Natural Interaction in Ambient Intelligence Environments

ICS-FORTH AmI Programme

Margherita Antona

Introduction

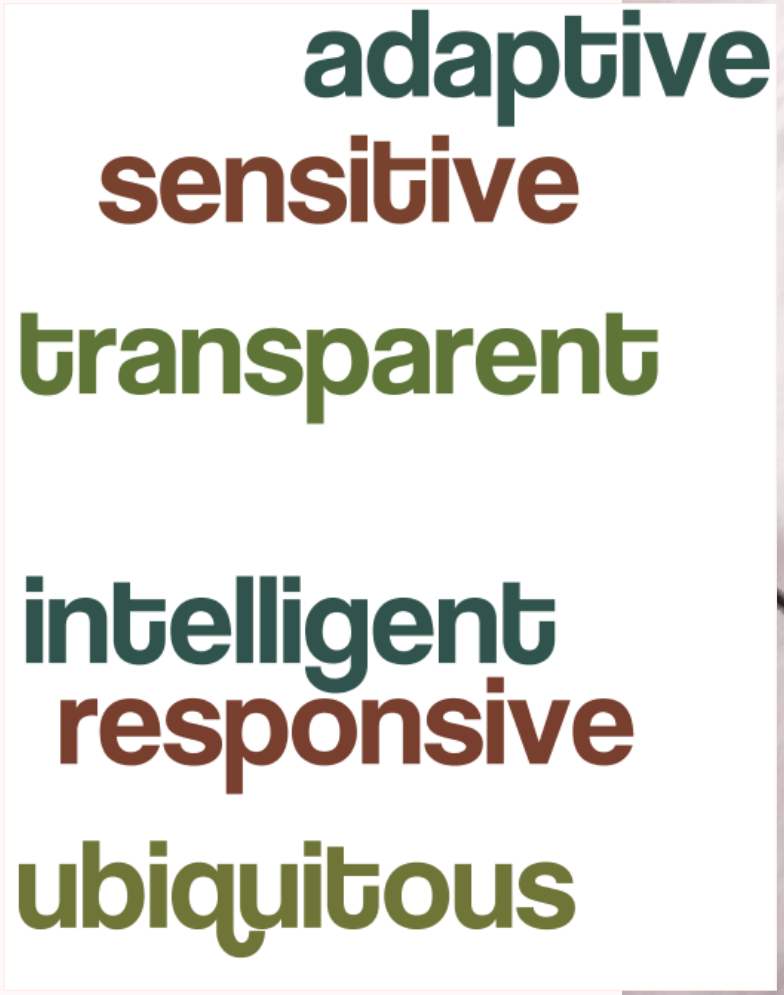
Ambient Intelligence: definitions

- *The concept of Ambient Intelligence provides a vision of the Information Society where the emphasis is on greater user-friendliness, more efficient services support, user-empowerment, and support for human interactions. People are surrounded by intelligent intuitive interfaces that are embedded in all kinds of objects and an environment that is capable of recognising and responding to the presence of different individuals in a seamless, unobtrusive and often invisible way."*

  Information Society Technologies Advisory Group of the European Commission (ISTAG, 2003)

