UNIVERSITY OF CRETE
COMPUTER SCIENCE DEPARTMENT

COURSE CS-564 (OPTIONAL)
ADVANCED TOPICS
IN HUMAN – COMPUTER INTERACTION

Course Convenor: Constantine Stephanidis

Recent and emerging interaction techniques: Speech, Head and Body Tracking

Speech Interaction

Human-Human conversations (1/2)

- Humans learn to perform spoken interaction **fluently**, **effortlessly** and **efficiently** about almost any topic and for almost any purpose
- Human-human conversation serves both to:
  - organize social life in general and
  - as the basis for more specific types of interaction
    - getting others to do something
    - obtaining information from them
    - solving problems together cooperatively

Human-Human conversations (2/2)

- The ability to perform human-human-quality conversation requires a large number of skills and other characteristics, such as:
  - recognition of complex spontaneous speech
  - a very large vocabulary
  - Syntactic-semantic parsing of the complex, non-fully-sentential grammar of spoken language, including characteristics of spontaneous speech input such as hesitations, repetitions, false starts and non-words
  - Inferential capabilities ranging over knowledge of the domain, the world, the shared situation etc.
  - dialogue turn-taking according to clues, semantics, plans etc.
  - unrestricted language and speech generation
  - etc.

Speech-based human-computer interaction

- However, humans' most natural forms of communication, **speech** and **language**, are also among the most difficult modalities for machines
  - despite, and perhaps because, these are the highest-bandwidth communication channels we have
- Speech-based interactions allow users to communicate with computers or computer-related devices without the use of a keyboard, mouse, buttons, or any other physical interaction device
- By leveraging a skill that is mastered early in life, speech-based interactions have the potential to be more natural than interactions using other technologies such as the keyboard

Speech Recognition

- Speech Recognition (also known as Automatic Speech Recognition (ASR), or computer speech recognition) is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program
- **Speech recognition** technology
  - has made it possible for computers to follow human voice commands and understand human languages
  - enabled **speech-based interaction**

Application Domains

- People are exposed to many situations in which they need to interact **hands-** and **eyes- free** with a computing device
  - the user's hands may be busy with another task (e.g., driving a car, conducting medical procedures)
- Traditional keyboard and mouse may be **inaccessible**, **inappropriate,** or **unavailable** making speech a compelling input alternative
- Voice interfaces are much more adapted to disabled people and the ageing population, than tactile interfaces (e.g., remote control) which require physical and visual interaction
- Speech interaction is particularly suited to **distress situations**
  - A person, who cannot move after a fall but being conscious, may have still the possibility to call for assistance while a remote control may be unreachable

Categorization

- Based on the input and output channels being employed, speech interactions can be categorized into three groups:
  1. Spoken dialogue systems
     - Utilize speech for both input and output
     - **Examples:** Telephony systems, speech-based environment control systems with voice feedback
  2. Speech output systems
     - Utilize speech for output while leveraging other technologies, such as the keyboard and mouse, for input
     - **Example**: screen access software, which is often used by individuals with visual impairments
  3. Speech recognition systems
     - Applications that utilize speech for input and other modalities for output
     - **Examples:** Speech-based cursor control in a GUI (Graphical User Interface), speech-based dictation systems

Speech input Vocabulary

- The **Vocabulary** of a speech input application is comprised of the **words** that can be recognized
- The size of the input vocabulary influences
  - the technology needed
  - the range of possibilities that the application can support
  - and the likelihood of recognition errors
- **Small Vocabulary Solutions**
  - can handle only a small, tightly defined vocabulary of terms
- **Large Vocabulary Solutions**
  - can handle whole sentences and extensive vocabularies

Small Vocabulary Solutions (1/2)

- Small vocabularies are mainly used in:
  - Speech-based **command and control** systems that allow users to interact with computers or computer-related devices by speaking predefined commands
  - Systems that permit **customization** of the available commands
    - E.g., voice dialing on a mobile phone allows users to record the name to be associated with a phone number. However, similar sounding names can lead to recognition errors
  - **Telephony applications** that handle information inquiries, order processing, and a variety of other tasks
  - Systems that use a **simple menu**, allowing users to select the appropriate option at each stage using simple spoken commands
  - Systems that try to establish a dialogue with the user in an attempt to gather the necessary information and complete the transaction as quickly as possible

Small Vocabulary Solutions (2/2)

- When designed effectively, such systems can:
  - be **easy to learn**
  - allow **many transactions** to be completed **quickly**
- The usage of small vocabularies can significantly **reduce recognition error rates**
  - Poorly designed command sets, which result in multiple commands which sound alike, will produce more recognition errors than well-designed distinct command sets

Large Vocabulary Solutions (1/2)

- Large vocabularies are mainly used in **speech-based dictation** systems that
  - allow users to generate **large quantities of text** via speech
  - often **exceed 20,000 words**
  - use a speech recognizer to translate an **audio signal** into **text**
  - provide a powerful **alternative to the traditional keyboard and mouse**
    - permitting the generation of a variety of documents such as emails, papers, and business reports
    - reducing the risk of keyboard or mouse based repetitive strain injuries

Large Vocabulary Solutions (2/2)

- As a consequence of the larger vocabulary, dictation applications tend to be less accurate than command and control systems
  - Vendors: recognition accuracy rates >99% for a number of years
  - Users: accuracy rates are substantially lower (during composition-oriented tasks)
- Research has also confirmed that speech-based solutions for correcting recognition errors have been slow, error prone, and frustrating
  - Users spent more time correcting errors than they did dictating
- When feasible for the user, environment, and task, effective multimodal error correction solutions can provide an efficient alternative
- It is normally recommended that users create a personal speech profile to improve recognition accuracy

Artificial Languages

- Mubin et al. (2009) focus on the design and implementation of an "Artificial Robotic Interaction Language"
- Through an artificial language, the speech recognition systems could easily identify commands
  - Avoiding false-positives that could arise from human-to-human conversations
  - However this is yet to be evaluated
- The goal is to find a balance between:
  - The effort necessary from the user to learn a new language and
  - The resulting benefit of optimized automatic speech recognition for a robot or a machine

Sample Words	
bifaf	bosib
tesif	aktubi
fasin	iptabe
pomi	tasime
kekof	itoka
ojsuko	otajki
wemupu	nobelu
miwawu	awejja
amop	famifo
ajamo	lipum
lowu	ajji
alalso	tomite
obonu	

Sample Vocabulary of 25 words

Learning Effects

- Speaking is quite natural, however users must:
  - learn how to dictate
  - memorize built-in commands
  - adapt to the presence of recognition errors
  - master error recovery solutions
- At the same time, users must also develop **interaction strategies**
  - Novices tend to fixate on a single interaction strategy
  - **Experience** leads the users to change their strategies depending on the situation
  - The adoption of inefficient strategies increases the likelihood that novice users will become frustrated and abandon the technology
- Little has been learned about how interaction strategies are developed or how they evolve as users gain experience
- A study suggests that when multiple alternatives are provided, each should have clear strengths and weaknesses such that users can learn when each approach should be used

