COMPARING STAGE METAPHOR INTERFACES AS A CONTROLLER FOR STEREO POSITION AND LEVEL

Brecht De Man, Nicholas Jillings and Ryan Stables

Digital Media Technology Lab
Birmingham City University
[brecht.deman, nicholas.jilings, ryan.stables]@bcu.ac.uk

ABSTRACT

Of all music production interfaces, the channel strip with a gain fader and pan pot is likely the most persistent, being found in nearly all digital audio workstations and hardware as the main way to adjust level and stereo position. A popular alternative to the channel strip is the stage view, or stage metaphor, in which the level and stereo position (and possibly other parameters) are modified using the position of a moveable icon on a 2D or 3D image of a stage. When designing a stage view, there are several configurations to choose from, and it is not yet established which model is most appropriate for effective music production. In this paper, we discuss the pros and cons of three stage-view configurations, and conduct a preliminary user-study to evaluate the importance of the stage configuration, when used for a music production task.

1. INTRODUCTION

The stage metaphor, first proposed by Gibson [1] as a "virtual mixer", represents the gain and pan parameters of each channel strip as an object in 2- or 3-dimensional space. In the 2-dimensional representation, the horizontal axis represents pan, and the vertical axis represents gain. In the 3-dimensional representation, the additional dimension represents the frequency content of the track. This form of visualisation is intended to offer a more intuitive mixing environment based on its likeness to a sound stage (i.e. musicians standing at different positions on a virtual stage).

This metaphor has been translated into a popular alternative to faders and pan pots, as a method to not only visualise but also control the mixing process. In this case, the user adjusts the positions of the sources in a virtual live venue, to the extent that some digital mixing interfaces are literally a stage with draggable instrument icons. In this scenario, the mix engineer is crafting the listener's illusion of attending an imaginary concert, who is similarly visualizing the instruments in the same places.

This method of mixing has been shown to provide very similar results to the mixing desk method, when subjects were given the task of matching the gain balance and pan positions of a reference mix [2]. Participants are able to converge on the same relative mix with no significant variation in speed or accuracy. Additionally, similar studies show the stage model out-performs the channel strip when performing complex mixing tasks [3], particularly when trying to match the panning positions of individual channels in a mix. The sound stage representation has been implemented in a number of alternative stand-alone mixing tools [4, 5]. Rodriguez and Herrera [6] extend the model to work as a VST plugin in a traditional DAW, based on Gibson's original 3-dimensional representation. Here, loudness and frequency are mapped to parameters of the spherical track objects.

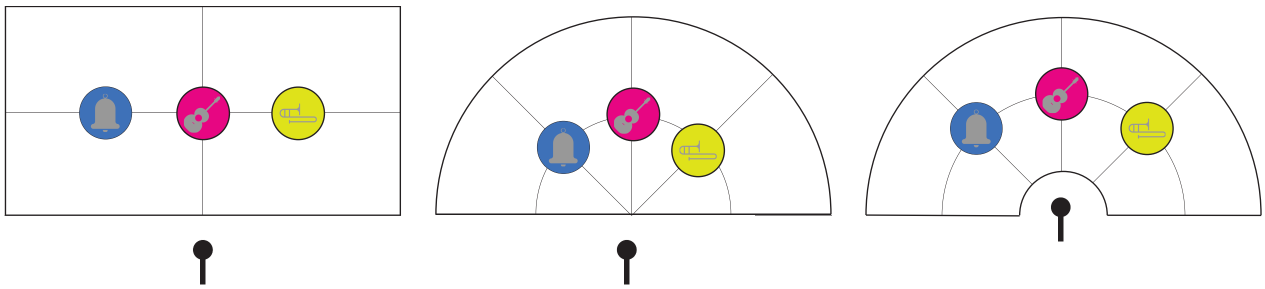
A reported issue with the stage visualisation is the potentially cluttered interface, particularly when objects occupy similar gain levels and pan positions. As more objects (tracks) are added to the mix, the stage representation becomes harder to navigate due to overlapping boundaries. To address this, Gelineck et al. [7] explore new methods of presenting channel information using features of the circular objects. Channel activity, level and auditory brightness are mapped to features such as the length of a circular line around the object, the noisiness of the object's boundary, and visual brightness. Whilst these features can provide additional information, the authors suggest that too much visual information can be overburdening to the user, adversely affecting the usability of the system.

