MIXPLORATION: Rethinking the Audio Mixer Interface

Mark Cartwright
Northwestern University
mcartwright@u.northwestern.edu

Bryan Pardo
Northwestern University
pardo@northwestern.edu

Joshua D. Reiss
Queen Mary University of London
josh.reiss@eecs.qmul.ac.uk

ABSTRACT
A typical audio mixer interface consists of faders and knobs that control the amplitude level as well as processing (e.g. equalization, compression and reverberation) parameters of individual tracks. This interface, while widely used and effective for optimizing a mix, may not be the best interface to facilitate exploration of different mixing options. In this work, we rethink the mixer interface, describing an alternative interface for exploring the space of possible mixes of four audio tracks. In a user study with 24 participants, we compared the effectiveness of this interface to the traditional paradigm for exploring alternative mixes. In the study, users responded that the proposed alternative interface facilitated exploration and that they considered the process of rating mixes to be beneficial.

Author Keywords
Audio; music; mixing; exploratory interfaces

ACM Classification Keywords
H.5.2. User Interfaces: Interaction styles; H.5.5. Sound and Music Computing: Systems

INTRODUCTION
Mixing refers to processing and combining multiple audio recordings (tracks) together into a single recording (the mix). Mixing is an integral part of how modern video and music production is done, where it is common to combine dozens of tracks into a single final mix.

In its most basic form, mixing consists of applying gain (a change in volume) to each track and summing all tracks together into the mix. Existing mixing interfaces in widespread use all start from the same underlying paradigm: the interface should provide one controller (fader) per track and this should control the gain applied to that track. Figure 1 shows an example of a typical mixing interface, whose design emulates existing hardware mixing boards.

If we think of each track as an independent dimension, and the gain of each track as the relevant feature, a mix of N tracks with a static gain for each track can be described as a point in an N-dimensional vector space. Similarly, a mix with varying gain on one or more tracks traces a path through a vector space. For simplicity we will assume static-gain mixes, where we are setting the rough volume levels for a set of N tracks.

Figure 1. The fader view of a mix in ProTools, a typical mixing interface.

Given this paradigm, we can now consider how one explores this N-dimensional space using the conventional N-fader approach. Typically, the user will set the faders to an initial position of roughly equal gain and then move one fader at a time to improve the mix. This is a form of N-dimensional hill-climbing where only a single dimension is varied at any one time. This is illustrated in Figure 2. Note that grouping tracks to be controlled by a single fader just changes this walk to allow diagonal travel at a fixed angle.

One common issue with hill-climbing approaches to optimizing a set of parameters is getting stuck in a local maximum which is not the global maximum. In search algorithms, this problem is typically ameliorated through multiple random restarts. Since mixing takes significant time, people do not take this approach. Instead, they either trust to luck or to the experience of a good mixing engineer (if they can afford one and the project allows for this) to ensure that they mix to at least a local optimum. This approach may miss artistically satisfying alternatives, since they may lie outside the space falling within the local maximum's basin of attraction or the mixing engineers prior experience.

In this work we rethink the interaction paradigm for mixing to facilitate the discovery of diverse, high-quality rough mixes. We define a "high-quality rough mix" as one whose gain and equalization parameters are set approximately correctly to achieve a pleasing sound, though there might still be some fine-tuning to do. Therefore, we seek an interface that facilitates high-level exploration of the mixing space (see Figure 2) so the user can quickly and easily reach places in the

IUI'14, February 24–27, 2014, Haifa, Israel.

http://dx.doi.org/10.1145/2557500.2557530

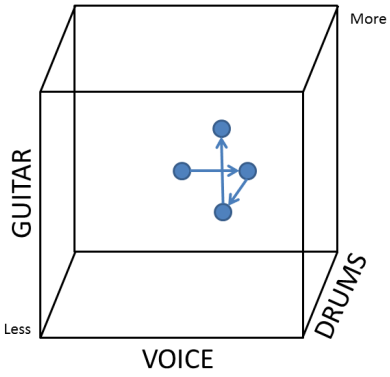

Figure 2. The 3-dimensional space representation of a three-track mixing session using faders. First, gain on the voice is raised, then drum gain is decreased, then guitar gain is raised.

mixing space they would be unlikely to reach with the conventional interface.

