Journal of New Music Research

ISSN: 0929-8215 (Print) 1744-5027 (Online) Journal homepage: http://www.tandfonline.com/loi/nnmr20

World Stage: Crowdsourcing Paradigm for Expressive Social Mobile Music

Ge Wang, Spencer Salazar, Jieun Oh & Robert Hamilton

To cite this article: Ge Wang, Spencer Salazar, Jieun Oh & Robert Hamilton (2015) World Stage: Crowdsourcing Paradigm for Expressive Social Mobile Music, Journal of New Music Research, 44:2, 112-128, DOI: 10.1080/09298215.2014.991739

To link to this article: http://dx.doi.org/10.1080/09298215.2014.991739

Journal of New Music Research, 2015
Vol. 44, No. 2, 112–128, http://dx.doi.org/10.1080/09298215.2014.991739

Routledge
Taylor & Francis Group

World Stage: Crowdsourcing Paradigm for Expressive Social Mobile Music

Ge Wang, Spencer Salazar, Jieun Oh and Robert Hamilton

Stanford University, CA, USA

(Received 26 January 2014; accepted 20 November 2014)

Abstract

The combination of powerful mobile devices and the connective potential of cloud-based computing are changing how, where, and when people use computers. No longer physically tethered, computers embedded in mobile phones and tablets freely roam in daily life alongside their human users while persistently connected to the network. At the same time, the pervasiveness of computing enables us to engage people to solve large-scale problems by leveraging human intelligence and the effect of large crowds. In this paper, we discuss a new model for crowdsourced musical interactions based on mobile devices. This model, which we call World Stage, is a pseudo-anonymous, location-aware ecosystem designed to connect hundreds of thousands of users in a social-musical game involving expressive musical performance and collaborative musical feedback. We describe the motivation and mechanics of the World Stage, and present a full implementation and case study around a commercial iPhone application: Smule's Leaf Trombone: World Stage. We also present the experiential design and the technological infrastructure around the World Stage and discuss the unique social/musical possibilities it affords. This work is about exploring, perhaps for the very first time, a crowdsourcing eco-system that incorporates expressive music-making and with game-like elements, aimed at inciting a mass audience.

Keywords: social aspects, performance, instruments, interfaces, audience, expression

'All the world's a stage, and all the men and women merely players...' – Shakespeare

1. Introduction

The age of mobile, ubiquitous computing is well underway. Powerful computers embedded in networked, mobile personal devices are transforming how, where, and when people use computing technologies. No longer tethered to the desktop, mobile phones and tablet computers roam with their human users through everyday life, while persistently connected to the network and aware of their location.

The most profound trend of this age may be the increasing pervasiveness of computers that yet remain obscured from our awareness and consciousness (Weiser, 1991). Computing is becoming less deliberate, less static, and increasingly ad hoc, on the fly, and context-sensitive to physical interaction, location, and social relationships. As computing paradigms evolve, opportunities arise for exploring new musical interactions that take advantage of the scale and pervasiveness of new technology, leading to musical experiences that are not only 'mobile', but also were perhaps never possible before.

This work was also inspired by the idea of: *human computation*. With pervasive mobile computing changing the way people use technology, new possibilities arise for leveraging people to solve tasks that are traditionally hard for computers, but easy for humans. For example, the notions around Human Computation (Von Ahn, 2005), Game With A Purpose (Von Ahn & Dabbish, 2004), and cloud-based services like Amazon Mechanical Turk (2014) place humans into the problem solving loop, and position technology to take advantage of human intelligence, expertise, and judgment (for example, labelling and identifying objects in photos, transcribing audio/music, researching data details). Given the undeniably human aspects of music (aesthetic, emotion, intent,

Correspondence: Professor Ge Wang, Stanford University, Center for Computer Research in Music and Acoustics (CCRMA) 660 Lomita Dr., Stanford, CA 94305, USA. E-mail: ge@ccrma.stanford.edu

expressiveness, etc.), might we explore how technology can leverage our musical intuition and judgment for new musical interactions? How might we use these ideas to infuse a human element into social-musical experience – for example, facilitating users to serve as judges to provide crucial feedback that naturally feels human?

1.1 What is the World Stage?

The World Stage (Figure 1) provides a massively social, game-like musical experience, leveraging the ubiquitous nature of mobile devices and ideas around human computation. It provides a platform where users can create musical content, perform using a mobile phone-based instrument, and present that performance to the greater community. As part of a social game, these users are then recruited to serve as juries in online, interactive judging sessions, giving feedback and ratings to the performers. This creates an ecosystem where a participant can take on the role of performer, judge, observer, and/or composer. The musical content, performances, and critique are all crowdsourced – that is, they originate from a large community of users, whose creative human output is brought to bear through technology. This paradigm is fully implemented and deployed in an iPhone application, Smule's Leaf Trombone: World Stage, which over its lifetime has had more than 800,000 users. We will use Leaf Trombone: World Stage as the primary case study to show the World Stage at work in practice, and as a way to evaluate its possibilities.

1.2 Why the World Stage?

The World Stage was conceptualized and created for one purpose: to leverage the collective intelligence and judgment of the crowd to provide musical critique and feedback for user-generated performances.

World Stage was motivated by the realization that it is difficult for computers to give 'deep' feedback on a musical performance, especially when it involves attributes such as expressiveness, virtuosity, musicality, and potentially unexpected qualities such as surprise and irony (e.g. it seems difficult for a computer to decide whether a performance is 'funny' or 'ironic' or 'epic' on some level). Yet, it is easy for a human to naturally grasp and evaluate these features, even if the human is not a trained musician. We can recognize and appreciate virtuosity without actually being virtuosos. We are able to judge whether we like or dislike a piece of music, even if we cannot quite say why. We cannot help but detect humour or irony. There is a certain universality that pervades our appreciation, understanding, and judgment of music – something that is innately human. In a sense, the World Stage is an attempt to use technology to put our human judgment and aesthetics into the experiential loop, enabling new large-scale social-musical interactions. To our knowledge, no one has previously created and deployed such a system – this was an experiment to see if it can be done and what would happen.

2. Related work

Researchers and new media artists have explored mobile phones for music-making long before the proliferation of app-based mobile devices like the iPhone and iPad. In this section we discuss related works that cover early mobile music research, to mobile music performance, collaboration and network music, and crowdsourcing.

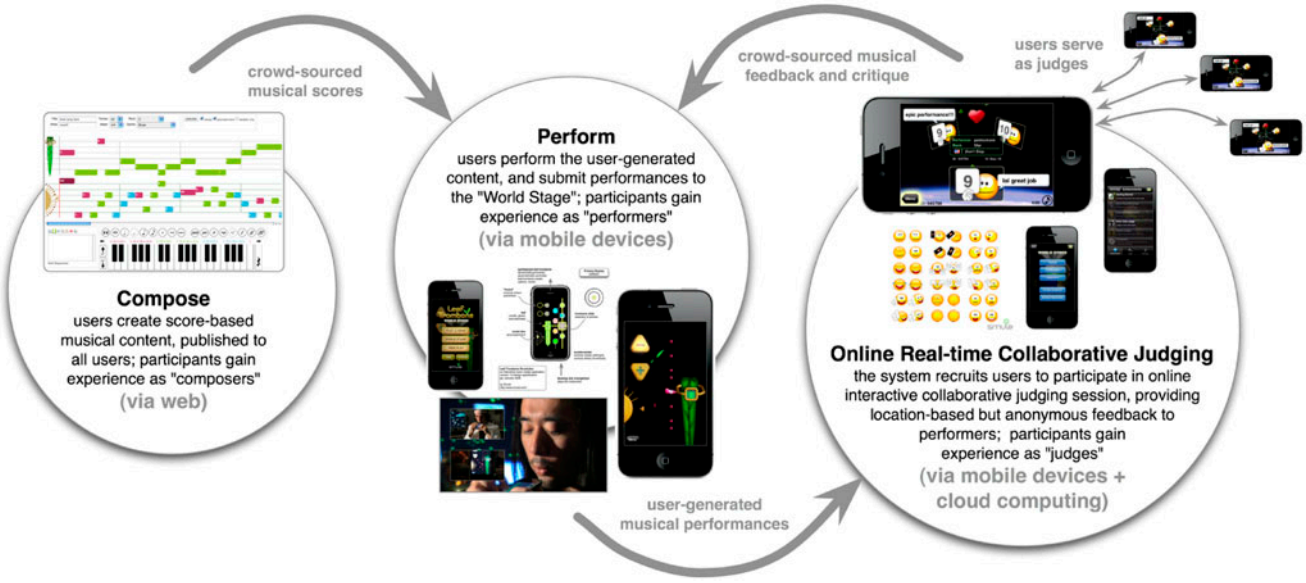

Fig. 1. The mechanics of the World Stage, combining elements of crowdsourcing, cloud-computing, and human computation to create a musical ecosystem where users take on symbiotic roles of composers, performers, and judges.

