Designing Smule's Ocarina: The iPhone's Magic Flute

Ge Wang
Center for Computer Research in Music and Acoustics (CCRMA)
Stanford University | SonicMule, Inc. (Smule)
660 Lomita Dr.
Stanford, CA 94305
ge@ccrma.stanford.edu

"*Any sufficiently advanced technology is indistinguishable from magic.*"
- *Arthur C. Clarke*

Abstract

The Smule Ocarina is a wind instrument designed for the iPhone, fully leveraging its wide array of technologies: microphone input (for breath input), multitouch (for fingering), accelerometer, real-time sound synthesis, high-performance graphics, GPS/location, and persistent data connection. In this mobile musical artifact, the interactions of the ancient flute-like instrument are both preserved and transformed via breath-control and multitouch finger-holes, while the onboard global positioning and persistent data connection provide the opportunity to create a new social experience, allowing the users of Ocarina to listen to one another. In this way, Ocarina is also a type of *social instrument* that enables a different, perhaps even magical, sense of global connectivity.

Keywords: Ocarina, mobile music, social, interface, multitouch, design, iPhone, ChucK.

1. Introduction

The Smule Ocarina is an expressive musical instrument created for the iPhone (Figures 1 and 2), re-imagining the ancient acoustic instrument while radically transforming it in the "kiln" of modern technology. Ocarina is sensitive to one's breath (gently blowing into the microphone controls intensity), touch (via a multitouch interface based on the 4-hole English Pendant ocarina), and movement (dual axis accelerometer controls vibrato rate and depth). It also extends the traditional instrument by providing precise intonation, extended pitch range, and key/mode mappings. As one plays, the finger-holes respond sonically and onscreen and the breath is visualized in pulsing waves. Sound synthesis takes place in real-time on the iPhone via Smule's audio engine, using the ChucK programming language and runtime [18].

However, the described interface is only half of the instrument. Ocarina is also a unique social artifact, allowing its user to hear other Ocarina players throughout the world while seeing their location – achieved through GPS and the persistent data connection on the iPhone. The instrument captures salient gestural information that can be compactly transmitted, stored, and precisely rendered into sound in the instrument's *World Listener*, presenting a different way to play and share music. In the first six months since its release in November 2008, Ocarina has been downloaded and played on more a million devices, while its users have collectively listened to more than 20 million performances.

Figure 1. An ensemble of iPhone Ocarina players, rendering the introduction to Led Zeppelin's Stairway to Heaven.

The online Ocarina forum offers user-created Ocarina tablature for more than 1200 (and counting) melodies. Most encouragingly, perhaps, is the observation that most Ocarina users are not musicians, and yet are able to be musically expressive in a unique global musical community. Overall, the Smule Ocarina serves as an experiment in making use of technology to explore new types of mobile, social, musical experiences.

Figure 2. Smule's Ocarina in action. The player blows into the microphone, while using different fingerings to control pitch, and the accelerometer to control vibrato. Players can hear one another around the world, via GPS and a centralized network.

2. Background and Related Works

2.1 Relate Work

The field of mobile music has been explored in several bodies of research, much of which has informed and inspired this work. Tanaka presented an accelerometer based custom-made augmented PDA that could control streaming audio [15]. Geiger designed a touch-screen based interaction paradigm with integrated synthesis on the mobile device using a port of Pure Data (PD) for Linux-enabled portal devices like iPaqs [9][10].

Using mobile phones for sound synthesis and live performance has been pioneered by Greg Schiemer [13] in his PocketGamelan instrument. At the same time there has been an effort to build up ways to allow interactive performance on commodity mobile phones. CaMus and CaMus2 introduced systems that use onboard cameras of mobile phones for tracking visual references for musical interaction [12].

The MobileSTK port of Perry Cook's and Gary Scavone's Synthesis Toolkit (STK) to Symbian OS [4] is the first full parametric synthesis environment available on mobile phones. It was used in combination with accelerometer and magnetometer data in ShaMus [5] to allow purely on-the-phone performance without any laptop. Golan Levin's DialTones performance is one of the earliest concert concepts which used mobile devices as part of the performance [11].

More recently, Stanford Mobile Phone Orchestra (MoPhO) [19] is exploring the combination of real-time sound synthesis, the ideas of "electronic chamber music" as pioneered by the Princeton Laptop Orchestra (PLOrk) and Stanford Laptop Orchesra (SLOrk), and the mobility of phones to create a new form of ensemble and classroom experience.

