

# FACIAL EXPRESSIONS AND BODY GESTURES AS METAPHORS FOR B2C INTERFACES

## *An Empirical Study*

Dimitrios Rigas and Nikolaos Gazepidis

*Department of Computing, School of Informatics, University of Bradford, Bradford, UK*

**Keywords:** E-Business systems, Multi-Modal stimuli, Usability, Facial Expressions, Body Language, User Interfaces.

**Abstract:** This paper describes an initial survey and an experiment that investigate the use of facial expressions and body language in e-Business applications. Our aim in this research is to investigate the usability aspects of B2C interfaces that utilise speech, non-speech, avatars with facial expressions and gestures in addition to the typical visual graphical metaphors that are currently used. The initial survey was performed with 25 users and the main experiment with 42 users. Results were encouraging and enabled us to understand some of the research issues involved in the use of multimodal metaphors in e-Business. The results indicated that avatars with facial expressions were most preferred by users for e-transactions and the full animated body was the second most preferred option by the users. The text was the least preferred option. These findings enabled us to understand several design issues of multimodal systems for e-Business applications. For example there were issues of multimodal presentation and combination of the different modalities, e.g. speech, avatars, facial expressions, graphics, and text.

## 1 INTRODUCTION

User interfaces for e-Commerce (EC) applications typically use text and graphics with the occasional or limited use of multimodal features. Interfaces for Business-to-Consumer (B2C) e-Commerce applications are used widely for online shopping. Many of these interfaces often use text and graphics to communicate information with the occasional use of auditory stimuli such as speech or earcons. On overall, the use of multimodal metaphors in these type of interfaces is limited. Multimodal metaphors may include the use of speech (synthesised or recorded), earcons (short musical stimuli), auditory icons (environmental stimuli) or even human-based avatars with facial expressions and body gestures.

Many of these different types of modalities were found to be particularly useful in other applications. Examples include the use of earcons to communicate program executions (Rigas, 1998) and to improve the usability of interfaces (Brewster, 1998a).

It is therefore believed that the use of different modalities in interfaces for B2C will provide additional benefits to the usability aspects to those interfaces. The paper describes the initial part of an

empirical study that investigated facial expressions and body gestures in e-Business applications.

## 2 RELEVANT WORK

Electronic Commerce (e-Commerce) is a general term that incorporates any process relating to buying, selling, or exchanging information electronically among buyers and sellers, goods or service providers and other third-party companies (Barnes – Vieyra, 2001, Turban, 2003). A broader definition of e-Commerce is e-Business which refers to all the strategic moves, managerial decisions and techniques of a business collaborating with other business partners in internet or in private networks. Interfaces for e-Commerce applications are often very important for the successful utilisation of e-Commerce transactions. The benefits for electronic commerce include (Wendler, 2001):

*Global impact:* A good e-commerce site can reach and appeal to any use of the World Wide Web.

*Borderless customer base:* Customers can be of a variety of cultural or ethnical background.

*Staff reduction and cost savings* (Tassabehji, 2003): All the expenses for dealing and signing documents from staff are lessened. Transactions are all based online and a well designed e-commerce solution, customers can resolve instantly any problem that appears.

*Presence in established or emerging virtual markets* (Mahadevan, 2002): New opportunities present in B2B and B2C virtual markets. Businesses in order to stay competitive they do a step further from a traditional market to an electronic commerce.

*Paperless transactions*: The traditional way of the use of paper is gradually becoming obsolete.

## 2.1 Multimodality

As e-commerce websites continue to expand the need for interaction and multimodal contents becomes noticeable (Böszörményi, 2001, Jalali-Sohi, 2001). The use of speech, earcons and auditory icons enhance user activities such as browsing and may well improve the capabilities of an application. Some different multimodal metaphors are briefly reviewed below.

*Auditory icons* are non-speech sounds that simulate physical sounds representing an event or an action that took place (Bjur, 1998). Users are familiar to these sounds and they know what they represent since they derive from the natural environment. In Gaver's research (Gaver, 1993) it is stated that "*Auditory icons add valuable functionality to computer interfaces, particularly when they are parameterized to convey dimensional information*". Gaver's SonicFinder is a well known auditory interface he developed for Apple Computers (Gaver, 1989). He used natural sounds to map specific operations for files such as copying, or deleting. The benefits of sound effects were major and added flexibility for the users in receiving information for the interface (Gaver, 1989).

