Natural interactivity

Recently, natural interactivity, or natural interaction, has become a buzzword which, not least in Europe but also elsewhere, and even by Bill Gates, is being used so frequently that there can be no doubt that natural interaction is viewed as *a good thing*. However, there has been precious little discussion yet as to the nature and limitations of natural interactivity. Similarly, once a new catchy term appears on the horizon destined to lead a short or sometimes longer life in the limelight, we wonder about what the term might mean in relation to the concepts which already hold, or puzzle, our imaginations of the future roles of computing and communication systems in our lives. Natural interactivity appears as if it has come to stay, so it may be worthwhile to take a look at what it is.

What is natural interactivity

The *interaction* part of 'natural interactivity' refers to interaction between humans and computer systems as well as to interaction between humans. For the time being, at least, all or most interaction between humans and computer systems consists in exchange of information through the use of various input and output devices, such as keyboards, screens, pens, cameras, microphones and various kinds of sensors. Sometimes people also interact with systems by smashing the hardware but such doings are not included in our mainstream concept of interacting with computer systems. What *natural* does is to qualify interaction in a particular way, as being a natural way of exchanging information with computer systems. Which way is that? It is *the ways in which humans normally, or by and large, exchange information with one another.*

Of course, just like the computer device user who, after a protracted phase of attempted peaceful negotiation looses patience and smashes the thing, humans interact to do many other things together in addition to exchanging information, such as making love, or war, but there is no doubt that exchange of information with other humans is a basic aspect of human life for which humans are naturally endowed. This aspect is supported by a range of skills all of which most humans have. When we exercise those skills, we exchange information in what to us are natural ways, perceptually, motorically, and in terms of the patterns of reasoning involved. And even if some individuals do not have all of those skills, they can still exchange information in natural ways by using the skills which they actually possess.

Natural interactivity exemplified

We are thoroughly familiar with natural human-human exchange of information, as in this scenario: two people discuss and solve an architectural design problem using photos and layout drawings, making sketches, hand-writing notes, inspecting typed memos and encircling important points with a pen and with their fingers, handling, modifying and labelling a 3D model, solving a geometrical problem together on paper, etc. They put red marks on the items which need to be discussed with colleagues later on. In the course of the discussion they nod, smile, look puzzled, hesitate, etc., all of which is being perceived by the interlocutor as an integral part of the information being exchanged. Towards the end of the session, they recognise the voice of a colleague in the hallway and call on her

to inform her of the progress they have made. They have not been using computing gear at all throughout the session.

Suppose that the two people in the scenario are supported by a system which participates in the discussion on more or less equal terms. The system takes part in the oral discussion, perceives what the humans perceive, more or less, handles 3D graphics versions of the objects which the humans handle physically, expresses surprise, support, and other mental states through a graphical speaking face, spots a puzzled face on occasion, etc. Actually or conceivably, the system could augment the problem-solving exercise in various ways, making it more efficient, more comprehensive, or better evaluated, for instance by rapidly retrieving from various networks almost any kind of information which is needed in the discussion; rapidly connecting the discussants with colleagues and experts from all over the world who could join into the shared workspace; performing highly complex computations on request; quickly generating a variety of solution options; storing the discussion and its results; summarising the discussion for later access; and much more. Incidentally, the humans would no longer have to be in the same location for everything to happen as in the "old days" scenario above. Clearly, the system could do many things faster and easier than the humans could by on their own. In other respects, the system would probably be inferior to the people who would want to make the important decisions themselves instead of leaving those to the system. The scenario just presented is an example of *natural human-human-system interaction* (HHSI) in which the system's role approximates that of a super-human assistant.

Natural interaction as vision

Today's systems cannot do all of the things described in the natural HHSI scenario above. For instance, we are not yet that far in conversational spoken language dialogue systems technology, in machine vision-based situation interpretation technology, in on-line understanding and expression by machine of prosody, facial expression and gesture, in agent technologies, in application sharing technologies, in multimodal input fusion and output fission technologies, in summarisation technology, or even in the handling of speech over the Internet. In fact, to get as far as described will require very substantial long-term research, partly in areas where we have only scratched the surface today.

Thus, natural HHSI expresses a *vision* about interaction (or about information exchange). According to this vision, interaction with computer systems will eventually become as natural as interaction among humans. What is more, the vision appears to be a *necessary* one. It is not just a vision amongst others but a necessary end-projection from the state-of-the-art, given the nature of human communication. This is probably why natural HHSI is becoming a powerful long-term target which provides an integral model for hitherto widely separate efforts and communities in research and technology development. One example is the European Industry's advisory document on how to implement the EU's 5th Framework Programme (FP5) from the year 2000 onwards (ISTAG 1999). In the world discussed here, 1999 is already a long time ago, of course, but the vision of natural interactivity is also perceptible in the first steps towards FP6, such as ISTAG's and the Commission's plans to brainstorm this autumn on "Scenarios for Ambient Intelligence - circa 2010" in order to articulate a vision for the Information Society Technologies (IST) Programme for FP6.

Corresponding to its potential for integrating hitherto separate research communities, and to its inherent complexity, the natural HHSI model invites a "think big" approach, or invites a transformation of systems research from small-to-medium scale science into medium-to-large scale science. We know where we want to end up, we know that the problem is a large and complex one, and we know what many of the necessary steps, each representing a serious research challenge, are going to be - so, let's get organised to achieve as large chunks of the vision as possible!

