AmbiLearn:
Ambient Intelligent Multimodal Learning Environment for Children

Jennifer Hyndman, B.Sc. (Hons.)

School of Computing and Intelligent Systems
Faculty of Computing and Engineering
University of Ulster, Magee, Derry.

Supervisors: Dr. Tom Lunney, Prof. Paul Mc Kevitt
100 Day Review Report
December, 2008

Abstract

Multimodality has real potential to improve interaction between computers and people. Research in the fields of pervasive computing and ambient intelligence has resulted in multimodal interfaces that enable the communication between technology and people to be more natural and efficient. Pervasive technologies, such as the Personal Digital Assistant (PDA) and smart phones are prime examples of where multimodal interfaces are of benefit, however these devices are rarely used by children for educational use. With many organisations 'harnessing technology' for educational and entertaining purposes, there is a growing field of edutainment software for children. The combined use of a multimodal interface based on a PDA with an edutainment application would allow a rich experience of collaborative learning where the task is not restricted to a physical location provided by a traditional desktop PC. This research aims to investigate the use of a PDA-based generic multimodal interface to facilitate children's education. Paying particular attention to the interaction between children and technology, this research also aims to provide guidelines for children's usability of mobile devices and multimodal interaction. The proposed system called *AmbiLearn* is a PDA-based multimodal system to enhance children's learning. The multimodality of the system will enable the user to use a combination of speech and pen as inputs and offer rich feedback through the use of speech, graphics and an embodied agent. Appropriate software tools are investigated with regard to the development of AmbiLearn. A research plan is outlined in the form of a Gantt chart and AmbiLearn is compared to similar systems using a range of key criteria. Focusing on multimodal input and output, it is expected that AmbiLearn will support collaborative learning and advance upon existing mobile interfaces improving interaction. In order to test and evaluate AmbiLearn, an educational application in the form of a serious game will be developed, e.g. a treasure hunt game/puzzles. Usability tests will be conducted with a range of children from different schools in the primary education sector.

Keywords: PDA, HCI, Multimodal interfaces, embodied agents, spoken dialogue systems, usability, serious games, learning environment.

1 Introduction

Ambient intelligence involves the convergence of several computing areas such as ubiquitous or pervasive computing, intelligent systems and context awareness with the added appreciation of social interactions of objects in the environment (Shadbolt, 2003). A key technology emerging from this field is intelligent user interfaces, more importantly natural user interfaces. Intelligent user interfaces are multifaceted in purpose and nature, and include capabilities for multimedia input analysis, multimedia presentation generation, and the use of user, discourse and task models to personalize and enhance interaction (Maybury, 1998). The advancement of spoken dialogue systems (Mc Tear, 2004; Delgado et al. 2005; Wahlster, 2006; Bernsen et al. 1998) and multimodal interfaces (Maybury, 1993; Mc Kevitt, 1995/96; Oviatt, 2007) offer new ways of interaction replacing the traditional keyboard and mouse. Handheld devices such as the Personal Digital Assistant (PDA) have now faster processing, increased storage and improved interconnectivity which have made them a universal tool for use in the education, entertainment, health and business sectors (Myers, 2003). This 100 day review report outlines a research proposal into a multimodal interface system for children on a mobile device. Research will focus on children HCI and investigate the use of multimodalities as a way of enhancing interaction between children and technology. Attention will be paid to the use of multimodal communication on a mobile device to support collaborative learning with children. A multimodal system, called AmbiLearn, will be developed combining speech and pen as inputs and graphics, embodied agents and speech as outputs. As a testbed application an educational game will be developed, e.g. a treasure hunt game/puzzles, which will integrate appropriate modalities to allow testing and evaluation of the system. A Literature review is provided in section 2 and section 3 outlines the project proposal. Comparisons to relevant

work in the field are outlined in section 4, section 5 highlights the project schedule for the proposed work and section 6 provides a conclusion.