*Earcons* are non-speech sounds that use abstract or synthetic sounds derived from instruments or synthesizers. Earcons (as well as other auditory stimuli) can attract the attention of the user or announce to users an action or a result of an action (Brewster, 1996). They do not appear visually on the screen and they can vary on their level of sound intensity (Brewster, 1998b). Brewster states that users are overloaded with visual information when they interact with computers (Brewster, 1997). His experiments showed that interfaces could be improved in terms of usability by integrating non-speech sounds to the user interface widgets such as menus, buttons, windows or scrollbars (Brewster, 1998c). Brewster, Wright and Edwards evaluated

and verified the benefits of auditory experiments (Brewster, 1993, Borden, 2002). They derived some guidelines on how a graphical interface can be improved by the use of sounds. Users can work more effectively when earcons are used in a graphical interface. An additional auditory modality can be used in the absence of users' attention.

*Speech metaphor* is often used in multimodal user interfaces so as to provide users with feedback along with the graphical environment about system's current state (Preece, 1994) and it is a very useful tool especially for visually impaired users (Lines, 2002). We distinguish two types of speech; natural and synthesised. Natural speech output is a digitally recorded message of a male or female spoken word. It is often useful for applications that require short sentences to be spoken but a dynamic use of recorded speech by incorporating short recorded messages (as building blocks) is a complex process given the need for grammatical structure, context, tone changes and phonemes. Large storage capacity is also required due to the vast vocabulary (Lines, 2002, Preece, 1994). Synthetic speech output is produced using a speech synthesizer. It can be generated mainly by two methods: *Concatenation* or *Synthesis by Rule* (also referred to as formant synthesis) (Lines, 2002, Preece, 1994). Using the concatenation method, digital recordings conducted by real human speech are stored and later on controlled as single words or sentences in a computer (Preece, 1994). An example based on concatenation is when someone uses a phone card and a recorded voice informs the user how many minutes left in the card. The audio message is digitally recordings of each digit separately, controlled by the computer system generating the spoken message. Synthesis by rule, involves the combination of synthesised words generated by rules of phoneme. It is useful for large vocabularies and as a result the quality of speech produced is poorer compared to the concatenation method (Lines, 2002). Janse (Janse, 2002) had studied the perception of natural speech compared to a synthetic speech and derived that although synthetic speech is becoming more and more intelligent, natural speech still more comprehensible for listeners. A number of studies on synthesised speech have shown that natural speech still more intelligible and comprehensible than the synthetic. Voltrax, Echo, DECTalk, Voder, are some good examples of speech synthesizers developed in the past but with poor quality (Reynolds, 2002, Lemmetty, 1999).

Computer-based systems offer a speech technology, known as *Text-to-Speech (TTS)*

synthesis technology (Dutoit, 1999). TTS systems have the ability to read any arbitrary text, they analyze it and after converting it, they output it as a synthesised spoken message comprehensible by the user (Schroeter, 2000, Wouters, 1999). TTS technology is widely used in software applications, and many corporations are taking into account the benefits of involving this technology in EC websites (Kouroupetroglou, 2000). TTS brings out new issues for the development of EC systems (Xydas, 2001) and provides the scope for new innovative applications.

We especially focus on the facial expressions and body gestures that give a more realistic interaction in human computer interfaces. The face is a mean of expressing emotions, feelings, and linguistic information and due to the improvement of computer hardware (high performance graphics and speed) instances of cartoon-like and human-like synthesized faces are under development and researched in depth for use in computer applications (Beskow, 1996a). Animated or realistic characters are used in spoken dialogue interfaces, conveying information for verbal and non-verbal communication by several means of facial modalities e.g. lip-synchronization, eyes gazing and blinking, turn taking and further advanced modeling capabilities such as the use of gestures and motion (Beskow, 1996b).