In addition to *facilitating broad exploration*, we designed our interface with two other goals in mind. We wanted users to *trust their ears* by listening to the whole mix, rather than trusting their eyes by focusing on individual parameter settings reported by a level meter or knob. We also wanted to make *explicit evaluation of alternatives* an integral and easy part of the process. We argue that an interface that supports exploration, evaluation and trusting one's ears will enhance and support the creativity of artists involved in mixing.

RELATED WORK

In the current digital audio workstation marketplace, products like Apples GarageBand may simplify the user interface, but the essential mixing paradigm is the same "hill-climbing" procedure as described earlier. One approach to simplify the mixing problem is to fully automate the mixing process [13, 14, 15, 20, 9, 2]. This however takes the user completely out of the loop, removing any creative input by the user, and assumes there is one ideal mix rather than multiple solutions to the problem. However, [7] shows that there may be multiple preferred mixes for a given piece.

Researchers in recent years have attempted to eliminate slider/knob-based interfaces for music and audio production by using machine learning and optimization to map gestures [3], examples [4, 11, 5, 6, 22], and language [16, 17] to control spaces. While these interfaces potentially allow users to explore parameter spaces without the distraction of a slider/knob-based interface, it is not clear that these interfaces would translate well to the mixing task.

Interfaces such as the Tenori-On [12] and TC-11 [19] provided inspiration for the two-dimensional interface in this work, but these interfaces are typically used for sequencing and synthesis rather than mixing, and there is no inherent tie-in to evaluation of what is being produced. There have been previous interfaces for controlling equalizers (a component of mixing) with a 2D space [18, 10]. Both of these interfaces mapped a high dimensional space down to a two-dimensional, square interface, but both of these interfaces were concerned with only the equalization of a single audio object, not mixing multiple objects. Further, their focus was on mapping descriptive terms (e.g. a "warm" sound) onto equalization, not facilitating exploration of mixing options.

A NEW MIXING INTERFACE

Our interface is illustrated in Figure 3. To *facilitate broad exploration*, we eliminated the one-dimensional sliders/knobs that one finds in a traditional interface. Instead the user changes the mix by moving a ball around a two-dimensional map, which changes multiple parameters at once. Each point in the map represents some setting of the gain and equalization parameters of all the audio tracks. This map is a two-dimensional reduction of the high-dimensional level and equalization parameter space. The map broadly covers the space of possible mixes, letting the user quickly move to very different points in the mixing space using a single control.

At any point in the map, the user hears the resulting audio, without seeing the individual parameter settings. This is done to encourage the user to *trust their ears* and listen to the whole mix, rather than trusting their eyes and focusing on individual parameter settings.

To encourage *explicit evaluation of alternatives*, the interface incorporates evaluation of mixes directly into the interface by encouraging users to rate each point on the map using a 9-level scale from *dislike* to *like*. The users rate mixes using their keyboard while navigating the map using their mouse. The instructions encourage the user to re-rate mixes as their preferences become more defined. The rating process may help the user to remember preferred mixes, and concretize their preferences. We believe it may also aid the user in transitioning from divergent thinking (exploring the diversity of mixes in the two-dimensional map) to convergent thinking (concretizing a specific mix idea).

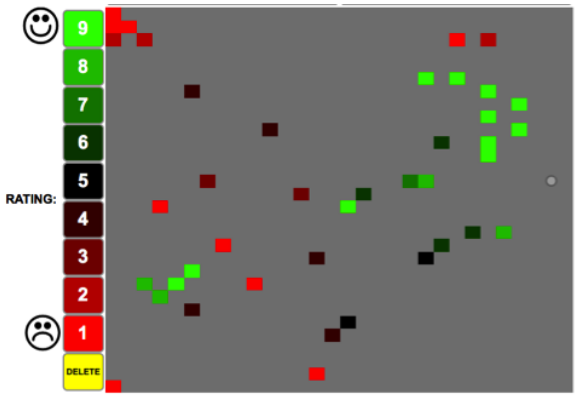

Figure 3. The proposed mixing interface.

