english

MOVET

Modernisation of VET through
Collaboration with the Industry

Ivan Minárik
Gregor Rozinaj
Renata Rybárová
Marek Vančo
Radoslav Vargic

Modern Ways of System Control

Erasmus+

This project has been funded with support from the European Commission.
This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Title: Modern Ways of System Control

Author: Ivan Minárik,
Gregor Rozinaj,
Renata Rybárová,
Marek Vančo,
Radoslav Vargic

Published by: Czech Technical University of Prague

Faculty of electrical engineering

Contact address: Technicka 2, Prague 6, Czech Republic

Phone Number: +420 224352084

Print: (only electronic form)

Number of pages: 34

Edition: 1st Edition, 2019

MoVET
Modernisation of VET through
Collaboration with the Industry

https://movet.fel.cvut.cz

Erasmus+

This project has been funded with support from the European Commission.
This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Explanatory Notes

	Definition
	Interesting
	Note
	Example
	Summary
	Advantage
	Disadvantage

ANNOTATION

Modern ways of system control are based on the latest technologies and using corresponding hardware. Modern system can be controlled without addition hardware as mouse or keyboard. Users can control or navigate system and application simply by using their hands (gesture navigation), voice (voice navigation) or eyes (eye tracking). Some systems may use brain-computer interface. This module will introduce listed technologies, to help understand basic principles which meet us in our daily lives.

OBJECTIVES

The main goal of the module is to introduce a student to the fundamental of modern ways of system control in different systems or application. The student is clearly acquainted with the base principles of gesture navigation, voice commands navigation, eye tracking, brain-computer interface and recommendation engine.

LITERATURE

[1] Vančo, Marek; Minárik, Ivan; Rybárová, Renata. Evolution of static gesture recognition. In: Redžúr 2014 proceedings; 8th International Workshop on Multimedia and Signal Processing; 13 May 2014, Dubrovnik, Croatia. Bratislava: Nakladateľstvo STU, 2014, p. 41-44. ISBN 978-80-227-4162-0.

[2] Rozinaj, Gregor, et al. Extending System Capabilities with Multimodal Control. Acta Polytechnica Hungarica, 2016, 13.4.

[3] https://medium.com/iotforall/how-gesture-control-will-transform-our-devices-32d4527a6d25

[4] http://www.thedrive.com/aerial/10674/djis-spark-is-a-hand-gesture-controlled-drone-that-flies-off-your-hand

[5] https://stfalcon.com/en/blog/post/intuitive-gestures-in-mobile-app-design

[6] Parrado Rollan, Marina; Posoldová, Alexandra; Rybárová, Renata. Recommendation engine design using Bayesian network for feature inference. In Redžúr 2016, 10th International workshop on multimedia and signal processing. Bratislava, Slovakia. May 24, 2016. 1. ed. Bratislava: Nakladateľstvo STU, 2016, CD-ROM, pp. 57-60. ISBN 978-80-227-4560-4

[7] Grau, C., Ginhoux, R., Riera, A., Nguyen, T. L., Chauvat, H., Berg, M., … Ruffini, G. (2014). Conscious Brain-to-Brain Communication in Humans Using Non-Invasive Technologies. PLoS ONE, 9(8), e105225. http://doi.org/10.1371/journal.pone.0105225

[8] Guy V. et al. Brain computer interface with the P300 speller: Usability for disabled people with amyotrophic lateral sclerosis, 2018, Annals of Physical and Rehabilitation Medicine, 61 (1) , pp. 5-11.

[9] Fernández, R. et al. Review of real brain-controlled wheelchairs, 2016, J. Neural Eng. 13 061001, https://doi.org/10.1088/1741-2560/13/6/061001

[10] Kosmyna, N., Tarpin-Bernard, F., Bonnefond, N., & Rivet, B. (2016). Feasibility of BCI Control in a Realistic Smart Home Environment. Frontiers in Human Neuroscience, 10, 416. http://doi.org/10.3389/fnhum.2016.00416

[11] Furness, D., The University of Florida just held the world's first mind-controlled drone race, available online: https://www.digitaltrends.com/cool-tech/mind-controlled-drone-race-university-of-florida/

[12] Chen,S., "Forget the Facebook leak': China is mining data directly from workers' brains on an industrial scale", available online: https://www.scmp.com/news/china/society/article/2143899/forget-facebook-leak-china-mining-data-directly-workers-brains

[13] Chen, A., "Brain-scanning in Chinese factories probably doesn't work — if it's happening at all", published 1.5.2018, available online: https://www.theverge.com/2018/5/1/17306604/china-brain-surveillance-workers-hats-data-eeg-neuroscience

[14] Martišius, I., Damaševičius, R., "A Prototype SSVEP Based Real Time BCI Gaming System," Computational Intelligence and Neuroscience, vol. 2016, Article ID 3861425, 15 pages, 2016. https://doi.org/10.1155/2016/3861425.

[15] Narayanan, A et al., "Toward domain-invariant speech recognition via large scale training," Google, USA, arXiv:1808.05312, 2018. https://arxiv.org/abs/1808.05312

Index

1 Introduction to system control

Nowadays information technologies are getting more and more into the foreground in our daily lives. The way of managing or controlling of devices and systems is getting more comfortable and user friendly. In following pages new and modern technologies for system control and navigation will be introduced.

Nowadays gestures are very popular way for application or systems control and many people are using it every day. Gesture controls feel natural because they are associated with the way people interact with real objects. Actually, we are using gestures in our mobile devices, computers applications, *augmented reality/virtual reality* (**AR/VR**) applications, game consoles, etc. Thanks to their everyday use, gestural interfaces are beginning to move into other areas of technology. It is expected that gestural interaction will be available in almost every device in just a few years' time. High popularity of gesture navigation forces researchers to improve these technologies. This is an obvious trend since computer performance is no longer the bottleneck of the more natural navigation and control using gestures [1].

Due to the existence of so much information available just one click away, it is not possible for an individual to keep up with all the data they can be interested in. By using technology, information can be selected for the users beforehand. That is why Recommendation Engine is a powerful tool which can make users stop wasting their time searching and scrolling information they do not really care about. Based in their preferences, the system is able to predict what we will be accordant to the user [6].

In the last few years, voice recognition has made considerable leap in speed and quality.