Dewey and Wakefield [8] investigate methods to prevent this reported clutter, by exploring a range of methods for visualising the track objects in the mix. The authors explore the use of six visualisation styles: text only, black and white circles, individually coloured circles, circles coloured by group, instrument icons, and circles with dynamic brightness. The study shows that track selection times increased across the board when the number of tracks increases, however the text only objects provided the fastest reaction times from participants in the study. Interestingly, whilst the simplistic interfaces provided the most efficient method for mixing, users actually expressed subjective preference for the more abstract interfaces.

2. STAGE CONFIGURATIONS

In the conception of the stage view, Gibson [1] proposes a 3-dimensional model of the stage, which was used primarily for the visualisation of tracks in a mix. Here, the dimensions of the stage were intended to be psychoacoustically representative, in which the depth dimension repre-

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).



(a) Rectangular configuration (b) Polar configuration (c) Half-doughnut configuration

Figure 1: Three stage-view layouts

sents level, the azimuth dimension represents pan, and the elevation represents frequency content. When this is translated to a 2-dimensional interface, the frequency content is omitted, and the x/y-coordinates of the interface are used to control the pan and gain of a channel strip. Whilst this view traditionally comprises a rectangle, this does not accurately represent the acoustic level changes and pan positions with respect to a static listener. Here, we consider the rectangular configuration, along with two alternatives.

2.1. Rectangular configuration

The rectangular configuration (Figure 1(a)) is the most widely used representation of a stage. The rectangular area is intended to be a literal representation of a stage. Here, the Cartesian coordinate system is used to map the stage dimensions to the pan positions (x) and gain (y). Here, we can map between the two systems using a rescaling process, shown in Eq. 1.

$$s = \frac{(v - v_{min}) * (s_{max} - s_{min})}{v_{max} - v_{min}} + s_{min} \qquad (1)$$

where v is the value (level or pan pot position) and s is the relative stage position on the rectangle (x or y value).

One of the issues with this rectangular configuration is the lack of correspondence with the physical world. For instance, at a constant panning value (X position) and changing level (Y position), the angle of incidence from a virtual object changes with respect to a listener, standing centrally in front of a stage. Thus, in the physical world, the perceived panning would change. Similarly, at a constant level (say $y = y_{max}$, i.e. the front of the stage) and changing panning, this listener would in reality experience a varying level as well.

2.2. Polar configuration

To resolve this issue, we propose the use of a polar-coordinate system (Fig. 1(b)), whereby the distance from a static listener position (i.e. the magnitude of the x/y coordinate) represents gain, and the angle with respect to the listener represents the pan. Eqs. 2 and 3 highlight the mapping between the pan and gain positions on the stage interface (p_s, g_s) and the values on a channel strip (p_v, g_v).

$$\begin{aligned} p_s &= \sqrt{{g_v}^2 + {p_v}^2} \\ g_s &= \tan^{-1}(\tfrac{p_v}{g_v}) \end{aligned} \qquad (2)$$

$$\begin{aligned} g_v &= g_s \cdot \cos(p_s) \\ p_v &= g_s \cdot \sin(p_s) \end{aligned} \qquad (3)$$

This has the advantage of corresponding well to the physical world, provided a suitable distance-to-level and angle-to-panning mapping is used. However, it becomes cumbersome to use especially when panning , and even more so when there are several overlapping sources in this area.

2.3. Half-doughnut configuration

Whilst the polar system more accurately represents the perceptual relationship between the object and the listener, the interface becomes difficult to use when the level values of each object are high, as the space in which the object can move laterally is reduced. To address this issue, we propose a second polar coordinate method (Fig. 1(c)), in which the maximum gain values still correspond with a non-zero distance from the listener, allowing greater angular movement (and therefore panning) of the object.

3. EXPERIMENT

To assess which of these configurations is the most suitable for music production, we conduct a preliminary experiment in which four experienced music producers (≥10 years experience) were asked to take part in a simple mixing task using each of the interfaces. To do this, we present all three configurations of the stage view through a web interface, using the audio engine developed in [9]. Participants were asked to match a target mix in the indie genre comprising six stems (rhythm guitar, lead guitar, bass, snare, overheads L/R). The excerpt was 30 seconds in length and had a reasonable variance in both pan position and level. Interfaces

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

were presented to each user in a randomised order, in order to avoid bias caused by the subject learning the mix. Gain and pan positions were reset in between each trial.

Each subject was given as much time as needed to match the target mix. To quantify this we measured the time taken to reach the desired mix, and the mean absolute error of each stem for each of the stage configurations. The mix exercise was followed by an informal discussion in which subjects were asked to provide comments on the suitability of each interface for music mixing.

