MadPad: A Crowdsourcing System for Audiovisual Sampling

Nick Kruge
Stanford University
CCRMA
660 Lomita Ct
Stanford, California USA
nkruge@ccrma.stanford.edu

Ge Wang
Stanford University
CCRMA
660 Lomita Ct
Stanford, California USA
ge@ccrma.stanford.edu

ABSTRACT

MadPad is a networked audiovisual sample station for mobile devices. Twelve short video clips are loaded onto the screen in a grid and playback is triggered by tapping anywhere on the clip. This is similar to tapping the pads of an audio sample station, but extends that interaction to add visual sampling. Clips can be shot on-the-fly with a camera-enabled mobile device and loaded into the player instantly, giving the performer an ability to quickly transform his or her surroundings into a sample-based, audiovisual instrument. Samples can also be sourced from an online community in which users can post or download content. The recent ubiquity of multitouch mobile devices and advances in pervasive computing have made this system possible, providing for a vast amount of content only limited by the imagination of the performer and the community. This paper presents the core features of MadPad and the design explorations that inspired them.

Keywords

mobile music, networked music, social music, audiovisual, sampling, user-generated content, crowdsourcing, sample station, iPad, iPhone

1. INTRODUCTION

MadPad is a social music and video creation system currently implemented for the Apple iPad and iPhone. The performer loads twelve independent video clips onto twelve virtual pads (Figure 1) which are laid out in a grid pattern similar to the sixteen pads of an Akai MPC-2000[1] (Figure 2). Upon tapping any pad, the associated audio and video play under the performer's fingertips. Up to eleven can be played simultaneously, and two finger drag gestures can be used to control the playback rate of any individual clip, allowing for expressive control of the content beyond basic re-triggering.

MadPad is a platform that employs the creativity of its users to make it come alive. The application is designed to be a transparent conveyance of user content where the content *is* the instrument. Clips can be created on a camera-enabled mobile device in either rapid succession or one-by-one. In the rapid mode, the performer can simply make

Figure 1: MadPad in action on the Apple iPad. The performer taps on the video clips to play the associated audio and video.

twelve separate sounds and they will be automatically distributed to the twelve slots, while the one-by-one mode allows for a more tailored approach, giving the performer as many takes as necessary to capture each desired sample individually.

Figure 2: The Akai MPC.

With the ease of creating a sample set in under a minute and the addition of video to the traditional MPC-like sampling paradigm, this system intends to give its user the feeling that there is a potential for music all around, and an instrument can be created out of anything in sight. Furthermore, these sample sets serve as a bridge between still

record audiovisual samples

MadPad server

share to central server

discover and **download** new audiovisual sets

play by tapping

empty pad template

full pad
audio visual elements
recorded via mobile device
and/or downloaded from server