The 2-dimensional map

When creating a 2-dimensional map, we wanted the topology of our 2-dimensional space to be similar to the topology of the original parameter space, i.e. the relative distance between points is similar in both spaces. We did not want the parameter values to jump dramatically and seem random. Instead we wanted the user to feel as if they were in control of their navigation through the space.

For the experiment in this paper, the interface controlled the equalization and gain parameters of 4 audio tracks. We reduced the dimensionality of the equalization parameters down to one parameter, the weighting coefficient for the first principal component of a 40 band graphic equalizer learned from data collected by Cartwright and Pardo [1]. This curve essentially represents a spectral tilt with a pivot point around 630 Hz. It is similar to the first dimension of the equalizer presented in [18]. Therefore, in total there were 8 mix parameters (one gain parameter and one EQ parameter per track).

We then used a self-organizing map (SOM a.k.a Kohonen map) [8] to map these 8 dimensions down to 2 dimensions (a 30x30 grid). The inputs to the SOM were 10,000 8-dimensional vectors sampled randomly from a 6-level quantized space. We used an initial neighborhood of 7 and allowed 400 iterations. Since our training examples were sampled from a uniform distribution, the goal of the SOM was not to learn a manifold as is typical with an SOM but rather to create a coarsely sampled but smooth map. While this map only contains 900 points from a much larger space, we think that such a broad, coarse sampling encourages the user to explore a wide variety of mixes before becoming focused on fine tuning one potential mix.

Refining the mix

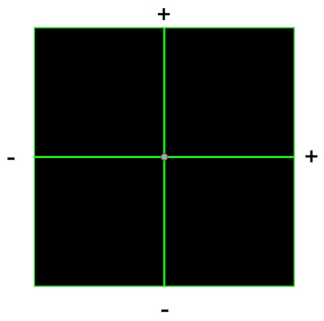

Figure 4. The refinement controller. The x and y axes respectively map to gain and equalization.

Having such a coarse map of the space encourages high-level exploration, but it does not allow for fine-tuning of the mix. Therefore, once a user rates several mixes and picks one *favorite mix*, the machine uses the ratings the user provided, along with their corresponding mix parameters (i.e. the 8-dimensional vector) to learn a weighting function of what the user finds important. This approach is similar to the equalization learning approach taken in [17]. For each individual control parameter, we perform a separate least-squares linear regression between the mix parameter values and the user's ratings of the mixes (i.e. find the least squares fit of α_i and β_i to the model $\mathbf{y} = \alpha_i + \beta_i \mathbf{x_i}$ where $\mathbf{y}$ is the vector of mix ratings and $\mathbf{x_i}$ is the vector of the i^{th} mix parameter's values (e.g. the gain parameters for track 1)). The slope of the resulting line, β_i, is used as the weighting coefficient for the i^{th} mix parameter.

We group the learned coefficients of the gain and equalization parameters into their respective weight vectors, $\boldsymbol{\beta_{gain}}$ and $\boldsymbol{\beta_{EQ}}$. We then provide the user with a 2-dimensional mix refinement controller (see Figure 4), where the x and y axes respectively control ω_{gain} and ω_{EQ} in the following equation:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z_{gain}} \\ \mathbf{z_{EQ}} \end{bmatrix} = \begin{bmatrix} \mathbf{x_{gain}} \\ \mathbf{x_{EQ}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\omega_{gain}}\boldsymbol{\beta_{gain}} \\ \boldsymbol{\omega_{EQ}}\boldsymbol{\beta_{EQ}} \end{bmatrix} \quad (1)$$

where $\mathbf{x} = \begin{bmatrix} \mathbf{x_{gain}} \\ \mathbf{x_{EQ}} \end{bmatrix}$ is the mix parameter vector of the user's chosen *favorite mix*, z is the mix parameter vector of the refined mix. Therefore, when the controller is set in the middle ($\omega_{gain} = \omega_{EQ} = 0$), the refiner has no effect, and the the *favorite mix* is left untouched. At other settings, this controller allows the user to refine their favorite mix by increasing/decreasing the gains and EQs the machine believes to be important (depending on the sign of ω_{gain} and ω_{EQ}).

Similar to how navigation through the coarse map could be considered a coarse tuning of the mix, this refinement stage could be considered a medium precision tuning of this mix since it changes multiple parameters at once. True fine-tuning of the mix is not the goal of this interface.

