Multitouch Interface for Audio Mixing

Juan Pablo Carrascal
Universitat Pompeu Fabra
Music Technology Group
Carrer de Tanger, 122-140
08018 Barcelona, Spain
juanpcarrascal@gmail.com

Sergi Jordà
Universitat Pompeu Fabra
Music Technology Group
Carrer de Tanger, 122-140
08018 Barcelona, Spain
sergi.jorda@upf.edu

ABSTRACT

Audio mixing is the adjustment of relative volumes, panning and other parameters corresponding to different sound sources, in order to create a technically and aesthetically adequate sound sum. To do this, audio engineers employ "panpots" and faders, the standard controls in audio mixers. The design of such devices has remained practically unchanged for decades since their introduction. At the time, no usability studies seem to have been conducted on such devices, so one could question if they are really optimized for the task they are meant for.

This paper proposes a new set of controls that might be used to simplify and/or improve the performance of audio mixing tasks, taking into account the spatial characteristics of modern mixing technologies such as surround and 3D audio and making use of multitouch interface technologies. A preliminary usability test has shown promising results.

Keywords

audio mixing, multitouch, control surface, touchscreen

1. INTRODUCTION

Even though today we are listening to very high quality surround systems - both in theaters and in our homes - and we might be about to enter the 3D audio revolution [19, 14, 16], the interfaces that we are using for mixing audio do not seem to have changed much since their introduction. In electrical terms, mixing implies the adjustment of variable-resistance controls (faders or potentiometers), which are standard components for electronic devices. Thus, these are the controls which have been traditionally used in mixing desk design [3]. This potentiometer-based interface design has been used up until our days, even if it is not necessarily ergonomical or adequate. In software, no big redesign has been proposed either [6, 11], and the majority of recent multitouch interfaces are simple adaptations of the same control schemes [18].

Maybe it is time to use what has been learnt with HCI research and question a trend which has ruled mixing console design for decades. We propose an initial prototype in which we emphasize these fundamental features:

- It gives importance to the spatial quality of sound. It may make use of position in a 2D space as a fundamental parameter in the mixing process, but should also have the possibility to control the z-axis position, making it ready for 3D audio applications.
- The use of a *listening point* (LP)
- The use a metaphorical interface: instead of channel strips controls and output busses, there are *channels* located in a *stage*.
- The use of multitouch technologies

2. STATE OF THE ART

Peter Gibson [7] suggests a "Virtual Mixer", a virtual 3D space in which sound sources can be located in three physical axes that correspond to perceptual sound parameters. The snapshots shown seem visually useful and didactic, but somehow cluttered and thus not too practical from a HCI point of view or for professional applications.

In professional audio mixing, there are some interesting options such as the Mackie DXB[1], which extends a digital mixing console with a pair of single-touch screens.

Multitouch technology have shown interesting possibilites. A brilliant example is the JazzMutant Lemur[2] interface, and an increasing number of applications for mobile platforms. The trend in these cases is to emulate the layout of mixers [18], which is precisely what we want to avoid and challenge.

In last year's NIME, the Cuebert mixing board was presented [12]. It heavily integrates a multitouch interface to enhance a mixing console for musical theatre applications. However it still uses the same channel strip approach as traditional mixers.

Vincent Diamante [5] suggests an interface which has some common features with the one presented in this paper. We think it could be seen more as a data visualization tool than as a new interface for professional audio mixing. Also, Diamante's work does not consider 3D mixing technologies. It's important noticing that its features and the arguments in his justification can be taken as a confirmation that HCI design for audio mixing is worth to be explored.

3. DESIGN

It is important to remark that we are not proposing a *mixer* but a *control interface*. As it has happened with several musical applications since the introduction of MIDI, and in some novel audio mixing technologies such as Meyer's D-Mitri[3], the control interface is separated from the functional engine [9]. This would allow a low-processing-power unit (such as a tablet computer) to control a digital audio processing engine. Because of its flexibility, we chose Open Sound Control (OSC) [21] as the communication protocol between the interface and the sound processing unit.

