A Multimodal Interaction Framework for Blended Learning

Nikolaos Vidakis[1], Kalafatis Konstantinos[1], and Georgios Triantafyllidis[2(✉)]

[1] Department of Informatics Engineering, Technological Educational Institute of Crete, 71500 Heralkion, Greece
nv@ie.teicrete.gr, kalafatiskwstas@gmail.com
[2] Medialogy Section, Aalborg University Copenhagen, 2450 Copenhagen, Denmark
gt@create.aau.dk

Abstract. Humans interact with each other by utilizing the five basic senses as input modalities, whereas sounds, gestures, facial expressions etc. are utilized as output modalities. Multimodal interaction is also used between humans and their surrounding environment, although enhanced with further senses such as equilibrioception and the sense of balance. Computer interfaces that are considered as a different environment that human can interact with, lack of input and output amalgamation in order to provide a close to natural interaction. Multimodal human-computer interaction has sought to provide alternative means of communication with an application, which will be more natural than the traditional "windows, icons, menus, pointer" (WIMP) style. Despite the great amount of devices in existence, most applications make use of a very limited set of modalities, most notably speech and touch. This paper describes a multimodal framework enabling deployment of a vast variety of modalities, tailored appropriately for use in blended learning environment.

Keywords: Multimodal human-computer interaction · Blended learning

1 Introduction

Multimodal interaction seeks to promote a more natural way of human-computer interaction. Despite studies proving that multimodal interfaces are not more efficient or quicker than standard WIMP interfaces [1], these interfaces were also proven to be more robust and stable [2]. Moreover, multimodal interfaces enjoyed greater acceptance from the vast majority of users. Multimodal interaction displays full support of naïve physics (multi-touch interaction), body awareness and skill (gesture and speech interaction), environment awareness and skills (plasticity), as well as social awareness and skills (collaboration and emotion based interaction).

Despite these benefits, multimodal interaction is quite demanding in terms of design and implementation. Multimodal interaction can make use of any type or number of modalities, requiring a wide range of skills in unrelated domains such as software engineering, human-computer interaction, artificial intelligence and machine learning. Moreover, a review of the literature reveals that few multimodal corpora currently exist,

A.L. Brooks and E. Brooks (Eds.): ArtsIT/DLI 2016, LNICST 196, pp. 205–211, 2017.
DOI: 10.1007/978-3-319-55834-9_24

and are targeted in specific modules and components of a multimodal interaction system. Moreover, most multimodal interaction systems, make use of a limited set of modalities, typically speech and gesture recognition.

Oviatt et al. [3] consider the following research directions: *new multimodal interface concepts,* such as blended interfaces that use both passive and active modes, *error handling techniques*, such as mutual disambiguation techniques, and dialogue processing techniques, *adaptive multimodal architectures*, that involve systems that automatically adjust to users and surroundings, and finally, *multimodal research infrastructures,* such as software tools that support the rapid creation of multimodal interfaces.

Garofolo [4] identifies another set of technological challenges, that is: *data resources and evaluation*, as the limited number of multimodal corpora in existence makes thorough evaluations unachievable, *core fusion research,* such as novel statistical methods and data representation heuristics and algorithms and, ultimately, *driver applications,* needed to guide research directions. Summarizing these findings, the following subset is derived, which is believed to be a representation of the most important issues in the field.

- *Architectures for multimodal interaction*: Because of the concurrent nature of these interaction types, tools that help in the rapid design and prototyping of multimodal interfaces are required, in order for multimodal interaction to become more mainstream.
- *Modelling of the human-machine dialog*: Because of the complex nature introduced by the large number of input and output modalities.
- *Fusion of input modalities*: A research domain, tightly coupled to human-computer dialog modeling, concerning effective fusion algorithms able to take into account multiple aspects of human-machine dialog.
- *Time synchronicity*: The ability to take into account, and adapt to multiple modal commands which can trigger different meanings, following their order, and delay between them.
- *Plasticity/adaptivity to user & context*: The capability of a human-machine interface to adapt both to the system's physical characteristics and environmental variables while preserving usability [5].
- *Error Management*: Being the weak link of multimodal interaction, since it is always assumed that users will behave in perfect accordance with the system's expectations of behavior, and that no unwanted circumstance will appear. Apparently, this is not the real life case, and error management will have to be carefully handled if multimodal interfaces are to be used broadly.
- *User feedback* is somewhat related to error management, in that the user is allowed to correct or adjust the behavior of the multimodal system in real time.