1.1 Objectives of the Research

The objectives of this research are to:

- Investigate the means for merging multiple modalities on a mobile device (PDA)
- Address the interface design and usability issues of a multimodal system for children
- Consider the possible use of serious games in children's education
- Demonstrate and test the potential of a multimodal, mobile game called AmbiLearn, as an educational tool

2 Literature review

This section provides background information and a literature review relating to the research areas. Multimodal interfaces are outlined with a focus on spoken dialogue systems and embodied conversational agents. The use of games in education is highlighted and literature relating to the field of Human computer Interaction is provided.

2.1 Multimodal Interfaces

Multimodal interfaces are interfaces to a computer which allow more than one mode of interaction. More specifically, the interaction is utilised by a combination of the natural forms of communication between two humans. Oviatt (2007, p. 414) describes a multimodal interface as a "computational interface capable of relatively human like sensory perception". The notion is that a multimodal interface will use more than one of the natural senses such as touch, taste, hearing, sight and smell. There are systems which have provided insight into the use of combined modalities such as Bolt's, *Media Room,* for natural human computer interaction (Bolt, 1980). The *Media Room* contained a large projected screen and a chair with a touch sensitive joystick on each arm. Having a combined use of speech and gesture allowed the user to say "put that there" while pointing to the 'object' and 'position' at the time of the utterances 'that' and 'there'. A more recent example of the combined use of modalities is *SmartKom* (Wahlster et al. 2001). As a multimodal dialog system *SmartKom* combines speech, gesture and facial expressions as well as using a friendly embodied agent 'Smartakus'. The SmartKom project has been a success regarding combining modalities and the research associated has proved to be invaluable to the field of multimodal systems. As of 2006, 52 patents concerning SmartKom technologies were filed, 59 new products released and 29 spin–off products have been developed (Wahlster, 2006). CHAMELEON (Brøndsted et al. 2002) is an intelligent multimedia software and hardware platform which demonstrates the combined use of multiple modalities. The IntelliMedia Workbench (Brøndsted et al. 2001) is an application which runs on CHAMELEON and uses gestures and speech as input and provides a laser pointer and speech as output. Using a physical table, the user can use a combination of spoken utterances while pointing to a position on the table gaining information on the architectural and functional layout of a campus building.

The use of multiple modalities is ideal within multi-user collaborative environments. A study on Children's collaborative interactions (Scott et al. 2003) suggests that sharing a physical display with multiple input devices may improve collaboration due to a heightened awareness of the other user's actions and intentions. Multimodal interfaces such as KidsRoom (Bobick et al. 1998a,b) and TICLE (Tangible Interfaces for Collaborative Learning Environments) (Scarlatos, 2002) are intended for multi-user participation supporting collaboration. KidsRoom developed at the MIT Media Lab, is a fully automated and interactive narrative play space for children (Bobick et al. 2000). The environment uses computer vision and speech recognition to identify users actions and based upon these actions provides output in the form of images, video, light, sound, music and speech. The TICLE system 'watches' as children work together on a Tangram puzzle (Scarlatos, 2002). TICLE uses computer vision to track the individual pieces of the puzzle. Based on the 'state' the puzzle is in, TICLE provides context aware feedback through speech and a graphical display.

2.1.1 Spoken dialogue systems

Speech is the most common method of communication between two humans and is being utilised at a growing pace in computer systems (Bernsen et al. 1998). Spoken dialogue systems contain five main components; Speech recognition, language understanding, dialogue management, language generation and text-to-speech synthesis (McTear, 2004). Each component individually has potential problems which causes difficulties when using speech as an interface modality. Speech recognition is the process concerned with speech as an input and many factors affect the technique implemented in each case. One factor includes the use of a speech system with children. Potamianos et al. (1998) identified extraneous speech patterns and linguistic variability within children's speech data. A further study (Narayanan et al. 2002) outlines the

differences in the acoustic and linguistic characteristics of speech by children compared with adults. The analysis shows that the age-dependant characteristic of the dialog, linguistic and acoustic, highlights the importance of developing children-specific spoken dialogue systems. Environmental conditions also affect speech recognition particularly when the user is mobile, as factors such as music, cross talk and traffic make it much harder for a spoken dialogue system to recognise utterances correctly (Delgado et al. 2005).

