Olegoru: A Soundscape Composition Tool to Enhance Imaginative Storytelling with Tangible Objects

Chuan-Che (Jeff) Huang*
chuanche@umich.edu

Yu-Jen Lin*
yjlintw@umich.edu

Xinda Zeng*
xinda@umich.edu

Mark W. Newman*
mwnewman@umich.edu

Sile O'Modhrain*
sileo@umich.edu

* University of Michigan
School of Information

Figure 1: Children could interact with Olegoru via speech, non-verbal voice and gesture to compose sound effects for their stories.

TEI '15, Jan 16-19, 2015, Stanford, CA, USA.
ACM 978-1-4503-3305-4/15/01.
http://dx.doi.org/10.1145/2677199.2687895

Abstract
Learning storytelling is beneficial for children's development. Various tools have been proposed to expand the set of materials children can use to compose their stories. However, most previous research focuses on enhancing the visual aspect of storytelling and underexplores the acoustic elements needed for children's stories. In this paper, Olegoru[1], a sound composition tool in the form of magic gloves and soul stones, is proposed to augment children's storytelling when using physical objects. Children can create contextual and regional sound effects as well as event-based acoustics through speech, non-verbal and gestural interaction, and could potentially enable children to create more immersive story-worlds. To investigate the technical feasibility of such tool, a preliminary prototype was built that accepts a limited number of vocalized sound effects and vocabulary.

Author Keywords
Storytelling; Physical computing; Sound tagging; Soundscape; Children

ACM Classification Keywords
H.5.2 User Interfaces

General Terms
Design; Human Factors

[1] Olegoru is a play on the word "Oregoru", which means music boxes in Japanese, and the word LEGO.

Introduction

Encouraging imaginative storytelling is beneficial for children's development––much research has demonstrated that the everyday social and physical dynamics of interaction with the world can be learned via storytelling, as well as aspects of living such as creativity, communication and motor skills. Therefore, one particular challenge for the Human-Computer Interaction researcher is to design tools that further expand the set of physical and virtual materials available for storytelling, thus allowing children to compose more complex, immersive and realistic stories while at the same time making the interface intuitive.

There are many digital storytelling tools [5, 10] that allow children to create stories with a high degree of freedom and more immersive environments. However, these digital tools distance children from the physical worlds and the physical objects they've already engaged with. They require children to adapt to the limited control afforded by devices such as computer mice and touch-pads for creating their stories and underutilize children's powerful motor and communication skills––their hands and voices. This limitation forces children to spend more time learning the control tools before they can engage in storytelling. In addition, research has demonstrated the potential benefits for children of using tangible materials for learning [3, 4]. For example, physical objects can support collaboration and allow children to manipulate objects in parallel easily.

To better enrich children's storytelling in the physical world, many researchers have developed physical interactive toys that augment the capabilities of physical objects such as [2, 6, 7, 8, 12]. For example, Lego Mindstorm [9] is a set of programmable Lego blocks that enables children to create the behaviors of physical objects. Children can thus make robotic animals that walk and respond to human commands in their stories. Storyrooms and Magic Wand [7], on the other hand, allow children to assign simple if-then rules for computer-augmented soft toys, such as blinking a stuffed star whenever it is touched. Although these tools have enhanced the interactivity of physical worlds, they focus more on augmenting the visual aspect of storytelling. The acoustic aspect, however, is underexplored in these projects. For instance, there is still no easy mechanism for children to compose the sound of the sea, a cannon, sword fighting and storms that are needed for their pirate stories.

Some researchers have explored sound composition tools for children (e.g., SoundBlock & SoundScratch [3]) and there have been a few commercial products available (e.g., LEGO SoundBrick and Sound Lab [15]) that allow children to manipulate sound effects easily using physical objects. However, these products either focus more on instrumental sound effects, or provide a very limited set of short sound effects. They are still far from the complexity of soundscapes such as those discussed by Schafer and Westerkamp [11, 14]. In addition to the introduction of rich natural and historical soundscapes, Schafer argues that the awareness of our sound environment is important as it pervades everything we do. As an outcome of this thinking, a series of curricula were developed by Schafer and others for elementary school children to encourage creation of soundscapes, in addition to more traditional music training. Inspired by Schafer, Westerkamp and the recent development of environmental sound synthesizers [13], we therefore propose Olegoru, an interactive toy that allows children to improvisationally compose various contextual sound and short sound effects using speech, non-verbal voice and gesture. In doing so, we strive to increase the richness and immersiveness of children's imaginative play.

