

“Voice Networks” – Exploring the Human Voice as a Creative Medium for Musical Collaboration

Gil Weinberg

Music Department, Georgia Institute of Technology
gil.weinberg@coa.gatech.edu

Abstract

I describe a musical installation that allows players to record, transform, and share their voices in a group. A central computer system facilitates the interaction as participants interdependently collaborate in developing their “voice motifs” into a coherent musical composition. Observations of subjects interacting with two different applications that were developed for the installation point to a number of underlying design concepts, such as interdependency, coherency, visualization and personalization that can be applicable to the development of effective collaborative musical applications in general.

1 Introduction

Music today is more accessible than ever thanks to technological developments in recording, compression and distribution. Unfortunately, most of the music that we listen to is consumed in an incidental, unengaged and/or utilitarian manner (DeNora 2000). Recent developments for the home studio address this problem by providing more engaged and creative musical experiences for novices. These developments, however, lead to undermining one of the most valuable traits of music – its inherent collaborative social attribute – by promoting private and isolated musical practice where hardly any musical instruments or even musicians are needed and where the value of live group interaction is marginalized. These observations led me to define two challenges for my research into designing and developing musical applications for novices and the general public. The first, inspired by the notion of constructionism (Papert 1980), is to design applications that would allow players an access to engaging, thoughtful and creative musical experiences that are personally meaningful for them. For this I decided to use the human voice as an idiosyncratic signature, owned by all, that can serve as malleable medium for personal creative exploration. My second research challenge is to try and enhance the social aspects of music making by concentrating and elaborating on the roots of music as a collaborative social ritual. To address this challenge, I decided to utilize the digital network as a framework that can provide unprecedented levels of interdependency, which can allow players to take

an active role in determining and influencing not only their own musical output, but also their peers’. The network in such systems can be likened to a habitat that supports its inhabitants (players) through a topology of interconnections and mutual responses, which can, when successful, lead to new crossbred musical “life forms.” (Bischoff, Gold and Horton 1975, Weinberg 2001).

2 Related Work

The algorithmic combination of digital input from various independent sources towards the creation of hybrid artistic artifacts (also referred to as digital “gene mixing”) has been explored in other media such as graphics and video (Sims 1998). In music, composers such as John Cage (1951), The League of Automatic Music composers and the Hub (Gresham-Lancaster 1998) and a variety of Internet musicians (see a survey in Weinberg 2002) explored similar interdependent interactions. These experiments, however, usually required advanced musical skills and experience from players as well as audiences, and often led to inaccessible “high art” musical products. More recent collaborative musical installations for novices on the other hand (for example Machover 1996, Iwai 2000, Levin 2001, Blain 2002), have rarely taken advantage of the unique interdependent possibilities offered by the network and have tended to preserve the musical autonomy of individual players.

3 “Voice Networks”

My main goal in designing “Voice Networks” was to create an interdependent musical collaboration that is easy to access for novices but is also rich, thoughtful and leads to coherent musical products, so that expert musicians can find it meaningful and challenging. As part of the development process I was also interested in identifying design schemes that would be relevant to the development of successful interconnected musical network applications in general. I based “Voice Networks,” on the concept of collaborative motif construction where players can record a personal “voice motif,” modify it, and share it with the group. “Voice Network” does not require musical skills or knowledge at the entry level as anyone who can use his/hers voice can

contribute a “motif.” But as more participants enter the interaction, the richer and more thoughtful the interaction becomes, as players transform each other’s motifs, insert their own musical ideas into the shared construction process, and attempt to synchronize and align all the materials into a cohesive musical composition. The role of the central system is to regulate players’ interactions and to coordinate between the constantly evolving interlocking motifs in an effort to support a meaningful musical experience.

4 The Platform

“Voice Networks” is installed around a 4’ high square podium with surface dimensions of 2’x2’. Four control stations, each consisting of a microphone and touchpad controller (the commercially available Korg Kaoss Pad), are installed on each side of the podium (see Figure 1). All stations are facing each other so that players can see and listen to their peers while playing. A flat screen monitor for visualization is installed on top of the podium in-between the stations. Four speakers, one per station, are located on the floor facing their respective players. A Macintosh computer running Max/MSP connected to a multi-port MIDI and audio interfaces are located inside the podium.