Very interesting technology for device control is the *Brain Computer Interface* (**BCI**). BCI represents (simply say) a direct communication pathway between human brain and external device. In this module we will introduce also eye tracking, where measuring eye activity is used to control device.

2 Technologies for system control

Technologies used for system control may differ based on the area where are used. For different system control (gesture navigation, voice navigation, eye tracking, etc.) also different hardware is used.

HW used in *gesture recognition process*:

- Gesture Control devices – wired gloves were used in the past to capture hand gestures and motions. The gloves use tactile switches, optical or resistance sensors to measure the bending of joints.
- Vision Based Gesture Recognition - uses a generic camera and/or range camera to capture and derive the hand gesture. There are multiple methods for camera-based gesture recognition.
- 3D cameras - can perceive depth. 3D cameras have become much more broadly available and cheaper in recent years [3].

Touchscreens

Generally, we can distinguish two type of touch screens: *resistive* and *capacitive*.

A resistive touchscreen consists of several layers, out of which the flexible plastic and glass layers are two important electrically resistive layers.

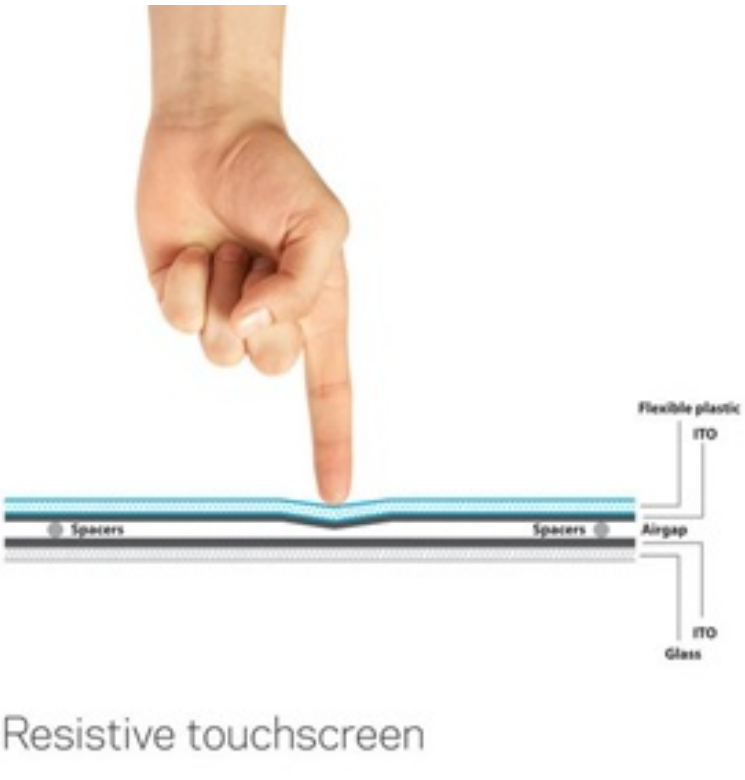
Both the layers face each other and between them is a thin gap. When a finger or stylus tip presses down on the outer surface, both the films meet. It is the measure of the resistance of both the layers on the place of contact and get an accurate measurement of the touch position.

Advantages of Resistive Touchscreen:

- High resistance to dust and water
- Best used with a finger, gloved hand or stylus
- Best suited for handwriting recognition

Disadvantages of Resistive Touchscreen:

- Not too sensitive, you have to press down harder
- Poor contrast because of having additional reflections from extra layer of material placed over the screen
- Does not support multi-touch



Resistive touchscreen

A capacitive touchscreen also consists of two spaced layers of glass, which are coated with conductor such as Indium Tin Oxide (ITO).

Human body is an electrical charge conductor. When a finger touches the glass of the capacitive surface, it changes the local electrostatic field. The system continuously monitors the movement of each tiny capacitor to find out the exact area where the finger had touched the screen.

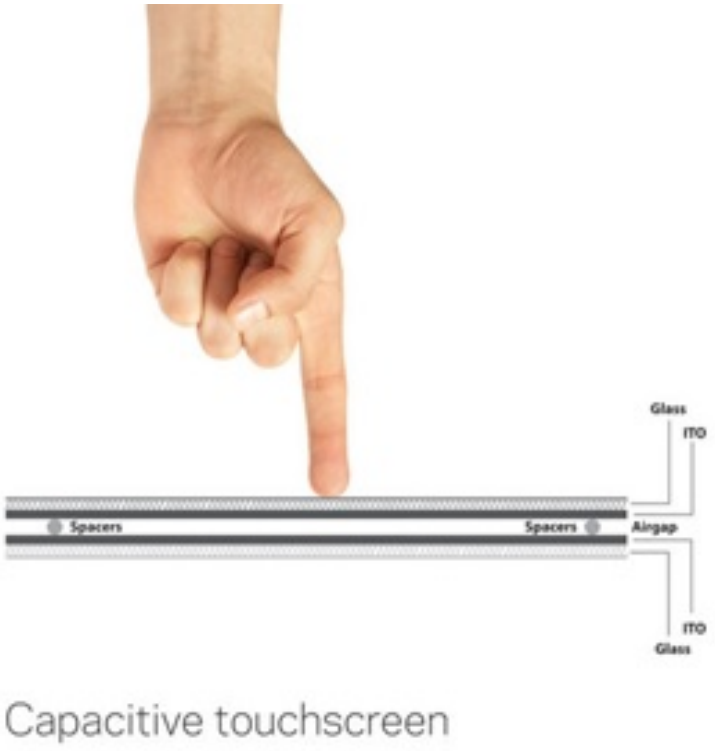
Advantages of Capacitive Touchscreen:

- Because capacitive touchscreen has glass layer instead of plastic, it looks brighter and sharper
- Highly touch sensitive and doesn't need a stylus
- Supports multi-touch

Disadvantages of Capacitive Touchscreen:

- Because the technology is dependent on the conductive nature of human body, it doesn't work if the user is wearing gloves
- Because of having a complex structure, these are quite expensive

