Musique Concrète Choir: An Interactive Performance Environment for Any Number of People

William Walker, PhD
Senior Manager
Mozilla Corporation
Mountain View, California, USA
(01) 650 903 0800
bwalker@mozilla.com, softwarewalker.com

Brian Belet, DMA
Professor of Music
San Jose State University
San Jose, California, USA
(01) 408 924 4632
brian.belet@sjsu.edu, www.BeletMusic.com

ABSTRACT
Using the Web Audio API, a roomful of smartphones becomes a platform on which to create novel musical experiences. As seen at WAC 2015, composers and performers are using this platform to create clouds of sound distributed in space through dozens of loudspeakers. This new platform offers an opportunity to reinvent the roles of audience, composer, and performer. It also presents new technology challenges; at WAC 2015 some servers crashed under load. We also saw difficulties creating and joining private WiFi networks. In this piece, building on the lessons of WAC 2015, we load all our sound resources onto each phone at the beginning of the piece from a stable, well-known web host. Where possible, we use the new Service Worker API to cache our resources locally on the phone. We also replace real-time streaming control of roomful of phones with real-time engagement of the audience members as performers.

1. INTRODUCTION
Mobile devices have become a ubiquitous part of most people's daily routine, and one role these devices have captured is that of roving personalized entertainment platform. When audio is included, as it often is, the accepted practice is to use individual headsets for increased audio quality. When mobile devices are included as part of a larger social context where shared audio output is acceptable, and even desirable, the hardware limitations of the devices soon become apparent. Creating meaningful artistic contexts for shared music making remains a challenge, one that can be addressed as smartphone hardware and software continue to evolve.

The authors began collaborating at the University of Illinois in the 1980's where a unique blend of computer audio art, science, and engineering found a home in the CERL Sound Group [http://www.cerlsoundgroup.org]. Walker worked with composer Sal Martirano creating a second generation of his seminal SalMar Construction using MIDI devices and a real-time Smalltalk-80 engine. This led to Walker's ImprovisationBuilder framework [Walker, Hebel, Martirano, and Scaletti, 1992], which was the basis for a joint 1996 composition with Belet for two Disklavier pianos, each controlled by a separate version of the software that affected the other performer's piano (*Cross-Town Traffic*). That collaborative research was presented at the 1996 International Computer Music Conference [Walker and Belet, 1996]. More recently they presented their geographically dynamic birdsong soundscape project at last year's Web Audio Conference in Paris, France [Walker and Belet, 2015].

Both authors continue to work in computing and composition: Walker leads a team of engineers at Mozilla focused on delivering great web experiences on mobile devices [http://softwarewalker.com]; Belet utilizes Smalltalk-80 within Symbolic Sound Corporation's Kyma digital audio workstation for his composing and performing platform [http://www.symbolicsound.com]. The authors draw on their musical experience to shape this current performance project into a deeper aesthetic experience for all participants in their shared composer, performer, and audience roles.

2. DELIVERING CUSTOMIZED, LOW FRICTION, HIGH PERFORMANCE AUDIO EXPERIENCES ON THE WEB
The open web is now becoming a compelling platform for audio experiences, thanks to the convergence of several trends: (1) unprecedented computational power due to advances in JavaScript performance and mobile CPU's; (2) wide availability of Web Audio API in modern web browsers; (3) reliable caching of app resources using the Service Worker API. Unlike native applications, composers can easily create and publish (and audiences can easily participate in) audio experiences built as web applications.

3. AUDIO SOURCE FILES
The authors composed, performed, and recorded the source audio files used for this project. Individual file durations range from ten to fifty seconds. The composition environment is structured with a total of sixteen source audio files: four audio files assigned to each of four performance sections. Walker recorded eight Hammond B3 organ gestures and Belet recorded eight viola gestures. Two of the *Musique Concrète* Choir performance sections contain B3 organ files and the remaining two sections have viola files. The audio files are further divided into gestural

Web Audio Conference WAC-2016, April 4–6, 2016, Atlanta, USA.

groups: half of the files contain long, sustained gestures while the other half contain short, percussive gestures. To maintain large-scale aesthetic unity, all of the audio files are derived from a single harmonic reference (D7#9). Performer manipulations of these audio files (including frequency shifts) are described below in Section 5 (Client-side Processing).