2.1 The rise of mobile music

An emerging community of mobile music was described in 2004 (Tanaka, 2004), and later in 2006 (Gaye, Holmquist, Behrendt, & Tanaka, 2006), in the context of New Instruments for Musical Expression (NIME). The first sound synthesis on mobile phones can be traced back to Geiger's PDa (Geiger, 2003), Pocket Gamelan (Schiemer & Havryliv, 2006), in which performers physically swung drone-generating mobile phones on tethers around their bodies, and Mobile STK (Essl & Rohs, 2006), a port of the Synthesis Toolkit (Cook & Scavone, 1999) to the SymbianOS mobile platform. Essl and Rohs explored the potential of mobile devices for music creation with Shamus and Camus2, both leveraging onboard sensors (e.g. camera and accelerometers) to create interactive mobile music instruments (Essl & Rohs, 2007; Rohs & Essl, 2007). The notions of interactivity, music expression, and physicality pervade their works (Essl & Rohs, 2009).

The musical interaction and physical interaction of Trombone were greatly influenced from experiences building musical instruments for the laptop in laptop orchestras since 2005 (Smallwood, Trueman, Cook, & Wang, 2008; Trueman, 2007; Wang, Bryan, Oh, & Hamilton, 2009), and later the mobile phone orchestra since 2008 (Wang, Essl, & Pentinnen, 2014; Bryan, Herrera, Oh, & Wang, 2010; Essl & Müller, 2010; Oh, Herrera, Bryan, Dahl, & Wang, 2010; Wang, Essl, & Pentinnen, 2014). This particular school of design emphasizes expressiveness, physical interaction design, and embraces the unique capabilities and limitations of commodity computing devices and their various sensors.

2.2 iPhone and Smule

Apple Inc. introduced the iPhone in 2007 and the iPhone SDK in 2008, forever altering the mobile device landscape. The iPhone, and subsequent mobile software platforms, thoroughly popularized the 'app' as a fundamental building block, with each app transforming the hardware mobile device into something new (e.g. a book, a camera, a game, a musical instrument). It also opened the doors for developers to easily create applications and distribute them to the masses through various 'app stores'. By 2012, more than 700 million new units of smartphones were being sold yearly, while more than a million distinct apps were available across phones, tablets running Apple's iOS and Google's Android mobile operating systems.

For music, this opened the doors for startups such as Smule (founded in 2008; Wang et al., 2009) to explore expressive mobile music, as well as social musical interactions through mobile devices. Ocarina, released in 2008 (Wang, 2014a), was one of the earliest mobile-musical and social-musical apps in a new era of app-based mobile computing. Ocarina presented breath-controlled, flute-like instrument designed for the iPhone. Users blow into the microphone of the iPhone, while using the multi-touch screen to control pitch and the accelerometer to control vibrato. Ocarina was also designed to be a social musical experience, providing a globe visualization that allows users to listen in on each other playing around the world. To date, Ocarina and its successor, Ocarina 2, have reached more than 10 millions users. Ocarina also served as precursor for Leaf Trombone: World Stage, both in terms of physical interaction and social interaction, which we shall discuss in depth in Section 3. Meanwhile, the creation of music-making apps necessitated ongoing investigations into data systems and representation for expressive mobile music (Hamilton, Smith, & Wang, 2011; Wang, Oh, & Lieber, 2011). Ocarina and Leaf Trombone were perhaps the first musical instruments designed to be easy to use by mainstream, non-expert users while still remaining expressive and interesting to proficient musicians (Wang, 2014b).

2.3 Networked musical performance

The wide-area, networked musical performance aspect of Leaf Trombone: World Stage has its roots even farther back, beginning with early network computer music ensembles such as The Hub (Lancaster, 1998), who explored the technology, methodology, and aesthetics of performing with interconnected computers. Networked musical performance (Kapur, Wang, Davidson, & Cook, 2005) is an ongoing investigation into the technology and art of live music performed by participants at many locations around the world. More recent networked musical performances have explored 'locative media' (Tanaka & Gemeinboeck, 2006), leveraging GPS and networking. Notable experiments and installations include Sonic City (Gaye, Mazé, & Holmquist, 2003), which cast the urban environment as a musical interface, GpsTunes (Strachan, Eslambolchilar, & Murray-Smith, 2005), using audio feedback to control navigation, and Net_Dérive (Tanaka & Gemeinboeck, 2008), where information collected via embedded sensors in participants on the streets was sent and visualized in a central art installation. Additional experiments and systems include Quintet.net (Hajdu, 2005), JackTrip (Caceres & Chafe, 2010), backend server and discovery mechanisms like DIAMOUSES (Alexandraki & Akoumianikis, 2010), and compass-based social mobile music (Tanaka, Valadon, & Berger, 2007).

Additionally, there have been a number of works exploring various aspects and interpretations of networked music. Kim-Boyle (2009) surveys a number of musical and aesthetic approaches to network music, including building instruments from the network itself, methods of audience feedback, and social interaction of participants. Makelberge (2012) addresses issues of creative autonomy and social interactions inherent in various technology-mediated music-making in wide-area network contexts, especially as related to sampling and mashups. While these are quite different in nature from Leaf Trombone World Stage, they nonetheless underscore the broadening possibilities of social networked music for a mass audience.

While the World Stage inherently requires a network (of people and machines), it operates in a different vein than these previous works on network music. For one, the World Stage

focuses the network not as a medium for the direct/real-time transmission of musical performance, but rather as a large-scale conduit/context for exchange of musical feedback as a form of social expression. Unlike the previous works, the World Stage was designed not only for musicians, but as an experience for the masses. Furthermore, the World Stage is a full ecosystem that includes content authoring, performance, judging, and observation/playback. It is about exploring, perhaps for the first time ever, a crowdsourcing ecosystem that incorporates expressive music-making and game-like elements, aimed at inciting a mass audience to engage musically.

2.4 Group-based music and crowdsourcing

Leaf Trombone: World Stage is an ecosystem that depends on a large crowd of users around the world playing a variety of roles, and leverages both aspects of crowdsourcing and game-like elements to power a complex social system. While both the nature and scale of Leaf Trombone's World Stage is unique to our knowledge, it draws inspiration from a number of works involving crowdsourcing and group-based musical performance. Golan Levin's DialTones is perhaps the earliest large group mobile phone performance (Levin, 2001). More recently, the aforementioned Mobile Phone Orchestra has explored music-making both in traditional ensemble settings as well as crowd-based settings (Kruge & Wang, 2011; Oh & Wang 2011a, 2011b). Gibson (2010) documents a system comprised of mobile devices that can instantiate performances in ad hoc physical locations; the system is scalable to the number of devices present and the music is intended to be created via social interaction. A key difference of this system from the World Stage is that the latter takes place in geographically disperse locations. Freeman and Godfrey (2010) crafted Flock, a multimedia work for saxophone quintet, dancers, video, and audience participation. CODES (Pimenta, Miletto, Flores, & Hoppe, 2011) synthesizes social network and network music concepts in a cooperative music creation tool aimed at novice musicians.

Leaf Trombone was also influenced by works using technology to unlock the computational capacities in large crowds of people (Von Ahn, 2005; Von Ahn & Dabbish, 2004). In these works, humans (or groups thereof) are put in a position to solve problems traditionally difficult for computers. Computers play an essential yet fundamentally supporting role. In a similar vein, Leaf Trombone: World Stage uses mobile and cloud computing to provide a game-like environment where groups of humans both create and evaluate musical performances, providing nuanced feedback to performers. Yet, to our knowledge, the World Stage is the first instance of applying ideas from crowdsourcing for social interchange of musical performances and feedback on those performances.

In contrast to these related works, the World Stage can be further differentiated by a set of unique criteria and challenges. For one, can we create a viable ecosystem comprised of multiple meaningful roles (performer, judge, composer, observer), and motivate users to fulfill them and naturally achieve a sustainable equilibrium (e.g. balance the ratio of judges to performers)? Secondly, can the system achieve a massive scale, given the World Stage's design goal of bringing musical expression to mostly novice and casual users? Thirdly, is it compelling to have crowdsourced human judges, instead of algorithms, provide feedback to performers?

3. World Stage concept

This section describes the design process of the World Stage. Notably, we did not actually start with the intention to build a large-scale crowdsourced musical experience; it grew out of necessity as we worked on the Leaf Trombone instrument, finding it imperative to provide a 'more human' form of feedback for how well people played the Leaf Trombone as both a game and instrument.

3.1 Designing Leaf Trombone

Leaf Trombone was first conceived, without any notions of World Stage, in early 2009, in the months after the creation of Ocarina, itself a breath-controlled, wind instrument for the iPhone. An Ocarina web forum was provided where users could create musical scores in the form of Ocarina tablature. This forum, while immensely popular, was entirely separate from the Ocarina app. Users displayed tablature from the forum on a second device while playing on their iPhone.

Learning from Ocarina, two motivations in building a new musical instrument emerged: (1) to capture the physicality of playing the iPhone as an instrument while integrating the musical content directly into the application experience, and (2) to remain fully expressive while engaging new audiences (including novice musicians) with game-like elements.

