# Personalized Song Interaction Using a Multi-Touch Interface

Jeffrey Scott, Justin Chapman†, Mickey Moorhead,
Ryan Schwabe†, Youngmoo E. Kim
Music and Entertainment Technology Laboratory (MET-lab)
Electrical and Computer Engineering
Music Industry Department†
Drexel University
{jjscott, jtc46, mmoorhead, rjs22, ykim}@drexel.edu

## ABSTRACT

Digital music technology is a catalyst for transforming the way people listen to music and creates new avenues for creative interaction and expression within the musical domain. The barrier to music creation, distribution and collaboration has been reduced, leading to entirely new ecosystems of musical experience. Software editing tools such as digital audio workstations (DAW) allow nearly limitless manipulation of source audio into new sonic elements and textures and have promoted a culture of recycling and repurposing of content via mashups and remixes. We present a multi-touch application that allows a user to customize their listening experience by blending various versions of a song in real time.

## Keywords

Multi-track, Multi-touch, Mobile devices, Interactive media

## 1. INTRODUCTION

Humans are becoming more exposed to hyper-customization of the experiences we have through the use of technology. From the games we play to the television shows we watch to the food we eat, customization is becoming more prominent and desirable in our everyday lives. This trend is particularly noticeable in the media we interact with and consume. Music videos are generated based on personal information and dynamic music in video games responds to the actions of players. The desire for more control and customization over media content has led to an expansion in the modes of interaction we have with music and audio.

While the digital domain resulted in entirely new channels for acquiring music, the expanding popularity of mobile devices has revolutionized the way we consume music on a more personal level. The amount of storage and processing power in mobile phones opens up the possibility to dynamically alter media content at the user level in real-time.

We introduce an application that puts control over the dynamics and evolution of a song in the hands of the listener. The interface enables the user to blend a series of versions of a single song into a personalized mixture, instantaneously updating the properties of the mixture as they listen. The result is that the listener explores the song over time and the parameter space we provide them, discovering nuances and interactions in the material that is not evident during an initial listening. By tracing a path on the touch screen, the characteristics of the song change from Rock to Acoustic to Electronic to Ambient. The application operates in a single-track mode where a single cursor blends the versions of the song and a multi-track mode where the listener has control over each instrument layer individually.

Several songs were created for the application using this four genre format. Although all of the songs follow the same template, this is certainly not a necessity and every artist or producer could parameterize the space as they see fit.

In the following section, we present background information on the technology and similar interfaces that led to the development of the MultiMixer. Section 3 describes the functionality and development of the application as well as the content generation process, followed by a discussion of the interface in Section 4.


*NIME'13,* May 27 – 30, 2013, KAIST, Daejeon, Korea.


## 2. BACKGROUND

The value of the mobile device for creating and interacting with music was explored in detail by Essl *et. al.* [2]. The wide variety of input sensors (accelerometer, gyroscope, proximity sensor, GPS, etc.) accelerated the development of increasingly interactive mobile music technologies. Multi-touch capability was specifically mentioned as one of the most significant drivers of innovation.

Savage *et. al.* report an interactive DJ interface for engaging a live audience [9]. Their application enables users to mix and apply effects to music based on simple gestures to help DJs maintain focus on their audience rather than on technical concerns.

Carrascal *et. al.* present a multi-touch mixing control surface deployed on the Reactable tabletop user interface [1, 4]. Their interface eschews the traditional mixing paradigm of knobs and sliders that is common in digital mixing applications to create *stages* that use spatial location in a parameter space to modify panning and auxiliary effects. The interface communicates via OSC to a digital hardware mixer that applies the proper settings to each audio channel. Each channel (stereo or surround) can be positioned freely in a two dimensional space and the user can select a listener position, easily creating the desired spatial parameterization for each track.

Forsyth *et. al.* introduce a remix application that leverages a dual time axis representation for manipulating audio [3]. This architecture allows a user to place simple shapes (squares and rectangles) to select regions of audio for playback. The dimensions of the shape simultaneously time-stretch or compress the length of the sample that is selected

via the dual time axis representation.