Design for Error

- When designing speech-based applications, it is critical to provide acceptable, if not ideal, solutions for correcting the consequences of recognition errors
- **Multimodal solutions for error correction**
  - designers should consider offering multimodal error correction solutions which allow users to correct errors using a variety of approaches such as speech, keyboard, mouse, stylus, or a touchscreen
- **Reduce the consequences of failed commands**
  - when those failures cannot be eliminated
- **Encourage effective strategy development**
  - designers should consider the various interaction strategies users may develop, and guide them towards more effective strategies

Speech Interaction in Smart Environments (1/5)

- What is necessary in order to make speech interaction a reality within the smart home?
- Much of the underlying technology already exists
  - Many home appliances, electrical devices, infotainment systems, sensors and so on are **sufficiently intelligent to be networked**
  - **Microphone** and **loudspeaker** technology is well established
  - Modern computers are highly **capable**, relatively **inexpensive**, and have already penetrated almost all parts of a modern home
- The bottleneck in the realization of smart home systems appears to have been the **automatic speech recognition** (ASR) and **natural language understanding aspects**

Speech Interaction in Smart Environments (2/5)

- Having in mind the advantages and disadvantages of small and large vocabulary systems, **a smart home system could be a mixture** of them:
  - Predefined **commands and menu navigation → a grammar-constrained** command-and-control vocabulary
  - Email **dictation and similar applications → large vocabulary** continuous speech recognition
- An ASR system in a smart home through its vocal interaction can be classified in two main categories:
  1. Specific control applications which form the essence of smart homes
  2. General vocal applications which any ASR system can carry out

Speech Interaction in Smart Environments (3/5)

- **Smart Home Control**
- The main feature of a smart home is the ability to vocally command the functions of the home and its appliances
- Smart home control could be achieved through a small vocabulary (0 to 50 words)
- For example, commands for controlling the lights might include 'on', 'off', the location, and perhaps a few more words, depending on what additional operations are available
- The device being controlled should be identified
- More complex commands could be managed through a set of alternatives, where the vocabulary is restricted and known, such as week days or times of the day

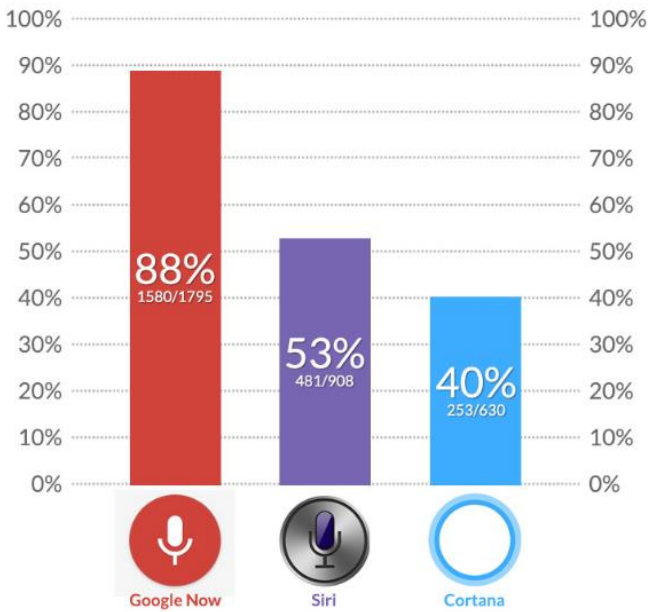
Speech Interaction in Smart Environments (4/5)

- **Dictation**
- Dictation involves the automatic translation of speech into written form
- It is differentiated from other speech recognition functions mostly because user input does not need to be interpreted (although doing so may well aid recognition accuracy)
- Usually there is little or no dialogue between user and machine
- Dictation systems imply large vocabularies
  - Domain specific systems can lead to increased accuracy

Speech Interaction in Smart Environments (5/5)

- **Open-ended Commands**
- Imagine a computer-based butler able to **interpret**, **understand**, and **carry out** complex **commands**
- **Example**: "please book me a night flight to Tokyo next Thursday and reserve a suitable suite at the Grand Hyatt Hotel"
- Challenges:
  - Significant **vocabulary size** in such commands
  - Need to **understand the meaning of the command**
  - Need to appreciate any **related factors** (e.g., words 'suitable' demands value judgments)

- This type of commands impinge on the area of natural language processing (NLP) research
  - Even today such open-ended commands are handled successfully by automated systems

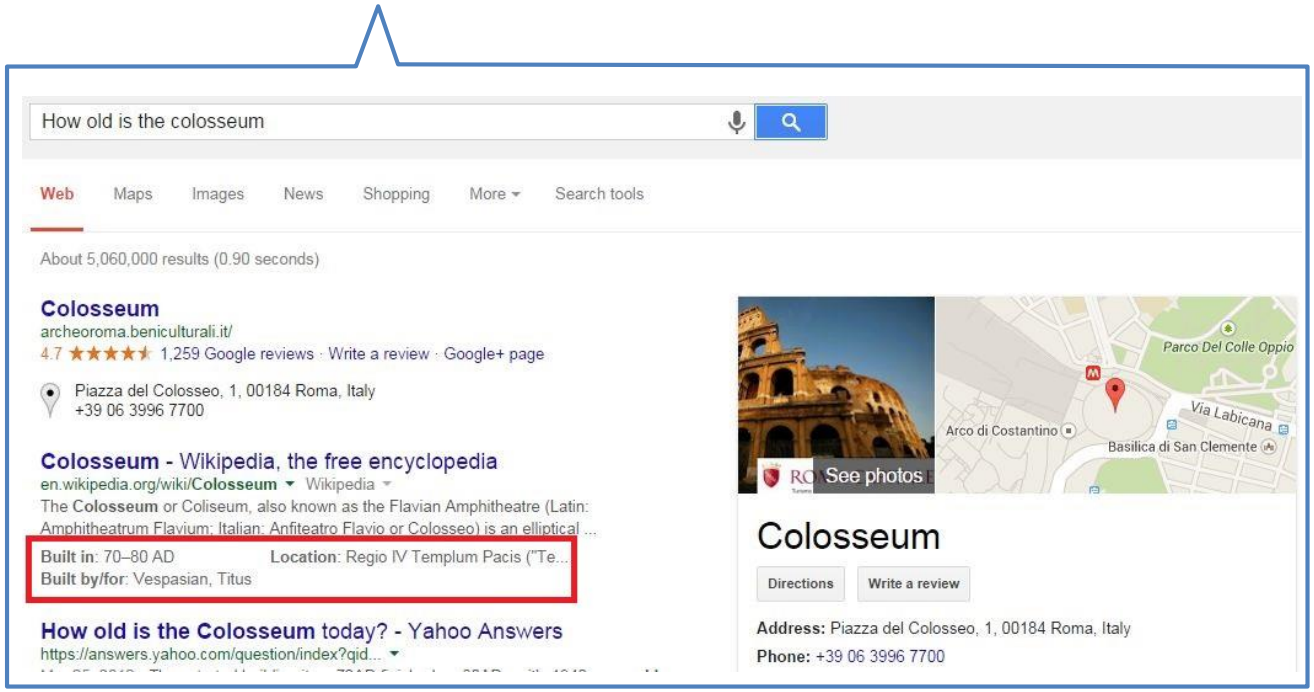


Great Knowledge Box Showdown (July '14)