Location and global positioning play a significant role in the Ocarina. This notion of "locative media", a term used by Tanaka and Gaye [16] has been explored in various installations, performance, and other projects. These include Wagenaar's "Kadoum", which GPS sensors reported heart-rate information for sonification from 24 participants. Gaye's explored this idea in Sonic City with location-aware, body sensors [8]. Tanaka et al has pioneered a number of projects on this topic, including Malleable Moble Music and Net D'erive, the latter leveraging a centralized installation that tracked and interacted with geographically diverse participants [17].

Many of these ideas and practices have been reviewed by Gaye et al [7], who work with the definition "Mobile music is a new field concerned with musical interaction in mobile settings, using portable technology."

2.2 Smule and the iPhone

Smule (a.k.a. SonicMule, Inc.) was founded in Summer 2008 by Jeff Smith and the author, intensely investigating a notion of "interactive sonic media", and starting with the iPhone. Smule serves as a unique platform for research and development, combining the state-of-the-art in computer music research with a unique potential to bring its visions to a wide population.

The initial catalyst for Smule stemmed from the iPhone and its more recent App Store [1], the combination of which, we believe, represents an inflection point in mobile computing. Until the iPhone, there has never been such an intersection of existing technologies, integrated into a single, personal mobile device, and deployed at such a pervasive scale. The iPhone contains a powerful CPU, GPU (graphics processing unit), multitouch (up to 5 points), dual-axis accelerometer, high quality audio pipeline (two speaker outputs, microphone headset), location, persistent data (via 3G, Edge, or 802.11). The iPhone software development kit contains API's to access all of these components, and provides libraries for concurrency, graphics (OpenGL ES), and user interface.

In terms of scale and reach, the iPhone, at the time of this writing, has an install base approaching 30 million users worldwide in over 70 countries (with a significant additional install base of iPod Touches). Meanwhile, more than 30,000 third party applications have been released in the App Store.

The arrival of such new technology is accompanied by exciting new opportunities to explore and discover novel uses that can change the way people make music and relate to each other. This is our research mission: to change how people think, do, and play through sound, afforded by new technologies. In the next sections, we apply these and other ideas to the design of the Smule Ocarina.

3. Design

3.1 Design Goals

The design for Ocarina aimed to achieve several goals. We wanted to create an expressive musical instrument. However, instead of taking a larger instrument (e.g., a guitar or a piano) and "shrinking" the experience into a small mobile device, we started with a small and simple instrument, the ocarina (more specifically, the 4 hole English Pendant ocarina), and fleshed it out onto the form factor of the iPhone. Secondly, we hoped to preserve as much of the physical interaction as possible, while leveraging the technology to extend the instrument in potentially useful ways. Thirdly, we wanted to explore location/GPS combined with persistent data connection on the iPhone to enable new social musical experiences.

3.2 Instrument Interface

The design for Ocarina (Figure 3) leverages the onboard microphone for breath input (located on the bottom right of the device). A ChucK shred analyzes the input in real-time via a custom envelope follower, tracking the amplitude and mapping it to the intensity of the synthesized Ocarina tone. This preserves the physical interaction of blowing from the acoustic instrument to the iPhone Ocarina. Multitouch is used to allow the player to finger any combination of the four fingerholes, giving a total of 16 different combinations. Animated visual feedback reinforces the engaging of the breath input and the multitouch fingering. Sound is synthesized in real-time via ChiP (ChucK on the iPhone).

The onboard accelerometer is mapped to vibrato. Up-down tilt is mapped to vibrato depth, while the left-right tilt is mapped to vibrato rate. This allows high-level expressive control, and contributes to the visual aspect of the instrument, as it requires the player to physically move the device.

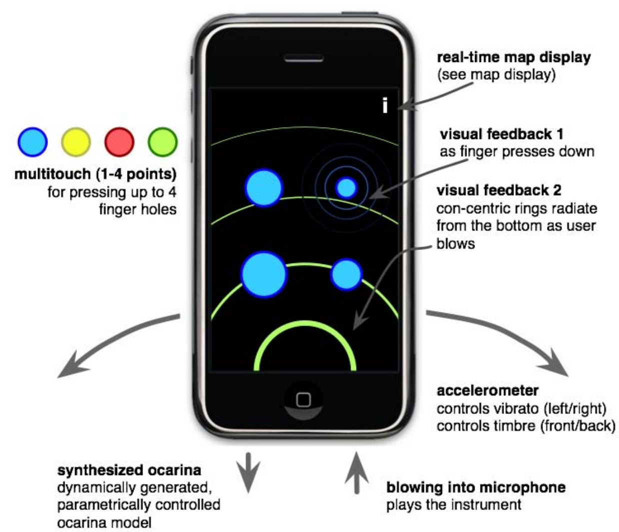

Figure 3. Design schematic for the instrument interface.