This paper presents a specific and efficient architecture and framework of multimodal interaction to be used in a blended learning environment. In the following Section, different interaction modalities will be investigated, while in Sect. 3 the main goals and principles of the blended learning approach will be presented. Section 4 presents the proposed multimodal interaction framework. Finally, Sect. 5 draws the conclusions.

2 Interaction Modalities

According to Bellik and Teil [6] modality is "a concrete form of a particular communication mode" where mode is defined as the five human senses (sight, touch, hearing, smell and taste) which constitute the receiving information, and the multifarious ways of human expression (gesture, speech, etc.) which constitute the product information. Furthermore, Bellik and Teil's definition characterizes the nature of information of human communication as visual mode, sound mode, gestural mode, etc. For example, noise, music and speech are modalities of the sound mode.

Nigay and Coutaz [8] formally present the modality as: m = (d,r)|(m,r), where "d" denotes the physical I/O device, "r" an interaction language (representational system) and "m" an interaction modality. For example, the speech modality can be defined using the "Microphone" as a physical device and "Pseudo-natural language" as an interaction language (Table 1).

Table 1. Examples of interaction modalities

Modality	Mode	Interaction language	Device
Acceleration	Gesture	Direct manipulation	Accelerometer
Speech	Voice	Natural language	Microphone
Hand motion	Gesture	Specialized sign language	RGB camera
Facial expression	Gesture	Specialized sign language	RGB camera
Pointing gestures	Tactile	Direct manipulation	Touch screen
Orientation	Gesture	Direct manipulation	Gyroscope
Speech synthesis	Audio	Natural language	Speakers

Modalities can also be classified according to the required user attention. A modality may be considered active if used consciously by the user, while it can be considered passive if used unconsciously. For example, using hand motions to control a specific element of the user interface is considered an active modality, while capturing the user location with a GPS is considered passive, since it does not need user attention. However, if the user moves on her own to go to a particular location by using the GPS location to create a path, the position using GPS modality may be then considered active.

A system is considered multimodal when it processes "two or more combined user input modes (such as speech, pen, touch, manual gesture, gaze and head and body movements) in a coordinated manner with multimedia systems output" [9].

3 Blended Learning

Over the years, pedagogical methods evolved and consistently improved compared to the educational systems of the past. A significant factor of this progress is unquestionably the increasing use of technology in teaching. Nowadays, most educators prefer to blend

traditional teaching with interactive software, in order to achieve the maximum involvement of their students and to consolidate their learning.

A sufficient description of blended learning could be the following: "Blended learning is the process of using established teaching methods merged with Internet and multimedia material, with the participation of both the teacher and the students" [10]. Still, there are a few matters in question regarding the process of creating blended learning environments [11]:

- Firstly, there is the issue of the importance of face to face interaction. Several learners have stated that they are more comfortable with the part of live communication in merged teaching methods, considering it more effective than the multimedia based part. Others are of the exact opposite opinion, which is that face to face instruction is actually not as required for learning, as it is for socialization. There are also those who believe that both live interaction and online or software material are of the same significance and it is the learner's decision which of the two is more suitable for their educational needs.
- Secondly, there is the question of whether the learners opt for combined learning exclusively because of the flexibility and accessibility of the method, without taking into account if they are choosing the appropriate type of blended teaching. Also, there are doubts regarding the ability of the learners to organize their own learning without the support of an educator.
- Another matter, is the need for the instructors to dedicate a large amount of time to guide the learners and to equip them with the necessary skills in order to achieve their goals. In addition, the instructors need to continue being educated themselves to be able to meet the requirements of blended teaching.
- There is also the argument that schools which have integrated technology into the curriculum are mostly addressed and beneficial to people in a comparatively favorable position in terms of economic or social circumstances. This, however, is refuted by the fact that blended teaching methods are quite affordable and easily accessible. The quick distribution of the multimedia material used in this type of instructing, is considered an advantage, but the universality of the system raises the need to modify the provided material, in order to make it more culturally appropriate for each audience.
- Lastly, a continuous effort is being made to proceed to the new directions given by the novelties of technology while maintaining a low-priced production of educational material. The uninterrupted evolution of communication and information technology is making this effort rather strenuous for the developers of such teaching models.