A recent trend is for musicians to release tracks or albums as applications for mobile devices that provide some means of interaction with the material through audio manipulation or video accompaniment. One of the first such releases is The Polyphonic Spree song *Bullseye*, which follows an animated character through a short narrative. Although the audio does not change, the user's actions effect the visuals on the screen. An early implementation of an album application is the Bjork release *Biophilia*, which comes with a separate visualization or simple game for each song. Passion Pit released two tracks with their *Gossamer* application, which provides a visualization for each track as well as a version that allows users to remix the song by activating various loops created from the source material.

There have also been several remix applications released for mobile platforms by companies such as Romplr and Future Audio Workshop for artists such as Soulja Boy and Deadmau5. An application by RjDj, the Rj Voyager, incorporates real-time effects with loops that can be manipulated in a 2D parameter space. Artists can create scenes using an accompanying software package to deliver a song in a format that Rj Voyager can read.

These applications form two general categories, those that are an accompaniment to the song and those where the user is required to generate the audio themselves using content from the song. Our application bridges the gap between these two classes. It is not a true remix application where the user is required to start from nothing and build a song with segments or loops from the artist content. It is also not a passive accompaniment where the song plays basically unchanged while the user interacts with some visual representation.

## 3. IMPLEMENTATION

The MultiMixer application presented here is designed to give a listener control over the evolution of a song. We present two versions, one where the user traverses the space via a single cursor and one where we take advantage of the multi-touch capabilities of the device to expose the multi-track nature of the recording to the user. We develop the application for deployment on Apple iOS devices, a demonstration of the application is available online[1].

### 3.1 Single-Track MultiMixer

By navigating a cursor throughout a continuous two dimensional space, the user seamlessly blends the various versions of the song. In this particular realization, four different versions of a song are created and associated with a corner of the application. Each colored region is associated with a specific instrumentation as well as a characteristic production technique. Figure 1 depicts the layout of the interface as well as the organization of the songs within the space.

The gain for each track is determined by the position in the space

$$g_k = 1 - d(p_k, c). \quad (1)$$

Here, $p_k$ is the point of origin for the $k$th version, $c$ is the location of the cursor and $d(\cdot)$ is the euclidean distance in the $xy$ plane created by the multi-touch surface. The coordinate space is normalized to be unit length on each side prior to calculating the gain values. This results in an attenuation of approximately $-6dB$ at the midpoint of the edges and an attenuation of about $-10.6dB$ at the center of the space for each track. At any given corner, the gain of the song associated with that corner is unity and the gain of all other tracks is $-\infty dB$.

[1]http://music.ece.drexel.edu/research/StructuredAudio

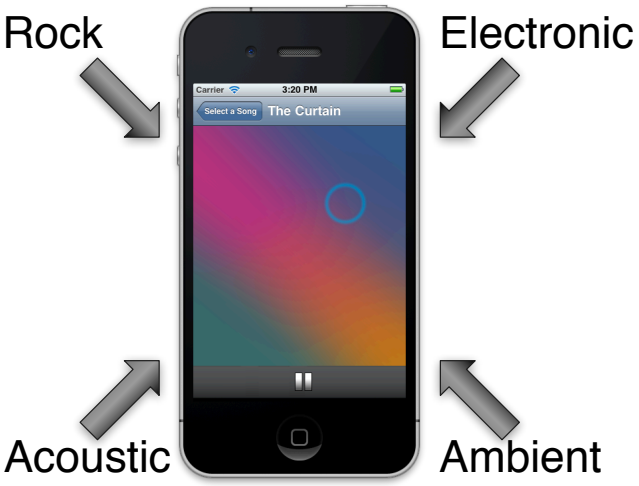


**Figure 1: Screenshot of the application interface. Each corner (color) is associated with a different version of the song. Moving the cursor thoughout the space determines the blend between the various audio streams.**

Each version plays continuously in a loop and has the gain value applied to it in real-time. It is essential that the content is time aligned for the application to function properly. Given that all of the audio must create a coherent mixture at all times in the playback phase, the content must be generated specifically for the application. We now discuss the process of creating and deploying content in the MultiMixer.

### 3.2 Creating Content

The content streams are created in full prior to being packaged in the application, therefore careful consideration must be made regarding the interaction between the various versions and mixes of the individual songs. The producer/mixing engineer and artist involved in developing the songs for the application work very closely to ensure that any combination of tracks associated with each version will fit well together when played back simultaneously.