- Glass is more prone to breaking



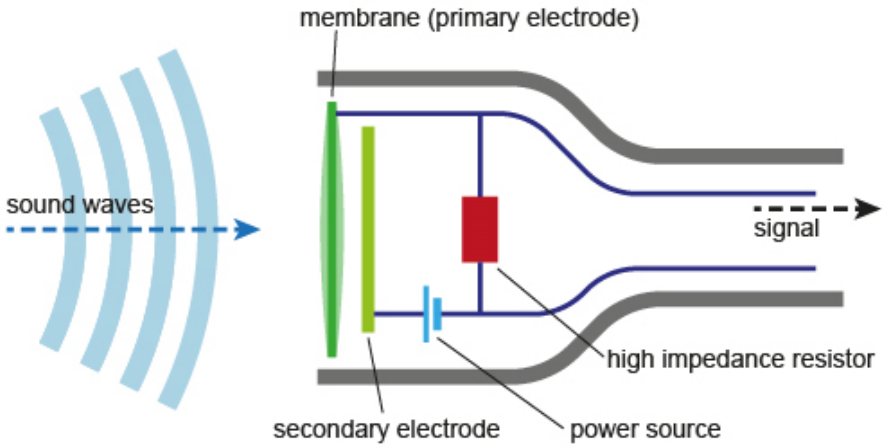
Capacitive touchscreen

Microphones

A microphone converts acoustic waves into electric signal. A diaphragm reacts to acoustic waves with vibration, which produce electric charges of corresponding intensity. There are several techniques to do so, for example condenser, dynamic, piezoelectric or even laser. Mobile phones usually use electret or **MEMS** (*MicroElectrical-Mechanical System*) microphones.

Condenser Microphone

Two plates are powered to create a condenser. One of the plates acts as a membrane and moves based on incoming acoustic waves. The movement changes output voltage, generating the signal.



Principal scheme of a condenser microphone

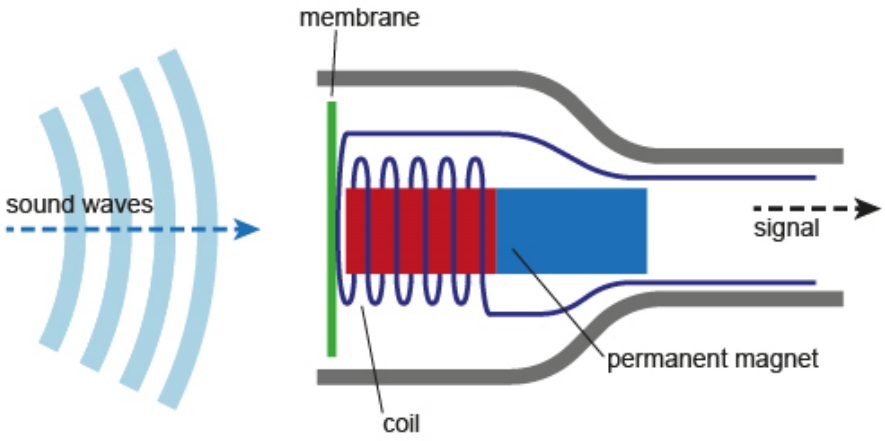
Usually a studio microphone, it is more sensitive than dynamic microphone. It is used to record musical instruments.

The microphone requires power source.

An electret microphone is a technological update of condenser microphone, making it more resilient. Electret microphone is used in majority of mobile devices nowadays.

Dynamic Microphone

Microphone membrane connects to a coil located around a permanent magnet. Pressure applied to the membrane forces the coil to move along the magnet and generate electric current.

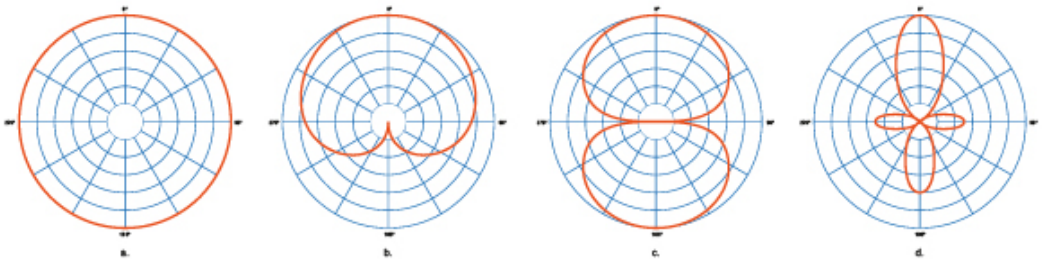


Principal scheme of a dynamic microphone

Dynamic microphone is less sensitive, making it more suitable for recording live stage singing.

Other technologies for sound acquisition include carbon, piezoelectric, ribbon, MEMS, liquid, or laser.

Based on the shape of basic components, a microphone can have different sensitivity to different angles of sound source. The most common is cardioid pattern, which makes microphone pick up sound waves in front of it but does not pick up sound from behind of it. Other sensitivity patterns include omnidirectional, bi-directional and directional.



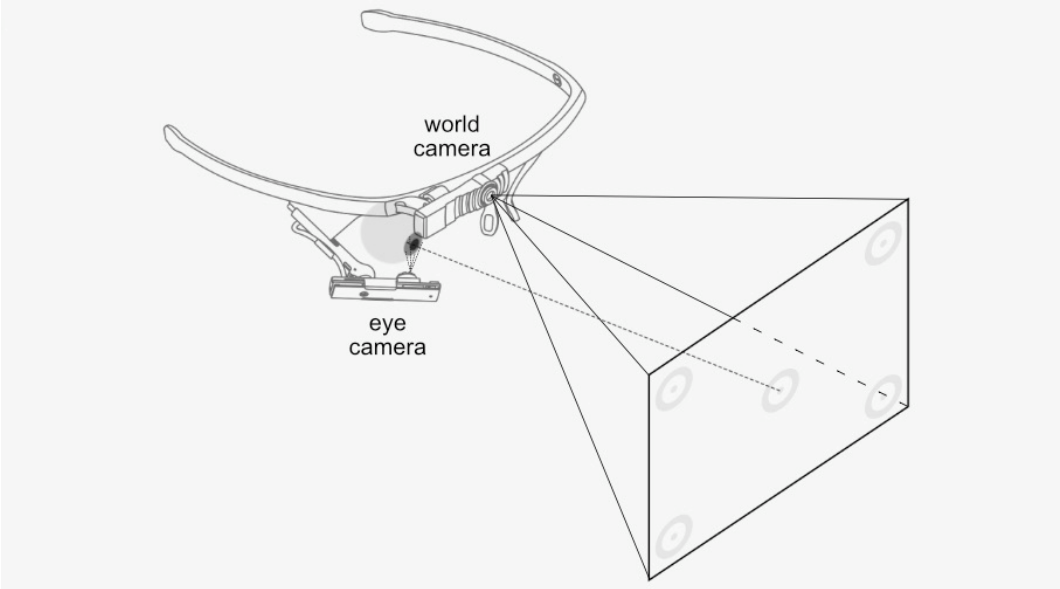
Sample Microphone Directional Characteristics. A. Omnidirectional. B. Cardioid. C. Bi-directional (Figure of 8). D. Directional (Shotgun).

