

LEARNING SOUNDING GESTURES

A. de Götzen L. Mion and
F. Avanzini

Dept. of Information
Engineering
University of Padova, Italy
<http://smc.dei.unipd.it/>

Stefania Serafin

Medialogy Department
Aalborg University in
Copenhagen, Denmark

ABSTRACT

Non-visual senses can be used in toys to enhance and enrich the play experience. Previous research has shown that – especially for young children developing sensory-motor skills – exploration and play are two tightly linked activities: everything is new and needs to be “studied”, and playful behaviors emerge from active exploration. The main idea of this paper is to provide a new approach in designing and realizing objects that elicit this type of behavior and encourage exploration by providing dynamic, real-time haptic, tactile, auditory feedback depending on a child’s gestures, movements, and emitted sounds. These toys provide interaction based on the enactive paradigm, where multimodal feedback is intimately tied to action – i.e. the human is “in the loop”. Moreover, these musical toys will teach how to perform musical gestures.

1. INTRODUCTION

According to the traditional mainstream view, perception is a process in the brain where the perceptual system constructs an internal representation of the world, and eventually action follows as a subordinate function. Two assumptions emerge from this view. First, the causal flow between perception and action is primarily one-way: perception is input from world to mind, action is output from mind to world, and thought (cognition) is the mediating process. Second, perception and action are merely instrumentally related to each other, so that each is a tool for the other. Recent theories have questioned such a modular decomposition and have rejected both the above assumptions: the main claim of these theories is that it is not possible to disassociate perception and action schematically, and that every kind of perception is intrinsically active and thoughtful. As stated in [8], only a creature with certain kinds of bodily skills (e.g. a basic familiarity with the sensory effects of eye or hand movements, etc.) can be a perceiver. One influential contribution in this direction is [12]. The authors present an “enactive conception” of experience, which does not occur inside the animal, but is rather something that the animal *enacts* as it explores its environment. In this view, the subject of mental states is the *embodied*, environmentally situated animal.

The animal and the environment form a pair in which the two parts are coupled and reciprocally determining. The term “embodied” highlights two points: first, cognition depends upon the kinds of experience that are generated from specific sensorimotor capacities. Second, these individual sensorimotor capacities are themselves embedded in a biological, psychological, and cultural context. The enactive knowledge is then stored in the form of motor responses and acquired by the act of “doing” [7]. An example of enactive knowledge is represented by the competence required by tasks such as typing, playing a musical instrument, sculpting objects, whistling, tying shoelaces etc. This type of knowledge transmission can be considered natural and intuitive, since it is based on the experience and on the perceptual responses to motor acts and it involves more than just one interaction modality. Multimodal Interaction for children however poses new specific challenges. Conceivably, the kind of support the children need is different from that of adolescents and adults. Pioneering works in this field are due to Seymour Papert, who developed the Logo programming language (the first children’s toy with built-in computation), and to Mitchel Resnick, whose research group developed the “programmable brick” technology that inspired the LEGO MindStorms robotics kit and the PicoCricket artistic-invention kit [13]. Toys for children are very often poor in terms of interactivity, while multimodal interaction should be the principal way of exploring the environment and learning from it. The importance of sound as a powerful medium has been largely recognized, up to the point that there are objects on the market that reproduce prerecorded sounds by pushing certain buttons or touching certain areas. However, such triggered sounds are extremely unnatural, repetitive, and ultimately annoying. As a consequence the interaction is unrealistic and un-engaging, and the learning pattern very stereotyped. The key for a successful exploitation of sounds in toys interfaces is to have models that respond continuously to continuous gestures, just in the same way as children do when manipulating rattles or other physical sounding objects, eliciting the enactive exploration of the world through multimodal interaction in order to help them to discover and recognize many different sounding gestures, each of them characterized by specific movement, force, velocity etc.

