Duet Musical Companion: Improvisational Interfaces for Children

David Ventura
Carnegie Mellon University
Entertainment Technology Center / ATR MIS Labs
5000 Forbes Avenue, 4301 Doherty Hall
Pittsburgh, PA 15213 USA
1-412-268-5351
dventura@cmu.edu

Kenji Mase
ATR MIS Laboratories / Nagoya University
2-2-2, Hikaridai
Seika-cho, Soraku-gun
Kyoto 619-0288 Japan
81-774-95-1440
mase@atr.co.jp

ABSTRACT
We present a sensor-doll interface as a musical outlet for personal expression. A doll serves the dual role of being both an expressive agent and a playmate by allowing solo and accompanied performance. An internal computer and sensor system allow the doll to receive input from the user and its surroundings, and then respond accordingly with musical feedback. Sets of musical timbres and melodies may be changed by presenting the doll with a series of themed cloth hats, each suggesting a different style of play. The doll may perform by itself and play a number of melodies, or it may collaborate with the user when its limbs are squeezed or bent. Shared play is further encouraged by a basic set of aural tones mimicking conversation.

Keywords
Musical improvisation, toy interface agent, sensor doll, context awareness.

1. INTRODUCTION
We have developed a musical entertainment system for a context-aware sensor doll with the aim of creating a responsive toy companion for children. By encouraging children to play with sound and make-believe, we hope to promote both learning and personal expression. The interface chosen is a bear for its traditional appeal to many ages, as well as having a set of familiar parts for tactile stimulation. The bear supports two roles: as an instrument for solo interaction, and a companion for collaborative performance. This duality abstractly illustrates the principles of both melody and harmony in experimentation. Feedback to physical contact is issued purely in the aural sense by chords, riffs, melodies and harmonic utterances. Varied play is encouraged through different timbre palettes associated with a series of cloth hats that trigger mode changes in the bear. Our overall objective is to provide an entertaining musical toy for children to play and experiment with.

2. RELATED WORK
The area of technologically augmented physical companions is a relatively new field as the required hardware begins to become affordable and compact-enough to place inside of children's playthings. There has been a growing offering in recent years both for commercial and research sectors. Embryonics' Munchkin produced an augmented stuffed animal Mozart Magic Orchestra Featuring Wolfgang that performs an instrument patch when one of its paws is squeezed [3]. Learning Curve's Lamaze has a similar doll Octotunes that plays a different note for a touch on each of its tentacles [5].

Figure 1. Bear dolls and themed hats.

Strommen and Alexander report on educational and emotional factors for interactive dolls such as ActiMates' Barney [8, 9] and Arthur [1]. There has also been interesting work recently examining methods for adolescent musical expression and learning. Browall and Lindquist have studied children's play and cooperation to develop a multi-user toy sound mixer [2]. Payling created a set of augmented blocks for children to provide an alternate interface for musical expression [6]. Gan studies a similar application but with particular focus to tactile feedback [4].

Our work leverages the hardware and principles of the com-music sensor doll developed by Yonezawa et. al [10] and Kazuyuki et. al [7]. In form, function and expressive capabilities our system is most similar, but we expand on the previous work by adding basic computer vision and combining it with emotional and design factors for children, much like methods described for the work above.

3. ENVIRONMENT
3.1 Design Issues
When constructing the interaction suite, there were a number of points considered that controlled the scope and design of the project.

3.1.1 Personality and Autonomity
To adequately serve the purposes of a play *thing* and a play *mate*, we added enough functional constraint to enforce basic musical principles and themes. Also, great amounts personality or intelligence were not invested to prevent the expectation of an autonomous figure. The doll does not speak,

but it does respond with musical murmurings if neglected or treated harshly. Input for the doll is received directly from physical contact with various parts of the body, and indirectly through a USB camera in its nose that processes images of the environment. The doll may play discrete audio samples and effects, or perform a backing track for the player to collaborate with. Aside from this simple feedback, however, much of the actual play is required from the child.