As a playable instrument, the Leaf Trombone consists of a leaf-like trombone slide accompanied by an animated Music Box (Figures 2 and 3). Like Ocarina, Leaf Trombone utilizes breath control for note onset and amplitude enveloping. Unlike Ocarina, it is also capable of smooth glissandi and features a pitch range of three octaves (changeable via two onscreen buttons). The placement and spatial mapping of pitches went through several iterations, aiming to maximize the limited screen-space and balance expressiveness and playability, with the challenge of screen visibility when holding the devices at a skewed angle near one's mouth. A vertical chromatic-pitch mapping was ultimately used, with markers denoting diatonic pitches. Three degrees of 'snap-to-pitch' were provided to aid users, which nonlinearly warped the continuous mapping space to interpolate towards nearest chromatic pitches, while still allowing smooth glissandi.

In game-play mode, players use the instrument to perform musical selections. As they play, animated scrolling leaf markers prompt where to position the Leaf Trombone slide next and when to articulate via breath. The music box provides pre-composed musical accompaniment, timed with the leaf markers. This mode integrated the score component of

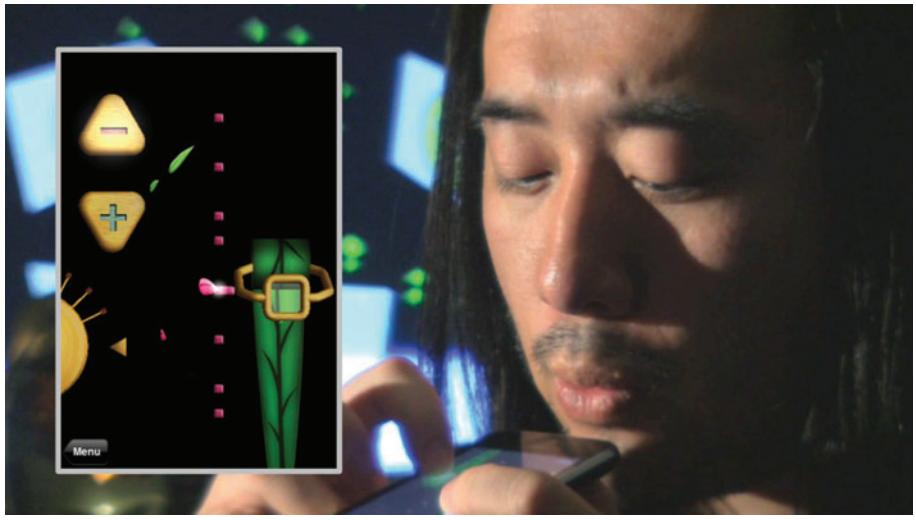

Fig. 2. Leaf Trombone instrument in action. Animated leaves hint where to position the green trombone slide and when to blow into the instrument, while expression is left to the user.

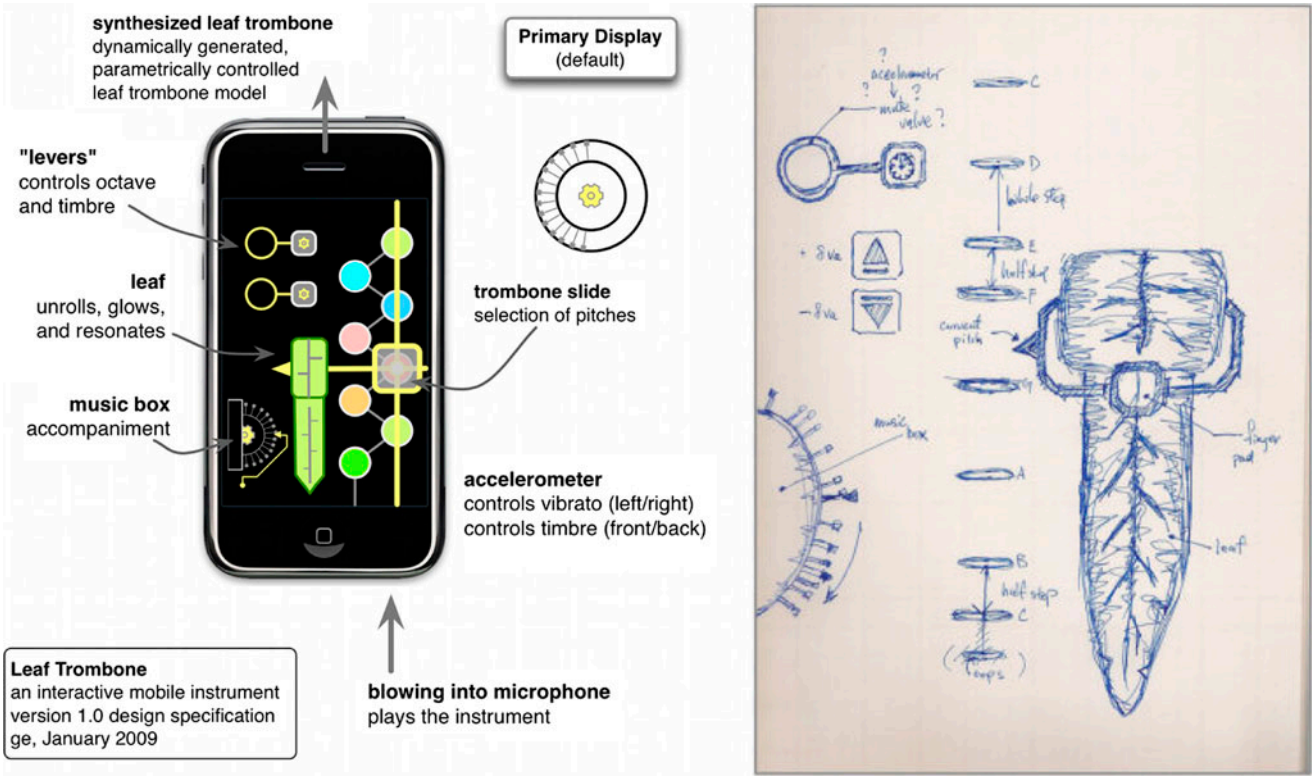

Fig. 3. Left: an early design of Leaf Trombone instrument; core components include leaf as trombone slide, blowing into microphone, music box accompaniment, physical model to synthesize the trombone sound, and game-like elements to prompt upcoming score; Right: a later design sketch that more closely resembles the final interface.

Ocarina's web forum into a single game-like experience. This is similar in broad concept to games like Guitar Hero (2005) and Rock Band (2007), but with one key difference: the audio generated in Leaf Trombone are always created by user interaction, and the user is free to express and embellish – there is no pre-recorded sound. In this context, it is helpful to think of Guitar Hero and Rock Band as games that 'masquerade' as instruments, whereas Leaf Trombone is designed to be an expressive musical instrument, augmented with game-like qualities.

3.2 Sound design

Leaf Trombone's audio is synthesized via relatively straightforward means in the ChucK programming language (Wang 2008); a two-operator FM patch controlled by a modulating and carrier envelope generates the brass-like sound of the trombone, while the music box accompaniment is produced by resonant-filtered impulses (Figure 4). The aural effect is that one is not playing an authentically simulated trombone, but rather a fun, cartoon-like brass instrument. The whimsical nature of the instrument is augmented by the ease of glissandi, pitch bending, and register-jumping. Performers are able to embellish melodies with grand musical gestures, which, in Leaf Trombone's synthetic brass blare, come across as both skillful and comical. The instrument similarly imbues poor performances with a sense of humour – missing a note somehow seems funnier when a performer then smoothly glides to the correct note.

3.3 Origin of the World Stage

A difficult problem presented itself at this point in the design process. We wanted Leaf Trombone to provide some type of rating or feedback to the user, and yet we hesitated to have the software automatically judge a performance. On one hand, it seemed straightforward to program the Leaf Trombone to track the 'accuracy' of how a piece was played (e.g. how many notes were 'correctly' played). However, given the whimsical, expressive, and often goofy personality of the Leaf Trombone instrument, there seemed to exist more interesting criteria in judging the value of a performance – qualities like expressiveness, surprise, context, and irony (for example, using Leaf Trombone to play Beethoven's *Ode to Joy* would likely require a very different mindset than playing, say, Europe's *Final Countdown*). Yet, these attributes seemed fundamentally difficult for a computer to judge.

From this observation, we attempted a new approach. We asked: instead of trying to use computers to solve these fundamentally hard problems, what if we leveraged the computer (and perhaps a network of computers) to connect and incorporate human intelligence and musical judgment into the system? After all, in judging Leaf Trombone performances, we are after the very human qualities surrounding a musical performance. It is easy and natural for people to grasp and evaluate these higher-level concepts (so natural it might be harder not to make a judgment) – we just need a way to harness this human musical intuition at a large scale. The realization of this concept soon developed into the World Stage, exploring how people inherently can make high quality (or at least interesting or nuanced) judges in the right conditions.

3.4 World Stage mechanics

After a player performs a song, he or she can submit it to the World Stage, where it will be judged and scored by a live panel of juries consisting of other Leaf Trombone players from around the world (Figure 5).