Another consideration in song selection and recording is that the structure and harmony must be malleable enough to be adapted into four distinct genres with different instrumentations and production techniques employed in recording and post-processing.

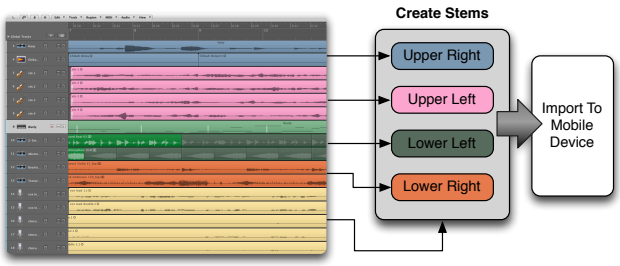


**Figure 2: Workflow process for creating the tracks to upload on the mobile device. Some instruments (yellow) are reused in all versions utilizing different processing effects. The final result is a set of stems for each version.**

| Genre/Style | Instrumentation | Techniques |
|---|---|---|
| Rock | Bass, Drums, Vocals, Keyboards<br>Strings, Elec/Acoustic Gtr | Parallel Drum Compression |
| Electronic | Synth Bass, Drum Machine, Percussion<br>Vocals, Keyboards, Strings | Glitch, Sidechain, Vocal Distortion<br>Panning Sweeps, Heavy Reverb/Delay/Filtering |
| Ambient | Piano, Strings, Vocals<br>Synth Pads | More Reverb, Light Compression |
| Acoustic | Acoustic Guitar, Vocals<br>Percussion | Dryer Vocals, Light Compression |

**Table 1: Characteristic instrumentations and production techniques for the different versions of the songs in the MultiMixer.**

There are three songs recorded for the MultiMixer (*Ghost Party, The Curtian* and *Tired Eyes*) each tracked and produced specifically for the interface. The four versions of each song utilized a variety of instrumentations as well as different processing techniques.

- **Version 1** (Rock) - This is the main version of the song and features mostly live instruments and an indie/rock feel.
- **Version 2** (Electronic) - A version containing mostly electronic and synthesized instruments.
- **Version 3** (Acoustic) - The instrumentation is sparse with acoustic guitar, vocals and a few select textures from the other versions.
- **Version 4** (Ambient) - Another sparse version but with a more airy, ethereal feel relying on ambient pads and piano.

Where possible, the same layers are used in multiple versions of the song, and once each version is finalized, the proper tracks are exported to create stem files for each representation. Table 1 shows the general trends in creating the versions of each song.

Although the tree songs created for the application adhere to the same template, this is by no means how the application must be used. This was a decision by the artist and producer to create the songs to span the four genres outlined above.

## 3.3 MultiMixer

The interface presented in Section 3.1 uses a single cursor to explore the space. We also offer an extended framework which exposes the multi-track representation to the user. In this realization, separate stem files are exported for each instrument class: bass, drums, vocals and guitar/piano. The guitar and piano tracks are associated with the same cursor since they are both instruments which generally occupy the mid-range of the frequency spectrum. Since not all versions of the song have guitar and not all versions have piano, these instruments were combined to avoid having a cursor that doesn't function throughout the entire space.

This alternate representation allows a user to manipulate the blend of each version on an instrument class level. It is possible to use the drums from the Electronic version and the bass from the Rock version or to use a blend between two or more versions. Whereas this mode allows higher customization of the user experience in terms of instrument and version specific transformations in real-time, the increased parameterization makes broad sweeping changes (i.e. switching between full versions of the song) more difficult to perform using a multi-touch gesture.