Ambient Intelligence: definitions

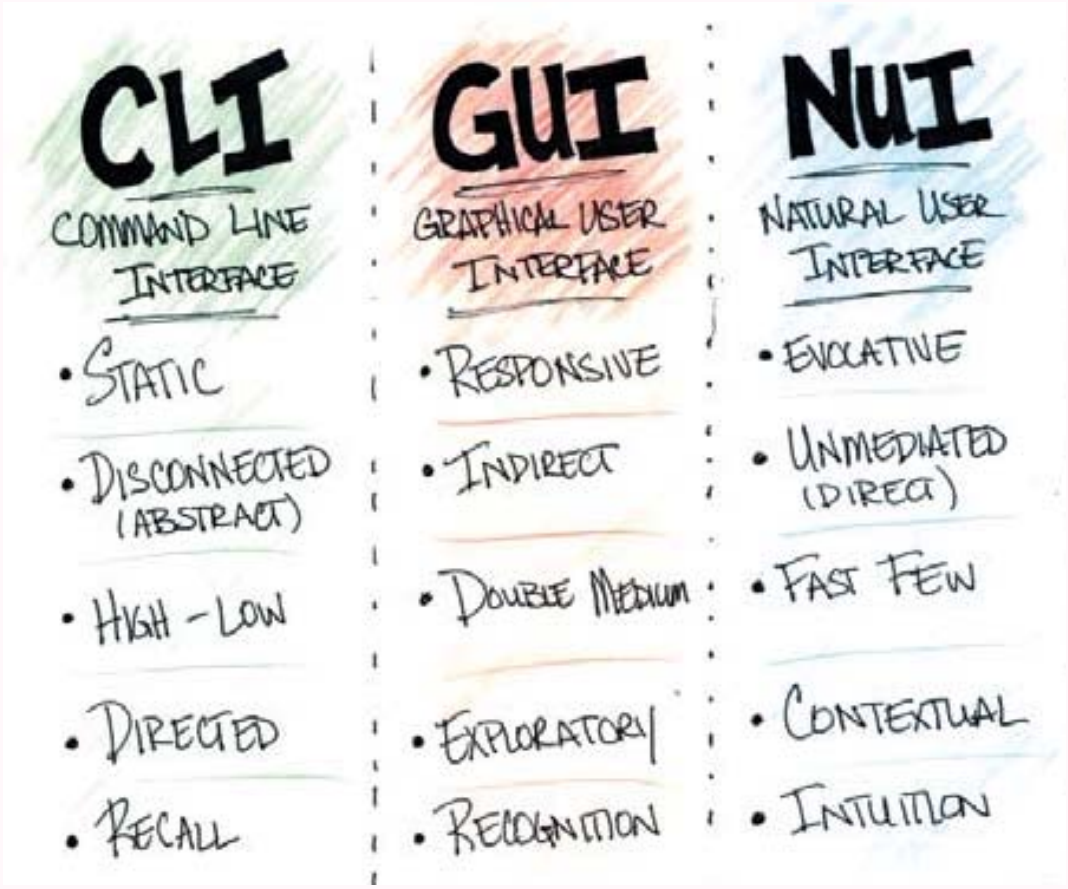
- A potential future in which we will be surrounded by intelligent objects and in which the environment will recognize the presence of persons and will respond to it in an undetectable manner
- The presence of a digital environment that is sensitive, adaptive, and responsive to the presence of people
- A new research area for distributed, non-intrusive, and intelligent software systems
- In an AmI environment people are surrounded with networks of embedded intelligent devices that can sense their state, anticipate, and perhaps adapt to their needs



AmI goal and design requirements

- While a wide variety of different technologies is involved, the goal of Ambient Intelligence is fundamentally dual:
  1. to hide the presence of its technological infrastructure from the end-users as much as possible
  2. to smoothly integrate in everyday objects, thus making it "disappear"
- The design requirements of an AmI system are:
  1. **unobtrusiveness** - devices are distributed in the environment, embedded into different physical objects, becoming invisible to humans unless visibility is needed
  2. **personalisation** - its behaviour can be configured to address individual user requirements
  3. **adaptation** – it is capable to automatically modify its behaviour relying on the recognition of user and context characteristics, including individual preferences, without requiring conscious mediation (ISO, 1999)
     - To this end, AmI systems support interactivity based on the continuous interpretation and processing of tasks, activities and contexts

Natural User Interfaces (NUI) (1/2)



The evolution of interfaces: interface types and their characteristics (Hinman, 2012)

Natural User Interfaces (NUI) (2/2)

- Building on the GUI notion that a graphic or an icon represents an information object, natural user interfaces (NUIs) depict information as objects in space. NUIs leverage human intuition; instead of what you see is what you get, NUIs rely on our innate sense of the physical world where what you *do is what you get*
- NUI systems understand and are responsive to the environments in which they are located
- NUI interactions are fast and few, and are based on the natural properties of the object and how you would expect those objects to behave
- Interactions are unmediated, allowing users to interact with information in a direct and natural way
- NUI systems are based on principles of contextualism, where there are no absolutes. Instead, events are analyzed in context and interpreted according to a frame of reference