In each judging session, three judges are selected from the pool of Leaf Trombone: World Stage users. Their task is to listen, comment, and rate a particular performance. While the performance takes place before it is judged, the three judges participate in real-time relative to each other, and thus are able to influence one another. As the performance is played back, judges are able to express themselves using textual commentary and emoticons, ranging from enthusiastic approval to boredom and bafflement (Figures 6 and 7). Notably, the emoticons are designed as purely expressive commentary and are not factored into a performance's final rating. At the end of each performance, each judge gives a score of 1 (weak) to 10 (strong). The judging sessions are fully recorded, stored on the central Leaf Trombone server, and available to the performer (and potentially any other user) for playback.

It was the exploration in physical interaction design and game-design that led to the creation of the World Stage. In the next section, we will look at the ecosystem and social model of the World Stage in depth.

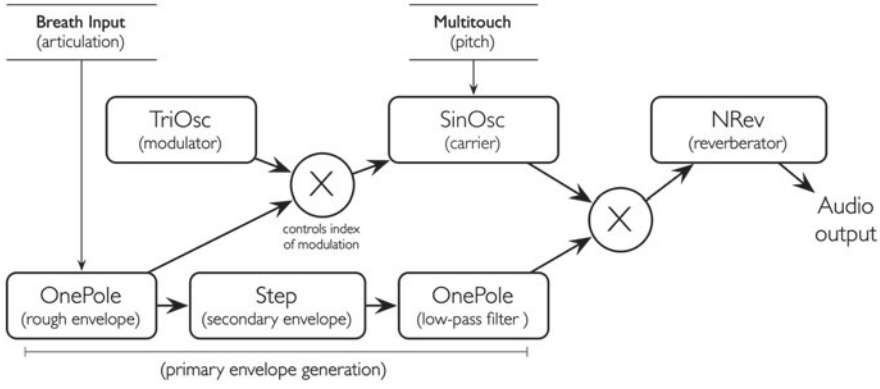

Fig. 4. ChucK block diagram for synthesizing the Leaf Trombone. In the two-operator FM setup, both the carrier oscillator and modulator oscillator are enveloped by breath input, thereby increasing the spectral brightness with the instrument's attack and creating a distinctively brass-like tone.

Fig. 5. Initial concept drawing for World Stage, including geo-location, three entities representing a panel of judges assembled from users (visualized on top), and the performer (visualized on the bottom).

4. Large-scale social model

4.1 Roles in the World Stage ecosystem

The World Stage as an ecosystem depends on several symbiotic roles that mutually reinforce each other. A member of this ecosystem can take on any and multiple roles.

Composer – any user can take on the role of a composer by creating and publishing musical content for others to perform. This is accomplished using a web-based interface (discussed in Section 4.5) that publishes directly to the application. A composer's unique World Stage username is credited with the work, allowing him or her to track the work's popularity. The overwhelming majority of songs available in Leaf Trombone: World Stage is created by composers from the community.

Performer – a user can play any track created by composers (in fact, the majority of users will at some point do this). A user is considered a performer after deciding to submit a performance to be judged on the World Stage. The performer is credited with that performance, and can gain feedback from judges and any observers of the performance. Over time, a performer can earn promotions in in-game rank, which is displayed in the user's profile information as a measure of his or her overall experience (see Figure 8).

Judge – judges are recruited by the World Stage to provide feedback and a final score to a performance. The World

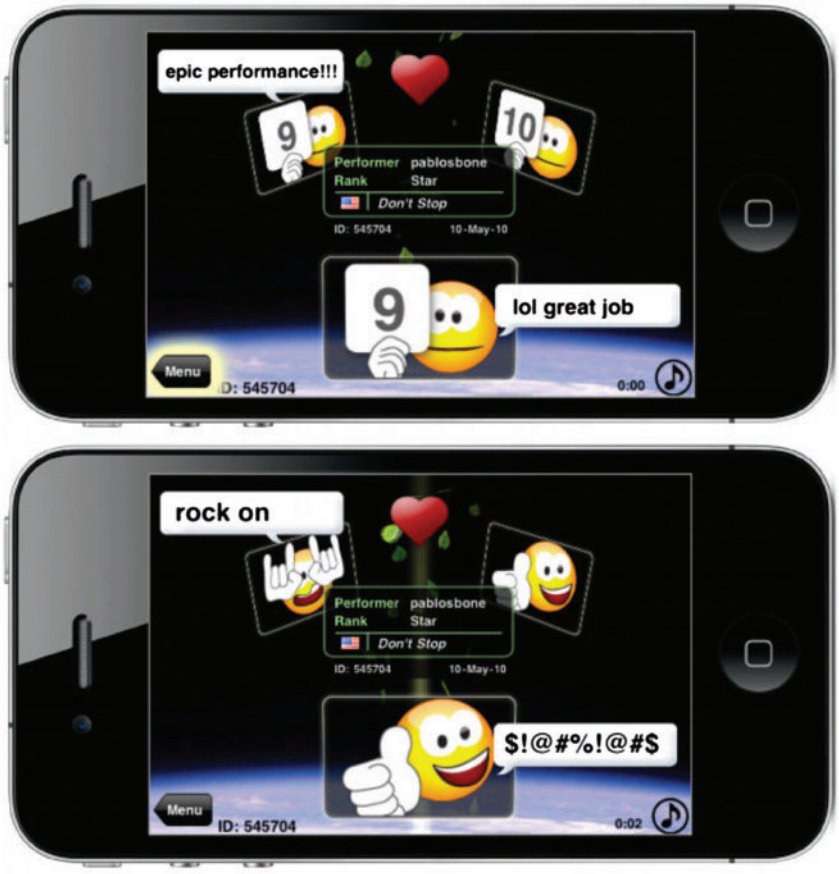

Fig. 6. Top: a panel of three online users judges a Leaf Trombone performance. The judges can confer and emote in real-time with fellow judges. Bottom: at the end of each judging sessions, the judges each provide a score on the scale of 1 (lowest) to 10 (highest).

Stage strives to maintain equilibrium between the number of performers and judges. Like performers, judges can earn promotions in rank.

Observer – a user can observe the full recording of any existing performance that has been judged, indexed by unique performance number. As an observer, no direct feedback can be added, though the observer can 'heart' a performance to show his or her support. Users can take on this role to view performances associated with a particular performer, judge, or song.

4.2 Balancing roles

In Leaf Trombone: World Stage, performing for the World Stage is meant to be a recurring but valuable opportunity. Users are encouraged to practice a song as much as they would like, but they have to expend a special 'performance token' each time they submit a performance to the World Stage, which both opens the possibility for other users to listen and give feedback on the performance, and allows the performer to progress through the game by gaining experience, levels, and renown. Performance tokens can be earned over time or by serving 'jury duty' (by participating as a World Stage judge). This creates a balance where users are motivated to both perform and critique. This approach works well – by adjusting the value of performance tokens, we can encourage judging or performances as needed, maintaining equilibrium in the system. In the World Stage, we arrived at a ratio of 2:1 (i.e. each player has to judge twice to have a chance to perform) through experimentation.

4.3 Anonymity and location

The system regularly invites randomly chosen online users to serve as judges. Alternatively, a user can volunteer to judge at any time. The system is anonymous in that real names or other real-world identifying information is not presented to users in the World Stage. However, the following individualized information is conveyed to fellow judges and any observers: the geographic location of each judge (this is displayed in an introductory globe visualization), the nationality of each judge (visualized by a flag icon), the rank of the judge, and the 'disposition alignment' of the judge (likelihood of giving a high or low score, in three degrees: fierce, fair, or friendly). Additionally, the unique World Stage usernames of all judges and performers are shown.

Fig. 7. Judging emoticons make available the range of emotions that judges can express in real-time, and is an important (and heavily used) aspect of the judging experience, possibly due to the universal and language-neutral nature of emoticons.

Experience Ranks

Rank	Judges	Performers
1	Nobody	Amateur
2	Peanut Gallery	Student
3	Aficionado	Bar Performer
4	Academic	Backup Player
5	Panelist	Lounge Act
6	Critic	Expert
7	Arbiter	Prodigy
8	Authority	Star
9	Judge	Int. Sensation
10	Master Justice	People Know Me

Fig. 8. Rankings given to users based on experience in judging and performing.

This arrangement of partial anonymity was designed to enhance the social experience in a number of ways. Firstly, the absence of real-life identities can suppress much of the potential reticence a would-be performer may have, thereby lowering the inhibition for musical expression and increasing the possibility of participation. At the same time, linking one's World Stage activity to a virtual identity keeps users vested in their progress as performers and judges, as does maintaining judge and performer rankings. Displaying geographic and nationality information about judges and performers provides a given user additional insights into who else is joining them on the World Stage, giving rise to a basic familiarity among users. Anonymity in an online community can potentially lead to a prevalence of unpunished brash or offensive behaviour, but we found the World Stage community to be largely encouraging and friendly, eluding such deleterious effects (nonetheless, there exists functionality for users to report abuse).

4.4 Massively multi-player music game

Presenting the musical experience of the World Stage as a game seems to accomplish several objectives. The initial perception of an experience as a 'game' is designed to lower an individual's inhibition to start making music. Secondly, through game mechanics, we can motivate certain behaviours in an ecosystem consisting of hundreds of thousands of geographically distributed users. For example, the aforementioned token system balances the number of performances

and judging sessions. Additional game elements, such as an achievements system (e.g. 'Judge 100 times on the World Stage') and global leader-boards, encourage players to explore the various musical roles in the World Stage and provide a sense of accomplishment over time.