2.1.2 Embodied agents

Embodied agents are virtual representations of a person which can make communication between the human and computer more natural and interactive. There are many potential benefits of using an embodied agent in multimodal systems which include: grabbing the user's attention, social interaction, naturalness and non verbal feedback (Ortiz et al. 2007). An early example of an embodied agent is Gandalf (Thórisson, 1997). Gandalf, applied to information on the solar system, is capable of real-time multimodal dialogue that involves perceiving a users intention through speech, body language, manual gesture and gaze and responding with speech, gaze, facial and hand gestures. A limitation of Gandalf was the inability to recognise and generate propositional information, which was the motivation of the development of REA (Real Estate Agent) (Cassell et al. 1999). REA is a full-sized agent capable of speech with intonation, facial gestures, body gestures and posture. As a real estate salesperson, REA is aware of the functions of verbal and non-verbal conversational behaviours thus improving interaction by helping the user remain aware of the state of the conversation (Cassell et al. 1999). Pedagogical Agents are embodied agents designed specifically for the purpose of teaching. The use of agents in educational software has many benefits as they can physically collaborate with the student. Steve is a pedagogical agent used for procedural training in virtual environments (Johnson et al. 1997). With the ability to give instructions and demonstrate activities Steve maintains an episodic memory of situations providing means to explain his actions when asked. Steve continuously monitors students' actions in the virtual world and so can offer context aware assistance when needed. Within the NICE (Natural Interactive Communication for Edutainment) project (Corradini et al. 2005), an embodied conversational character namely Hans Christian Anderson (HCA), was used to interact with children. The children could ask HCA questions about his life, historical period and his fairy tale characters through natural language. Using spoken language and 2D gestures, the user can follow HCA around a fairy tale land and gain context aware information. The Persona Effect (Lester et al. 1997) was an empirical study which outlined that well crafted lifelike agents have an exceptionally positive impact on students. Students perceived the agents as being helpful, creditable and entertaining. Lester et al. (1997, p. 359) states, "the captivating presence of the agents can motivate students to interact more frequently with agent-based educational software. This in turn has the potential to produce significant cumulative increases in the quality of a child's education over periods of months and years."

2.2 Human Computer Interaction (HCI)

The success of any system or new technology is largely dependent on whether the user accepts it. For this reason HCI still plays a major role in all technology related research. HCI is related to the study of interaction between human and computers and as new technologies emerge, new interaction styles and modalities become available. Child-Computer Interaction (CCI) (ChiCI group, 2007) is a sub field of HCI and its attention is focussed on the interaction between children and technology. Research is this field is relatively limited compared to that of mainstream HCI, however just as important when the user group targeted is children. As a growing field CCI researchers share a common belief that the usability of technology for children is worth special consideration (ChiCI group, 2007). Bruckman et al. (2007) identified several characteristics which differ between children and adults and therefore need to be considered when developing systems for children. These include: dexterity, speech, interaction style, reading, and background knowledge. Research has been conducted in the field of designing technology for children and it has been proposed that children should play an active role in the development process. Druin (2002), outlines 4 key roles a child could play to enhance the user needs. These include: the child as a user, tester, informant and design partner. The emphasis here is that although there are certain challenges when using the child in each role, the benefits and rich information that can be retained outweigh the difficulties.

2.3 Games in education

A growing field of interest within education is the use of games as a method of learning. Providing a fun, playful approach games facilitate learning in three ways: learning as a result of tasks stimulated by the content of the games, knowledge developed through the content of the game and skills arising as a result of playing the game (McFarlane et al. 2002). As the computer games industry grows, research is being carried out into the use of these as an educational tool, however, the challenge lies in identifying the pedagogical features needed to balance the 'fun and enjoyable' aspect of a game with the appropriate content needed to raise understanding of a particular domain. Studies suggest that offering appropriate and adaptive feedback, the

