

# Sony CSL

Sony Computer Science Laboratory Paris  
<http://www.csl.sony.fr>

**President :**  
Dr. Mario Tokoro

**Director :**  
Prof. Dr. Luc Steels

**Contact:** Sophie Boucher  
[sophie.boucher@csl.sony.fr](mailto:sophie.boucher@csl.sony.fr)

**Date and location:**  
Open House 2004  
6<sup>th</sup> October,  
9h30-16h00,  
Sony CSL Paris,  
6, rue Amyot,  
75005 Paris

**Web site:**  
<http://www.csl.sony.fr>



**The Playground Experiment (2004)**  
<http://playground.csl.sony.fr>



**The Talking AIBO Project (2002)**

## Sony CSL Paris Open House 2004

### Recent advances in robotics, cognitive science, music and computational neuroscience

#### Sony CSL Paris: a laboratory for fundamental research

Sony CSL Paris was founded in 1996 and is a spin-off of Sony CSL Tokyo, founded in 1988. The objective is to contribute extensively to the advancement of science through original fundamental research, unrestrained by commercial needs. It is our policy to make the results of our research public and to collaborate with the academic institutions. The researchers in CSL Paris come from diverse countries, and all have a strong research background achieved in universities. Their work is interdisciplinary, crossing the domains of robotics, cognitive science, music and computational neuroscience with computer science at its core.

#### Understanding human communication, building social robots

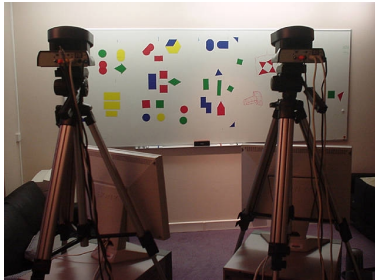
One of the founding themes of Sony CSL Paris is communication. It entails both the scientific issue of understanding human communication and the technological challenge of building robots with which we can have complex social interactions. We believe that this synthetic methodology is a powerful approach towards understanding human language, and that this understanding will in turn provide insights for building robots that can speak.

Our first major achievement in this area was the **Talking Heads experiment (1999)**. This experiment has shown how a population of robotic agents can self-organize a shared lexicon as well as a shared set of categories through cultural evolution. This experiment validated a new theory for the origins and evolution of language based on viewing language as a complex adaptive system, in which local interactions can lead to the formation of global cultural patterns through self-organization. This theory was published in reputed journals (1), and the setup was demonstrated in several museums in Paris, London and Cambridge.

A few years later, we developed the **Maïdo and Gurby experiment (2001)**, which showed that the same theoretical framework could account for the cultural formation of shared speech sounds (2), which are the physical carriers of language. This also provided an explanation of both the diversity of speech sounds in human languages and the statistical regularities of

their phonetics. This robotic demonstration was shown for two years in a public exhibition at *La Cité des Sciences et de l'Industrie* in Paris.

Beyond mere scientific advances, these two experiments also lead to innovative technological realizations. For example, we used the knowledge acquired in the Talking Heads experiment to build a system and an interaction framework within which a human could teach a robot how to give names to some objects in their shared world. This was called the **Talking AIBO Project** (3). We also developed technologies for emotional speech synthesis and recognition (4), which sprang naturally from the Maido and Gurby experiment. This technology was then used in the speech system of the Sony Qrio entertainment robot (5).



**The Talking Heads Experiment (1999)**

We are now starting a new phase of research in which we broaden the scope of these past achievements. These new advances will be presented at the Open House 2004. Our work is expanding in two fundamental directions. On the one hand, we are now building populations of robots using and sharing grammatical constructions. The challenge is to show how complex syntax as well as non-trivial semantic structures can emerge and evolve culturally. This lead us to develop a new formal framework called **Fluid Construction Grammars** (6), which breaks the limits imposed by former systems, and associated mechanisms for parsing, production, invention and adoption of grammatical constructions.