The acoustic ocarina produces sound as a Helmholtz resonator, and the size of the finger holes are carefully chosen to affect the amount of total uncovered area as a ratio to the enclosed volume and thickness of the ocarina – this relationship directly affects the resulting frequency. The pitch range of a 4-hole English pendant ocarina is typically one octave, the lowest note played by covering all four fingerholes, and the highest played by uncovering all fingerholes. Some chromatic pitches are played by partially covering certain holes. Since the Smule Ocarina is digitally synthesized, a certain amount of flexibility becomes available. No longer coupled to the physical parameters, the digital Ocarina offers precise intonation for all pitches, and is able to remap and extend the fingering. For example, the Smule Ocarina allows the player to choose the root key and mode (e.g., Ionian, Dorian, Phrygian, etc.), the latter offering alternate mappings to the fingering.

Figure 4. Screenshots of the instrument and the *World Listener*.

3.3 World Listener

In addition to the instrument interface, the Smule Ocarina presents a *World Listener* view, where one can see the locations of other Ocarina players (as indicated by white points of light), and hear one another. If the listener likes the snippet, he/she can "heart" the snippet by tapping the heart icon. The "snippet" being heard is chosen via an algorithm at a central Smule Ocarina server, and takes into account recentness, popularity, geographic diversity of the snippets, as well as filter selections by the user. The listener can choose to listen to 1) the World, 2) a specific region, 3) snippets that he/she has loved, and 4) snippets he/she has played.

The snippets are captured on the device, as the instrument is played. An algorithm decides when to record, captures the information, tags with the current GPS location (given the user has granted access), and sends it to the Smule Ocarina server. The musical recording consists of precisely timed gestural information (breath pressure, finger-hole state, tilt), simultaneously compact and rich. During playback, the Ocarina audio engine interprets and renders the gestural information into sound in real-time. ChucK's *strongly-timed* features lend themselves naturally to this endeavor.

4. Social Experience and Community

Ocarina is perhaps the first instrument in history that allows its players to hear one another. Given that there are now over a million Smule Ocarina players around the world, this is significant. Over 20 millions snippets have been created and shared, each with precise timing, key, melody information. We have only begun to mine this significant body of musical data.

The anonymity of the social interaction is also worthy of note – everyone is only identified via a self-chosen handle (e.g., Link42), their GPS location, and through his/her music. And yet, according to overwhelming user feedback, this seems to be compelling in and of itself.

Figure 5. Screenshots of in the *World Listener*, rendering snippets from actual users. Ocarina now has more than one million users worldwide.

In addition to the experience on the mobile device, Smule's Ocarina has a web portal [14] dedicated for users to generate and share musical scores. Since November 2008, users of the Ocarina have authored more than 1200 scores using our custom Ocarina tablature (Figure 6), serving millions of views. User-generated scores include video game music (e.g., *Legend of Zelda Theme Song*, *Super Mario Bros. Theme*), western classical melodies (e.g., *Ode to Joy*, *Blue Danube*, *Adagio for Strings*), to rock classics (e.g., *Yesterday* by the Beatles, *Final Countdown* by Europe), movie tunes (Star Wars *Imperial March*, *Superman* theme), to show tunes, holiday music, and more.

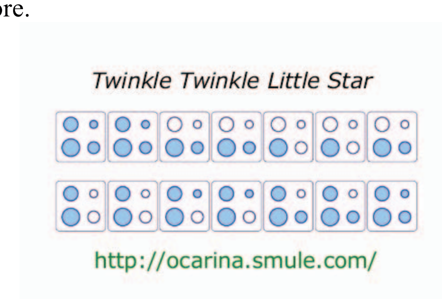

Figure 6. An example of Smule's Ocarina tablature: the beginning of Twinkle, Twinkle, Little Star.

It is worthwhile to note that, as far as we can tell, most of our users are not "musicians", and yet they seem to playing the Ocarina as an expressive instrument, and many learning to play an instrument for the first time. It also serves as a point of social interaction. People play Ocarina over dinner, at family gatherings, to show off the iPhone to their friends. Hundreds of user generated Ocarina YouTube videos have appeared [20] (search for "Smule Ocarina"). Additionally, Ocarina resembles a traditional instrument in that players are practicing in front of scores (in this case, on their computer monitors) while playing a physical artifact.