EXPERIMENT

To validate our proposed interface, we compared it against the traditional mixing interface, with a focus on answering the following questions:

1. Which interface facilitates creating more satisfying mixes?
2. Which interface better facilitates exploration of alternative mixes?

The Mixing Interfaces

To evaluate whether the explicit rating of mixes or the refinement controller added value, we evaluated two variants of the *proposed* interface: 1) the complete proposed interfaces with ratings and refinement (*2D rater*) and 2) just the exploratory 2-dimensional map portion of the proposed interface (*2D map*), with no explicit rating of mixes and no refinement controller. We compared them to a traditional mixer interface (*traditional)*, similar to that in Figure 1. The *traditional* interface had three controls for each audio track: an overall gain slider, a low frequency gain knob, and a high frequency gain knob.

Audio Sources

We used one musical excerpt from each of three different genres for the mixing source material: one pop (electro-pop), one rock (shoe-gaze), and one electronic (techno) excerpt. All excerpts were between 17 and 32s long musical sections obtained from [21]. Each excerpt consisted of 4 temporally aligned, stereo "stems" (i.e. subgroup recordings).

We chose these genres to support a variety of mixing styles. Genres like jazz or classical were not included since the engineer may likely strive for a realistic-sounding mix, recreating how it would be heard in a live setting. We instead chose genres which we believe can support a variety of artistic mixes without the constraint of "realism" as an aesthetic .

Participant Pool

Since we seek a fair test for the new interface, we chose not to focus on professional mixing engineers, who would be experts in using the standard paradigm. Instead, we recruited critical listeners without significant mixing experience. We defined critical listeners as either experienced musicians, audio researchers, or music enthusiasts who passed our critical listening pre-test, in which they had to identify small differences in mixes. We believe this population is capable of judging the quality of mixes but does not have years of experience to bias their judgment of interfaces. Participants were recruited through personal contacts.

Mixing Procedure

Each participant took part in one session that lasted about an hour. Sessions were conducted in a quiet room using a laptop computer and high-quality headphones. Each session consisted of two trials: one trial with the *traditional* interface and one trial with one of the proposed interfaces. Half the participants were assigned the *2D rater* interface and half of the participants were assigned the *2D map* interface. The order of presentation of interfaces was random, with half of participants using a proposed interface first and half using the traditional interface first.

Prior to each trial, participants were given a minimum of one minute (no maximum) of training on the interface used in the trial. As there were three musical excerpts and only two trials per participant, each trial used a unique excerpt and the training was conducted on the third excerpt. This prevented learning the details of a musical excerpt on one interface affecting the results for the next interface. Combinations of musical excerpt and interface were balanced across participants so that all excerpt/interface pairs were equally represented.

In a single trial, the participant was presented with one of the three song excerpts (e.g. the "pop" excerpt) and asked to create three diverse, but "good," mixes with the given interface (e.g. three distinct "pop" mixes with the *2D rater*). Participants were given 10 minutes to create each mix.

After each mix, participants completed a survey regarding the diversity, satisfaction, and objectives of their mix. At the end of the entire session, participants were asked additional questions regarding their preferences and experiences working with the interfaces.

For the *traditional* interface, prior to the start of each mix, the settings of the sliders and knobs were all randomized. Similarly, the mapping function for the proposed interfaces was randomly chosen from a set of 24, prior to each mix. The map was randomized to prevent associating particular map locations with particular sounds because we want the users to mix with their ears not their eyes. Controllers on the *traditional* interface were randomized to make things fair.

RESULTS

24 participants performed the experiment. There were two trials per participant and each trial had three mixes. Therefore each participant created 6 mixes. Ideally this would yield 24 mixes for each excerpt/*traditional* pair and 12 mixes for each excerpt/*proposed* pair, for a total of 144 mixes. However, two participants did not finish their mixes in time, reducing the total number of mixes to 142. The participants reported having an average of 169 (SD=148, median=204) months of experience playing music, 23 months (SD=37, median=3) mixing audio, and 56 (SD=63, median=24) months using audio recording equipment. Three participants reported significantly more experience mixing than the others, causing a large difference between median mixing experience (3 months) and mean mixing experience(23 months). However, these participants did not skew the data and therefore were not removed/replaced.