embedding of cognitive strategies and animated graphics which reduce task time and increase achievement, have a positive effect on student learning (Rosas et al. 2003). A feature which contributes to the success of games is having a clear goal (Rosas et al. 2003). Storytelling contributes to this feature by illustrating the goal of the game through a mission statement or brief history and game plot. Larson and Peterson (1999) developed and interactive storytelling environment where a player can choose the outcome of the story like a game environment by choosing a certain path. The narrative features available through storytelling offer means to provide the cognitive strategies and appropriate feedback suggested by Rosas et al. (2003). Pervasive technologies and Augmented Reality (AR) has provided many opportunities within educational games. Complementing the real world by proving information or allowing objects to be manipulated, AR does not alter real world experiences, but rather enhances them. Stock et al. (2005, p. 253) state, "Augmented reality systems generate a composite view for the user: a combination of the real scene perceived by the user and a virtual scene generated by the computer that augments the scene with additional information". By augmenting real objects with pervasive devices such as radio frequency identification devices (RFID) and embedded sensors, virtual information can be provided based on location and context. Hinske et al. (2007) used a Playmobil Knights' Castle toy as a platform for teaching about the Middle Ages through AR. By using a PDA and embedded RFID technology, the Augmented Knight's Castle provides verbal commentary, music and visual feedback based on the child's play. Read-It is a multimodal tangible interface to aid in children's reading (Weevers et al. 2004). Based upon the game 'Memory', a child turns a blank block over which has an embedded sensor, the Read-It system then provides visual and auditory feedback on the spelling and pronunciation of the virtual pictographic word associated with that block. Virtuoso (Wagner et al. 2006) is a PDA based Art History game where players must sort a collection of artworks according to their date of creation. Based within a museum, the PDA and embodied agent provide virtual hints and information on a particular piece of artwork through audio and visual display. All of these examples demonstrate the potential of games as an educational tool.

3 Project proposal

The focus of this research is to investigate the use of multimodal communication supporting collaborative learning for children using mobile devices. Problems in spoken dialogue technology, use of embodied agents, modality fusion and HCI limitations on PDAs will be investigated and resolved. Requirements analysis will provide guidelines for developing an application for children and the use of multimodalities within the interface. The overall aim is the design and development of, AmbiLearn, a PDA based multimodal system to facilitate children's education. The multimodality of AmbiLearn will enable the user to use a combination of speech and pen as input and offer rich feedback through the use of speech, graphics and an embodied agent. In order to test and evaluate AmbiLearn, an educational application in the form of a serious game will be developed, e.g. a treasure hunt game/puzzles. Usability tests will be conducted with a range of children from different schools in the primary education sector.

3.1 Methodologies, Software Analysis and Prospective Tools

The development of AmbiLearn may follow a Logical User-Centred Interactive Design (LUCID) methodology (Cognetics Corporation, 2008). The LUCID methodology focuses on ease of understanding and use by assessing user requirements and developing a look, feel and navigational flow which supports the functional requirements of the system. The Object Oriented Analysis and Design (OOAD) methodology has an emphasis on reusability and offers stability as its approach structures the system as interacting objects with key elements such as Information hiding, encapsulation, inheritance and polymorphism. Key design issues including Child Computer Interaction (CCI) guiding principles (ChiCi group, 2008; Druin, 2002; Bruckman et al. 2007) will be considered as well as limitations and constraints of PDA development.

Prospective software tools are being investigated with regard to the development of the project. These include the use of toolkits and platforms in aiding the development of spoken dialogue systems, multimodal systems, embodied agents and fusion of the different modalities. The CSLU Toolkit is a comprehensive suite of tools to enable exploration, learning, and research into speech and human-computer interaction (CSLU, 2004). Developed at the Centre for Spoken Language Understanding (CSLU) at the Oregon Graduate Institute of Science and Technology, the toolkit provides core technologies for speech recognition and text-to-speech synthesis, as well as a graphically-based authoring environment for designing and implementing spoken dialogue systems (McTear, 1999). Microsoft's .Net Speech SDK offers developers powerful APIs and tools to build their own speech-enabled applications using the Speech Server Developer tools for Visual Studio 2005 (Microsoft, 2008). XFace is an MPEG-4 based open source toolkit for 3D Facial Animation. This toolkit offers a platform independent suite of components including an XFace library, XFace Editor and an XFace Player (Balci, 2004). Microsoft Agent is a tool which can be used to develop agents. Speech is generated using a TTS engine and is compatible with windows applications and development within Visual Studio (Microsoft Agent, 2008). COLLAGEN (for COLLaborative AGENt) is Java Middleware for building collaborative interface