A combination of several microphones (for example, ordered in a row) creates a microphone array. The array can better focus on selected area while ignoring the unnecessary area. An important feature of microphone arrays is the ability to deduce direction from which comes the speaker's voice. This helps targeting other systems of the multimedia system.

Eye tracking and BCI

Eye tracking is the process of measuring the gaze point position (where the tracked person is looking) or the motion of eyes relative to the head.

The systems for eye tracking are mostly based on cameras that capture the eye or eyes images and based on this evaluate the gaze position. The cameras capture the eye image typically with framerate starting with 30Hz (entry level, gaming) up to 1200Hz (research grade). There are basically two constructions of eye trackers – mobile variant, where the cameras that capture the eyes position are mounted on glasses or built into HMD, or fixed variant, where the cameras are placed in box below screen/monitor. The mobile binocular versions (each eye is tracked by dedicated camera) of the eyetracker are generally more precise than monocular (only one eye is tracked) and allow bigger eye movements. Besides eye cameras, in the mobile variant is used also "world" camera, that captures the environment and allows mapping the gaze position on the surrounding image.



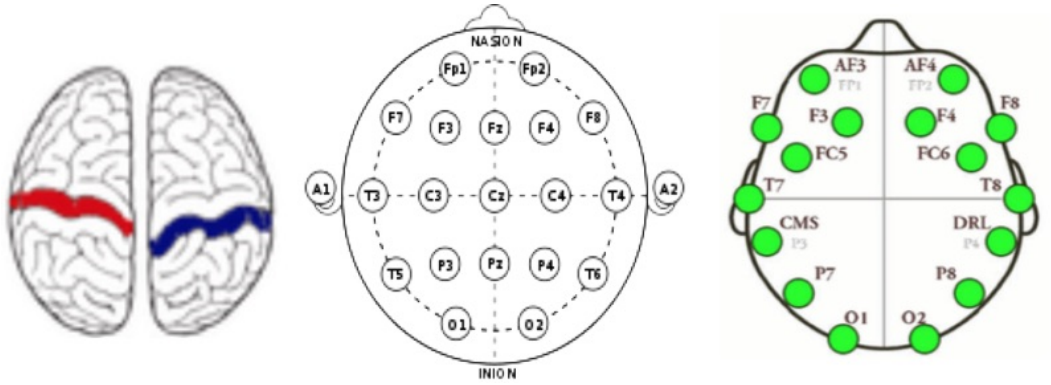
Principle of mobile monocular eye tracker, the viewport is scanned by the world camera and the pupil position by the eye camera

There are basically two constructions of eye trackers – mobile variant, where the cameras are mounted on glasses or built into HMD, or fixed variant, where the cameras are placed in box below screen/monitor. Besides above mentioned "eye" cameras, in the mobile variant is used also "world" camera, that captures the environment and allows mapping the gaze position on the surrounding image.

A Brain Computer Interface (BCI) is a technology that allows communication between a human brain and an external system (typically computer based). BCI can refer to a technology that reads signals from the brain to external system and/or a technology that sends signals to the brain.

For device control is primary interest to read the brain signals and interpret the user intention. The sending signals to brain can be used as feedback channel. BCI technology for sending the signals to the brain can use e.g. *transcranial magnetic*

stimulation (**TMS**) [7]. TMS is a non-invasive approach in which a changing magnetic field is used to cause electric current in the target brain region via electromagnetic induction. BCI technology for reading the brain signals mostly uses *electroencephalography* (**EEG**) – electrical signals created by neurons and captured on the skin over the skull using electrodes, which are typically gold plated or wet. Typically, BCI systems use 2 electrodes (entry level, gaming) – 4 and above, up to 256 (research grade). The important part of captured signals (brain waves) lie in frequency in the 2Hz-30Hz band, are very weak (2-30mV) and need to be amplified. This frequency range is split among more sub-bands (called also brain waves), as beta, theta, etc ... Presence of energy in these sub-bands can indicate various situations. Depends also on measurement location. For example Delta waves are up 4 Hz, located in frontal part and for adults it is present in more sleep stages. Alpha waves are form 7 Hz to 13 Hz located on posterior regions of head, both sides, they emerge with eye closing and with relaxation, and attenuates with eye opening or mental exertion. Mu waves are from 8Hz to 13 Hz are located on sensorimotoric cortex (central upper part on scalp on both sides) and are present during motoric actions or even imagination of motoric actions. One problem is, that also myo-signals (signals generated because of muscle movements) are captured and these are in one magnitude order stronger (10-300mV). So, the careful postprocessing is necessary. There are also systems that are based on myo-signals, e.g. captured e.g. on wrist or forearm to capture the gestures. Mostly the myo-signal based systems actually focus to be used in the neuroprosthetic solutions. One special case is oculography (capturing the signals from eye movement muscles). These systems are now widely replaced by the above-mentioned camera-based eye tracking systems.



BCI Spatial locations relevant to the signal acquisition using EEG: a) Location of SMR cortex (red – motoric part, blue - sensorics part) b) 10/20 system electrode placement c) Emotiv EPOC BCI device electrode placement

3 Different ways of system control

3.1 System control via gestures

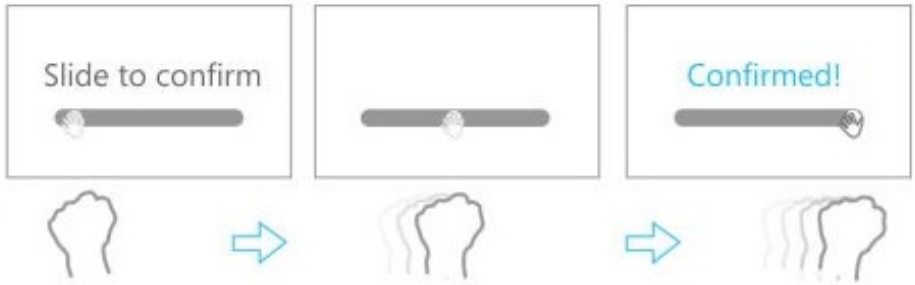
Currently, the most widely used input devices for human–computer communication are keyboard, mouse, or touch tablet. These devices are far from an idea of natural communication with a computer, and rather represent human adaptation to computer limitations. In the last few years a requirement began to pop up that humans need to communicate with machines in the same way as they do with each other: by speech, mimics or gestures, since they conceive much more information than peripheral devices approach.

