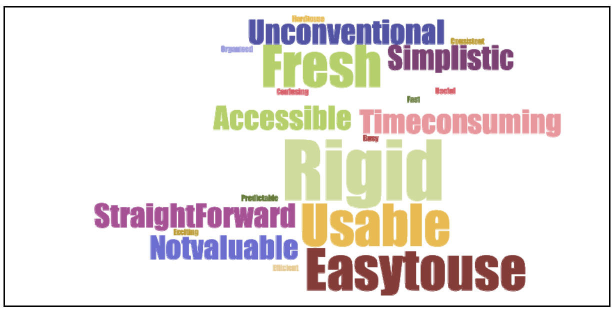

Figure 22 Wordcloud for interface 6

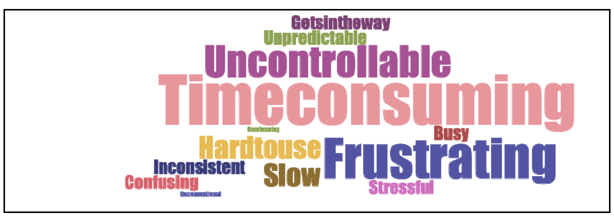

Figure 23 Word-cloud for interface 7

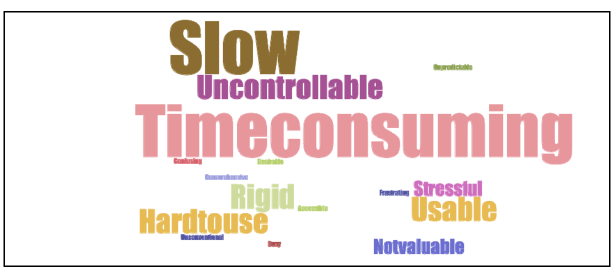

Figure 24 Word-cloud for interface 8

Figure 25 Word-cloud for interface 9

Figure 26 Word-cloud for interface 10

Figures 17 – 26 present the keywords selected by the subjects for each interface as word-clouds with the size of each word proportional to the number of times it was selected. These visualisations clearly support the preference score results with interfaces 1, 2 and 10 generally being assigned positive keywords and interfaces 3-9 assigned more negative keywords. Amongst the positive words for interface 10 it is interesting to see the prominence of the words ‘Fun’ and ‘Appealing’ against the more conservative positive words for interface 1, namely, ‘Familiar’, Predictable’ and ‘Straightforward’. The keyword ‘Timeconsuming’ is prominent on interfaces 3, 4, 7 and 8. Interfaces 7 and 8 use ‘colour saturation’ for level and this keyword is consistent with the efficiency assessment in the next section. The word-cloud for interface 4 appears slightly at odds with the other prominent keyword ‘Easytouse’ and its apparently good efficiency in the next section. All ‘common scale’ interface word-clouds prominently feature the keyword ‘Getsintheway’. The authors believe this is because the interactive elements can and do overlap each other on the display confirming the interface clutter issues raised in Section 2.

6.2. Efficiency

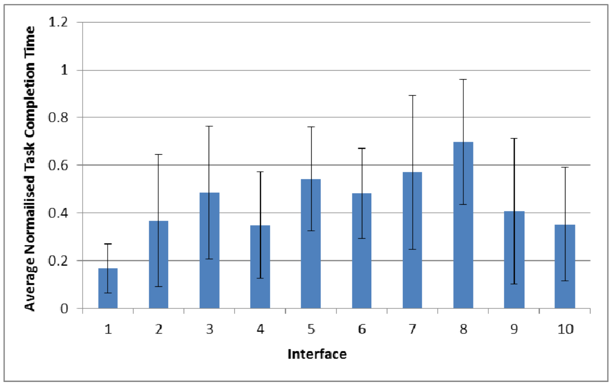

Figure 27 Average normalised task completion time for each interface tested

A normalised task completion time (NTCT) for each interface was calculated per subject by subtracting the subject's fastest completion time from the completion time in question and dividing this value by the subject's slowest completion time minus their fastest time. The average results for all interfaces are presented in Figure 27. Due to the small number of subjects tested and wide distributon of times it is difficult to deduce anything to conclusive from these NTCTs although the results do tentatively support the findings of the satisfaction measures (see interfaces 1, 2, 9 and 10). In contrast interface 4 appears out of line with the satisfaction results (also see comment regarding this interface in previous section).

During the tests subjects were observed having to 'search' for the desired audio channel to edit when using the pan 'common scale' interfaces by cursoring over the interactive elements to reveal the track name. This observation may account for the apparently increased average NTCTs (see interfaces 3 , 5 and 9 in Figure 27) when compared with the corresponding pan 'non-aligned' implementations in which track names were displayed as static elements towards the right of the interface (interfaces 4, 6 and 10). However, interfaces 7 and 8 do not appear to support this assessment, although this was the slowest interface overall and potentially the main issue with this interface is using colour saturation and not the alignment.

