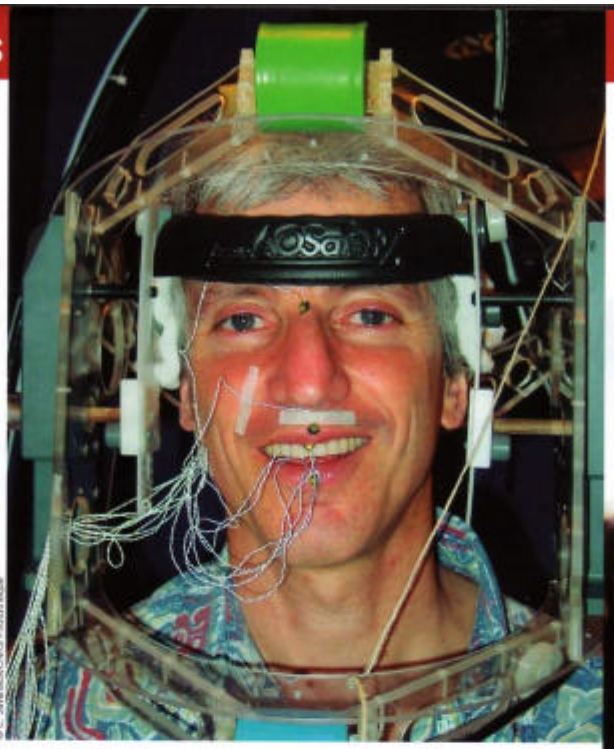
In Grenoble, in the heart of the Alps, researchers from a variety of disciplines are exploring human speech: a phenomenon at the interface of the intimate and the social domains, and whose complexity is often forgotten. They take us on a scientific excursion into the world of language.



© C. Servaz/CNRS Photothèque

The apparatus on the left is able to measure all speech-related articulatory events with precision. It is called an articulograph.

This device, unique in the world, functions through a water pressure system, enabling ICP researchers to study the vocal cords with precision.

© C. Servaz/CNRS Photothèque

LANGUAGE SCIENCES

Speech communication research speaks up

The researchers at the Institut de la communication parlée' (ICP) in Grenoble talk passionately about their research: they have been working on the enigma of human speech for over twenty years. Their laboratory looks at speech in every possible way, from the study of the vocal tract to the analysis of emotions, from phonetics to the study of cerebral processes. Not to mention the different applications developed by the ICP such as robotics and telecommunications as well as the treatment of speech disorders. Are they perhaps trying to cover too many topics at the same time? Not at all, according to physicist Jean-Luc Schwartz, who is the director of ICP. "Looking into all these areas is the only way to achieve our dream, that is, to someday understand how speech 'works'."

The first obstacle to this seemingly simple quest is how air moves through our body to produce a sound and then a word. To discover this, a team from the ICP, in collaboration with the University of Eindhoven (Netherlands), created an artificial vocal tract going from the lungs (a 1m³ box which distils compressed air) to the vocal cords (two small pieces of metal). "This prototype, which is the only one like it in the world, may look rather simple at first, but in fact it allows us to reproduce a mechanical speech gesture based on the human model which can be modified and reproduced forever," explains Xavier Pelorson, who directs the Acoustics team at the ICP. "Thanks to this machine, we have demonstrated the still largely unexplained phenomena of swirls and turbulence at the larynx exit." The researchers have also managed to reproduce the phenomenon of oscillation in the vocal cords which makes up the sound of a voice. Another major exploit has been to measure the vibrations of the cords using a laser. The medical field has already started to use this knowledge in the treatment of voice disorders. Somewhat more unexpected is the connection with sleep apnea. "In this disorder, breathing stops due to partial obstruction of the upper airways," explains Annemie Van Hirtum, a researcher with the Acoustics team. With the ICP device it is possible to study the disorder in detail. Collaboration between the TIMC laboratory[2] and the Grenoble teaching hospital will hopefully lead to the development of software to help surgeons operate on such cases – the current success rate is only 50%.

Medicine is not the only domain concerned though. "Traditional methods of voice synthesis give excellent results for male spoken voices, but not for those of women or children, nor for singing," observes Xavier Pelorson. "In this case, physical modeling is probably the best option." Of course, there is a long way to go before we hear a song leave the "mouth" of the prototype. "In particular, we need to analyze the contact mechanisms of the vocal cords," explains Nicolas Ruty, a Ph.D. student at the ICP. "They are what generate the high frequencies which are so characteristic of the human voice," he adds.