Natural interaction in AmI environments (1/2)

- The pervasiveness of interaction in AmI environments requires the elaboration of new interaction concepts that extend beyond the current user interface concepts like the desktop metaphor and menu driven interfaces
- AmI will therefore bring about new interaction techniques, as well as novel uses and multimodal combinations of existing advanced techniques, such as, for example, gaze-based interaction, gestures, and natural language
- Interaction will be embedded in everyday objects and *smart artifacts*
- The interaction resulting from tangible user interfaces is not mediated and it supports direct engagement of the user with the environment
  - Consequently it is considered more intuitive and natural than the current keyboard and mouse-based interaction paradigm

Natural interaction in AmI environments (2/2)

- Interaction in AmI environments inherently relies on multimodal input, implying that it combines various user input modes, such as speech, pen, touch, manual gestures, gaze and head and body movements, as well as more than one output modes, primarily in the form of visual and auditory feedback
- In this context, *adaptive multimodality* is prominent to support natural input in a dynamically changing context-of-use, adaptively offering to users the most appropriate and effective input forms at the current interaction context
- Multimodal input is acknowledged for increasing interaction accuracy by reducing uncertainty of information through redundancy

Interaction Techniques

Gestures (1/4)

- The strength and efficiency of gestures as an expression medium is indisputable
  - Gestures have been studied since the first century A.D. as a means for complementing orators' speeches

Definitions

A gesture is a motion of the body that contains information. Waving goodbye is a gesture. Pressing a key on a keyboard is not a gesture because the motion of a finger on its way to hitting a key is neither observed nor significant. All that matters is which key was pressed.

Kurtenbach and Hulteen (1990)

Any physical movement that a digital system can sense and respond to without the aid of a traditional pointing device such as a mouse or stylus. A wave, a head nod, a touch, a toe tap, and even a raised eyebrow can be a gesture

Saffer (2008)

Gestures (2/4)

- The current peak in gesture-based interaction started with game consoles
  - Nevertheless, this interaction technique has been used since 1980s
  - Initially gesture-based interaction was feasible with the use of gloves
    - For example, the Z-Glove and the DataGlove (Zimmerman et al., 1987) were two cotton gloves containing flex sensors which could measure finger bending, positioning and orientation systems, and tactile feedback vibrators

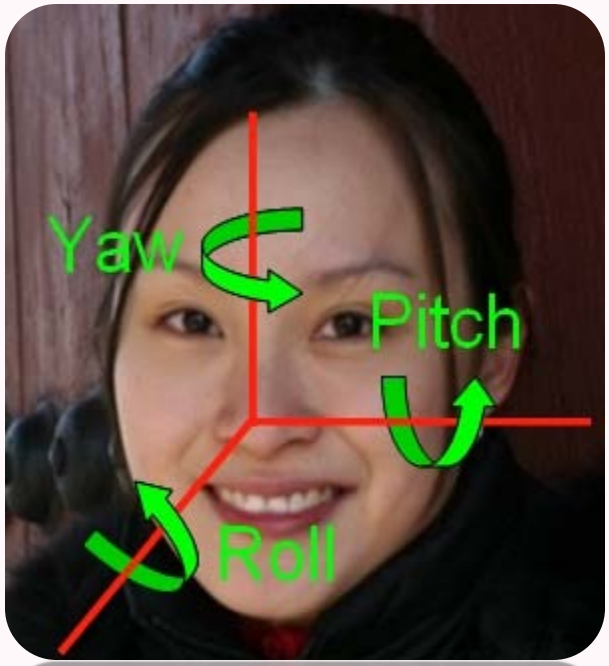
Gestures (3/4)