[1] `http://www.mackie.com/products/digitalxbus/`
[2] `http://jazzmutant.com/`
[3] `http://www.meyersound.com/products/d-mitri/`

After some preliminary user tests with a paper prototype [17], a hi-fi prototype was developed to allow us to apply user tests and evaluate our ideas. The hardware platform we used was the Reactable [10], from which we only used its multitouch capabilities, without the fiducials. The prototype created allows the detection of up to four fingers, but further developments should be able to handle more. A finalized product could be implemented on a very lightweight and portable platform such as an Apple's iPad.

3.1 GUI

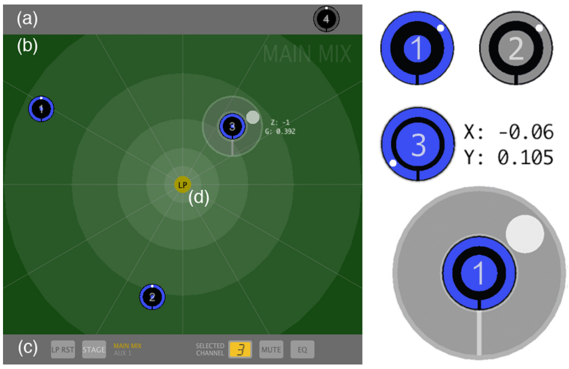

Figure 1: The GUI and the *channel*

The prototype has a simple yet flexible layout (Figure 1). There is a number (four in the prototype) of *channels* which can be dragged around the screen. There is an upper *inactive zone* (a), where all channels are initially located before being moved to the desired place. When channels are in this position, they remain muted and do not generate any control data. As soon as they are moved inside the *stage* (b), they become active. And just below the stage, there's a *control zone* (c) with buttons that control general functions: resetting the *Listening Point* (d) to it's default position, selecting between different stages, and displaying / muting / showing the equalizer for the currently selected channel.

3.1.1 The channel

We considered the channel to be the main element which should be changed from the traditional scheme. In standard mixers, it consists of a *channel strip* with lots of individual controls, mapped in a one-to-one basis to mixing parameters. Thus we took special care to create a control unit for it. We wanted to take advantage of the multitouch platform used, and we propose a versatile, multiparametric control scheme [8]. Of course, a control that can be dragged around a surface is going to have the inherent capability of controlling at least two parameters [4]. In our prototype, every channel (Figure 1, right) has these features:

- It can be dragged freely across x and y axes and Its position values are proportional to its panning and volume (in stereo mode) or to its surround (left-right and front-rear) panning (in surround mode).
- It has a gain control, by means of a surrounding "halo" with a marker that can be adjusted by moving it around the channel center.
- It has an internal circle whose diameter is proportional to the *z-axis* position parameter (for 3D audio environments). If the circle has a lower diameter, the channel is at a lower height, and viceversa.
- The value of the parameters is shown as they are adjusted.
- By watching the different channels in the stage it should be easy to spot and compare the values of their parameters at a glance.

Also, a fully functional, multitouch-enabled 4-band parametric equalizer interface is included for every channel.

3.1.2 The stage

Speaking in traditional mixing terms, the destination of the mix of a number of channels is called a *mixing bus* (the main mix, an auxiliary send, a recording bus, etc.). Electrically, this is just the cable that takes the electric signal from the summing circuit to the output connector [3]. We suggest the use of a metaphor for the bus in the interface. When mixing, we are going to locate sound sources in a sound space, or a *stage* (which represents the physical space, such as a studio or a live stage). Therefore, in our interface, each *stage* represents a possible destination for a sum of *channels*. For example, the main mix is one stage, an auxiliary mix for instrument monitoring could be another one, an auxiliary mix for a reverb effect could be yet another one and so on. Two stages were implemented in the prototype, a "Main Mix" and an "Aux 1" mix for adding a reverb effect. Every stage has a *listening point* (LP), and the panning of every channel present in the stage is determined by its position, relative to the position of the Listening Point. The Listening Point can be dragged around the stage with a two finger drag (to avoid accidental moves with a single finger). This way, it is easy to create custom mixes based on a reference mix without having to move every channel (Figure 2).