agents (Cassell et al. 2001). Developed at the MIT Media Lab, COLLAGEN aids in the development of agents which both communicate with and observe the actions of the user (Rich et al. 1998). Psyclone is a platform which bridges platforms and programming languages. Based on modularity principles inherited from the CDM (Constructivist Design Methodology), Psyclone simplifies the design of and implementation of systems with many components that interact in complex ways. Psyclone incorporates OpenAIR, which is an information and exchange and routing specification. OpenAIR allows flexible component interaction as it provides a language-independent messaging format and a network-independent routing protocol (Thórisson, 2007; Mindmakers, 2007). Webtops such as Glide, Nivido, eyeOS and g.ho.st support collaboration by proving file storage, sharing and access online (Miller, 2008). As AmbiLearn will be deployed on a PDA the features such as memory and processor power will be important in deciding which PDA is best suited. Table 3.1 shows a comparison of features of the leading PDA's.

Product	Operating system	Internal memory	Display resolution	Touch screen	Processor speed	Battery
HP iPAQ 114 Handheld computer	Windows Mobile 6 Classic	64 MB	320 x 240	yes	624 MHZ	Li -ion
Palm Tungsten TE2 Handheld Computer	Palm OS v5	32 MB	320 x 320	yes	200 MHZ	Lithium polymer
PALM ONE TX Handheld Computer	Palm OS V5.4	128MB	320x480	yes	312MHZ	Li - ion
HP iPAQ 214 Enterprise	Windows mobile 6	164 MB	640x480	yes	624 MHZ	Li -ion

Table 3.1 PDA feature comparison

4 Comparison to other work

Table A.1 in Appendix A compares the proposed system AmbiLearn with existing multimodal learning environments. The table provides features of existing systems and highlights with the use of (✓) or (✗) for features which the corresponding system has, or does not have, respectively. Table A.1 shows that the use of Speech as an input modality is widely available in desktop applications with most supported by an embodied agent. The development of AmbiLearn aims to follow this trend and based on a PDA it is envisioned that AmbiLearn will be extended to mobile environments. AmbiLearn improves upon these existing systems by supporting collaborative learning indirectly thought the use of a game. With emphasis on multimodal input and output it is expected that AmbiLearn will advance upon existing mobile interfaces improving interaction.

5 Project Schedule

Appendix B, Table B.1 provides a Gantt chart outlining the plan of proposed work for the duration of this project. The deliverables of the project include a confirmation report and presentation scheduled for June, 2009. A poster will be presented at the research dinner in April, 2010 and there is a 3rd year presentation in January, 2011 at the graduate school research conference. The final thesis will be completed and due for submission in September, 2011. Throughout the project, background literature will be investigated and activities involving the development and testing of AmbiLearn carried out. Expected submissions to conferences such as the International Symposium of Electronic Arts (ISEA, 2009) and the Irish Conference on Artificial Intelligence and Cognitive Science (AICS-09) are noted.

6 Conclusion

This research aims to advance knowledge on the use of multimodal mobile games within children's education, multimodal interfaces on mobile devices and collaborative learning through mobile devices. A review of multimodal interfaces, HCI and the use of games in education has provided background knowledge of this research. Further investigation will be conducted into each of the modalities of AmbiLearn and suitable means for integrating them investigated. Prospective software tools have been outlined which may aid in developing AmbiLearn and problems in child centred design have been discussed. An educational game, possibly treasure hunt/puzzles, will be designed and developed to test AmbiLearn. The testing will include field work with a selection of primary school children to identify the usability of multiple modalities on a mobile device as well as the pedagogical effect it will have in this specified educational domain.

References

Balci, K. (2004) Xface: MPEG-4 Based open Source Toolkit for 3D Facia Animation. *Working Conference on Advanced Visual interfaces, Gallipoli, Italy, May.* New York: ACM, 399-402.