Two applications were developed for this platform, “Voice Patterns,” and “Silent Pool,” which differ in their approach for interdependent group play. The individual interaction in both applications is identical – players can press the record button on the Kaoss pad and record their voices into a pre-designated buffer. Microphones’ height and location were chosen to encourage participants to use their voice. A second press on the button stops the recording and immediately starts playing the recorded motif in a loop through the respective speaker. Players can then transform their voice motif by moving their fingers on the touch pad (see Figure 2). All four stations are based on identical transformation algorithms, programmed in Max/MSP. These transformations were designed to allow easy access to basic musical parameters such as pitch, amplitude, rhythm and timbre. The touch pad is divided into four quadrants (see figure 3); each features a unique transformation effect with two real-time user-controllable parameters. At the bottom right quadrant players can change the pitch and amplitude of their looped voice. The bottom left quadrant offers pitch and amplitude control as well, but here the



Figure 2. A player records her singing voice and manipulates it by moving her finger over the touch pad

motifs are played in reverse. The top right quadrant allows players to manipulate the parameters of a low-frequency oscillator (LFO) mapped to modulate the amplitude of the recorded audio. Players can change the LFO’s amplitude and frequency, which leads to a variety of rhythmic pulsation effects. At the top right quadrant players can interact with a delay line algorithm in a range that leads to a variety of timbre transformations such as flanging and chorus. These simple mappings were chosen to cover a wide range of basic musical elements that are simple to follow, but that can create quite elaborate transformations when interdependently controlled by a group.

5 Application 1 – “Voice Patterns”

The group interaction in “Voice Patterns” is aimed at encouraging players to align their gestures with each other, therefore synchronizing their voice transformations. The result of successful synchronization is a more coordinated musical outcome, which culminates in the trading of voices between the two synchronized players. This allows player to collect and manipulate their peers’ voices. The audible effect of the synchronization depends on the specific musical parameter assigned to each quadrant on the pad. After the trading, players can manipulate and develop the voice motif that they received and try to trade it further by synchronizing their gestures with other participants. The musical output of the system, therefore, is quadraphonic propagation of motifs and variations, which successively gets in and out of synchronization. To support this



Figure 1. “Voice Networks” Platform – a microphone and a touch pad serve as input controllers for each station

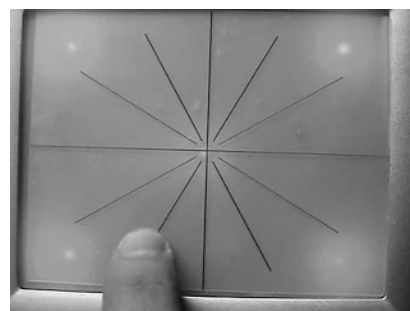


Figure 3. The Kaoss touch pad is divided into four quadrants, each features two real-time audio transformation controllers

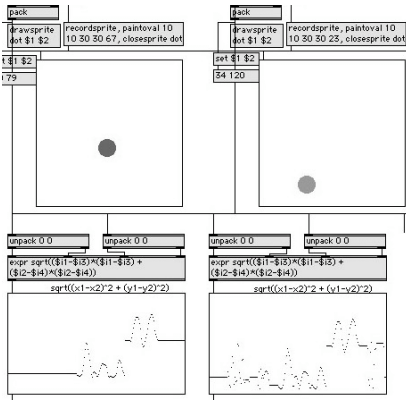


Figure 4. The synchronization engine in “Voice Patterns” - checks for synchronized finger gestures by players

interaction “Voice Patterns” utilizes four audio buffers that can be recorded to and manipulated by MIDI commands sent from the Kaoss touch pad. A synchronization engine written in Max/MSP (see figure 4) constantly checks for matching MIDI input patterns from the pads and executes a voice trade when the synchronization lasts more than 5 seconds. Players can record and re-record their voices to the assigned buffer at their leisure and manipulate the sound by moving their finger on the touch pad as the system is looking to match their patterns on the pad with their peers. Animated visualization at the central monitor shows participants their and their peers’ location on each pad. When a match is detected, a yellow connection bar appears between the matching stations. During the synchronization, the yellow connection bar slowly turns into red, a transformation that culminates with the trading of the voices between players (See figure 5). The couple can then play in “duo mode” as the other two participants are muted. After 5 seconds all players become active again and can participate in the interaction until the next trade.

