Head-Computer Interface: A Multimodal Approach to Navigate through Real and Virtual Worlds

Francesco Carrino[1,2], Julien Tscherrig[1], Elena Mugellini[1],
Omar Abou Khaled[1], and Rolf Ingold[2]

[1] College of Engineering and Architecture of Fribourg, Switzerland
{Francesco.Carrino,Elena.Mugellini,Omar.AbouKhaled}@Hefr.ch,
Julien.Tscherrig@Edu.Hefr.ch
[2] University of Fribourg, Switzerland
Rolf.Ingold@Unifr.ch

Abstract. This paper presents a novel approach for multimodal interaction which combines user mental activity (thoughts and emotions), user facial expressions and user head movements. In order to avoid problems related to computer vision (sensitivity to lighting changes, reliance on camera position, etc.), the proposed approach doesn't make use of optical techniques. Furthermore, in order to make human communication and control smooth, and avoid other environmental artifacts, the used information is non-verbal. The head's movements (rotations) are detected by a bi-axial gyroscope; the expressions and gaze are identified by electromyography and electrooculargraphy; the emotions and the thoughts are monitored by electroencephalography. In order to validate the proposed approach we developed an application where the user can navigate through a virtual world using his head. We chose Google Street View as virtual world. The developed application was conceived for a further integration with a electric wheelchair in order to replace the virtual world with a real world. A first evaluation of the system is provided.

Keywords: Gesture recognition, Brain-Computer Interface, multimodality, navigation through real and virtual worlds, human-computer interaction, psychophysiological signals.

1 Introduction

Smart environments, augmented reality and human body used as controllers are just some examples of how the technology is getting introduced into our everyday life activities, increasing demand for natural ways of interaction and communication. On the one hand, we are observing a proliferation of interactive devices, user interfaces, and wearable technologies (such as Nintendo Wii®, Microsoft Kinect® and PlayStation Move®); on the other hand the human has to adapt, to learn, to master and to use those new devices.

In order to make human-computer interfaces truly natural, we need to develop technologies that track human movement, body behavior and facial expressions, and interpret this information in an "affective" way (this means taking into account also the

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2011, LNCS 6762, pp. 222–230, 2011.

subject's emotional state)[1]. The gesture-based interaction is a natural way to interact and can represent a substitution/complement of other forms of communications or in special context (such as concerning impaired people interaction capabilities) [2].

The usual approaches to address the gesture recognition challenges go into the direction of computer vision and image processing [3-5]. These methods are limited by some typical environmental constraints such as sensitivity to lighting changes, reliance on camera position, etc. Moreover, the images elaboration is very expensive in terms of computer processing power, making the real-time analysis a difficult challenge [6].

Additionally psycho-physiological sensors such as Electroencephalogram (EEG), Electromyogram (EMG), Blood Volume Pressure (BVP) or Galvanic Skin Response (GSR) (just to mention some of the well-known technologies) give important information about the cognitive, affective or subject's health conditions.

Furthermore, non-verbal information, such as facial expression, gaze and gesture, plays an important role in human communication [7]. The exchange of non-verbal information is important in all forms of communication and, in some specific contexts (e.g. impaired people), it is even more important than verbal information.

According to this, our paper focuses on the Human-Machine-Interaction with particular attention to the non-optical, non-verbal techniques. In this work we interlace the concepts of Gesture Recognition, Context Awareness and Multimodal Interaction in order to enhance our communication and control capabilities, and offer impaired people new ways of interaction.

In order to validate our idea, this paper presents a prototype, called Virtual Move, which allows users to navigate through Google Street View using the head only. The herein adopted concept of "head" includes movements and expressions, as well as thoughts and emotions.

This paper is organized as follows: Section 2 describes the state of the art of related researches. Section 3 describes the presented prototype focusing on the conception, the adopted interface and the evaluation. Finally, section 4 concludes the paper and discusses future work.

2 Related Projects

Multimodal approaches are often adopted in gesture recognition systems. For instance face and gestures recognition are combined in [8], where the author describes three examples of systems developed using a design-for-all strategy: a gesture sensing devices to replace the mouse; a speech to lip conversion for the hearing impaired; a brain-computer interaction based on the analysis of electroencephalographic recordings in response to gestural movements for severely disabled people.