4.5 Crowdsourcing musical content

Leaf Trombone: World Stage's musical content is intended to be created and renewed by its user community. Each Leaf Trombone song consists of a score for the Leaf Trombone instrument with accompaniment from the Music Box. To date, more than 7000 user-generated scores have been published for use within the game. These scores are created using a web interface (Figure 9), which presents a notation system created specifically for the instrument. A user can compose by entering notes for the melodic trombone part and for the background accompaniment to be played on the Music Box. Scores can be previewed by listening to a MIDI version of the track, and saved to a user's online account. When it is completed to satisfaction, a user publishes the score, instantly making it available to all users of the game. Additionally, the score is simultaneously posted to the Leaf Trombone online forums for discussion and viewing. Modifications can be made by the original composer or other composers, and derivative works created from another user's score are tracked as such in the central database.

5. Implementation

The web-based composition interface, essentially a specialized MIDI composer, is implemented using the Google Web Toolkit, a high-level API for developing rich Internet applications. When a user finishes a composition, the specified sequence of notes for the melody and accompaniment (the 'score') is serialized into a custom data format built on top of JSON, a widely supported, cross-platform data meta-format. This representation allows a potentially broad array of devices, platforms, and application developers to interact with the World Stage.

The score data and related metadata are stored in a central database, managed by MySQL. The iPhone client application dynamically and regularly updates the catalogue of songs presented to the user by querying this database over the network. When a user selects a song to play, the iPhone client retrieves the appropriate score data from the database. These and other related network calls are formalized using an HTTP middleware layer implemented in Django, a Python-based web framework. Once downloaded by the client, the score data is parsed by an internal ChucK program. The system synthesizes

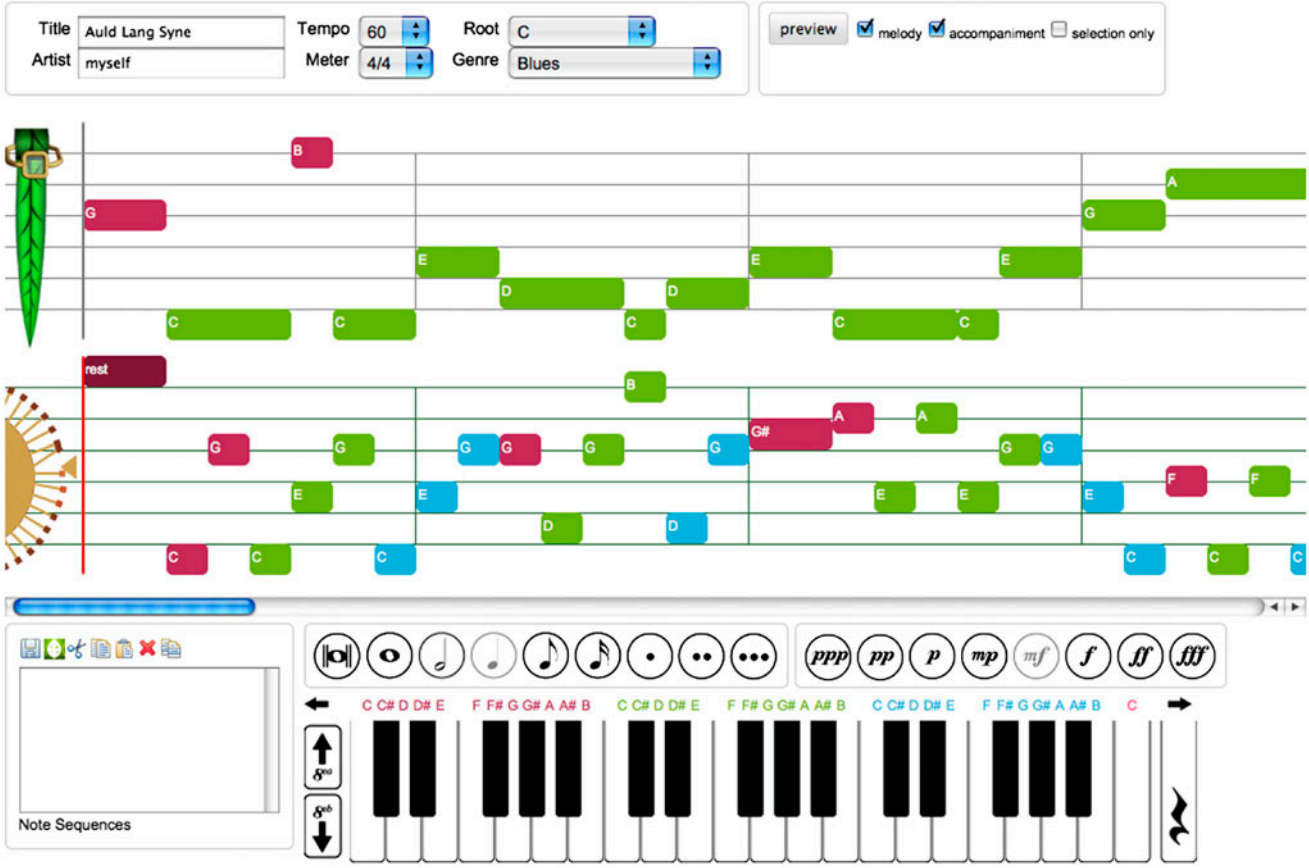

Fig. 9. Composing interface for Leaf Trombone and the accompanying music box. This web-based application publishes directly to the World Stage.

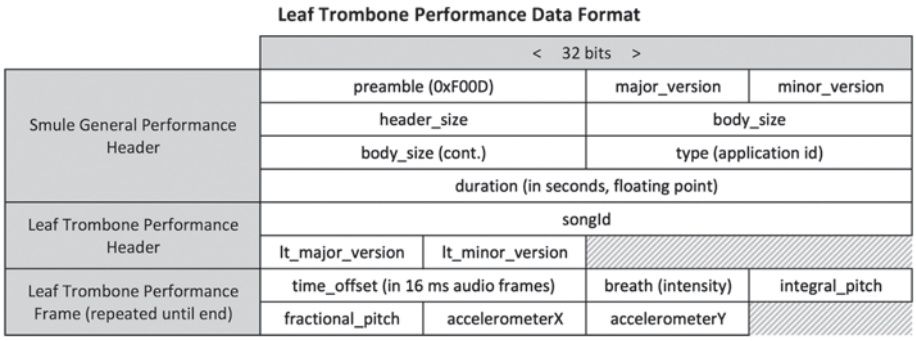

Fig. 10. The data format for a Leaf Trombone: World Stage performance. Each row represents 4 bytes (32 bits), with 1-, 2-, or 4-byte subdivisions thereof for individual data fields.

the Music Box accompaniment, visualizes graphically what notes the user should play and when to play them, records control data from the pitch slide and microphone 'blow' input, and synthesizes the corresponding Leaf Trombone audio.

The recorded control input, along with the original score file, represents a complete Leaf Trombone 'performance', and if the user desires, it is uploaded to the World Stage. Due to data size concerns, we used a custom binary data format for recording performances (Figure 10). Input is encoded in a series of packets, each representing the blowing intensity and pitch at a specified time offset from the beginning of the performance. When offered to the World Stage, this performance data is stored in the central database and becomes available for judging.

The World Stage judging experience, as a multi-modal real-time interaction between multiple networked human participants, requires a degree of responsiveness and statefulness difficult to attain using classical web programming environments such as PHP, Django, or Ruby on Rails. For this reason we were obliged to implement a custom stateful HTTP server to mediate between judges. When the user of the iPhone client opts to enter a judging session, the application registers its interest to the server through an initial network call. When three clients have registered in this way, the server groups them into a judging session and assigns to the session a performance over which to adjudicate. If no unjudged performances are available, the server selects from older performances that have already been judged.

Once three judges have been assembled, the session begins. The server sends each judge the performance data and score data, which is then played back through ChucK in loose synchrony across each participant's device. The client executes network calls to the server to register a judge's change in emoticon or textual comment. Throughout the session, the client also polls the server for emoticon and comment updates from the other judges. The judging server timestamps and records each emoticon and comment into the database, enabling future playback of judging sessions. Upon completion of the performance, each judge offers a final numerical rating and comment, which the server records and broadcasts to each other judge.

To play back a previous judging session, the iPhone client application retrieves from the database a serialized list of the time-stamped judge updates for the session, and replays these synchronously with the recorded performance and Music Box accompaniment. The sound is rendered from recorded gesture data, using the same real-time synthesis engine of the Leaf Trombone instrument.