To add to the doll's charm, subtle characterizations are embedded. We have produced two dolls: first, a brown bear, designated as a boy, warbles and sings in a deep and gruff tone, secondly we have a more colorful blue bear, which has a higher pitched and more feminine vocal timbre. Both, however, contain the same melodies for play and respond vocally to stimuli such as pulling on their mouths.

3.1.2 Props and Color Choice

Each bear may play four different "games", where a game is defined as six complementary samples and a melody. Musical themes associated with the games are linked to the hat each bear uses to activate them. A further degree of iconization is encapsulated in the hat's color. Using basic color psychology principles, different adolescent emotional archetypes are bound to the music. A baseball cap signifies sports, so it triggers patches associated with a brass band and a tune resembling "Take Me Out to the Ballgame". The hat is colored red, which commonly is associated with aggression and excitability. A nightcap is connected to a lilting lullaby and soporific, long decaying major chords of wind instruments. The nightcap is made from quilted, yellow fabric, suggesting infancy and a strong dependence on parental protection. With a painter's beret, the bear supports creative expression, inner reflection and emotional balance. The aural scheme consists of a tonally balanced set of string instruments and a meditative melody. Lastly, a blue knight's helmet formed in the shape of a dragon's head is used to represent daydreaming and imagination, and emotional maturity. A wistful celestial song is employed in conjunction with a set of surreal warm pads and bells. With these separations and groupings, we unobtrusively reinforce the connections and roles between several levels of music and the subconscious. (See table 1.)

Table 1. Themed color and musical separations

Hat Type	Color	Audio Patches
Baseball cap	Red	trumpet, tuba, drums
Nightcap	Yellow	flutes, air vocals
Painter's beret	Green	violin, cello
Knight's helmet	Blue	pads, bells

4. IMPLEMENTATION

4.1 Physical Components

We have constructed two dolls, each a stuffed bear approximately 14 inches in height. In the bear's core we have installed an off-the shelf 400MHz Celeron CPU, an A/D PCMCIA converter card to process sensor data, and a USB speaker. Interaction is supported by an USB CCD camera in the nose, a G-Force sensor in the belly, proximity sensors in the face and hips, and touch/bend sensors in the arms, legs, back, front and head. The system runs with Windows 2000 and DirectX 8.1, and was developed using Visual C++ 6.0, DirectMusic, the Microsoft Vision SDK, and an open source image library. The system is completely self-contained, and executes automatically following connection to an internal rechargeable battery. The process for sensor input to musical output inside of the doll occurs at a rate of 4Hz.

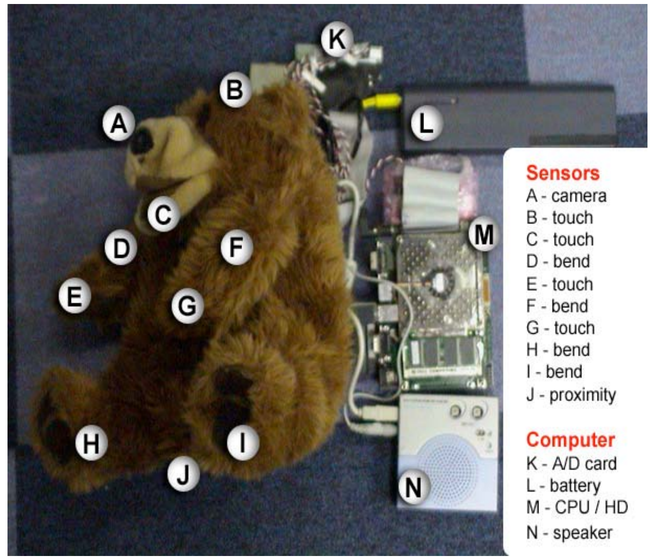

Figure 2. Hardware components.