Hands-free manipulation systems have also been studied in several works. Takahashi [7] presents a non-verbal approach based on bio-potential signals (electrooculargraphy (EOG), electromyography (EMG)) for simple gesture recognition and he proposes a prototype that aims to control a moving object in a virtual space. A similar approach is adopted in [9], where the two produced experiments aim to control at first a walkthrough application in a 3D virtual space and then an active camera. In [10] Hashimoto uses expressions and gaze to control an electric wheelchair in a non-verbal oriented approach.

A combination of mechanical (accelerometers) and physiological (EMG) sensors is used [11, 12] to identify hand gestures in order to control various devices or in substitution of videogames controllers.

Several works [1, 5, 13] propose head gestures (e.g. nodding, shaking, tilting, etc.) recognition using a computer vision approach or speech recognition techniques.

With respect to these works we can state that multimodal use of different sensing technologies can allow an effective and functional interaction for people with limited mobility/coordination abilities, or under specific conditions where it is not possible to use neither the whole body nor the voice.

However, what is currently missing is a system that jointly exploits all the interaction capabilities offered by the head in order to allow users to interact using non-optical and non-verbal techniques. Especially brain signals, such as the EEG, are often not considered.

This paper presents a system that allows the user to communicate and control the navigation in a virtual environment using head movements, facial expressions, thoughts and emotional states. The proposed approach doesn't make use of optical or verbal techniques.

3 Virtual Move Concept and Prototype

In this paper we describe our prototype, called Virtual Move, a multimodal application that aims to allow the users to navigate in Google Street View (from now GSV) using the head. EEG, EMG, EOG and a biaxial gyroscope signals are collected using the Epoch Emotiv headset [14] in order to exploit all the communication capabilities of the human head (i.e. expressions, head's movements, emotions and thoughts that, from now on, we will refer to as "head's actions"). Next sections show details about the conception and the interface that allow the user to communicate in a multimodal way with the application. Finally an evaluation about the used technology and the developed application is provided.

3.1 Concept

Virtual Move is based on the following main ideas:

- Multimodality – in order to exploit all the human head interaction capabilities (thoughts and emotions included).
- Flexibility – in order to allow a large number of users to use the application, Virtual Move has to adapt to the different possible users' needs. In the case of a subject that can not perform a specific movement or expression, the application should be able to replace the head's action with another, or add redundancy allowing the user to use at the same time different head's actions in order to transmit the same command to the navigator.
- Reusability (from the software engineering point of view) – that facilitate the forthcoming adaptation to an electric wheelchair, in order to shift from a virtual world to the real world.

In order to do that, Virtual Move is placed as bridge between the EmoEngine that analyzes the data coming from the headset, and the GSV API (see Fig. 1). Moreover, Virtual Move functions as an interface including on its Main View the GSV map's navigator and other tools to make the navigation using the head easier.

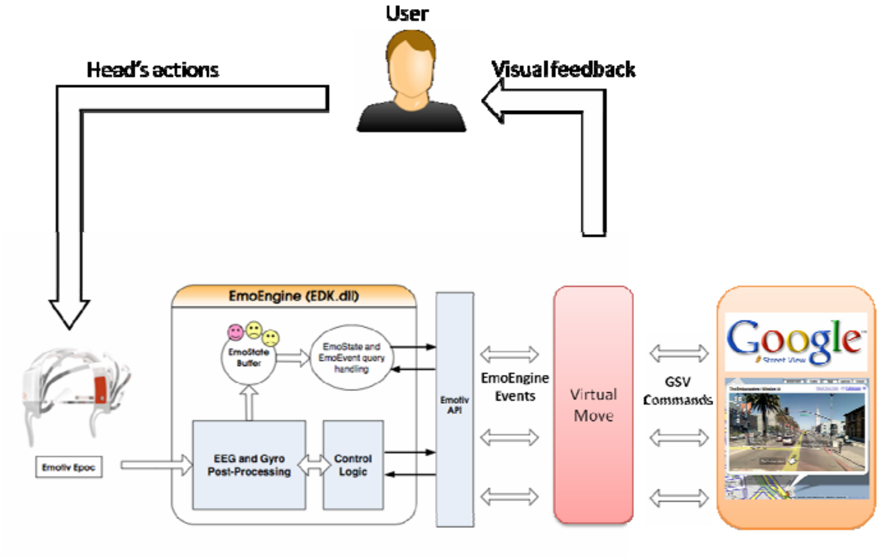


Fig. 1. The Communication Flows between the user, the Epoch headset, EmoEngine, Virtual Move and Google Street View