6. Evaluation and discussion

6.1 Community

Leaf Trombone: World Stage has had over its lifetime an estimated base of more than 800,000 users on the iPhone and iPod Touch. Users have participated in nearly one million judging sessions. Users have generated more than 7000 individual songs, all of which are available to other users. The most played and judged selections include renditions of Auld Lang Syne, Zelda's Lullaby (from the video game, The Legend of Zelda: The Ocarina of Time), Titanic – The Portrait (from the motion picture *Titanic*), The Lion Sleeps Tonight, Amazing Grace, Hallelujah (by Leonard Cohen), Yesterday (by The Beatles), You Are My Sunshine, and the Star-Spangled Banner (Figures 11 and 12). While the experienced musician may find many of these selections simplistic, one must consider that the core demographic of Leaf Trombone users is not musically trained. For many users, a World Stage rendition of Twinkle Twinkle Little Star (the default song played in the app's tutorial) may be his or her first public musical performance. In this way, the World Stage presents new musical possibilities to users who would have previously had few chances to express themselves musically. More generally, the World Stage has broadened networked music making to a community of performers both casual and seasoned.

With regards to engagement with the World Stage, over 125,000 Leaf Trombone users have judged at least one performance, and 1326 users have judged 100 or more performances (see Figure 13). Nearly 80,000 Leaf Trombone users have submitted at least one performance to be judged on the World Stage, and about 1200 users have submitted 25 or more such performances. The most prolific composer in the World Stage

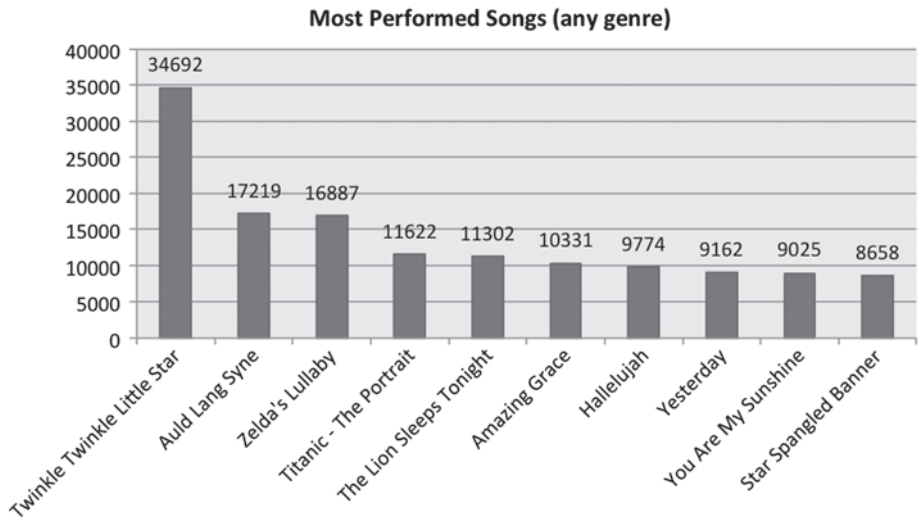

Fig. 11. Most performed songs, across all genres of music available in the app.

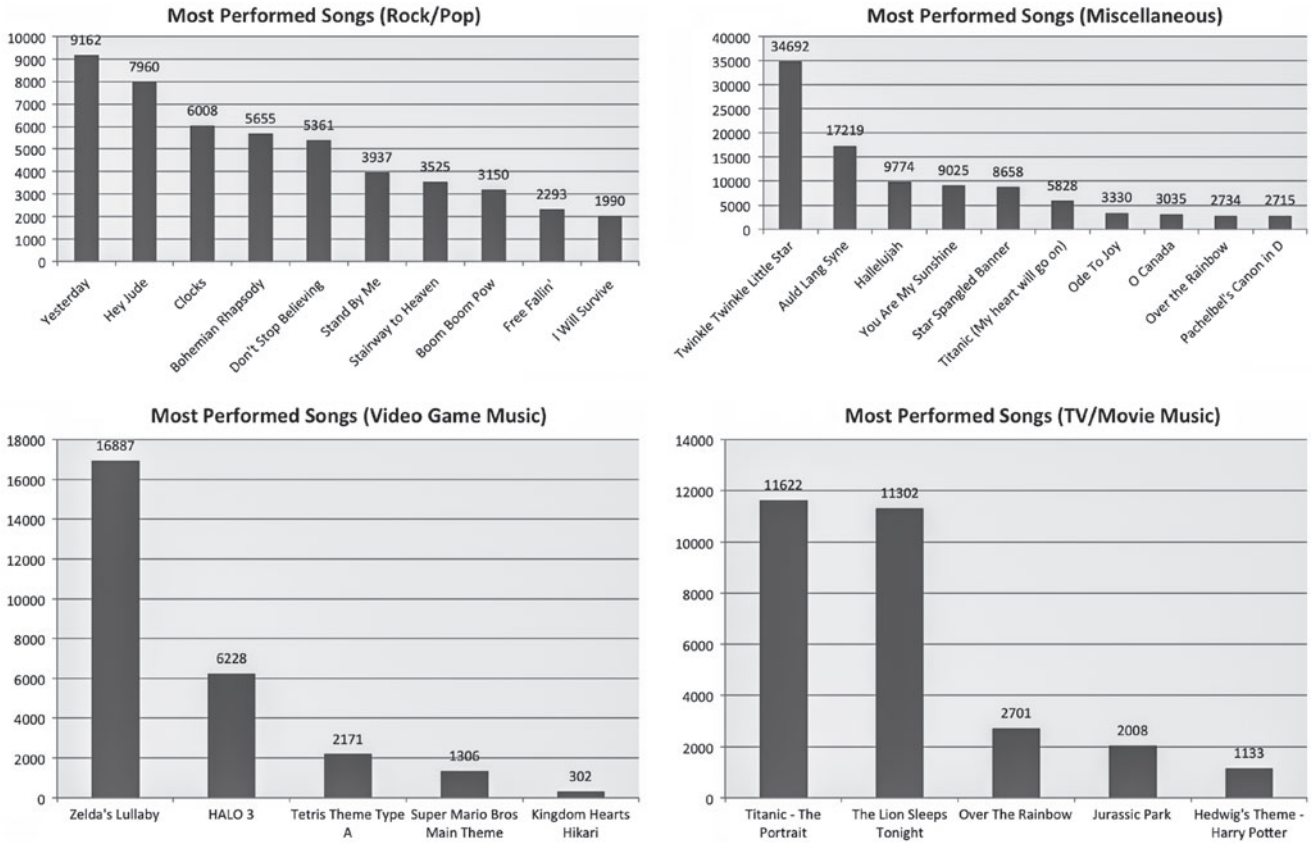

Fig. 12. Most performed songs within Leaf Trombone's four most popular genres (by number of performances of songs of that genre): Rock/Pop, Miscellaneous, Video Game Music, and TV/Movie Themes.

single-handedly contributed 177 different songs to the community; one particular user contributed 2639 performances to the World Stage, while, incredibly, one judge presided over 10,537 distinct sessions.

631,718 judging sessions proceeded on the World Stage. Within these, emoticon-based feedback was used 2,923,353 times (approximately 4.5 emoticons per session), and 1,865,634 text comments about the performances were shared (approximately three comments per session). We view these statistics as convincing of the level of engagement within the World Stage's real-time judging framework. The top 100 words or consecutive word sequences in the text comments

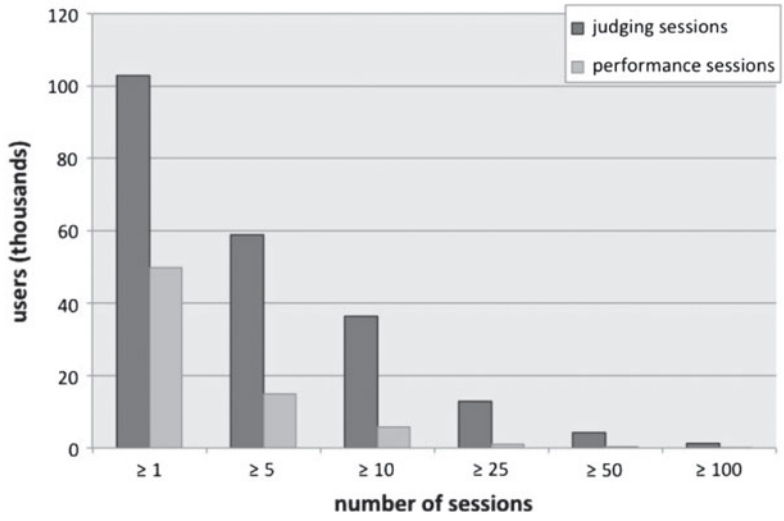

Fig. 13. Levels of engagement of Leaf Trombone performers and judges. For example, 21% of all Leaf Trombone users have judged at least one performance, 12% of all users have judged at least five performances, and 1% of all users have submitted at least 10 performances.