- The current peak of gestures started with the Wii game console which was deployed in 2006
  - An important feature of the Wii console was its remote control, Wiimote, which allowed users to interact with the console through gestures, achieving this with the use of accelerometers and a single infrared camera for tracking, while data were transmitted wirelessly using Bluetooth
- In 2009, Sony released Move , which uses a camera to track a glowing sphere on top of a wand to follow a player's horizontal and vertical movement, as well as the distance from the screen, while accelerometers and gyroscopes sense rotation
- In 2010, Microsoft released Kinect , which features a pair of depth-sensing range cameras, a system of infrared structured light sources, a multiarray microphone, and a regular RGB camera
  - Kinect provided a hands-free solution

Gestures (4/4)

- In summary, technology-wise, there have been various approaches to gesture recognition, employing for example:
  - Wearable devices (e.g. wristwatches)
  - Accelerometer based information, available in numerous consumer electronics, such as the Nintendo Wii remote or the Apple iPhone
  - Depth cameras (Microsoft Kinect sensor, Asus Xtion sensor)
  - Computer vision techniques for recognizing users' bare hands, coloured markers, etc.
    - Provide more natural, non-contact solutions

Head movements (1/2)



- Head movements and head pose estimation have been used in recent research efforts
  - as indicators of the user's attention, assisting for example automotive applications and conversations with embodied agents, and
  - as a means for controlling computers and wheelchair
- Furthermore, the role of head gestures in the context of conversations is very important and as a result several research efforts have focused in interpreting these gestures as a means for modeling conversations, for inferring users' mental states, as well as in the context of Sign Language communication

Head movements (2/2)

- In the context of a conversation, head gestures are classified as
  - Spatial head gestures (where the head moves in one of the general directions: left, right, up, or down): allow for spatial references and indication of directions
  - Semantic head gestures (like nodding and shaking the head): are used to express agreement or disagreement

Interacting with the body (1/2)

- Users' bodily information can be used as a context indicator
- Embodied interaction notion (Dourish, 2004): context and activity are mutually constitutive, and therefore users negotiate and evolve systems in the course of their interaction
- Body movements can be traced with the use of a wearable device or with vision techniques, using cameras
- Interaction through body movements without the use of any additional equipment is an attribute that is anticipated in Ambient Intelligence systems, where the environment can determine the context in which certain activities take place, thus providing meaningful information about persons and the environment, such as positioning and identification

Interacting with the body (2/2)

- KidsRoom is an interactive and immersive story environment, where children's actions (e.g., dancing with a monster, rowing a boat) are recognized and used to drive the story and control the narrative
- Macrographia supports the exploration of digital representations of large-scale museum artifacts through non-instrumented, location-based, multi-user interaction
- SoundMaker is an interactive audio environment which employs users' movements in space as a means for controlling the sequencing of percussive sounds and for changing musical parameters of those sounds collaboratively
- Body-Brush is a system which turns the human body into a dynamic brush, allowing users to create a rich variety of visual forms within a virtual 3-D canvas
- Ambient Assisted Living systems may exploit users' bodily information for home care and especially for fall detection

Touch

- Touch interfaces feel so intuitive because the sense of touch is quite possibly the most innate and intimate sense we humans possess
  - the sense of touch develops before all other senses in embryos, and it is the main sense that newborn infants use to learn about their environment
- Most everyday devices employ nowadays the touch modality (e.g. smartphones, tablets, refrigerators, kitchens, TVs, etc.)
- Touch is also a gestural interaction technique

- While touch interfaces enable more intuitive and direct manipulation of information, there are three important design considerations to take into account when creating a mobile touch UI:
  - Optimize for generous touch targets
  - Lead with the content, not the interface
    - How a design element, such as a photo, communicates a message, but also how it can convey affordances for interaction

Design Tip

Gestural UIs (1/3)

- Gestural UIs are inherently intuitive: They build on our innate sense of movement to trigger interactions
  - Gesture-based interaction refers to hand, head, and body gestures
- The most successful gestural UIs are those that leverage movements we have already learned. Gestures like sliding or flicking a page in a digital application, nudging a digital object, or rotating a digital photograph with our fingers feel intuitive because we have real-world experience with those movements
  - They're committed to our long-term muscle memory so we don't have to learn them. We simply have to associate the movement with the experience and proceed

Gestural UIs (2/3)

- An important concern regarding the use of gestures as an interaction medium is that of cultural differences
- Gestures are unconstrained and they are apt to be performed in an ambiguous manner
  - Each system can make its own assumptions regarding the expected user input
  - However, as gestural interfaces mature, some gestures have become “standard” and are rather easy for users to discover

☑ When new gestures have to be introduced:

- Introduce new gestures with the help of familiar ones
- Provide feedback with an additional sense
- Be creative and patient!