4. LIVE PROCESSING

4.1 Audio Playback

The sixteen source audio files are loaded asynchronously into memory upon first visiting the website for audience participants (hereafter audience performers). The sounds are allocated to pre-assigned performance section categories. Whenever an individual user selects a specific audio file within that group, a random index is generated to determine a unique starting point within that sound file.

The smartphone version of the performance interface allows control over one sound at a time and uses monophonic audio files.

The audience choir is supplemented by a single performer with laptop (principal performer, described below in Section 4.2), whose version of the interface controls multiple copies of each sound. This laptop is connected to the main performance audio system and provides reinforcement of the choir.

The source code is available at https://github.com/wfwalker/mc-choir; the smartphone interface is available at https://wfwalker.github.io/mc-choir/.

4.2 Laptop User Interface (Principal Performer)

The principal performer has access to all sixteen audio files and is able to add any one or more of these sound files to the overall audio mix as the performance progresses (see Figure 1). This polyphonic capability contrasts with each audience performer's monophonic capability.

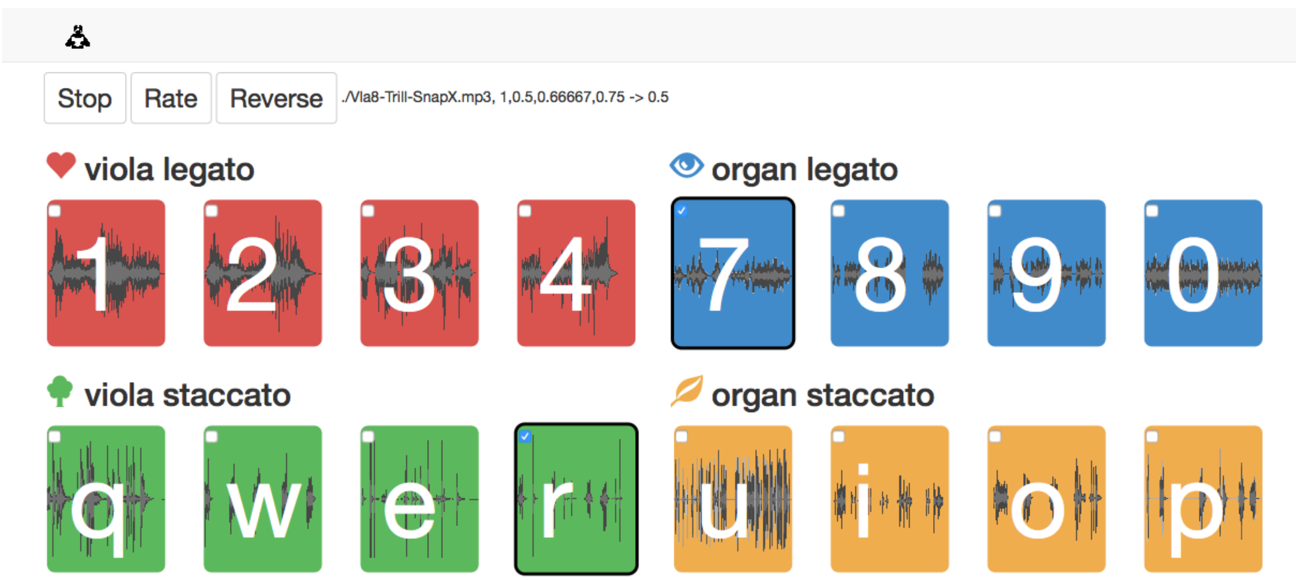

Figure 1. ***Musique Concrète*** **Choir polyphonic master control user interface for the Principal Performer's laptop.**

4.3 Smartphone User Interface (Audience Performer)

The audio files are organized into four performance sections, each containing four related audio files. Each performance section is assigned a color, and the available audio files are identified with numbered buttons, although not further described. This performer's smartphone interface is a streamlined version of the more comprehensive controlling laptop interface (see Figure 2).