In contrast the subjects were observed performing more fluid interactions when using interfaces 9 and 10. It is our assertion that the test subjects intuitively used the icons to correctly differentiate and identify channels enabling them to interect more effectively with these two interfaces. This potentially supports the view that icons reduce cognative load on the user. Furthermore, the subjects were observed spending more time using these interfaces because they were more immersed in their interactions which indicates that efficiency may not always be the best way to evaluate on interface.

6.3. Effectiveness

When devising the test scenario we considered the task of recreating a mix to be focused in nature and originally intended to measure effectiveness in terms of accuracy by comparing the subject's final channel settings to the channel setting of the reference mix. A preliminary analysis of the kick drum (see Figure 28) shows that the average error in its level was higher than might be expected and there was even an error for the traditional interface (CSP) that all the test subjects are very familiar with. This result was not anticipated. This may be due to subject's mix replication and critical listening skills as much as their ability to interact with an interface.

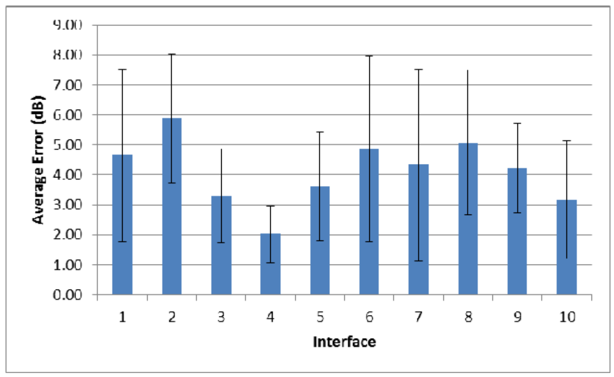

Figure 28 Average kick drum level error (dB)

Generally the subjects created the similar mixes with each interface although in some cases this differed significantly from the reference mix. In order to use accuracy as a measure of effectiveness in future tests we believe these results show that we must simplify the tasks undertaken by the subjects.

7. CONCLUSIONS

The results suggest that AMIs that combines the 2D Stage Paradigm and icons could have potential. Various combinations of these will be explored in future work.

Measuring satisfaction by asking subjects to score each interface using a screen that featured a slider for each interface under test was more informative than previous work which involved subjects ranking interfaces in order of preference [17].

Basing the design of level and pan only AMIs on data visualisation first principles does not appear to have been that fruitful. This may be due to the nature of the panning and level balancing task and may not be the case for other aspects of AMIs.

On reflection, normalised task completion time is probably a more appropriate measure of interface efficiency when the task is very clearly defined and does not depend on other factors, such as, critical listening skills. These tasks should focus on timing specific aspects of interaction; for example timing how long it takes subjects to identify, select or edit specific audio tracks. Additionally such tasks should help with the errors seen in the effectiveness results. For more complex tasks, such as creating a mix, interfaces may be better evaluated in terms of immersion.

8. REFERENCES

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

[2] Ratcliffe, J. (2014). Hand motion-controlled audio mixing interface. In Proceedings of NIME.

[3] 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.

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

[5] Iosono, (2013, November). AnyMix 1.5 AAX User Manual.

[6] Gibson, D. (1997). The Art Of Mixing: A Visual Guide To Recording, Engineering, And Production. Boston, MA: ArtistPro Press.

[7] Halliday, A. (2005, May). Audio Dementia: A Next Generation Audio Mixing Software Application. In Audio Engineering Society Convention 118. Audio Engineering Society.

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

[9] Carrascal, J. P., & Jordà, S. (2011). Multitouch interface for audio mixing. In Proceedings of New Interfaces for Musical Expression (pp. 100-103).

[10] Gelineck, S., Overholt, D., Büchert, M., & Andersen, J. (2013). Towards an Interface for Music Mixing based on Smart Tangibles and Multitouch. In New Interface for Musical Expression. Daejeon/Seoul, Korea Republic: Association for Computing Machinery.

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

[12] Shneiderman, S. B., & Plaisant, C. (2005). Designing the user interface 4th edition. Pearson

[13] Green, M. (2006). Toward a perceptual science of multidimensional data visualisation: Bertin and beyond. ERGO/GERO Human Factors Science, 1-30.

[14] Bertin, J. (1983). Semiology of graphics: Diagrams, maps networks. Madison WA: University of Wisconsin Press.

[15] Myatt, G. J., & Johnson, W. P. (2009). Making sense of data II: A practical guide to data visualisation, advanced data mining methods, and applications. John Wiley & Sons.

[16] Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American statistical association, 79(387), 531-554.

[17] Dewey, C., & Wakefield, J. (2014, April). A guide to the design and evaluation of new user interfaces

for the audio industry. In Audio Engineering Society Convention 136. Audio Engineering Society.

[18] X32 Digital Mixer User Manual. (2016) (1st ed.). Retrieved from https://media.music-group.com/media/PLM/data/docs/P0ASF/X32_M_EN.pdf

[19] User Focus. (n.d.). ISO 9241 Part 11: Guidance on usability. Retrieved from http://www.userfocus.co.uk/resources/iso9241/part11.html

[20] Benedek, J., & Miner, T. (2002). Measuring desirability: New methods for evaluating desirability in a usability lab setting. Proceedings of Usability Professionals Association, 2003, 8-12.