The EmoEngine is the main application provided with the Emotiv Epoch headset. This software does the whole signal processing and then creates "events". These events are related to all the needed information: from the detected expressions to the gyroscope information, from the detected emotion to the "thoughts". Obviously the EmoEngine can not recognize each kind of thought. The detected cognitive events are related to the so called Motor Imagery [15-17]. The Emotiv application allows the user to train the system to recognize maximum four gestures simultaneously in addition to the neutral state (no movements, no thoughts). The performance of the classifier and then the usability of this interaction modality are strongly related to the entrainment rate achieved by the user. However, according to the tests we have performed, using more than two cognitive actions is practically impossible, due to the low precision of the EEG sensors.

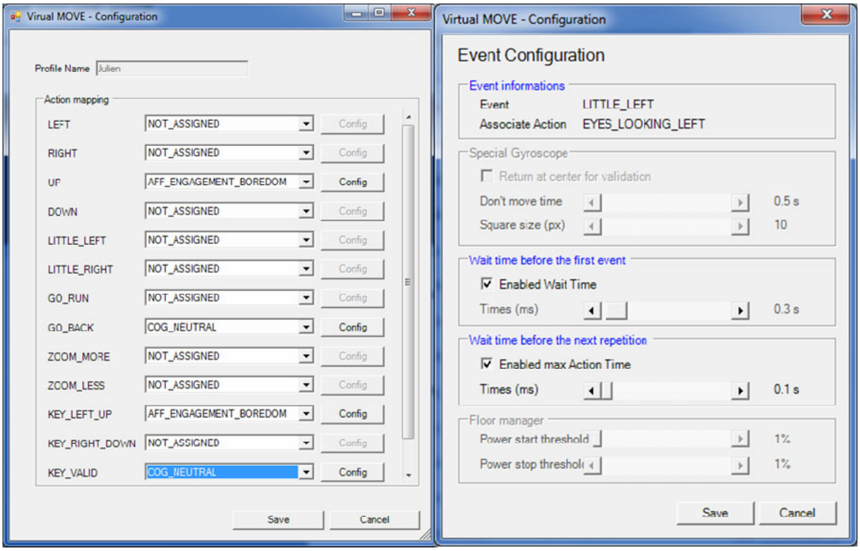
The events created by the EmoEngine are read by Virtual Move and sent to the GSV API allowing the user to navigate as he wishes.

Fig. 1 shows the communication flows between Virtual Move, EmoEngine and Google Street View. The Emotiv Epoch headset recognizes the head's actions processing data such as gyroscope coordinates and EEG, EMG, EOG signals. From the biaxial gyroscope it can detect heads movements such as nodding and shaking (tilting is not recognized); EEG gives information about emotional state and cognitive tasks; EMG detects muscular face activity giving information about the user's expressions, while the EOG detects the eyes activities. It is important to remark that all the psychophysiological signals (EEG, EMG and EOG) are detected by the same type of sensors. EMG and EOG signals are much stronger than EEG signals (i.e. measured EMG/EOG potentials range from about 100µV to 100 mV, while the EEG potentials range between 0.5-100 µV). This means that using commercial EEG devices, cognitive signals are less accurate. For this reason, applications that make use of EEG data have to be more error-tolerant. In our application it is suggested to use cognitive actions and emotions carefully or for less important activities so as not to impair navigation. For the same reason, in a pure BCI approach, EMG and EOG are considered as annoying noise.

After the signal processing the EmoEngine generates events related at the detected head's actions. These events are handled by Virtual Move, which transforms them into commands for GSV. For example, a left wink can be associated to a GO_LEFT command, and a right wink to a GO_RIGHT. Head movements and cognitive actions are handled in the same way. The only difference concerns emotions that trigger GSV commands when they exceed a given threshold. Based on the same principle, we can easily imagine replacing the GSV API with the API of an electric wheelchair. However, the associations and thresholds configurations are managed in the Virtual Move Settings Windows, independently from the type of device (GSV, electric wheelchair or any other physical or virtual device).

3.2 Interface

Virtual Move is composed by two principal modules: the Settings Windows (fig. 2) and the Navigation (or Main) Window (fig. 3).



