How to Provide "Live Help": The Effects of Text-To-Speech Voice and 3D Avatar on Perception of Presence in Electronic Shopping

by

Lingyun Qiu

Master of Engineering, Shanghai JiaoTong University, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN BUSINESS ADMINISTRATION

IN

THE FACULTY OF GRADUATE STUDIES

(Faculty of Commerce and Business Administration)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 2002

Abstract

With the prevalence of online shopping, companies began to provide real-time communications on their websites to facilitate human-to-human interactions between sales representative and online customers. This study investigates the interface design of such "Live Help" functions. More specifically, it attempts to understand whether the implementation of Text-To-Speech (TTS) voice and 3D avatar in the user interface of "Live Help" affects user's views of the interaction, the service staff, and the website.

A laboratory experiment was designed and conducted to empirically test the hypotheses that TTS voice and avatar will have significant effects on user's perception of presence, flow, and trust. A 3 by 2 full factorial design (with 3 different levels on the dimension of voice and 2 different levels on the dimension of avatar) was adopted and 72 university students and staff members were recruited for the study. Results showed that TTS voice had significant effects on increasing user's perceptions of flow and trust, while 3D avatar enhances user's feeling of telepresence. These findings not only provide a good starting point for studying the virtual communication experience in online shopping, but also offer practitioners useful guidelines on how to integrate and improve the interface design of real time human-to-human communications on the shopping websites.

Keywords: Electronic Shopping, Virtual Communication, Text-To-Speech, Avatar, Social Presence, Telepresence, Trust

Table of Contents

List of Tables

List of Figures

Acknowledgments

The author wishes to express his gratitude to Professor Izak Benbasat for his guidance and supervision in this thesis. In addition, the author would like to say thanks to Weiquan Wang, Kevin Chen, and Jack Jiang for their insightful inputs. Finally, this study was completed with the help provided by two MSc fellows, Gary Schmidt and Ken Guo, who did a great job as the research assistant.

1. Introduction

1.1 Overview

The purpose of this study is to explore a new and interesting phenomenon in the field of Business-to-Consumer (B2C) electronic commerce — the "Live Help" feature provided by a shopping website to facilitate the real-time interactions between sales representatives and online customers. More specifically, it tries to understand how the user interface with such features would affect the user's experience of the interaction, the service staff, and the website. This work attempts to answer the following research questions:

Q1: Will the addition of computer-generated voice supported by Text-to-Speech (TTS) technology and/or 3-D avatar designed to embody a real salesperson increase the user's perception of presence, the perception that a mediated experience is real rather than mediated?

Q2: Will the addition of TTS voice and/or avatar improve the user's interaction with the website?

Q3: Will the addition of TTS voice and/or avatar increase the user's trust of the service representative?

1.2 Background

1.2.1 Customer Service on the Web - Emergence of "Live Help"

With the fierce competition in the B2C e-commerce industry, online retailers are struggling to reach and retain more customers by making their website easy to use and implementing more customer support features, but both academic and anecdotal evidence shows that there is still a lot of room for improvement. For example, studies suggest that customer service for the pre-order part of the shopping process has a positive influence on user attitudes toward internet catalog shopping (Vijaysarathy and Jones 2000), but such support is limited in electronic commerce systems (Jarvenpaa and Todd 1997). A research report (Datamonitor 2002) by Datamonitor, a market research company, further showed that 69% of all potential online transactions in 2001 were abandoned. While most of the popular reasons for abandoning carts centered around processing time for purchases, security concerns, and high shipping/handling costs, 8% of that 69% of stopped transactions resulted from retailers failing to offer consumers a method by which to have their queries answered while shopping.

In fact, online retailers have already realized the importance of real-time human to human contact in establishing and maintaining the customer relationship; they are seeking ways to implement such communication on their websites. Some pioneering companies, such as Landsend.com and Nordstrom.com, have deployed a feature called "Live Help" to help online shoppers find what they are looking for and get

questions answered instantly through real-time communications with a customer service representative. It is reported that such instant interactions successfully bolstered Lands' End's web success (Dukcevich 2002). For example, the average value of an order increases by 6% when a surfer uses the "Live Help" function; an online visitor who uses Lands' End's instant messaging is 20% more likely to make a purchase than a customer who does not.

Currently, most "Live Help" services are implemented through real-time text chat between the shopper and a customer service representative. When a shopper visits a store equipped with "Live Help" software, the merchant is immediately notified that a visitor is at the "front door". As customers move around the web site, the merchant can engage them by proactively greeting them and inviting them to chat online. During the chat, the salesperson could answer any questions raised by customers and even recommend products. The chat could also be initiated by the customer by clicking the "Live Help" button on the significant position of every webpage within the site. Realizing the value of providing close interaction with potential customers, vendors of Customer Relationship Management (CRM) software, such as Liveperson and Facetime, have begun to integrate Instant-Messaging (IM) based customer service into their software.

Although not much academic research work has been conducted in this area, it is assumed that a website with real time text-based interactions could provide customers a better experience compared with those "silent" sites. For example, research (Aberg and Shahmehri 2001) showed that integrating human assistants into web systems is a

way to provide efficient user support. Furthermore, this integration makes a web site more fun to use and increases the user's trust in the site. Such support also improves the site's "store" atmosphere.

In the industry, "Live Help" had also been proved to be both feasible and worthwhile. A quick glance at the customer list of major software vendors like RightNow, Liveperson and Facetime showed that three types of businesses would significantly benefit from this implementation:

1. A service-intensive online business, such as an online retailer and financial service provider. This type of online business requires intensive human-to-human communication and a large amount of real-time exchange of information. Actually, customer-care is indeed the most important assets in those industries. Although providing human service online may incur some additional costs, the benefit or the risk of not providing such service could never be overestimated.

2. A small online business. "Live Help" function would enable the shop owner to provide basic online customer service and good opportunity of cross-selling at a relatively low cost, compared with alternatives like toll-free calls.

3. An online or brick-and-click business with an existing call center. Actually, the department which provides "Live Help" service is a web-based call center, which could be easily integrated with current call center after software deployment and staff training. Compared with phone-based customer service, online "Live Help" not only costs less, since the customer service representative could multi-task and no phone

call is made directly, but it also provides richer support and more chances of selling by giving product recommendations and analyzing customer shopping behavior.

Research reports from the industry also support such claims. An analysis report from Gartner Group (Elliot 2000) projected that "*...by 2002, only 20 percent of call centers will have integrated live Web contacts or email response management systems with their telephone-based agents; by 2005, 70 percent of call centers in geographic areas with high Internet adoption rates, such as North America and parts or Europe, will support integrated live Web contacts and e-mail response management systems for their telephone-based agents.*" Another report from Gartner Group (Elliot, Kolsky et al. 2001) compared different ways of providing customer service, including both synchronous channel, such as phone and text chat, and asynchronous channel, such as email. It is estimated that the assisted service costs for email range from $5 to $10 per email; the live phone interaction is on average $5.50 per call and text chat costs are approximately $7 per session, although some websites appear to achieve much lower costs, possibly because the interaction is short and the agents can handle many at a time.

1.2.2 Big Picture - Online Shopping Experience

Researchers and practitioners have been studying how to help e-commerce customers enjoy the convenience and efficiency of online shopping, while at the same time reserving or replicating the elite part of their shopping experience in the physical world. In other words, how can we restore a full-scale e-shopping experience in the

virtual space? In this study, we propose that there are four types of interactions taking place during the shopping activities in the "real" world: 1) customer-product interaction, 2) customer-store interaction, 3) customer-salesperson interaction and 4) customer-customer interactions. These interactions have been partially duplicated in the current online shopping environment, but there is still a long way to go before a perfect e-shopping experience could be realized.

1) Customer-product interaction (also known as Virtual Product Experience (VPE)). Researchers have found that through virtual control functions implemented by multimedia software, like Macromedia Shockwave and Apple QuickTime, the customer could acquire more diagnostic information of the products than from pure text and plain images (Jiang and Benbasat 2002). Although the product category applicability of VPE is still under further investigation, such interactions could surely provide to the customer some "look-and-feel" perceptions, the lack of which is one of the most frequently cited disadvantages of online shopping (Rose et al 1999), especially on products' experiential attributes.

2) Customer-store interaction (also known as Virtual Store Experience (VSE)), that is, the experience customers get from a store itself. Just as observed in the popular book "*Why We Buy: The Science of Shopping*" (Underhill 2000), both the layout of the store and the arrangement of the products can greatly affect a shopper's behavior and experience. The corresponding part in the online store is how the customer perceives and responds to the design of the online store per se. Factors to consider in the design include: the text color and font size, the design of the product

page and the navigation structures. Studies on website usability (Nielsen et al. 2000) mainly focused on this dimension.

3) Customer-customer interaction. Shopping, to a great extent being a social activity, inevitably involves interactions between people. The experience of hanging out with friends or family members could also be partially duplicated in the online world either through an asynchronous channel like online forum, where visitors could share their opinions on the merchant and products, or through more advanced co-browsing features supported by an online collaboration software such as Groove. Such interaction becomes part of the online customer's virtual communication experience (VCE). In this paper, Virtual Communication Experience refers to the customer's perception of interactions that occurred between her and other real people (including other customers and service staff); those interactions are mediated by a website or software.

4) Customer-serviceperson interaction. The other half of virtual communication happens between the customer and the salesperson or other service personnel. There are currently software agents partially fulfilling such roles, for example, a product recommendation agent. However, human-to-human interactions still need to be reestablished on the web, especially for the three types of online businesses aforementioned. Although the traditional communication channels like E-Mail and phone calls are still applicable, the face-to-face-type synchronous interactions are missing in most online stores, which has not only degraded the "store atmosphere", but also induced negative influences on the merchant's level of service and

credibility. We think the "Live Help" function would play an important role in remedying this problem.

The research framework is summarized as Figure 1-1. This figure shows the relationship between different components of user's online shopping experience. This study will focus on the issues of Customer-Serviceperson Interaction, which consists part of the user's Virtual Communication Experience.

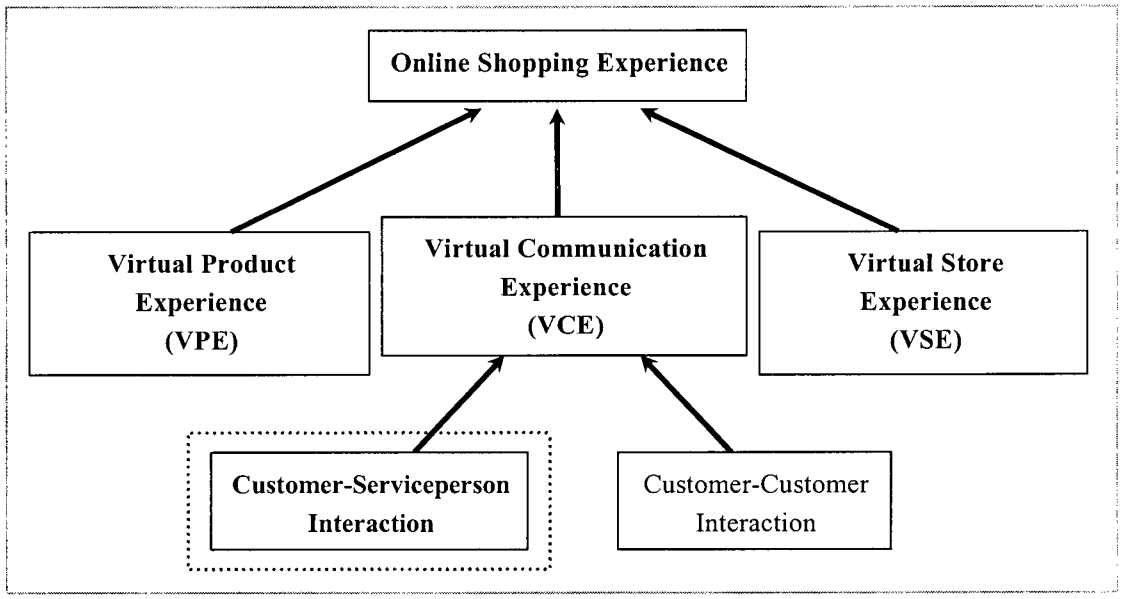


Figure 1-1 Research Framework of Online Shopping Experience

1.3 Motivation

This study will mainly focus on the interactions which occur between a customer and a service representative through "Live Help" features provided by online shopping websites. Although real-time collaboration technologies have come a long way and proved helpful, it is still tedious to communicate via text-only instant chat. The availability of relatively low-cost multimedia and virtual reality (VR) technology such as Text-To-Speech Voice and 3D avatar could make it possible to enrich both the interface and the user's shopping experience. This availability leads to the following questions: *Would the customers appreciate the changes? Would such multichannel and multimedia interface really improve the interactions between online customers and the salespeople? Could the customer perceive a more face-to-face like communication with the help of those technologies? How could such communications benefit online merchants?*

The purpose of this study was to answer these questions through an empirical test and thus contribute to the theory and practice of computer-mediated communication (CMC) and Human-Computer Interaction (HCI) in the context of electronic shopping. The theoretical contributions will aid in understanding and predicting the relationship between presence and shopping experience, employing text-to-speech technology and 3D avatars. Based on the theory of social presence and communication modality, a controlled lab experiment was conducted to investigate whether or not the presence of a computer-generated voice and an anthropomorphic animation character on the Live

Help function would help people perceive more sociable interactions, and whether such enrichment would improve the general experience of online shopping in areas such as flow and trust. Understanding of these relationships can also guide B2C e-commerce practice. In the following sections, we will fully explore and test the research model and hypotheses. The next chapter builds the theoretical foundations and reviews relevant previous research in this area. The research model and hypotheses are presented in Chapter 3. Chapter 4 describes the experimental design and methodologies used during the data collection phase of the study. Chapter 5 summarizes and discusses the data analysis process. Chapter 6 discusses the implications, limitations, and conclusions of the study as well as suggestions for future research.

2. Theoretical Foundations and Literature Review

2.1 Computer-Mediated Communication

Researchers have studied person-to-person communication mediated by different technologies for a long time (for a historical review, see Short 1976). There are also numerous empirical studies comparing face-to-face and computer-mediated communications (for a detailed summary, see Bordia 1997, Hoffman et al 1996). But most of these studies are conducted in the field of Computer-Supported Cooperated Work (CSCW) and the principal goal is to study communication media's impact on people's collaboration behavior in the working environment. Few research works were aimed at investigating communications between customers and serviceperson in the environment of online shopping. Among them, Aberg and Shahmehri (2001) used a field study to explore the impacts of integrating text-chat into customer support of a Web Information System. They concluded that such integration is a way to provide efficient user support. It makes a web site more fun to use, increases the user's trust in the store; and improves the store atmosphere. Basso et al. (2001) compared four types of shopping website interfaces: 1) a standard web interface with no audio or real time interpersonal communication; 2) a web interface with text and audio descriptions of the product; 3) a web interface with instant messaging system between user and salesperson; 4) a TV-based application that allowed a salesperson to talk to the customer while "pushing" catalog items to the customer's TV. They found that

real-time interactivity, but not information expressiveness like voice, increased judgments of friendliness and of the trustworthiness of the salesperson.

In designing customers' virtual communication experience, it is not enough to focus only on the media characteristics, but rather on how to provide the user enough "perceptual bandwidth" (Reeves and Nass 2000). Reeves and Nass (2000) proposed that there are two different sources of information that a computer could present as a perceptual stimulus. Firstly, the computer would be nothing more than a conduit to another person or object, which is the classic case of computer-mediated communication. In these cases, the perceptions of other people and objects may change because the stimulus information about them filters through the representational capabilities of a machine. From this perspective, we could compare the communication effectiveness and information richness of different media used to connect the web customer and the salesperson, for example, text chat, text chat with TTS voice and text chat with 3D avatar.

Second, a computer could also present a virtual social actor, such as the human-controlled 3D avatar representing a virtual salesperson, which is "semantically" controlled by humans while partially powered by software. In this category, the traditional communication theory may not be fully applicable because the aim is not to replicate the real human-human communication, which in this study may be realized even better by real-time web-cam motion pictures and Internet Phone, but create a virtual representation that could not exist in the real world (with 3D avatar as its outlook and text-to-speech synthesis as its voice). The ultimate goal is not

to replicate the real image or voice of the "behind curtain" salesperson, but design an acceptable (in the future, a perfect) embodiment of her, who would stand on the front stage and interact with several customers simultaneously. The rationale behind this proposal is that although the faithful replication of face-to-face interaction is preferred to other mediated communication, it may not be the most efficient one in the environment of online shopping. Online retailer could greatly reduce its labor costs while improve customer's communication experience if it could use an avatar to generate an as-real-as-it-get face-to-face interactions, while at the same time enable the serviceperson to communicate with several customer simultaneously.