Interaction Concepts

Olegoru contains a fingerless magic glove, and a set of intelligent bricks called "soul stones". Children ages 8-12 can use Olegoru while constructing their LEGO

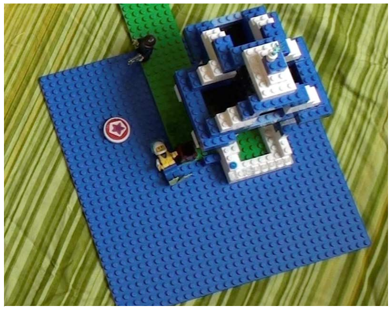

Figure 2: A soul stone could be placed at a region of a story-world. It allows children to assign sound effects to a specific region (around 30cm.)

Figure 3: A child could attach a soul stone to a LEGO mini-figure and assign a laser shooting sound to it. After that, once the child shakes the mini-figure, the real laser gun sound effects would be played.

stories. They can issue speech and non-verbal commands to the magic glove to create various sound backdrops. In addition, sound effects can be assigned to a specific region or object. We first introduce the then illustrate how they are implemented.

Form Factors of the Design

The design metaphor for our system is a magic glove. The storyteller (in this case the child) wears the magic glove, which has a microphone embedded in it. The glove metaphor is inspired by movies and fantasy stories, such as Frozen, where gloves are used to either hide or augment the magic capabilities of a protagonist. Therefore, for children familiar with such stories, gloves often incorporate a sense of magic and thereby become an object for casting spells; such spells might be accompanied by 'magic words', utterances that are designed to evoke a magical world, complete with a magical soundscape. The advantage of this magic glove design is that children's hands are usually occupied when playing Lego. The glove design allows children to focus their hands on manipulating objects while still being able to compose the soundscape.

Global Soundscape Creation

Using Olegoru, children are able to create the sound backdrop for their stories. This feature allows children to set the general tone of the stories. For example, in a pirate story, the sound backdrop could be the sound of sea waves or the sound of a storm. In Olegoru, Children can compose a global soundscape through casting the spell "Oleg", and then provide a command, such as "give me some wind", or "this is a sea". The keyword "Oleg" activates the speech recognition system, thus allowing the subsequent sentence to be recognized and matched to a specific sound effect.

Regional Soundscape Tagging

Aside from global soundscape creation, Olegoru also supports regional soundscape creation––contextual sounds that are limited to certain small areas of the play space. For example, in a pirate story, a child might want to create the sound for a treasure island, and separate it out from the sound of the surrounding sea. To enable sound assignment for distinct regions, we designed a kind of intelligent LEGO brick, which we called a soul stone that is embedded with RF-based (e.g, Bluetooth) proximity sensors. A child can first tap a particular soul stone, and issue a speech command to assign a sound effect to a soul stone corresponding to a small region. This regional sound effect will appear when their hand, wearing the magic glove, moves close to the soul stone, and will disappear when the glove moves away.

Imperative and Declarative Commands

Sometimes it is un-natural for children to issue imperative commands such as "I want to hear the sound of wind" while they construct their stories. Children usually just assign a meaning to a LEGO set, "this is a forest", they might say. To incorporate this observation into the design, thus enabling more intuitive sound composition, we allow children to use both imperative and declarative commands for playing with different sound effects. The commands "This is a forest" and "I want to hear the sound of wind", for example, might result in similar sound effects.

Non-verbal Sound Tagging

Olegoru also enables children to tag a sound effect to a specific object using non-verbal voice commands, and allow the activation of the tagged sound by shaking the object. Often, when children are immersed in storytelling, they make various sounds that serve as an imitation of the characters in their imagined scenario. For example, a child might say "Shu" to imitate the sound of laser guns, or "Wuf" to imitate the sound of a dog barking. While we don't want to discourage children from making these sounds themselves, as this is also a means for children to express their creativity; in fact we want to augment this non-verbal aspect to allow for more complex and immersive sound