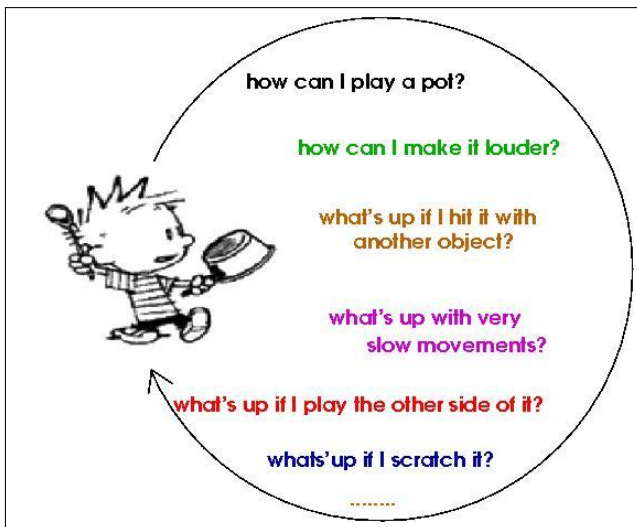


Figure 1. The action perception closed loop

2. WHAT DO WE LEARN?

Gesture and sound seem naturally connected in a clear and obvious way: the image of instrument players learning to use their body in order to produce sound is indeed widespread and compelling enough. Musical gesture can be simply thought as a gesture that produces sounds in an continuous feedback loop: this is a general definition that can be used in many interactive contexts besides the musical ones. In this respect it can be useful to design interactive sounding toys that do not have to be the exact replica of original instruments, while they have to help in acquiring some basic musical skills. Children get tired soon of traditional teaching methods e.g. for bowing instruments, since they have to spend many hours before the teacher get satisfied by the sounds and modulations produced. In this case the arc bowing can be taught - along with aesthetics - at a more gestural level by means of multimodal interfaces. Controllers, for instance can be tuned in order to make children have fun during their learning, adding a visual/haptic feedback that can engage the child in exercising/playing with the instrument. Moreover very young babies could start their musical training with basic toys, learning that is not enough to kick an object to produce sound, but that it might sound in a more pleasant way by just caressing it, or finding the needed force to squeeze it. The idea is to teach musical gestures by a simple interaction mediated by the child's body. Each object/instrument has its own way to be played: through an enactive exploration of the object children can learn the effects of their gestures, learning what are the 'musical gestures' needed to produce the sound they want (as simply sketched in figure 1). Sound and gestures are indeed very important for children in preschool age.

In fact, cognitive sciences focus on how humans interact with their environment, searching the connection between perception and action to bridge the semantic gap that humans experience in their everyday life: as sound and music are linked to their physical energy, the content

of auditory information has to be linked to meaningful actions that we can use to access the encoded high-level information. To face this gap problem we pursue the idea that the human mind is embodied, so the relation between meaning and physical energy is mediated by the human body. This relation is crucial for non-verbal communication in particular; implicit messages (e.g. expressive contents) are indeed the basis of the communication process in different social situations, specially for children whose language is based on sounds and gestures, organized by semantics and constructs only at a later stage. Those sounds and gestures can be very expressive and rich of emotional content, as music can be.

Humans use recognition and expression of affect to detect meaning [9], and communication by means of vocalization, facial expression, posture, and gesture express affect (emotion) and convey information more powerfully and efficiently than spoken language. Concerning the communication between children, the tactile/auditory perceptions are the major actors for emotional response and affect: the sound-making gestures of infants are the earliest attempts for separating basic emotions [4], and earlier exposure to sound patterning has profound effects on perceptual and emotional development, while deprivation can lead to future weak development of linguistic and musical skills [10]. The understanding of the emotional response related to sensory experiences and object relationships is then a crucial issue, and a new generation of expressive toys will exploit the idea of embodied-expressive knowledge; moreover expressive paradigms based on affective and sensorial adjectives will be used to provide expressive feedback to children according to their input. The idea is to create interfaces in which children can associate well known feelings and basic emotions to audio feedback, expressed by physical metaphors which can be directly mapped to higher emotional labels [5]. Applications in this direction can be imagined for teaching/educating to musical gestures rather than to the musical language itself. Gestural skills can be developed by means of interfaces for controlling in real time the expressive information by tactile interaction, controllers to map and to transform audio data, even promoting and to stimulating the communication process at the same time.

3. HOW DO WE LEARN?

In light of embodied perception theories, it is clear that developing "enactive interfaces" implies developing techniques for multimodal feedback and input, including sound, touch and gesture. Sound and touch are inherently tied to movement. Without movement there would be no sound, and the sounds we perceive are influenced by the way our ears move within the world. Most of the information received by touch is also a result of movement, this being particularly true for proprioception and kinesthesia. This is well known for children who explore the objects around them by touching moving themselves and the objects, hearing the results of their actions etc. Thus the

study of haptic and sound becomes particularly interesting from our point of view since we are focusing on the dynamic properties of the interaction and on the learning process that is elicited by the action.