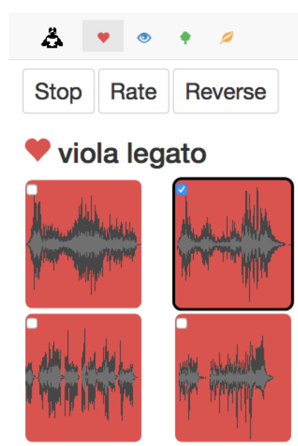

Figure 2. ***Musique Concrète*** **Choir monophonic user interface for performers' smartphones.**

User controls are basic and direct, designed to invite audience play and participation. In a traditional concert setting, whether a symphony orchestra reading traditional music notation or a smaller jazz combo reading lead sheets with chord symbols, each performer needs all the relevant information required to realize their part of the performance, but no extraneous information. A separate conductor may have access to a larger information data set in the form of a full score. This paradigm guides our approach here; the conductor needs access to the overall structure of the composition, while the audience performers only need control over their individual device audio.

The four performance sections are identified by color and icon. The conductor gives silent, visual instructions to the audience performers using these colors and icons. For example: red section performers start, and/or green section performers stop their sound files on the next specified visual signal. These signals can be communicated using the projected laptop screen, as well as by hand signals and color-coded cards. In performance, the audience performer selects a performance section and then selects any one of the four audio files assigned to that group. Selecting a sound file starts the file playback, with either forward or reverse direction, a random start index, and an initial playback rate randomly selected from a restricted set of values. A Stop button stops the current sound, avoiding the need for the performer to hunt through all four tabs for this function.

The user interface for the smartphone and for the laptop use the same JavaScript code and differ only in their HTML presentation. All the behavior differences between the two user interfaces are encoded in HTML attributes, allowing a loose coupling of appearance and implementation.

4.4 Concert Performance

In addition to creating the sixteen source audio files, the authors have composed the large-scale plan of the composition. Titled *Cross-Town Traffic 2.0*, this composition structure guarantees a cohesive time plan for the music performance, with a beginning, middle, and ending to the abstract narrative within a set overall time plan of ten minutes. The audience performers need not know this plan; only the principal performer and conductor need to know. Using very simple signals, which are explained and rehearsed briefly prior to the performance, the principal performer and conductor guide the performers through the concert (see Figure 3).

Figure 3. Conductor cuing the Red section, with the current performance status displayed on the projection screen.

Within the overall structure, individual performers retain performance freedom over several audio functions, including selecting which audio file to use within their group, Stop/Start, forward or reverse direction, jumping to a new playback rate (but not selecting the specific rate), and even switching sections by selecting another color and its new set of four audio files.

5. CLIENT-SIDE PROCESSING AND OPTIONS

When a performer selects a sound, playback begins at a random point within that sound file. This results in a cloud of individual sonic paths within the selected sound files whose complexity and density increase as the number of performers increases, even if only a single sound file is selected. As the sound continues to play, the performer can choose to randomly alter the playback rate. To preserve the harmonic structure of the overall composition, the playback rate options are restricted to the original recorded speed (1x), twice the speed (2x: resulting in an octave shift up), half the speed (0.5x: an octave shift down), plus a few near-related overtone series relationships (1.5: a perfect fifth higher; 0.667: a perfect fifth lower; 1.333: a perfect fourth higher; and 0.75: a perfect fourth lower). This small number of playback rate options, compounded by the number of performers, results in a very large number of simultaneous and sequential permutations of the selected sound files, while still maintaining a unified large-scale musical aesthetic sense to the performance.

The audio files are monophonic MPEG-3 files, an audio compression format that decreases loading times while providing sufficient audio quality for the smartphone's built-in speaker. Performance spatialization is determined by the physical location of the performers in the concert space. Performers are initially organized into color sections at the beginning of the performance: one quarter of the physical room is designated the red section, one quarter is the green section, and so on. This initial spatial arrangement creates a soundscape context with clearly defined sonic characteristics in each section of the room, analogous to seated sections in an orchestra or band. Entropy can invade the initial organization when individual performers switch sections during the performance, or if they physically walk from one section's original location to another.