Design Tip

Gestural UIs (3/3)

☑ when designing for movement-based interaction
- consider how the movement quality might influence the user's experiences
- take into account that specific movements are more or less appropriate in certain situations and environments

Design Tip

☑ When designing, keep in mind that the main questions people have when using a product are:
- What can I do?
- Where? How?
- What happened?
- How do I get back (undo)?

Design Tip

Speech (1/3)

- Voice UIs are likely the least greedy interface type of all because they are largely invisible
  - Instead of sight, they harness one of the most natural forms of communication between humans: speech
- It's their invisibility, however, that makes voice-based interfaces challenging to design and use
- While speech is one of the most natural forms of communication between humans, most people find using speech to communicate with machines anything but natural

- Creating a successful speech interface involves using it as a solution to the right design problem
  - Speech input is generally well suited to circumstances in which the user's hands are busy
  - Speech output is generally well suited to circumstances in which the user's eyes will be busy
  - Do not use speech when privacy is a concern

Design Tip

Speech (2/3)

- Problems with speech-based applications
  - Adoption rates for speech-based applications are largely disappointing
  - User-based evaluations are often discouraging
  - Many users are frustrated during their initial interactions with speech applications due to unnatural dialogs, frequent recognition errors, and difficulty in navigation
    - Perhaps the most critical obstacle to date has been recognition errors and the cumbersome associated recovery process
  - Using devices in a noisy environment, or an environment where the background noise changes, can have a significant impact on the accuracy of the speech recognition process
  - Many users have unrealistic views regarding speech recognition, expecting their initial interactions to be as smooth as a conversation with another person

Speech (3/3)

- Limitations of working memory require voice-only menus to be fairly short
  - Both hands-on experience and cognitive psychology indicate that menus should contain only three or four items so users can remember the choices
  - Nevertheless, in a multimodal application—where choices can be displayed on a screen—the screen can supplement the user's memory

Design Tip

- Communicate clearly, concisely, and consistently
  - Phrase all prompts consistently
  - Enable the user to speak keyword utterances rather than natural language sentences
  - Use pauses to divide information into natural "chunks"

Design Tip

Eye-gazing (1/2)

- The direction of gaze of a person not only allows observation by that person of the world around them, but also reveals and communicates their focus of visual attention to the wider world
  - In some cases, eye gaze may be the only communication option available for a person
- A common way of implementing eye control is to use eye movements to control the mouse cursor
  - Binding eye movements directly to mouse movements to create an 'eye mouse' may seem an easy solution; however, there are several issues that have to be taken into account
    - Eyes move constantly, and make small corrective movements even when fixating
    - If eye movements are followed faithfully without any smoothing, the movement would appear very jerky

- Auditory feedback is very important in eye-gazing solutions
  - Appropriate feedback also increases performance and improves accuracy
  - When physically clicking a button, the user also feels and hears the button "click". Such extra confirming (auditory or tactile) feedback is missing when an "eye press" is used to click, and so must be provided

Design Tip

Eye-gazing (2/2)

- One should not expect eye control to become an appealing option in mainstreaming applications, such as applications that are widely used by the general public
- Instead of using the eyes as an explicit control medium, an application can make use of the information of the user's eye movements subtly in the background (being 'eye-aware') without disturbing the user's normal, natural viewing
- Extending this concept, some of the most promising areas for eye tracking applications are so called attentive interfaces
  - Attentive user interfaces benefit from the information of the user's area of interest (AOI), and change the way information is displayed or the way the application behaves depending on the assumed attentive state of the user
  - By monitoring the user's eye movements the application "knows" more about the user's state and intentions, and is able to react in a more natural way, thus helping the user to work on the task instead of interacting with the computer