Despite the difficulties faced when designing blended learning techniques, the advantages of combining face to face instructing with technology are many. Some of enormous importance are [11]:

1. The learner is intrigued by the procedure itself and as a result, the material being taught seems more interesting to them. This leads to the easier accomplishment of the educational goals that the teacher has set.

2. There is no limitation regarding the time or the place of the lesson. A student can attend remote classes being offered online, frequently having the opportunity to record and watch again the lesson.
3. The cost of the teaching process is greatly reduced. Every school unit that uses blended learning, has access to a variety of Internet and software material which would require a lot of time and effort to be independently produced, resulting in additional expenses.
4. In industrial applications of blended learning, it has been observed that the desired results have been achieved twice as fast as with the established instruction methods.

Blended learning should not be conceived solely as a method of enriching the class with technology or making the learning process more accessible and engaging. Its main purpose is to modify and adjust the teaching and learning interaction in order to upgrade it. It assists and enables the growth of critical thinking, creativity and cognitive flexibility [12]. In this context, the following section presents a framework of multimodal interaction to be used in a blended learning environment.

4 Multimodal Interaction Framework for Blended Learning

Based on the goal for an efficient "educational platform for blended learning that enables multimodal player interaction" a framework has been developed that lays between the interaction devices deployed by the users and the engine used by the system platform, endowing the application with multimodal interaction capabilities. Figure 1 demonstrates the framework's architectural overview.

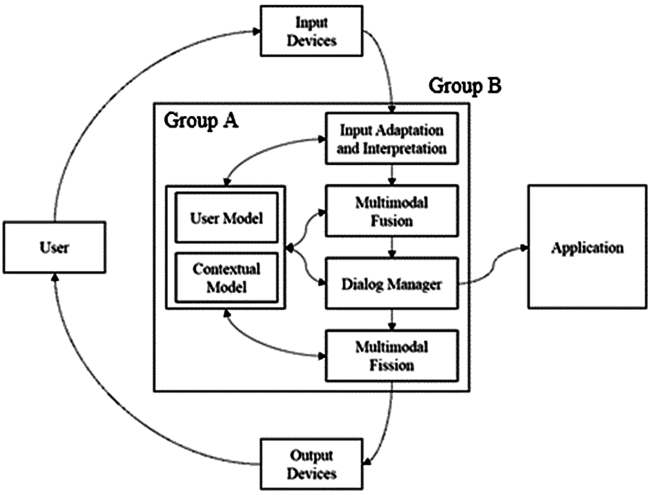

Fig. 1. Framework's architectural overview.

The framework consists of two main modules namely the Context Management module (group A) and the I/O Management module (group B). In more detail the Context

management module includes the "User Model" and the "Contextual Model" components and the I/O Management module includes "Input Adaptation and Interpretation", "Multimodal Fusion", "Dialog Management" and "Multimodal Fission" (see Fig. 1). In the next few paragraphs the framework modules and components are described in detail.

The Input Adaptation and Interpretation component, is responsible for the recognition of data provided by the user. This recognition can be both direct, e.g. speech recognition directly from the microphone, and indirect, e.g. gesture recognition from the Microsoft Kinect sensor.

The Multimodal Fusion component is responsible for interpreting data from the recognizers into meaningful commands. In a blended learning platform, which is deployed in a set-box environment, with often limited processing power, it is crucial to minimize the workload of each component framework. For that reason, the fusion component is centered towards decision level fusion, which assigns a major amount of responsibility to the various recognizers, which provide data that must be interpreted and merged to achieve a final interpretation. A frame-based strategy was chosen due to its simplicity and the ease of augmentation. These augmentations, among others, include attribute constraints and modality prioritization.

