Visually Representing and Interpreting Multivariate Data for Audio Mixing

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

ABSTRACT

The majority of Digital Audio Workstation designs represent mix data using a channel strip metaphor. While this is a familiar design based on physical mixing desk layout, it can lead to a visually complex interface incorporating a large number of User Interface objects which can increase the need for navigation and disrupt the mixing workflow. Within other areas of data visualisation, multivariate data objects such as glyphs are used to simultaneously represent a number of parameters within one graphical object by assigning data to specific visual variables. This can reduce screen clutter, enhance visual search and support visual analysis and interpretation of data. This paper reports on two subjective evaluation studies that investigate the efficacy of different design strategies to visually encode mix information (volume, pan, reverb and delay) within a stage metaphor mixer using multivariate data objects and a channel strip design using faders and dials. The analysis of the data suggest that compared to channel strip designs, multivariate objects can lead to quicker visual search without any subsequent reduction in search accuracy.

1. INTRODUCTION

The majority of Digital Audio Workstation (DAW) designs represent mix data using a channel strip metaphor where individual controls are mapped on a one-to-one basis to mixing parameters. So, for example, equalisation, pan position, volume and effects (such as reverb) are all represented by different virtual controls. This can result in an increasingly complex interface [1, p.1] leading to a fragmented and disjointed approach to mixing [2]. Furthermore, the use of dials to represent the mix information (a major design element of channel strip designs) can be hard to interpret due to the fact that the human eye has difficulty comparing angles, specifically underestimating acute angles and overestimating obtuse angles [3 p. 49].

Within other areas of data visualisation (such as medial visualisations, geo-spatial and cartographic displays)

many-to-one mappings are used to simultaneously represent a number of parameters within one graphical object, by assigning data to specific visual variables such as position size, shape, hue, saturation, texture, opacity and dynamics [4,5,6]. This can 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 [7]. Indeed, research by Dewey et al, [8] has shown that the use of icon based mixers can not only reduce cognitive load on the user but also increase immersion. However, due to the limits of human visual perception, there are constraints on the design of multivariate data objects [5]. For example, while colours can be interpreted easily when displayed at reduced sizes they are liable to certain caveats such as the range of colours that can be effectively differentiated and the potential issue of 'colour blindness' among users [9]. Furthermore, some studies suggest that visually representing several streams of information at the same time can increase cognitive processing load [10,11].

In DAW design alternatives to the channel strip metaphor exist which make use of many-to-one mappings. For example, in [12] a virtual microphone position is used to represent the relative fader settings for multiple microphones around a sports stadium. Another alternative is the *stage metaphor,* a design which visualises channels as sound sources on a virtual stage where one can control pan position (relative left right position in the stereo field) and volume within a single User Interface (UI) object using its *x* and *y* positions [13,14,15]. Previous work by the authors has found that the consequent reduction in UI objects can minimise the need for navigation, allow significantly quicker visual search of mix parameters and improve concurrent critical listening tasks compared to an equivalent channel strip design [16]. However, a typical channel strip mixer will represent equalisation and audio effects as well as pan and volume position [17]. Being able to represent these within a stage metaphor design is therefore necessary in order to convey important attributes of the mix.

This paper therefore evaluates the efficacy of different designs to visually represent further mix parameters within a stage metaphor mixer by assigning mix parameters to multivariate data objects and comparing the visual search

times and accuracy to a channel strip mixer. By so doing the authors hope to convey mix information in a way which is perceptually and cognitively efficient and which optimally supports the interpretation of visual mix data.

2. STUDY ONE: REPRESENTING AN ADDITIONAL MIX PARAMETER

2.1 Participants

Participants were comprised of staff and 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 DAWs (with a minimum of five hours a week exposure to DAWs and mixing). Sixteen participants were selected (10 male, 6 female aged 17-19). The details of the study were approved by the ethics department of QMUL.
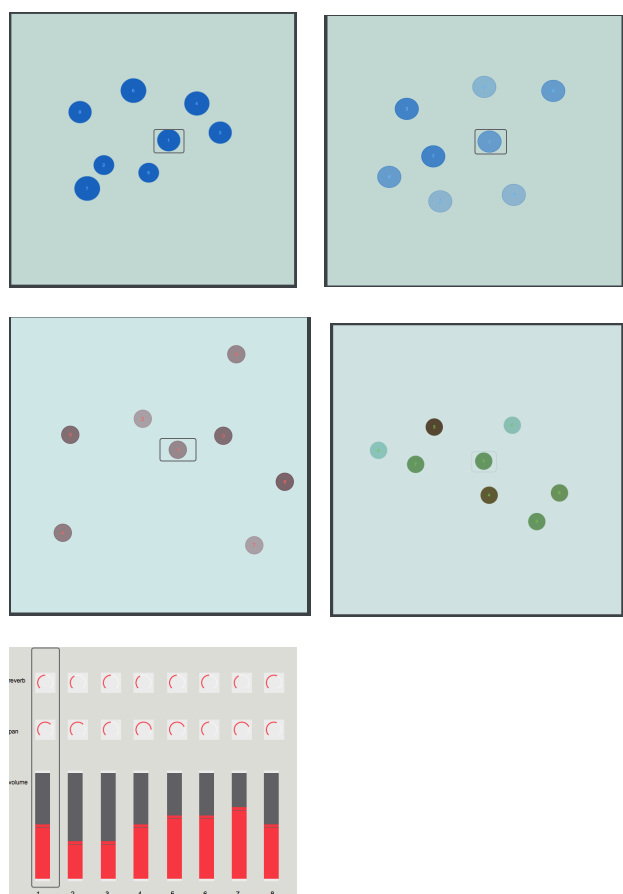