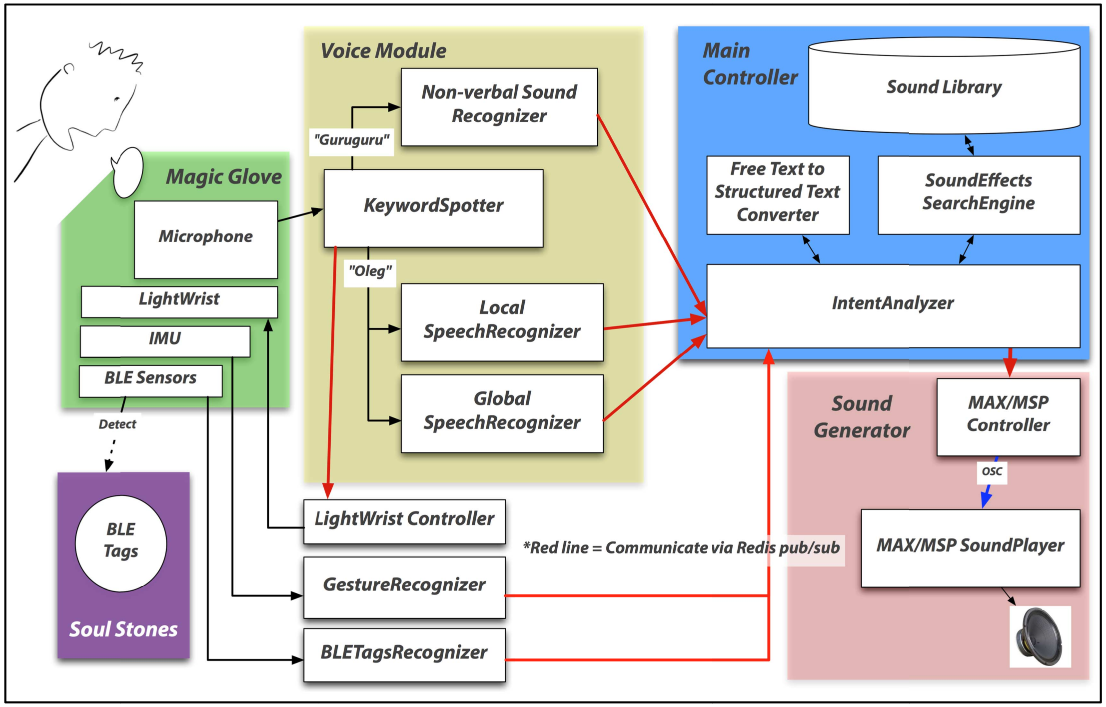

Figure 4 & 5: (Top) System Architecture. (Left) The first magic glove prototype.

composition. We therefore introduce an interaction called non-verbal sound tagging. A child could first tap on a LEGO module that is a soul stone, cast the spell "Guruguru", and then assign a sound effect using non-verbal voice utterances such as "Shu" for a laser gun or "Wuf" for a barking dog. The real sound effects would be attached to this module, and the child could activate the sound by shaking the object.

Preliminary Prototype

To better demonstrate the design concept, we implemented a preliminary prototype that consists of four components, namely: magic glove & soul stones, the voice module, a main controller, and the sound generator (See Figure 4).

Magic Glove & Soul Stones

Magic glove and soul stones are the input interfaces of the system. We embedded several sensors and actuators within the magic glove: a lavaliere microphone, an inertial measurement unit (IMU), an LED strip and a Bluetooth Low Energy sensor. We used RadioShack's 33-3013 Omnidirectional Microphone to capture the audio input. Also, an MPU-6050 IMU module was adopted for gesture detection. A shaking gesture is detected when the value from any axis constantly exceeds a pre-defined threshold over a period of time. We also sewed an LED strip onto the magic glove to provide visual feedback about the state of the magic glove. The LED lights up for a short period of time when either of the keywords "Oleg" or "Guruguru" are detected. This provides an indication of whether the magic glove is listening to the commands. Finally, to allow sound tagging, we instrumented the glove with a Bluetooth Low Energy sensor to track the proximity of soul stones assigned with sound effects. The soul stones are currently implemented using Bluetooth Stick N' Find tags. These soul stones can be either attached to an object or a region. The proximity of a soul stone is inferred via the Received Signal Strength Indication (RSSI) given by Bluetooth. Since RSSI does not provide centimeter-level accuracy, we wrap the BLE sensor with multiple layers of tin foil [1] to limit the range of detection from 10 meters to within 30-50 cm. We then simply set a RSSI threshold to determine if a soul stone is close or not.

Voice Module

The voice module is primarily designed to process audio information. A keyword spotter implemented with the CMU Sphinx [16] is continuously listening to children's utterances. If the keyword "Oleg" is detected, the verbal recognizer will be activated while the non-verbal recognizer will be triggered if "Guruguru" is uttered. We built a non-verbal recognizer using HTK [17]. Several Hidden Markov Models were trained for recognizing our pre-defined non-verbal sounds. For the speech recognizer, we implemented both a local and a global recognizer. The local recognizer is also based on the CMU Sphinx and has a small dictionary that contains mostly words related to sound effects. In addition, a grammar-based language model is used. The global recognizer is based on Google Web Speech API. Using both local and global recognizers allows us to boost the detection accuracy, as the local recognizer performs well for spotting pre-defined sound-related keywords, while the global recognizer is better at inferring out-of-grammar sentences.

Main Controller

The main controller handles the input and output information in order to understand the player's intent and then to instruct the system to take corresponding actions. For speech, we built a natural language processing module with wit.ai [18] to structurize free texts. Specifically, this module resolves a verbal dictation to a pre-defined intent, such as the playing of a sound effect, as well as extracting the sound-related keywords from the dictation. For non-verbal sounds, the keywords of intended sounds are already identified in the recognition process. The main controller then takes such keywords and communicates with a sound effect search engine to retrieve related sound files from a sound effects library. For our initial prototype, a small set of sound effects were extracted from the BBC Sound Effects Library [19] and annotated with textual information related to the context of the sound. The main controller also accounts for gestural events and the proximity of soul stones to allow object sound tagging and sound activation.