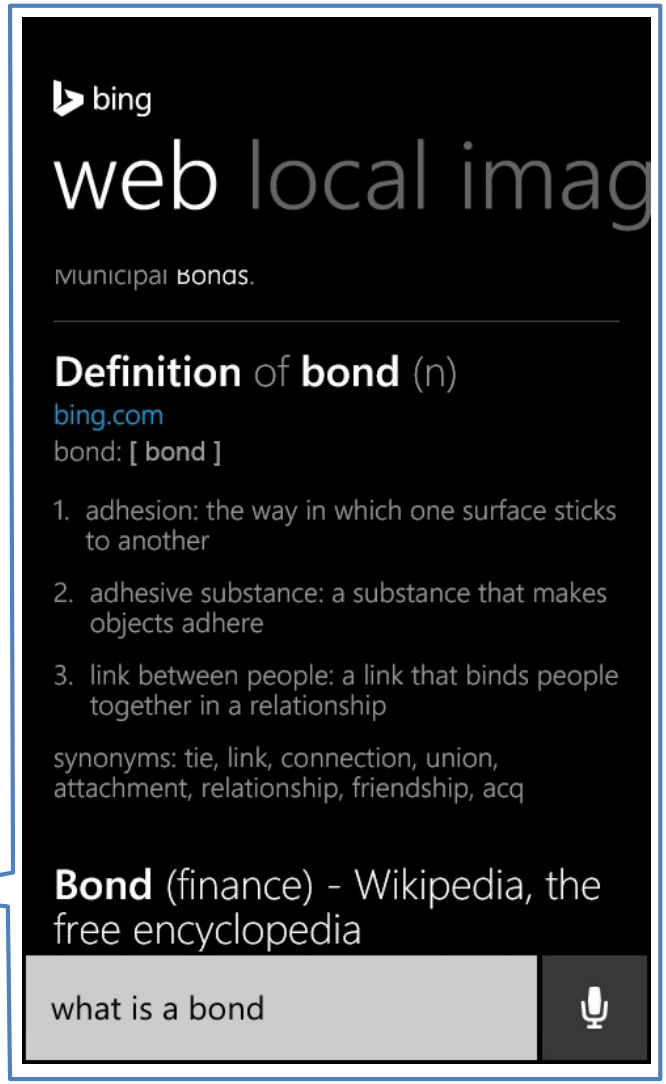
Available at: https://www.stonetemple.com/great-knowledge-box-showdown/

Advancements in speech Interaction (1/3)

*"**Google** how old is the Colosseum?"*

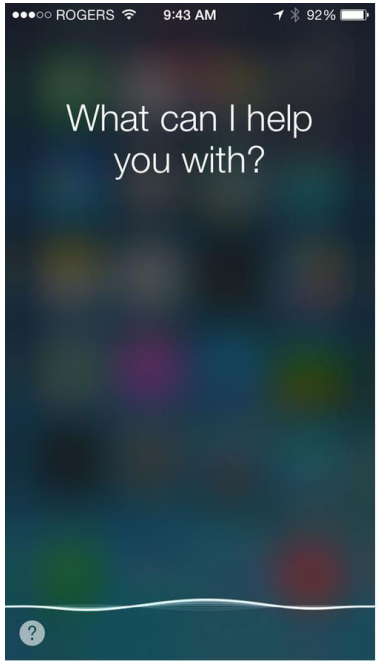


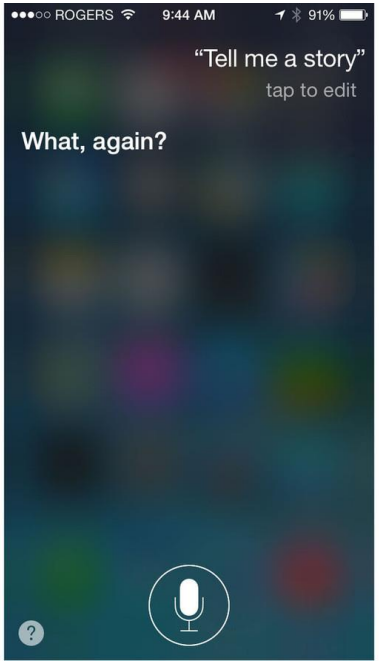
*"**Cortana** what is a bond?"*

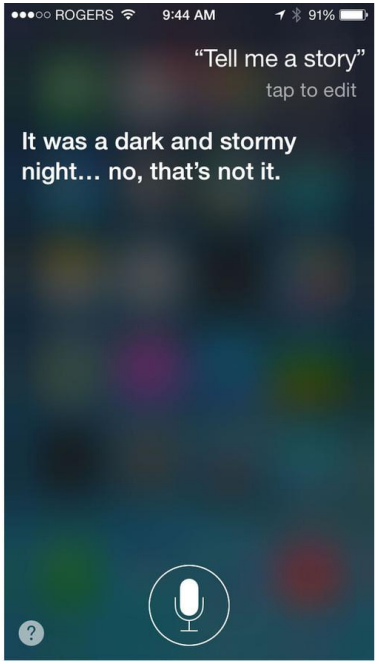


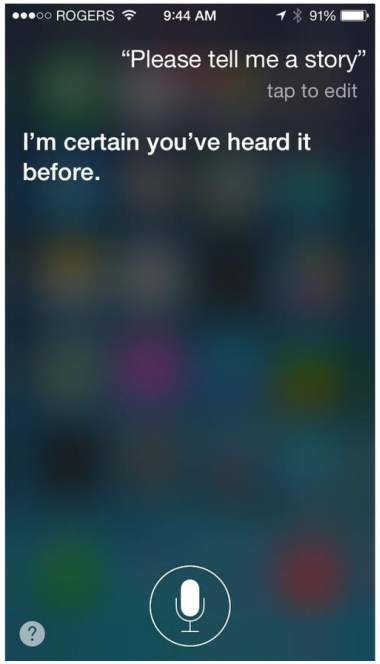
Advancements in speech Interaction (2/3)