On the other hand, we are studying prerequisites for the emergence of communication. We try to show how certain mechanisms which were pre-programmed in the **Talking Heads Experiment** as well as in the **Maïdo and Gurby experiment**, can be developed autonomously by the robots.



**Maïdo and Gurby (2001)**

This involves for instance the study of **joint attention**. We work on the development of techniques allowing several autonomous robots to build social protocols by themselves, that enable them to interpret and manipulate the attention of each other using the AIBO platform.

Another example is our work on the **self-organized origins of vocalizations**. We will present a model of the origins of primitive but complex speech sounds in which agents do not yet possess the ability to sustain coordinated social interactions as required in the Maido and Gurby experiment. This work (7), achieved in the context of a PhD thesis, puts forward a theoretical breakthrough in the academic field of the origins of language, and will be published soon as a book. It has also been awarded the "*Prix Le Monde de la recherche universitaire*" (election of the 15 best Ph.D Thesis in France).

We are also trying to understand how agents could develop coordinated interactions and be interested in communication without this being pre-programmed. This boils down to the even more general question: how do human infants develop? What is

the basic mechanism that drives them through activities of increasing complexity, from simple sensorimotor skills to high-level linguistic interaction? We believe that the key is in motivational systems, specifically in an abstract drive such as “**curiosity**”. We have developed an operational model of curiosity, called “Intelligent Adaptive Curiosity”, implemented on a robot. At the Open House, we will present **The Playground Experiment** (8), in which a robot equipped with this drive is playing on a baby playmat. We show how this robot develops autonomously capabilities and engages in sensorimotor activities of continuously increasing complexity.

These new projects are part of a larger **European Project**, called **ECAgents**, launched in January 2004. This project develops concepts, tools, and models for analysing collections of both natural and artificial agents, and algorithms for designing artefacts that consist of evolving populations of interacting and communicating embodied agents. The consortium gathers leading scientists in the area of embodied cognition and complex systems from Italy, France, Switzerland, Belgium, Hungary, Sweden, Germany, Spain and Japan. More information is shown on <http://www.ecagents.org/>.



**A child playing with the Continuator**

### **Enhancing the music experience: access, listening and playing**

The main goal of Music research at Sony CSL Paris is to invent technologies that enhance the musical experience from the user viewpoint. To reach this goal we develop various prototypes that blur the traditional distinction between listening and performing, to invent new modes of “music-ing”. These works are divided into two main areas: music interaction systems, where we try to develop new ways of performing music, and music access systems, where the goal is to harness the complexity of large and evolving music collections.

The Music Interaction projects have come up with systems such as **MusicSpace** (10), a real time control system for spatialized music in which listeners can modify in real time the location of sound sources such as instruments. A representation of the know how of sound engineers through a constraint mechanism ensures that all user actions are corrected to always produce a “good mix”. **The Continuator project** (11) involves a real time interaction with a system that learns musical styles. With the Continuator, users can play music as they wish, and the system will automatically engage in a dialogue by producing musical phrases designed as continuations of user input. Gradually, the dialogue becomes more and more interesting and challenging as the system continuously learns from all the previous interactions. We are interested in the forms of excitement produced, and the links with, e.g. the theory of Flow (12).

The Music Access projects deal with the management of large music collections using content-based techniques (13). How to build representations of music titles (and of catalogues