individual sample
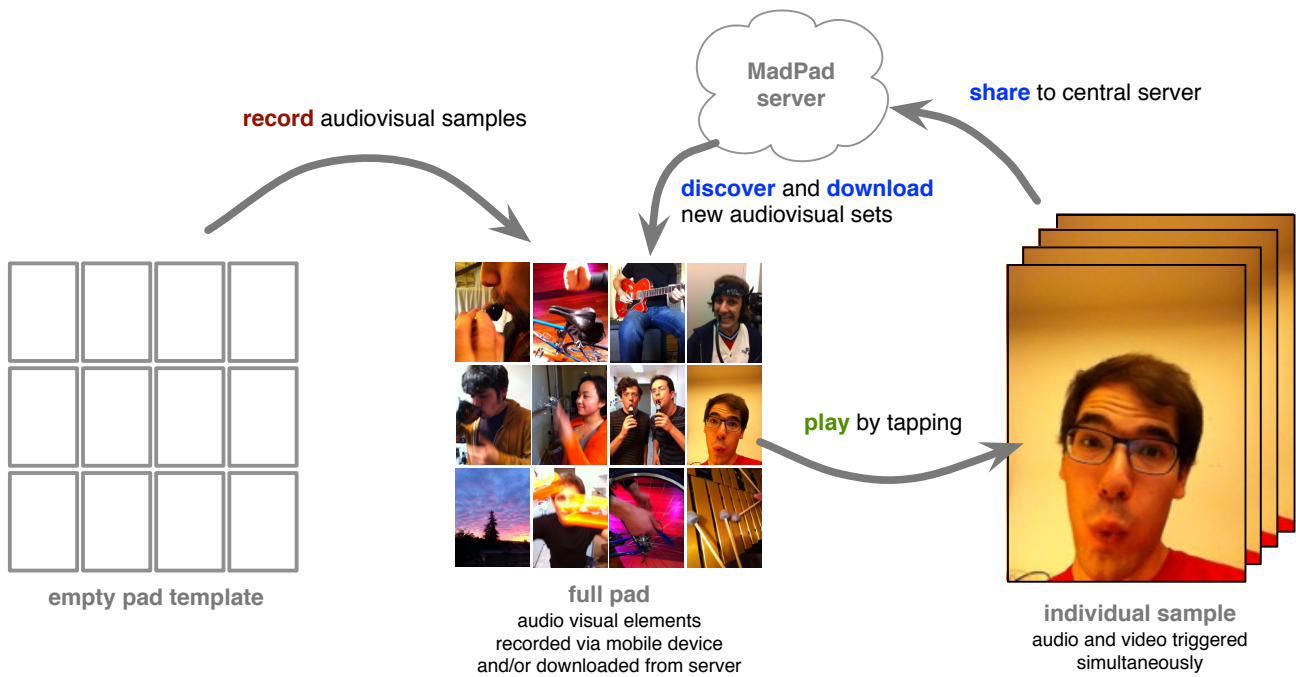audio and video triggered
simultaneously

Figure 3: Overview of the MadPad system.

photography and video, allowing moments to be captured, shared, and relived in a novel and interactive way. Once created, there is an online forum that allows users to share their creations with others, enabling a rich community of user-generated content to develop. All of MadPad's features can be experienced by a performer with a single camera-enabled iPad or iPhone connected to the internet, making the entire experience very portable.

2. RELATED WORK

Figure 4: Frame from a VideoSong: Featuring Pomplamoose and Ben Folds

The proliferation of mobile music technology[7] as well as the increasing number of performance outlets for mobile musicians[14] have set the stage for MadPad. A number of existing works, both academic and artistic, have influenced the design and implementation.

MadPad shares aesthetic similarities with a style of music video called the "VideoSong" (Figure 4), which employs repeatedly triggered video samples in multiple panes for effect. Just as in MadPad, these video samples are literal depictions of the associated audio, and display the actual

recording of the sound the listener is hearing. In 2006, Norwegian artist Lasse Gjertsen released *Amateur*[9], an audiovisual piece that reuses a handful of tightly cut single drum and piano hit videos in what he refers to as his "hyperactive editing style" to create a full song. In 2009, *Pomplamoose*, an indie rock duo from the San Francisco Bay Area, sold roughly 100,000 songs thanks to several viral online videos [16] and coined the term VideoSong. Although Pomplamoose tends to use longer cuts and more varied layouts than Gjertsen or MadPad, the visual aesthetic is similar. Jack Conte of Pomplamoose defines it with two rules: 1.) What you see is what you hear. 2.) If you hear it, at some point you see it[4].

The basic user interaction of MadPad draws from the Akai Music Production Center (Figure 2), commonly referred to by the acronym *MPC*. The main interaction of the MPC uses 16 finger pads to trigger single audio samples when tapped. With this tool, a larger audio clip can be chopped up quickly and distributed to the pads as different subsets of the original sound[1]. Also, individual and potentially unrelated sound clips can be loaded and triggered singularly. These possibilities combined allow for a large number of sonic sample sets to be formed even with just a few seconds of initial material, giving performers a quick way to move from sound clips to expressive, playable instruments[12]. MadPad uses the large, multitouch display of the iPad to offer this same interaction with videos in place of the pads. Additionally, it uses dragging and multi-touch gestures to take further advantage of the expressiveness and playability offered by touch screens [8].

