Touchless Interaction

Øyvind Kurisaki-Sagberg
Department of Product Design
Norwegian University of Science and Technology

ABSTRACT

Recent years have seen a growth in the introduction of new input devices for touchless interaction. This article discusses the characteristics of touchless interaction in light of a re-evaluation of "naturalness" and "intuitiveness", arguing that the search for applications for touchless interaction should begin in analogies with human-human interaction.

KEYWORDS: Touchless Interaction, Gestures

1. BACKGROUND

Technologies for touchless interaction have existed in various stages of development for many years, but in recent years devices have begun to find their way out of laboratories, and appear on the consumer market at price points that make them a viable alternative to traditional input devices. Some well known examples are Microsoft's Kinect, the Leap Motion sensor, or the natural language processing of Apple's Siri.

However, the interfaces that these devices are part of often leave something to be desired. Aside from entertainment applications, the search for what these modes of interaction are really suited for is still ongoing. Speaking or gesturing in mid-air are activities vastly different from typing on a keyboard or pushing a mouse around. Yet this seems to be ignored in many interfaces, where one input device is simply swapped for another without further changes to the software side of the interface. In other cases the designers seem to be so enthralled by the novelty of the input device that they use it without considering whether the characteristics of the input modality are appropriate to the task.

A user interface is the totality of input device, output device, and the logic that interprets user action and produces feedback. In designing an interface, one needs to ensure that all parts of the interface constitute a meaningful whole, where the characteristics of each part are appropriate to the task, and in correspondence with the characteristics of the rest of the system. In order to accomplish this, we need a thorough understanding of the characteristics/properties of these modalities; we need an understanding of how they are perceived by a user in a social and cultural context; and we need a common and clear language to discuss them.

This article begins with a discussion of some analytical concepts, which will then be applied to a summary of the most important types of touchless input methods and a comparison of and reflection on their respective characteristics.

2. IN SEARCH OF ANALYTICAL CONCEPTS

2.1 Touchless Interaction

By touchless interaction we refer to human-computer interaction where the user is not in physical contact with, and does not mechanically manipulate a device, but rather uses his own body to generate signals that are perceived by sensors at a distance and interpreted by a computer.

2.2 Device vs Modality

It is important to distinguish between the input device and the input modality. An input device is an electromechanical apparatus (e.g. a mouse, a camera), while an input modality is a means of encoding information (e.g. sound, vision, smell). For a user, it is irrelevant whether a computer is able to see him using stereoscopic vision or using an IR-based depth camera (all else being equal). The user merely knows that he can be seen, and acts accordingly. For the purpose of designing an interaction, therefore, the modality is what shapes the experience, and the device should be selected after the modality has been considered.

Interactions can be unimodal, meaning they employ only one modality, or multimodal, meaning they employ several. There are two main ways of using modalities together; they can be redundant, with the same function being accessible via different modalities, or they can be complementary, with each modality accessing a set of functions. A well known example of this is the seminal «put that there»-experiment, which combined voice input and deictic gesture in the manipulation of graphics on a large display (Bolt 1980).

2.3 Suitability

Much of recent research deals primarily with improving the precision of different sensing methods, either through new hardware or improved software and algorithms. These studies are usually accompanied by some kind of evaluation of performance metrics.
Given the assumption that problems of fidelity - precision, resolution, segmentation - all the technical challenges of touchless interaction, will be overcome, there still remains the question of suitability. In the same way that a mouse is not very suitable for text input, or a keyboard is rather impractical when navigating a traditional GUI, there are some tasks that gestures, or natural language input, are more or less suited for, even though they function perfectly and without error technically.

2.3 Desired Qualities of an Interaction

In order to discuss how these touchless interaction modalities should be employed, what they are suited for, we need some kind of description of the wanted qualities of an interaction. Both in academic writings and mass media these new devices are often claimed to offer a more «natural» and «intuitive» way of interaction - words of general praise, but usually without further definition. These rather elusive terms have indeed been used by so many to mean so much that some are calling for them to be banned altogether (Arnall 2013).

However, the fact that they are so frequently employed to describe positive aspects of an interaction must mean that they do carry associations to certain qualities - qualities to which an interaction apparently should aspire. Therefore, instead of banishing these terms, I would like to clarify them and try to offer meaningful descriptions that capture and identify these implied qualities.