Bernsen, N.O., H. Dybkjær and L. Dybkjær (1998) *Designing Interactive Speech Systems. From First Ideas to User Testing.* London: Springer.

Bobick, A., Intille, S., Davis, J., Baird, F., Pinhanez, C., Campbell, Y., Schütte, A. and Wilson, A. (1998a) Perceptual user interfaces: The KidsRoom. *Communcications of the ACM*, 43(3), 60-61.

Bobick, A., Intille, S., Davis, J., Baird, F., Pinhanez, C., Campbell, L., Ivanov, Y., Schütte, A. and Wilson, A. (1998b) *The kidsroom: A perceptually-based interactive and immersive story environmant.* Technical report, MIT Media Laboratory, The MIT Media Laboratory, 20 Ames Street, Cambridge, MA 02139.

Bobick, A., Intille, S., Davis, J., Baird, F., Pinhanez, C., Campbell, Y., Schütte, A. and Wilson, A. (2000) Perceptual user interfaces: The KidsRoom. *Communcications of the ACM*, 43(3), 60-61.

Bolt, R. A. (1980) Put-that-there: Voice and gesture at the graphics interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, Seattle, Washington, United States, July.* New York: ACM, 262-270.

Brøndsted, T., P. Dalsgaard, L.B. Larsen, M. Manthey, P. Mc Kevitt, T.B. Moeslund and K.G. Olesen (2001) The IntelliMedia Workbench – an environment for building multimodal systems, In *Cooperative Multimodal Communicaion*. LNCS 2155, Berlin/Heidelberg: Springer, 217-233.

Brøndsted, T., P. Dalsgaard, L.B. Larsen, M. Manthey, P. Mc Kevitt, T.B. Moeslund and K.G. Olesen (2002) Developing Intelligent MultiMedia applications, In *Multimodality in Language and Speech Systems (MiLaSS*), Bjorn Granstrom, Inger Karlsson and David House (Eds.), Dordrecht, The Netherlands: Kluwer Academic Publishers, 149-171.

Bruckman, A., A. Bandlow and A. Forte (2007) HCI for Kids. In: A. Sears and J. Jacko, (eds.) *The Human-Computer Interaction Handbook. Fundamentals, evolving technologies and emerging applications.* Second edition. New York, USA: CRC press Inc. 793-810.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H. and Yan, H. (1999) Embodiment in conversational interfaces: REA. *SIGCHI Conference on Human Factors in Computing Systems: the CHI is the Limit, Pittsburgh, Pennsylvania, United States, May.* New York: ACM, 520-527.

Cassell, J., Nakano, Y.I., Bickmore, T., Sidner, C.L. and Rich, C. (2001) Non-Verbal Cues for Discourse Structure. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, Toulouse, France, July.* NJ: ACL, 114-123.

ChiCI Group (2007) *Child Computer Interaction* [online]. Available from: http://www.chici.org/about_us.php [Accessed 04/11/08].

Cognetics Corporation (2008) *The LUCID Framework* [online]. Available from: http://www.leadersintheknow.biz/UserExperienceDesign/LUCIDFramework/tabid/178/Default.aspx [Accessed 14/11/08].

Corradini, A., Mehta, M., Bersen, N. And Charfuelan, M. (2005) Animating an interactive conversational character for an educational game system. In *Proceedings of the 10th international Conference on Intelligent User Interfaces, California, January.* New York: ACM, 183-190.

CSLU (2004) *CSLU Toolkit* [online]. Available from: http://cslu.cse.ogi.edu/toolkit/ [Accessed 23/10/08].

Delgado, R.L.C. and M. Araki (2005) *Spoken, Multilingual and Multimodal Dialogue Systems, Development and Assessment.* Chichester, West Sussex, England: Wiley.

Druin, A. (2002) The role of children in the design of new technology. *Behaviour and Information Technology,* 21(1), 1-26.

Hinske, S. and Lampe, M. (2007) Interactive Educational Play with Augmented Toy environments. *ECRIM News,* 71, 38-40.