Gestures are naturally transformed into our smartphones, tablets, computers, etc. Their mission is to make easy human-computer communication respectively control. Gesture can be touch or touchless but the main principles still the same.

Gestures can be divided into two basic categories by user experience.

- Innate gestures are based on general experience of all users such as to move an object to the right by moving hand to the right, catch an object with closed fingers, etc. Naturally, the innate gestures can be affected by habits or culture. With the innate gestures there is no need for a user to study them in order to get good gesture experience, they just need to be showed to him.
- The second category are learned gestures, which need to be learned. The gestures can also be divided into three categories based on the notion of motion.
    - Static gestures represent shapes created by gesturing limbs, which carry a meaningful information. The recognition of each gesture is ambiguous due to the occlusion of the limb's shape and, on the higher level of recognition, the actual meaning of the gesture based on local cultural properties [1].
    - The second category, continuous gestures serve as a base for an application interaction where no specific pose is recognized, but a movement alone is used to recognize the meaning of a gesture [1].
    - Dynamic gestures consist of a specific, pre-defined movement of the gesturing limb. Such gesture is used to either manipulate an object, or to send out a control command [1].

Static gesture



Dynamic gesture

Using gestures for navigation and system control will provide *Natural User Interface* (**NUI**), which completely removes dependency on any mechanical devices like a keyboard or mouse. The key contributor to NUI is touch-less gesture control which allows manipulating virtual objects in a way similar to physical ones. NUI let users quickly immerse in the 'new world' – applications with master control with minimum learning, what is very important for AR/VR applications and ambient intelligence systems. In burgeoning applications like autonomous drone control and in-car infotainment navigation, NUI can greatly increase the usability [3].

3.2 System control via voice commands

Voice recognition represents an uprising trend in interaction with consumer devices [2]. Voice is the most natural form of human-to-human communication and contains most of communicated information.

Voice commands are a valuable tool to control devices and systems when gestures or touch interfaces are not suitable. Their usage ranges from home entertainment systems to car infotainment control to control for the physically impaired.

Voice recognition covers several sub-fields, namely speaker identification and voice command recognition. The latter is in focus of today's researchers thanks to significant advances in neural network technology.

Generally, a voice recognition system works in these two modes:

- Learning
- Recognition

During learning, the system learns about all the possible inputs and their meaning. This usually happens in a parametric domain; whether are they parameters for individual voice commands or speaker-specific information. During recognition, an unknown input pattern is matched to a closest match from the learned parametric patterns. Both of these steps perform better with higher quality and quantity of input data.

Speech recognition is prone to incorrect recognition due presence of noise or other speakers talking simultaneously.

However, the more data a system has to process, the more time it takes. And time is crucial when we want to achieve pleasant, seamless speech recognition.

If we look back a few years, most speech recognition systems allowed recognizing only a limited set of isolated commands, or a speaker from a limited database. This would lead to highly specialized command set.

With cloud-based services becoming available widely and affordably, speech recognition systems could make use of fast server solutions. This, combined with widely available high-speed Internet connection, allows current user interfaces to process more complex voice inputs (generally, this applies to any input signal pattern). The combination allows utilizing complex decisions performed by neural networks server-side, which eliminates need for powerful user hardware and software preparation. Moreover, neural networks make recognition of isolated commands so efficient that they can now be used to recognize complex commands comprising multiple commands or command types.

The progress in utilizing neural networks over more and more powerful hardware allows improvements in several areas. Firstly, the system grows more environment-independent. The deep speech parameters are distinguishable in changing audio conditions [15]. Next, the system is able to recognize not only words or specific phrases, but to recognize whole sentence utterances, with nuances and variations of the used words. Moreover, by incorporating the previously recognized speech, systems can deduce the meaning of the current sentence or command, even if they are vague, unspecific. Systems currently start to understand actual context in which the speech was recognize and allow reacting more appropriately. This means systems start to comprehend not the actual speech, but the idea hidden behind the words.

3.3 System control via eyes tracking

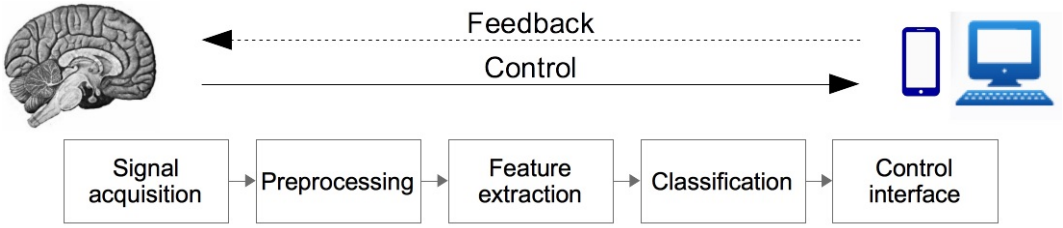
The systems for eye tracking evaluate position of gaze point based on pupil position. They use the video signals from cameras to track the pupil position. The video signals are processed by central unit, where the position of the pupil is evaluated, and user gaze position is estimated. To transform the recognised pupil position into gaze position an eye model for particular user is maintained. The parameters of the eye model are estimated during the calibration process.

3.4 System control via brain-computer interface (BCI)

Systems control using BCI is one of the one challenges for future systems control options.

BCI has more advantages as privacy (no loud sounds, no visible gestures) and low computation (only very few data are captured and processed when comparing e.g. to video).

But BCI has at the moment also major drawbacks as comfort issue (the headband, headset or headcap is needed) and mental effort (certain mental effort in most cases necessary to "generate" the control signals).