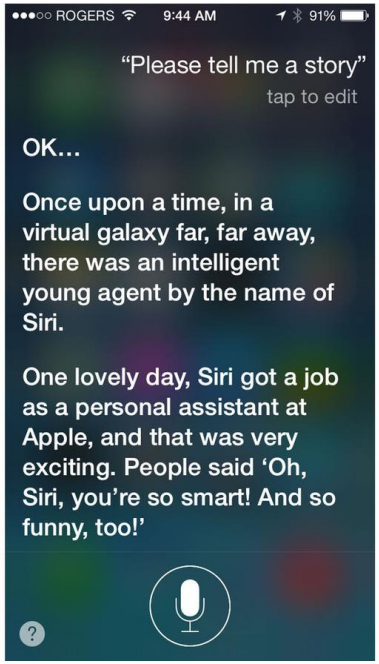
- Complex dialogues between the user and the computer





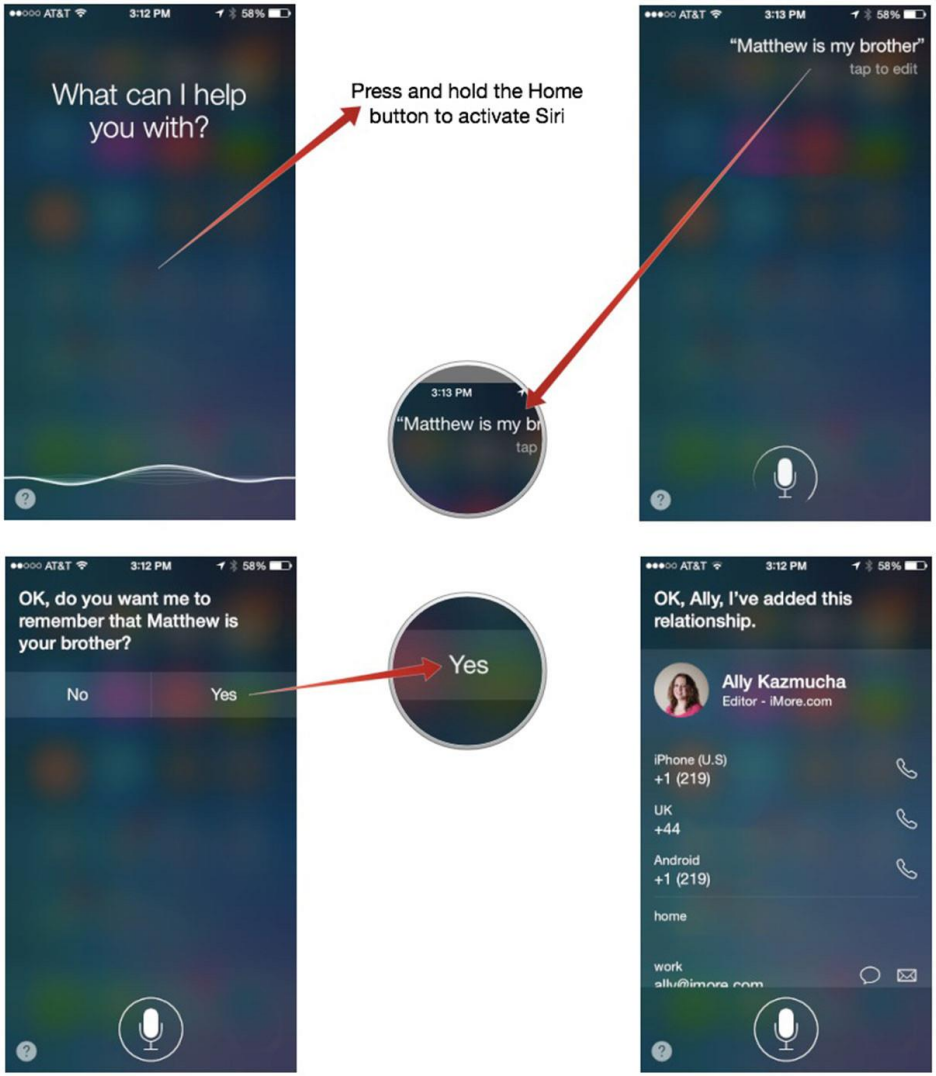






Advancements in speech Interaction (3/3)

- Intelligent systems can collect contextual knowledge over time



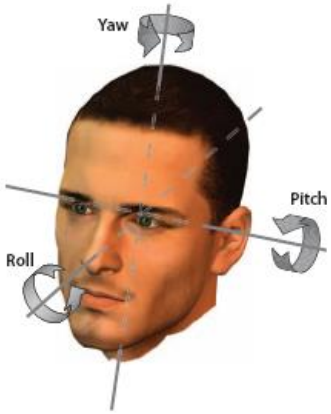
Issues

- Handling speech that is not in the recognition vocabulary
  - Even for an isolated word recognizer, users will often say “yes, please” when asked to respond “yes” or “no”
  - For large vocabularies, the issue is compounded since most users will be unaware they are using words that the recognizer does not understand
- Recognizing speech in noisy environments
  - The noise level in cars, airplanes, train stations, etc. is high, and little can be done to reduce it
- Cognitive challenges are introduced
  - users may be required to remember commands instead of recognizing the desired alternative as in menu-based systems
- Privacy and Security concerns

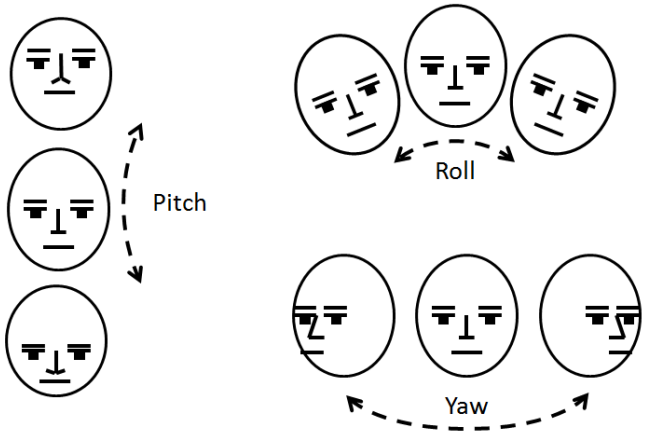
Head Pose Tracking

The role of the head in communication (1/4)

- People use the orientation of their head to convey rich, interpersonal information
  - A person will point the direction of his head to indicate who is **the intended target of a conversation**
  - In a dialogue, head direction is a nonverbal communique that cues a listener when to switch roles and **begin speaking**



- There is an important meaning in the movement of the head as a form of **gesturing** in a conversation
  - People nod to indicate that they **understand** what is being said and they use additional gestures to indicate **dissent**, **confusion**, **consideration**, and **agreement**
  - Exaggerated head movements are synonymous with **pointing a finger**, and they are a conventional way of **directing** someone **to observe** a particular location