The concept of crowdsourcing musical creation through mobile technology has been explored previously, perhaps starting in 2001 with *Dialtones - A Telesymphony* by Golan Levin[10], where phones in the audience were dialed by the performers using custom control software that allowed up to 60 phones to be dialed simultaneously. Moving from local to networked, *World Stage* has been explored in both research and products by Smule, a mobile software developer

focused on social music applications. World Stage offers a place for a community to score arrangements for each other, perform music to one another, and even anonymously judge performances[15], all on a mobile device.

3. DESIGN EXPLORATION
3.1 The Interactive Album

The initial intent of this research was to create an interactive album. The goal was to give the user a sense of interactivity that, unlike a mash-up or remix, did not overtake the composer's structure. Another objective was to compose and experience interactive music written specifically for the "popular music" domain, meaning that it would be accessible to a large audience outside of the realm of computer music. For this to work, the experience would need to be powerful no matter how much the listener chose to interact with it, creating a unique but always desirable output with every listen. Initial research left numerous questions ranging from psychology to system design: *What does it mean to write a "hook" in a non-linear, event-triggered soundscape? What is the balance between the control one maintains as a composer versus as a listener? What interactions can one leverage from available devices to manipulate a composition?*
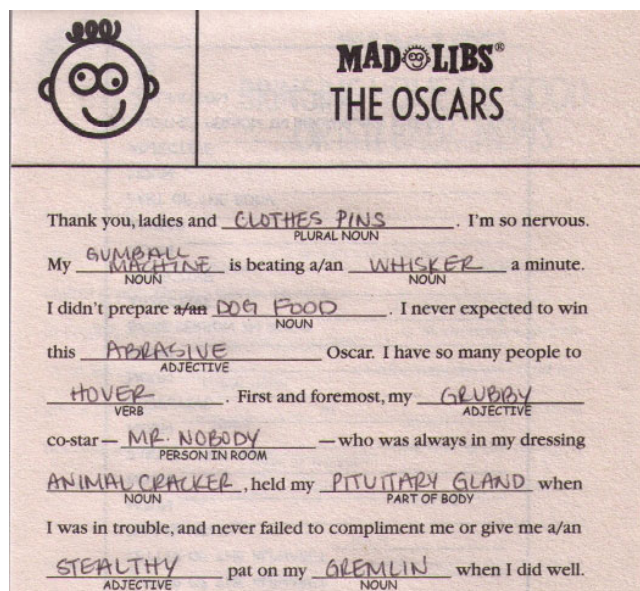

Figure 5: An Excerpt from a completed original Mad Libs. This page is only revealed *after* the words have been chosen.

3.2 From Mad to Madder

The first stab at addressing these questions was entitled *Madder Libs.* It was designed through the metaphor *Mad Libs for audio*[1]. The basic premise of Mad Libs (Figure 5) is that the player is given a page with several blanks to fill in, and a basic category for a word that he or she will choose to fill in each blank. Although it is known that these choices will fill in key words for a small story, nothing about the structure or content of that story is revealed, so the player must choose the words almost blindly, based on the given hints [11]. *Madder* Libs is very similar to this. A composer creates a song that does not produce any sound, but is a musical blueprint that pictorially hints at what sounds are to

---

[1]Hopefully it is easy to see why one might consider this to be Madder.

be used for each note (Figure 6). It is the listener's responsibility to record a sonic interpretation of the picture for each note without knowing the structure or content of the song, and upon completion the user can hear the song played with these new personalized sounds. In this way, the structure of the composition is maintained while still allowing the user to have a novel and personalized experience. At its core, Madder Libs is a non-traditional notation system that utilizes audio technology to make quick recordings rather than have the sound for each note repeatedly performed live. An extension that this technology offers is the ability to manipulate and replay these clips with accuracy, repetition, and speed beyond the limits of human ability, allowing the composer to write, for instance, extremely fast or lengthy passages without needing to worry about the limitations of the performer.