ISEA (2009) *International Symposium Electronic Art* [online]. Available from: http://www.isea2009.org/wordpress/?p=27 [Accessed 22/10/08].

Johnson, W.L. and Rickel, J. (1997) Steve: an animated pedagogical agent for procedural training in virtual environments. *SIGART Bulletin, December,* 8 (1-4), 16-21.

Larson, C. and Peterson, B. (1999) *Interactive Storytelling in a Multimodal Environment*. Masters Thesis, Aalborg University, Institute of Electronic Systems, Aalborg, Denmark.

Lester, J.C., Converse, S.A., Kahler, S.E., Todd Barlow, S., Stone, B.A. and Bhogal, R.S. (1997) The Persona Effect: Affective Impact of Animated Pedagogical Agents. In *Proceedings of the SIGCHI conference on Human factors in computing systems, Atlanta, Georgia, United States, March*. New York: ACM, 359-366.

Maybury, M. (1993) Intelligent Multimedia Interfaces. California, USA: AAAI Press.

Maybury, M. (1998) Intelligent user interfaces: an introduction. In *Proceedings of the 4th international conference on Intelligent user interfaces, Los Angeles, California, United States, January*. New York: ACM, 3-4.

McFarlane, A., Sparrowhawk, A. And Heald, Y. (2002) *Report on the educational use of games: An exploration by TEEM of the contribution which games can make to the education process* [online]. Available from: http://www.teem.org.uk/publications/teem_gamesined_full.pdf [Accessed 03/11/08].

Mc Kevitt, P. (Ed.), (1995-96) *Integration of Natural Language and Vision Processing, (Vols. I – IV),* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Mc Tear, M. (1999) Software to Support Research and Development of Spoken Dialogue Systems. In *Eurospeech '99, Budapest, Hungary.* 339-342.

Mc Tear, M. (2004) *Spoken Dialogue Technology. Towards the conversational user interface*. London, England: Springer.

Microsoft (2008) *Speech Technologies* [online]. Available from: http://www.microsoft.com/speech/speech2007/default.mspx [Accessed 02/11/08].

Microsoft Agent (2008) *Introduction to Microsoft Agent* [online]. Available from: http://msdn.microsoft.com/en-us/library/ms695902(VS.85).aspx [Accessed 11/11/08].

Miller, M. (2008) *Cloud Computing: Web-based applications that change the way you work and collaborate online*. USA: Que Publishing.

Mindmakers (2007) *Psyclone* [online]. Available from: http://www.mindmakers.org/projects/Psyclone [Accessed 15/11/08].

Myers, B.A. and Beigl, M. (2003). Handheld computing. *Computer,* 36(9), 27-29.

Narayanan, S. and Potamainos, A. (2002). Creating Conversational Interfaces for Children. *IEEE Transactions on Speech and Audio Processing,* 10(2), 65-78.

Ortiz, A., Carretero, M., Oyarzun, D., Yanguas, J.J., Buiza, C., Gonzalez, M.F. and Etxeberria, I. (2007) Elderly Users in Ambient Intelligence: Does an Avatar Improve the Interaction?. *Universal Access in Ambient Intelligence Environments.* LNCS 4397, Berlin/Heidelberg: Springer, 99-114.

Oviatt, S. (2007) Multimodal Interfaces. In: A. Sears and J. Jacko, (eds.) *The Human-Computer Interaction Handbook. Fundamentals, evolving technologies and emerging applications.* Second edition. New York, USA: CRC press Inc, 413-432.

Potamianos, A. and Narayanan, S. (1998) Spoken dialog systems for children. In *Proceedings of the 1998 IEEE International Conference on Acoustic, Speech and Signal Processing, Seattle, USA, May.* IEEE, 197-200.

Rich, C., Lesh, N., Rickel, J. And Garland, A. (1998) A Plug-in Architecture for Generating Collaborative Agent Responses. In *Proceedings of the First international Joint Conference on Autonomous Agents and Multiagent Systems, Italy, July.* New York: ACM, 782-789.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., Grau, V., Lagos, F., Lopez, X., López, V., Rodriguez, P. and Salinas, M. (2003) Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers & Education,* 40(1), 71-94.