2.4 «Naturalness»

While often seen in the context of so-called Natural User Interfaces, it lacks clear definition, and its use has been criticized by among others Dan Norman (Norman 2010). Sadly, Norman offers little in the way of clear definition, although he seems to equate NUI with gestural interfaces.

The Merriam-Webster Online Dictionary defines «natural» as «existing in nature and not made or caused by people : coming from nature». (Dictionary 2013) The common phrase «human nature» also implies that nature is something unchangeable, something with which we are born.

2.5 «Intuitiveness»

Jef Raskin (Raskin 1994) argues that «intuitive», as it is often used, can be considered a synonym for «familiar». Phil Turner (Turner 2008) expands on Raskin's arguments and discusses intuitiveness in the context of embodiment.[1] He argues that

«… intuitive systems work because they make use of pre-existing or 'pre-compiled' action-perception (motor) routines and socially (and culturally/historically) acquired 'know-how', which frees us (…) of having to engage with the interacting rather than using the system to achieve our ends». (Turner 2008) To Turner, then, «intuitive» encompasses both physical skills and social skills.

This concept of being able to perform better by not expanding mental energy on the mechanics of an action, is mirrored in discussion of «peak performance» among athletes or musicians, or other kinds of «expert» knowledge (Dreyfus and Dreyfus 1986).

While there is significant overlap in the way the two terms are used, there are clear differences in their fundamental meaning which can guide us in trying to disentangle them.

Some of the skills to which Turner refers are based on our interaction with the physical world, interactions shaped by human biology and the laws of physics, and can be regarded as universal and unchangeable. The other set of skills - skills in interpreting specific symbols, skill at specific tasks, or knowledge of procedures or structures - can and will depend on cultural background. Within a given cultural/social group, this knowledge will also change over time.

I would offer the suggestion that in order to distinguish between the universal knowledge of the physical world (including general action-perception motor routines), and the culturally and socially acquired knowledge (including task-specific action-perception motor routines), we refer to the former as «Natural», and the latter as «Intuitive». In the much-discussed nature-nurture dichotomy, then, Naturalness comes from nature, while Intuitiveness resides in the domain of nurture.

3. DESIGNING FOR NATURALNESS AND INTUITIVENESS

It is important to note that neither «intuitiveness» nor «naturalness» are properties of the interface in itself, but rather are results of the combination of interface, task and the user in a given context so that the user's tacit knowledge becomes applicable.

Designing for «intuitiveness» thus means constructing an interface that is consistent - externally with the intended users acquired knowledge of similar tasks and actions, and internally, in that acquired knowledge of one part of the interface is applicable to the rest of the interface as well.

«Naturalness» depends on designing an interface that is in accordance with the biological preconditions of the user's muscular and sensory systems in a world governed by the laws of physics. A natural interface takes advantage of the strengths and mitigates the weaknesses of our physical bodies and mental capacities.

For a touchless interaction to be considered natural and intuitive it must be both in accordance with the human body and mind and with previous experiences. Given that the majority of users currently have little or no

[1] *For a comprehensive account of embodiment, see Dourish: «Where the Action Is» (Dourish 2001)*

previous experience with touchless human-computer interaction, I would argue that we need to search for usage patterns in real world interactions, or find applicable metaphors that relate to our previous real-world interaction experience.

4. TWO FORMS OF INTERACTION

Humans interaction with nature is primarily mechanical: We apply force to physical objects, and they move (or remain motionless) accordingly. Talking to a brick wall is not expected to produce a very fulfilling conversational experience. Likewise, we tend to interact with other humans largely without touch. With a few exceptions, touching other humans are seen as inappropriate, and applying mechanical force is usually punishable by law.

In the light of the terms discussed initially it is interesting to note that most human-computer interactions are currently touch-based (mechanical), unimodal, and classified according to device. Human-human interactions, on the other hand, are primarily touchless, inherently multi-modal, and usually analyzed by modality.

The notion of different forms of interaction being acceptable with intelligent and inanimate objects is an interesting one, and its implications for the larger context of human-computer interaction could be the subject of another article.