In response to the numerous questions raised about interactive album making in the previous section of this paper, this single interaction was in no way a complete answer. However, the insight gained proved valuable, and provided the foundation and the etymology for MadPad.
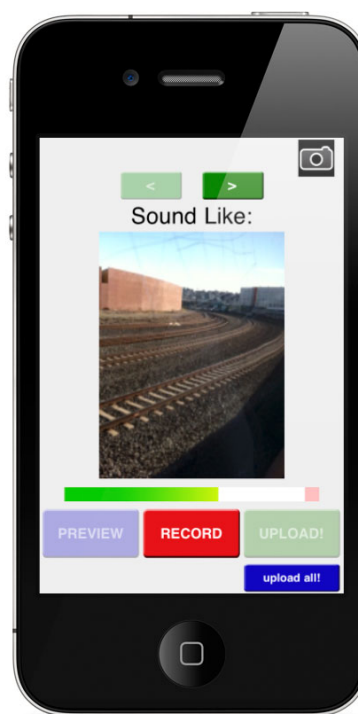

Figure 6: The Madder Libs recording interface, a predecessor to MadPad.

3.3 Bringing In the Network

At an early stage of the Madder Libs design process, it was entered into the program for The Stanford Mobile Phone Orchestra's[14] Fall concert, which was themed around audience participation. With the possibility of many audience members recording sets of audio samples for the same composition, the new goal was to amass these sets quickly to a single location, at which point they could be called back down to the audio player and added to the song. This was achieved by creating a networked database to which all sound clips for a song were submitted. The database could then be queried, and the desired samples could be downloaded and dropped into the song at any point. The resulting performance was no longer solely a personal experience, but rather it was one shared by the audience, whose members contributed the content. To ensure that
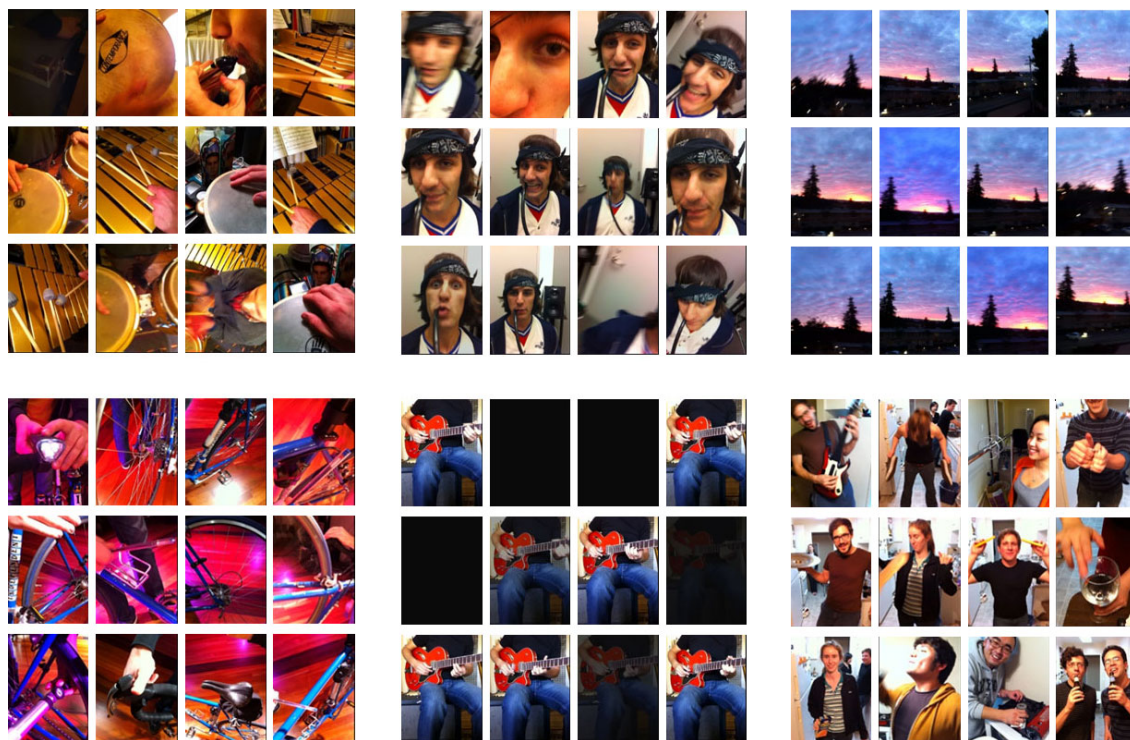
Figure 7: Six screenshots of user generated MadPads. Top left: Sampling a band. Top Center: Sampling a person playing a talkbox. Top Right: Capturing a sunset. Bottom Left: Turning a bike into an instrument. Bottom Center: Sampling a guitar (Some images have are faded because "Ghost Mode" playback is enabled, which fades away videos when they have stopped playing) Bottom Right: Remembering a dinner party.