Profile Name: Julien

Action mapping

LEFT	NOT_ASSIGNED	Config
RIGHT	NOT_ASSIGNED	Config
UP	AFF_ENGAGEMENT_BOREDOM	Config
DOWN	NOT_ASSIGNED	Config
LITTLE_LEFT	NOT_ASSIGNED	Config
LITTLE_RIGHT	NOT_ASSIGNED	Config
GO_RUN	NOT_ASSIGNED	Config
GO_BACK	COG_NEUTRAL	Config
ZOOM_MORE	NOT_ASSIGNED	Config
ZOOM_LESS	NOT_ASSIGNED	Config
KEY_LEFT_UP	AFF_ENGAGEMENT_BOREDOM	Config
KEY_RIGHT_DOWN	NOT_ASSIGNED	Config
KEY_VALID	COG_NEUTRAL	Config

Event informations
Event: LITTLE_LEFT
Associate Action: EYES_LOOKING_LEFT

Special Gyroscope
☐ Return at center for validation
Don't move time: 0.5 s
Square size (px): 10

Wait time before the first event
☑ Enabled Wait Time
Times (ms): 0.3 s

Wait time before the next repetition
☑ Enabled max Action Time
Times (ms): 0.1 s

Floor manager
Power start threshold: 1%
Power stop threshol: 1%

Fig. 2. Settings Windows

In our application each command made available from the Google Street View (Google Maps) API (i.e. “GO_FORWARD”, “GO_BACK”, “TURN_LEFT”, “TURN_RIGHT”, “UP”, “DOWN”, etc.) can be associated with the user’s head’s actions detected by the Emotiv headset. Each user can set his profile, however a default profile is provided. These associations are set in the Settings Windows.

In these panels is embedded the application flexibility and multimodality. In fact, each Google Street View command can be associated with one or more head’s actions going in the direction of a “Concurrent” multimodal approach (CASE model [18]). This was done because patients with affected motor functions may have difficulties with some kind of actions. For instance, in the case of a user unable to perform a rotation to the left, the application allows him to use another movement (e.g. nodding) as well as another modality (e.g. a facial expression). Moreover, modulating thoughts or emotions in order to continuously command a system requires constant concentration. Hence, this could compromise the reliability of the cognitive approach. Combining inputs from several modalities adds redundancy to the system (CARE model [19]) allowing user to interact easier.

Moreover, in the Settings Windows, the user can set the thresholds for commands triggered by the emotions. For example, a scroll bar allows choosing the necessary relax level to produce an action such as make appear on the screen a virtual keyboard to type (always using head’s actions) the name of the city to visit.

In Virtual Move the use of the gyroscope can become pretty advanced. In fact, the user can chose to trigger a GSV command not only by a single head movement but also by a sequence of movements (e.g. a possible sequence can be “up-down-up-down”). This approach considerably increases the interaction possibilities given from the gyroscope allowing the association of several GSV commands.

Other parameters can also be modified for a refined configuration.

The Navigation Window is where the user, principally, can see the Google Street View maps and interact with them by his head’s actions. In this window besides the maps, all the feedbacks from the selected modalities are showed to the user. On the left (see fig. 3) several panels, bars and graphics allow the control of the gyroscope position, the emotion level and, in general, the detected head’s actions.

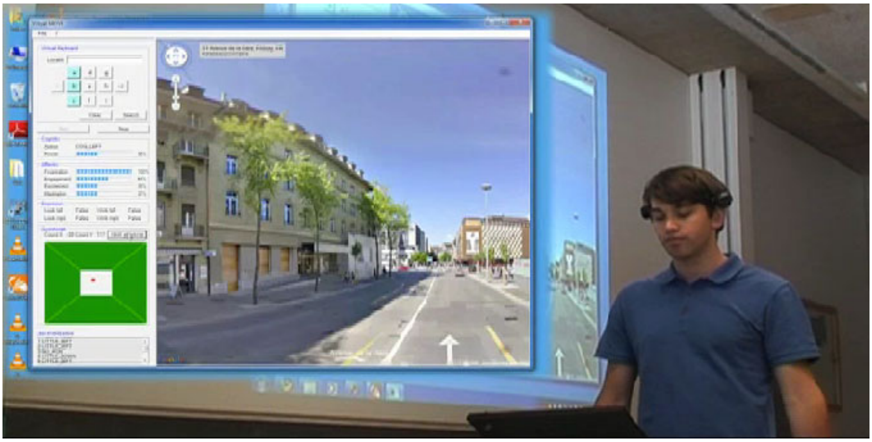


Fig. 3. User interacting with the Navigation Window

In this window it is also possible to enable a virtual keyboard, making possible to type the name of the city and/or street to visit.