Scarlatos, L.L. (2002) TICLE: using multimedia multimodal guidance to enhance learning. *Information Sciences,* 140 (1-2), 85-103.

Scott, S.D., Mandryk, R.L. and Inkpen, K.M. (2003) Understanding children's collaborative interactions in shared environments. *Journal of Computer Assisted Learning,* 19(2), 220-228.

Shadbolt, N. (2003) Ambient Intelligence. *IEEE Intelligent Systems,* 18(4), 2-3.

Stock, O. and Zancanaro, M. (2005) Multimodal Intelligent Information Presentation. Dordrecht, the Netherlands: Springer.

Thórisson, K.R. (1997) Gandalf: an embodied humanoid capable of real-time multimodal dialogue with people. *First international Conference on Autonomous Agents, Marina del Rey, California, United States, February.* New York: ACM, 536-537.

Thórisson, K.R. (2007) Integrated A.I. systems. *Minds and Machine,* 17(1), 11-25.

Wagner, D., Schmalstieg, D. and Billinghurst, M. (2006) Handheld AR for Collaborative Edutainment. In *Advances in Artificial Reality and Tele-Existence.* LNCS 4282, Berlin/Heidelberg: Springer, 85-96.

Wahlster, W. (2006) Dialogue Systems go multimodal: The SmartKom experience. In: W. WAHLSTER, (ed.) *SmartKom: Foundations of Multimodal dialogue Systems.* Berlin/Heidelberg: Springer – Verlag, 3-28.

Wahlster, W. Reithinger, N. and Blocher, A. (2001) SmartKom: Towards multimodal dialogues with anthropomorphic interface agents. In: *Proceedings of International Status Conference "Human-Computer Interaction", Berlin, Germany, October.* 23-34.

Weevers, I., Sluis, W., Van Schijndel, C., Fitrianie, S., Kolos-Mazuryk, L. and Martens, J.B. (2004) Read-It: A Multi-modal Tangible Interface for Children Who Learn to Read. *Entertainment Computing – ICEC.* LNCS 3166, Berlin/Heidelberg: Springer, 226-234.

Appendix A: Comparison of Multimodal educational systems

System	Year	Device		Input Modality			Output modality			Embodied Agent	Support collaboration	Educational Domain
		Desktop PC	PDA	Speech	Touch	visual	Speech	Graphics	Text			
Gandalf	1997	✓	×	✓	×	✓	✓	✓	✓	✓	✓	Solar system
Steve	1997	✓	×	✓	✓	×	✓	✓	×	✓	✓	Virtual Procedural Training
KidsRoom	1998	✓	×	✓	×	✓	✓	✓	✓	×	✓	Imaginative Story Telling
Storyteller	1999	✓	×	✓	×	✓	✓	✓	×	×	×	Story Telling
TICLE	2002	✓	×	×	✓	✓	✓	✓	×	×	✓	Tangram Puzzle
Read-It	2004	✓	×	×	×	✓	✓	✓	✓	×	✓	Reading and Language
NICE	2005	✓	×	✓	✓	×	✓	✓	×	✓	×	Life and Works of Hans Christian Anderson
Virtuoso	2006	✓	✓	×	✓	×	✓	✓	✓	✓	✓	Art History
Augmented Knights Castle	2007	×	✓	×	✓	✓	✓	✓	✓	×	✓	History (Middle Ages)
AmbiLearn	2011	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	Treasure Hunt / Puzzles

Table A.1 Comparison of Multimodal educational systems

Appendix B: Proposed Research Schedule

	Activities
	Submissions
	Deliverables

Research Activities	2008	2009				2010				2011		
	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep
Perform literature review												
100 Day Review and VIVA												
Develop AmbiLearn												
Confirmation report												
ISEA 2009 conference												
AICS 2009 conference												
2nd Year Poster												
Test and Evaluate AmbiLearn												
3rd Year Presentation												
Thesis write up												
Submit Thesis												

Table B.1 Proposed research schedule