all participants were given a chance to be included and to make each performance unique, samples were programmed to swap throughout the song. The concept of crowdsourcing content and making everyone feel like a part of the performance would become an important feature of MadPad, extending this notion from a local crowd to a global system.

3.4 From Madder Libs to MadPad

As preparations for the concert continued, we decided to add a visual component to the Mad Libs metaphor. Not only would we record a sample of the participant's voice, we'd also record the corresponding video–acquiring plenty of fodder for our projector, but more importantly giving the audience a way to connect the sounds they were hearing with the people who performed them. This emergence of the "What you see is what you hear" concept would become a main pillar of the MadPad experience. At this point the samples were laid out in a grid pattern[2] on a computer screen and triggered by precomposed MIDI messages, but it wasn't long after seeing this arrangement that the desire grew to trigger those clips on the fly and on-the-go. The whole system was ported to the iPad, utilizing both the large, multitouch surface and brilliant color display, as well as making the cloud-based social aspects of the system mobile. The concert was performed successfully on the iPad as a combination of precomposed MIDI, live performance on the device, and random audience-sourced sample swapping.[3]

---

[2]Originally there were plans to offer a wider array of layouts and transitions and this is still being considered as a future implementation.

[3]After the performance, the ability to play precomposed MIDI was removed from the feature set, as it was considered a potential source of confusion to the average user. Future work intends to include abilities to sequence and

4. CORE FEATURES
4.1 Adding Video To The Mix

The primary interaction of MadPad[4] employs the extension of sample-based, tap-triggered music to include both audio and video. As a general concept, triggering video samples on-the-fly existed before MadPad, but not for multitouch devices. Tapping in to this new interface is what separates MadPad from its audiovisual sampling predecessors. For one, the recent ubiquity [3] of multitouch devices makes the interaction much more accessible to everyday people, and reaching a large audience has always been a primary goal. Additionally, multitouch screens give performers the ability to control the videos under their fingertips, as if the pads of an MPC were replaced with individual video screens. One can infer the interaction almost instantly–touch a picture to make it play. The system itself is intended to be a generic platform[6], and recedes into the background, allowing the content to shine through and encouraging a natural sense of wonder and exploration.

4.2 Make An Instrument Out Of Anything!

When using the MadPad to create content, video clips of anything can be loaded into the sample slots, and the possibilities are only bound by the user's surroundings and imagination. This finds shared ground with the concept of musique concrète, wherein *(translated from French) "The compositional material is not restricted to the inclusion of sounds derived from musical instruments or voices, nor to*

---

record shareable performances.