Sound and haptic interaction are related in a number of different ways. Actions produce sounds by direct, physical manipulation of physical objects. There is a physical energetic consistency between action and produced sounds: sounds can be produced by instant object manipulation (the sound starts after the end of action), or by continuous object manipulation (the sound continues during the manipulation). Everyday sounds are used to get information from the environment, in order to know what things are, where they are, and what happens. They can be used to inform the environment about our actions or intentions, in order to show what we are doing, where and when we are doing it. Studies on the interplay between touch and audition concerning object properties have mainly focused on contact properties such as hardness, stiffness and texture. For surface roughness and stiffness it has been shown [6], that touch dominates over audition, but both of them can improve the perception or even create an illusion.

In light of the perceptual studies mentioned above, simultaneous audio-haptic rendering is a particularly interesting problem in the development of enactive multimodal interfaces. Recent literature has proposed *physically-based models* for sound synthesis, i.e. sound synthesis algorithms based on a physical description of the sound generating mechanisms. Since the resulting computational structures respond to physical input parameters, they automatically incorporate complex responsive acoustic behaviors. A second quality of physically-based approaches concerns interactivity and ease in associating motion to sound control. As an example, the parameters needed to characterize collision sounds, e.g. relative velocity at collision, can be directly used to control a physically-based model, and the sound feedback consequently responds in a natural way to gestures and actions. Various approaches have been proposed in the literature for contact sound modeling. Modal synthesis [1] was proposed in [11] as a framework for describing the acoustic properties of objects; the modal representation is naturally linked to many *ecological* dimensions of the corresponding sounds: modal frequencies depend on *shape* and geometry of the object, *material* determines the sound decay characteristics, and so on. Physically-based models for real-time synchronous haptic-sound rendering is an approach that will ensure synchronization and perceptual similarity between haptic and audio feedback. A significant amount of recent literature deals with this problem. In [3] the modal synthesis techniques described in [11] were applied to audio-haptic rendering. A related study was recently conducted in [2]: physically-based sound models were integrated into a multimodal rendering architecture, and the setup was used to run an experiment on the relative contributions of haptic and auditory information to bimodal judgments of contact stiffness.

4. SCENARIOS

The first years of a baby life is a continuous discovery. A baby starts to learn the reactions he produces in the world around him, how things sound, move, smell and how things can be used. Everything can become a toy and the boundary between a tool, a toy, and a simple gadget is never clear and determined. In particular, pre-scholar babies spend hours playing with very simple objects that become whatever they wish, according to their shape or properties. A big pillow can become a spaceship, a big empty box can be knocked with the hands or with a spoon, it can become a drum or the door of a little house. Every object can elicit the imagination of babies, and in general the simpler the object is, the bigger are the transformations that it can perform in the baby's mind. The idea of this work is to realize simple toy-objects that, once explored, show their multimodal properties, suggesting basic and complex reactions and interactions, improving the learning process and the motor skills of children.

In the following we describe four scenarios which materialize our concepts.

4.1. The sounding car-box

The sounding car-box is a plastic cube on top of which children can sit. It is meant for children from three years old onwards. On one side of the cube a crank can be attached. By rotating the crank, children play different tunes, like in a musical box. The box is also provided with several wheels. One of the wheels can be attached on top of the box, and provides haptic and auditory feedback when manipulated. This is obtained by having sensors which detect the rotation of the wheel and the pressure at the center of the wheel. Four more wheels can be attached in the bottom of the box, two for each side. When all the wheels are connected, the box becomes a small car which children can use to move around.

4.2. The multimodal lego village

The multimodal lego village is a construction kit made of lego blocks. The kit is similar to traditional Duplo lego kits, but designed in such a way to enhance children's awareness of everyday life's multimodal feedback. In the kit some of the different lego components are enhanced with sensors. As an example, the kit is provided with plastic pets. Such pets respond to the action of the children while touched. Kids can caress the lego cat and hear it purring. People in the village are enhanced with pressure sensors on their feet, which are used to create synthetically simulated footsteps while the children make them walk in the different locations of the environment. The village is also provided with some flowers whose smell is artificially reproduced using the smell actuators.

4.3. The musical toys kit

The musical instrument kit is a collection of musical toys designed to enhance musical skills and collaboration among children. As an example, the kit contains a drum built with lego blocks. The drum is a drum shaped object with two plastic mallets. The mallets are embedded with accelerometers, one for each side and with a vibrator which is activated when the drum is hit. The surface of the drum has a piezo sensor. It is covered with damped material, so no acoustics sound is produced when the drum is played.