1	song	21	wtf	41	ya	61	hear	81	man
2	lol	22	off	42	notes	62	love this	82	did
3	hi	23	pretty	43	little	63	f*ck	83	horrible
4	bad	24	suck	44	job	64	yea	84	love this song
5	ok	25	cool	45	here	65	bit	85	need
6	hey	26	haha	46	yes	66	dude	86	best
7	this song	27	omg	47	i dont	67	a little	87	tune
8	nice	28	heard	48	sucks	68	keep	88	i hate
9	stop	29	play	49	long	69	needs	89	okay
10	better	30	i love	50	hate	70	never	90	much
11	oh	31	more	51	my ears	71	ouch	91	sorry
12	do	32	pretty good	52	please	72	same	92	u suck
13	dont	33	hello	53	hard	73	try	93	its ok
14	love	34	guys	54	rock	74	whats	94	good job
15	again	35	sounds	55	god	75	playing	95	sh*t
16	too	36	very	56	the song	76	right	96	where
17	great	37	practice	57	thats	77	guy	97	got
18	wow	38	awesome	58	why	78	hmm	98	sound
19	yeah	39	really	59	wat	79	yo	99	idk
20	not bad	40	ears	60	sup	80	song is	100	far

Fig. 14. Top 100 words or word sequences used in Leaf Trombone comments (excluding the 100 most common words in the English language). Some words in this table have been censored due to profanity.

are shown in Figure 14. (The 100 most frequent words in written English, according to the Oxford English Corpus[1], were excluded from this analysis, as were sequences of such words, common abbreviations and contractions, and alternate forms of the most frequent verbs. Additionally, words were considered equivalent even if letter case did not match, and apostrophes within words were discarded before testing equivalence.) Within this list, there is a spectrum of both positive and negative tones, varying to either direction in intensity. Despite the presence of needlessly critical commentary, a common trend in many pseudo-anonymous online communities, significant pockets of constructive feedback can be found in this table. This confirms the authors' anecdotal experience that the World Stage community is by and large a positive and welcoming one, and, in any case, has nuance and personality.

6.2 Visualization

To better understand and evaluate the user experience in the World Stage, we created a visualization that dynamically fetches recent judging sessions from the server and renders key features within them. The visualization was implemented as a desktop application using OpenGL and the C++ programming language.

The visualization elucidates the judging context by displaying the timestamp and latitude-longitude information of the performer and judges. The timestamp is shown relative to the present moment, such that viewers see how many minutes and

[1] http://www.oxforddictionaries.com/words/the-oec-facts-about-the-language

Fig. 15. The visualization of *Dichotomous Harmonies* during a performance.

seconds ago the judging session occurred (according to the records in the database). For displaying the location, we use reverse geo-coding to determine the city and country names corresponding to the latitude and longitude information, making the information more meaningful. Such information helps us better understand the user base, and to ask questions about emerging patterns relating the time of day, user locations and nationalities, and types of songs performed.

The actual performance that is being judged in a session can be triggered from within the visualizer, and re-synthesized and played back using ChucK. The synchronized audio and visuals give us a clearer view of judges' behaviours. Other features of the judging session, including the average score, emoticons used, and comments given out by the three judges, are also rendered using colourful textures in such a way that the visualization serves as an artistic and informative summary of what was actually experienced by users moments ago. The visualizer was also adapted for a crowdsourced musical performance.

6.3 *Dichotomous Harmonies*, a meta-composition

Performance data generated by World Stage users forms the basis of the musical meta-composition, *Dichotomous Harmonies*. Premiered in Milan, Italy, at the 2010 Milano-Torino Festival in a special concert event entitled *Play Your Phone!*, *Dichotomous Harmonies* combines a pre-rendered soundscape with a live improvisatory duet between a remote trombonist (located in Bournemouth, England) and an iPad performer. At the same time, a video artist controls the World Stage visualization engine with a customized set of performance excerpts, providing a visual and conceptual backdrop to the music (Figure 15).

The central unifying element of *Dichotomous Harmonies* is a melodic excerpt from musician Leonard Cohen's *Hallelujah*, combining a short verse section with a single chorus. Leaf Trombone performance data for one-thousand user renditions of this excerpt were extracted from Smule's database and mixed into an 8-channel soundscape. This rendering is played back while a live performer manipulates excerpts from the same set of performances via an iPad control interface to ChucK processes running on a laptop. The improvisatory element of the iPad performance is mirrored by a networked live trombone improvisation, lending clarity to the melodic material itself while reinforcing the disembodied nature of Cohen's haunting melodic line. Combining the live trombone with the massive number of World Stage performances creates a striking dichotomy between the tangible and intangible nature of melody, performance, and presence.

6.4 Conclusions

The World Stage is a social and musical experiment at an intersection of mobile music, cloud computing and human computation. The system is complete in design, implementation, and deployment. While fully implemented in Leaf Trombone, we believe the concept and architecture of the World Stage is applicable to other instruments and social-musical experiences.

Let us consider the criteria and challenges outlined in Section 2.4. The World Stage can be seen as a proof of concept for a new type of ecosystem with multiple roles that feed into one another. Composers (the smallest group) add content into the system; performers provide the primary World Stage commodity: performances; users serve as judges to give critique and feedback to each other; judging sessions are part of the World Stage 'public record' for any user (also known as observer) to playback and experience. Balancing such a system is less about engagement or how many users use it, and more about equilibrium. For example, we were able to create a stable and sustainable supply of judges by tweaking the performance token requirements to incentivize about three times as many people to judge than to perform.

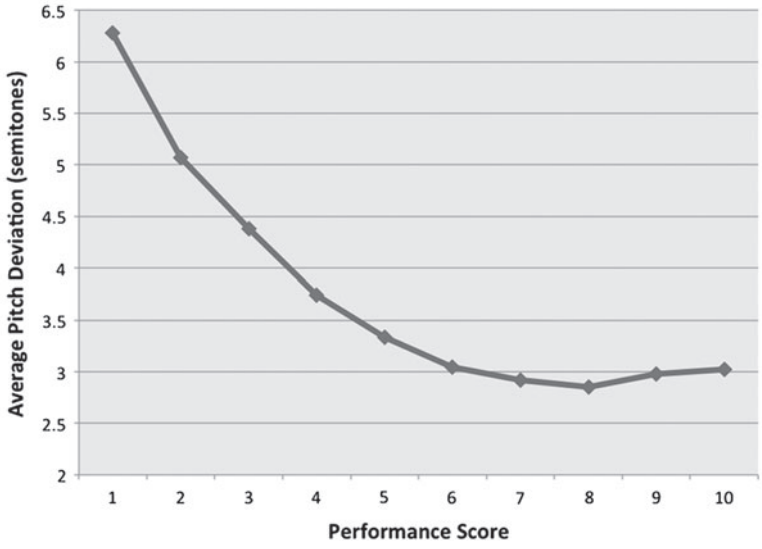

Fig. 16. Comparison of average pitch deviation (from the mean performance) of 15,000 renditions of Auld Lang Syne, grouped by average score of the performance.

The World Stage is a system designed to bring musical expression for mostly novice, casual users, and was envisioned to work with many users (tens of thousands or more). In the first year of the World Stage, more than 800,000 users tried the app, and of those 160,000 (21%) users were considered engaged users for having experienced both performer and judging roles. The system operated smoothly during this time and at various scales when the number of active users fluctuated.

The experiment of using crowdsourced human judges (instead of algorithms) to judge and provide feedback for the performances has been interesting. Perhaps first and foremost, the World Stage shows that it is possible to create an environment where users are compelled to participate both as novice performers and ad hoc judges. Experientially, and given the number of people who have contributed to the large numbers of judging sessions (and frequent users, including those who have judged hundreds of times or more), the cycle of performance and jury duty seems to be an activity that users engage with. Furthermore, the World Stage was not designed to be a forum for exchanging serious musical critique, and from the casual and often amusing nature of feedback from judges, the World Stage suggests the simple idea that it is nonetheless valuable for a user to have someone, 'somewhere out there' listen to a performance they have made and react to it.

An additional takeaway from the World Stage initiative is that incorporating human intelligence into a social-musical system can be powerful, as demonstrated by abandoning computer-based judging in favour of more crowdsourced, human participation. Additionally, by adding game elements to expressive musical experiences, we can motivate more musical participation and engagement. We can view one interpretation of the results of this experiment in Figure 16. In this chart, we have used pitch deviation of a given World Stage performance from the mean performance of that song as a correlate for musical variation within that performance. Pitch deviation for a set of performances M is calculated according to the formula:

$$\bar{D}_M = \frac{1}{|N|} \sum_{n \in N} \sqrt[n]{\frac{1}{|M_n|} \sum_{m \in M_n} (x_{m,n} - \bar{x}_n)^2},$$

where N is a set of discrete-time steps that encompasses the maximum duration of all performances under consideration, M_n is the subset of M that have a sample at time n (in case the performer didn't play to the end of the song), $x_{m,n}$ is the pitch of performance m at time n, and $\bar{x}_n$ is the mean pitch across all M_n at time n. Additionally, a pitch sample for performance m is not considered (i.e. it is excluded from M_n) if the performance's amplitude is below perceptible levels. Overall this metric can be thought of as song set M's average deviation in pitch at any given point in the song. A higher deviation indicates more variation, and vice versa – this variation can be thought of as the mean difference from a statistically 'perfect' performance, so even canonically virtuosic gestures, like vibrato on longer notes, are penalized under this model.