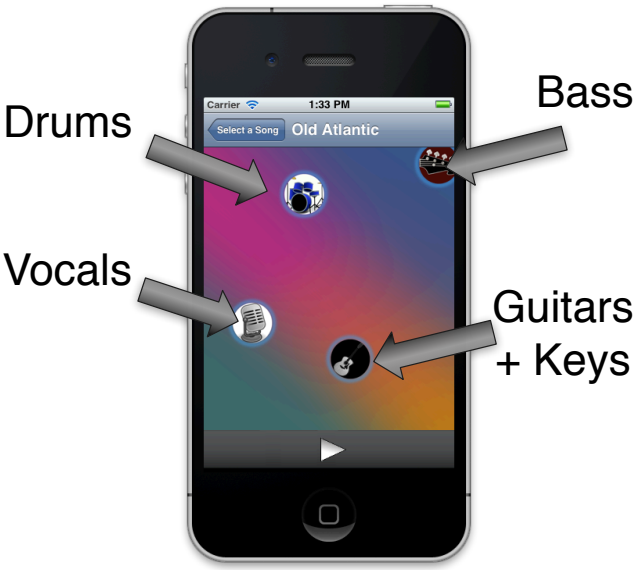


**Figure 3: Screenshot of the multi-track version of the MultiMixer application interface.**

Figure 3 depicts an example of the multi-track version of the MultiMixer application. The configuration shows the user selecting the bass from the Electronic version, a blend of mostly Rock/Electronic for the drum track, a blend of Acoustic/Rock for the vocals and predominantly Ambient/Acoustic for the remaining layer.

# 4. DISCUSSION

The format of the multi-track MultiMixer presented in the previous section opens up a wide range of possibilities to integrate intelligent analysis of audio into the way we listen to music. Since we are now working with a track/stem representation of each version of the song, it would be possible to perform some of the processing effects on the device. This would be ideal for extending automatic mixing techniques, using a learned or pre-defined model for each genre/style used in the space [11, 5, 6, 7]. It would be useful to provide a framework where the user could interact with the audio via higher level representations such as timbre or emotion rather than low level, less user-friendly signal processing constructs [10, 12, 8].

Such systems may provide the ability to upload unmixed stems or multi-track files to the application and automatically perform the necessary processing to correctly parameterize the space according to the desired genres/styles. This

would be a step toward an interactive structured audio representation using high level information about genre, style, instrumentation and production techniques to not only reduce the information necessary to represent the song but also provide a versatile parameterized model that can be easily tailored for a desired level of interaction [13].

## 5. ACKNOWLEDGMENTS

Special thanks to James Larson of The Old Atlantic for the content used in the MultiMixer application.

## 6. REFERENCES

[1] J. P. Carrascal and S. Jordà. Multitouch Interface for Audio Mixing. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 100–103, 2011.

[2] G. Essl and M. Rohs. Interactivity for mobile music-making. *Organised Sound*, 14:197–207, 2009.

[3] J. Forsyth, A. Glennon, and J. P. Bello. Random Access Remixing on the iPad. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 487–490, 2011.

[4] S. Jorda, M. Kaltenbrunner, G. Geiger, and R. Bencina. The reactable*. In *Proceedings of the international computer music conference (ICMC), Barcelona, Spain*, pages 579–582, 2005.

[5] S. Mansbridge, S. Finn, and J. D. Reiss. An Autonomous System for Multitrack Stereo Pan Positioning. In *AES 133rd Convention*, 2012.

[6] E. Perez-Gonzalez and J. D. Reiss. *DAFX: digital audio effects, second edition*. Wiley Online Library, 2 edition, 2011.

[7] J. D. Reiss. Intelligent systems for mixing multichannel audio. In *17th International Conference on Digital Signal Processing*, pages 1–6, July 2011.

[8] A. T. Sabin and B. Pardo. 2DEQ: An intuitive audio equalizer. In *Proceedings of the 7th ACM Conference on Creativity and Cognition*, pages 435–436, New York, NY, USA, 2009.

[9] N. S. Savage, S. R. Ali, and N. E. Chavez. Mmmmm : A Multi-modal Mobile Music Mixer. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 395–398, 2010.

[10] E. M. Schmidt and Y. E. Kim. Modeling musical emotion dynamics with conditional random fields. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, Miami, Florida, 2011.

[11] J. Scott, M. Prockup, E. M. Schmidt, and Y. E. Kim. Automatic multi-track mixing using linear dynamical systems. In *Proceedings of the 8th Sound and Music Computing Conference*, Padova, Italy, 2011.

[12] J. Scott, E. M. Schmidt, M. Prockup, B. Morton, and Y. E. Kim. Predicting time-varying musical emotion distributions from multi-track audio. In *Proceedings of the International Symposium on Computer Music Modeling and Retrieval*, London, U.K., 2012.

[13] B. Vercoe, W. Gardner, and E. Scheirer. Structured audio: Creation, transmission, and rendering of parametric sound representations. In *Proceedings of the IEEE*, pages 922–940, 1998.