The interaction between a web customer and customer service representative is explained in Figure 2-1:

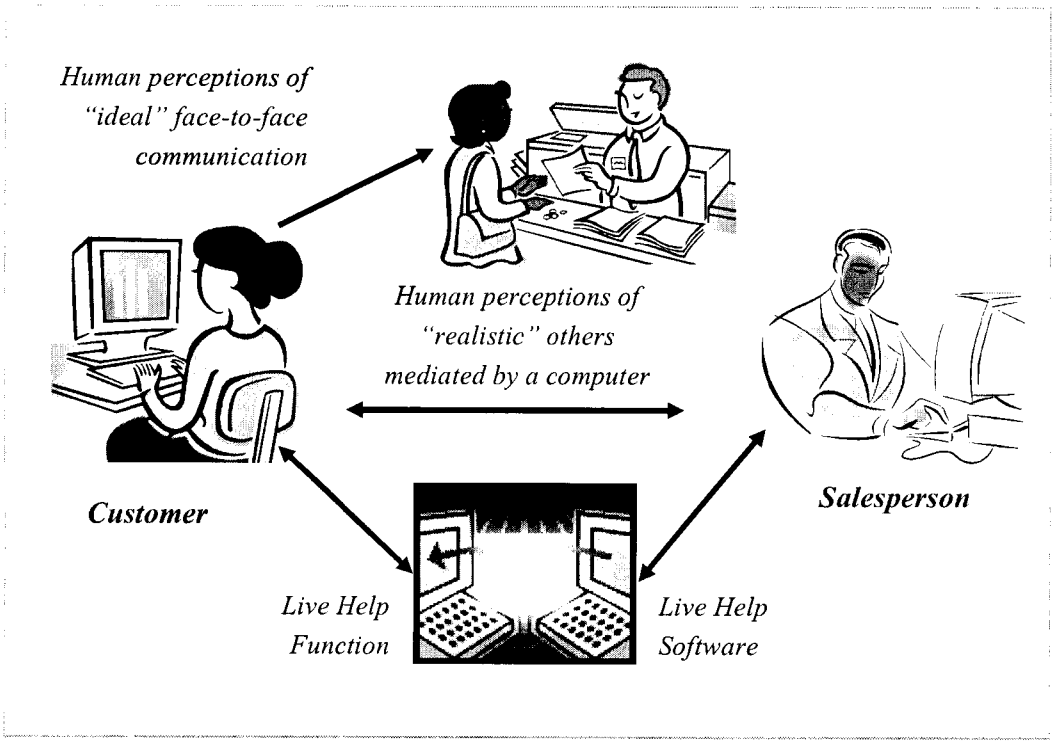


Figure 2-1 Interactions between Online Customer and CSR

2.2 Computerized Voice and Text-To-Speech (TTS) Technology

One way to increase user's perceptual stimuli is to add voice into the interaction. In the context of online shopping, there are various ways to fulfill voice interaction between customer and service staff, such as:

- The "Call Back" function which lets the customer leave his phone number on the website so that some serviceperson could call her later;
- The "Talk through Computer" function by which the two parties could talk through Voice over Internet Protocol (VoIP) technology; (this feature enables the dial-up users to talk with a serviceperson without disconnecting the internet access to free up her phone line).
- The third example, which is the research interest of this paper, is to use a Text-To-Speech (TTS) synthesizer.

According to Dutoit (1997), a TTS synthesizer is a computer-based system that should be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system. The fundamental difference between the TTS system and any other talking machine (for example, a cassette-player) is the automatic production of new sentences. In other worlds, the TTS system is not simply concatenating isolated words or parts of sentences but automatically producing speech "through a grapheme-to-phoneme transcription of the sentences to utter." Currently, most TTS

products in the market are powered by the software engines from Lernout & Hauspie, Microsoft and AT&T.

Compared with TTS, there are some significant drawbacks for either "Call Back" or "Talk through Computer" for online customer service; those are also the reasons why we argue that TTS may be a good alternative to realize voice communication in online customer service:

1) The customer service representative cannot multi-task, in other words, she cannot interact with several customers at the same time. This would directly increase online merchant's operating costs;

2) "Call Back" cannot provide real-time interactions since it would be inconvenient for the user to wait for an incoming phone call to solve some trivial questions encountered on the website, and the shopping experience as a whole would be fragmented;

3) The voice quality of "Talk through Computer" is subjected to bandwidth limitation. In a research by Zhang et al. (1998), it is shown that the delay and choppy sound caused by VoIP on dial-up internet connections can have detrimental effects on communication.

Although TTS systems are increasingly used to develop more human-like interfaces and are widely implemented in the area of automated voice response services, Interactive Voice Response (IVR) systems, computer games and even some Virtual Reality Environment (VRE) like Cybercity.com and Blaxxun communities,

there is limited research work exploring people's perception of the computerized voice. In the studies by Nass and Lee (2000), it was found that participants would regard the computer voice more attractive, credible and informative and evaluate the product more positively if the perceived personality of the voice matches their own. It is also demonstrated that users employ social cues even they know the voice is not coming from a real person. Jensen et al. (2000) compared four forms of communication and found that TTS voice is perceived significantly more positively than text-only interaction, although the researchers had previously assumed that TTS voice might be artificially lacking cues that people use in predicting the veracity of utterance.

In this paper, we are going to investigate TTS as the channel of voice interaction. One big difference between TTS-based "Live Help" and other communication channels is that the voice is computer-generated rather than from a real human being, even though the script (text) is prepared by a real person. The other major distinction of TTS in this study from those of CSCW research is that our user understands that she is interacting with a real human being (the customer service representative), rather than with computer software. Nass and Lee (2000) found out that users employ social cues even with highly-presented reminders that the voice is not coming from a person but a computer, which make it more like a Human-Computer interaction rather then Computer-Mediated communication. What we want to know in this paper is whether the user would respond positively to a voice communication even though they knew the voice was generated by computer rather than from the salesperson herself.

In general, TTS systems are evaluated according to two dimensions: 1) intelligibility and 2) naturalness of the resulting speech (Nass and Lee 2000). Currently, the speech generated by the best TTS systems could score almost as perfect in intelligibility as human speech with a score of 97%; but in naturalness the best system could only be considered as "good", indicating that it still does not completely match the quality and prosody of natural human speech. In this study, we are going to use the most advanced and sophisticated TTS engine from AT&T Natural Voices, which is arguably "the most human sounding" product on the market (McDonald 2001).

2.3 Anthropomorphic Agent and Avatar

The other question we are going to explore in regards to the "Live Help" interface is whether a 3D avatar should be used to visualize a customer service representative in the process of communication. In the literature, there are various definitions of either an avatar or an anthropomorphic agent. According to Badler (1997), an anthropomorphic agent is defined as a virtual human figure representation that is created and controlled by computer programs while an avatar is a virtual human controlled by a live participant. In this paper, we will adopt this definition whenever we mention anthropomorphic agent or avatar.

2.3.1 Anthropomorphic Agent

There are numerous research papers about application of anthropomorphic agents and their interaction with human beings. But just as Isbister et al. (2000) mentioned, although the "bots" on some websites could answer questions and direct visitors in a friendly way, these agents are designed to *simulating* human-to-human interactions, rather than *facilitating* human-to-human interactions. Furthermore, researchers raise concerns on whether the current Artificial Intelligence (AI) technology could truly provide the user a similar feeling of interactions compared with talking with real person. For example, King (1996) argued that although anthropomorphic forms provide a great degree of subtlety and afford social interaction, these forms may be problematic since they may be inherently interpreted as having a high degree of agency and intelligence; Brennan (1992) wrote "*...it's irritating to interact with a*

system that's superficially anthropomorphized or cute. And in an electronic medium, where communication takes place over a channel much narrower than face-to-face, imitating a human being can be misleading." While technologies do improve quickly and significantly over time, and a current software-powered agent could even chat with a human in a quite reasonable way (for example, the virtual host "Luci" on the website of artificial-life.com), it is still difficult for them to pass the Turing Test. This is a popular way to test whether a user can, only by asking questions, differentiate the computer from a real human. Some major trials of business applications of anthropomorphic agents have also proven unsuccessful, such as the Office Assistant included in Microsoft Office software. Currently, most software-based customer service applications on B2C websites are actually not very advanced, offering mainly searchable FAQs, which hardly provide enough support and interactions for online shoppers.

Based on the above discussion, we will focus our investigation in this research project on human-controlled interaction interface – the avatar.

2.3.2 Avatar

The word "avatar" is most frequently used in the study of Collaborative Virtual Environment (CVE), meaning a 2D or 3D image representing computer users. In this way, the participants of a CVE would be directly visible to themselves and to other participants, which creates a more immersive experience and has greater potential to provide important information to a user about its partners and the environment. There

are two aspects which need to be addressed for the implementation of an avatar: function and representation.

1) Function. Avatar has been used in CVE to represent user's presence, orientation and location. Salem and Earle (2000) proposed that an avatar could also be used to facilitate communications between users. In our study, the avatar integrated in an online shopping website would be used to represent a customer service representative. We would consider the avatar as part of the communication medium which could be used to assist the interactions of ideas and emotions rather than an orientation tool in a virtual environment.

2) Representation. According to Salem and Earle (2000), avatars could be realistic, abstract or naturalistic in form:

Realistic avatars are designed to provide an accurate representation of the user. For example, we could use a real time video image captured by a video camera as the face of the avatar. But this implementation is restricted by certain technical factors such as the bandwidth and the computational power of the user's PC.

Abstract avatars use some cartoon image or fictitious entity, for example, an animated talking dog or even a paper clip. Although such implementations have long been used in software-powered interface agents, such as in Microsoft Office and Windows XP, its effects in supporting human-to-human communication have not been empirically supported. In the context of online shopping, we assume it would be unnatural and confusing for the user to interact with the avatar as the embodiment of a

salesperson, although the validity of this assumption still needs further testing in future research.

Naturalistic avatars are usually humanoid in form, but have a degraded level of detail. This type of avatar can emulate natural paradigms just enough to achieve recognition of familiar features such as a smile, a waving hand and nodding. We advocate this type of avatar in our study because it is not only technically feasible, but also because it preserves the multi-modal communications and conventional conversational habits of face-to-face interaction. The use of non-verbal communication capabilities, facial expressions, hand gestures and body postures would be easily perceived and understood by the customer while at the same time enriching their experience. Some preliminary products have been developed to support computer-mediated communications, but few of them are designed for e-commerce settings. For example, Microsoft has produced virtual communication software, named Microsoft Chat, where symbols were extracted from text messages to automatically generate body gestures, facial expression and background, although in 2D form.

2.4 Virtual Communication Experience in Online Shopping

Just as mentioned in Kumar and Benbasat (2002), "*shopping has become an experience and malls have become entertainment centers with communities.*" In the environment of online shopping, the only thing a consumer can interact with is a networked computer and a software-powered website. How can the real-world experience be replicated or even be enhanced through this computer/website-mediated interaction? Will the customer get the same experience as they would in a physical store? In this paper, we will see whether customer's perceived interaction with a salesperson in a physical store can be created or partially replicated through the integration of TTS and a 3D avatar with the website, and we will primarily investigate their effects on the customer's perception of Presence.

2.4.1 Concept of Presence

There are numerous explanations for the concept of presence in the literature. Lombart and Ditton (1997) provided a complete review in which they summarized six categories:

1) *Presence as social richness*: the extent to which a medium is perceived as sociable, warm, sensitive, personal or intimate when it is used to interact with other people. Most relevant MIS and organizational behavior research has been done under this category, including the social presence theory proposed by Short (1976) and the media richness theory by Rice (1992).

2) *Presence as realism*: the degree to which a medium can produce seemingly accurate representations of objects, events, and people – representations that look, sound, and feel like the "real" thing.

3) *Presence as transportation*: the illusion generated by the media that: a) users feel as though they are being transported to another place; b) another place and objects being transported to the user; c) two (or more) communicators are being transported together. One frequently used concept in this category is "telepresence" (Biocca, 1997; Steuer, 1992), which was defined as "*the experience of presence in an environment by means of a communication medium.*"

4) *Presence as immersion*: the compelling virtual reality experience that users' senses are immersed in the virtual world.

5) *Presence as social actor within medium*: the treatment of mediated entities as social actors. That is, the mediated nature of the interaction is ignored and the media personality is incorrectly perceived as a social actor.

6) *Presence medium as social actor*: the social responses of media users to cues provided by the medium itself. As Reeves and Nass (1996) demonstrated in their studies, it is found that because computers use natural language, interact in real time, and fill traditionally social roles, even experienced computer users tend to respond to them as social entities. Kumar and Benbasat (2002) proposed the theory of Para-Social Presence, in which the website itself is considered a social actor.

In this research, we have adopted Lombard and Ditton's (1997) definition that presence is the perception that a mediated experience is real rather than mediated, in other words, "an illusion of non-mediation." In other words, we are testing whether the use of TTS and a 3D avatar would make the customer's experience more like a non-mediated face-to-face interaction. More specifically, we will investigate from the perspective of social richness and transportation, in other words, the dimension of social presence and telepresence.

2.4.2 Source of Social Presence

It is generally believed that the more human senses involved in a communication medium, the stronger the feeling of social presence. For example, Short et al. (1976) found that participants reported greater social presence after an audio-visual task-based interaction than after one which was audio-only. According to Christie (1974, as cited in Lombard and Ditton, 1997), research has shown that visual media have more social presence than verbal (audio) media, which in turn have more social presence than written media. Lombard and Ditton (1997) also mentioned that the number of sensory output channels is not the only important factor in generating the sense of presence. The consistency of information in different modalities is also vital.

However, the aforementioned conclusions are also influenced by the factor of task. Douglas (1957, as cited in Short et al. 1976) proposed that two distinguished aspects of exchange, called "Interparty" exchange and "Interpersonal" exchange, happened in any interaction. Morley and Stephenson (1969, as cited in Short et al.

1976) argued that the balance of these two could be affected by the communication medium, and the emphasis on one aspect would only be realized at the expense of the other one. Based on these arguments, Short et al (1976) proposed that the verbal channel contains the "Interparty", task-oriented, cognitive material while the key roles of multimodal cues are in the communication of "Interpersonal" attitudes.

In the context of online shopping, the extent that the shopping activity emphasizes "Interparty" exchanges over "Interpersonal" exchanges will have a big impact on deciding whether we should add more non-verbal communication cues into the interaction.

2.4.3 Source of Telepresence

The perception of telepresence can be influenced by the structure of the mediated presentation, the content in the medium, and the media user's characteristics. Steuer (1992) argued that a sense of telepresence is based in large part on two determinants: a medium's vividness and its interactivity.

1) Vividness

According to Steuer (1992), vividness means "*the representational richness of a mediated environment as defined by its formal features; that is the way in which an environment presents in formation to the senses.*" Two generalized factors would contribute to vividness: a) sensory breadth, which "*refers to the number of sensory dimensions simultaneously presented*", b) sensory depth, which "*refers to the resolution within each of these perceptual channels.*"

In this study, we consider that the degree of vividness could be a very important factor that differentiates the effects on the perception of telepresence. Whether or not the TTS and the 3D avatar are integrated into the Live Help interface would greatly change the vividness of the communication medium. For example, Nass et al. (1994) found that the use of computerized voice could evoke gender stereotypes; and anthropomorphic properties are important because they facilitate the real time transmission of some of the body's communication cues (Biocca, 1997).

2) Interactivity

Interactivity is another presence factor, although there is little agreement on a specific set of conceptual and operational definitions. Choi et al. (2001) classified it into two perspectives: mechanical view and interpersonal view.

From a mechanical point of view, interactivity is defined (Steuer 1992) as "*the extent to which users can participate in modifying the form and content of a mediated environment in real time.*" Lombard and Ditton (1997) further extend this into five subsidiary variables:

1) The number of inputs from the user that the medium accepts and to which it responds;

2) The number of characteristics of the mediated presentation or experience that can be modified by the user;

3) The range or amount of change possible in each characteristic of the mediated presentation or experience;

4) The degree of correspondence between the type of user input and the type of medium response;

5) The speed with which the medium responds to user inputs.

From an interpersonal communication perspective, researchers use face-to-face communication as the standard of interactivity and evaluate the interactivity of mediated communication by how closely it simulates face-to-face communication. Burgoon et al. (1999) uses medium's "structural affordance" and "experiential properties" to conceptualize interactivity. They summarized following properties of affordance: participation, mediation, contingency, median and information richness, geographic propinquity, synchronicity, identification, parallelism and anthropomorphism. The experiential properties include interaction involvement, mutuality and individuation.