Speech and music are obviously closely linked. Helene Loevenbruck, a linguist at the ICP, knows all about

LANGUAGE AT CNRS

Linguistics is a major research area for CNRS. On November 26, 2004, scientists met at CNRS headquarters in Paris to give an overview of their field. You can find the names of the 40 laboratories which took part, the presentations and activities, together with the photographs taken during the meeting at the following link: **www.cnrs.fr/SHS/actions/linguistique.php**

M. R.

the melody that shapes our utterances. This is known as prosody, a term covering phenomena such as intonation, stress, rhythm and phrasing. "One sentence can have four or five different meanings. It's the prosody which allows us to generally determine the right one," she comments. Each language has a specific prosody, learned from birth. This is used to prioritize information and, more surprisingly, to decide whose turn it is to speak. "Margaret Thatcher, for example, always used falling intonation at the end of her sentences to make them more substantial," says Loevenbruck, who does a wonderful imitation of the famous Iron Lady. "But she was always being interrupted because her interlocutors thought she'd finished speaking at the end of every sentence!" On an everyday basis, Loevenbruck works on focus, in other words, how an element in a statement is emphasized. Intonation, articulation and the level of the voice are all used to make others understand. "One of our studies showed that people perceive focus as key information and understand it extremely well." Perhaps that is why this skill is developed early. "A baby points its finger at an object and says 'more'," Loevenbruck explains. "Gradually, the finger is replaced by the vocal apparatus, but both types of 'pointing' are similar." This was further demonstrated when the ICP researchers recently proved that the activated areas of the brain were the same in both cases. In adults, the process also involves a number of facial movements. At the ICP, all these 'mini gestures' are analyzed using an electromagnetic 'articulograph'.

A few doors away, a team of researchers are working on 'speaking machines'. "By 'speaking' we mean two things," says Gerard Bailly, who leads the team. First of all sound, obviously, but also everything you can observe in the faces of the two speakers. A dominating factor in a face to face discussion is mutual attention. "When I speak to you, for example I watch your eyes to see whether you are following what I say, if you understand me and the type of effect my words have on you," the researcher explains. "But when I listen to you, my glance moves between your eyes and your lips." No, your interlocutors are not going to point out to you that you have a piece of lettuce wedged between your teeth. They are just trying to confirm what they have heard. "We all know how to lip-read," Schwartz points out. "And we use this skill in every conversation."

This woman listens to you and watches you closely. But don't be fooled, she is entirely virtual.

© E. Perez/CNRS Photothèque

WHEN THE BRAIN SPEAKS

For linguistics, progress in brain imaging has been a real godsend. Observations have now made it possible to understand many of the mechanisms related to speech. Jean-Luc Schwartz gives us an example. "If I say to you: life, life, life, life, life, life, it's highly likely that, at some point, you'll hear fly, fly, fly." ICP researchers recently highlighted the cerebral networks' involved in this shift in meaning, a well-known phenomenon for specialists, who have named it 'the verbal transformation effect.' Their discovery is part of a long-term ongoing project. Indeed a considerable amount of research still remains to be done on the link between the systems of production and perception of speech.

M. R.

1. Sato M. et al. Neuroimage, vol. 23, 2004, pp. 1143-1151.

Scientists are now trying to create virtual characters capable of this type of exchange, and the demonstration given to visitors is impressive. On the screen, a woman's face answers your questions, all the while keeping her eyes fixed on you. The secret lies in laser sensors positioned all around the screen which can locate the back of your eyes. To create this character, the scientists covered the face of a real woman with almost 400 microspheres and then measured the movements for all possible combinations of two phonemes, in other words nearly a thousand. For the lips, there's no problem: the ICP has developed a highly effective method to study and model lip movements (see pictures). And that's not all. The ICP is currently working on creating a character for the Arte TV channel[3]. Its role will be to replace the teletext system for the deaf and hard of hearing. The virtual presenter will not use sign language but a form of "cued speech", where "hand gestures act as cues to complement the movements of the lips," Schwartz explains. To model these gestures, the same method was used: our modern Pygmalions covered a woman's face and hands with microspheres and got her to read around 230 sentences which covered virtually all sounds, using cued speech. At the moment they are trying to equip their Galatea with a virtual tongue that you would almost swear you could see speaking. In 2002 the researchers developed one of the most powerful functional tongue models in the world. But they have an even more ambitious dream. "Our aim is to make two people communicate at a distance using virtual clones which echo their discussion," Bailly explains with a smile. To do that, the ICP is preparing a very special experiment room where human subjects will be locked in with a virtual character on a screen. Behind the observation windows, the researchers will be able to draw valuable conclusions to improve their computer clones. And no doubt they will share the results over a hot drink between two experiments with their 60-odd ICP colleagues. At the entrance to the laboratory, the coffee room is permanently filled with the sound of lively discussions. One thing is sure: in this bustling research center, speech will be around for a good many years to come.

Matthieu Ravaud

1. The Institute for Spoken Communication. Joint unit: CNRS / the Institut National Polytechnique de Grenoble (INPG) and Stendhal University of Grenoble.
2. Techniques de l'imagerie, de la modélisation et de la cognition (TIMC – Imaging, modeling and cognition techniques)
3. In collaboration with the Thales group.

© C. Savariaux/CNRS Photothèque

To model the movements of the lips, the researchers have found a simple solution: blue makeup. Using a traditional system of image processing and the well-known method of placing a backdrop behind the weather forecaster, they can obtain a perfect image of their movements.

CONTACTS

ICP, Grenoble

→ **Gérard Bailly**
bailly@icp.inpg.fr

→ **Hélène Loevenbruck**
loeven@icp.inpg.fr

→ **Xavier Pelorson**
pelorson@icp.inpg.fr

→ **Jean-Luc Schwartz**
schwartz@icp.inpg.fr

→ www.icp.inpg.fr