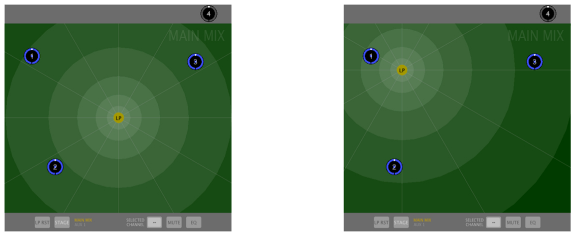

Figure 2: Two different mixes achieved by moving the Listening Point.

3.2 Functional Design

3.2.1 Software

We believed that a modular design process will make additional development or porting easier. Having that in mind, we chose Apple's *Quartz Composer* [2] for the development of our prototype. Quartz Composer (QC) is a visual, node-based programming language for graphical applications. It is released by Apple as part of the Xcode development tools.

Some of the patches (functional blocks used to create QC compositions) used in this project are not part of the basic QC distribution. These additional patches (Mansteri OSC sender[4], Kineme Structure Tools and Kineme Spooky Patch[5]) are, however, freely downloadable tools.

3.2.2 OSC address space

Currently, there are two main types of OSC messages generated by the interface, *stage* and *eq*, which are associated with stages and channels, respectively. This is the format of the *stage* messages (*muttmix* was chosen as the identifier for this project's OSC messages):

```
/muttmix/stage/N/ch/n A d E x y z g a
```

[4]http://www.mansteri.com/software

[5]http://kineme.net/QuartzComposerPatches/

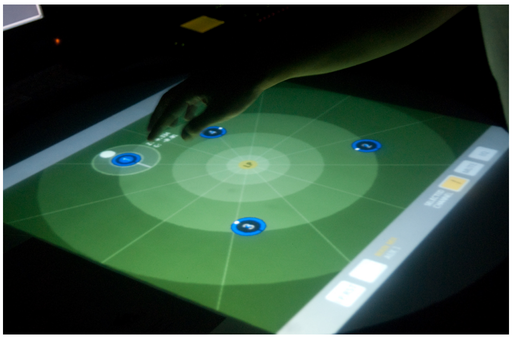

Figure 3: The prototype.

Where:

`N`: stage number
`n`: channel number
`A`: azimuth angle (float)
`d`: euclidean distance from LP (float)
`E`: elevation angle (float)
`x,y,z`: position - rectangular coordinates (float)
`g`: gain (float)
`a`: channel active (integer)

The position of the channel is sent as rectangular coordinates (x,y,z) as well as a 3D coordinate system compatible with the one used in Ambisonics [1] encoding and decoding. If the Listening Point is moved, every channel would send its updated position information, since their relative positions would change. This way the whole mix is readjusted to the new listening position.

The *eq* messages have this format:

```
/muttmix/eq/channel/n/b f g q G
```

Where:

`n`: channel number
`b`: band id (0 = low, 1 = mid-low, 2 = mid-high, 3 = high)
`f,g,q`: filter frequency, gain and Q (floats)
`G`: compensation gain (float) (Still not implemented)

3.2.3 Communication with a mixer

A Yamaha 01V96 [22] mixer, which has surround capabilities, was set up to be controlled by the proposed interface. For this particular setup, an additional tool was needed for translating and map the OSC messages generated by the multitouch interface into the appropriate MIDI continuous controllers understand by the 01V96. A custom Pure Data [15] patch was created for this purpose. The mixer was connected to a well calibrated surround monitoring system.