In this study, since we want to examine the real-time communications between customer and service staff conducted on the website, we mainly anchor our study from the interpersonal perspective. In other words, we will try to find out whether and to what extent the implementation of TTS voice and 3D avatar would make the user feel like she is having face-to-face communication with the customer service representative.

3. Research Framework

In the real-world shopping environment, as soon as a customer enters into a store, she is greeted by a salesperson. She may request information about the product she is interested in and the ordering process. She can even chat casually with the salesperson. The interaction between customer and salesperson comprises a significant portion of "store atmosphere". No matter how usable the website being designed is and how easy it is to follow the navigation structure, lots of people still need human-to-human interaction when shopping or just browsing, not to mention those new comers who have all kinds of concerns for credit card security, shipping charges and product specifications. In the study by Aberg and Shahmehri (2001), the customer reported experiencing better quality support, having more fun using the website, and a higher level of trust, when the website was implemented with human assistant aid. Moreover, this service not only attracts users who need help with a task that they cannot solve on their own, but is also of value for other purposes such as saving customers' time or effort.

This study examines the effects of TTS and the 3D avatar in the implementation of online "Live Help" from three perspectives: 1) the effects on the consumers' perception of presence, which is further classified into social presence and telepresence; 2) the effects on the consumers' involvedness during the communication, which is represented by flow and the frequency of interactions; 3) the effects on the customer's trust in the salesperson.

3.1 Dependent Variables

3.1.1 Social Presence

In Short et al. (1976), social presence is regarded as being a quality of the communication medium. They propose that social presence is a single dimension representing a cognitive synthesis of all the factors about communication as they are perceived by the individual to be present in the medium. The factors that contribute to the social presence of a communication medium include facial expression, direction of looking, posture, dress and other non-verbal vocal cues. However, the weights given to all these factors were determined by the individual. To measure social presence, they used a series of bipolar, seven-point semantic differential items including impersonal-personal, unsociable-sociable, insensitive-sensitive, and cold-warm. The social presence ranking thus depended on three independent factors: the interaction of the medium; the task at hand and the user's subjective judgment. Rice (1992) proposed in his "media richness" theory that the media richness or information richness could be measured less subjectively by examining a medium's "*capacity for immediate feedback, the number of senses involved, personalization and language variety*".

Lombard and Ditton (1997) summarized that presence as social richness is related to two concepts originally applied to nonmediated interpersonal communication: intimacy and immediacy. Intimacy behaviors include posture and arm position, trunk and body orientation, gestures, facial expression, body relaxation, touching, laughter,

speech duration, voice quality and others. A medium high in social presence allows the actors to adjust more of these variables and therefore more precisely adjust the overall level of intimacy. Immediacy or psychological closeness is more about choices of language, thus it can be varied within any medium that can transmit language, but Short (1976) still expects immediacy and social presence to be correlated.

3.1.2 Telepresence

Presence as transportation, which is also defined as telepresence, is another dependent variable in our study. Lombard and Ditton (1997) summarized that there are three distinct dimensions of transportation: 1) "You are there", in which the user is transported to another place; 2) "It is there", in which another place and the objects within it are transported to the user; 3) "We are together", in which two communicators are transported together to the same space. In this study, we focus on the first and third dimensions of telepresence.

Researchers use the phrase "you are there" or "being there" to describe how viewers experience the environment they encounter through a media. For example, Kim (1996) studies the feeling of telepresence with regard to television and defines telepresence as a "*feeling of being a part of the phenomenal environment created by television and not being a part of the physical environment surrounding the view and the television set*". The concept of transporting users to remote physical places is often

used in the study of virtual reality, for example, the immersive experience of a tour of art exhibits and museums on the multimedia platforms.

In this study, we adopted part of Kim's (1996) concept and measurement of telepresence, that is, the subjects in our study would be asked to report how often they had had the following perceptions: "I felt I was in the world the website created", "the website-generated world seems to me to be more like 'somewhere that I visited" rather than 'something that I saw'", and "my body was in this room, but my mind was inside the world created by the website". Our goal was to test whether the addition of TTS voice and the 3D avatar would generate a stronger feeling of "being in the virtual store".

Another dimension of telepresence we study is the users' perception of "We (User and the salesperson) are together". Such a definition of telepresence could be found in literature studying video conferencing as well as virtual reality. In Kim (1996), this was measured by asking the participants to report the degree to which they agreed or disagreed with statements such as "I felt we were in the same room" and "It felt like a face-to-face meeting". In this study, we also test whether and to what extent TTS voice and the 3D avatar would help creating illusion such that the user would feel as though they were having a face-to-face conversation with the customer service representative during the process of online shopping.

The third type of transportation, which is the perception that the object in another world is being transported to the user, would be more applicable in the study of customer-product interaction, or Virtual Product Experience, rather than in this study.

3.1.3 Flow

Flow is a concept which describes the state in which people are involved in certain activities (Csikszentmialyi 1975). To be more concrete, flow represents a subjective psychological experience that characterizes the human-computer experience as playful and exploratory (Webster et al 1993). Webster (1993) further decomposes flow into three dimensions: control, attention focus, and cognitive enjoyment. More importantly, they found that flow was positively correlated with perceived communication quantity and effectiveness.

Hoffman and Novak (1996) introduced the concept of flow into Internet marketing. They found that four antecedent conditions were necessary for the flow state to be experienced: (1) matched skill and challenge, (2) focused attention, (3) interactivity, and (4) telepresence. It was also proposed that higher flow would lead to better consumer learning, higher perceived behavior control, more exploratory behavior, more positive subjective experience, and stronger distortion in time perception. Agarwal and Karahanna (2000) examined the role of cognitive absorption in the use of the World Wide Web. Similar to the concept of flow, cognitive absorption is identified as being composed of five dimensions: temporal dissociation, focused immersion, heightened enjoyment, control, and curiosity. It was found that after controlling for self-efficacy perception, cognitive absorption had a significantly positive effect on perceived ease of use and perceived usefulness of information technology.

In this study, we use flow and frequency of interactions between customer and salesperson to measure the user's involvedness in the communication.

3.1.4 Trust

Trust is a very important concept in the research of shopping behavior (for a comprehensive review, see Xiao and Benbasat 2001). There are also various definitions of trust. In this research, we adopt Mayer et al.'s (1995) definition that trust is a "*willingness to be vulnerable, based on positive expectations about the action of others*". Prior research on customer trust in traditional commerce mainly focuses on interpersonal trust, especially customer trust in salespeople (Swan, Bowers et al. 1999). They propose that customer trust in the salesperson has two components: affect and cognition. "*Affect is the feeling secure or insecure about relying on the salesperson, and cognition is the belief that the salesperson has both the necessary competence and motivation to be relied upon*". Xiao and Benbasat (2001) further define customer trust as a five-entity, three-dimensional, and three-level concept. The three-dimensions refer to (1) cognitive trust in competence, (2) cognitive trust in goodwill, and (3) emotional trust. Specifically, they define emotional trust as "*customer's feelings about relying on an entity*".

Trust is also an important issue in the field of CSCW and CMC. For example, researchers found that interpersonal trust is likely to be affected by mediated communications. Previous research shows that it can be more difficult to develop trust in an online setting than face-to-face. Bos et al. (2002) studied the emergence of trust

in four different communication situations: face-to-face, video, audio, and text chat. They found that all three of the richer conditions were significant improvements over text chat. Video and audio conferencing group were almost as good as face-to-face.

Not many studies have been devoted to investigating communication media's effects on consumer's trust in the context of online shopping, and conflicting results are observed from these studies. For example, Greenspan et al. (2000) used a lab experiment to study prospective homebuyers who selected houses of interest using web, telephone, or PhoneChannel (Telephone plus TV). They found that telephone or PhoneChannel led to higher trust; on the other hand, Basso et al (2001) concluded that real-time interactivity, but not voice, increased judgments of friendliness and of the trustworthiness of the salesperson.

In this study, we investigate the effects of TTS voice and the 3D avatar on customer's cognitive trust in competence and emotional trust on the customer service representative.

3.2 Research Model

To the best of our knowledge, this paper is the first one to study virtual communication with both TTS and the 3D avatar in the context of online shopping. Thus, most of our research hypotheses are exploratory.

Previous research suggests that social presence is determined by three factors: the interaction of the medium; the task at hand and user's subjective judgment (Short, et al. 1976). In this study, we would focus on the interaction of different media. We examine whether the real-time communication implemented through a different communication channel (text chat, TTS, 3D avatar and the combinations of them) has significant effects on the perception of social presence.

As Steuer (1992) proposed, medium's vividness and interactivity would largely determine the sense of telepresence. Since either TTS voice or the 3D avatar would provide different sensory breadth and width from text-chat, we want to know their effects on the perception of telepresence as well.

Given that flow is highly correlated with perceived communication quality and effectiveness (Webster 1993), it is reasonable to hypothesize that different communication media may influence the perception of flow and involvedness. We would like to further test whether such a difference is significant enough to affect user's behavior, for example, the frequency of their interactions with the salesperson.

In addition, although the content of the communication may be the most important factor to judge the trustworthiness of a salesperson, there are still no clear indications

of the communication media's effects on the emergence of trust from the previous literature. We would like to reinvestigate whether the media expressiveness or medium characteristics would have significant effects on either cognitive trust or emotional trust.

The research model is shown in Figure 3-1:

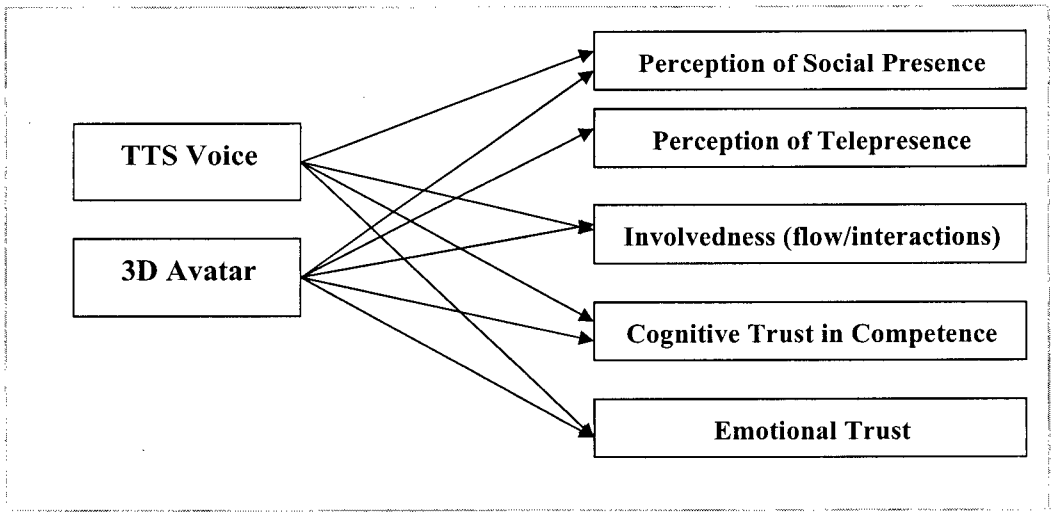


Figure 3-1 Research Model

3.3 Hypothesis Development

Currently, "Live Help" functions were only provided through textual chat, which is real-time and simple to implement, but narrow-bandwidth and monotonous. From the previous research work of social presence, it is projected that the higher level of vividness, which includes the number of sensory channels (Short et al. 1976) and the quality of sensory fidelity (Reeves and Nass., 1996), the higher the degree to which the senses are engaged, and the higher the level of social presence. The factors that may contribute to the social presence of a communication medium include facial expression, direction of looking, posture, dress and other non-verbal vocal cues (Short et al. 1976). The theory of social presence and information richness both assume that face-to-face communication and video telephones are inherently superior to telephone because they provide more channels of communication. Compared with face-to-face conversation, electronically mediated communication, for example, telephony, videotelephony, instant messaging and email provide less feedback to the speaker and fewer sensory cues. Thus, face-to-face is more "media rich" than audio-only or visual-only communication media. Of the two, audio is considered "richer" because it conveys cues about the speakers' meaning through inflection, pitch, tone and pause. With the addition of TTS voice and/or the 3D avatar, customers are being exposed to more audio-visual communication cues. Furthermore, when a user interacts with a human-like virtual character, talking and moving, in a vividly simulated audio-visual environment, more sensory cues will be involved and perceived. Applying this

analysis to the interface design of online customer service, we developed the first set of hypotheses:

H1a: Live Help function with TTS would generate a higher level of social presence than the one without TTS;

H1b: Live Help function with a 3D avatar would generate a higher level of social presence than the one without the avatar.

On the other hand, when a user interacts with a humanoid form avatar, the avatar is able to deliver nonverbal cues such as gazing, nodding and waving hand, which would more closely resemble to face-to-face communication. Literature concerning video conferencing as well as virtual reality uses the word "telepresence" as "the degree to which participants of a telemeeting get the impression of sharing space with interlocutors who are at a remote physical site" (Muhlbach et al. 1995). In Lombard and Ditton (1997), this was also explained as the feeling of "we are together". HCI researchers have also discussed and demonstrated some benefits of the avatar in one-on-one task settings such as taking an educational tutorial and looking at real estate. Cassell et al. (1999) further discussed an embodied conversation partner with the proper human verbal and nonverbal communication skills. Therefore, humanoid avatars for the communicator are an efficient design since it allows the use of everyday conventions such as gestures, postures and body language as part of the communication channels between communicators. This would significantly enhance the feeling of vividness and interactivity. Based on the above analysis, the second hypothesis we are proposing is:

H2: Live Help function with a 3D avatar would generate a higher telepresence than the one without an avatar.

As summarized in the previous section, the perception of flow represented a subjective psychological experience that characterizes the human-computer experience as playful and exploratory. We would consider that multimodal communications, which involved more sensory cues like voice and an animated 3D avatar, would generate higher sense of interactivity. As proposed by Hoffman and Novak (1996), one of the antecedents of flow is telepresence, which means the medium generating higher feeling of telepresence may induce higher level of flow too. We propose the third set of hypotheses as following:

H3a: Live Help function with TTS would induce a higher level of flow than the one without TTS;

H3b: Live Help function with a 3D avatar would induce a higher level of flow than the one without an avatar.

Another good indication of a user's involvedness in the communication is the frequency of interactions. We count the total interactions and the interactions initiated by the customer during a specific shopping task as a complementary measure of involvedness. Thus here comes the fourth set of hypotheses:

H4a: Live Help function with TTS would induce more interactions than the one without TTS;

H4b: Live Help function with a 3D avatar would induce more interactions than the one without an avatar.

As mentioned in the pervious chapter, there is much literatures examining several modes of communication and their relative effects on task performance and collaboration, but most of these studies are in the field of CSCW and in the video-conferencing domain. It has been proven that more immediate forms of communication (like face-to-face or video) are more effective in promoting cooperation than less immediate forms such as text chat. But would such examination of communication modality be applicable to the domain of electronic commerce, which relies on establishing a trusting relationship between strangers? Jensen et al. (2000) use the game of "Prisoner's Dilemma" to compare four types of communication conditions in promoting the development of trust and cooperation, including no communication, text chat, TTS, and voice. The conclusion is voice condition is significantly better than text-chat and no-communication condition. And the TTS group is marginally better than text-chat and no-communication. This suggests that current TTS technology is good enough to make a difference. We would also like to examine the additional modalities like the 3D avatar's effects. Therefore, the last set of our hypotheses are:

H5a: Live Help function with TTS would generate a higher level of trust (both cognitive trust and emotional trust) than the one without TTS;

H5b: Live Help function with a 3D avatar would generate a higher level of trust (both cognitive trust and emotional trust) than the one without avatar.

4. Research and Experimental Design

4.1 Overview

We used a controlled laboratory experiment to test the theory and proposed hypotheses. We choose this research method because it gives the researcher the greatest control over the manipulation of the independent variables (in this case, TTS voice and the 3D avatar). Extraneous variables are minimized or controlled in a laboratory setting. Thus, causality is more likely to be inferred using this method. In this study, we want to see whether the implementation of TTS Voice and 3D avatar in "Live Help" interface will have significant effects on user's perception of social presence, telepresence, involvedness and trust.

A mock-up e-commerce website integrated with six different interfaces of "Live Help" feature was developed to test and account for the two independent variables (TTS Voice and 3D avatar) and all combinations of their interactions. To further explore the potential interaction effects between text and voice, a 3 by 2 factorial design (three levels for TTS and two levels for 3D avatar) is adopted: the voice dimension is divided into three levels: text only, voice only, and text plus voice; the avatar dimension can be present or absent (i.e. two levels). Subjects were then randomly assigned to one of six interfaces. They were given specific instructions to complete a set of shopping task with the "Live Help" provided by a salesperson. A questionnaire was then administered after the subject completed all of the tasks.