The role of the head in communication (2/4)

- Observing a person's head could reveal interesting information
  - For instance, quick head movements may be a sign of **surprise** or **alarm**
  - These head movements commonly trigger reflexive responses from an observer, which is very difficult to ignore, even in the presence of other contradicting stimuli (e.g., auditory)
- Head pose estimation is intrinsically linked with visual **gaze estimation***
  - Head pose provides a coarse **indication of gaze**
  - When humans pay attention to an external object they orient themselves towards that object to have it in the **center of their visual field**

* the ability to characterize the direction and focus of a person's eyes

The role of the head in communication (3/4)

- Direction of gaze is highly influenced by the pose of the head
  - Wollaston illusion: Although the eyes are the same in both images (a, b), the perceived gaze direction is dictated by the orientation of the head
- Physiological investigations have demonstrated that a person's prediction of gaze comes from a **combination of** both **head pose** and **eye direction**

(a) (b)

The role of the head in communication (4/4)

- Establishing the **visual focus of attention**
  - Visual attention is influenced by external stimuli
  - Estimating the **user's direction** reveals the **visual focus of attention**
  - If two people focus their visual attention on each other, sometimes referred to as **mutual gaze**, this is often a sign that two people are engaged in a discussion
- Mutual gaze can also be used as a sign of **awareness**
  - For instance, a pedestrian will wait for a stopped automobile driver to look at him before stepping into a crosswalk
- Observing a person's head direction can also provide **information about the environment**
  - If a person shifts his head toward a specific direction, there is a high likelihood that it is in the direction of an object of interest
  - Children as young as six months exploit this property, known as gaze following, by looking toward the line-of-sight of a caregiver as a saliency filter for the environment

Head Pose Tracking

- Head pose systems seek to make use of the information conveyed through **head orientation** to
  - create an **alternative input stream** or
  - to create a **complimentary secondary input stream** in a multimodal system
- In a computer vision context, head pose estimation
  - Is the process of inferring the orientation of a human head from digital imagery
  - It requires a series of processing steps to transform a pixel-based representation of a head into a high-level concept of direction
- Tracking of head movements can be achieved:
  - either with a wearable device
  - or through vision technologies

Spatial and Semantic Head Gestures

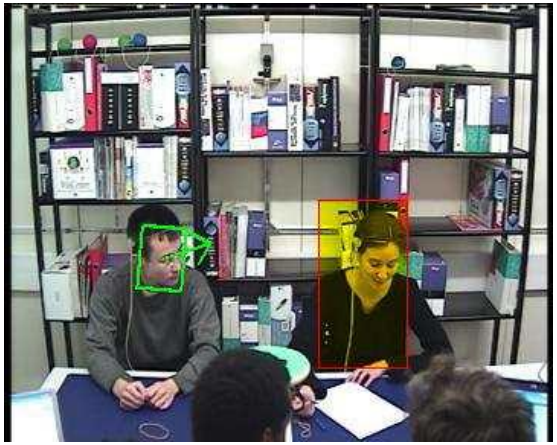
- Head gestures can be classified as spatial head gestures and semantic head gestures
- **Spatial head gestures** (where the head moves in one of the general directions: left, right, up, or down) allow for spatial references and indication of directions
  - The user's head orientation can provide information regarding a person's direction of attention, which is very useful for locating objects or persons with which a user is interacting in the context of human-robot interaction
- **Semantic head gestures** like nodding and shaking the head are used to express agreement or disagreement
  - Head nods are used as backchannels and agreements in conversation with robotic agents

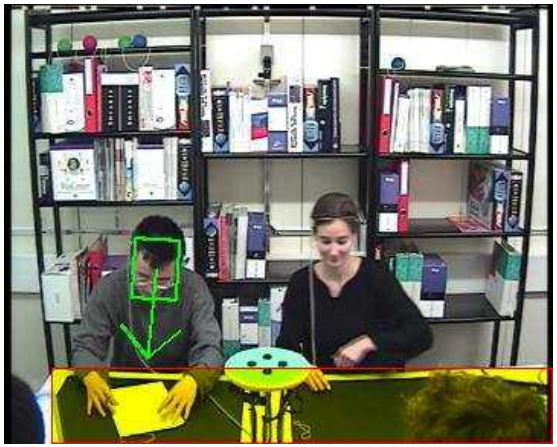
Applications of head pose estimation systems

- Head pose estimation systems play a key role in:
  - **Intelligent environments**
  - **Alternative input interfaces**
  - **Human-robot interaction**
  - **Automobile Safety**
  - **Advertising effectiveness**
  - **Sign language communication**

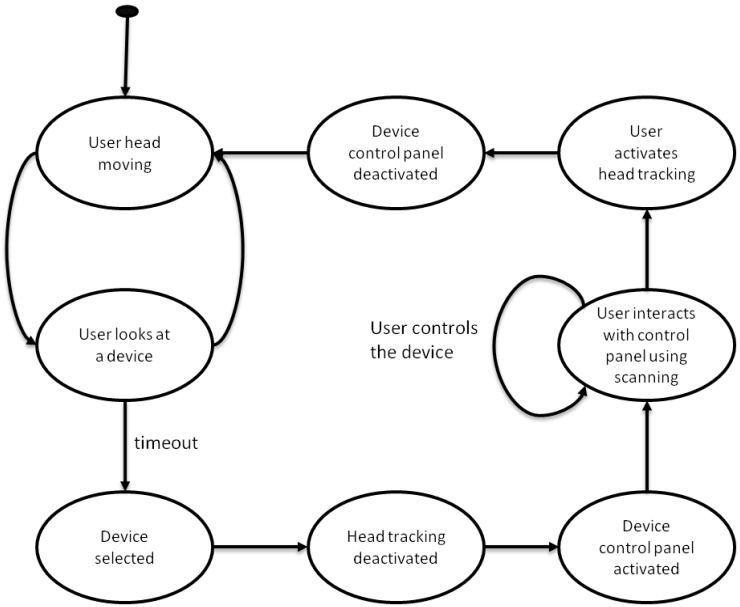
Intelligent Environments (1/2)