There is, however, one conception of man interacting with and manipulating inanimate nature without touch that is seen as both acceptable and efficaceous, and it has been in existence for millennia: Magic. The fact that touchless interaction promises to make this age-old dream come true could go a long way towards explaining the fascination many of us experience when exploring these interfaces.

5. CHARACTERISTICS AND PROPERTIES

In this section, I present some subcategories of the modalities under consideration, leading up to a discussion of their respective perception and characteristics.

5.1 Main Categories

Human communication is largely visual and auditory: Consequently, the two most important types of touchless input are computer vision and computer audition.

Audition can be used to recognize the location of the user, the identity of the user, or natural language utterances. Vision can be used to detect user location, gaze and gestures, as well as emotion and identity through facial recognition.

Gestures deserve special attention, as it is a wide term, and has been used to refer to everything from any kind of body movement (Cadoz 1994, referenced in Buxton 2009-present) to finger movements on a touch-sensitive surface (everyday usage).
Buxton(Buxton 2009-present) summarizes several taxonomies of gesture and correlates them with a model of speech-framed gestures called Kendon's Continuum (fig.1)

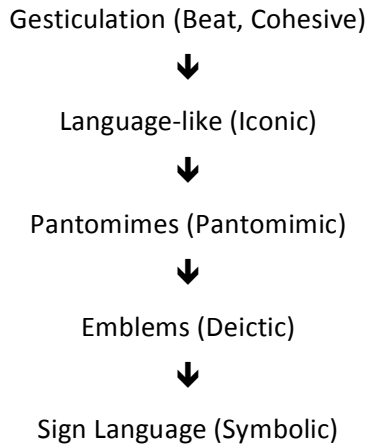

Figure 1: Kendon's Continuum (Kendon 1988)

If we consider gestures unimodally, i.e. without the speech component, we see that the model

can be further simplified into two main categories: deictic gestures (comprising deictic and iconic gestures), which indicate position or scalar value, or symbolic gestures (including sign language, pantomimic gesture), which must be interpreted in order to extract meaning.[2]
Furthermore, we see that deictic gestures have a direct mapping to a parameter (virtual or physical), whereas symbolic gestures do not.

These two categories could also be broken down further into static-dynamic subcategories. While this is highly relevant when it comes to the actual pairing of a specific task to an input modality, it does not affect the general nature - deictic or symbolic - of the interaction from the user's point of view; whether a symbol is static or dynamic is a question of interpretational complexity, not of what is being communicated. Likewise, a user perceives pointing at a single position or pointing at several positions in succession to indicate a group or an area as fundamentally the same action, and different from, say, giving a thumbs up sign. In this article, therefore, gesture will not be further divided than deictic vs. symbolic.

5.2 General Characteristics

One thing that all methods of touchless interaction have in common is the fact that they do not require the user to be in physical contact with the sensor. While this is, in essence, a tautology, the fact is it has several important implications for the practical applications of touchless interfaces.

1. The locus of interaction becomes indeterminate
2. The user is given freedom of movement
3. Physical constraints or indicators are removed
4. There is a decoupling of action-reaction

The first two points have a bearing on the type of scenario in which these technologies can find a space, and it would seem that this technology is ideally suited for reactive environments (Buxton 1997), the ambientROOM and metaDESK concepts developed at MiT's Media Lab (Ishii and Ullmer 1997), or as modes of interaction with ubiquitous technology (Weiser 1991)(Greenfield 2006).
It also allows for peripheral/background interaction, as discussed by Buxton (Buxton 1995)

The latter two points indicate some of the challenges that these modalities present. Unlike e.g. a WIMP (Windows, Icons, Mouse, Pointer) interface, there are no inherent means of discovering possible actions and mechanisms. The decoupling between action and reaction also means that the method of providing the user with feedback and input confirmation must be carefully considered.

5.3 Individual Characteristics

There are fundamental differences to the way video and audio is used, perceived and processed. For example, determining location by computer vision requires no conscious action from the user, while doing the same using audio requires the user to actively produce sounds.
There is one other important difference as well: vision has a direction and a point of view, while audio is omnidirectional. Granted, these differences are not without modifications - audio can be attenuated or occluded by objects in space, and omnivision can be attained by distributing visual sensors over a room. But from a human experience-based, «natural» perspective, this is how the modalities are understood.