Participant Feedback

To answer our first question ("Which interface facilitates creating more satisfying mixes?"), each participant was asked "How satisfied are you with the quality of this mix?" after completing each mix. Responses were chosen from a seven-level scale: *completely satisfied, mostly satisfied, somewhat satisfied, neither satisfied nor dissatisfied, somewhat dissatisfied, mostly dissatisfied, completely dissatisfied*. The median value for all of the three tested interfaces was the same: *somewhat satisfied*. However, if we perform a Kruskal-Wallis sum-rank test on the group of distributions, we reject the null hypothesis that the location parameters of the distributions are equal ($p=0.0017$). If we then look at the two distributions we are most concerned with, mix satisfaction of *traditional* and the *2D rater*, and perform a one-sided Wilcoxon sum-rank test, we find that the distribution of *traditional*'s mix satisfaction is greater than that of the *2D rater* mix satisfaction ($p=0.0357$). However, while this difference is statistically significant, with the medians the same, it is not discouraging. Recall that the goal of the proposed interface was not to make an interface that supports fine-tuning of mixes, but rather one that facilitates broad exploration of the mixing space. If a user can get close enough with the proposed interface, they can always use the traditional mixer after the *2D rater* in order to fine-tune a novel mix.

At the end of the experiment participants were also asked to complete a survey in which they were asked a number of questions regarding their interface preferences, the physical/mental demands of the interface, etc. The results of these questions are shown in Figure 5. In this survey we directly asked them our second question ("Which interface better facilitates exploration of alternative mixes?"). As shown in the plot, it seems clear that participants think that the proposed interfaces facilitate exploration and that the traditional interface facilitates precise mixing. The results on which interface is less distracting are not clear, but there does seem to be some agreement that the proposed interface is more mentally demanding. The participants were only a bit more inclined to think that the traditional interface is more physical demanding and the proposed interface. Unfortunately, the majority of participants preferred working with the traditional mixer over the proposed mixers, but they gave plenty of feedback as to why, which will help in a future iteration.

From the participants' feedback, it seems that many users found the proposed interfaces great for "exploring possibil-

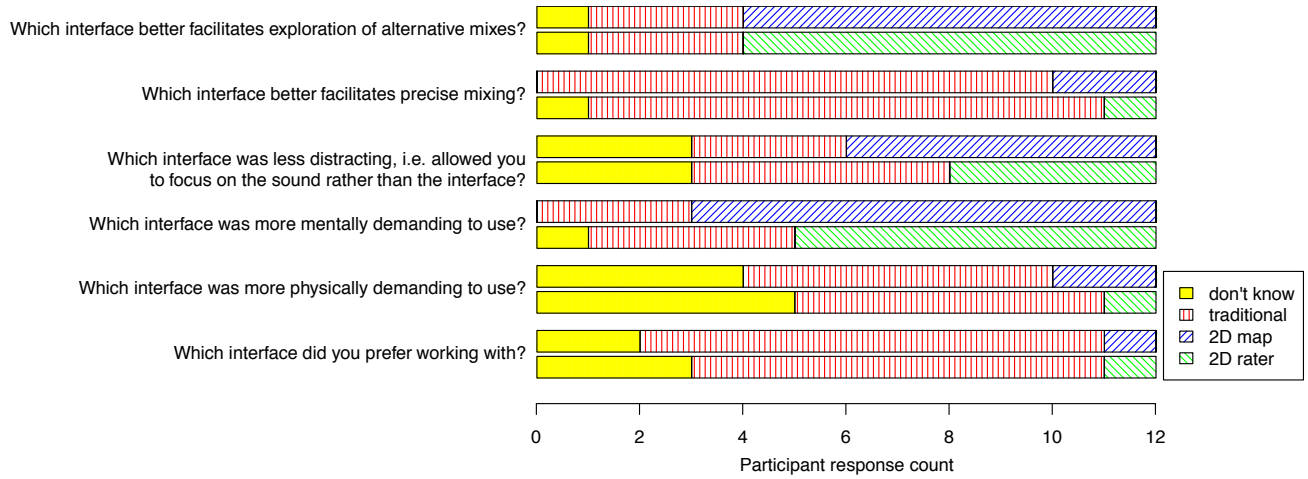

Figure 5. Interface survey response data. The first stacked bar for each question is the response data from participants given *2D map*, and the second is from those given *2D rater*.