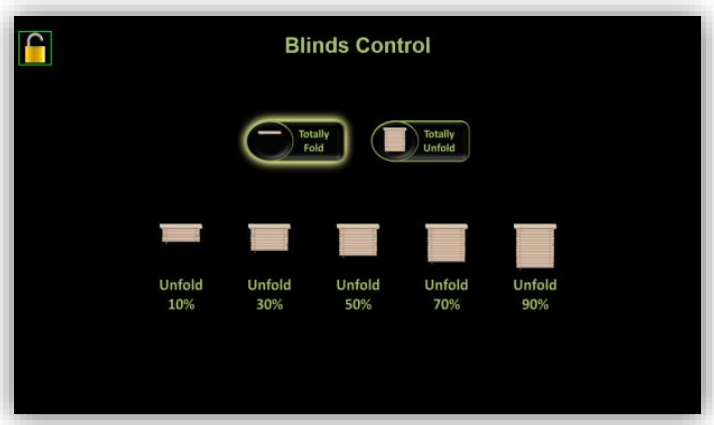
- **Smart rooms** that monitor the activities of the occupants and identify the visual focus of attention
- **Smart meeting rooms** that support humans in any kind of meeting situations (Ba & Odobez 2011, Stiefelhagen et al. 2001)
  - Track towards whom / what the participants of a meeting are looking
  - Recognize the participants' visual attention in order to introduce (if needed) context-dependent interaction models that relate to group activity and the social dynamics of communication
- **Smart lecture environments** that automatically detect and track the face of the lecturer (Voit et al. 2006)
  - The lecturer's head orientation can be a valuable cue to determine his or her focus of attention and interaction partners





Intelligent Environments (2/2)

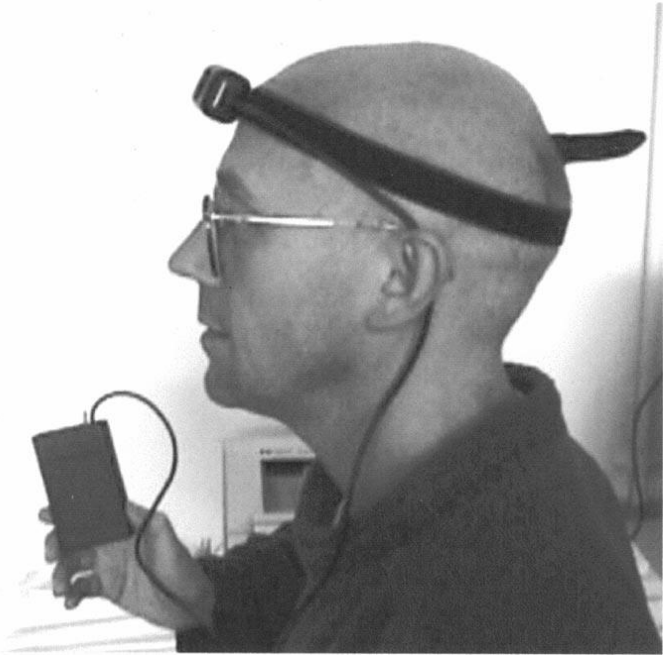




- **Domotic Control** systems that allow users to gain full control of the surrounding environment's devices and interactive components (e.g. doors, window blinds) (Ntoa et al. 2013)
  - The head scanner system utilizes a motion sensing input device (e.g. Microsoft Kinect)
  - Users can use only their head for selecting the desired device
  - Selection is accomplished by directing their head towards the specific device and "looking" at it for a specified time
  - The device is then controlled through a binary input assistive technology, such as sip and puff or binary switches

Alternative Input Interfaces (1/2)

- Some existing examples include systems that allow a user to control a computer mouse using head movements
  - Addressing mainly disabled or situationally disabled users
- Malkewitz (1998) introduced a system combining speech input and head movements to **simulate keyboard** and **mouse controls**, using a tracking device, with a sensor probe worn on the top of the user's head for tracing head movements
- Evans et al. (2000) created a head-operated joystick using infrared light emitting diodes (LED's) and photodetectors to determine head position, achieving thus a low cost **mouse emulation device**



Head-operated joystick

Alternative Input Interfaces (2/2)

- Morency et al. (2002) proposed an algorithm for stereo-based head tracking
  - They utilized a stereo camera and a head-mounted tracking device
  - Their approach included cursor control and target selection tasks in desktop and interactive room environments
  - The conclusion was that it was a robust approach even in cases of large head rotation and illumination variations
- Alternative approaches suggested using the nose for cursor control and for hands-free games and interfaces in general
  - Given that the nose, as the most protruding and the furthest from the axes of head rotation part of a person's face, has the largest degree of motion freedom

Human-Robot Interaction (1/2)

- Katzenmaier et al. (2004) revealed that head pose estimation is more reliable that acoustic and visual cues for identifying the addressee in human-human-robot interactions
- Hommel & Handmann (2011) focus on improving human-robot interaction
  - Head pose becomes classified to **attention** or **inattention**
  - This visual attention estimation will be used to
    - Get feedback whether the user is interested in the current dialog
    - For correct interpretation of the current emotional condition

Human-Robot Interaction (2/2)

- Gaschler et al. (2012) analyzed how humans use head pose in various states of an interaction
  - The analysis showed:
    - customers follow a defined script to order their drink - attention request, ordering, closing of interaction
    - customers use head pose to nonverbally request the attention of the bartender, to signal the ongoing process, and to close the interaction
  - Based on these findings, the robot bartender of the European JAMES project uses head pose estimation to infer the state of interaction of human customers in a multi-party bar scenario

Automobile Safety (1/2)

- There is great interest in driver assistance systems that act as virtual passengers, using the driver's head pose as a cue to visual focus of attention and mental state
- Head pose estimation and head movements have been used as indicators of a driver's **fatigue**, **attention**, or even intention to change lane
- Recent research efforts present systems for monitoring driver's inattention, focusing on the drowsiness or fatigue category, by evaluating several parameters:
  - percent eye closure
  - eye closure duration
  - blink frequency
  - nodding frequency
  - face position
  - fixed gaze
- The level of inattentiveness is a result of the combination of the aforementioned parameters using a fuzzy classifier
  - If the driver faces other directions than frontal for extended periods of time, this means that he is distracted
  - Frequent head tilts indicate the onset of fatigue

Automobile Safety (2/2)

- In the same context of driver assistance, but from a different perspective, **head localization** and **head pose estimation** algorithms can be used for pedestrians' detection and for estimating their awareness of a moving vehicle
  - For example, a recent approach proposed an algorithm for such a detection of pedestrians using a moving, vehicle-mounted camera, based on low resolution, lack of color images, with high variations in illumination and background [Schulz et al., 2011]

Advertising effectiveness

- Smith et al. (2008) define and address the problem of finding the visual focus of attention for a varying number of wandering people
  - A single video camera is employed to monitor the attention passers-by pay to an outdoor advertisement
  - To determine if a person is looking at the advertisement or not head pose estimation and location information is evaluated
  - Results indicated good performance on sequences where up to three mobile observers pass in front of an advertisement

Sign language communication

- Head gestures recognition is important for sign language communication
- Erdem & Sclaroff (2002) have proposed a system that uses a 3D head tracker to recover head rotation and translation parameters from monocular video
  - It successfully identifies head shakes in the context of American Sign Language (ASL)
  - Its evaluation has yielded positive results