All the subcategories have their own diverse characteristics; gaze tracking and symbolic hand gesture recognition are both accomplished by means of computer vision, but they have very different meanings and characteristics from a user's perspective.

[2] *Beat and cohesive gestures do not appear without speech, and are therefore not relevant in a gesture-only scenario*

Table 1: Input Modalities and Key Characteristics

	Sender	Recipient	Organ used	Key action
Visual				
Location	Unconscious	Little interpretation, low contextual knowledge	whole body	Standing next to an object you interact with, walking towards a door
Gaze	Unconscious	low contextual knowledge	eyes	Looking at an object of interest
Deictic gestures	Conscious	Little interpretation, low contextual knowledge	arms, hands, fingers. less frequent: head, legs, eyes	Indicate physical object or property.
Identity	Unconscious	Advanced interpretation. Low contextual knowledge	Primarily face, whole body	Establishing common context for references by identifying person
Symbolic gestures	Conscious.	Advanced interpretation, high degree of contextual knowledge	arms, hands fingers, head, less frequent eyes, mouth	Waving to a person to signal «come here», Shaking head to indicate «no». Holding up three fingers to indicate the number «three».
Emotion	Conscious and unconscious.	Advanced interpretation, high degree of contextual knowledge	Face, arms, whole body	Smiling :-)
Auditory				
Location	Conscious/unconscious	little interpretation, little contextual knowledge	Voice	Calling attention to location of speaker
Identity	Unconscious	advanced interpretation, contextual knowedge	Voice, (and on an advanced level, patterns of utterances)	Establishing common context for references by identifying person
Natural language utterance	Conscious	Advanced interpretation, extremely high contextual knowledge.	Voice	Communicate complex message to someone.

Table 1 summarizes the most important touchless interactions, along with observations on both sender and receiver. It also shows some key actions that are representative of human-human interactions.

These key actions could serve as a starting point in the search for tasks with which touchless modes of interaction should be paired. Once this initial survey is done, further analysis should be done, comparing task characteristics with action characteristics, using the criteria of naturalness and intuitiveness as described in this article.

6. CONCLUSION

Applications of touchless modes of interaction must be approached from a user-perception perspective. This article offers a re-evaluation of the concepts of naturalness and intuitiveness, and suggests using these terms as guidelines for adopting human-human-interactions as models for touchless interaction.

REFERENCES

Bolt, R. A. (1980). ""Put-that-there": Voice and gesture at the graphics interface." ACM SIGGRAPH Computer Graphics 14(3): 262-270.

Buxton, B. (2009-present). Human Input to Computer Systems: Theories, Techniques and Technology.

Buxton, B. (1997). Living in Augmented Reality: Ubiquitous media and reactive environments. Video Mediated Communication. K. Finn, A. Sellen and S. Wilber. Hillsdale, N.J., Erlbaum: 363-384.

Buxton, B. (1995). Integrating the Periphery and Context: A new taxonomy of telematics. Proceedings of Graphics Interface '95.

Dreyfus, H. and S. Dreyfus (1986). Mind Over Machine: The power of human intuition and expertise in the era of the computer. New York, The Free Press.

Dourish, P. (2001). Where the Action Is. The Foundations of Embodied Interaction. Cambridge, Massachusetts, MIT Press.

Greenfield, A. (2006). Everyware: the dawning age of ubiquitous computing. Berkeley, Calif., New Riders.

Ishii, H. and B. Ullmer (1997). Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. CHI '97.

Kendon, A. (1988). Sign Languages of Aboriginal Australia: Cultural, Semiotic and Communicative Perspectives. Cambridge, Cambridge University Press.

"Merriam-Webster Dictionary: Natural." Retrieved 09.12.2013, 2013, from http://www.merriam-webster.com/dictionary/natural.

Norman, D. A. (2010). "Natural user interfaces are not natural." interactions 17(3): 6-10.

Raskin, J. (1994). "Intuitive Equals Familiar." Communications of the ACM 37(9).

Turner, P. (2008). "Towards an account of intuitiveness." Behaviour & Information Technology 27(6): 475-482.

Weiser, M. (1991). "The Computer for the 21st Century." Scientiic American 265(3): 66-75.