[4]"MadPad" as a name was initially just a joke. The project filename was the hasty concatenation of "Madder Libs" and "iPad" when we were just doing an initial test to see if the iPad could even load this amount of data to memory. Naturally, the name stuck.
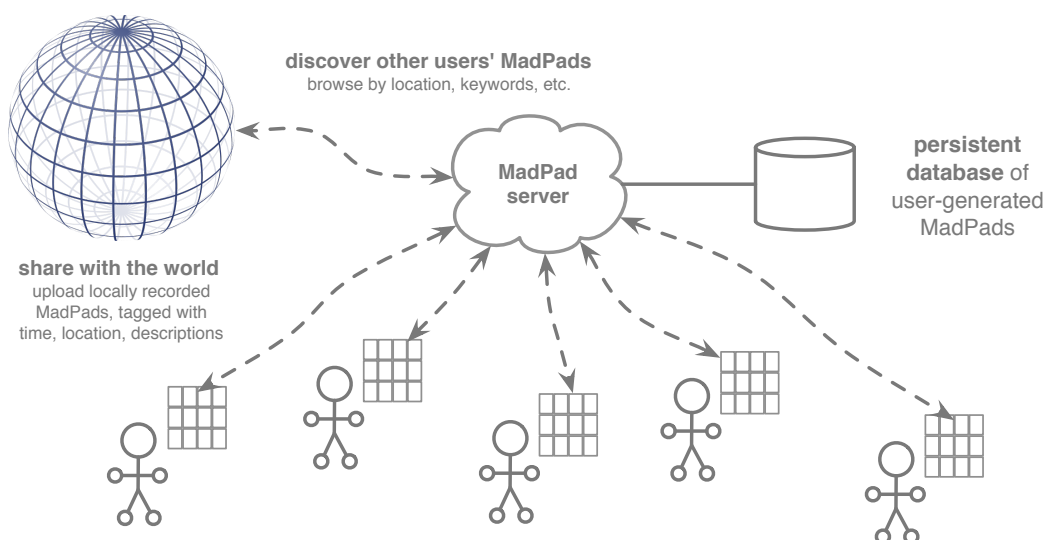
Figure 8: MadPad Community: Users serve as both content producers and consumers.

elements traditionally thought of as 'musical' (melody, harmony, rhythm, metre and so on)."[5]. However, whereas the notion of acousmatic sound in musique concrète is to intentionally separate the sound from its origin[2], MadPad's philosophy is quite the opposite. Although the concept of "found sound" is central to both musique concrete and MadPad, by its very nature MadPad encourages the association of the "found video" as well. Rather than concealing the sound source, MadPad exhibits the antithesis: What You See Is What You Hear (WYSIWYH)[5]. Interestingly, one can choose to deliberately rebel against this aesthetic and create samples where the audio and the visual have no direct correlation (e.g., a video of a marimba being struck while hearing the sound of a duck).

The WYSIWYH concept provides an additional level of personal connection to traditional sampling. By elucidating the sound source with highly contextual visuals, the system aims to generate a more holistic and immersive experience. It transforms the act of *sampling* into a type of *instrument design*. By allowing this interaction, WYSIWYH on MadPad attempts to open up the minds of its users to view everything they interact with as a potential source of music–perhaps its the sounds on the bus or the footsteps of different shoes. In addition to recording objects not intended to be musical, one can sample notes from an actual musical instrument and play them in a different way. (See figure 7 for more examples.) The MadPad platform thrives on the creation of unique, personalized musical instruments from anything important or interesting in the lives of each individual user.

4.3 A Social Sampler

Creating an environment to share content is another important feature of MadPad. Although standard sample sets[6] are readily accessible, fresh user-generated content is available through a MadPad Community. (Figure 8). The concept of social music content generation is explored in the design of Smule's Ocarina, an iPhone application that transforms the phone into an expressive, flute-like instru-

ment[13]. The Ocarina community uses a simple tablature notation system to share popular melodies with its users in the form of an online songbook, with over 2,000 songs currently viewable. Thus the value of the Ocarina is constantly being enhanced due to the dedication of the user base. Similarly, in MadPad the social aspect autonomously extends the available content. In addition to creating samples, the user can browse clips from users around the world. The community serves as a forum for sharing creative ideas and collaboratively developing new ways in which the MadPad platform can be used for musical expression. This adds value for the viewer of the content, and it also adds a new level of drive for the creator. Knowing that one's concept of a musical instrument will be viewed by anonymous people around the world can motivate the production of more content and the innovation of more ideas for what these twelve empty slots can do, continuing to enrich the community.