Details of the design and implementation of these procedures will be expanded in subsequent sections. The six types of Live Help interface are shown in Table 4-1.

Table 4-1 Six Types of Live Help Interface

Condition	***Live Help Interface Features***
1.	No Avatar + Text Only
2.	No Avatar + Voice Only
3.	No Avatar + Text + Voice
4.	3D Avatar + Text Only
5.	3D Avatar + Voice Only
6.	3D Avatar + Text + Voice

4.2 The Interface

The experimental interface, which consists of the shopping site interface and "Live Help" interface, was developed over a period of 2 months. Prototypes were developed and revised over this period. A computer programmer was hired to assist in the implementation process. Depending on the condition being assigned, subject would experience part or all of four components of the interface, the shopping site, text chat, TTS Voice, and 3D avatar.

4.2.1 The Shopping Site

We developed a mock up shopping website with HTML (Hyper Text Markup Language), ASP (Active Server Page), and JavaScript. The look and feel of the interface was designed to resemble actual EC web sites as closely as possible. The website is selling digital cameras, camera memories and photo printers. The outlook of the website is shown in Figure 4-1.

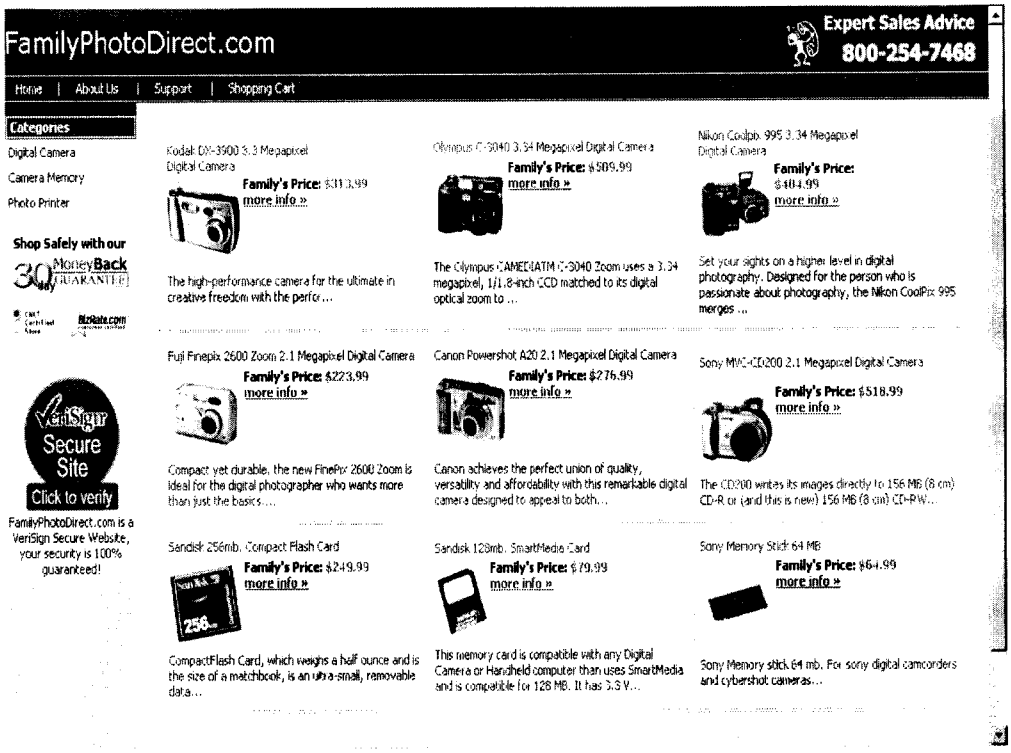


Figure 4-1 Homepage of Experimental Website

As shown in Figure 4-1, on the left side of the homepage, a list of product categories is displayed along with typical components on an EC website. Hyperlinks of the product category lead to sub-categories where products are further classified. This hierarchy continues until a specific product page is displayed. Three related products (digital camera, camera memory and photo printer) were used because we wanted to ensure there would be enough interactions taking place between the subject and salesperson, since most participants would ask the salesperson some questions about the compatibility and specifications of these products. These products were chosen due to their popularity in the B2C EC industry and among university students, the major source of our experiment samples.

4.2.2 Implementation of Live Help

With the use of "frame" tags of HTML, we are able to integrate the shopping site with the "Live Help" interface, which is implemented separately while giving the subject an illusion that the "Live Help" feature is part of the shopping site.

4.2.2.1 Blaxxun Contact

We used free software called Blaxxun Contact and its online virtual space to realize text chat and implement 3D avatar. Blaxxun Contact is a Virtual Reality Markup Language (VRML) plug-in for designing virtual community. The Blaxxun platform provides several ways of communication which could be easily integrated with the experimental website. In this study, it was used in every condition for text chat and viewing and controlling of a humanoid 3D avatar. We also revised several other settings to further customize the "Live Help" functions, such as chat macro and body language.

Chat macros are predefined texts that can be chosen from a list. Answer to some frequently asked questions were prepared in advance in order to reduce delay. Actually, this is already a standard function in most commercialized online customer support systems.

Body language is the most important characteristics of a 3D avatar. Blaxxun platform enables its avatars with several body movements, including waving hand, nodding, shaking head, frown and others to make the communication more interactive. All these body movements are activated through some key words in the

text, that is, the content of the text chat message which, in our case, refers to the answers and advices provided by the salesperson. We prepared those frequently used emotion-conveying words and matched them with a specific body movement.

4.2.2.2 AT&T Natural Voices and AT&T Instant Messenger

AT&T Natural Voices is the best TTS engine we could obtain in the commercial market. As a standard Speech Application Program Interface (SAPI) compliant component, it could be integrated with other software to "read" the designated text. The only obstacle we were facing is that the Blaxxun platform and AT&T TTS engine were not compatible, thus we employ another piece of software, the AT&T Instant Messenger which is running on the background, to duplicate the messaging texts and read them aloud.

4.3 Experiment Settings

Two research assistants were recruited for the experiment. One would stay with the subject to walk her through the whole process; the other would stay in the other room and use another networked computer playing the role of customer service representative. The setting is shown in Figure 4-2:

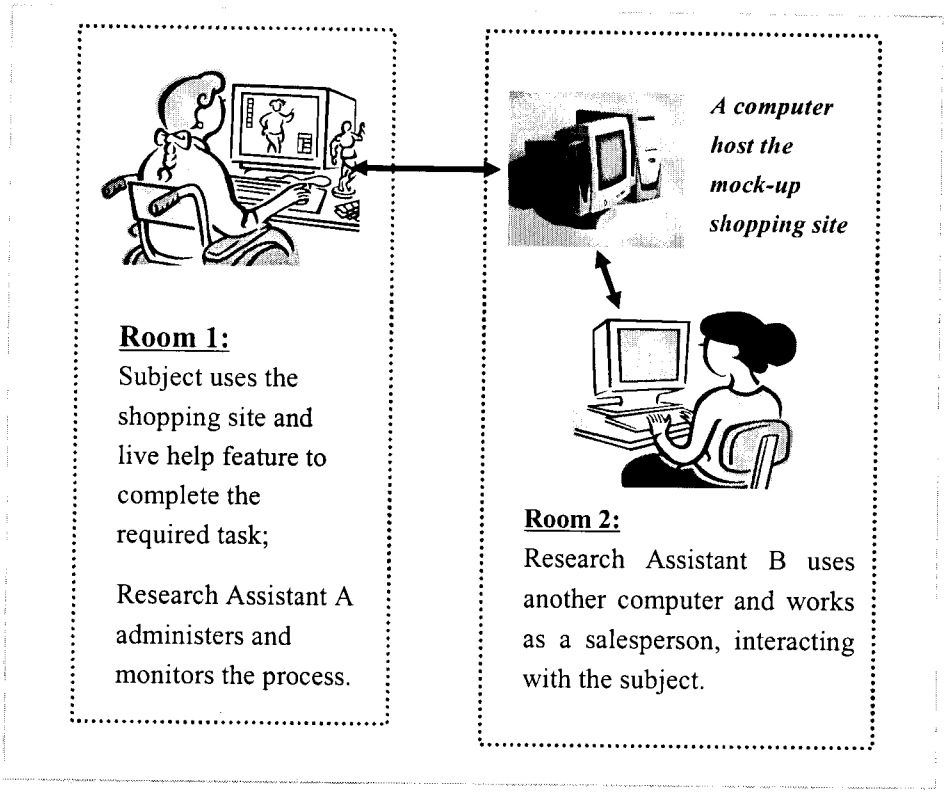


Figure 4-2 Experiment Settings

4.3.1 Condition 1: No Avatar + Text Only

Under this condition, on the computer screen (19 inch monitor with 1024*768 resolution) subject will see the shopping site on the left two thirds of the screen and a chat window, which would be used for subject interacting with salesperson through text messaging, on the right side. The TTS function would be turned off. The only

communication channel between the subject and customer service representative is textual chat. Whenever the salesperson gives a response or initiate a conversation, subject would hear a beep as notification.

The interface is shown in Figure 4-3.

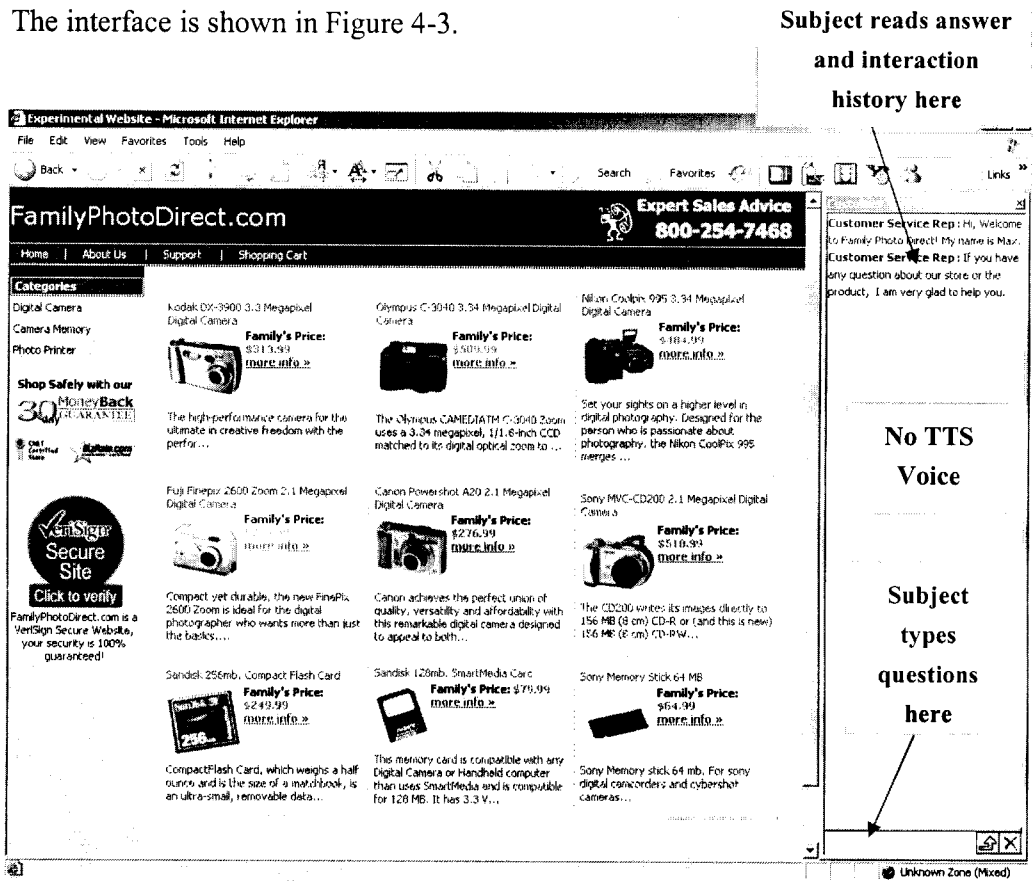


Figure 4-3 Interface of Condition 1

4.3.2 Condition 2: No Avatar + Voice Only

Under this condition, subject will still see the shopping site on the left two thirds of the screen; on the right side, subject would use a similar text window to ask questions or give comments, and check her own messaging history. But she will get

the answer or any other feedback from the salesperson through TTS voice generated by AT&T Instant Messenger running on the background. In other words, she will "listen" to the answer through computer-generated voice rather than "read" them on screen.

The interface is shown in Figure 4-4.

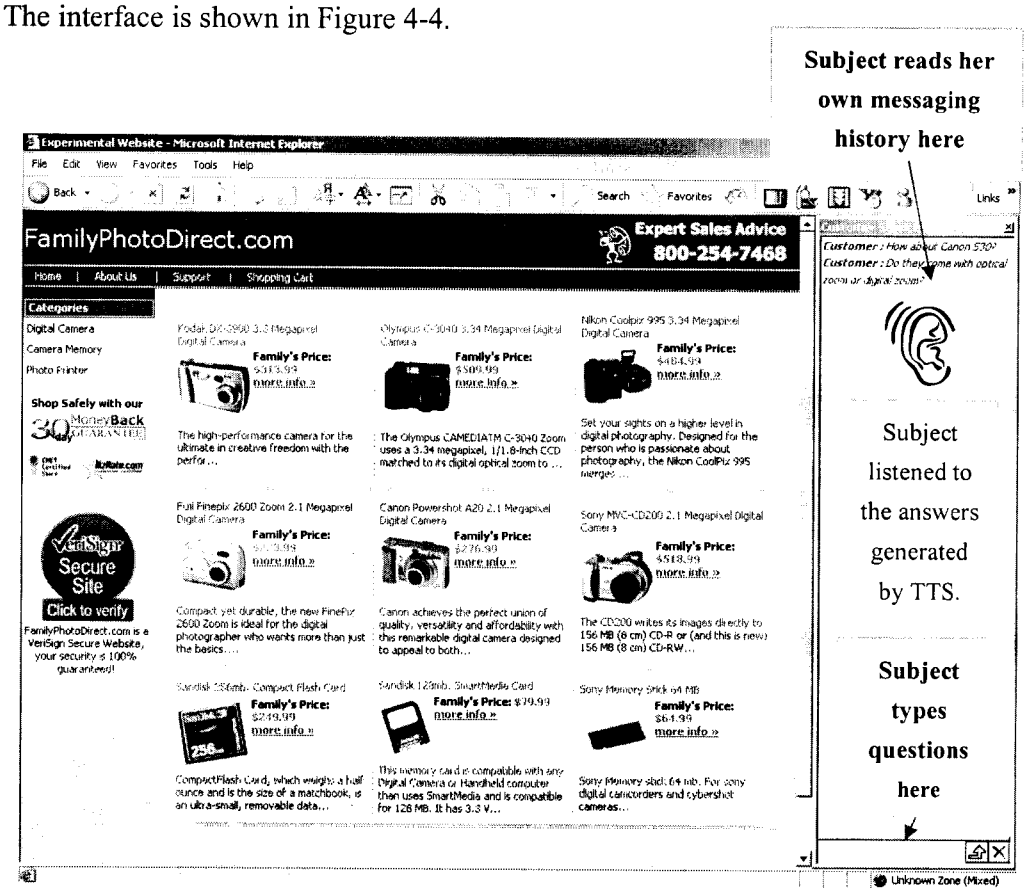


Figure 4-4 Interface of Condition 2

4.3.3 Condition 3: No Avatar + Voice + Text

The interface is quite similar to that of Condition 2. The only difference is that subject could read answers and messaging history of both interactants on the screen while at the same time listen to what the salesperson input through TTS voice. The text and voice would be synchronized.

The interface is shown in Figure 4-5.