A conceptual scheme of a BCI

General scheme of the simple BCI system is depicted on Fig. 1. The first stage is a signal acquisition. EEG measures electric brain activity during synaptic excitations in the neurons. EEG signals are captured non-invasively using electrodes on the scalp. After the signal acquisition, signals are to be pre-processed. In general, the acquired brain signals are contaminated by noise and other artefacts caused by bio signals or external signals like power line, etc. After obtaining the "noise-free" signals in the signal enhancement phase, essential features from the brain signals are to be extracted. The most common feature extraction methods used with EEG signals include the discrete orthogonal transforms. Once the proper features are extracted, there must be a method that classifies the signal into desired classes. There are many categories of classification techniques as: generative (*Gaussian Mixture Model* -**GMM**), discriminative (*Neural networks*- **NN**, *Support Vector machines*- **SVM**), non-parametric i.e. sample based (*K nearest neighbour* - **KNN**), etc. Each method has its pros and cons and thus must be selected based on the application requirements. The Control interface stage of BCI uses the classification output as a control signal. The approaches can be divided into: endogenous (based on self-regulation of brain rhythms and potentials without external stimuli) and Exogenous (uses the neuron activity elicited in the brain by an external stimulus). The most frequently used methods include *slow cortical potentials* (**SCP**), sensorimotor rhythms, *visual evoked potentials* (**VEP**) *including steady-state VEPs* (**SSVEPs**), and P300. Particular devices or processes can be operated using the control interface. BCI using sensorimotor rhythms use Mu brain waves present imagination of motoric actions (e.g. hand movement). This is endogenous BCI. On

the other hand, the SSVEP is exogenous BCI. SSVEP is based on the property that the visual cortext "resonated" according to the frequency of the visual stimulus that the subject observes. Thus, using this method there are needed e.g. on screen movement control buttons, each blinking with different frequency. As the user looks at the particular button, the signal captured on visual cortex contains this frequency, so the estimation at which button the user is looking can be done. Another frequently used method is the P300. It is based on the fact, that positive peak that appears in the EEG approximately 300 ms after the presentation of a rare stimulus. Though we have an external stimulus here, P300 is considered to be an endogenous BCI, as the occurrence peak links not to the physical attributes of a stimulus, but reflect processes involved in stimulus evaluation or categorization.

3.5 System control via recommendation engine

In the introduction was mentioned, that recommendation engine can predict user's preferences and so save his time. The system can do this based on what users have previously shown to be attracted to (what users like the most), what users similar to them have liked before (users are divided into multiple segments based on their preferences) or a mix between the former options. By collecting all these data, we can have information that is relevant for their interest selected.

The recommendation system becomes more accurate as the number of items rated by users increases. Also, the more precise the engine is, the more users will be encouraged to use it. As a result, is seems that building a valid system is a decisive task.

A Recommendation System is a scheme that predicts the item rating of a user and, by doing so, can choose the best option for him or her.

There are two main approaches: *collaborative filtering* (**CF)** and *content-based* (**CB**). In the former, CF, recommendation is based on ratings from users similar to the current one. In the later one, CB, recommendations are based on item descriptions the user previously rated.

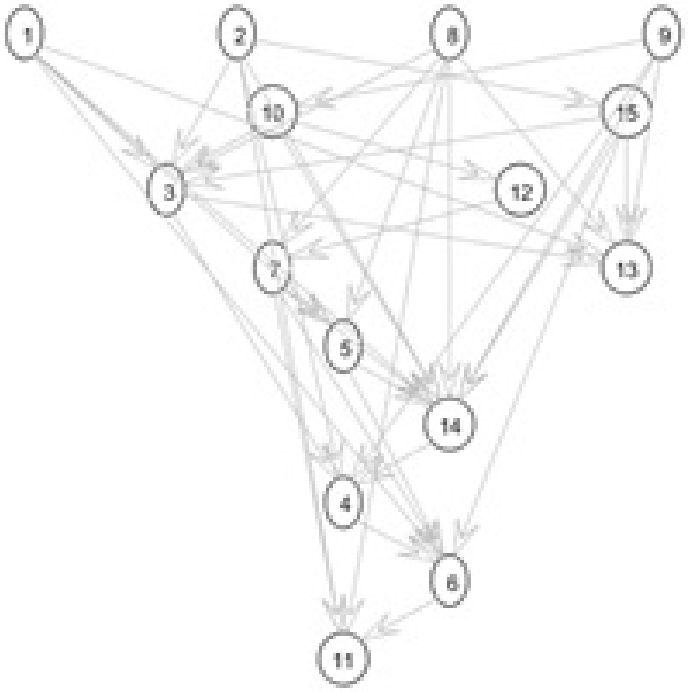
Both approaches have pros and cons.

CF suffers from cold start problem. It rises when a new item is released, no user has previously rated it so the system will not recommend it, or when there is a new user who has not rated any item, the system cannot compare to any other user. Also, finding a set of users likeminded is not always an easy task as the probability of having several users rating the same items is low. CB lacks in novelty and needs to save content description for each item.

But the recommendations uncovered by CF are serendipitous. However, for CB the cold start problem is solved by making new user fill a short survey.

Efforts have been made to combine both approaches in several ways in order to minimize the drawbacks. These techniques include switching between the two approaches, weighting the output of both schemes, using them in cascade, etc.

Bayesian networks become very useful when thinking about modelling our Recommendation system. A Bayesian Network is a directed acyclic graph where the nodes represent a set of random variables and the arcs depict direct dependencies between the variables. The strength of the relationship between them is quantified by conditional probability distribution associated with each node.



Example of Bayesian network

4 System control in applications

4.1 Voice commands for mobile devices

Mobile devices are currently the most used piece of consumer computing technology. They cover a wide range of possible usages, from making a simple call to work tasks to media creation and consumption. Even though most use cases rely on a touchscreen display as the primary command input method, there are situations in which hands-free control is necessary.

Driving

In a vehicle, the driver is obliged to pay visual attention to the situation on the road. This means he/she cannot control the phone using hands, as they are on the steering wheel or other car control. Utilization of voice commands comes in several areas:

- Requesting navigation to desired destination
- Conducting a phone call
- Controlling music playback
- Other application-specific control

The biggest players in the field include Apple, Google, Microsoft or Amazon whose speech driven assistants can access information located in the device memory, as well as on the Internet. This allows to place a call, launch a map application with pre-set destination, perform play, pause, next or previous song task and many other basic functions. Interaction with either of the assistants is mostly fluent and successful in performing the desired command, with each of them understanding the inputs in various forms of utterance.

