What is Natural Interactivity?

Niels Ole Bernsen

Natural Interactive Systems Laboratory
University of Southern Denmark
Science Park 10, 5230 Odense M, Denmark
nob@nis.sdu.dk

Abstract

Recently, natural interactivity, or natural interaction, has become a buzzword which is being used so frequently that there can be no doubt that natural interaction is viewed as *a good thing.* This paper analyses the nature, realisation and limitations of natural human-human-system interaction. It is argued that natural interaction represents a necessary and integral vision for large and hitherto separate areas of advanced interactive systems research.

1. Introduction

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.* This paper analyses the nature, realisation and limitations of natural human-human-system interaction. It is argued that natural interaction represents a necessary and integral vision for large and hitherto separate areas of advanced systems research. The effects of the vision can already be seen in a series of large-scale research efforts world-wide. As few people seem to know what is the difference, or even the relationship, between natural interactivity and multimodality, this potential confusion is addressed. Finally, it is argued that, like any other paradigm of interaction, natural interactivity is not a perfectly general one which offers a goalpost or, at least, a sense of direction for the solution to any conceivable interaction design problem.

2. Natural Interaction Defined and Illustrated

The 'interaction' part of the phrase 'natural interactivity' refers to interaction between humans and computer systems. For the time being and in the sense of 'interaction' we are after, interaction between humans and computer systems consists in, and only in, exchange of information through the use of various input and output devices, such as keyboards, screens, pens, cameras and microphones. It follows that the term 'natural' (interaction) is intended to qualify interaction in a particular way, i.e. as being a natural way of exchanging information with computer systems. So, which way is that? Speaking generally, it is *the way(s) in which humans normally, or by and large, exchange information with one another.* Humans do other things together than 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 basic aspect is supported by a range of skills all of which most humans have. When humans exercise those skills, they exchange information in what to them are natural ways – perceptually, motorically and in terms of the patterns of reasoning involved. And even if particular human individuals do not have all of those skills, they can still exchange information in natural ways by using the skills which they actually possess.

The following is a familiar scenario which demonstrates natural human-human information exchange. Two people discuss and solve an architectural design problem using photos and layout drawings, making sketches, handwriting 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 part of the information that is being exchanged. At the end of the session, they recognise the voice of a colleague in the hallway and call on her to inform her on the results they have reached. In this scenario, the humans don't use computer systems at all.

Now suppose that the two people in the scenario are supported by a computer system which participates in the discussion on more or less equal terms. The system participates in the 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, etc. Actually or conceivably, the system could augment, that is, make more efficient, more comprehensive, better evaluated, etc., the problem-solving exercise in various ways: by rapidly retrieving almost any kind of information which is needed in the design discussion via various networks; 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; rapidly generating a variety of solution options for inspection; storing the discussion and its results; summarising the discussion for later access; and more. In many cases, the computer system could do things much faster and easier than the humans. In other respects, the system would probably be inferior to the people present 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 an extremely capable assistant or servant.

3. Natural Interaction as a Vision

Today's computer 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 scene interpretation technology, in the 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 in the scenario above will require very substantial long-term research, partly in areas where we have only scratched the surface today.

This means that natural HHSI expresses a *vision* about interaction (or about exchange of information). According to the vision, interaction with computer systems will eventually become as natural as interaction between humans. Moreover, the vision appears to be a *necessary* one. It is not just a vision amongst others but one which is 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 appears set to provide an integral model for hitherto widely separate efforts in research and technology development. One example of this focusing process 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). Corresponding to its integration potential and inherent complexity, the natural HHSI model straightforwardly invites a "think big" approach, or invites a transformation of ITC research from small-to-medium scale science into medium-to-large scale science. Why? Because 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 are going to be - just like when mankind wanted to put a man on the moon some decades ago. So, let's get organised on a large scale to achieve as large chunks of the vision as possible!

4. The Vision Chunked

Some chunks of the vision are clearly visible in a series of "think medium-to-big" research programmes world-wide. This section briefly presents some of them.

4.1. DARPA Communicator

The DARPA Communicator (http://fofoca.mitre.org/) is a US stab at a chunk of the natural HHSI vision. The goal is to foster the next generation of intelligent multi-party conversational interfaces to distributed information, i.e. to support the creation of emerging standards-conformant, 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 and involved at its launch four leading US research labs and four US companies. The DARPA Communicator is based on a state-of-the-art platform/-architecture (Galaxy) provided by MIT's Speech Lab. The Communicator architecture will build on emerging commercial standards in the speech and language areas and extend these to support intelligent multi-party conversational interaction through the use of telephones, mobile wireless, PDAs etc. The architecture which is to be further developed together with infrastructure and a software repository, will act as a focal point for rapid development of new interfaces, providing access to new sources of information, as well as for competitive development and evaluation of systems and components.

4.2. Oxygen

The Oxygen project is an MIT Computer Science Lab. project which was announced in August 1999 (Scientific American). Like the DARPA Communicator, Oxygen is based on the research platform provided by MIT's Speech Lab. Oxygen, however, does not focus squarely on natural interaction with computer systems as does the Communicator. Rather, Oxygen takes the Communicator programme for granted and focuses on the development of a global infrastructure for technology-mediated human-to-human communication. This will involve building what is claimed to become an entirely new form of hand-held device, the Handy 21, which combines cellular phone technology with a visual display, a camera, infrared detectors and a computer; and a new local device, the Enviro 21, which does what Handy 21 does, but faster, and in addition keeps track of people locally. Based on novel communication protocols, a novel form of network, Net 21, will link the Handy and the Enviro.