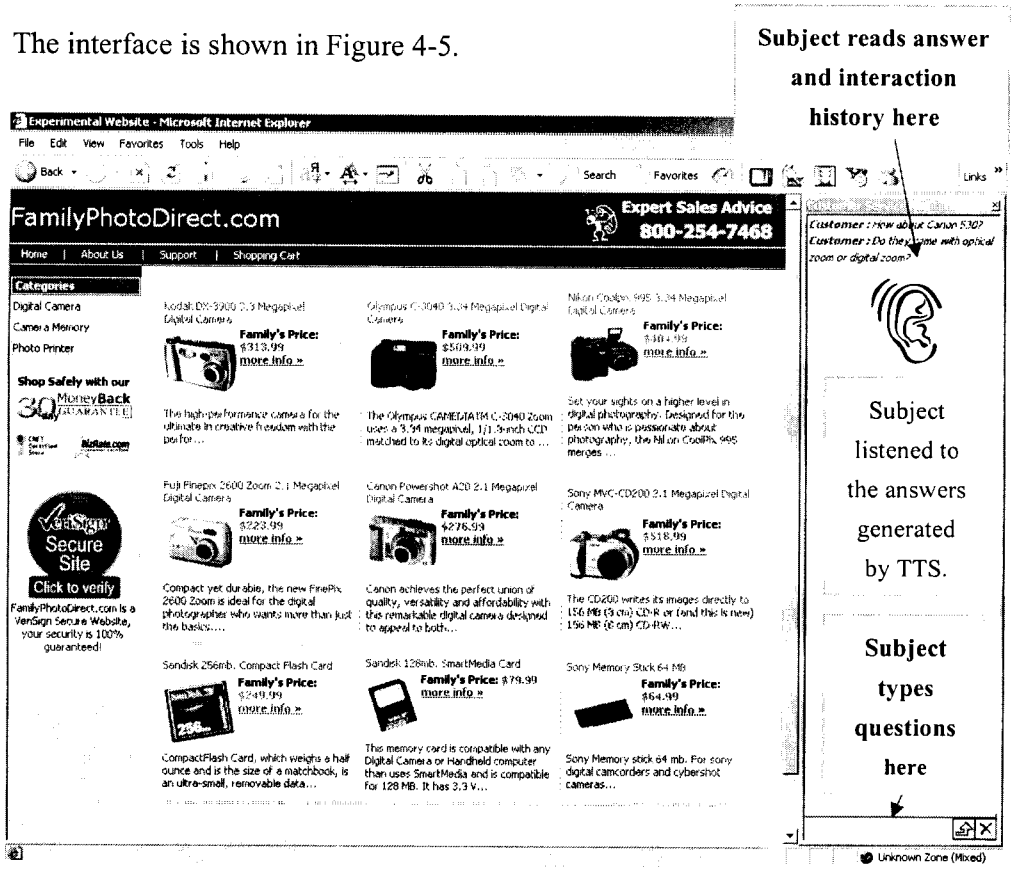


Figure 4-5 Interface of Condition 3

4.3.4 Condition 4: Avatar + Text

Under this condition, in addition to the shopping site and text chat window, subject see on the right upper corner of the screen a 3D avatar – the embodiment of the customer service representative she is interacting with. Subject could still input questions and comments on the right bottom corner and read the answer and interaction history on the screen, as well as seeing the avatar move his body accompanying with the text message. TTS function would be turned off under this condition. In addition to viewing body movement of the avatar, subject would also

hear a beep as notification whenever the salesperson gives a reply or initiate a conversation.

The interface is shown in Figure 4-6.

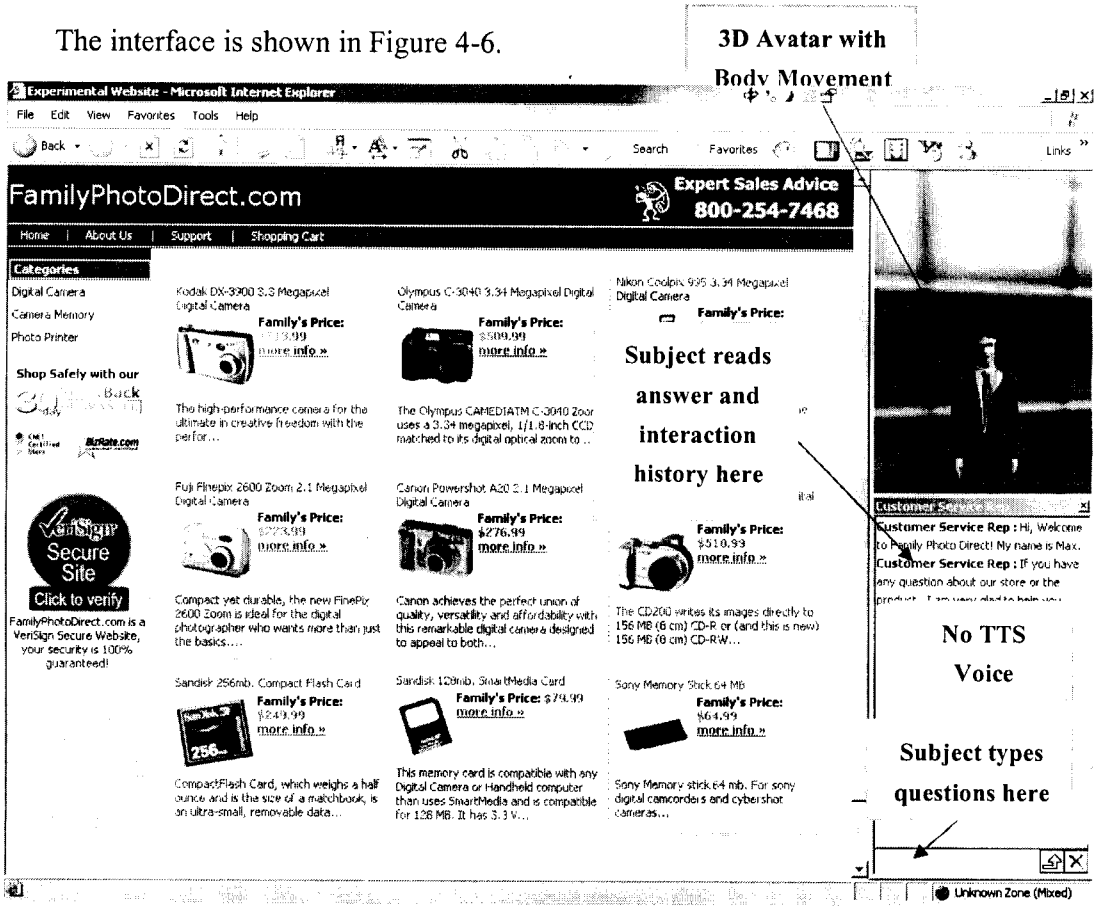


Figure 4-6 Interface of Condition 4

4.3.5 Condition 5: Avatar + Voice

Under this condition, subject could get the information provided by the salesperson through the voice generated by TTS, while at the same time view the body movement of the avatar. Body language of the avatar would be synchronized with the voice. On the chat window, subject could only read her own messaging

history. The only channel subject could get information from the customer service representative is through TTS voice.

The interface is shown in Figure 4-7.

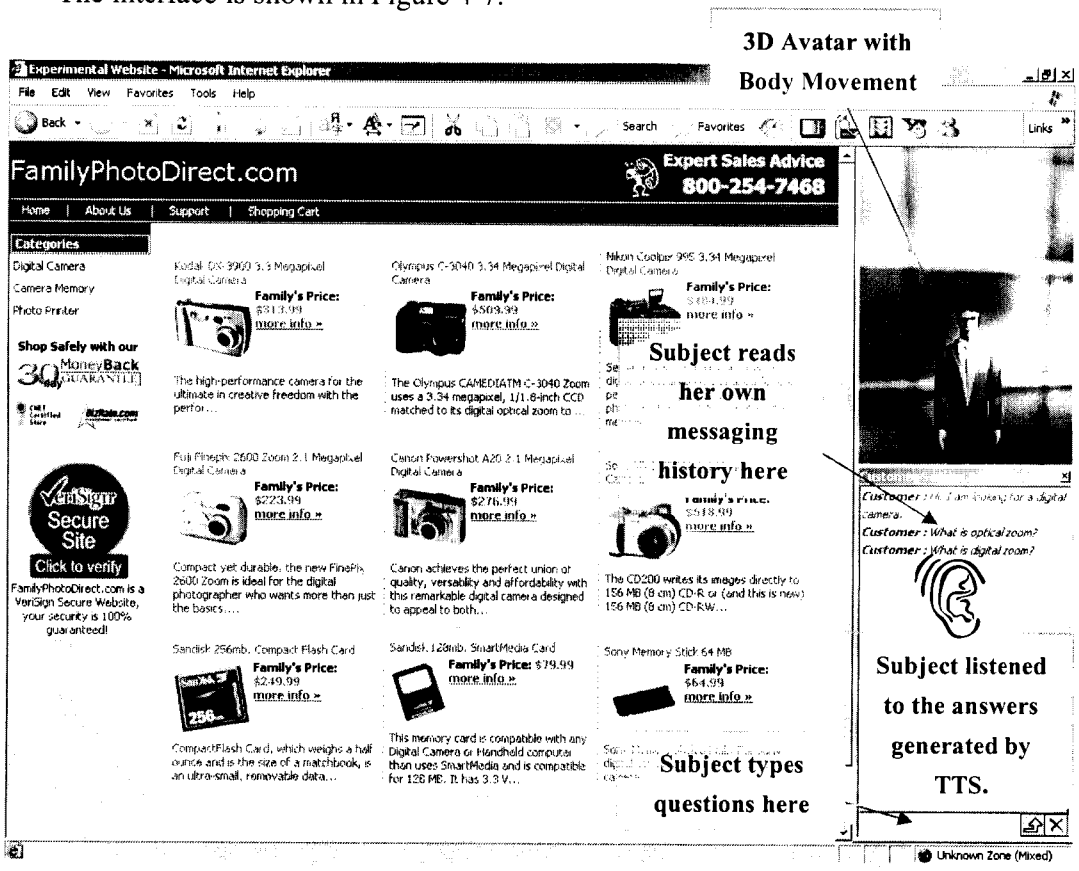


Figure 4-7 Interface of Condition 5

4.3.6 Condition 6: Avatar + Voice + Text

Under this condition, the interface would be almost the same as that of condition 5, except that subject would see avatar body movement, read the answer on the screen and hear the corresponding voice generated by TTS all at the same time. Avatar's body movement, voice and text would be synchronized.

The interface is shown in Figure 4-8.

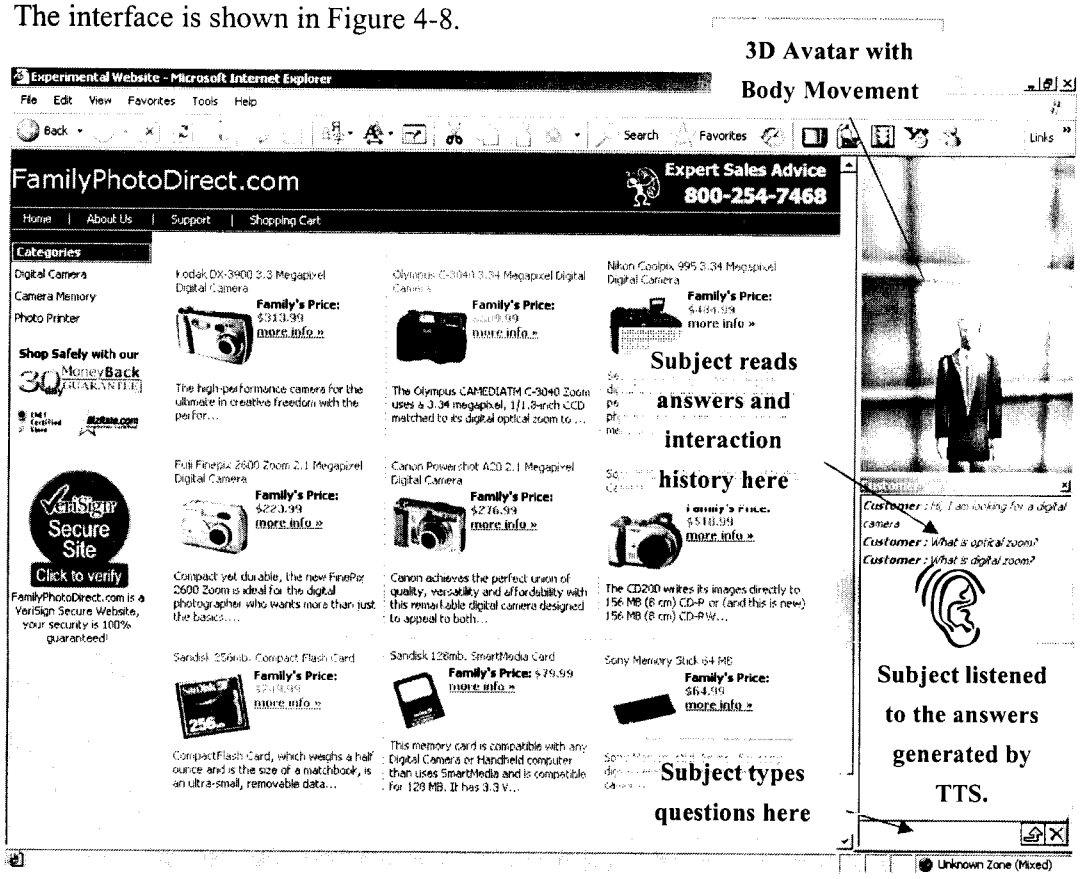


Figure 4-8 Interface of Condition 6

4.4 Measurement

Two types of measurement are used in this study. The first one is a self-report questionnaire, which is required to be filled out after subject completed the shopping task; the other is the chat log recording all of the interactions happened between the subject and customer service representative.

4.4.1 The Questionnaire

In our questionnaire, five dependent variables would be measured by Likert-scale questions, including Social Presence (7 questions), Telepresence (9 questions), Flow (10 questions), Cognitive Trust in Competence (5 questions) and emotional trust (5 questions). The complete questionnaire is enclosed in Appendix A.

4.4.1.1 Measurement of Social Presence

The measurements of social presence used in this study are adapted from the scale used by Short et al. (1976). The adapted version had been validated in other studies (for example, Burke and Chidambaram, 1999). The items include participants' evaluations of the extent to which they feel the communication be personal, expressive and emotional. Seven semantic pairs would be used, include: impersonal/personal, warm/cold, distant/close, dehumanizing/humanizing, expressive/inexpressive, emotional/unemotional, and insensitive/sensitive. The participants would be asked to choose one position of 7 between every contrasting pair of words. For example,

Impersonal * * * * * * * **Personal**

4.4.1.2 Measurement of Telepresence

The second part of the questionnaire is to measure participant's experience of telepresence, which means to what extent the subject feel: 1) she was actually "being in the virtual store" and 2) to what extent she and the customer service representative are "being together". Based on the definition of telepresence, Kim and Biocca (1997) developed a scale of self-reported telepresence for the television viewing situation. Our measurements were created by adapting those scales into online shopping context. The measurement consists of 9 Likert-scale items.

4.4.1.3 Measurement of Flow

Webster et al. (1993) studied the dimensionality of flow. Their research indicates that flow has three dimensions: control, attention focus, and cognitive enjoyment. Because this measurement has enjoyed long-term credibility and been widely used in IS research, we use them to measure flow after some minor changes in wording of some items in order to adapt them to the experimental context.

4.4.1.4 Measurement of Cognitive Trust and Emotional Trust

According to Xiao and Benbasat (2001), Cognitive trust in a trustee's competence is defined as a trustor's rational assessment that a trustee will have the ability to fulfill its obligations as understood by the trustor; Emotional trust is defined as a trustor's feeling secure and comfortable about relying on a trustee. Based on card sorting and

factor analysis, they develop a complete set of measurement for both cognitive trust in competence (5 items) and emotional trust (5 items). Although in the study by Xiao and Benbasat (2001), these definitions and measurement are used towards user's judgments of recommendation agent, we think they are also applicable in evaluating user's feeling of trust toward a real person – the customer service representative answering questions and providing advices. Therefore, we adopted their definition and corresponding measurement.

4.4.2 Chat Log Analysis

To further investigate user's involvedness in the online interaction with the customer service representative, we analyzed the chat logs which record all of the verbal communication between the two parties. From the chat log, we counted in each session (the complete process for a subject to finish her shopping task) the total number of interactions happened and the number of interactions that were initiated by the subject.

4.4.3 Reliability and Validity of Measurement

Since most of the measurements used in this study are borrowed from current IS and communication literature, their reliability and validity had been proved in previous research works. What's more, we used personal interview to locate and correct weaknesses in the questionnaire instrument. Interviews were conducted during the pilot testing periods with a select number of peers (MIS graduate students). Their feedbacks were used to improve the scales and their presentation. Comments from

these interviews were also used to improve the experimental procedures. (Changes made as a result of the feedback are described in section 4.5 – Experimental Procedures).

4.5 Experimental Procedure

Pilot tests were conducted for this study before actual data collection took place. These procedures took place over a course of two months. The experiment went through one major change before it was used for actual data collection. The details of these developments are described in this section.

4.5.1 Pilot Test and Experimental Revision

Fifteen graduate and undergraduate students were invited to participate in the pilot test and paid $10 as a reward. By examining their answers to our questions and their verbal report, we were able to identify some problems with our initial experimental design, including those on website and on questionnaire. As a result, we revised our design.