4. EVALUATION

A preliminary usability test was arranged, and two setups were made available in order to compare the performance of the users with both of them. The first setup is the same one described in the previous section. The other one involved using the controls of the 01V96 mixer directly. A four-track song was prepared, consisting of percussion, guitar, piano and voices. The basic working principles of both interfaces (01V96 and multitouch interface) were explained to all users (three men, three women, with ages ranging from 25 to 35; one of the men was a sound engineer, the rest had no previous experience in audio mixing). Users were asked to try to mix the song with the 01V96 mixer and with the prototype (in that order) trying to achieve certain specific spatial positioning of instruments. Though mixing has an inherent technical component, its aesthetic aspect is difficult to measure; there's not a precise "good" or "bad" way of doing it (specially taking into account that most of the users were non-experts). Because of this, users were asked to work as long as they want with both interfaces and try to achieve what they considered to be equally satisfactory results. The time required to finish the task was measured for both systems. A questionnaire with a Likert scale of 1 to 5 was applied afterwards to evaluate the interface, and users were asked to write their comments and observations. A video was shot during the tests, and users were also asked to sign a permission to use it in the context of this project.

4.1 Results

The times measured during the tests are shown in Table 1. The first column shows the times required by every user for completing the mix with the Yamaha 01V96, and the second column the times required with the proposed multitouch interface.

Table 1: Times used for mixing

User	Mixer	Multitouch
1	9:57	4:10
2	7:43	3:24
3	5:36	3:21
4	6:40	4:18
5	4:40	3:23
6	3:07	2:38
Average	5:33	3:25

5. DISCUSSION

An analysis of the comments written by the users showed preference for the multitouch interface over the mixer. Also, the multitouch interface seemed to be more time-efficient. For non-expert users, experimental results and user comments suggest that the multitouch interface was easier to learn. Interestingly, some users felt that it encourages creativity and playing more than the standard mixer. This might suggests further development for audio mixing education, especially in surround environments.

One user commented that the multitouch interface, more than the mixer, encouraged the use of both hands. At least two users were observed using both hands with the proposed interface and just one with the mixer. This might be due to the physical distribution of the mixer in the location, but it is an interesting point that should be further investigated.

Interestingly enough, the only expert user (first one in Table 1) had the longest time for the mixer and the second longest time for the multitouch interface. The user explained it saying that he took the time to make a very polished mix with both interfaces. This was not the case for other users who said to have felt a bit overwhelmed by some controls and in some cases opted for ignoring them.

We think that many considerations should be supported on expert users experience. Mixing is a task which involves technical and artistic components, both equally important. An expert should feel comfortable with the tool he or she is using in order to do a good job. Also, some studies suggest that the aesthetical features of an interface can affect its usability, and thus its perceived performance [20, 13]. So if a specific tool is deeply established in certain work context, a new one which offers not only a different set of functions, but also a different interface and aesthetic appeal will be initially hard to accept. However, after showing the paper prototype and explaining the goals of this project to a well-known and experienced sound engineer, it was found that some of our concerns might have also appeared in the professional audio context. Projects such as the Reactable and others, many of them from the NIME conferences, have called for the attention of professional musicians towards new technologies. We believe that sound engineers might share the same interest.

6. CONCLUSIONS AND FUTURE WORK

We presented the prototype for a novel multitouch mixing control interface, supported with a preliminary evaluation by means of user tests and questionnaires.

The emphasis on spatial control, relating mixing parameters with physical position of the input channels is one of the strong points of the interface. The "channel+stage" approach seemed intuitive for novice users, and offered a true metaphor-based interface as opposed to traditional mixers.

The literature and the test results suggest great possibilities for interfaces like the one suggested. The biggest drawback of the proposed system, as in any touchscreen, is the lack of tactile feedback. Traditional mixing consoles, with lots of physical controls, have a great advantage in this aspect. Hopefully a fully developed prototype, a better implementation, and good demonstration strategies would make the proposed approach more competitive.

6.1 Future Work

The shape and size of the Reactable are not ideal for the context of professional audio mixing, so it would be desirable to port our prototype to a multitouch platform with a smaller size, a higher graphical resolution, and a rectangular surface, such as a tablet computer.