The Vision Chunked

Chunks of the vision are apparent in a series of "think medium-to-big" research programmes world-wide. Here are some of them.

DARPA Communicator

DARPA Communicator (http://fofoca.mitre.org/) is a US stab at a chunk of the natural HHSI vision. The goal is to build the next generation of intelligent multi-party conversational interfaces to distributed information by creating speech-enabled interfaces that scale gracefully across modalities, from speech-only to multimodal interfaces that include graphics, maps, pointing and gesture. This 20 Mio. $ US/year programme was launched by DARPA and NSF in 1998. A positive innovation is that Communicator has invited a number of European affiliates to join. The Communicator architecture which is based on MIT Speech Lab's Galaxy, will extend emerging speech and language standards to support conversational interaction through the use of telephones, mobile wireless, PDAs etc.

Oxygen

Oxygen is an MIT Computer Science Lab. project which was announced in August 1999 (Scientific American). Oxygen does not focus squarely on natural interaction with computer systems as does the Communicator. Rather, Oxygen takes Communicator for granted and focuses on the development of a global infrastructure for technology-mediated human-human communication. This involves building what is claimed to be a new form of hand-held device which combines cellular phone technology with a visual display, a camera, infrared detectors and a computer; and a new local device which does what the hand-held one does, but faster, and keeps track of people locally. A novel form of network will link the devices.

Jointly, DARPA Communicator and Oxygen address an important chunk of the natural HHSI challenge.

SmartKom

Launched in 1999, SmartKom (http://www.dfki.de/smartkom/) is a German project worth approx. 50 Mio. Deutschmarks. SmartKom focuses on natural interactivity and multimodal interfaces, starting from spoken dialogue like the DARPA Communicator. A minor difference from Communicator is SmartKom's emphasis on individual adaptivity and cartoon-like presentation agents. SmartKom envisions three different human-human-system communication technologies: the Public Booth, offering videophone and web access; SmartKom Mobile, offering web access; and SmartKom Home/Office, offering enhanced functionality compared to current PCs. SmartKom is at the intersection of

Communicator and Oxygen. The i3 project Magic Lounge launched in 1997 is among the origins of SmartKom (http://www.dfki.de/imedia/mlounge/).

CLASS

True to form, EU's IST programme generally takes a distributed, rather than a chunking, approach to natural HHSI. As natural interactivity has now made top priority in FP5, most special IST programmes invite projects which include natural interactivity aims. The project closest to the chunking approach may be CLASS (http://www.class-tech.org/), a Human Language Technologies experimental project which started in July, 2000. CLASS coordinates technical cooperation among 25+ research projects launched in 2000 and organised into three clusters. One cluster in particular, on Natural and Multimodal Interactivity, includes projects which address natural HHSI in the same general domain as DARPA Communicator. The cluster will specify a reference platform and architecture for next generation natural interactive systems and investigate best practice in development and evaluation for natural interactive systems.

Conclusion

None of the above endeavours will achieve the vision of natural HHSI. Significantly, some of the programmes appear to represent a new creature in the systems research world, namely that of *market-driven fundamental research.* In this kind of research, it is a matter of getting the technology out there fast and before anyone else in the hope of setting de facto standards, with only back-seat space being provided for investigating the multitude of complex and fascinating, unsolved issues that currently prohibit full natural HHSI. This having been said, the chunking approach does seem appropriate when a research vision can be systematically decomposed in an ordered series of steps.

Natural Interactivity and Multimodality

With good reason, many people are confused by the relationship between 'natural interactivity' and another buzzword, 'multimodality', which has been around for a decade. Natural interactivity is multimodal most of the time, because it involves a range of modalities for information exchange, such as speech, pointing gesture and facial expression. A multimodal system, on the other hand, is not necessarily a natural interactive system. Multimodality in a system merely signifies that users may, or must, exchange information with the system using several different input and/or output modalities (Bernsen, 1994; Benoit et al., 2000). The modalities themselves need not be natural for humans. There is nothing particularly natural about a double-click with the mouse, yet this haptic notation input modality forms a necessary part of many multimodal interactive set-ups. Perhaps, multimodal systems development can be said to address a fraction of the much taller research agenda imposed by the vision of natural HHSI.

"How about natural interactivity and the Disappearing Computer?", I can hear somebody asking. They are overlapping research agendas.

References

Benoit, C., Martin, J. C., Pelachaud, C., Schomaker, L., and Suhm, B., 2000. Audio-Visual and Multimodal Speech Systems. To appear in D. Gibbon (ed.), *Handbook of*

Standards and Resources for Spoken Language Systems - Supplement Volume. Kluwer.

Bernsen, N. O., 1994. Foundations of multimodal representation. A taxonomy of representational modalities. *Interacting with Computers* 6, 4, 347-71.

CLASS: http://www.class-tech.org/

DARPA Communicator: http://fofoca.mitre.org/

Information Society Technologies Advisory Group (ISTAG): Orientations for Workprogramme 2000 and beyond. Draft Report July 1999.

Magic Lounge: http://www.dfki.de/imedia/mlounge/

Oxygen: *Scientific American,* August 1999, 36-47.

SmartKom: http://www.dfki.de/smartkom/

About the author

Prof. Niels Ole Bernsen is director of the Natural Interactive Systems Laboratory at the University of Southern Denmark, coordinator of i3net, affiliate of DARPA Communicator, partner in Magic Lounge and coordinator of CLASS.