To avoid the problem of lacking referencing point, which was reported in Jiang and Benbasat (2001), our original design in the pilot test was to let subject complete two sets of shopping tasks: purchase a digital camera and its accessories (camera memories and photo printers) and a Personal Digital Assistant along with some accessories (PDA memories and other optional parts) as a gift for a friend. The subject was required to complete one task under the base condition (condition 1: no avatar + text only) and the other under treatment condition (randomly selected from all other 5 conditions). In the questionnaire, they were asked to compare the two interfaces they had used and give their preference on one of them for all the questions. The rationale behind such design is to let subject experience not only the interface

corresponding to his/her group but also the basic textual chat interface so that even for the user who never used "Live Help" features before (which is true among most of our participants), there would be lower possibility of ceiling effects. But during the pilot test, we found there are other significant confounding factors existing in the design. The first one is that subject's incentives toward the two products (digital camera and PDA) are significantly different. Actually, most of the participants have stronger interests in digital camera. Although we have balanced the order of the shopping for a specific product, it is found that subjects asked fewer questions for the PDA task, which would have big impacts on their judgment of the communication quality. The other concern is that as the time of completing two tasks is longer than a hour, some subjects showed some signs of fatigue by the end of the second shopping task and thus influence the measurement of involvedness.

Based on the above observation and analysis, we revised the research design and corresponding questionnaires. In the revised design, we adopted a pure between-group design, that is, the subject is asked to complete only one shopping task --- purchase a digital camera and some camera accessories --- on the shopping site integrated with one "Live Help" interface randomly selected from the six conditions. Questionnaire from the initial design stage was thus re-designed to be in line with the change (see Appendix II).

4.5.2 Formal Experiment

After the pilot tests were completed, formal experimental data collection took place over a one-month period. Altogether 72 student subjects participated in the study. All of them were recruited by campus advertisement. Their reward was $10 plus one-in-ten chance to win a gift certificate valued at 50% of the price of the selected camera (up to $200). Subjects were randomly assigned to the six interface groups. During the experiment, subjects were asked to behave as if they were shopping for a digital camera and accompanying accessories as a gift for friend. Time was not limited but was normally around 40-60 minutes. The final experimental procedure was designed as follows:

Step 1 (training): A subject was randomly assigned to one of the six interface groups. After reading the information sheet and understand the shopping task, he/she would be trained by a research assistant on how to use and navigate in their assigned web interface and how to use the "Live Help" interface. The subject did not proceed to the second stage until he/she understood how to use both of them.

Step 2 (main test): The subject was asked to examine and choose a digital camera, camera memory and a photo printer. They are encouraged to use the "Live Help" function with the type of interface corresponding to his/her group and interact with the customer service representative if they have any questions about the website or the products. After selecting the preferred products, the subject was not required to go through the checkout process but only put the selected products into shopping cart. A

research assistant was there to observe whether the subject used the "Live Help" feature.

Step 3 (Questionnaire): After the shopping task being completed, subject was asked to fill out a questionnaire with all of the measurements on the aforementioned dependent variables: social presence, telepresence, flow, cognitive trust and emotional trust.

Step 4: Debriefing and dismissal.

5. Data Analysis and Discussion

This chapter describes the processes that were undertaken in analyzing the data collected from the subjects' interactions with different "Live Help" interface. An overview of the analytical process is provided, briefly explaining the process and the reasoning behind the statistical tests chosen. Then, data from the different statistical tests is presented. Finally, the significance of the findings is disclosed with a detailed discussion.

5.1 Data Analysis Overview

The data were checked for their integrity and completeness before statistical tests were performed. All the questionnaires were checked for missing data and none had been found.

Next, several statistical tests were conducted to validate the measurement of dependent variables. Means for the dependent variables were first calculated according to the six treatment conditions. Correlations between different dependent variables were then calculated. Correlation between the variables *spmean* (mean of social presence), *tpmean* (mean of telepresence), *flmean* (mean of flow), *etmean* (mean of emotional trust), *ctmean* (mean of cognitive trust in competence), *interactions* (times of interactions initiated by subject) were presented in Table 5-1. Means and correlations are useful tests for the initial analysis of the data. Means are particularly convenient for revealing initial differences among dependent variables between treatment conditions. Correlations are effective for determining whether two

variables are closely related to each other. This is useful for relationships between independent and dependent variables. As well, we can discover whether dependent variables are closely tied to one another. This is helpful in eliminating unnecessary variables.

Table 5-1 Correlations between Dependent Variables

		SPMEAN	TPMEAN	FLMEAN	ETMEAN	CTMEAN	Interactions
SPMEAN	Pearson Correlation	1	.468**	.573**	.521**	.548**	.015
TPMEAN	Pearson Correlation	.468**	1	.540**	.406**	.315**	.145
FLMEAN	Pearson Correlation	.573**	.540**	1	.479**	.576**	.192
ETMEAN	Pearson Correlation	.521**	.406**	.479**	1	.754**	-.049
CTMEAN	Pearson Correlation	.548**	.315**	.576**	.754**	1	-.080
Interactions	Pearson Correlation	.015	.145	.192	-.049	-.080	1

** Correlation is significant at the 0.01 level (2-tailed).

The analysis of correlation shows that although the correlation between most dependent variables are significant, only one of them, the correlation between emotional trust and cognitive trust ($r = 0.754$), is high enough to alert us that these two dependent variables may actually collapse into one.

To further test these two dependent variables, we combined the 10 measurements of both variables and conduct an exploratory factor analysis. Results are shown in Table 5-2.

Table 5-2 Factor Analysis of Trust Measurement

	Factor	
	1	2
Emotional Trust:secure	.806	.334
Emotional Trust:comfortable	.798	.346
Emotional Trust:happier	.826	.204
Emotional Trust:confident	.858	.308
Emotional Trust:content	.750	.386
Competence Trust: expert	.490	.706
Competence Trust: competent	.197	.794
Competence Trust: expertise	.605	.561
Competence Trust: ability	.394	.768
Competence Trust: knowledge	.269	.848

Factor analysis showed that all emotional trust measurements are loaded into the first factor, most cognitive trust factor are loaded into second factor except for the third measure (expertise) which loads almost equally on both factors. According to Hair, et al. (1998), loadings of greater than + .30 are considered to meet the minimal level; loadings of + .40 are considered more important; and if the loadings are + .50 or greater, they are considered practically significant. We therefore believe that emotional trust and cognitive trust are actually two separate dependent variables.

Next, we calculated reliability of these measurements. The value of Cronbach Alpha was presented in Table 5-3. Results show that the Cronbach Alpha of all of our dependent variables is over 0.7, which suggests the measurements are valid.

Table 5-3 Reliability Analysis

Sample Size: 72	
Dependent Variables	**Cronbach Alpha (a)**
Social Presence (7 items)	0.7823
Telepresence (9 items)	0.7471
Flow (10 items)	0.8570
Cognitive Trust (5 items)	0.8869
Emotional Trust (5 items)	0.9185

For these dependent variables, a Multivariate Analysis of Variance (MANOVA) test was chosen to analyze the aggregate effects of TTS voice and 3D avatar according to treatment conditions. It is essentially an F test in which an estimate of the between-groups variance is compared with an estimate of the within-groups variance. As recalled in chapter 4, this study is based on a 3 x 2 factorial design (3 levels in the dimension of Voice, 2 levels in the dimension of Avatar). MANOVA allows us to assess two variables' independent and interaction effects at the same time. This is also a test used to examine multiple dependent variables according to treatment conditions. It is a more sensitive statistical test over ones which investigate just one factor at a time mainly due to its ability to reduce overestimating of error variance, which is commonly the case in one-factor analyses.

We also wanted to examine the individual effects of TTS voice and 3D avatar on the dependent variables. MANOVA test could also provide univariate tests of between-subjects effects. This is very similar to MANOVA. The only difference is that only one dependent variable is analyzed each time.

Since our design has 3 different treatments on the factor of TTS voice ("text only", "voice only", and "text + voice"), a Scheffe test would be conducted to further examine all possible linear combinations of the difference in group means (that is, the difference on all those dependent variables between different treatment of TTS voice), not just pairwise comparisons.

5.2 Test Results

The statistical analysis results from this study are presented here. A significance level of 0.05 was used throughout the course of statistical analysis. In addition, the observe power levels are also calculated. This is useful in giving insight into cases where significance is not found.

5.2.1 MANOVA Results

An MANOVA test was fist run on all of the dependent variables based on treatment conditions. The results are shown in Table 5-4.

Table 5-4 Multivariate MANOVA Results

Effect	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	735.379	6.000	61.000	.000	.986	4412.274	1.000
AVATAR	**2.113**	**6.000**	**61.000**	**.065**	**.172**	**12.675**	**.714**
VOICE	1.131	12.000	124.000	.341	.099	13.576	.624
AVATAR * VOICE	.710	12.000	124.000	.739	.064	8.523	.394

a Computed using alpha = .05

b Exact statistic

c The statistic is an upper bound on F that yields a lower bound on the significance level.

d Design: Intercept+AVATAR+VOICE+AVATAR * VOICE

As seen here, only avatar control is shown to be marginally significant. However, since there are as many as 6 dependent variables in this study, the aggregate analysis does not indicate clearly TTS voice and avatar's effects on each dependent variable

individually. Therefore, we decided to run univariate test on each dependent variable separately.

5.2.2 Social Presence

An univariate test was run on the variable *spmean* (the average of the 7 social presence items) based on treatment conditions. The results are shown in Table 5-5.

Table 5-5 Univariate MANOVA Results of Social Presence

Tests of Between-Subjects Effects
Dependent Variable: SPMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	2.121	5	.424	.681	.639	.232
Intercept	1809.971	1	1809.971	2906.269	.000	1.000
AVATAR	2.281E-02	1	2.281E-02	.037	.849	.054
VOICE	.313	2	.156	.251	.779	.088
AVATAR * VOICE	1.786	2	.893	1.434	.246	.297
Error	41.104	66	.623			
Total	1853.196	72				
Corrected Total	43.225	71				

a Computed using alpha = .05
b R Squared = .049 (Adjusted R Squared = -.023)

This test revealed that there was no significant difference among six treatment groups, that is, TTS and 3D avatar had no significant effects on the perception of social presence. The first set of our hypothesis is not supported. A detailed discussion is provided in the next section of this chapter.

5.2.3 Telepresence

An univariate test was also run on the variable *tpmean* (the average of the 10 telepresence items) based on treatment conditions. The results are shown in Table 5-6.

Table 5-6 Univariate MANOVA Results of Telepresence

Tests of Between-Subjects Effects

Dependent Variable: TPMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	4.080	5	.816	1.145	.346	.382
Intercept	1365.544	1	1365.544	1916.311	.000	1.000
AVATAR	**3.557**	**1**	**3.557**	**4.992**	**.029**	**.595**
VOICE	.163	2	8.165E-02	.115	.892	.067
AVATAR * VOICE	.360	2	.180	.252	.778	.088
Error	47.031	66	.713			
Total	1416.655	72				
Corrected Total	51.111	71				

a Computed using alpha = .05

b R Squared = .080 (Adjusted R Squared = .010)

This test revealed that avatar had significant effects on subject's perception of telepresence, that is, subject reported significant stronger feeling of telepresence when a 3D avatar was integrated into the "Live Help" interface. The other interesting finding is that although the results of both univariate test ($p = 0.892$) and Scheffe test (*Sig.* = *0.822*) showed that TTS voice's effect is not significant, the descriptive statistics showed that when avatar is not present, there is almost no difference among three treatments of voice control (text only, voice only, text + voice); when avatar is implemented, subjects in the "voice only" group reported higher score of telepresence

than those of "text only" group or "text + voice" group. Altogether, our second hypothesis: *Live Help function with 3D avatar would generate higher telepresence than the one without avatar* is fully supported by the experimental data.

5.2.4 Involvedness

We used two different measurements to test subject's involvedness under different treatment of "Live Help" interface. The first measurement is flow, which was adopted and adapted from Webster et al. (1993); the other measurement is interaction frequency, which is used to represent how often and how likely the subject would like to use the "Live Help" interface to communicate with a customer service representative. This measurement was acquired by analyzing the chat log and counting the "chat blocks", especially the interactions initiated by the subject.

5.2.4.1 Flow

An univariate test was also run on the variable *flmean* (the average of the 10 flow items) based on treatment conditions. The results are shown in Table 5-7.

Table 5-7 Univariate MANOVA Results of Flow

Tests of Between-Subjects Effects

Dependent Variable: FLMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	5.813	5	1.163	2.186	.066	.681
Intercept	2148.401	1	2148.401	4040.208	.000	1.000
AVATAR	1.201	1	1.201	2.259	.138	.316
VOICE	**3.591**	**2**	**1.795**	**3.376**	**.040**	**.617**
AVATAR * VOICE	1.021	2	.510	.960	.388	.210
Error	35.096	66	.532			
Total	2189.310	72				
Corrected Total	40.909	71				

a Computed using alpha = .05

b R Squared = .142 (Adjusted R Squared = .077)

This test revealed that TTS voice had significant effects on subject's perception of flow, that is, subject reported significant stronger feeling of flow when a TTS Voice was integrated into the "Live Help" interface. Again, the surprising finding is that subject reported highest score of flow in the "Voice Only" group, no matter avatar being present or not; "voice + text" group scored even lower than "text only" group when the avatar is not present and almost the same when the avatar is present. The Scheffe test further revealed that there is marginally significant difference (*Sig. = 0.064*) on the measurement of flow between the "Voice Only" group and "Text + Voice" group. The descriptive statistics also showed that avatar group reported higher average score of flow, although the difference is not significant ($F = 2.26$, $p = 0.14$).

Our third set of hypothesis is partially supported, which is *H3a: Live Help function with TTS would induce higher level of flow than the one without TTS.* A detailed discussion is provided in the next section of this chapter.

5.2.4.2 Interaction Frequency

A univariate test was run on the variable *interactions* (the number of chat blocks initiated by the subject in the chat log of each session) based on treatment conditions. The chunking of chat log is conducted from the perspective of linguistics, that is, the meaning and context of the communication. The results are shown in Table 5-8.

Table 5-8 Univariate MANOVA Results of Interaction Frequency

Tests of Between-Subjects Effects

Dependent Variable: Interactions Initiated by Subject

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	369.736	5	73.947	2.009	.089	.637
Intercept	8602.347	1	8602.347	233.748	.000	1.000
AVATAR	**292.014**	**1**	**292.014**	**7.935**	**.006**	**.793**
VOICE	42.528	2	21.264	.578	.564	.142
AVATAR * VOICE	35.194	2	17.597	.478	.622	.125
Error	2428.917	66	36.802			
Total	11401.000	72				
Corrected Total	2798.653	71				

a Computed using alpha = .05

b R Squared = .132 (Adjusted R Squared = .066)

This test revealed that subjects in the avatar group initiated significantly more interactions than those of non-avatar group. In other words, when a 3D avatar is

present, customer is more likely to communicate with the customer service representative. Our forth set of hypothesis is partially supported, which is *H4b: Live Help function with 3D avatar would induce more interactions than the one without avatar.* A detailed discussion is provided in the next section of this chapter.

5.2.5 Trust

An univariate test was also run on the variable *ctmean* (the average of the 5 competence trust items) and *etmean* ((the average of the 5 competence trust items) based on treatment conditions. The results are shown in Table 5-9 and Table 5-10.

Table 5-9 Univariate MANOVA Results of Cognitive Trust

Tests of Between-Subjects Effects

Dependent Variable: CTMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	8.858	5	1.772	1.747	.136	.567
Intercept	2195.636	1	2195.636	2164.588	.000	1.000
AVATAR	.269	1	.269	.265	.608	.080
VOICE	**6.248**	**2**	**3.124**	**3.080**	**.053**	**.575**
AVATAR * VOICE	2.341	2	1.171	1.154	.322	.245
Error	66.947	66	1.014			
Total	2271.440	72				
Corrected Total	75.804	71				

a Computed using alpha = .05

b R Squared = .117 (Adjusted R Squared = .050)

Table 5-10 Univariate MANOVA Results of Emotional Trust

Tests of Between-Subjects Effects

Dependent Variable: ETMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power
Corrected Model	7.318	5	1.464	1.122	.357	.375
Intercept	1816.036	1	1816.036	1392.299	.000	1.000
AVATAR	.320	1	.320	.245	.622	.078
VOICE	**6.988**	**2**	**3.494**	**2.679**	**.076**	**.514**
AVATAR * VOICE	1.000E-02	2	5.000E-03	.004	.996	.051
Error	86.087	66	1.304			
Total	1909.440	72				
Corrected Total	93.404	71				

a Computed using alpha = .05

b R Squared = .078 (Adjusted R Squared = .009)

This test revealed that TTS had marginally significant effects on both cognitive trust in competence and emotional trust. In other words, subjects tend to have more trust in the serviceperson's competence and also more likely to trust that person emotionally when the "Live Help" interface is implemented with a TTS voice. Consistent with other results, "Voice Only" group is the highest one within different voice treatments while there is no noteworthy difference between "Text Only" group and "Text + Voice" group. The Scheffe test further revealed that there is marginally significant difference (*Sig. = 0.073*) on the measurement of cognitive trust between the "Voice Only" group and "Text Only" group and also marginally significant difference (*Sig. = 0.098*) on the measurement of emotional trust between the "Voice

Only" group and "Text + Voice" group. There is no significant difference between the avatar group and non-avatar group.