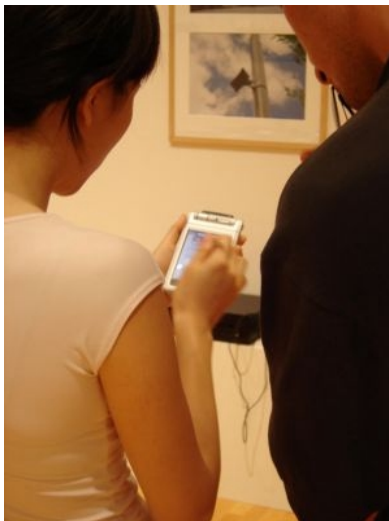


The EDS system

themselves) so that users can find interesting music even when they do not know the content of the collection? This research involves two main sides. First we try to extract automatically meaningful information from music titles, from the acoustic signal, such as rhythm structure or timbre (14). **The EDS project** (15) is an attempt to generalize this complex feature extraction task by using a novel approach, consisting in letting the system evolve automatically problem-dependent audio features from the analysis of audio signals. We have shown that EDS performs better than standard approaches in feature extraction on a series of well-known feature extraction problems (e.g. detection of singing voice). The second aspect concerns man-machine interfaces, that let users find music using similarity search engines. We have developed several music browsers, including **the Cuidado music browser**, and more recently gesture-based control music access systems.

### Creative communities and content sharing

Digital technologies and the Internet have accelerated modes of content sharing and content production that rely on collaboration and that lack central authority. Our research on **Creative Communities** aims at a better understanding of these new modes of production and seeks to develop technologies that enhance them.



Creative communities and content sharing (2004)

Consider, for example, the sharing of data in peer-to-peer networks. Participants in these networks typically organise their data differently, which complicates the formulation of search queries. One solution to this problem is to impose a standard information organisation on all users. However, the **Talking Heads experiment** has shown how agents could self-organise a shared lexicon as a side effect of their interaction. At the OpenHouse, we will demonstrate how the use of **emergent semantics** enables the information exchange in such distributed systems.

Another demo that will be shown focuses on audience participation in the content production. It features the system we experimented at the exhibition "Social Capital: Forms of Interaction," organised by the Whitney Independent Study Program in New York. In this case study, we created a shared, online space in which the visitors of the exhibit can post pictures of the displayed works and annotate them with text.

**Malleable Mobile Music** takes social dynamic and mobility as inputs to an advanced audio re-composition engine to enable listeners to experience familiar music in new ways. The system consists of mobile terminals, a sensor subsystem, GPS simulator, and music re-composition engine. A community of listeners is connected via a hybrid Peer-to-Peer/Client-Server infrastructure. They choose to listen to a single piece of music that is adaptive and responsive to their situation. This transforms music from a fixed entertainment medium into a malleable content form that enables shared experiences.

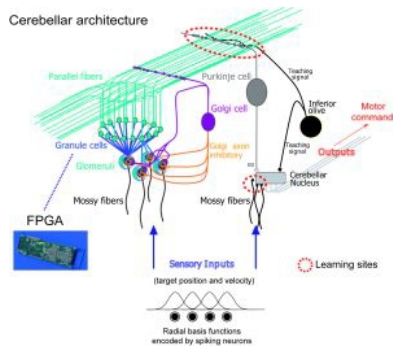


**Malleable Mobile Music  
(2004)**

The malleable music engine reconciles incoming data from the listener group and transforms a network musical stream. A generalized time domain re-sequencing technique allows structural reorganization of the music from the high level of song form to the low level of rhythm and melody variation. Social mechanisms are used to nurture *Trust*. Each listener selects a part in the music to be his *musical avatar*. Geographical localization is used to create a *social re-mix* of a familiar pop song. Users interact with the music through subconscious actions. Rather than being incited to become active in the music making process, the listener's seemingly passive act of listening is captured by a responsive music re-generation system.

### Understanding brain processing

Although advances have been made recently in machine learning and artificial systems, major issues remain unresolved. These regard abilities that are difficult to mimic by machines but that humans and animals display ubiquitously, such as adaptation, generalization, continuous learning with experience, and conceptualization. Do neural brain cells provide a computational platform with characteristics and representations that permit such abilities to be expressed in machines and be applied in practice? Do neural brain processes use unique methods for performing 'computation' in general, in contrast to those used in current computer and integrated circuit technologies? The Neuroscience Group attempts to answer these questions by investigating the processing used by the brain and its neural substrate.