Figure 1. Screens for Study One: (a) top left; size, (b) top right; transparency, (c), middle left, saturation, (d) middle right, colours (e) bottom, the channel strip mixer.

2.2 Study Design

Five eight-channel mixers; a channel strip design and four stage metaphor mixers (figures 1, a-e) were designed using Max/MSP showing each channel's volume, pan and reverb amount (reverb is a commonly used audio effect which is often 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 frame of reference.

For the channel strip design, faders were used for volume, while dials were used for the pan position and the reverb. For the stage metaphor, 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). Rate of flashing (dynamics) was not used due to concerns that this might trigger seizures among people with photosensitive epilepsy [18]. Shading was discounted due to the difficulty of interpretation at the high zoom levels required to analyse the overview [19] and shape was not included since it is chiefly a categorical data set [20].

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. 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 [21]. Furthermore, some displays will not accurately display small colour differences due to varying visual display characteristics (ibid).

To represent increments of five reverb values using colour and saturation, twenty gradients were used (fig 2); gradient 1 showing reverb values of 1-5, gradient 2 showing 5-10 etc. For increments of ten, alternate gradients were used with each one representing a range of ten values (0-10, 10-20 etc.). For increments of twenty, only five gradients were used, each representing a range of twenty values (0-20, 20 40 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 five, ten and twenty sizes. To represent increments of five reverb values, twenty circle sizes were used (the smallest circle showing values of 1-5, the second smallest showing value 5-10 etc.), to represent increments of ten, alternate sizes were used (each depicting a range of 10 values) and to represent increments of twenty, five circle sizes were used (each representing a range of 20 values). In all cases smaller circle sizes represented less reverb. 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 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 at set +/- 20 (increments of 20). This created a total of fifteen screens for the study.

2.3 Procedure

Each participant was presented with each mixer design at the three increment differences between target and other channels (which were randomized for each participant). 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 with the exception of 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 randomized for each participant and they were presented one after the other.

The mapping of the designs to reverb amount (e.g. smaller circle size to less 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). Immediately after the study each participant was asked about their experience of using the different designs.

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

2.4 Analysis and Results

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 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.

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 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.

The participants were also asked to comment on using the different designs. Several of the participants said they found the transparency design very difficult, as it was hard to tell the difference between the reverb values, even at differences of 20%. A source of confusion for the colour design was the mapping of the reverb values; a number of participants expressed confusion over which way it was mapped, e.g. did lighter colours represent more or less reverb. This issue did not occur with size, where all participants were happy with the "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.

3. STUDY TWO: ADDING A FURTHER MIX PARAMATER

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 had at least one year's experience mixing on DAWs (with a minimum of five hours a week exposure to DAWs and mixing). For Study

Two, 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. The details of the study were approved by the ethics department of QMUL.

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 channel strip and stage metaphor designs. As with Study One, no audio was used, as the aim of the study was to evaluate the efficacy of visual representation and interpretation. As with Study One, 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 One. As outlined in section 2.4, size had performed best, while colour and saturation had been evenly successful. Transparency however had shown a significantly higher error rate (table 1); a result which corresponds with research suggesting that colour and size are the dominant visual channels and are most efficiently interpreted [6]. Transparency therefore was discounted for Study Two. Between colour and saturation, the latter was taken forward due to it being a colourblind safe design and due to the fact that multiple colours had resulted in some confusion from users over mapping.

Again, a channel strip design using faders and dials was included so that a direct comparison could be made between design outcomes of multivariate objects and current design paradigms. For the stage metaphor 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 One, the reverb and delay parameters were given values of 100 steps, and the mixers represented these in increments of 20, 10 and 5 divisions.

3.3 Procedure

Participants were presented with both designs of an eight-channel mixer (figure 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? The task was chosen as it required the simultaneous analysis of all four visual channels (*x* and *y* position, size and saturation).