Some of the findings gathered during the paper prototype stage have not been implemented yet, and some of them, specially the ones coming from expert users, are crucial. Additional functionalities, such as the possibility to control dynamic processors and external plug-ins, would add value to the package. In general, a competitive prototype should allow the user to do anything he could do with a hardware mixer, in order to perform a fair comparison. This was out of the scope of this project, but would be highly desirable for a final implementation.

Last but not least, more statistically significant tests are yet to be done. These should be done after further refining the prototype, and involving more expert users. However, so far, the results are promising.

A comparison with the software mixer of a digital audio workstation such as Pro Tools was planned but not done due to technical and time constraints. It would be good to include this option in further tests.

7. REFERENCES

[1] T. W. Abhayapala and D. B. Ward. Theory and design of high order sound field microphones using spherical microphone array. In *Proceedings of ICASSP*, 2002.

[2] Apple. *Introduction to Quartz Composer User Guide.* Apple, Inc.

[3] G. Balou. *Handbook for Sound Engineers. The new audio cyclopedia.* Howard W. Sams and Company, 1 edition, 1987.

[4] W. Buxton. Lexical and pragmatic considerations of input structures. *SIGGRAPH Comput. Graph.*, 17:31–37, January 1983.

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

[6] M. Duignan, J. Noble, P. Barr, and R. Biddle. Metaphors for electronic music production in reason and live. In *6th Asia-Pacific Conference on Computer-Human Interaction*, 2004.

[7] D. P. Gibson. *The Art of Mixing.* ArtistPro Press, 1997.

[8] A. Hunt, M. M. Wanderley, and R. Kirk. Towards a model for instrumental mapping in expert musical interaction. In *San Francisco: International Computer Music Conference*, 2000.

[9] S. Jordà. New musical interfaces and new music-making paradigms. In *Proceedings of the 2001 conference on New interfaces for musical expression*, NIME '01, pages 1–5, Singapore, Singapore, 2001. National University of Singapore.

[10] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner. The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, TEI '07, pages 139–146, New York, NY, USA, 2007. ACM.

[11] G. Levin. Painterly interfaces for audiovisual performance. Master's thesis, Massachusetts Institute of Technology, 1994.

[12] N. Liebman, M. Nagara, J. Spiewla, and E. Zolkosky. Cuebert: A new mixing board concept for musical theatre. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression*, 2010.

[13] S. R. McDougall and I. Reppa. Why do i like it? the relationships between icon characteristics, user performance and aesthetic appeal. In *Proceedings of the Human Factors and Ergonomics Society 52th annual meeting.*, 2008.

[14] N. Peters, T. Matthews, J. Braasch, and S. McAdams. Spatial sound rendering in max/msp with vimic. In *Proceedings of the 2008 International Computer Music Conference*, 2008.

[15] M. Puckette. Pure data. In *Proceedings of the International Computer Music Conference, (ICMC).*, 1996.

[16] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. In *The journal of the Audio Engineering Society*, volume 45, 1997.

[17] M. Rettig. Prototyping for tiny finger. *Communications of the ACM*, 37(4):21–27, 1994.

[18] C. Roberts. Multi-touch, consumers and developers. Technical report, Media Arts and Technology Program - University of California, 2008.

[19] J. K. Thompson. The allobrain: An interactive, stereographic, 3d audio, immersive virtual world. In *International Journal of Human-Computer Studies*, volume 67, 2009.

[20] N. Tractinsky, A. S. Katz, and D. Ikar. What is beautiful is usable. *Interacting with Computers*, 13(2):127 – 145, 2000.

[21] M. Wright and A. Freed. Open sound control: A new protocol for communicating with sound synthesizers. In *Proceedings of the International Computer Music Conference*, 1997.

[22] Yamaha. *Manual of the Yamaha 01V96 digital mixing console.* Yamaha Corporation.