5.1 Observations

In preliminary observations that were conducted in an open house at MIT Media Lab, players spent anywhere between 1 to 5 minutes interacting with the system. Using the voice allowed almost any visitor to have some degree of meaningful group interaction. Players sang, spoke, clapped, whistled, or just tapped on the microphone. The most successful interaction, however, was the use of the voice as players were intrigued to follow the different transformations that their personal and familiar input was about to undergo. The effort to create a meaningful musical outcome from a collage of unrelated audio segments was only partly successful, as some of the musical parameters proved to be difficult to coordinate and synchronize. The most successful parameter in that regard was the pulsation effect generated by the LFO synchronization, which provided a unified beat to the collage. The synchronization of other musical parameters, such as pitch and timbre did not always lead to a coherent musical outcome. Better audio analysis and harmonization algorithms can be used to improve these results. Another problematic aspect of the interaction was the central role given to the

visual display, which led players to focus on the graphical aspects of matching their patterns on the screen rather than on the musical aspects of their actions. When the graphical display was disconnected in an effort to allow players to concentrate on the audio transformations, unplanned voice trading occurred when two players were exploring the same area at the same time.

6 Application 2 – “Silent Pool”

“Silent Pool” was developed in an effort to improve some of the deficiencies that were detected in “Voice Patterns.” The main goal here was to create a more abstract music-centered experience by minimizing the role of goal-oriented activities and by creating a less literal and less distracting visualization scheme. In addition, the application was designed to provide a more dynamic musical result, which is based on the appearance and reappearance of familiar motifs and transformations over time, and to allow more autonomous control for individual players by preventing unwanted interactions with the group. “Silent Pool,” therefore, utilizes eight audio buffers as opposed to the four buffers in “Voice Patterns.” In an idle mode, four of the buffers are assigned to each station and the other four are assigned to a “silent pool” where they are muted. Trading operations are performed with a random muted buffer from the silent pool and do not require direct gesture synchronization with other players. Colorful circles that represent the different voice buffers (See figure 6) visualize the network architecture and activity as opposed to the more literal finger-location visualization in “Voice Patterns.” After a sound is recorded into a buffer its respective circle starts to “breathe,” performing cyclic expansion and contraction movements that correspond to the input from the touch pad. When players are ready to trade their transformed sounds, they can leave the pad and the circle automatically moves into the silent pool, shrinking in size as it fades-out in volume. When the circle reaches the silent pool, a random circle from the pool, representing a buffer that contains a previously recorded and transformed voice motif, automatically moves toward the station to complete the trade. The receiving player can then manipulate the new voice or record a new motif into the buffer.

6.1 Observations

The four extra buffers in the silent pool, which allow players to trade with the system rather than trading directly

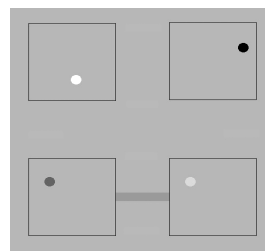


Figure 5. Visualization of “Voice Patterns”

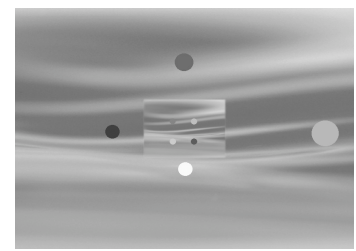


Figure 6. Visualization of “Silent Pool”