Samples can also be discovered based on location. They are loaded as a conglomeration of the twelve closest samples made by distinct users. This mode allows a user to load in a set of samples recorded in a chosen geographical region and play an audiovisual instrument based on the collaboration of people who might be complete strangers to each other, but all share a similar proximity (e.g. loading in twelve samples from twelve different users in downtown Chicago).

In addition to offering anonymous sharing, MadPad also offers an ability to share locally without a network. For instance, friends at a party can take samples throughout the night just as one might snap photos. The result is an instrument that documents small snippets of the events that transpired. This type of scene capturing is a novel form of persistent media that bridges the gap between a photo album and a video, in that it offers a quickly digestible and *interactive* way to relive the moment. In another example, many people can pass around the camera and take turns recording the samples which will ultimately become a final instrument. This collaboration allows each performer to give individual input into what the instrument should be, and the result is a unique mix of personality and imagination representative of the group (Figure 7, upper left, is an example from a social gathering). After performing *for* the recording, the group can immediately gather around and continue to perform *with* the recording, closing the loop of the MadPad system.

---

[5]And commutatively, What You Hear Is What You See (WYHIWYS)

[6]Standard instrument sets and quirky idea sets are permanent, downloadable links, bundled with the application when downloaded.

5. CONCLUSIONS

MadPad began with the desire to create interactive compositions and evolved into a social/mobile platform for audiovisual creativity and collaboration. The ability to tap on a picture and make it play serves to bring creativity out of those who are not familiar with a traditional audio sampler, while also giving those who are familiar with it a new dimension to their creativity. Using the platform to create an instrument out of anything one sees encourages people to view the world as a more musical place. Giving people a place to share their creations allows them to learn from each other, and see the musical world that exists in every person's life–that is always present.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Akai Pro. http://www.akaipro.com/mpc. Retrieved January 2011.

[2] M. Chion. *Audio-Vision Sound on Screen*. Columbia University Press, July 1994.

[3] J. Colegrove. The state of the touch screen market in 2010. DisplaySearch Touch Panel Market Analysis. Retrieved 2011-01-25.

[4] J. Conte. VideoSong 1 - Push - Jack Conte. Online video clip. YouTube, March 2008. Retrieved 2011-01-25 from http://www.youtube.com/watch?v=FUVgPjnEMzw.

[5] J. Dack. Technology and the instrument. *musik netz werke - Konturen der neuen Musikkultu*, 2002.

[6] G. Essl, G. Wang, and M. Rohs. Developments and Challenges turning Mobile Phones into Generic Music Performance Platforms. In *Proceedings of the Mobile Music Workshop*, Vienna, Austria, 2008.

[7] L. Gaye, L. E. Holmquist, F. Behrendt, and A. Tanaka. Mobile music technology: Report on an emerging community. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 22–25, Paris, France.

[8] G. Geiger. Using the Touch Screen as a Controller for Portable Computer Music Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Paris, France, 2006.

[9] L. Gjertsen. Amateur. Online video clip. YouTube, November 2006. Retrieved 2011-04-25 from http://www.youtube.com/watch?v=JzqumbhfxRo.

[10] G. Levin. Dialtones - a telesymphony, September 2001. Retrieved 2011-04-25 from http://www.flong.com/projects/telesymphony/.

[11] Penguin Group USA. Mad libs. Retrieved 2011-01-25 from http://www.madlibs.com.

[12] Two Hand Band. History and Significance of the MPC. Documentary Film, August 2008.

[13] G. Wang. Designing Smule's iPhone Ocarina. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburgh, USA, 2009.

[14] G. Wang, G. Essl, and H. Penttinen. Do Mobile Phones Dream of Electric Orchestras? In *Proceedings of the International Computer Music Conference*, Belfast, 2008.

[15] G. Wang, J. Oh, S. Salazar, and R. Hamilton. World Stage: A Crowdsourcing Paradigm for Social / Mobile Music. In *Proceedings of the International Computer Music Conference (under review)*, Huddersfield, UK, 2011.

[16] L. Werthheimer. Pomplamoose: Making A Living On YouTube, April 2010. National Public Radio.