Our fifth set of hypothesis is partially supported, which is *H5a: Live Help function with TTS would generate higher level of trust (both cognitive trust and emotional trust) than the one without TTS.* A detailed discussion is provided in the next section of this chapter.

5.3 Discussion of Results

From the above data analysis, we found that some of our hypotheses are significantly supported by the experiment results, some marginally supported and others not. For those insignificant or marginally significant results, it is necessary to do some further analysis to investigate whether these results originated from the theoretical foundations or experimental design and process.

5.3.1 Social Presence

To our surprise, we found no significant effects of either TTS or avatar on user's perception of social presence. As mentioned in Chapter 3, social presence is considered as a single dimension representing a cognitive synthesis of all the factors about communication as they perceived by the individual to be present in the medium (Short et al. 1976). Two major factors could explain why subjects didn't give the interface with TTS and avatar a higher evaluation on social presence:

1) *Limitation of technology*. According to Short et al. (1976), the factors that contribute to the social presence of a communication medium include facial expression, direction of looking, posture, dress and other non-verbal vocal cues. In our study, although the 3D avatar, which was built on Blaxxun Contact software, could generate basic body language like eye gaze, nodding, and shaking head, the lack of facial expression greatly degraded the fidelity and warmness the subject could acquire from a face-to-face communication. Despite our TTS engine could provide a quite realistic voice, the monotonous speed and lack of variety of tones to some extent

reduced the emotion-caring ability of voice communication. As well, the synchronization of the voice and facial expression is also missing in the current settings.

2) *Product familiarity and shopping task.* As mentioned in Chapter 3, social presence is also influenced by the factor of communication task. If two distinguished aspects of exchange, "Interparty" exchange and "Interpersonal" exchange, also happened in the shopping activity and interactions between the customer and salesperson, the balance of these two could be affected by communication medium, and the emphasis on one aspect would only be realized at the expense of the other one (Douglas 1957; Morley and Stephenson 1969). In other words, if the customer pays more attention to the "Interparty" information exchange, which is more related to the task-oriented, cognitive content of the communication, she will care less about the communication cues which was mainly used to convey the "Interpersonal" attitudes (Short 1976). A post hoc analysis showed that among the 72 subjects who accomplished the task of shopping for a digital camera in our experiment, more than 75% of them rated herself/himself below 4 (on the scale of 1 to 7) on the item of "expertise of photography" and "expertise of digital camera". This may suggest that a significant number of subjects had based their evaluations of social presence more on the cognitive and task-oriented part of the communication, that is, the content of the interaction, rather than the communication medium itself. This conclusion does not necessarily diminish the importance of implementing a media-rich communication channel in the customer service of online shopping, but remind us that the product

categorization and customer expertise may also have some effects on user's perception. Information-intensive products, like digital camera and other electronics, may require different implementation of "Live Help" interface from emotion-intensive products, such as clothes and cosmetics.

5.3.2 Telepresence

From the experimental results we found that the implementation of a 3D avatar in the "Live Help" interface can contribute to a sense of telepresence in viewers. Vividness and interactivity were implied to be primary telepresence boosters, especially the interactivity (for example, avatar's gestures), from simulated face-to-face communication, is an important component for telepresence. As discussed in the section of social presence, being short of the synchronization between voice and facial expression, communication modality (for example, TTS voice) didn't make a significant difference in this study (actually the mean value of telepresence of in "Voice Only" and "Voice + Text" group is higher than that of "Text Only" group when avatar is implemented), it is very likely that with the development of more realistic speech synthesis technology, we will be able to incorporate more expressive speech to even strengthen the effects of avatar.

5.3.3 Flow

As summarized in Chapter 3, flow represented a subjective psychological experience that characterizes the human-computer experience as playful and exploratory; flow could be further decomposed into three dimensions: control,

attention focus, and cognitive enjoyment (Webster et al 1993). In our study, we found that TTS voice would significantly enhance user's perception of flow during the interaction with a customer service representative. More interestingly, the "Voice Only" group averaged the highest compared with the other two voice settings. Following arguments could be used to explain this finding:

1) Modes of communication that included voice were more expressive than those that did not. Although the voice interaction in our study is actually a one-way communication (subjects used text chat while listened to TTS Voice), voice, as a richer media including more cues like pitch and tone, is more enjoyable to use. The enjoyment may come from: 1) No need to spend effort on reading the answer from computer screen, 2) More efficient in understanding the meaning.

2) "Voice Only" treatment induced more focused attention. Under the setting of "Voice Only", the only channel that a subject could get the information from a customer service representative is through TTS voice, thus she/he had to be more focused during the interaction.

In our results, avatar group scores non-significantly higher in the measurement of flow than non-avatar group. Two major factors may contribute this insignificance:

1) Lack of control. One major component of flow is the feeling control; in our experiment, subject had very few control on the interaction. Actually, there are two types of control which had not been realized in our experiment:

a) Control over the avatar of the customer service representative. In our study, the subject was not able to select the gender, age and appearance of an avatar; neither could she control the body movement of the avatar. The only thing the subject could control is the text chat window that was used to input questions and other comments.

b) Control over the embodiment of subject. In our experiment, there is no avatar or representation of customer herself on the interface, this would also decrease the feeling of control compared with similar studies in the field of CSCW and CMC, where people used avatar (representing themselves) to communicate in a virtual environment.

2) Statistical limitation. Our sample size is still not large enough to get sufficient power.

5.3.4 Interaction Frequency

To our surprise, although the subjects from avatar group and non-avatar group reported no significant difference in their perception of social presence, we found that subject in avatar group initiated much more ($p < 0.001$) interactions than those in non-avatar group. In other words, avatar was more engaging than text. Why would people spend more time, respond more, ask more questions to avatar although they didn't feel stronger feeling of social presence? Two possible explanations are:

1) Engaging does not mean social presence or flow. Social psychologists have found that the presence of another person usually serves to increase arousal and motivation on the part of someone asked to perform a task (Zajonc 1965). In our

study, the task is to shopping for a digital camera, which means the subject had stronger incentive to use the "Live Help" feature to accomplish the task.

2) Self-report does not completely reflect the true feeling of subject's perception of social presence. As mentioned in Reeves and Nass (1996), although people would not admit verbally they will treat computer as a social actor, their behavior manifests this inclination. In our case, it is possible that despite people didn't report different perception of social presence in the written questionnaire, the significant difference in their inclination to interact with the customer service representative prove the existence of stronger perception of social presence when the avatar is implemented. But this explanation was only a surmise, the support of which requires broader literature review and further investigations.

5.3.5 Trust

Our experiment results showed that TTS Voice had marginally significant effects on user's perception of both cognitive trust in salesperson's competence and emotional trust. Consistent with the findings by Jensen et al (2000), more immediate forms of communications like TTS showed a greater impact on the development of trust.

As summarized in Xiao (2002) and Muhlfelder et al (1999), trust consists of cognitive and emotional factors. For retail transactions, cognitive-oriented factors increase confidence that a transaction will be successfully completed. Emotional-oriented factors increase feelings of attraction and loyalty. These factors

help reduce awkwardness, complexity and uncertainty, and increase confidence in the goodwill and ability of the other person.

In our study, although the interactivity and interaction content is similar between different treatment groups, user report higher level of engaging and flow under the TTS condition. This indicates that voice component can indirectly influence evaluation on salesperson's competence and willingness to trust. Media expressiveness (the presence of voice) could influence the initial formation of positive interpersonal trust. This finding is also in line with Short et al (1976)'s argument that voice is more important than interactivity in establishing trust.

6. Conclusions and Limitations

In this chapter, we will summarize the conclusions and claims for this study. In addition, we will discuss the limitation of the study from the perspective of both internal and external validity. Last, some suggestions for future research work are presented.

6.1 Conclusions

Here, we provide a quick summary of the purpose and procedures of the study. In addition, we present the conclusions and claims from this study.

6.1.1 Experiment Recap

This thesis presented a study conducted in an effort to understand the effects of TTS voice and 3D avatar on customer's perception of social presence, telepresence, flow and trust when these information technologies are implemented in the "Live Help" function of an online shopping website. As mentioned in Chapter 2, we would like to know whether and to what extent would these technologies be used to improve customer's online shopping experience.

Because this paper is the first one in studying both TTS voice and 3D avatar in the context of electronic shopping, it absorbed research work in other fields like Human Computer Interactions, Computer Mediated Communications, and Computer Supported Collaboration Work. More specifically, it focused on the concept and measurement of presence, flow and trust. This study is an attempt to explore and

validate the effects of TTS Voice and 3D avatar on these constructs not only based on the prior theories like those of social presence, media richness and communication modality but also extend those theories onto the online shopping environment. We conducted a laboratory experiment with undergraduate and graduate students to test an experimental shopping website integrated with different "Live Help" interfaces. Students were instructed to complete a shopping task, interact with the customer service representative to get necessary information, and fill out several self-assessment scales on a questionnaire.

6.1.2 Conclusions

From the data analysis results of the experiment, we have deduced the following conclusions. We have found TTS voice's main effects on flow, emotional trust and cognitive trust; as well as avatar's main effects on telepresence and interaction frequency. These are explained in detail in the following.

Firstly, we found empirical support for the effect of TTS voice on flow and trust. Within the three levels of TTS Voice factor, the mean score of "Voice Only" group is significantly higher than those of "Text Only" and "Text + Voice" group. We also found marginally significant effects of TTS voice on cognitive trust and emotional trust. Similarly, the mean score of "Voice Only" group is also higher than those of "Text Only" and "Text + Voice" group. As discussed in Chapter 5, two major reasons could contribute to the difference in flow: cognitive enjoyment and focused attention. One possible explanation for the little difference between "Text Only" and "Voice +

Text" group is that when text and voice are presented at the same time, user would treat the voice as reading aloud the script in stead of the speech given by a person, thus the TTS voice would render itself less useful and even distracting. As for the TTS's contribution to both cognitive and emotional trust, our study supports the findings mentioned by Jensen et al (2000) that more immediate forms of communications like voice had a greater impact on the development of trust. One point need to be pointed out is that since all our subjects had only used the experimental website for about an hour, we could only conclude that TTS voice's effects on the initial development of trust, i.e., the first impression, whether this effect would fade away after repeated usage still need further investigation.

Secondly, we conclude that a 3D avatar has a significant effect on user's perception of telepresence. Support for this claim come from the significant difference in univariate results. The implementation of a realistic avatar with humanoid gestures, although at a degraded fidelity, successfully enhanced the sense of telepresence in viewers. Vividness and interactivity were proved again to be primary telepresence boosters, especially the interactivity (for example, avatar's gestures), by simulating face-to-face communication. This finding also supported the claims that non-verbal language play an important role in interpersonal communications, and the interactions of non-verbal language could significantly reduce the feeling of distance and detachment in the environment of online shopping.

Thirdly, we found a very interesting phenomenon that although our subjects from the avatar and non-avatar group reported no significant difference in either social

presence or flow; they did initiate considerably more interactions with the customer service representative when the avatar is present. As discussed in Chapter 5, if users didn't feel stronger feeling of social presence or flow, why would they spend more time, respond more, and ask more questions to the avatar? This finding also supported the theory by Reeves and Nass (1996) that people would treat computer/website/software as social actors and thus behave sociably although they will not admit this verbally. Although the difference from self-report questionnaire and behavior is not the main focus of this paper, it would be of great interests if further research could be conducted to investigate this issue.

6.1.3 Practical Implications

To the best of our knowledge, there is currently no online shopping website which had implemented both TTS voice and 3D avatar in its "Live Help" interface. There are several factors that may have caused this delay:

1) Limitations of software cost. Because both TTS voice and 3D avatar are client-side programs, they require customer to download and install some plug-ins before she could use them to interact with the customer service representative. Although there are some free or open-source software products in the market of TTS voice and avatar, those of higher quality are still in proprietary status, such as AT&T Natural Voices used in this study. The software license fees may incur extra costs for online vendors who are very budget-sensitive in the current period.

2) Limitations of communication quality. We have to admit that at this time neither the TTS voice nor the 3D avatar is totally satisfactory in its "expressive richness" and fidelity. Online vendors may still have to wait for some time before a fully-commercialized solution could be provided.

Nevertheless, we consider that the results from this research still provide several important implications for the practitioners in evaluating and designing the "Live Help" interface features:

1) Voice interaction would make the real-time interactions between a serviceperson and online visitors more attractive and absorbing; more importantly, voice interaction could boost user's cognitive trust and emotional trust toward the serviceperson. Actually, just as mentioned in a report (Wingfield 2002) on Wall Street Journal, although people did enjoy the convenience of online shopping, long-distance purchases made without human contact still make them cautious before putting things into their shopping cart. When more media-rich and face-to-face-like communication channels such as TTS voice would be added to the interface, people would be less concerned that they may be scammed. This implies that the extra costs of implementing the TTS voice are justifiable from the increasing of merchant's trustworthiness and reputation.

2) People tend to communicate more with a serviceperson when that person is embodied with a 3D avatar. Although the specific design of the avatar still need further exploration, our research clearly showed that online visitors initiated significantly more interactions with the avatar even though they did not report

verbally much difference in their feeling of social presence. As mentioned in the first chapter of this paper, Landsend's successful story proves the significant correlation between customer's usage of the "Live Help" feature and her purchasing likelihood and value of selected products. This implies that the extra costs of implementing the 3D avatar are justifiable from the perspective of cost-efficiency.

6.2 Limitations

There are several threats to the overall validity of this study. They can be classified into threats to either internal or external validity of this research.

6.2.1 Internal Validity

This study proved TTS and avatar's effects on the perception of presence. However, several factors may affect the proposed causality relationship. The first threat is novelty effects and ceiling effects. Because there is currently no shopping site had integrated TTS voice or 3D avatar into its "Live Help" features, for most of our subjects, these interface features are really new; whether these effects would be washed out after repeated usage is still an open question. Ceiling effects may also affect the results because subjects from "Text Only" group may have rated too high on the perception of presence if they didn't realize the existence of other rich communication media and thus lack a reference point.

Second, the effects of TTS and avatar may depend on product category. As mentioned in Jiang and Benbasat (2002), products had different level of diagnosticity and different portion of experiential and search attribute, the perception of social presence is also up to the cognitive effort necessary to accomplish task. The effects of TTS and avatar may be different when shopping for information-intensive products like electronics and for emotion-intensive products like clothes and cosmetics.

Also, user's personal traits may have some effects on the results. For example, people's extrovertness and trust propensity will affect whether she would like to use the "Live Help" function to interact with others during shopping and how easy the feeling of trust would be developed.

6.2.2 External Validity

External validity is about the ability to generalize this study to other people, time, and place (Benbasat, 1998). The main concern is the restricted degree of realism in an information system study conducted in a laboratory setting using student subjects.

In our study, although being told the shopping site is real and required to treat the shopping task as a real one, subjects were aware that they would not make the actual purchase during the study. The lack of risk exposure may affect the measurement of trust.

Also, the use of a convenience sample (university students) reduces the external validity of this study. Students are not the best representation for the general public because they are more computer/internet literate but have comparably less disposable income. We recognize these limitations but we have chosen these methods as a tradeoff to enhance the internal validity in our study.

Last, the sample size (n = 72) is still limited, which may have decreased the value of observed power. This could also be the reason why some of our hypotheses are only marginally supported.

6.3 Future Research

As an interesting application field in both HCI and CMC, future research is expected to fully explore the virtual communication experience in online shopping. One important question is how the mediating variables such as perception of social presence, telepresence, flow and trust would affect users' purchasing decision. Would the user purchase more value of products or more likely to make a purchase when she feel strong sense of presence? Would the media-rich "Live Help" interface increase the likelihood that the user would make repeated visits in the future? Are the real-time image and voice of a customer service representative more favorable than a perfect substitute which supported by computer generated voice and avatar? Should we implement different interface features for different category of products and for users who have different level product expertise and shopping experience? We expect more research will be done in these areas.

Bibliography

Aberg, J., and Shahmehri, N. "An Empirical Study of Human Web Assistants: Implications for User Support in Web Information Systems," CHI'01, ACM Press, 2001.

Agarwal, R., and Karahanna, E. "Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage," *MIS Quarterly* (24:4) 2000, pp 665-694.

Badler, N.I. "Real-Time Virtual Humans," IEEE - The Fifth Pacific Conference on Computer Graphics and Applications, 1997, pp. 4-13.