The Dialog Manager component, manages changes in the application state, and is responsible for the communication of the framework with the application. It is also responsible for providing output information to the Fission Engine for communication between the framework and the user.

In our approach the dialog manager module uses a finite state machine to identify the command created by the user, and maps the results in a way that is understandable by a system engine. The data that trigger transitions in the finite state machine are generated by the Fusion Engine as described above. This approach is chosen due to the relatively small computational footprint.

The Dialog Manager component, uses a mapping between commands generated by the user (e.g. Command, Location, and Selection) and commands that are understandable by the system engine. This way the Dialog Manager can manipulate the commands communicated to the system engine in order to meet the commands issued by the user. Due to this architectural choice the blended learning platform can augment the functionality of the chosen system engine in order to support multimodal interaction and/or expand the device repertoire of the engine. Also this architecture allows the platform to function with any given system engine as long as a proper mapping of commands is created.

In addition, the dialog manager is responsible for generating data that are passed to the Fission Engine for the communication of messages generated either by the framework itself (e.g. an incomplete command) or the system engine (e.g. auditory notifications, for visually impaired players).

The Fission Engine component. The Fission Engine component is responsible for generating appropriate output messages directly to the user, in a format that is compliant with the user and application context, as well as the environmental variables.

5 Conclusion

In this paper we have presented the ongoing work on a framework that supports multimodal interaction to assist blended learning. Our primary design target is to set up a framework that supports multimodal interaction on educational games according to available I/O modalities, user needs, abilities and educational goals.

Ongoing work covers a variety of issues of both technological and educational engineering character. Some of the issues to be addressed in the future include: (a) Run various use cases in vivo with the guidance and involvement of users and (b) Elaborate further on the Multimodality Amalgamator module to involve more input and output modalities so that the roles between game player and machine are reversed and the player performs gestures, sounds, expressions etc. and the machine responds.

References

1. Oviatt, S.: Ten myths of multimodal interaction. Commun. ACM **42**(11), 74–81 (1999)
2. Oviatt, S.: Advances in robust multimodal interface design. IEEE Comput. Graph. Appl. **23**(5), 62–68 (2003)
3. Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., Ferro, D.: Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. Hum. Comput. Interact. **15**(4), 263–322 (2000)
4. Garofolo, J.: Overcoming barriers to progress in multimodal fusion research. In: Multimedia Information Extraction: Papers from the 2008 AAAI Fall Symposium Arlington, Virginia, pp. 3–4. The AAAI Press (2008)
5. Thevenin, D., Coutaz, J.: Plasticity of user interfaces: framework and research agenda. In: Proceedings of INTERACT, vol. 99 (1999)
6. Elouali, N., Rouillard, J., Pallec, X.L., Tarby, J.-C.: Multimodal interaction: a survey from model driven engineering and mobile perspectives. J. Multimodal User Interfaces **7**(4), 351–370 (2013)
7. Bellik, Y., Teil, D.: Définitions Terminologiques pour la communication multimodale. In: presented at the Proceedings of Interface Hommemachine (IHM) (1992)
8. Nigay, L., Coutaz, J.: Multifeature systems: the CARE properties and their impact on software design. In: Multimedia Interfaces: Research and Applications, Chap. 9 (1997)
9. Oviatt, S.: Advances in robust multimodal interface design. IEEE Comput. Graph. Appl. **23**(5), 62–68 (2003)
10. Friesen, N.: Report: Blended Learning (2012)
11. Bonk, C.J., Graham, C.R. (eds.): Handbook of Blended Learning: Global Perspectives, Local Designs. Pfeiffer Publishing, San Francisco, Chap. 1.1, Blended learning systems: Definition, current trends, and future directions (2005)
12. Singh, H., Reed, C.: A white paper: achieving success with blended learning. Centra softw. **1**, 1–11 (2001)
13. Garrison, D.R., Kanuka, H.: Blended learning: uncovering its transformative potential in higher education. Internet High. Educ. **7**(2), 95–105 (2004)