4.2 Data Transfer

Our software system builds on existing functionality used with similar projects for the given hardware, but is designed to be reconfigurable and robust, so its overall structure operates with three modules supporting the main component.

The core of the application monitors the bear's current play state and internal attributes, such as sensor stimuli and active sound buffers. A timer polls the sensor array for new data four times a second to prevent backlog of audio feedback events sent to the sound manager. Requests for camera acquisition and analysis are also handled by the main application.

At initialization, a virtual camera object is created by the image processing unit, which then maintains the connection to the USB camera and grabs low-resolution JPEG images when the pressure sensor in the bear's head is activated. This functionality is triggered by the guest when he/she shows the bear a new hat for changing musical themes. The image is then saved to disk with a timestamp. A dominant-hue analysis is performed to determine which color hat is being shown to the bear, and consequently which game is to be played next. Acceptable returns for the hue analysis are extreme amounts of pure red, green, blue or yellow (which the hats are discretely comprised almost entirely of, see figure 2). The algorithm works with a 90% success rate in identifying the correct hat under most indoor lighting conditions. If a dominant color cannot be extracted (if no hat is occupying a majority of the camera's viewing frustum), the bear returns to an idle state of murmuring conversation and touch-activated chattering.

The tactile sensor array polls the hands, arms, legs, base, head and mouth of the bear constantly and sends updated data to the AI's state machine four times a second to keep in sync with sound playback requests. Each request issued maps to a registered audio sample that may be played in a distinct MIDI channel. Samples consist of a riff, one-time hit, or chord, all of which are in key with the main song associated with the musical game. For all four themes, sensor events map to similar tonal components. Squeezing the left or right hand results in a supporting chord, and bending the left or right arm inwards a descending or ascending riff up the scale, respectively. Bending of a leg provides a warm overtone or short melody that backs the bear's main tune. Each sample loops if the corresponding body part is held. Players may experiment with the doll for individual performance by touching the mouth, limbs, and head. Accompaniment mode is invoked when the bear is picked up/held, or laid on its back.

4.3 System Flexibility

All of the doll's sounds are general MIDI files, so samples and melodies may be changed by replacing the files and restarting the system. Additionally, since all of the images taken by the camera are stored to disk in JPEG format, the images may be extracted from another computer via a wireless or 10 base-T Ethernet connection, both of which are embedded inside the bear. To provide reliability in sensor on/off thresholds, we have implemented a calibration utility as well that allows the maintainer to physically touch the doll and define for the system what is an acceptable "on" and "off" amount of pressure. All of the data received through the bear's sensors is also capable of being recorded and logged to a Microsoft Access database, which may then be later viewed and played back through another supporting application.

5. PRELIMINARY EXPERIMENTS

5.1 Reaction Testing

We presented our dolls to a group of 38 7-12 year olds at a local daycare center to evaluate initial reactions from our target demographic. We selected primary school children so that we may observe a wide range of motivations for interacting with the doll, from simple curiousity to structured play. This also gave us the fortuitous opportunity to watch the differently aged children interact and learn from each other. We introduced the bears to children, and explained our intention to design an interactive musical toy to be played with. After demonstrating the basic features of the dolls, the children were allowed to take turns playing with them, first one at a time and then collaboratively. Following this, we asked the each of the children to draw a picture of the bears, write a few sentences about what they did, and then list some other things they would like the bears to be able to do.

5.2 Observations and Guest Feedback

We found that the children were very receptive to the dolls, and could interact successfully with little difficulty. Unfortunately, one of the bears was too heavy for some of the children to pick up and play with, as it had a protective metal casing around the computer's core. Some of the children enjoyed the "dress-up" aspect of the dolls more, while others preferred experimenting and testing the limitations of the hardware. Children aged in the median of our demographic appeared to enjoy playing the most, and exhibited the strongest sustained and focused play. They experimented with all of the hats and actuators, in addition to occasionally dancing back and forth when another student would start a new song. These All of the children seemed to appreciate the human-like warbling exhibited when the bears' mouths were pulled on or if the dolls were left alone for a time.