Obtaining information

Speech assistants are able to perform search on the Internet and report the results in a logical, fast and comprehensive way via the speech synthesizer. Tasks can be uttered in a sentence-like, complex way and the neural network will deduce the intention.

Home automation

The mobile device may well serve as a home control hub. While connected to the home network, the speech driven assistants can receive commands to control various aspects of one's home, given that these aspects are connected on-line. Common tasks may include switching (specific) lights on or off, closing the curtains, up to setting a specific mood with combination of several smart devices.

Special needs

People with special needs often require interfaces adjusted to their physical condition. When there is difficult to use keyboard or touch enabled devices, such as visual impairment or damage to the limbs, voice-controlled environment is an essential aid in operating any multimedia or information system. Such system usually consists of both speech command recognizer and speech feedback.

4.2 Gestures for modern TV control

One of the greatest drawbacks of wider use of natural user interfaces is their lack of usability and human-centered design. While other modalities (i.e. voice command navigation) seem to adapt rather quickly, gesture recognition still cannot deliver truly natural experience, especially on touch-less devices. There are several factors that determine intuitiveness and naturalness of gesture recognition. Firstly, it is the hardware limitations that limit sensor algorithms to recognize more specific details in gesture performance. This causes gestures to be recognized incorrectly and force users to perform gestures that require plenty of effort and lack comfort. Secondly, gesture sets currently proposed in touch-less systems aren't inherently intuitive. System designers tend to overcome sensor limitations by introducing gestures that are easily recognizable but are often far from simple [2].

Apple TV

Apple TV uses its remote controller to catch gestures. Remote's touch surface detects a variety of intuitive, single-finger gestures. There are three types of gestures.

Swipe. Moves focus up, down, left, or right between items. Swiping lets the user scroll effortlessly through large volumes of content with movement that starts fast and then slows down, based on the strength of the swipe.

Click. Activates a control or selects an item. Clicking is the primary way of triggering actions. Clicking and holding is sometimes used to trigger context-specific actions. For example, clicking and holding an interface element may enter an edit mode.

Tap. Navigates through a collection of items one-by-one. In apps with standard interfaces based on UIKit, tapping different regions navigates directionally. For example, tapping the top of the touch surface navigates up. Some apps use tap gestures to display hidden controls.

Differentiate between click and tap, and avoid triggering actions on inadvertent taps. Clicking is a very intentional action, and is generally well-suited for pressing a button, confirming a selection, and initiating an action during gameplay. Tap gestures are fine for navigation or showing additional information, but keep in mind that the user may naturally rest a thumb on the remote, pick it up, move it around, or hand it to someone.

Apple TV controller

SingleCue

The original Singlecue launched in late 2014 and offered an early glimpse of what could be possible with gesture control. About the size of an Xbox Kinect, that device worked via infrared, and allowed you to turn on a device with a wave, quiet the volume by putting your finger to your lips, or switch between devices with a variety of other gestures.

The second-generation Singlecue builds upon that work by adding new gestures such as a wave of the hand, a pinch of a finger, and a palm click to make the device even more useful to the end user.

For example, playing and pausing video can now be controlled by opening and closing your hand, while volume can be controlled by moving a pinched finger from left to right. These gestures would work at any time, meaning the user wouldn't necessarily need to be in a specific menu to access that functionality.

SingleCue sensor

Samsung TV / LG TV

Samsung TV uses a simple gesture control to access your favourite movies, sports, apps and other Smart Content in Samsung Smart TV. Samsung uses a simple camera to monitor an environment ahead.

User can forget the remote and use your hands to control TV functions by swiping to navigate and grabbing to select. It's as smart as it is easy. Using Motion Control to change the channel, adjust the volume, move the pointer, and control other TV functions. Supported gestures are swipe, zoom, like, grab.

This set of gestures enable basic control of Smart TV. Motion may be limited by:

- dark or too bright light conditions,
- you are too close to or too far from the camera,
- your fingers cannot be detected in the case you are wearing gloves or a bandage,
- your hand is in front of your face during motion recognition,
- there is direct sunlight,
- using other fingers instead of index finger.

Samsung TV gesture control

4.3 Gestures for smart phones (and other applications)

Many applications in mobile phones have designed and implemented use of intuitive gestures that would allow users to guess which movement they should make in order to run a specific command. Gestures also let designers develop nice interfaces by leaving more space for professional ideas. New interfaces are now usually designed without clickable buttons and offer space for professional ideas. Buttons cannot disappear from the mobile application for good as they play a crucial role in driving calls-to-actions. However, in case where gestures feel more natural and intuitive, and they simplify user interaction, they should be implemented [5].

An example of mobile application, controlled with gestures, can be Google Maps or any navigation system used in mobile phone. A Google Maps app provides users with an opportunity to apply various gestures to control its certain functions. For example, to zoom in or zoom out the map on the screen, you can use your finger to move up and down, respectively.

Other application is Clear, an iOS task managing mobile application. The very amazing fact about this application is that is has no buttons, so it is completely based on gestures control. It uses taps and swipes to add and remove tasks from a to-do list.

There are definitely more applications based on gesture navigation and control, but we listed only the most wildly used.

Google maps with gesture navigation

Other applications using gestures control

Many applications are developed for professional coaching applications (e.g. golf, baseball). User do not need any additional hardware like keyboards and joysticks. Applications can provide three-dimensional body and hand motion capture capabilities in real-time. For this purpose, Kinect is used. Kinect uses depth camera for motion control. While Kinect primarily focuses on capturing body pose, Leap Motion developed a short-range gesture capture device using a stereo infrared camera. Leap Motion can track fine gestures of two hands at high frame rate. It enables applications like drawing and manipulating small objects in virtual space. Some PC vendors partnered with Leap Motion to provide the user natural user interface (NUI) in desktop applications like *Computer Aided Design* (**CAD**) [3].

Several software vendors are also providing **SDK** (*software development kit*) or middleware for application developers to easily integrate gesture and pose recognition to their applications (e.g. Gestoos, eyeSight).