6. MUSIC AESTHETICS AND COMPOSITION

6.1 Historical Lineage

This current composition project shares a historical connection to *musique concrète* in a way that is similar to the authors' WAC 2015 birdsong soundscape [Walker and Belet, 2015; http://birdwalker.com/quartet.html]. One of the two original approaches to electronic music (developed in the 1930s and 1940s), *musique concrète* uses real (or 'concrete') sounds recorded from both the traditional musical and environmental realms as source material. All sounds and all manipulations of these sounds are fair game for compositional exploration.

6.2 Aesthetics Determined by Physical Factors

There are several aspects of this current performance environment that cannot be predicted or precisely controlled, resulting in a unique realization for each performance. The first human factor is the number of performers. A minimum of four performers is required so that each sound group can be represented. The maximum number is limited only by the size of the performance space. The next human variable is the performance skill of each user. We have designed the user interface and group performance paradigm to be relatively simple, but we cannot control a given user's skill (technical and musical) in the context of the performance itself.

The audio output of current smartphones is quiet and tinny, constraints that composers of this kind of work must confront. In response to these real technical limitations and human variables, and considering the larger joy of group composition and performance, we welcome the musical aesthetic championed by John Cage (1912-1992), acknowledging that for any performance ''[A]nything … may happen. A 'mistake' is beside the point, for once anything happens it authentically is." [Cage, 1961]

6.3 The Composer-Performer-Audience Paradigm

In classical music the roles of composer and performer are usually addressed by different people. Even when one person plays both roles, the work of composing and performing are usually separated in time and space. This has developed over hundreds of years due the high level of complexity and the tremendous amount of time required to master each craft. In jazz or rock music, the composer is very often also the performer (often joined with other performers in a band for this re-creation part of the process). Many genres outside classical music lend themselves to combining the composer and performer roles together (e.g., Blues progression, play the 'head' and then improvise variations, play central riffs and then improvise solos over this foundation), as the tasks required for both parts of the process, while still requiring talent and specific skills, are not so specialized to preclude this combination. The more adventurous aspects of the jazz and rock worlds still require either a separation of these creative tasks, or are produced in tandem by only a few individuals who have developed their diverse musical skill levels to a very high degree.

While the relationship between composer and performer (as creator and re-creator) is one relationship under consideration, determining the role of the listening audience is another matter entirely. There are many music contexts where the listener is expected be only that – the last step in the music-making process, a task of receiving the finished sound without contributing to its construction. This listening process can be either passive or active, depending on the social context of the music, the related expectations of the listening activity, and the skill and musical training of the individual listener. There are those musics where the audience is invited, even expected, to participate. These are usually either semi-formal contexts (e.g., singing the chorus of a hit rock song during a stadium concert, or attending *The Rocky Horror Picture Show*) or ritualized formal events (e.g., congregational singing of a church hymn). In both contexts care is taken by the composer to structure the audience participation so that it can be successfully performed with little or no rehearsal, and by individuals who are not necessarily trained music performers. This amateur performance paradigm is usually supported by the professional performers at some level (e.g., the rhythm section of the band keeps playing, or the church organ and trained choir sing the hymn).

The increasing technical capabilities of mobile communication devices and the widespread use of social media have encouraged a renewed interest in audience participation in music and video contexts. Recent advances in web browser technology have introduced yet another reflection on this paradigm. Stanford University researchers Jieun Oh and Ge Wang cite individual user convenience and direct access (i.e., the user's own smartphone becomes a performing instrument) as primary reasons for the growing audience interest in this activity, yet they acknowledge that this is aimed at the amateur rather than at the professional performer [Oh and Wang, 2011]. Wang and Oh continue their experiments within the Stanford Mobile Phone Orchestra (http://mopho.stanford.edu/), which was formed in 2008. Formerly associated with the Stanford project, Georg Essel directs the Michigan Mobile Phone Orchestra (http://mopho.eecs.umich.edu/). Other such ensembles are beginning to appear in academic research centers (e.g. Aalto University, Helsinki). As with any new technical development (including new musical instrument design) much of the work to date is largely focused on developing the tool itself. The existing mobile phone ensembles primarily replicate the synthesis approach to sound, using the phone to model basic oscillator-filter modules or to replicate existing instruments. Our approach is to use pre-recorded sound samples, and then use the smartphone to manipulate these sounds in performance.