ities quickly", "finding ideas I hadn't anticipated", and ideas that they "would not necessarily think of" themselves. However, once they did have an idea in mind, they preferred the traditional interface because they could express their idea in "terms of balance and EQ" and "easily isolate the sounds" to play with. The proposed interface's strength for exploration was its weakness for precision. As one participant put it, "Whenever I found something I liked I, but wanted to tweak one thing, I couldn't find that one tweak." Or as another participant reported, "While at first interface B was really cool in the way that I just wiggle my finger around and get different mixes without actually mixing, I felt frustrated when trying to achieve something specific." This seems to be because many users found the lack of labeled axes on the proposed interface frustrating and caused the mixes to seem random. From this feedback it seems that a combination of the two interfaces may be a good approach for a future iteration. In fact, participants reported this preference as well: "I'd use B (*the 2D rater*) followed by A (*the traditional mixer*) in an ideal mixing situation" reported one participant. The *2D map* and *2D rater* responses were similar, with only a couple of more participants listing *2D map* as more mentally demanding, but less distracting.

Figure 6 shows the results of questions specific to the *2D rater* variant of the proposed interface. It seems that there is little agreement on whether rating mixes helps user concretize their ideas. For instance, one participant reported that "it helped most when my objectives were quite vague to begin with." Whereas other participants reported that their "mixing objectives were not influenced by the ratings" and that "actually rating mixes makes your mixing objectives less clear". There also seemed to be little consensus on how burdensome the rating process was. Some found rating mixes very easy and others didn't. One participant reported, "Rating a mix is difficult when you have no criteria of how to quantify it. It's a difficult task to say I like this, or I don't like this." However, participants did generally agree that rating mixes was useful, especially for the purpose of providing visual feedback in order to remember where and what preferred mixes were. This

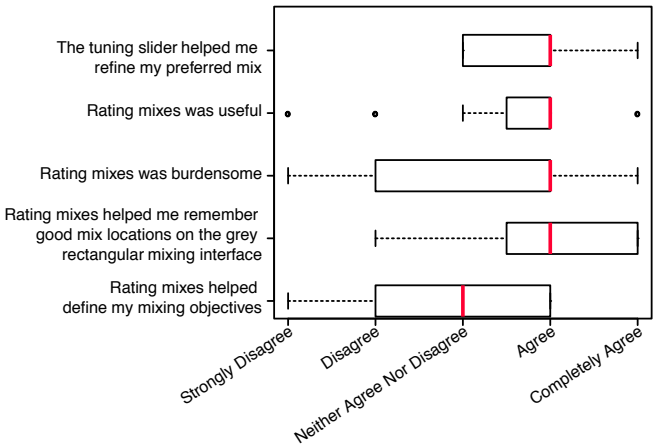

Figure 6. *2D rater*-specific survey response data.

benefit could also be why participants reported the *2D rater* as mentally demanding less often than the *2D map*. Lastly, participants found the refinement controller helpful. One participant reported that it "definitely helped me get closer to something I was happier with." However, participants also reported that they "would have liked more control over separate tracks." In general, the perceived benefit of rating mixes seems to be very participant dependent. Some people finding it useful, others not. This implies that making both the rating stage and the learned refinement stage optional would be a good approach, especially in a hybrid traditional/2D map interface.

DATASET

The mixing data from the study is available for download at http://music.eecs.northwestern.edu/data/mixploration/.

This dataset includes:

1. source audio files
2. mixing parameters of final mixes

3. survey response data of the final mixes
4. time series of mixing parameters during the mixing process

CONCLUSION

This paper presented a new mixing interface that facilitates broad exploration of the mixing space. From the survey response data, it seems that most participants were able to explore the mixing space more easily with the proposed interface than a traditional mixer. Participants found rating mixes to be useful for both remembering good mixes and creating the refinement controller. This let them to get closer to their preferred mix. While the refinement controller helped, participants generally preferred the traditional interface, especially for precision mixing due to its clearly defined dimensions. Participant feedback indicates that in a future iteration we should combine the two interfaces and provide more guidance and feedback when navigating the 2-dimensional map.