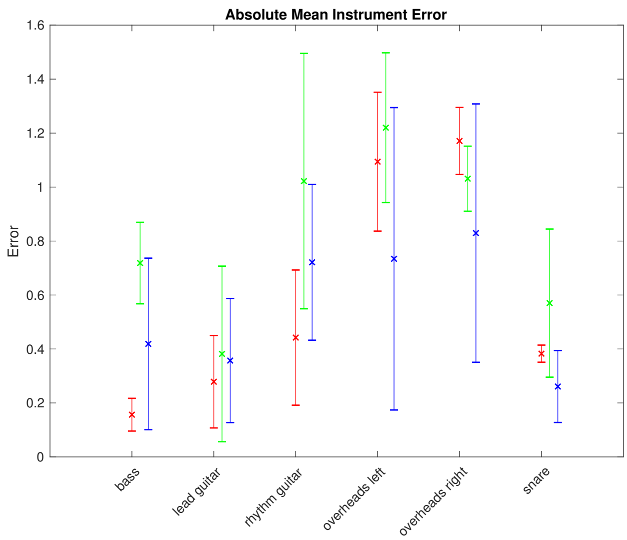


Figure 2: Absolute Mean Error, where the red line represents the rectangular configuration, green represents polar, and blue represents the half-doughnut method

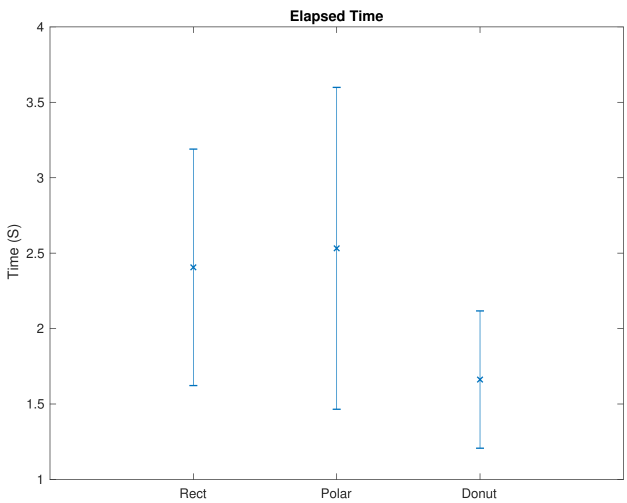


Figure 3: Elapsed time for each of the stage configurations

4. RESULTS AND CONCLUSION

As illustrated in Figure 3, the tests show that generally the half-doughnut method allows users to more accurately achieve the target mix, with the least error, however the variance in results does not indicate significance ($p \geq .05$). This is also true for the elapsed time measurement, as shown in Figure 3. Here, subjects generally performed quicker when using the half-doughnut interface. Of the four participants, one expressed a preference for the rectangular interface due to increased on-screen real-estate, and two expressed a difficulty to use the polar method, due to restrictions on panning at high gain. To conclude, the half-doughnut methods appears to resolve usability issues with the polar method, however more extensive user-tests are required in order to fully understand the benefits of each approach.

5. REFERENCES

[1] D. Gibson, *The Art Of Mixing: A Visual Guide to Recording, Engineering, and Production*. Thomson Course Technology, 2005.

[2] S. Gelineck, D. M. Korsgaard, and M. Büchert, “Stage- vs. channel-strip metaphor: Comparing performance when adjusting volume and panning of a single channel in a stereo mix,” in *NIME*, 2015.

[3] J. Mycroft, T. Stockman, and J. D. Reiss, “Visual information search in digital audio workstations,” in *AES Convention 140*, Audio Engineering Society, 2016.

[4] A. Holladay and B. Holladay, “Audio dementia: A next generation audio mixing software application,” in *AES Convention 118*, Audio Engineering Society, 2005.

[5] S. Gelineck and D. Korsgaard, “Stage metaphor mixing on a multi-touch tablet device,” in *AES Convention 137*, October 2014.

[6] B. J. Rodriguez Nino and M. Herrera Martinez, “Design of an algorithm for vst audio mixing based on gibson diagrams,” in *AES Convention 142*, May 2017.

[7] S. Gelineck and A. K. Uhrenholt, “Exploring visualisation of channel activity, levels and EQ for user interfaces implementing the stage metaphor for music mixing,” in *2nd AES Workshop on Intelligent Music Production*, September 2016.

[8] C. Dewey and J. Wakefield, “Formal usability evaluation of audio track widget graphical representation for two-dimensional stage audio mixing interface,” in *Audio Engineering Society Convention 142*, May 2017.

[9] N. Jillings and R. Stables, “An intelligent audio workstation in the browser,” in *Web Audio Conference (WAC 2017)*, London, UK, 2017.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).