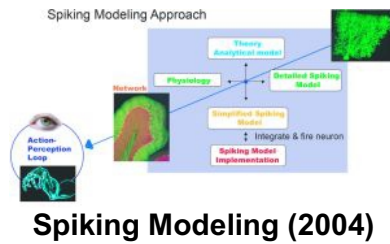
**Cerebellum modeling  
(2004)**

Neurons in the brain process information in the analogue domain (continuous values), whereas their communication in the form of impulses or *spikes* is essentially digital (0 or 1) and asynchronous in time. Our research focuses on how **spiking neurons** enable rapid **decision-making and learning**, with the ultimate goal of constructing real-time adaptive systems (machines) that can learn continuously.

The Neuroscience group of the SONY Computer Science Laboratory (CSL) in Paris and its **SpikeFORCE** project partners presented a novel technology at the Neural Information Processing System (NIPS) conference, held in Vancouver, Canada in December 2003. The new system was a brain-inspired software-hardware hybrid system for adaptive motor control. This system constitutes one of the first systems to have a spiking brain model that processes information in a complete sensory-motor loop of action and perception.

The system mimicked the response of neurons in the **cerebellum** by computing their impulses (spikes), which neurons use to transmit information. The cerebellum is a brain area well known in particular for its importance in the learning and coordination of motor activities (movements). The hybrid system combined the responses of different types of neurons

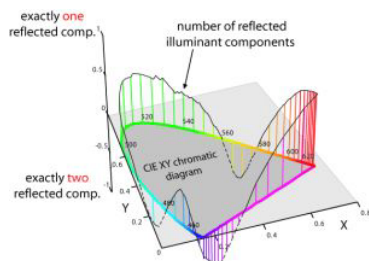
that were generated on a hardware chip and on a personal computer software that communicated the occurrence of spikes through the PC bus interface. More details are available at the SpikeFORCE site [www.spikeforce.org/nipsdemo.html](http://www.spikeforce.org/nipsdemo.html).



### Spiking Modeling (2004)

The SpikeFORCE consortium of neuroscientists, physicists, and engineers from four European institutions aims to produce a model of the cerebellum based on known physiology and latest analytical and computational results that can be implemented efficiently in software/hardware for running real-time robotic experiments. The consortium also investigates the advantages of spiking representations for computation and physical implementation, with the objective to provide learning mechanisms that would permit their use in constructing real-time adaptive systems.

We have constructed a program for simulating real-time spiking neural networks. It is a table-based event-triggered simulator, which makes use of predictive threshold computation to achieve real-time simulation of an adaptive cerebellar model composed of thousands of neurons learning the smooth tracking of different target trajectories. **Real-time simulation of a cerebellar network learning** will be available at the Sony CSL Open House.



### The sensorimotor laws of color perception (2004)

#### Understanding sensorimotor laws

Perception and motor control have been considered as separate issues for a long time. More and more scientists realize that this is not the case, but a framework to bring them together has remained difficult to establish.

In the past year, we have been working on this problem and have been leading the emergence of a framework to unify perception and action. The framework suggests general methods to formalize the *sensory-motor laws* from which we can identify synergies between the different sensory modalities and motor control involved.

This framework has recently been used to characterize the human sensorimotor laws associated with the manipulation of coloured surfaces. As it turns out, a novel way to categorize colours has emerged from analyzing these laws. More interestingly even, the categories correspond to *natural* human colour categories. These findings, although based on colours only, point out that we may need to reconsider some assumptions about the nature of perception in general.

This work will be presented and debated at the Sony CSL Open House and is being done in collaboration with the CNRS and the ENS in Paris.

Another recent breakthrough on colour concerns the interaction between colour categorisation and colour naming. This work has been accepted as a target article in the prestigious Behavioral

and Brain Science Journal (17). In a series of systematic experiments in collaboration with researchers of the University of Brussels AI Laboratories, the language game interactions and cognitive mechanisms developed in the Talking Heads experiment have been shown to lead to the coordination of colour categories between agents and thus explain a fundamental puzzle in cognitive science: how is it possible that individuals share the same categories if these are not innate.