4.3. SmartKom

SmartKom (http://www.dfki.de/smartkom/) is a German national project worth approx. 50 Mio. Deutschmarks (30 MEuros) which was launched in 1999 and involves 11 partners from academia and (mostly) industry. SmartKom focuses on natural interactivity (starting from spoken dialogue) and multimodal interfaces. A minor difference to the DARPA Communicator is that SmartKom emphasises individual adaptivity and cartoon-like presentation agents. SmartKom envisions three different human-human-system communication technologies, i.e. the Public Booth offering videophone, web access etc., the SmartKom Mobile, offering web access etc., and the SmartKom Home/Office offering strongly enhanced functionality compared to current PCs. In general, SmartKom's focus is at the intersection of the DARPA Communicator and Oxygen. Like Oxygen, SmartKom is a project in search for new basic platforms, architectures and devices.

4.4. CLASS

CLASS (Bernsen, 1999) is a European Human Language Technologies project which is starting in June, 2000. The innovative experiment embodied in CLASS is to coordinate technical cooperation among 18 RTD projects launched in 2000 and organised into three clusters. In particular, one CLASS cluster, on Natural and Multimodal Interactivity, includes six projects which address natural HHSI in the same general domain as the DARPA Communicator. This cluster aims to specify a reference platform and architecture for next generation natural interactive systems as well as to develop a best practice development methodology for natural interactive systems.

4.5. Conclusion

Evidently, none of the projects or programmes briefly reviewed above will achieve the vision of natural HHSI. I know of colleagues who do not wish to join the more centrally organised among those programmes even though they could if they wanted to. Also, to some significant extent it might be said that some of the programmes seem to represent a new creature in the IT research world, i.e. that of *market-driven fundamental research.* In this kind of research, it's a matter of getting the technology out there fast and before everyone else, and of setting de facto standards, rather than of investigating the multitude of complex and fascinating, unsolved issues which currently prohibit full natural interactivity. Still, all of those programmes can be viewed as being strongly motivated by the vision of natural HHSI and there is little doubt that they will generate new knowledge.

5. Natural Interactivity and Multimodality

What is the relationship between natural interactivity and another buzzword, one which has been around for about ten years now, i.e. multimodality? Natural interactivity is multimodal most of the time, of course. But a multimodal system 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 at all. 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. The next section mentions several other examples of useful albeit less-than-natural multimodal interactivity.

6. Limitations of the HHSI Vision

There is an issue about the limitations to the vision of natural HHSI. In its unlimited version, the vision would amount to the claim that, eventually, any exchange of information with computer systems for any conceivable purpose can be done through natural HHSI or by communicating with systems in the same ways as we communicate with our fellow humans during problem solving and otherwise.

It is not evident, though, that all HHSI will eventually become natural HHSI. For instance, we don't know whether or not the workstation will go away completely or, alternatively, whether or not it will morph into something fully natural. Moreover, it is well known that multimodality offers a wealth of opportunities for designing useful interaction for people with reduced interactive skills. Examples are text editing for the blind using Braille in combination with speech, mobile communication for people with speaking difficulties, using new kinds of keyboard instead of speech input, etc. We know that speech is not well suited for conveying spatial information, and that (keyboard-typed) text is inferior in some respects to situated linguistic communication, such as spoken discourse (Bernsen, 1997). However, the point is that when people cannot use the modalities which are optimal for conveying particular kinds of information in order to carry out a certain task, new multimodal input/-output combinations can make it possible for them to accomplish the task nevertheless. It is more important to get the task done than to remain within the confines of natural interactivity.

Furthermore, we may not always be able to tell whether or not a particular form of HHSI is natural. Even if natural interactivity appears to be a necessary new paradigm for human-human-system interaction, that does not necessarily make it a universal classification procedure for interaction designers to use. Many useful future forms of interaction might be such that we cannot really tell whether or not they are natural. Consider the contemporary case of someone providing navigational input information into a 3D graphics output environment. The person is visiting, say, the local department store in virtual space looking for a pair of sports shoes. On the screen, the store looks just like it is in reality - nothing out of place on all three floors of it including the sports goods shop on the third floor. Now, to get to the third floor, will the visitor put the haptic pointer on the escalator and wait for it to move two floors up to the sports shop? Hardly! Whatever the chosen solution for navigating the department store, the navigation is not likely to copy a real person's navigation in a real building. Does that make the navigation a case of *non-natural* interaction? It is not clear that there is a 'yes' or 'no' answer to that question.

In other words, like all other interaction paradigms to date, the paradigm of natural HHSI has its limitations. In the case of natural HHSI, there seem to be both conceptual limitations and possible limitations of scope. This means that the paradigm cannot guide the design of interaction in general.

7. 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.

Bernsen, N. O., 1997. Towards a tool for predicting speech functionality. *Speech Communication*, 23:181-210.

Bernsen, N. O., 1999. CLASS – Collaboration in LAnguage and Speech Science and technology. *Elsnews* 8, 4, 6-7.

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

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

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

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