Challenges-Limitations

- The difficulties in creating head pose estimation systems are:
  - immense variability in individual appearance
  - differences in lighting, background, and camera geometry

Body Tracking

Interacting with the Body (1/3)

- "we know more than we can tell"

  - Michael Polanyi

- Consider riding a bicycle: the bicyclist is simultaneously navigating, balancing, steering, and pedaling
  - Yet it is not possible for bicyclists to articulate all of the nuances of an activity that they successfully perform
  - Riding a bicycle is just one of thousands of activities that our bodies can do

Interacting with the Body (2/3)

- Many theories support the importance of body engagement during interaction:
  - Physical interaction with the world facilitates a child's **cognitive development** (Piaget, 1953)
  - **Kinesthetic learning** approach (Begel et al., 2004)
    - Learning through doing, humans learn about the world and its properties by interacting within it
  - **Embodied cognition** (Wilson, 2002; Anderson, 2003)
    - Cognitive processes are deeply rooted in the body's interactions with the world
  - **Motor / Muscle Memory** (Seitz, 1999)
    - People are able to sense, store and recall their own muscular effort, body position and movement to build skill

Interacting with the Body (3/3)

- Increased focus on new domains of interaction such as:
  - Experience centers
  - Sports
  - Museums
- For instance, a bodily experience at a museum can be much more present and palpable when the user can **move around freely** in the physical space of experience compared to exploring a system through an **index finger**
- Involving the body more directly in the design of interactive systems may bring a solution to:
  - Health problems
  - Work related problems
  - Life style problems
    - such as obesity and back problems due to sedentary work

Tracking Body Movements (1/2)

- Body movements can be traced with the use of:
  - a wearable device or with
  - vision techniques, using cameras
- Moen (2007) introduced BodyBug, a wearable device able to climb up and down, or stay still, according to the wearer's body movements
  - The results of their study indicate that when designing for movement-based interaction one must consider how the movement quality might influence the user's experiences and take into account that specific movements are more or less appropriate in certain situations and environments

Tracking Body Movements (2/2)

- Nintendo Wii & Sony Move both use a wearable device to track the user
  - The system does not take advantage of users' full body movement, but only of specific parts of the body where the controller is attached
  - For example an experienced player could manage playing tennis from the comfort of his couch by just moving appropriately his hand holding the console controller
- Microsoft Kinect provides a hands-free solution and allows applications to exploit full-body movement, featuring a variety of sports, fitness and dancing games

Context Indication

- Users' bodily information can be used as a context indicator
- For example, Schmidt (2000) proposes a **sensor-augmented tie** for:
  - Perceiving users' movement and
  - Identifying as to whether a user can be interrupted visually
- In the same work, Schmidt studies how context information can benefit the design of mobile applications regarding their output and input

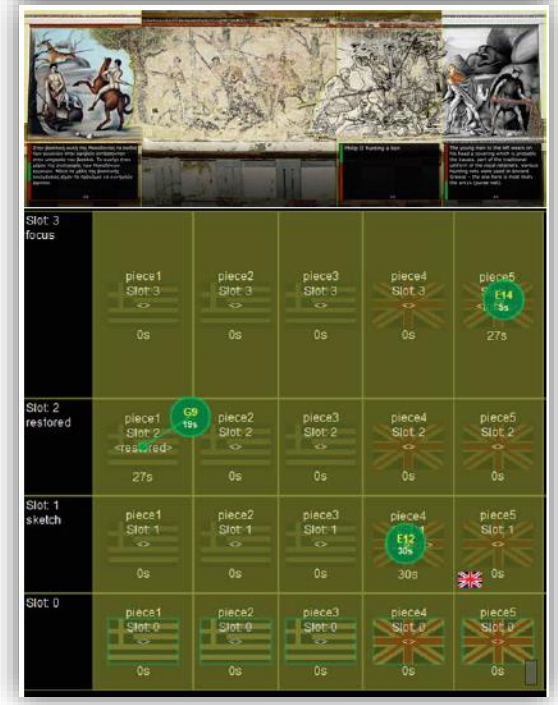
Context-aware tie

KidsRoom

- 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
- The basic experience of the KidsRoom is inspired by numerous famous children's stories (such as Peter Pan and Where the Wild Things Are) in which children's rooms are transformed into magical places
- Throughout the experience the room appropriately responds to the actions of the children by changing projected video images on the two video walls, moving sound effects around the room, providing guiding narration and hints, varying the lighting, and playing individually scored musical accompaniments

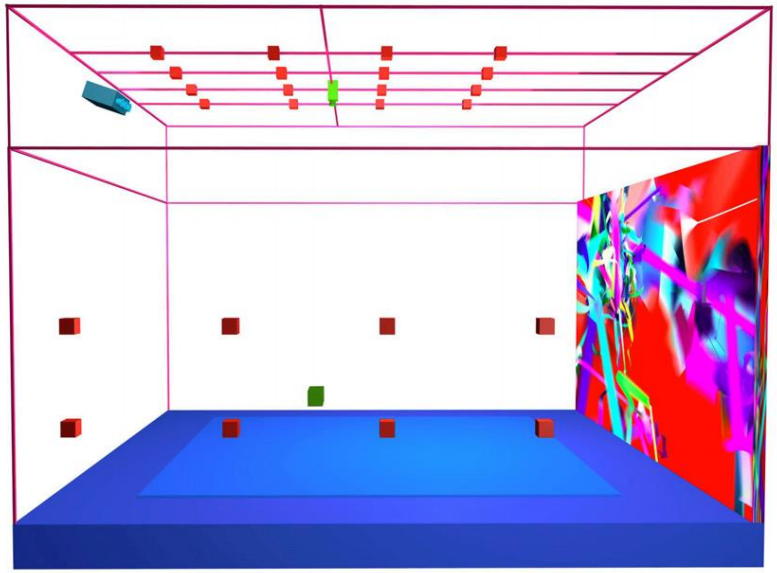
Macrographia

- A system that supports the exploration of digital representations of large-scale museum artifacts through non-instrumented, location-based, multi-user interaction
- The room is conceptually split in 5 zones perpendicular to the display, determined by an equal number of themes presented in the wall painting and corresponding to 5 vertical regions in it. Furthermore, the room is also split in 4 horizontal zones parallel to the wall painting, which are determined by their distance from it. Thus, a 5x4 grid is created, comprising 20 interaction slots

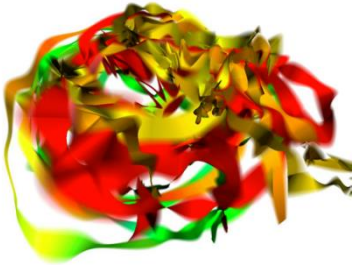


BodyBrush

- Body-Brush turns the human body as a whole into a dynamic brush and provides an intuitive tool for creating visual aesthetics.
- By controlling the mathematical centres of the body, the motion speed and acceleration, the gestural spatial dimensions, and the body trajectory, the user can create 3D drawing and painting within a virtual 3D canvas.
- As a result, the sketch in the real space by this body-as-the-brush is transformed into the shaping of the virtual visual space