On the other hand, the use of gestures such as the movement of the body, of hands and head, play a major role in everyday communication. These are often used either to give emphasis to our speech, to point at an object or to illustrate the size or shape of it. Gestures are also known as ‘body language’ communication (Benyon, 2005). In e-Commerce, anthropomorphic agents introduce the use of speech and gestures scoping to interact with the user. They illustrate a mean of natural communication between the human and the machine. For instance, an animated character reading a text would differ from a character which uses gestures as well, since gestures themselves are movements conveying information. It makes the character more dynamic, more personal and believable motivating and it affects users’ decisions since it combines synchronised speech and gesture modalities (Heckman, 2000).

*Avatars* are visual representations of an interactive character in a virtual space displayed in real time virtual environments (VEs) (Dix, 2004, Benyon, 2005, Bartneck, 2004). These VEs have been used for many web-based applications, and they have been evolved in Collaborative Virtual Environments (CVEs) (Fabri, 1999), where users from any physical location they can interact,

communicate and cooperate with each other (Burford, 1999, Prasolova-Førland, 2002, Krenn, 2004). Fabri and Moore (Fabri, 2005) argue that when CVEs incorporates with facial expressive avatars, can be beneficial for people with special needs (e.g. autism), regarding their achievements and their performance (Prasolova-Førland, 2002). More particularly, avatars can be programmed adopting artificial intelligence from the developer. These animated characters interact with the user and basically represents the presence of a mean of facial communication in an application or in a website. Avatars can be categorised and characterised in abstract, realistic or naturalistic (Qiu, 2005, Salem, 2000):

Abstract avatars are non-humanoid avatars. They are mainly embodied by cartoon or animated characters, they have predefined actions and their role is interactive between a computer application and the user. Examples of abstract avatars are the animated characters in Microsoft office suites by giving tips and suggestions to the user during his work process (Qiu, 2005, Hoorn, 2003).

Realistic avatars are highly accurate representation of users. They are used in computer graphics and animation, multimodal applications, console games or even teleconferencing environments so as to give a new high level of realism (Maria-Cruz Villa-Uriol, 1994). They take a set of static images or real time video images in order to be used for the avatar. The costs however of the technology and the hardware that should be used are high (Qiu, 2005).

Naturalistic avatars are a humanoid approach, but with low-level details. They are used in CVEs and they ‘mimic’ some basic humans’ actions like smiling, walking, or waving hands. These virtual environments are three-dimensional and users are entering the world represented by the form of an avatar and a unique ID which interacts with other users-avatars (Dickey, 2003). The most popular 3D virtual worlds are Active Worlds (Active Worlds) and blaxxum (Blaxxum).

### 3 RESEARCH PROGRAMME

User interfaces for B2C applications typically use visual metaphors with the occasional or limited use of multimodal features. This experimental programme aims to investigate usability aspects of B2C interfaces that utilise speech, non-speech, avatars with facial expressions and gestures in addition to the typical visual graphical metaphors that are currently used. More specifically, some of the research questions include:

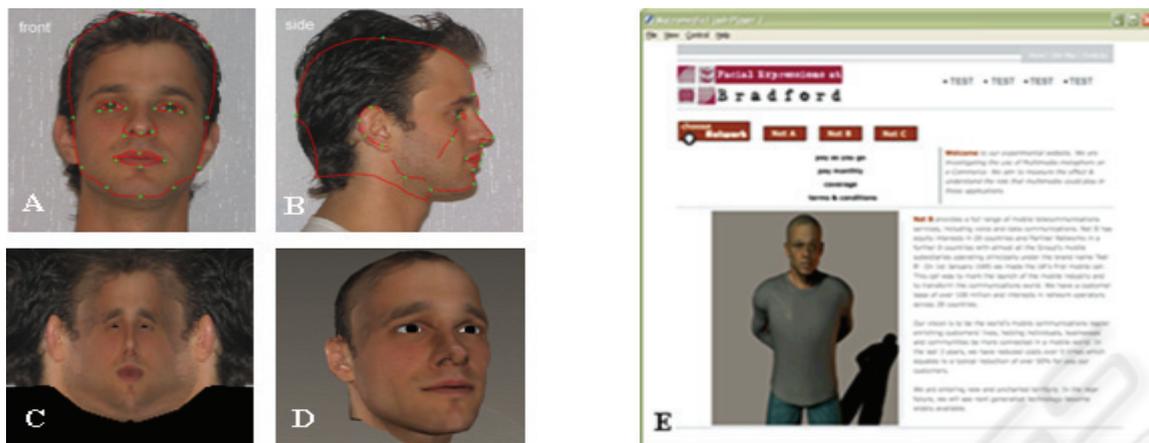


Figure 1: The experimental platform used for the experiments. A front view, B side view, C texture preview, D face sculpting, E multimodal approach.

1. Does the use of natural or synthesised speech have an effect upon the usability of an interface? If so, which of the two increases the usability of the interface?
2. Does the use of auditory icons and earcons have an effect upon the interface?
3. Does the use of avatars with human like facial expressions and different types of voice tones have an effect upon the usability of an interface?
4. Does the use of avatars with human-like gestures have an effect upon an interface?
5. Does multimodal overall provide less or more efficient options for a candidate buyer through the WEB upon his final decision?

Taking into account the recent experimental work (as described in section 2) on multimodal metaphors, including facial expressions and gestures, the following hypothesis can be made:

*"A multimodal e-Commerce website using avatars with facial expressions and body gestures would be more usable and desirable for the users' decisions in terms of accomplishing e-transactions when compared with text and graphics based e-Commerce application"*.

The initial survey and experiment described in the sections below is the starting point to investigate the above hypothesis.

#### 4 AN INITIAL SURVEY

First, an initial survey was performed with 25 users (20 males and 5 females) were used (most of them were students from the University of Bradford, from different departments). Each user answered questions related to some personal details, education

level, and internet experience prior to the demonstration of two websites for which their views were elicited. A textual-based and a multimodal website were demonstrated to the users individually. The text-based version incorporated graphics and the multimodal version incorporated multimodal features such as speech output, face and body animation. The aim of this survey was to explore the usability, believability and likeability aspects of a website that utilises multimodal features, whether is preferable and more interesting or not. After the demonstration of both versions (i.e., multimodal and textual) of the website, users were required to answer a short questionnaire relating to those two demonstrations.

Figure 1 shows the face model and its various stages of development that were used as a basis for the animated face and body. A front view and side view required in digital pictures format so as to create a texture preview that will be attached on the animation model and initiate the face sculpting. The interface design consisted of two versions, i.e. the multimodal and the textual. The multimodal application was using of speech and animation and the textual application was using text and simple graphics.

The data gathered was checked and ensured for its integrity and totality prior to further analysis. Among the users a 60% had an average age of 22-30 years old, another 36% was between 18 and 21 and only a 4% was over 30 years old. 80% of users were males and 20% were females. 64% of the users were in an undergraduate degree level, 20% were in a doctorate degree and the percentage is lower for people with a master degree and college education that is 12% and 4% respectively. On the question