Children play the drum as a traditional drum, i.e., by hitting the surface of the drum alternating left and right mallet. The gesture of the children is tracked by the accelerometers and by the piezo sensors. The sound of the drum is produced by a physical model of an impact connected to a two-dimensional mesh representing the drum. It is possible to vary the type of drum by changing the parameters of the physical model (tension of the drum and dimensions). The force feedback provided by the actuators enhances the perception of hitting different materials.

A software application trains children to perform different drum strokes. The application starts by producing a simple stroke, which the children are requested to imitate. An algorithm checks that the stroke has similar dynamics as the target strokes (force of impact and velocity) by comparing the accelerometers' data and the piezo data of the children to the target stroke. The application then introduces game patterns, which the children is asked to imitate. A pattern recognition algorithm checks that the child has performed the correct rhythmic pattern.

The drum improves existing toy drum interfaces (such as Taiko: drum master for Playstation 2), where the act of hitting a drum is considered as an on/off action. The enactive drum enhances this idea by tracking the gestures performed by the child while hitting on the object.

5. CONCLUSIONS

This investigatory work wants to propose a new way to think about toys for children, taking the particular perspective of enaction and multimodal interaction. The basic idea is to use enactive toys interfaces to teach by playing how a gesture can produce a sound, and how different gestures affect the quality of the produced sound. Basic interaction strategies have been proposed, starting from squeezable toys for small children to more musical instruments that can help young musicians to approach the study of an instrument.

Acknowledgment

This research was partially supported by the EU Sixth Framework Programme – IST Information Society Technologies (Network of Excellence “Enactive Interfaces” IST-1-002114, <http://www.enactivenetwork.org>).

6. REFERENCES

- [1] Jean-Marie Adrien. The missing link: Modal synthesis. In Giovanni De Poli, Aldo Piccialli, and Curtis Roads, editors, *Representations of Musical Signals*, pages 269–297. MIT Press, Cambridge, MA, 1991.
- [2] Federico Avanzini and Paolo Crosato. Integrating physically-based sound models in a multimodal rendering architecture. *Comp. Anim. Virtual Worlds*, 17(3-4):411–419, July 2006.
- [3] Derek DiFilippo and Dinesh K. Pai. The AHI: An audio and haptic interface for contact interactions. In *Proc. ACM Symp. on User Interface Software and Technology (UIST'00)*, San Diego, CA, November 2000.
- [4] J. Dore. *Feeling, form, and intention in the baby's transition to language*. In Golinkoff, R.M. (Ed.). *The transition from prelinguistic to linguistic communication*, 1983.
- [5] G. G. De Poli, F. Avanzini, A. Rodà, L. Mion, G. D'Incà, C. Trestino, C. Pirrò, A. Luciani, and A. Castagne. Towards a multi-layer architecture for multi-modal rendering of expressive actions. In *Proceedings of 2nd International Conference on Enactive Interfaces (Enactive'05)*, Genova, Italy, November 2005.
- [6] S.J. Lederman, R.L. Klatzky, T. Morgan, and C. Hamilton. Integrating multimodal information about surface texture via a probe: relative contribution of haptic and touch-produced sound sources. In *Symp. Haptic interfaces for virtual environment and Teleoperator Systems (HAPTICS 2002)*, Orlando, FL, 2002.
- [7] A. Luciani, J.L. Florens, and N. Castagne. From action to sound: a challenging perspective for haptics. In *Workshop on Enactive Interfaces (Enactive'05)*, Pisa, Italy, January 2005.
- [8] Alva Noë. *Action in perception*. MIT press, Cambridge, Mass., 2005.
- [9] R. Picard. *Affective computing*. MIT Press, Cambridge, MA, 1997.
- [10] A. Sabbadini. On sounds, children, identity and a 'quite unmusical' man. *Br. J. Psychother.*, Vol. 14(2):189–196, 1997.
- [11] Kees van den Doel and Dinesh K. Pai. The sounds of physical shapes. *Presence: Teleoperators and Virtual Environment*, 7(4):382–395, August 1998.
- [12] Francisco Varela, Evan Thompson, and Eleanor Rosch. *The Embodied Mind*. MIT Press, Cambridge, MA, 1991.
- [13] P. Wallich. Mindstorms: not just a kid's toy. *IEEE Spectrum*, pages 52–57, 2001.