5. Concluding Remarks

We've learned much from our Ocarina adventure. One takeaway, for us, is that the integration of the technologies on the iPhone is indeed compelling for expressive music-making (albeit Ocarina is a relatively simple instrument). Another takeaway is that there is a sense of "magic" in wide-area, massive scale location, and furthermore, identity is perhaps not crucial (and anonymity can be just as powerful as it encourages different types of social interactions). Finally, the sheer number of Ocarina users at large shows that perhaps with the right approach and settings (e.g., mobile, personal, easy), we can encourage a large population to engage in expressive music making, and even create global communities virtually overnight.

At the same time, we have a long way to go in terms of truly unlocking the potential of the technology. In this context, it is useful to never forget that it is *people* we are ultimately designing for. Technology, almost by definition, will evolve, rise, and become obsolete, while human nature changes much more slowly (if at all). We hope to do our part to explore music-making in the unfolding landscape of mobile computing.

`http://ocarina.smule.com/`

6. Acknowledgments

This work is the collaborative and creative effort of many folks at Smule and CCRMA, including Jeff Smith, Spencer Salazar, David Zhu, Mattias Ljungstrom, Arnaud Berry, Rob Hamilton, Perry Cook, Georg Essl, Jennifer Wu, Rebecca Fiebrink, Jonathan Berger, Chryssie Nanou, Turner Kirk, Tina Smith.

References

[1] Apple, Inc. "iPhone." http://www.apple.com/iphone/. Retrieved January 2008.

[2] Apple, Inc. "iPod Touch." http://www.apple.com/ipodtouch/. Retrieved January 2008.

[3] P. Cook. *Real Sound Synthesis for Interactive Applications*. A. K. Peters. 2005.

[4] G. Essl and M. Rohs. "Mobile STK for Symbian OS." *In Proceedings of the International Computer Music Conference*, New Orleans, Nov. 2006.

[5] G. Essl and M. Rohs. "ShaMus - A Sensor-Based Integrated Mobile Phone Instrument." *In Proceedings of the International Computer Music Conference*, Copenhagen, 2007.

[6] R. Fiebrink, G. Wang, and P. R. Cook. "Don't Forget the Laptop: Using Native Input Capabilities for Expressive Musical Control." *In Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 164–167, New York, NY, 2007.

[7] L. Gaye, L. E. Holmquist, F. Behrendt, and A. Tanaka. "Mobile Music Technology: Report on an Emerging Community." *In Proceedings of the Conference on New Interfaces for Musical Expression*, pages 22–25, June 2006.

[8] L. Gaye, R. Maze, and L. E. Holmquist. "Sonic City: The Urban Environment as a Musical Interface." *In Proceedings of the International Conference on New Interfaces for Musical Expression*, Montreal, Canada, 2003.

[9] G. Geiger. "PDa: Real Time Signal Processing and Sound Generation on Handheld Devices." *In Proceedings of the International Computer Music Conference*, Singapure, 2003.

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

[11] G. Levin. "Dialtones - a telesymphony." www.flong.com/telesymphony, Sept. 2, 2001. Retrieved on April 1, 2007.

[12] M. Rohs, G. Essl, and M. Roth. "CaMus: Live Music Performance using Camera Phones and Visual Grid Tracking." *In Proceedings of the International Conference on New Instruments for Musical Expression*, pages 31–36, Paris, June 2006.

[13] G. Schiemer and M. Havryliv. "Pocket Gamelan: Tuneable Trajectories for Flying Sources in Mandala 3 and Mandala 4." *In Proceedings of the 2006 conference on New Interfaces for Musical Expression*, pages 37–42, Paris, June 2006.

[14] *Smule Ocarina Forum*. http://ocarina.smule.com/ Retrieved January 2008.

[15] A. Tanaka. "Mobile Music Making." *In Proceedings of the 2004 conference on New Interfaces for Musical Expression*, pages 154–156, June 2004.

[16] A. Tanaka and P. Gemeinboeck. "A Framework for Spatial Interaction in Locative Media." *In Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 26–30, Paris, June 2006.

[17] A. Tanaka and P. Gemeinboeck. *net derive*. Project web page, 2006.

[18] G. Wang. *The ChucK Programming Langauge: a Strongly-timed, On-the-fly Environ/mentality*. PhD Thesis. Princeton University Press. 2008.

[19] G. Wang, G. Essl, and H. Pentinnen. "MoPhO: Do Mobile Phones Dream of Electric Orchestras?" *In Proceedings of the International Computer Music Conference*. Belfast, August 2008.

[20] YouTube. http://www.youtube.com/ retrieved January 2008.