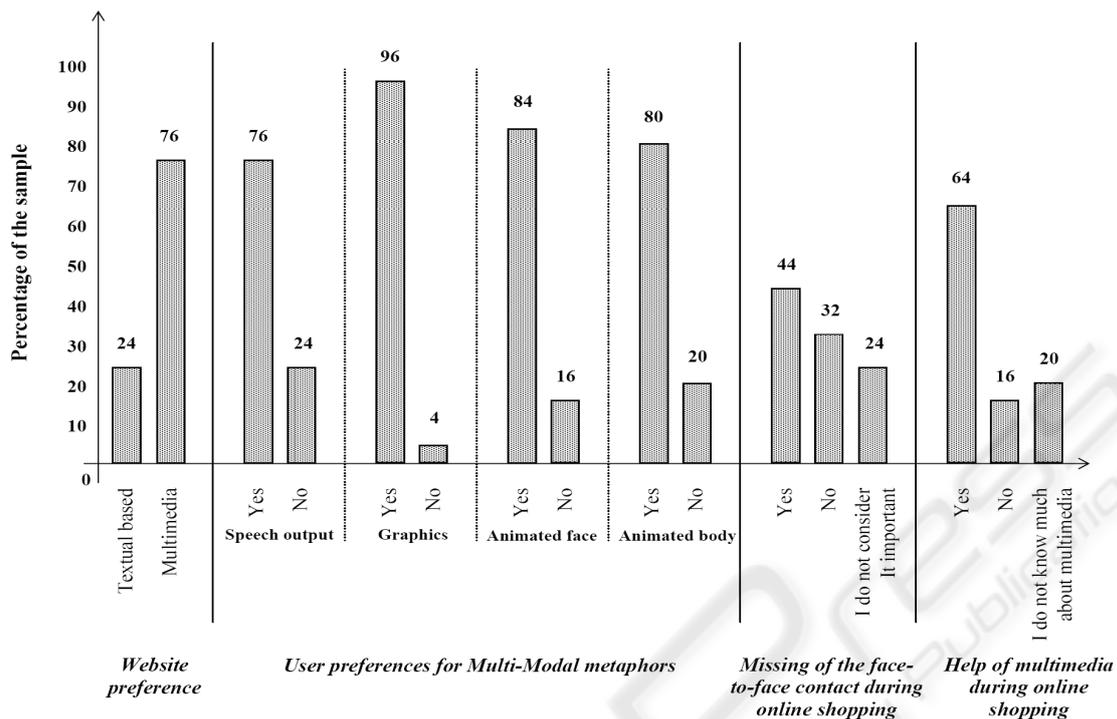


Figure 2: Results in percentages of textual vs. multimodal website, multimodal metaphors and multimodal users' preferences for EC websites.

“How many hours do you spend on the WEB?” most of users answered more than 5 hours a corresponding number of 44%. The rest of the percentages are equally distributed to the rest of the choices. These results can be seen in Figure 2. Users were also asked if they had ever purchased items from the WEB and 72% answered positive whilst 28% negative. The statistical analysis of the survey showed that 76% of them preferred the multimodal version than the textual where almost all of them were very excited with the use of the multimodal features referred above. In the question if they would have been happier with a substitute of a face-to-face contact when shopping online, 44% answered ‘yes’, 32% answered ‘no’ and the remaining 24% of the users did not consider it important.

The results of the initial survey encouraged the development of an experiment in which facial expressions and body gestures were examined further in a B2C application.

## 5 EXPERIMENT

The experiment evaluated three different interface metaphors for a B2C application. These metaphors

involved the use of speech, animation, facial expressions and body gestures for the presentation of three products using three different ways (textual, face, and full body animation). The experiment measured aspects of usability, interactivity and likeliness of the animating application. The experiment was conducted with 42 users (34 males and 8 females).

The B2C application developed involved three products. The brand names of the products were removed in order to avoid brand name influencing user choice. Each product had a textual description which was either presented visually for the textual version or spoken for the face and full body animated versions. In the spoken version, parts of the textual description were emphasised.