Virtual Move offers to the user two interaction possibilities: gyro-A and gyro-B. The first one is strongly related to visual feedbacks. The user, watching in a rectangular zone, can see in real time the gyroscope coordinates (visualized as a red dot) and act consequently. Touching with the red dot different regions, separately or in sequence, he can trigger the desired GSV command. This approach presents three main problems: the user is forced to focus on the gyroscope panel; the user's head is never free to move; the drift present in the gyroscope signals (a typical problem of inertial systems). For this reason we developed the second interaction possibility. In gyro-B modality, instead of the gyroscope position we consider the velocity. The command is detected only when the head is moving faster than a defined threshold (configurable in the Settings Windows). This approach avoids the previous problems but we have to consider that people with coordination difficulties could not be able to perform such kind of movements. For this reason both approaches are available. However, tests on healthy people have demonstrated a preference for the second approach.

3.3 Headset and Prototype Evaluation

During the prototype development we did two evaluations. Each one was composed by a first part where the users were testing the system following a specific test protocol, followed by a questionnaire. In the first evaluation, we tested the adopted technology, the Emotiv Epoch headset, with thirteen healthy subjects. In the second evaluation we tested the Virtual Move application on a smaller sample. For the sake of brevity we will show only the result of the second, since testing the device is not the focus of our work. However, it is important to highlight some points. Testing technologies such as the Emotiv Headset poses several difficulties. Firstly, we are dealing with the brain whose behavior is quite variable from person to person [20]. Secondly, we don't know what happens within the EmoEngine "black box", because we do not have information about the signal processing actuated by EmoEngine. For all these reasons the results of these experiments were not unequivocal. What we can affirm is, on the one hand, a concrete difficulty to reproduce/detect a cognitive action without a long training for the user. On the other hand, the possibility to interact using emotions and thoughts is very appealing and all the subjects marked the system as "not frightening".

Virtual Move application was tested with five healthy subjects. The starting configuration was the same for all the users: the same head's actions were related to the same GSV commands. After five minutes of free navigation we switched the gyroscope to the second modality (gyro-B, see the previous section). After five minutes we asked the subjects to adapt the settings according to their preferences. Finally, we asked them to attain three specific locations.

Results came as the following:

- All the subjects attained the requested locations.
- The most relevant commands (GO_FORWARD", "GO_BACK", "TURN_LEFT" and "TURN_RIGHT") were assigned to the modality felt more reliable: the gyroscope. Followed by expressions, emotions and cognitive actions.

- Four subjects out of five preferred the gyro-B modality. The fifth didn't care.
- Relax was the only emotion used (in order to make appear and disappear the virtual keyboard, as in the default configuration)
- Since the required relax level to trigger the command was often achieved after closing the eyes, it was suggested to add an acoustic feedback for this type of interaction.
- Cognitive interactions were used with skeptical curiosity.
- The virtual keyboard was described as "functional but boring to use".
- The subjects were doubtful about an application in the real world.
- Mouse and keyboard, if available, remain a better solution.

In conclusion, the Virtual Move tests were satisfying, showing good system usability and the application resulted flexible and user friendly. However, as expected, our subjects expressed preference for the traditional "mouse and keyboard" interaction, waiting for a strong improvement of the brain-related technology. This was not surprising, being our test's subjects healthy. An important future step will be testing the application with severely impaired people and compare the result.

4 Conclusion and Future Work

In this paper we proposed a novel multimodal approach to navigate within a virtual world using the head. In order to avoid problems related to computer vision, the proposed approach does not make use of optical techniques. Furthermore, in order to avoid other environmental artifacts and make human communication and control smoother the used information is non-verbal.

Finally a prototype was developed and evaluated, as well as the adopted technology.