The feedback we received was unilaterally positive and constructive. Many remarks made were that the bears were cute, and their audio feedback was entertaining and interesting. Many children commented they enjoyed touching the bears and wanted to play with them again. Requests for modification included conversational responses from the bears, more hats to dress with, and walking or automated movement of the limbs.

6. CONCLUSION

We designed an interactive musical toy application for an augmented sensor doll. Using principles of color psychology and basic music theory, we created an interaction that supports both solo improvisation and assisted performance. Our system is one step towards creating digitally augmented toys that support traditional interaction while encouraging experimentation and personal expression.

There are several areas for further expansion that persuade examination. Increasing the doll's sense of direct awareness in regards to its user and its environment may allow it to respond more intelligently and shift closer to playing the role of being a companion. Enhanced imaging functions and voice recognition methods could be used to control the bear's mood and timbre synthesis, either directly or in a more subtle fashion. We chose to make use of the USB camera-internal PC combination for the physical object recognition and sound processing because we strove to keep the system design as simple, cheap and off-the-shelf as possible. A radio frequency tag system may have been more accurate, but its addition could surpass the physical space limitations inside of the bears, and also significantly raise the total system cost. Similar tradeoffs for performance may arise with the use of a custom hardware microcontroller. Regardless, future investigation of such methods may allow greater freedom in design. The doll may also be enhanced in its role of an instructor, teaching musical concepts through playback games a la a purely audio version of *Simon*. Overall, we hope to see further research in creating toys instructive and compelling enough that they stimulate and promote growth in children, both scholastically and emotionally.

7. ACKNOWLEDGMENTS

Our thanks to Hiromi Fukunaga for helping us with our user testing and feedback, and to Rodney Berry for his continued consultation and advice. Most of all, we'd like to thank the children of the Ikoma Gagudo daycare center for trying out our application and giving us such wonderful feedback. This research was made possible in part by the Telecommunication Advancement Organization (TAO).

8. REFERENCES

[1] Alexander, K., and Strommen, E. Evolution of the Talking Dinosaur: The (Not So) Natural History of a New Interface for Children. In Proceedings of CHI' 98, ACM Press, 7-8.

[2] Browall, C., and Lindquist K. Our Little Orchestra: The development of an interactive toy. In Proceedings of DARE 2000 on Designing augmented reality environments, ACM Press, 149-150.

[3] Embryonics, Mozart Magic Orchestra Featuring Wolfgang. Munchkin Incorporated. 2000.

[4] Gan, S., Squeezables: Tactile and Expressive Interfaces for Children of All Ages. MS Media Arts and Sciences Dissertation, Massachusetts Institute of Technology School of Architecture and Planning. September 1998.

[5] Lamaze, Octotunes. Learning Curve International. 2003.

[6] Payling, D., Musical Blocks for Children. MSc Music Technology Dissertation, The University of York. August 2001.

[7] Saitoh, K., Yonezawa, T., Mase, K., Awareness Communications by Entertaining Toy Doll Agents. In Proceedings of the International Workshop on Entertainment Computing 2002, Kluwer Publishing, 326-333.

[8] Strommen, E. When the Interface is a Talking Dinosaur: Learning Across Media with ActiMates Barney. In Proceedings of CHI' 98, ACM Press, 288-295.

[9] Strommen, E. Emotional Interfaces for Aardvarks: Designing Affect into Social Interfaces for Children. In Proceedings of CHI' 99, ACM Press, 528-535.

[10] Yonezawa, T., Clarkson, B., Yasumura, M., Mase, K., Context-aware Sensor-Doll as a Music Expression Device. In Extended Abstracts of CHI' 01, ACM Press, 307-308.