There were 18 screens in total. Nine stage metaphor screens and nine channel strip 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 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 between 70 or 80, with the exception of one other channel which was assigned the same reverb and delay settings as the target.

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.

3.4 Results and Analysis

The amount of errors (incorrectly identified channels) were calculated for each participant in all 18 screens. From this the total number of errors made on each screen by all participants could be calculated (table 2). The results show that the percentage of errors in selecting the correct channel was higher when analysing the mix using faders and dials than the multivariate data objects.
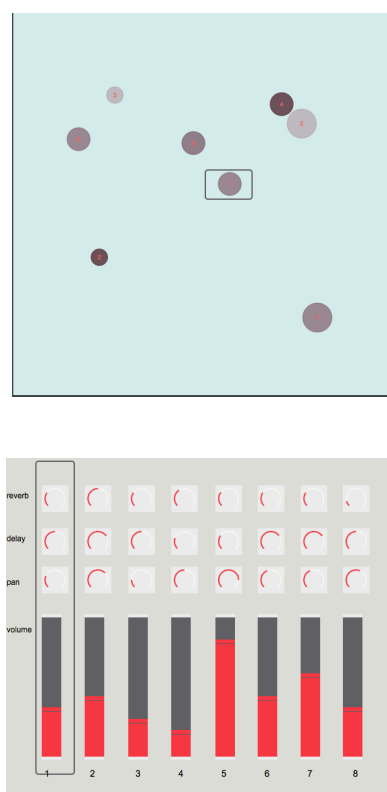

Figure 3. Top (a), the stage metaphor mixer; *x* and *y* positions show pan and volume, saturation of red colour shows delay amount and size shows reverb amount. Bottom (b) channel strip mixer; faders show volume, dials show pan, reverb and delay.

These results were analysed using a *z*-test for proportions to see if the error rate between designs was significant (at 95% CI). The analysis revealed that while the error rates for the multivariate design were much lower than the channel strip at 5% differences, there was no statistical difference between the two, which may be due to the increased visual load required to analyse colour, size and position attributes simultaneously [10,11]. At 10% increments, however, there was a significant difference in error rates in favour of the stage metaphor design. As with study one, no significant difference was found at 20 % differences, possibly due to the fact the perceptual difficulties found in estimating angles in dials ceased to be an issue when the difference was increased to this level.

The time 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 standard deviation were calculated. This was used to generate Confidence Intervals at 95%. The analysis revealed significant time differences in identifying the correct channels between the channel strip and stage metaphor designs with the former taking significantly longer at all increment levels (figure 4). The analysis suggests that the stage metaphor multivariate mixer allows users to find visual information significantly quicker without any subsequent increase in error rate.
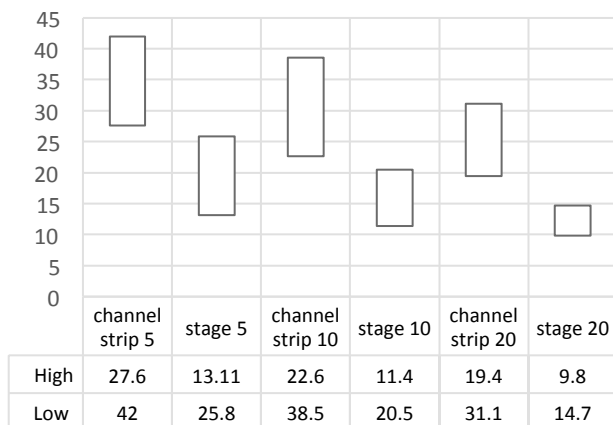

	channel strip 5	stage 5	channel strip 10	stage 10	channel strip 20	stage 20
High	27.6	13.11	22.6	11.4	19.4	9.8
Low	42	25.8	38.5	20.5	31.1	14.7

Figure 4. The visual search time (seconds) was significantly quicker in the stage metaphor design. 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%

Table 2. Error rates for both design at different value differences. The stage metaphor design was significantly more accurate at increments of 10 percent.

4. CONCLUSION AND FUTURE WORK

The results of these studies suggest that mapping mix attributes within a single multivariate object can result in significant improvements in visual search time and accuracy compared to a channel strip design. 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 a channel strip design. Given the increase in mix capacity and the reduction in screen size found in tablet computers this may prove useful in reducing screen clutter and helping users better analyse and interpret the visual information presented.

However, the results of these studies suggest 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. For example, 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 [21]. Multiple colours, while prone to mapping confusion, may be suitable to more ordinal tasks such as identifying which channels are grouped together (e.g. vocals, drums, percussion instruments etc.) [23]. 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 as mapping schemes become better understood [6, p.2].