The setup of Body-Brush



Pictures from Body Brush

Fall Detection

- Smart environments may exploit users' bodily information for home care and especially for fall detection
- It is important to note that the significance of inactivity changes with context. A person lying on a sofa, as she often does, is probably only resting. In contrast, a person lying on the floor where she has not previously lain may have fallen and require assistance
- When this information is combined with body pose and motion information this should provide a useful cue for fall detection

Virtual Environments

- Person tracking has been used in the context of virtual environments
- For example, the ALIVE system places the user at a particular position in a virtual room populated by virtual occupants and allows wireless full-body interaction to control artificial life creatures

References

- Murphy-Chutorian, E.; Trivedi, M.M., "Head Pose Estimation in Computer Vision: A Survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.31, no.4, pp.607,626, April 2009 *(Available through the course website)*
- Scott R. Klemmer, Björn Hartmann, and Leila Takayama. 2006. How bodies matter: five themes for interaction design. In *Proceedings of the 6th conference on Designing Interactive systems* (DIS '06). ACM, New York, NY, USA, 140-149 *(Available through the course website)*
- McLoughlin, Ian, and Hamid Reza Sharifzadeh. "Speech recognition for smart homes." *Speech Recognition, Technologies and Applications* (2008): 477-494 *(Available through the course website)*
- Mubin, Omar, Christoph Bartneck, and Loe Feijs. "What you say is not what you get: Arguing for Artificial Languages Instead of Natural Languages in Human Robot Speech Interaction." *the Spoken Dialogue and Human-Robot Interaction Workshop at IEEE RoMan*. 2009.

Additional reading (1/2)

- SO Ba, JM Odobez (2011). Multiperson visual focus of attention from head pose and meeting contextual cues. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (1), 101-116
- R Stiefelhagen, J Yang, A Waibel (2001). Estimating focus of attention based on gaze and sound. In Proceedings of the 2001 workshop on Perceptive user interfaces, 1-9
- M. Voit, K. Nickel, R. Stiefelhagen (2006). Estimating the Lecturer's Head Pose in Seminar Scenarios – A Multi-view Approach. In Machine Learning for Multimodal Interaction, Vol.3869, 230-240, Springer
- S. Ntoa, M. Antona, G. Margetis, C. Stephanidis (2013). Addressing learning disabilities in ambient intelligence educational environments. In Universal Access in Human-Computer Interaction. Applications and Services for Quality of Life, 231-240, Springer
- Rainer Malkewitz. 1998. Head pointing and speech control as a hands-free interface to desktop computing. In Proceedings of the third international ACM conference on Assistive technologies (Assets '98). ACM, New York, NY, USA, 182-188
- Evans, D.G., Drew, R., Blenkhorn, P. (2000). Controlling mouse pointer position using an infrared head-operated joystick. In Rehabilitation Engineering, IEEE Transactions, 107-117
- Morency, L., Rahimi, A., Checka, N., Darrell, T. (2002). Fast stereo-based head tracking for interactive environments. In Automatic Face and Gesture Recognition, 2002. Proceedings, 390-395
- M. Katzenmaier, R. Stiefelhagen, and T. Schultz (2004). Identifying the addressee in human-human-robot interactions based on head pose and speech. In Proceedings of the 6th international conference on Multimodal interfaces (ICMI '04). ACM, New York, NY, USA, 144-151
- Hommel, S., Handmann, U. (2011). Realtime AAM based user attention estimation. In Intelligent Systems and Informatics (SISY), 201-206, IEEE
- Gaschler, A., Jentzsch, S., Giuliani, M., Huth, K., de Ruiter, J., Knoll, A. (2012). Social behavior recognition using body posture and head pose for human-robot interaction. In Intelligent Robots and Systems (IROS), 2128-2133, IEEE
- A. Schulz, N. Damer, M. Fischer, R. Stiefelhagen (2011). Combined Head Localization and Head Pose Estimation for Video–Based Advanced Driver Assistance Systems. In Pattern Recognition, Vol. 6835 of Lecture Notes in Computer Science, 51-60, Springer

Additional reading (2/2)

- Smith, K., Ba, S.O., Odobez, J., Gatica-Perez, D. (2008). Tracking the Visual Focus of Attention for a Varying Number of Wandering People. In Pattern Analysis and Machine Intelligence, IEEE Transactions, vol.30, no.7, 1212-1229
- Erdem, U.M., Sclaroff, S. (2002). Automatic detection of relevant head gestures in American Sign Language communication. In Pattern Recognition, 2002, vol.1, 460-463
- J. Piaget (1953). The origin of intelligence in the child, Routledge & Kegan Paul, London
- Begel, A., Garcia, D.D. and Wolfman, S.A. (2004). Kinesthetic learning in the classroom. Proceedings of the Technical Symposium on Computer Science Education, Special Session
- Wilson M. (2002). Six views of embodied cognition. Psychon. Bull. Rev. 9:625–36
- Anderson, M.L. (2003). Embodied cognition: A field guide. Artificial Intelligence, 149, 91–130.
- J. Moen, (2007). From hand-held to body-worn: Embodied experiences of the design and use of a wearable movement-based interaction concept. In Proceedings of TEI'07, pp. 251–258, NY: ACM, 2007.
- Schmidt, A., (2000). Implicit Human Computer Interaction Through Context. In Personal Technologies Volume 4(2&3), 191-199
- Bobick, A. F., Intille, S. S., Davis, J. W., Baird, F., Pinhanez, C. S., Campbell, L. W., Ivanov, Y.A., Schutte, A. & Wilson, A. (1999). The KidsRoom: A perceptually-based interactive and immersive story environment. Presence: Teleoperators and Virtual Environments, 8(4), 369-393
- Zabulis, X., Grammenos, D., Sarmis, T., Tzevanidis, K., Padeleris, P., Koutlemanis, P., & Argyros, A. A. (2013). Multicamera human detection and tracking supporting natural interaction with large-scale displays. Machine vision and applications, 24(2), 319-336
- Ip, H. H., Hay, Y., & Tang, A. C. (2002). Body-Brush: a body-driven interface for visual aesthetics. In Proceedings of the tenth ACM international conference on Multimedia (pp. 664-665). ACM.
- Maes, P., Darrell, T., Blumberg, B., & Pentland, A. (1995). The ALIVE system: Full-body interaction with autonomous agents. In Computer Animation'95., Proceedings. (pp. 11-18). IEEE.

The End

Questions?