Several limitations and concerns remain, both technical and aesthetic. Technical issues include the existence, speed, and reliability of mobile device networks as well as the current capabilities of the various web browsers. This is extremely important in live music performance, as time synchronization is a crucial factor. Aesthetic concerns include the requisite skill levels of each performer, and the degree of formality intended for the performance context. Balancing the communal joy of music making and the artistic control over the final audio result is another issue to consider. Having the listener participate in the music making process as performer (and sometimes also as composer) is indeed a socially engaging experience, an activity that carries its own positive value. The task for the composer or master-performer (this latter term used by Oh and Wang, which corresponds to our principal performer) is to direct the overall performance experience, and to ensure that the group effort serves a singular artistic goal. Establishing an effective means of communication between the master-performer and the performance group becomes a necessity for both technical and aesthetic concerns.

7. SUMMARY

7.1 Current Status of Performance Environment

We intend all our decisions to support the live performance experience. Technologically, our performance interfaces are resilient in the face of unreliable networks. Aesthetically, working with a fully composed musical structure provides artistic continuity. In the moment, individual performers are free to make numerous decisions that shape their audio output, providing an element of playful participation and experimentation while serving a larger artistic plan. The true test of the success of this endeavor is a live performance, followed by evaluations of the technical issues (objective) and artistic results (subjective).

7.2 Performer (Client-Side) Capabilities for Future Development

We hope to extend our smartphone interface by using the Device Orientation API to control sound playback based on data from the phone's accelerometer.

8. ACKNOWLEDGMENTS

Our special thanks to Rainer Cvillink and Spencer Hui from Mozilla for producing the performance demonstration video for this project, and to Harald Kirschner, Michael Ellis, Havi Hoffman, Matthew Claypotch, for play testing the interface and performance experience. Dr. Belet thanks the School of Music and Dance, San Jose State University, for partial travel support to attend WAC 2016.

9. REFERENCES

[1] Cage, J. 1961. Composition: To Describe the Process of Composition Used in *Music of Changes* and *Imaginary Landscape No. 4*. In *Silence* (Middletown, CT: Wesleyan University Press) 59. https://archive.org/details/silencelecturesw1961cage.

[2] Oh, J. and Wang, G. 2011. Audience-Participation Techniques Based on Social Mobile Computing. In *Proceedings of the 2011 International Computer Music Conference*, (Huddersfield, UK, July 31 – August 5), International Computer Music Association, San Francisco, CA, 666-72.

[3] Walker, W., and Belet, B. 1996. Applying ImprovisationBuilder to Interactive Composition with MIDI Piano. In *Proceedings of the 1996 International Computer Music Conference* (Hong Kong, China, August 19-24). International Computer Music Association, San Francisco, CA, 386-89.

[4] Walker, W., and Belet, B. 2015. Birds of a Feather (Les Oiseaux de Même Plumage): Dynamic Soundscapes using Real-time Manipulation of Locally Relevant Birdsongs. In *Proceedings of the 2015 Web Audio Conference* (Paris, France, January 26-28). IRCAM, Paris. http://wac.ircam.fr/pdf/wac15_submission_22.pdf

[5] Walker, W., Hebel, K., Martirano, S., and Scaletti, C. 1992. ImprovisationBuilder: Improvisation as Conversation. In *Proceedings of the 1992 International Computer Music Conference* (San Jose, CA, October 14-18). International Computer Music Association, San Francisco, CA, 190-93.