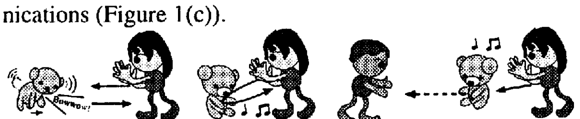
Context-aware Sensor-Doll as a Music Expression Device

Tomoko Yonezawa [1,2], Brian Clarkson [3], Michiaki Yasumura [2], Kenji Mase [1]
[1] ATR MIC Research, [2] Keio University, [3] MIT Media Lab.
{yone, mase}@mic.atr.co.jp, clarkson@media.mit.edu, yasumura@sfc.keio.ac.jp

ABSTRACT
We present a *sensor-doll capable of music expression* as a sympathetic communication device. The doll is equipped with a computer and various sensors such as a camera, microphone, accelerometer, and touch-sensitive sensors to recognize its own situation and the activities of the user. The doll has its own internal "mind" states reflecting different situated contexts. The user's multi-modal interaction with the passive doll is translated into musical expressions that depend on the state of mind of the doll.

Keywords
Context-aware doll, multi-modal interaction, music expression

INTRODUCTION
We have developed a context-aware sensor-doll with the eventual aim of a human-human communications device to support non-verbal channels between humans. We emphasize two important roles of dolls, i.e., as a partner (second person) and as another self (first person), which are often seen in children's playing of house where the roles are switched back and forth. An active robot, in contrast, would show its own existence with a stronger personality, and possibly disturb the expressions and activity of the user in the communications (Figure 1(a)). With an ambient entity like the proposed doll not only dominated and controlled by the user but also maintaining its own character, we expect the doll to lead the user to perform actions by its new communication environment and enrich the conventional verbal communications channel. Our system was prototyped for the initial situation of human-doll interaction (Figure 1(b)); this is the first step for assisting human-human communications (Figure 1(c)).



(a) (b) (c)

Figure 1: Types of Communications

We adopt music expressions as the only actuators. We regard that music can play an important role as a new communication channel since it does not provide the absolute interpretation for an expression; instead, the interpretation depends on the context and environment. For example, the same melody can make different harmonics when affected by different back chords.

The doll is fully equipped with a stand-alone computer and various sensors. They are used to recognize the contexts of the doll, such as environmental events, and to interpret each input from the user based on its current internal state. We envision that the more the doll can come to be an intimate object for a human partner, the more it will be able to capture the partner's context.

RELATED WORKS
The most related work to our system is the "Swamped!" testbed [3] which was built for the "sympathetic interface" using a sensor-doll. The doll is mainly used as a controller for the character's behavior in its own graphically presented virtual world. We propose using sounds and music as the ambient actuators. ActiMates Barney [2] is a commercialized sensor doll available as a scaffolding of a kid's playmate for learning through interaction. In contrast to this system with prerecorded reactions, we use sounds and music parameters to extend the expressive capability of the actuators. A wearable sensor system [1] uses a camera and microphone as sensors and adopts Hidden Markov Models (HMMs) to recognize the context of the user and various events. We adopt their context sensing mechanism.

INTERACTION MODEL & SYSTEM DESIGN
We first designed the context-aware sensor-doll as a reactive communicator between the user and the doll. The sensor-doll has several internal modes and accepts two kinds of reactive controls: (1) context-recognition for mode switching and (2) direct input translation in each mode. The internal modes of the doll are divided into five states representing the doll's internal mind states such as moods. Each state nearly corresponds to the strength of activities and is represented by Interaction Levels from Level0 (IL0) to Level4 (IL4); IL0 is the sleeping state but interested in environment where calm breathing sound is generated, IL1 is the user encounter and awake state with a voice-like sounds, IL2 is a state for warm/familiar communications with music-tabled voice and breathing sounds, IL3 is a state for rhythmical and musical communications with musical instrumentation, and IL4 is a non-communicative and out-of-control state with confusing music and sounds. The transitions between states are controlled by the interaction with an automaton model, which is a finite state machine. A different recognition module is activated for each event based

on the current internal state. For instance, when the "lift up doll" event is sensed at the IL1 state, the internal state moves to IL2, or the internal state changes to IL3 while the internal state is IL2, if the doll recognizes a rhythmical input event. The finite state machine internal model might also be modified by learned events and response correspondences (Figure 2).