To eliminate a driver distraction and to increase traffic safety the automobile manufacturers are coming up with a touch-less hand gesture interface (for example BMW's camera-based gesture control system). This interface reduces the need for drivers to reach out to the dashboard control panel. It is more natural way to control the infotainment system and helps to keep the driver's eye on the road.

Hand gestures are the viable way to also guide drone operations outdoor, so drones can fly autonomously from the remote control (e.g. summoning the drone back by waving hands). Example can be a new drone from DJI called Spark. Spark is a hand-gesture controlled drone, where all you need is your hands to command it. You can order it to distance itself, snap a selfie of you, or freely explore the skies in any way you choose—all with gestures. Spark can sense objects ahead of it from up to 5m away, to automatically avoid any unpleasant collisions [4].

Spark

4.4 Eye tracking (PC mouse control, navigation between options in screen)

Eye tracking option for mouse control has emerged in recent years as an accessibility option when controlling the PC. It is integrated already in some operation systems e.g. in Windows 10 from Microsoft.

Windows 10 supports specific eye tracking hardware. Some key capabilities can be controlled via Eye movement. These capabilities are focused on accessing applications, entering information, and communicating. The basic Windows 10 Eye Control *user interface* (**UI**) element is the launchpad, which allows users to simply look at icons to access the mouse, keyboard, and text-to-speech features as well as to move the launchpad to the top or bottom of the display. Interacting with the Eye Control UI is simple but requires some practice: just look at the screen and focus, or "dwell" as Microsoft calls it, on a button or other element until it engages. The system provides feedback throughout that lets you know what you're doing.

As replacement for the mouse and general user the precision of the systems is not enough, but in cases for user with special needs, who magnifies the screen and slower operation is not issue, this can be great choice.

When used for control in games that doesn't require the fine control of a first-person shooter, the player can achieve very immersive experience, navigating simply by looking at where the player wants to go. Camera can allow the player to look where he wants or focus on a specific character. In such situations the player could feel like to be in the game.

There are many more application areas, not only computer control. E.g. currently being developed eye tracking technology named 'wearable cockpit', for the future fighter jets will provide fighter jet pilots with a virtual display projected through the helmet. It will enable pilots to quickly access, assess and act on critical information, providing easy control on the cockpit of the aircraft. E.g. just by looking at something it can be highlighted it and then by making a gesture to 'press' it can be pressed.

There are also some programs that track using notebook camera the position of the head and use it for the mouse control (e.g. iTracker).

4.5 BCI (wheelchair navigation, navigation in games)

There are more examples where the BCI is used. This is in many cases the enthusiastic solution, mostly to control some drone, robot, car model. In other cases, this is an assistive device for people with impairment as wheelchair control or spelling machine. In the following text some representative cases are given.

- P300 speller. In P300 speller is heavily used the P300 method (positive peak that appears in the EEG approximately 300 ms after the presentation of a rare stimulus). I consist of on-screen virtual keyboard. The user looks at desired letter. The groups of keys are sequentially flashing until the letter is recognised. In some classical P300 speller solutions the groups are formed by key rows and columns, in the newer solutions these are pseudorandom groups of letters designed to minimize the consecutive flashing of the same characters, and the simultaneous flashing of neighbour characters. With these systems the around 3.5 letter/minute can be typed [8] with 95% correct letters.

- Wheelchair navigation. For the wheelchairs navigation are used mostly the P300, SSVEP and sensorimotor based BCI systems [9].

- Smart home control. There are more studies where the BCI is integrated in some Smart Home automation system. E.g. lighting, TV set, coffee machine and the shutters of the smart home were controlled in study, where about 80% task accuracy was found. The most used methods in the studies are P300 and SSVEP [10].

- Model control. There is huge amount of reports and studies where the BCI was used for model control (drone, car, robot, ...). No commercial solution is available until today, it is still active research area. Tough wide range of solutions is available, the usability is disputable. There were in last years organised also more races, e.g. car races, drone races [11] using BCI based control.

- Emotions reading. Though BCI systems can up to certain degree “read emotions”. From time to time there appear articles in newspapers about such technology availability and usability, even in large deployments [12], but then appears another ones, full of doubts [13]. BCI is usable for determination if someone is awake or asleep, but for complex emotional states like depression and anxiety it is not yet enough understood which patterns of brain activity match which emotional stages [14].

- Gaming control. There are many projects that aimed to introduce the BCI based game control interface. Among the most successful are the SSVEP based solutions [14].

In most cases there is hard to judge, if the system for the particular use case is really usable. Due to the major issues related to comfort, speed and reliability they are definitely not yet competitive alternative for healthy users. However, this can change quite soon as e.g. the signal acquisition tage could make lot of progress.

5 Conclusion or Quo vadis, System Control??

Majority of tools, people have developed in the history, consists of two main parts, functional part of the tool and in general a handle. Typical example is a knife having a blade and a grip. The more complex is the tool from functionality point of view, the more sophisticated is the interactive interface for the system control.

Computer applications represent a very specific type of tool with usually very complicated functionality. This fact results in necessity of specific interface for system navigation. Hexadecimal code may be considered as one of the first interface followed by various programming languages for handling computers. Development of modern operating systems is a remarkable effort to make working with computers as comfortable as possible. Introducing windows philosophy and mouse was a first step to make the communication simple for humans rather than computers.

Research in the field of human natural communication has led to biometric systems which have an origin of human communication in real life. Typical modalities in the direction human-computer are voice navigation, gesture navigation, speaker identification, face recognition, body motion recognition etc. In the opposite direction, modalities like system of warning/agree/disagree sounds, speech synthesis for voice feedback, graphical avatar in form of various creatures or humanoid avatars for visual interaction and in recent years virtual reality (VR) and augmented reality (AR) are used to deliver information to the user. Lot of effort has been observed to develop and use such modalities, which follow human expectation on how we understand natural communication. All types of outputs which are natural to computers are not natural to humans and vice versa.

Developing the quality human-computer interface (HCI) is very often more complex problem, than the method in the application itself. Achieving natural user-friendly HCI, where user can communicate in a natural manner is not a final stage of HCI development. The next step is a system, which observes the user and his behaviour either as a short-term or better long-term activity. The so called recommendation engine is able to predict not only future behaviour of the user but his wish and actual needs and activate relevant functionalities/commands to satisfy the user. The computer shows the functions of artificial intelligence and it becomes not only a machine but a partner.