This work was supported by National Science Foundation Grant Nos. IIS-1116384 and DGE-0824162 and EPSRC grant EP/K007491/1, "Multisource audio-visual production from user-generated content."

REFERENCES

1. Cartwright, M., and Pardo, B. Social-eq: Crowdsourcing an equalizaiton descriptor map. In *Proc. of International Society for Music Information Retrieval* (Curitiba, Brazil, 2013).
2. De Man, B., and Reiss, J. D. A knowledge-engineered autonomous mixing system. In *Audio Engineering Society Convention 135*, Audio Engineering Society (Year).
3. Fiebrink, R. A. *Real-time human interaction with supervised learning algorithms for music composition and performance*. PhD thesis, Princeton Univ., 2011.
4. Garcia, R. Growing sound synthesizers using evolutionary methods. In *Proc. of ALMMA 2002 Workshop on Artificial Life Models for Musical Applications* (Cosenza, Italy, 2001).
5. Heise, S., Hlatky, M., and Loviscach, J. Automatic adjustment of off-the-shelf reverberation effects. In *Proc. of Audio Engineering Society Convention 126* (2009).
6. Heise, S., Hlatky, M., and Loviscach, J. Automatic cloning of recorded sounds by software synthesizers. In *Proc. of Audio Engineering Society Convention 127* (2009).
7. King, R., Leonard, B., and Sikora, G. Consistency of balance preferences in three musical genres. In *Proc. of Audio Engineering Society Convention 133* (2012).
8. Kohonen, T. The self-organizing map. *Proceedings of the IEEE 78*, 9 (1990), 1464–1480.
9. Mansbridge, S., Finn, S., and Reiss, J. D. Implementation and evaluation of autonomous multi-track fader control. In *Proc. of Audio Engineering Society Convention 132*, Audio Engineering Society (2012).
10. Mecklenburg, S., and Loviscach, J. subjeqt: controlling an equalizer through subjective terms. In *Proc. of CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montreal, Canada, 2006).
11. Mintz, D. Toward timbral synthesis: a new method for synthesizing sound based on timbre description schemes. Master's thesis, University of California, 2007.
12. Nishibori, Y., and Iwai, T. Tenori-on. In *Proc. of New Interfaces for Musical Expression*, IRCAM (Paris, France, 2006), 172–175.
13. Perez-Gonzalez, E., and Reiss, J. Automatic equalization of multichannel audio using cross-adaptive methods. In *Proc. of Audio Engineering Society Convention 127*, Audio Engineering Society (2009).
14. Perez-Gonzalez, E., and Reiss, J. Automatic gain and fader control for live mixing. In *Proc. of Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*, IEEE (2009), 1–4.
15. Perez-Gonzalez, E., and Reiss, J. Automatic mixing. In *DAFX : digital audio effects*, U. Zlzer, Ed., 2nd ed. Wiley, Chichester, U.K., 2011.
16. Reed, D. A perceptual assistant to do sound equalization. In *Proc. of the 5th international conference on Intelligent user interfaces*, ACM (2000), 212–218.
17. Sabin, A., Rafii, Z., and Pardo, B. Weighting-function-based rapid mapping of descriptors to audio processing parameters. *Journal of the Audio Engineering Society 59*, 6 (2011), 419–430.
18. Sabin, A. T., and Pardo, B. 2deq: an intuitive audio equalizer. In *Proc. of ACM Creativity and Cognition*, ACM (Berkeley, USA, 2009).
19. Schlei, K. Tc-11: A programmable multi-touch synthesizer for the ipad. In *Proc. of New Interfaces for Musical Expression* (Ann Arbor, USA, 2012).
20. Scott, J., Prockup, M., Schmidt, E., and Kim, Y. Automatic multi-track mixing using linear dynamical systems. In *Proc. of Sound and Music Computing* (Padova, Italy, 2011).
21. Senior, M. The 'mixing secrets' free multitrack download library. `http://www.cambridge-mt.com/ms-mtk.htm`.
22. Yee-King, M. J. *Automatic sound synthesizer programming: techniques and applications*. PhD thesis, Univ. of Sussex, 2011.