The lack of significant improvement in error rates between the multivariate designs and channel strip designs at 5% increments may have been due to the increased visual load required to analyse colour, size and position attributes simultaneously [10,11]. Previous work by the authors has shown that the use of Dynamic Query (DQ) filters (UI objects such as sliders that facilitate real time visual display of query formulation and results) resulted in a higher amount of correctly completed visual and aural tasks compared to versions of the same interface without them [24]. DQ filters may be applicable to displaying multivariate data; allowing the user to visually explore and filter the information while continuously viewing the changing results.

Lastly, the authors acknowledge that this paper is preliminary in the sense that it focuses exclusively on visual aspects. Future studies should incorporate audio tasks alongside existing and multivariate designs to assess the extent to which they ameliorate potential difficulties in simultaneously analysing multiple data and help keep the users' attention optimally focused on interpreting both visual and auditory mix data.

5. 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] J. Mycroft, T. Stockman, J.D. Reiss. 'The Influence of Graphical User Interface Design on Critical Listening Skills'. In: *Sound and Music Computing (SMC),* Stockholm, 2013.

[3] N. Robbins. *Creating More Effective Graphs*. Wiley, 2005.

[4] J. Bertin. *Semiology of Graphics.* University of Wisconsin Press, Madison WI 1981.

[5] W. Cleveland. *Visualizing Data*. Hobart Press, Summit, New Jersey, 1993.

[6] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M.Chen. "Glyph-based visualization: Foundations, design guidelines, techniques and applications". In: *Eurographics 2013–State of the Art Reports*, pages 39–63. The Eurographics Association, 2012.

[7] M. Ward. Multivariate data glyphs: "Principles and Practice". In: *Handbook of Data Visualisation*, 2008.

[8] C. Dewey and J. Wakefield. 'Novel designs for the audio mixing interface based on data visualaisation forst pronciples'. In: *Proceedings of the 140th AES Convention*, Paris, June 4-7, 2016.

[9] M. Stone . "Choosing Colors for Data Visualization" http://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf. Accessed March 2016.

[10] R. Gudur, A. Blackler, V. Popovic and D.P Mahar, D. P. "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* 2009.

[11] S. Gelineck and D. Overholt. 'Haptic and Visual Feedback in 3D Audio Mixing Interfaces'. In: *Proceedings of the 9th Audio Mostly Conference,* Thessaloniki, Greece, 2015.

[12] H. Bourne, J. D. Reiss, "Evaluating iBall—An Intuitive Interface and Assistive Audio Mixing Algorithm for Live Football Events," 135th AES Convention, New York, Oct., 2013

[13] C. Dewey and J. Wakefield. 'A guide to the design and evaluation of new user interfaces for the audio industry'. In: *136th Audio Engineering Society Convention*, 26th-29th, Berlin, Germany, April 2014

[14] S. Gelineck and D. Korsgaard. "Stage metaphor mixing on a multi-touch tablet device". In *Audio Engineering Society Convention*, 2014.

[15] P. Gibson. *The Art Of Mixing: A Visual Guide To Recording, Engineering, And Production*. ArtistPro Press, 1997.

[16] J. Mycroft, J.D. Reiss, T. Stockman, T. "The effect of differing user interface presentation styles on audio mixing." In: *International Conference on the Multimodal Experience of Music (ICMEM),* Sheffield, 23-25 March 2015.

[17] M. Cartwright, B. Pardo and J.D.Reiss. "Mixploration: Rethinking the audio mixer interface". In: *19th Int. Conf. on Intelligent User Interfaces,* pages 365–370, Jan. 2014.

[18] About Epilepsy. Available:http://epilepsy.org/about-epilepsy/.org.

[19] M. Hlatky, K. Gohlke, D. Black and J. Loviscach. "Enhanced Control of On-Screen Faders with a Computer Mouse". In: *Proc. Audio Engineering Society.* Munich, Germany 2009.

[20] D. Chung, R. Laramee, J. Kehrer, H. Hauser and M. Chen. "Glyph-based Multifield Visualisation". In: *Scientific Visualisation*, Springer 2014.

[21] M. Harrower and B. Sheesley. "Designing Better Map Interfaces: A Framework for Panning and Zooming". In: *GIS*, 9, 77–89, 2005.

[22] H. Leitte and C.Heine. *Visual Data Analysis*. http://www.iwr.uni_heidberg.de/groups/CoVis. Pdf.

[23] D. M. Ronan, B. De Man, H. Gunes and J. D. Reiss. 'The impact of subgrouping practices on the perception of multitrack mixes,'. In: *139th AES Convention,* NY, 2015

[24] J. Mycroft, T. Stockman, J.D. Reiss 'Visual information Search in Digital Audio Workstations'. In: *Proceedings of the 140th AES Convention*, Paris, June 4-7, 2016.