Comparing the average pitch deviation of 15,000 performances of *Auld Lang Syne* to the average rating of each of these performances, we can see a correspondence between user appraisal of musical content and musical variation. The trend roughly ascribes poorer evaluations to more varying performances. The slight upswing in deviation for 9- and 10-rated performances (scored out of 10) could be explained as statistical noise, but the trend does suggest that 'perfection' is not the only criterion for a judge's evaluation on the World Stage. While far from conclusive, these results are insightful into the potential of human computation for large-scale music performance appraisal.

Another idea not to be overlooked is that anonymous interactions can be more powerful than ones where identities are known. For example, performers may be more likely to submit performances to be judged by other users if they know the feedback is anonymous. While the validity of this point deserves more investigation, it seems clear the World Stage operates as an ecosystem without the need for traditional identities.

Finally, this experiment views crowdsourcing as marriage between technology and masses of people. We believe this is only a beginning. By using the technology to set the right conditions, we can take deeper advantage of human intelligence, judgment, and intuition in creating new musical experiences.

Acknowledgements

This work has been the results of many people working together, and is indebted to Jennifer Wu, David Zhu, Mattias Ljungstrom, Tricia Heath, Arnaud Berry, Jeffery Smith, Mike Earhart, Tony Ventrice, Ryan Pfenninger, Turner Kirk, Jonathan Berger, Elon Berger, Michael Wang, Stefan Kotes, and the entire Smule team. Special thanks to Clara Valenstein for analytics and analysis, David Kerr for support and suggestions, and Gina Collecchia for insights into data analysis and signal processing. We would also like to thank the journal reviewers and editor for their feedback in improving and strengthening this article.

References

Alexandraki, C., & Akoumianakis, D. (2010). Exploring new perspectives in network music performance: The DIAMOUSES framework. *Computer Music Journal, 34*(2), 66–83.

Amazon Mechanical Turk. (2014). http://aws.amazon.com/mturk/ Extracted January 2014

Bryan, N., Herrera, J., Oh, J. & Wang, G. (2010). MoMu: A mobile music toolkit. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Sydney, Australia, pp. 174–177.

Caceres, J., & Chafe, C. (2010). JackTrip: Under the hood of an engine for network audio. *Journal of New Music Research, 39*(3), 183–187.

Cook, P.R. & Scavone, G. (1999). The Synthesis Toolkit (STK). *Proceedings of the International Computer Music Conference*, Beijing, China, pp. 164–166.

Essl, G. & Rohs, M. (2006). Mobile STK for Symbian OS. In *Proceedings of the International Computer Music Conference*, New Orleans, LA, pp. 278–281.

Essl, G. & Rohs, M. (2007). ShaMus - A sensor-based integrated mobile phone instrument. In *Proceedings of the International Computer Music Conference*, Copenhagen, pp. 200–203.

Essl, G., & Rohs, M. (2009). Interactivity for mobile music making. *Organised Sound, 14*(2), 197–207.

Essl, G. & Müller, A. (2010). Designing mobile musical instruments and environments with urMus. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Sydney, Australia, pp. 76–81.

Farago, P. (2012). App engagement: The matrix reloaded. *Flurry Insights*. http://www.flurry.com/bid/90743/App-Engagement-The-Matrix-Reloaded

Freeman, J., & Godfrey, M. (2010). Creative collaboration between audiences and musicians in Flock. *Digital Creativity, 21*(2), 85–99.

Gaye, L., Mazé, R. & Holmquist, L.E. (2003). Sonic City: The urban environment as a musical interface. *Proceedings of the International Conference on New Interfaces for Musical Expression*, Montreal, Canada, pp. 109–115.

Gaye, L., Holmquist, L.E., Behrendt, F. & Tanaka, A. (2006). Mobile music technology: Report on an emerging community. In *Proceedings of the International Conference on New Instruments for Musical Expression*, Paris, France, pp. 22–25.

Geiger, G. (2003). PDa: Real time signal processing and sound generation on handheld devices. In *Proceedings of the International Computer Music Conference*, Singapore, pp. 283–286.

Gibson, I. (2010). Collaborative music software for mobile phones. *International Journal of Research and Reviews in Applied Sciences, 2*(2), 104–107.

Hajdu, G. (2005). Quintet.net: An environment for composing and performing music on the Internet. *Leonardo, 38*(1), 23–30.

Hamilton, R., Smith, J., & Wang, G. (2011). Social composition: Musical data systems for expressive mobile music. *Leonardo Music Journal, 21*, 57–64.

Kapur, A., Wang, G., Davidson, P., & Cook, P. (2005). Interactive network media: A dream worth dreaming? *Organized Sound, 10*(3), 209–219.

Kim-Boyle, D. (2009). Network musics: Play, engagement and the democratization of performance. *Contemporary Music Review, 28*(4–5), 363–375.

Kruge, N. & Wang, G. (2011). MadPad: A crowdsourcing system for audiovisual sampling. *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, pp. 185–190.

Lancaster, S. (1998). The aesthetics and history of the hub: The effects of changing technology on network computer music. *Leonardo Music Journal, 8*, 39–44.

Levin, G. (2001). *Dialtones – a telesymphony*. http://www.flong.com/projects/telesymphony/

Makelberge, N. (2012). Rethinking collaboration in networked music. *Organised Sound, 17*(1), 28–35.

Oh, J., Herrera, H., Bryan, N., Dahl, L. & Wang, G. (2010). Evolving the mobile phone orchestra. *Proceedings of the International Conference on New Interfaces for Musical Expression*, Sydney, Australia, pp. 82–87.

Oh, J. & Wang, G. (2011a). Audience participation in mobile music performance. *Proceedings of the International Computer Music Conference*, Huddersfield, UK, pp. 665–671.

Oh, J., & Wang, G. (2011b). Converge: An omni-biographical composition. *Computer Music Journal Emille, 9*, 71–78.

Pimenta, M., Miletto, E., Flores, L., & Hoppe, A. (2011). Cooperative mechanisms for networked music. *Future Generation Computer Systems, 27*(1), 100–108.

Rohs, M. & Essl, G. (2007). Camus2 – Collaborative music performance with mobile camera phones. *Proceedings of the International Conference on Advances in Computer Entertainment Technology*, Salzburg, Austria, pp. 190–195.

Schiemer, G. & Havryliv, M. (2006). Pocket Gamelan: Tuneable trajectories for flying sources in Mandala 3 and Mandala 4. *Proceedings of the International Conference on New Interfaces for Musical Expression*, Paris, France.

Smallwood, S., Trueman, D., Cook, P. R., & Wang, G. (2008). Composing for laptop orchestra. *Computer Music Journal, 32*(1), 9–25.

Strachan, S., Eslambolchilar, P. & Murray-Smith, R. (2005). GpsTunes: Controlling navigation via audio feedback. *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*, pp. 275–278.

Tanaka, A., (2004). Mobile music making. *Proceedings of the International Conference on New Interfaces for Musical Expression*, Montreal, Canada, pp. 154–156.

Tanaka, A. & Gemeinboeck, P. (2006). A framework for spatial interaction in locative media. *Proceedings of the International Conference on New Interfaces for Musical Expression*, Paris, France, pp. 26–30.

Tanaka, A., Valadon, G. & Berger, C. (2007). Social mobile music navigation using the Compass. *Proceedings of the International Mobile Music Workshop*, pp. 34–37.

Tanaka, A., & Gemeinboeck, P. (2008). Net–Dérive: Conceiving and producing a locative media artwork. In G. Goggins, & L. Hjorth (Eds.), *Mobile technologies: From telecommunications to media*. London: Routledge.

Trueman, D. (2007). Why a laptop orchestra? *Organised Sound, 12*(2), 171–179.

Von Ahn, L. (2005). Human computation (PhD thesis), PA: Carnegie Mellon University.

Von Ahn, L. & Dabbish, L. (2004). Labeling images with a computer game. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. *Association for Computing Machinery*, pp. 319–326.

Wang, G. (2008). *The ChucK audio programming language* Ph.D thesis, Princeton University.

Wang, G. (2014a). Ocarina: Designing the iPhone's Magic Flute. *Computer Music Journal, 38*, 2.

Wang, G. (2014b). Improvisation of the masses: Anytime, anywhere music. In G. Lewis, & B. Piekut (Eds.), *Oxford handbook of improvisation studies*. Oxford: Oxford University Press.

Wang, G., Bryan, N.J., Oh, J. & Hamilton, R. (2009). Stanford Laptop Orchestra (SLOrk). *Proceedings of the International Computer Music Conference* (pp. 505–508). Montreal, Canada

Wang, G., Essl, G., & Pentinnen, H. (2014). The mobile phone orchestra. In S. Gopinath, & J. Stanyek (Eds.), *Oxford handbook of mobile music studies*. Oxford: Oxford University Press.

Wang, G., Essl, G., Smith, J., Salazar, S., Cook, P.R., Hamilton, R., et al. (2009). Smule = Sonic Media: An intersection of the mobile, musical, and social. *Proceedings of the International Computer Music Conference*, Paris, France, pp. 283–286.

Wang, G., Oh, J. & Lieber, T. (2011). Designing for the iPad: Magic Fiddle. *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, pp. 197–202.

Weiser, M. (1991). Computers in the 21st century. *Scientific American Special Issue on Communication, Computers, and Networks, 265*(3), 94–103.