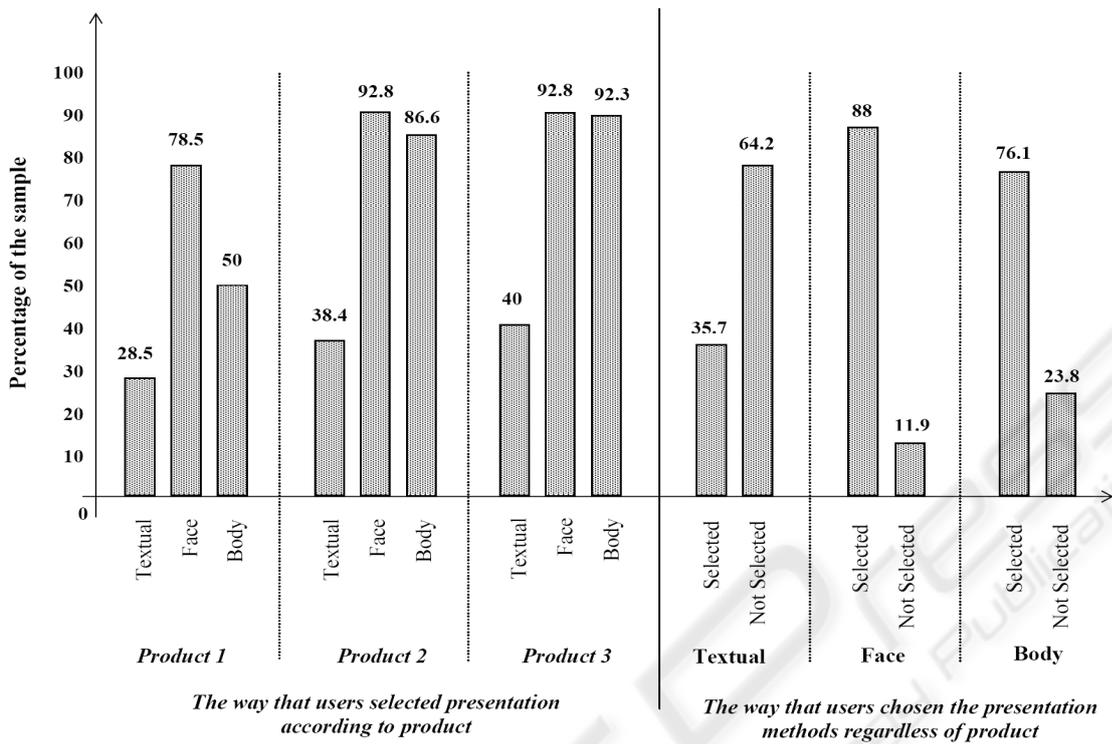


Figure 3: Results in percentages of the way users selected the presentation according to product and regardless of it.

Table 1: Key parameters for product description.

Key parameters	Notation
Speech emphasis	italics
Speech pause	!
Head turning right	} }
Head turning left	{ {
Head movement	[ ]
Eye blinking	( )
Hands clenching	# #
Open right palm	> >
Open left palm	< <
OK gesture	√
Hands on waist	* *
Finger gestures right	//
Finger gestures left	\\

Table 1 shows the notation for the speech emphasis, gestures and body postures of the avatars at particular places in the example of the text given below.

“\*The [(first)] and *smallest* handheld with integrated 3-way (wireless) }capabilities} for your data and voice communication that can be used } worldwide\*. > It{features}> \*Microsoft (Windows) for a familiar [(environment)] to the user\*. >It [{provides}> (wireless) connectivity giving />access{ to the Internet, email and [wireless] [(phone calls)]>/ at [very high-speeds]. \<[(Bluetooth)] technology is also available for easy (data) synchronisation, file [(sharing)], [printing] and [lastly]! it connects to GPRS (networks)<.”

### 5.1 Results

Almost 50% of the users described the *textual* presentations of the products as poor. However, only 33% of the users described the textual presentation as good and 12% described it as very good. During the post experimental interviews, users commented upon the interactivity, attraction, eye-catching, interesting and impressive aspects of the multimodal approach. Some users also commented that it was

easier to interact with a multimodal application compared to the one that is dominated with text and graphics only. The presence of the avatar gave a positive impression to users. Therefore the face-to-face approach was perceived by users that went somehow to simulate a real face-to-face shopping experience.

As far as the multimodal metaphors are concerned, users stated that *speech output* was considered very important. Speech output was selected by 87% of the users. All 42 users regarded graphics as essential and that played an important role in the interaction of the e-Business transactions. *Animated face* was preferred by 80% of the users, and *animated body* was preferred by 67% of the users.

## 6 DISCUSSION & CONCLUDING REMARKS

Many users expressed an interest to the use of avatars with facial expressions and body language. They thought that it would be helpful if such a technique could be used in B2C applications. Users preferred products that were presented using avatars with facial expressions and body gestures than the products presented using text and graphics. Therefore the approach of using avatars with facial expressions and body gestures has provided sufficient scope for further experiments.

The experimental results so far have demonstrated that there is some evidence to support that this simulated face-to-face communication between an online shopper and human like avatar with human expressions and body gestures might help on the usability of the online system.