Sound Generator

We built a sound generator for the playback of sound effects. Although the preliminary prototype only supports the playback of one sound effect at a time as the decision logic of the main controller is relatively simple, the sound generator, implemented using MAX/MSP can technically layer multiple sound effects together to provide an immersive auditory experience.

Conclusion and Future Work

In this paper, we highlight the opportunities for developing soundscape composition tools to enhance children's imaginative storytelling while playing with physical objects. An interactive toy, Olegoru is introduced to allow children to improvisationally compose various contextual sound effects and event-based acoustics using speech, non-verbal voice and gesture.

In the form of magic glove and soul stones, Olegoru supports four interaction concepts––global and regional soundscape creation, imperative and declarative command recognition, and non-verbal sound tagging. However, our first prototype has some limitations. The local speech recognizer uses phoneme models that were trained using adult's voices, thus it might not currently perform well for children. In addition, the non-verbal recognizer only supports five different sound effects, and the models were also trained on adult's voices. We are interested in using a children's speech corpus to enhance our speech recognition system and to collect more non-verbal utterances for training the non-verbal recognizer. The proximity detection for soul stones needs to be improved as well, as we currently use close-source Stick N' Find tags that have a slow sensor update rate, which results in slower detection and reduced accuracy. We therefore would like to explore other open-source BLE tags that allow the adjustment of the low-level BLE settings to increase the detection accuracy. Finally, we would like to gain knowledge about how children interact with this toy in a user study. We are interested in how our four interaction concepts might increase the complexity of the soundscapes children could create, thus potentially making children's story-worlds more immersive and engaging. In addition, we are curious about how children's knowledge and awareness of their sound environments might be improved through using Olegoru.

Acknowledgements

We thank Emily Provost for providing insights on the implementation of our speech and non-verbal recognizers.

References

[1] Cheung, K.C., Intille, S.S., and Larson, K. An inexpensive bluetooth-based indoor positioning hack. *Proc. UbiComp06 Extended Abstracts*, (2006).

[2] Glos, J.W. and Cassell, J. Rosebud: Technological Toys for Storytelling. *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, ACM (1997), 359–360.

[3] Harrison, J. SoundBlocks and SoundScratch : tangible and virtual digital sound programming and manipulation for children, 2005. http://dspace.mit.edu/handle/1721.1/33885.

[4] Marshall, P. Do Tangible Interfaces Enhance Learning? *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, ACM (2007), 163–170.

[5] MineCraft. https://minecraft.net.

[6] Montemayor, J., Druin, A., Chipman, G., Farber, A., and Guha, M.L. Sensing, storytelling, and children: Putting users in control. (2003).

[7] Montemayor, J., Druin, A., Chipman, G., Farber, A., and Guha, M.L. Tools for Children to Create Physical Interactive Storyrooms. *Comput. Entertain. 2*, 1 (2004), 12–12.

[8] Montemayor, J., Druin, A., Farber, A., Simms, S., Churaman, W., and D'Amour, A. Physical Programming: Designing Tools for Children to Create Physical Interactive Environments. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2002), 299–306.

[9] Resnick, M. Behavior Construction Kits. *Commun. ACM 36*, 7 (1993), 64–71.

[10] Resnick, M., Maloney, J., Monroy-Hernández, A., et al. Scratch: Programming for All. *Commun. ACM 52*, 11 (2009), 60–67.

[11] Schafer, R.M. *The Soundscape.* Destiny Books, Rochester, VT, United States, 1993.

[12] Umaschi, M. Soft Toys with Computer Hearts: Building Personal Storytelling Environments. *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, ACM (1997), 20–21.

[13] Verron, C., Aramaki, M., Kronland-Martinet, R., and Pallone, G. A 3-D Immersive Synthesizer for Environmental Sounds. *IEEE Transactions on Audio, Speech, and Language Processing 18*, 6 (2010), 1550–1561.

[14] Westerkamp, H. Linking soundscape composition and acoustic ecology. *Organised Sound 7*, 01 (2002), 51–56.

[15] Molecule Synth:An Electronic Musical Instrument for Everyone. https://www.kickstarter.com/projects/travisfeldman/molecule-synth.

[16] CMU Sphinx - Speech Recognition Toolkit. http://cmusphinx.sourceforge.net.

[17] HTK. http://htk.eng.cam.ac.uk.

[18] wit.ai. https://wit.ai.

[19] BBC Sound Effects Library. http://www.sound-ideas.com/sound-effects/bbc-1-40-cds-sound-effects-library.html.