## References:

(1) Steels, L. and Kaplan, F. [Bootstrapping grounded word semantics](#). In Briscoe, T., editor, Linguistic evolution through language acquisition: formal and computational models, pages 53-73, Cambridge University Press. Cambridge, 2002.

Steels, L., Kaplan, F., McIntyre, A. and Van Looveren, J. [Crucial factors in the origins of word-meaning](#). In Wray, A., editor, The Transition to Language, pages 252-271, Oxford University Press. Oxford, UK, 2002.

Steels, L. The Talking Head Experiment, Limited edition for the Laboratorium exhibition.

Kaplan, F. La naissance d'une langue chez les robots, Hermès Science, 2001.

The Talking Heads Experiment web site : <http://talking-heads.csl.sony.fr/>

(2) Oudeyer, P-Y. Origins and Learnability of Syllable Systems, a Cultural Evolutionary Model, in Artificial Evolution, pp. 143-155, eds. P. Collet, C. Fonlupt, J.K. Hao, E. Lutton, M. Schonenaus, LNCS 2310, Springer Verlag.

Videos available on: <http://www.csl.sony.fr/~py/videos.html>

(3) Steels, L. and Kaplan, F. [AIBO's first words: The social learning of language and meaning](#). Evolution of Communication, 4(1):3-32 2001.

More information of the Talking AIBO Project: <http://www.csl.sony.fr/~kaplan>

(4) Oudeyer P-Y. (2002) The production and recognition of emotions in speech: features and algorithms, International Journal in Human-Computer Studies, vol. 59/1-2, pp. 157-183, special issue on Affective Computing.

(5) Qrio web site: <http://www.sony.net/SonyInfo/QRIO/>

(6) Steels, L. [Constructivist Development of Grounded Construction Grammars](#). In Daelemans, W., editor, Proceedings Annual Meeting of Association for Computational Linguistics, 2004.

(7) Oudeyer P-Y. (2005) From Holistic to Discrete Vocalizations, to appear in the volume "Language Origins: Perspectives on Evolution", ed. Tallerman M., Oxford University Press.

Oudeyer, P-Y. (to appear as a book) The Self-Organization of Speech Sounds, Studies in the Evolution of Language, Oxford University Press.

The Origins of Speech Sounds Project web site: [www.csl.sony.fr/~py](http://www.csl.sony.fr/~py)

(8) Oudeyer, P-Y. and Kaplan, F. [Intelligent adaptive curiosity: a source of self-development](#). In Luc Berthouze and Hideki Kozima and Christopher G. Prince and Giulio Sandini and Georgi Stojanov and G. Metta and C. Balkenius, editor, Proceedings of the 4th International Workshop on Epigenetic Robotics, vol. 117, pages 127-130, 2004. Lund University Cognitive Studies.

The Playground Experiment web site: <http://playground.csl.sony.fr>

(10) Pachet, F. and Delerue, O. (2000) [On-The-Fly Multi-Track Mixing](#). Proceedings of AES 109th Convention, Los Angeles, 2000. AES.

(11) Pachet, F. (2004) [Enhancing Individual Creativity with Interactive Musical Reflective Systems](#). Psychology Press.

(12) Pachet, F., Addressi, Anna-Rita (2004) [When Children Reflect on Their Playing Style: The Continuator](#). ACM Computers in Entertainment, 1(2)

(13) Pachet, F. (2003) [Content Management for Electronic Music Distribution: The Real Issues](#). Communications of the ACM, April

(14) Aucouturier, J.-J., Pachet, F. and Sandler, M. (2004) [The Way It Sounds : Timbre Models For Analysis and Retrieval of Polyphonic Music Signals](#). IEEE Transactions of Multimedia.

(15) Zils, A. and Pachet, F. (2004) [Automatic Extraction of Music Descriptors from Acoustic Signals using EDS](#). Proceedings of the 116th AES Convention, May

(17) Steels, L. and T. Belpaeme (2005) Coordinating Perceptually Grounded Categories through Language. A Case Study for Colour. Target article Behavioral and Brain Sciences. To appear.