It is believed that the ability of the users to remember the information communicated by avatars using the facial expressions and body gestures, could be better from when the information was communicated textually. Additional experiments are needed in order to establish further the above point.

The anticipated benefits of continuing this research will be to produce a series of empirically derived guidelines for the designing of B2C interfaces that incorporate avatars with facial expressions and full body gestures in a multimodal interaction. The use of the various facial expressions and gestures would depend upon the different interface instances and it is expected that guidelines will be produced to address at least some of the issues discussed above.

This research work described a survey and an experiment to mainly investigate the use of facial expressions and gestures in avatars that communicate information about products in a typical B2C application. The results demonstrated that the multimodal approach taken was valid and that human-like avatars with facial expressions and body gestures have the potential to benefit the interaction process in applications for e-Business transactions. The avatars with the facial expressions was the most preferred modality by the users (followed by full body gestures). The lessons learned during these experiments enabled us to identify issues of multimodal presentation and combination of the different modalities. A number of experiments are currently designed to investigate these identified issues.

## REFERENCES

- Active Worlds. <http://www.activeworlds.com/>
- Barnes – Vieyra, P., Claycomb, C., 2001. *Business-to-Business E-Commerce: Models and Managerial Decisions*. Business Horizons / May-June.
- Bartneck, C., Takahashi, T., Katagiri, Y., 2004. Cross-Cultural Study of Expressive Avatars.
- Benyon, D., Turner, P., Turner, S., 2005. *Designing Interactive Systems – People, Activities, Contexts, Technologies*. Addison-Wesley.
- Beskow, J., 1996a. Talking heads – communication, articulation and animation. Department of Speech, Music and Hearing, KTH.
- Beskow, J., 1996b. Animation of Talking Agents. Centre for Speech Technology, KTH, Stockholm, Sweden.
- Blaxxun. <http://www.blaxxun.com/en/site/index.html>
- Borden, G., 2002, An Aural User Interface for Ubiquitous Computing. Sharp Laboratories of America.
- Böszörményi, L., Hellwagner, H., Kosch, H., 2001. Multimedia Technologies for E-Business Systems and Processes. Institut für Informationstechnologie, Universität Klagenfurt.
- Bjur, J., 1998. Auditory Icons in an Information Space. Department of Industrial Design, School of Design and Craft, Göteborg University, Sweden.
- Brewster, S., Wright, P., Edwards, A., 1993. An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces. Department of Computer Science, University of York, U.K.
- Brewster, S., Raty, V.K., Kortekangas A., 1996. Earcons as a Method of Providing Navigational Cues in a Menu Hierarchy.
- Brewster, S., 1997. Using Non-Speech Sound to Overcome Information Overload. Glasgow Interactive Systems Group, Department of Computing Science, The University of Glasgow, U.K.