with each other, were successful in turning players' attention from the graphical synchronization to the musical aspects of their actions. The new visualization scheme, representing the network topology rather than each player's precise location on the pad, was instrumental in achieving this goal. By directly interacting only with the central system, participants were granted full control over trading their sounds and were not concerned with having unwanted interactions with the group. The new design also led to richer motif-and-variation musical results where a voice motif that was created and transformed in one station could have been silenced for seconds or even minutes before coming back to the foreground in another station, allowing other motifs to evolve and dissolve in the meantime. In an effort to enhance this effect and to support a more dynamic musical outcome players were provided with a wider variety of input sources such as a set of percussive instruments that were available for them in the room (see figure 7). Players used these instruments to support and complement their voice-based material. But "Silent Pool's" new design also led to some deficiencies in comparison to "Voice Patterns." The sequential interaction, which aimed at maintaining more coherent and undisturbed interaction for each player, compromised the quality of the interpersonal group collaboration. In particular, players found the random routing scheme disturbing, as it did not allow for choosing a particular playing partner. Another problematic feature was the removal of the trading reward in an effort to provide a more abstract musical experience. The lack of a clear goal for the experience compromised the drive and interest of players, which in general spent less time interacting with the system in comparison to "Voice Patterns."

5 Conclusion

Some of the system's weaknesses, as identified in both applications, call for specific solutions such as adding an audio analysis phase, which can lead to better aligning and synchronization of motifs. The different collaborative approaches that were experimented with can also point to a number of network design considerations that can be applicable to the development of collaborative musical networks in general. The concept of digital "gene mixing", for example, proved to be effective in generating interesting crossbred musical products. But in order to prevent



Figure 7. Players using their voice, as well as musical instruments in "Silent Pool"

confusion and loss of control it was clear that designers of collaborative networks should carefully choose the parameters that would be controlled interdependently and those that should stay autonomously controlled by individuals. In "Voice Networks," influencing the timbre of peers' voices proved to be more coherent and less confusing than interdependently changing pitch. Another transferable finding is that sequential networks, in which players create and manipulate their input without external influence before sharing their products with the group, can lead to more coherent but less engaging interaction than synchronous networks that allow players to manipulate the same material in real-time. The two different approaches for user interaction that were taken also show that the balance between goal-oriented activities and abstract artistic experiences should be carefully maintained. A similar balance should be maintained when designing a visualization scheme for the interaction, which should help represent the interaction for players without distracting them from the musical aspects of their actions. And lastly, providing a personal connection for novice participants to the musical material they create, such as in allowing them to experiment with their own voices as a creative medium, proved to be an effective tool in creating a personal, engaging, and challenging collaborative musical experience.

Acknowledgments

I would like to thank Tod Machover and Tristan Jehan for their support and advice.

References

- Bischoff, J., R. Gold, and J. Horton. (1975). Microcomputer Network Music. *Computer Music Journal*. MIT Press: 2(3), 24-29.
- Blain, T (2002). "Multi-player Musical Controller into the Jam-O-Whirl Gaming Interface," *Proceedings of NIME-02*.
- Cage J. 1951 "Imaginary Landscape no.4"
- Cage, J. Silence. (1961). Wesleyan University Press, Middletown, CT
- DeNora, Tia. (2000). *Music in Everyday Life*. Cambridge University Press, Cambridge, U.K.
- Gresham-Lancaster, S. (1998). "The Aesthetics and History of the Hub: The Effects of Changing Technology on Network Computer Music." *Leonardo Music Journal* vol. 8, , pp. 39-44.
- Iwai, T. *Composition on the Table*. Exhibition at Millennium Dome 2000, London, UK.
- Levin, G. Telesymphony web site. <http://www.flong.com/telesymphony>
- Machover, (1996) T. *The Brain Opera*, web site. <http://brainop.media.mit.edu>.
- Papert, S. *Mindstorm*. (1980). Basic Books, New York.
- Sims Karl (1998) *The "unofficial" retrospective web site*. <http://www.biota.org/ksims>
- Weinberg G., and Gan S. (2001). *The Squeezables: Toward an Expressive and Interdependent Multi-player Musical Instrument*. *Computer Music Journal*. MIT Press: 25(2), 37-45.
- Weinberg, G. (2002). *The Aesthetics, History, and Future Challenges of Interconnected Music Networks*. *Proceedings of the International Computer Music Conference*, pp. 174-177. Gothenburg, Sweden: International Computer Music Association.