Pen-based input

- The original vision of pen-based computers was that they would bring the benefits of physical paper and pen to computer interaction, allowing people to interact more ‘naturally’ with the computer instead of typing
  - In 1964, Ian Sutherland’s famous SketchPad system allowed user to perform gestures with a light pen, providing coordinate input for positioning picture parts on a computer drawing and demonstrative input for pointing to existing picture parts to make changes
- The pen remains essential for some tasks, like sketching and free form idea input, and in these applications it has found success

Interaction with physical objects

- In Ambient Intelligence environments, interaction is expected to be embedded in everyday objects and smart artifacts
- As a result, employing physical objects as a means for interaction is an important attribute of natural interaction in AmI environments
  - Physical objects should be used with caution, making sure to involve objects relevant to the context of use, e.g.:
    - a plate on a restaurant table
    - a book on a reading table
    - etc.

Multimodal interfaces

- Multimodal systems process two or more combined user input modes – such as speech, pen, touch, manual gestures, gaze and head and body movements – in a coordinated manner with multimedia system output
- Design guidelines:
  - Avoid unnecessarily presenting information in two different modalities in cases where the user must simultaneously attend to both sources to comprehend the material being presented
  - Maximize the advantages of each modality to reduce user's memory load in certain tasks and situations
  - Integrate modalities in a manner compatible with user preferences, context, and system functionality
  - Match output to acceptable user input style
  - Ensure system output modalities are well synchronized temporally
  - Ensure the current system interaction state is shared across modalities

Challenges

- Further advances in tracking technologies are only expected to allow for more natural interactions, posing however novel challenges under an HCI perspective, in order to ensure that already established design guidelines (such as recognition rather than recall, or user control and freedom) will not be neglected in the name of the excitement and thrill that these new interactive environments promise
- Future approaches should consider the issue of “ambient” multimodality under a Universal Access perspective, ensuring that multimodal systems are accessible by all potential users, including people with disabilities and older adults

References

- Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S. & Harvill, Y. (1987). A hand gesture interface device. In: Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface. ACM Press, pp. 189–192
- Sung, K., (2011). Recent Videogame Console Technologies. In Computer, 44 (2), pp. 91-93
- Kurtenbach, G., & Hulteen, E., (1990). Gestures in Human-Computer Communication. In B. Laurel (Ed.), The Art and Science of Interface Design, Addison-Wesley, pp. 309-317
- Saffer, D. (2008). Designing Gestural Interfaces: Touchscreens and Interactive Devices. O'Reilly Media, p.2
- Murphy-Chutorian, E., Trivedi, M., M. (2009). Head Pose Estimation in Computer Vision: A Survey. In IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (4), pp. 607-626
- Dourish. P. (2004). What we talk about when we talk about context. Personal Ubiquitous Comput. 8 (1), pp. 19-30
- Norman, D. (2010). Natural User Interfaces Are Not Natural. Interactions, 17 (3), pp. 6-10
- Hinman, R. (2012). The Mobile Frontier: A Guide for Designing Mobile Experiences. Rosenfeld Media, Brooklyn, New York, 282 pages
- Grammenos, D., Michel, D., Zabulis, X., & Argyros, A. (2011). PaperView: augmenting physical surfaces with location-aware digital information. In the Proceedings of the 5th International Conference on Tangible, Embedded, and Embodied Interaction (TEI 2011), Funchal, Portugal, 23-26 January (pp. 57-60). New York: ACM Press
- W3C, Common Sense Suggestions for Developing Multimodal User Interfaces. Available at: http://www.w3.org/TR/mmi-suggestions/
- Zabulis, X., Grammenos, D., Sarmis, T., Tzevanidis, K., & Argyros, A.A. (2010). Exploration of large-scale museum artifacts through non-instrumented, location-based, multi-user interaction. In the Proceedings of the 11th VAST International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST'2010), Palais du Louvre, Paris, France, 21-24 September (pp. 155-162).