- Brewster, S., 1998a. Using Earcons to Improve the Usability of a Graphics Package. Glasgow Interactive Systems Group, Department of Computing Science, University of Glasgow, Glasgow, G12, 8QQ, UK.
- Brewster, S., 1998b. Using Earcons to Improve the Usability of Tool Palettes. Glasgow Interactive Systems Group, Department of Computer Science.
- Brewster, S., 1998c. The Design of Sonically-Enhanced Widgets. Glasgow Interactive Systems Group, Department of Computing Science, The University of Glasgow, U.K.
- Burford, D., Blake, E., 1999. Real-Time Facial Animation for Avatars in Collaborative Virtual Environments Collaborative Visual Computing, Department of Computer Science, University of Cape Town.
- Dickey, M., 2003. 3D Virtual Worlds: An Emerging Technology for Traditional and Distance Learning Miami University.
- Dix, A., Finlay, J., Abowd, G., Beale, R., 2004. *Human-Computer Interaction*. 3<sup>rd</sup> Edition, Prentice Hall, pp. 379-381.
- Dutoit, T., 1999. A Short Introduction to Text-to-Speech Synthesis. TTS research team, TCTS Lab.
- Fabri, M., Gerhard, M., Moore, D., Hobbs, D., 1999. Perceiving the Others: Cognitive Processes in Collaborative Virtual Environments. Virtual Learning Environment Research Group, School of Computing, Leeds Metropolitan University.
- Fabri, M., Moore, D., 2005. The use of emotionally expressive avatars in Collaborative Virtual Environments ISLE Research Group, School of Computing, Leeds Metropolitan University.
- Gaver, W., 1989. The SonicFinder: An interface that uses auditory icons Human-Computer Interaction. vol. 4, pp. 67-74.
- Gaver, W., 1993. Synthesizing Auditory Icons. Rank Xerox Cambridge, EuroPARC, U.K.
- Heckman, C., Wobbrock, J., 2000. Put Your Best Face Forward: Anthropomorphic Agents, E-Commerce Consumers, and the Law.
- Hoorn, J., 2003. Personification: Crossover between Metaphor and Fictional Character in Computer Mediated Communication. Free University, Amsterdam, The Netherlands.
- Jalali-Sohi, M., Baskaya, F., 2001. A Multimodal Shopping Assistant for Home E-Commerce Fraunhofer Institute for Computer Graphics, 2001.
- Janse, E., 2002. Time-Compressing Natural and Synthetic Speech. University of Utrecht, The Netherlands.
- Kouroupetroglou, G., Mitsopoulos, E., 2000. Speech-enabled e-Commerce for Disabled and Elderly Persons, University of Athens, Greece.
- Krenn, B., Neumayr, B., Gstrein, E., Grice, M., 2004. Life-like Agents for the Internet: A Cross-Cultural Case Study.
- Lemmetty, S., Karjalainen, M., 1999. Review of Speech Synthesis Technology. Department of Electrical and Communications Engineering, Helsinki University of Technology.
- Lines, L., Home, K., 2002. Older Adults' Evaluations of Speech Output. Department of Information Systems & Computing, Brunel University, Uxbridge, Middlesex, U.K.
- Mahadevan, B., 2002. *Emerging Market Mechanism in Business-to-Business E-Commerce: A framework*. Rome, Italy.
- Prasolova-Førland, E., Divitini, M., 2002. Supporting learning communities with collaborative virtual environments: Different spatial metaphors. IDI, Norwegian University of Science and Technology.
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carrey, T., 1994. *Human-Computer Interaction*. Addison-Wesley, pp. 252-253.
- Qiu, L., Benbasat, I., 2005. Online Consumer Trust and Live Help Interfaces: The Effects of Text-to-Speech Voice and 3D Avatars. Sauder School of Business, University of British Columbia.
- Reynolds, M., Isaacs-Duvall, C., Haddox, M. L., 2002. A Comparison of Learning Curves in Natural and Synthesized Speech Comprehension. Marshall University, Huntington, West Virginia.
- Rigas, D., Alty, J., 1998. Using Sound to Communicate Program Execution. Euromicro, p. 20625, 24th. EUROMICRO Conference Volume 2.
- Salem, B., Earle, N., 2000. Designing a Non-Verbal Language for Expressive Avatars. University of Plymouth, U.K.
- Schroeter, J., Ostermann, J., Graf, H.P., Beutnagel, M., Cosatto, E., Syrdal, A., Conkie, A., Stylianou, Y., 2000. Multimodal Speech Synthesis. AT&T Labs – Research, New Jersey, USA.
- Tassabehji, R., 2003. *Applying E-Commerce in Business*. SAGE Publications.
- Turban, E., King, D., 2003. *Introduction to E-Commerce*, Prentice Hall. Upper Saddle River, New Jersey, pp. 3.
- Villa-Uriol, M-C., Sainz, M., Kuester, F., Bagherzadeh, N., 1994. Automatic Creation of Three-Dimensional Avatars. University of California. Irvine.
- Wendler, S., Shi, D., 2001. E-Commerce Strategy on Marketing Channels – A benchmark approach. Industrial and Financial Economics.
- Wouters, J., Rundle, B., Macon, M., 1999. Authoring Tools for Speech Synthesis Using the Sable Markup Standard.
- Xydas, G., Kouroupetroglou, G., 2000. Text-to-Speech Scripting Interface for Appropriate Vocalisation of e-Texts, University of Athens, Greece.