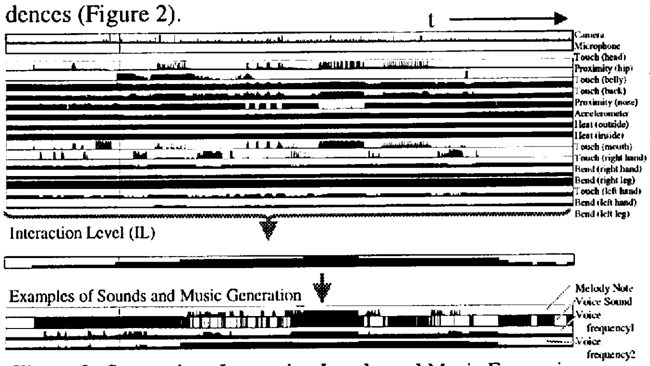


Figure 2: Sensor data, Interaction Levels, and Music Expression

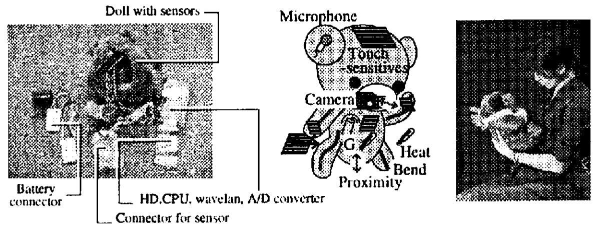


Figure 3: Sensor-doll Setup and Snapshot of Playing

The sensor-doll contains a small PC with wireless networking capability, battery, A/D signal converter, and 16 sensors of seven types attached to the shell of the plush bear-like doll; touch-sensitive sensors, bend sensors, a camera, a microphone, an accelerometer, and two infra-red proximity sensors (Figure 3). The sensor values and recognized gesture data are transmitted over the network to a PC station in the MIDI format. In the current implementation, the internal state automaton control and sound and music synthesis generation are performed with MAX/MSP at the PC station. The system generates and outputs sounds and music from a context-based interpreter which receives (i) the internal state, (ii) recognized events and gestures, and (iii) raw sensor data. The sounds and music are sent out to the room's loudspeakers as well as the doll's internal wireless loudspeaker. The room's loudspeakers play ambient music while the wireless speaker plays the doll's voice.

There are five control categories of music expressions: 1) global loudness, harmony, key, and tempo, 2) breathing sound interval, loudness, resonance filter intensity, and the harmony structure, 3) voice sound loudness, filtering frequency, speed, and delay time, 4) melody musical notes, length, and loudness, and 5) rhythm loudness and pattern. We map them depending on the interaction level as the interpreted contexts. Some expressions are real-time responses to inputs and others are autonomous displays of the doll's state. Consequently, the same input to the doll can result in different expressions appearing, depending on the context. In IL3 in particular, the doll performs as a musical controller allowing its partner to play music with it. We also employ a rhythm detection algorithm from tap signals via tactile sensors to use the rhythm as 1) the current tempo and 2) trigger to move to IL3 during IL2.

PRELIMINARY EXPERIMENTS

We set the doll in the free running mode and observed its interaction with people being allowed to play with it. It changed its internal states and generated sounds and music through the various types of interaction almost as designed (Figure 2). In the playing of house, we observed that some of the people sometimes treated it familiarly and sometimes ignored the character as a controller. In another example, some of them who knew that a camera and microphone were installed tried to interact with the doll through tactile inputs, not visual or auditory inputs. We could tell such the tactile interaction with the doll was important in the preliminary experiment.

CONCLUSION

The unique part of the system is that the user's action is interpreted based on the doll's embedded internal character. The doll displays its own built-in autonomous behaviors when responding to external approaches.

The remaining issues for further study may be summarized as follows; 1) stand-alone hardware configuration, 2) HMM-based gesture recognition, and 3) intensive user evaluations. We plan to put all functions including the music generation into the doll's PC. Although only simple event recognition can be achieved at present, we are training the HMMs to recognize more complex gestures and environmental changes in the doll's PC.

We consider that the doll will become a musical communications media for its partner, providing the user with some musical expression controls and harmonizing their combinations. The doll can also be a playmate toy for entertainment purposes, a part of a music education system, a human-human communications support device over a network, etc.

ACKNOWLEDGMENTS

We thank Ryohei Nakatsu and other ATR MIC members for their support and discussions in this research.

REFERENCES

1. Clarkson, B., Mase, K., Pentland, A. (2000), Recognizing User Context via Wearable Sensors, ISWC2000 Proceedings, pp. 565-569.
2. Johnson, M. P., Wilson, A., Kline, C., Blumberg, B., Bobick, A. (1999), Sympathetic Interfaces: Using Plush Toys to Direct Synthetic Characters, Proceedings of CHI 99, pp. 152-158.
3. Strommen, E. (1998), When the Interface is a Talking Dinosaur: Learning Across Media with ActiMates Barney, Proceedings of CHI 98, pp. 288-295.