As next step we would like to use the Virtual Move application to control an electric wheelchair. Obviously, errors and inaccuracies assume greater importance in the real world. In addition, this technology (overall the BCI aspect) must be firstly accepted and understood by the end user. In this sense our tests were encouraging, with the shared subjects' opinion that BCI technologies are not frightening. However, new tests involving impaired people will be very useful for the transition to the real world.

Finally, aiming to achieve more independence from the EmoEngine "black box", the next projects will interface raw data directly.

References

1. Gunes, H., Piccardi, M., Jan, T.: Face and body gesture recognition for a vision-based multimodal analyzer. In: Proceedings of the Pan-Sydney area, vol. 36, pp. 19–28 (2004)
2. Mitra, S., Acharya, T.: Gesture Recognition: A Survey. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews) 37, 311–324 (2007)
3. Davis, J., Shah, M.: Visual gesture recognition. Vision, Image and Signal Processing. In: IEE Proceedings, IET , pp. 101–106 (2002)
4. Je, H., Kim, J., Kim, D.: Vision-Based Hand Gesture Recognition for Understanding Musical Time Pattern and Tempo. In: 33rd Annual Conference of the IECON 2007, pp. 2371–2376. IEEE Industrial Electronics Society (2007)

5. Ng, P.C., De Silva, L.C.: Head gestures recognition. In: International Conference on Image Processing, Proceedings, pp. 266–269. IEEE, Los Alamitos (2002)
6. Keir, P., Payne, J., Elgoyhen, J., Horner, M., Naef, M., Anderson, P.: Gesture-recognition with Non-referenced Tracking. The Computer Journal, 151–158 (2006)
7. Takahashi, K., Nakauke, T., Hashimto, M.: Remarks on Simple Gesture Recognition of Bio-Potential Signals and Its Application to Hands-Free Manipulation System. In: TENCON 2006 - 2006 IEEE Region 10 Conference, pp. 1–4 (2006)
8. Reilly, R.B.: Applications of face and gesture recognition for human-computer interaction. In: Proceedings of the Sixth ACM International Conference on Multimedia Face/Gesture Recognition and their Applications - MULTIMEDIA 1998, pp. 20–27 (1998)
9. Takahashi, K., Hashimoto, M.: Remarks on EOG and EMG gesture recognition in hands-free manipulation system. In: 2008 IEEE International Conference on Systems, Man and Cybernetics, pp. 798–803 (2008)
10. Hashimoto, M., Takahashi, K., Shimada, M.: Wheelchair control using an EOG- and EMG-based gesture interface. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 1212–1217 (2009)
11. Zhang, X., Chen, X., Wang, W.-h., Yang, J.-h., Lantz, V., Wang, K.-q.: Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors. In: Proceedings of the 13th International Conference on Intelligent User Interfaces - IUI 2009, vol. 401 (2008)
12. Chen, X., Zhang, X., Zhao, Z.-Y., Yang, J.-H., Lantz, V., Wang, K.-Q.: Hand Gesture Recognition Research Based on Surface EMG Sensors and 2D-accelerometers. In: 11th IEEE International Symposium on Wearable Computers, pp. 1–4 (2007)
13. Morimoto, C., Yacoob, Y., Davis, L.: Recognition of head gestures using hidden Markov models. In: Proceedings of 13th International Conference on Pattern Recognition, pp. 461–465 (1996)
14. Emotiv - Brain Computer Interface Technology, `http://www.emotiv.com/`
15. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. Proceedings of the IEEE 89, 1123–1134 (2001)
16. Pfurtscheller, G., Neuper, C.: Motor imagery activates primary sensorimotor area in humans. Neuroscience Letters 239, 65–68 (1997)
17. Scherer, R., Lee, F., Schlögl, A., Leeb, R., Bischof, H., Pfurtscheller, G.: Towards self-paced (asynchronous) Brain-Computer Communication: Navigation through virtual worlds. IEEE Transaction on Biomedical Engineering 55, 675–682 (2008)
18. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: a survey of principles, models and frameworks. Human Machine Interaction, 3–26 (2009)
19. Nigay, L., Coutaz, J.: A design space for multimodal systems: concurrent processing and data fusion. In: Proceedings of the INTERACT 1993 and CHI 1993 Conference on Human Factors in Computing Systems, pp. 172–178. ACM Press, New York (1993)
20. Guger, C., Edlinger, G.: How many people can control a brain-computer interface (BCI). In: Proc. BrainPlay, pp. 29–32 (2007)