Basso, A., Goldberg, D., Greenspan, S., and Weimer, D. "First Impressions: Emotional and Cognitive Factors Underlying Judgments of Trust E-Commerce," EC'01, ACM Press, 2001, pp. 137-143.

Biocca, F. "The cyborg's dilemma: Progressive embodiment in virtual environment," *Journal of Computer Mediated Communication* (3:2) 1997.

Bordia, P. "Face-to-Face Versus Computer-Mediated Communication: A Synthesis of the Experimental Literature," *Journal of Business Communication* (34:1), January 1997 1997, pp 99-120.

Bos, N., Olson, J., Gergle, D., Olson, G., and Wright, Z. "Effects of Four Computer-Mediated Communications Channels on Trust Development," CHI'02, ACM Press, Minneapolis, Minnesota, USA, 2002, pp. 135-140.

Brennan, S., Laurel, B., and Shneiderman, B. "Anthropomorphism: From Eliza to Terminator 2," CHI'92, ACM Press, 1992, pp. 67-70.

Burgoon, J.K., Bonito, J.A., Bengtsson, B., Ramirez, A., Dunbar, N.E., and Miczo, N. "Testing the interactivity model: communication processes, partner assessments, and the quality of collaborative work," *Journal of Management Information Systems* (16:3) 1999, pp 33-56.

Burke, K., and Chidambaram, L. "How much bandwidth is enough? A longitudinal examination of media characteristics and group outcomes," *MIS Quarterly* (23:4) 1999, pp 557-580.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., and Yan, H. "Embodiment in Conversational Interfaces," CHI'99, ACM Press, 1999.

Choi, Y.K., Miracle, G.E., and Biocca, F. "The effects of anthropomorphic agents on advertising effectives and the mediating role of presence," *Journal of Interactive Advertising* (2:1) 2001.

Csikszentmihalyi, M. *Beyond Boredom and Anxiety* Jossey-Bass, San Francisco, 1975.

Datamonitor "eShopping in the US." in: http://www.datamonitor.com

Dukcevich, D. "Lands' End's Instant Business," in: *Forbes*, 2002.

Dutoit, T. *An Introduction to Text-To-Speech Synthesis (Text, Speech and Language Technology, V. 3)* Kluwer Academic Publishers, 1997.

Elliot, B. "Call Center Internet Contact Integration: Business Issues," Gartner Group.

Elliot, B., Kolsky, E., and Cain, S. "Contact Center Self-Service Costs," Gartner Group.

Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C. *Multivariate Data Analysis, fifth edition* Prentice-Hall, Inc., 1998.

Hoffman, D.L., and Novak, T.P. "Marketing in Hypermedia Computer-Mediated Enviroments: Conceptual Foundations," *Journal of Marketing* (60:July) 1996, pp 50-68.

Isbister, K., Nakanishi, H., Ishida, T., and Nass, C. "Helper Agent: Designing an Assistant for Human-Human Interaction in a Virtual Meeting Space," CHI'00, ACM Press, The Hague, Amsterdam, 2000, pp. 57-64.

Jarvenpaa, S.L., and Todd, P.A. "Consumer Reactions to Electronic Shopping on the World Wide Web," *International Journal of Electronic Commerce* (1:2) 1997, pp 59-88.

Jensen, C., Farnham, S., Drucker, S., and Kollock, P. "The effect of communication modality on cooperation in online environments," CHI'00, ACM Press, 2000, pp. 470-477.

Jiang, Z., and Benbasat, I. "Effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping," 2002.

King, W.J., and Ohya, J. "The representation of Agents: Anthropomorphism, Agency, and Intelligence," CHI'96, ACM Press, 1996, pp. 289-290.

Kumar, N., and Benbasat, I. "Para-Social presence and communication capabilities of a web site: A theoretical perspective," 2002.

Lombard, M., and Ditton, T. "At the Heart of It All: The Concept of Presence," *Journal of Computer Mediated Communication* (3:3) 1997.

McDonald, T. "AT&T Unveils Voice Re-Creation Technology," in: *New York Times*, 2001.

Muhlbach, L., Bocker, M., and Prussog, A. "Telepresence in videocommunications: A study on stereoscopy and individual eye contact," *Human Factors* (37:2) 1995, pp 290-305.

Muhlfelder, M., Klien, U., Simon, S., and Luczak, H. "Teams without trust? Investigations in the influence of video-mediated communication on the origin of trust among cooperatin persons," *Behaviour & Information Technology* (18:5) 1999, pp 349-360.

Nass, C., Steuer, J., and Tauber, E. "Computers are social actors," CHI'94, ACM Press, 1994.

Nass, C., and Lee, K. "Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction," CHI'00, ACM Press, 2000, pp. 329-336.

Nielsen, J., Snyder, C., Molich, R., and Farrell, S. "E-Commerce User Experience," Nielsen Norman Group, 2000.

Reeves, B., and Nass, C. *The media equation: How people treat computes, television, and new media like real people and places* CLSI Publication, Stanford, CA, 1996.

Reeves, B., and Nass, C. "Perceptual bandwidth," *Communication of the ACM* (43:2) 2000, pp 65-70.

Rice, R.E. "Task analyzability, use of new medium and effectiveness: A multi-site exploration of media richness," *Organization Science* (3:4) 1992, pp 475-500.

Rose, G., Khoo, H., and Straub, D.W. "Current Technological Impediments To Business-To-Consumer Electronic Commerce," *Communications of the Association for Information Systems* (1:16) 1999.

Salem, B., and Earle, N. "Designing a Non-Verbal Language for Expressive Avatars," CVE'00, ACM Press, 2000.

Short, J., Williams, E., and Christie, B. *The Social Psychology of Telecommunication* John Wiley & Sons, London, 1976.

Steuer, J. "Defining virtual reality: Dimensions determining telepresence," *Journal of Communication* (42:Autumn) 1992, pp 73-93.

Swan, J.E., and Bowers, M.R. "Customer Trust In The Salesperson: An Integrative Review And Meta-Analysis Of The Empirical Literature," *Journal of Business Research* (44:2) 1999, pp 93-107.

Underhill, P. *Why We Buy: The Science of Shopping* Touchstone Books, 2000.

Vijaysarathy, L.R., and Jones, J.M. "Print and Internet catalog shopping," *Internet Research: Electronic Networking Applications and Policy* (10:3) 2000, pp 191-202.

Webster, J., Trevino, L.K., and Ryan, L. "The Dimensionality and Correlates of Flow in Human-Computer Interactions," *Computers in Human Behavior* (9) 1993, pp 411-426.

Wingfield, N. "A Question of Trust," in: Special Report on E-Commerce, *The Wall Street Journal*, September 16, 2002.

Xiao, S.Y., and Benbasat, I. "Understanding Customer Trust in Agent-mediated Electronic Commerce, Web-mediated Electronic Commerce, and Traditional Commerce," the Sixth INFORMS Conference on Information Systems and Technology, CIST 2001, Miami Beach, Florida, 2001.

Zajonc, R. "Social Facilitation," *Science* (149) 1965, pp 269-274.

Zhang, Q., Wolf, C.G., Daijavad, S., and Touma, M. "Talking to Customers on the Web: A Comparison of Three Voice Alternatives," CSCW 98, ACM Press, 1998, pp. 109-117.

Appendices

Appendix 1: Experimental Questionnaire

This study is to investigate an implementation of "Live Help" interfaces in electronic shopping.

Please answer the following questions only after you finish the Task. Please read the instructions carefully before answering the questions

Instructions: The following questions would ask you to evaluate the "Live Help" interfaces you have just experienced. Please indicate to what extent you agree with the following statements.

The scale would be like this: (except for Part I)

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Mildly Disagree	Neither Agree or Disagree	Mildly Agree	Agree	Strongly Agree

Please circle one number (only) to indicate your choice.

Note:

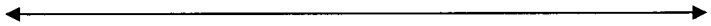
- ***Please make sure you have read carefully and understood the meaning of each statement before indicating your evaluation.***
- ***If you have any questions about those statements, please don't hesitate to ask the assistant.***

Appendix 1 (continued)

Part I: (please indicate your perception by identify a position between each contrary pair)

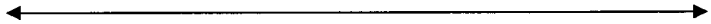
1.1 I felt that the interaction with CSR was

Impersonal * * * * * * * ***Personal***



1.2 I felt that the interaction with CSR was

Cold * * * * * * * ***Warm***

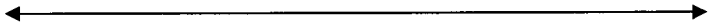


1.3 I felt that the interaction with CSR was

Distant * * * * * * * ***Close***

1.4 I felt that the interaction with CSR was

Dehumanizing * * * * * * * ***Humanizing***



1.5 I felt that the interaction with CSR was

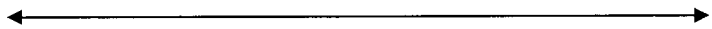
Inexpressive * * * * * * * ***Expressive***

1.6 I felt that the interaction with CSR was

Unemotional * * * * * * * ***Emotional***

1.7 I felt that the interaction with CSR was

Insensitive * * * * * * * ***Sensitive***



Appendix 1 (continued)

Part II: (Likert-Scale Questions)

2.1 When the experiment ended, I felt that I **came back to the "real world".**

2.2 I felt that I was **in the virtual store the website created**.

2.3 I felt that I was talking with the CSR **face to face**.

2.4 I felt that I was **in a same space** with the CSR.

2.5 I **never** forgot that I was in the middle of an experiment.

2.6 I felt that **my body was in the room, but my mind was inside the world created by website**.

2.7 I felt that I was **looking at or listening to a real sales person**.

2.8 I felt that the website-generated world was **more real or present** compared to the "physical world" around me.

2.9 I felt that the website-generated world seemed to me only "**something I saw**" rather than "**somewhere I visited in person**."

Part III:

3.1 My imagination was **aroused** when interacting with the CSR.

3.2 I felt **curious** when interacting with the CSR.

3.3 I thought about **unrelated things** when interacting with the CSR.

3.4 The interaction with the CSR was **interesting**.

3.5 I felt **bored** when interacting with the CSR.

3.6 I was **absorbed** in the interaction with the CSR.

Appendix 1 (continued)

3.7 I was **distracted by unrelated things** when interacting with the CSR.

3.8 My **curiosity was excited** when interacting with the CSR.

3.9 It's **fun** to interact with the CSR.

3.10 I was **not focused** when interacting with the CSR.

Part IV:

4.1 I felt **secure** about relying on the CSR for my decision.

4.2 I felt **uncomfortable** about relying on the CSR for my decision.

4.3 I felt **happy** about relying on the CSR for my decision.

4.4 I felt **unconfident** about relying on the CSR for my decision.

4.5 I felt **content** about relying on the CSR for my decision.

4.6 I felt that the CSR was **like a real expert** in assessing products.

4.7 I felt that the CSR was **incompetent** in evaluating available product choices.

4.8 I felt that the CSR had **expertise** to understand my needs and preferences.

4.9 I felt that the CSR was **unable** to capture my needs and preferences.

4.10 I felt that the CSR had **good knowledge** about products.

Appendix 2: Data Analysis Details

Sample Size: 72

Background:

- Gender: Male (34), Female (38)
- Language: Native Speaker (28), Non-Native Speaker (44)
- Age: Min: 18, Max: 37, Mean: 23.15, Std: 4.076

Reliability:

Construct	No. of Items	***Cronbach Alpha***
Social Presence	7	0.7823
Telepresence	9	0.7471
Flow	10	0.8570
Emotional Trust	5	0.9185
Competence Trust	5	0.8869

2.1 Multivariate MANOVA Test of Dependent Variables

Between-Subjects Factors			
		Value Label	
Avatar Control	0	No Avatar	36
	1	Avatar	36
Voice Control	0	Text Only	24
	1	Voice Only	24
	2	Text + Voice	24

Multivariate Tests[d]

Effect		Value	F	ypothesis d	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	Pillai's Trace	.986	735.379[b]	6.000	61.000	.000	.986	412.274	1.000
	Wilks' Lambd	.014	735.379[b]	6.000	61.000	.000	.986	412.274	1.000
	Hotelling's Tr	72.332	735.379[b]	6.000	61.000	.000	.986	412.274	1.000
	Roy's Largest	72.332	735.379[b]	6.000	61.000	.000	.986	412.274	1.000
AVATAR	Pillai's Trace	.172	2.113[b]	6.000	61.000	.065	.172	12.675	.714
	Wilks' Lambd	.828	2.113[b]	6.000	61.000	.065	.172	12.675	.714
	Hotelling's Tr	.208	2.113[b]	6.000	61.000	.065	.172	12.675	.714
	Roy's Largest	.208	2.113[b]	6.000	61.000	.065	.172	12.675	.714
VOICE	Pillai's Trace	.197	1.131	12.000	24.000	.341	.099	13.576	.624
	Wilks' Lambd	.807	1.150[b]	12.000	22.000	.327	.102	13.805	.632
	Hotelling's Tr	.234	1.168	12.000	20.000	.314	.105	14.020	.640
	Roy's Largest	.207	2.144[c]	6.000	62.000	.061	.172	12.863	.723
AVATAR * V(	Pillai's Trace	.129	.710	12.000	24.000	.739	.064	8.523	.394
	Wilks' Lambd	.873	.713[b]	12.000	22.000	.736	.066	8.558	.395
	Hotelling's Tr	.143	.716	12.000	20.000	.734	.067	8.587	.396
	Roy's Largest	.127	1.308[c]	6.000	62.000	.267	.112	7.847	.475

a.Computed using alpha = .05

b.Exact statistic

c.The statistic is an upper bound on F that yields a lower bound on the significance level.

d.Design: Intercept+AVATAR+VOICE+AVATAR * VOICE

2.2 Univariate Results of MANOVA Test of All Dependent Variables

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Model	Interactions Initiated by Subject	369.736[b]	5	73.947	2.009	.089	.132	10.047	.637
	SPMEAN	2.121[c]	5	.424	.681	.639	.049	3.406	.232
	TPMEAN	4.080[d]	5	.816	1.145	.346	.080	5.726	.382
	FLMEAN	5.813[e]	5	1.163	2.186	.066	.142	10.932	.681
	ETMEAN	7.318[f]	5	1.464	1.122	.357	.078	5.610	.375
	CTMEAN	8.858[g]	5	1.772	1.747	.136	.117	8.733	.567
Intercept	Interactions Initiated by Subject	8602.347	1	8602.347	233.748	.000	.780	233.748	1.000
	SPMEAN	1809.971	1	1809.971	2906.269	.000	.978	2906.269	1.000
	TPMEAN	1365.544	1	1365.544	1916.311	.000	.967	1916.311	1.000
	FLMEAN	2148.401	1	2148.401	4040.208	.000	.984	4040.208	1.000
	ETMEAN	1816.036	1	1816.036	1392.299	.000	.955	1392.299	1.000
	CTMEAN	2195.636	1	2195.636	2164.588	.000	.970	2164.588	1.000
AVATAR	Interactions Initiated by Subject	292.014	1	292.014	7.935	.006	.107	7.935	.793
	SPMEAN	2.281E-02	1	2.281E-02	.037	.849	.001	.037	.054
	TPMEAN	3.557	1	3.557	4.992	.029	.070	4.992	.595
	FLMEAN	1.201	1	1.201	2.259	.138	.033	2.259	.316
	ETMEAN	.320	1	.320	.245	.622	.004	.245	.078
	CTMEAN	.269	1	.269	.265	.608	.004	.265	.080
VOICE	Interactions Initiated by Subject	42.528	2	21.264	.578	.564	.017	1.156	.142
	SPMEAN	.313	2	.156	.251	.779	.008	.502	.088
	TPMEAN	.163	2	8.165E-02	.115	.892	.003	.229	.067
	FLMEAN	3.591	2	1.795	3.376	.040	.093	6.753	.617
	ETMEAN	6.988	2	3.494	2.679	.076	.075	5.357	.514
	CTMEAN	6.248	2	3.124	3.080	.053	.085	6.159	.575
AVATAR * VOICE	Interactions Initiated by Subject	35.194	2	17.597	.478	.622	.014	.956	.125
	SPMEAN	1.786	2	.893	1.434	.246	.042	2.868	.297
	TPMEAN	.360	2	.180	.252	.778	.008	.505	.088
	FLMEAN	1.021	2	.510	.960	.388	.028	1.920	.210
	ETMEAN	1.000E-02	2	5.000E-03	.004	.996	.000	.008	.051
	CTMEAN	2.341	2	1.171	1.154	.322	.034	2.308	.245
Error	Interactions Initiated by Subject	2428.917	66	36.802					
	SPMEAN	41.104	66	.623					
	TPMEAN	47.031	66	.713					
	FLMEAN	35.096	66	.532					
	ETMEAN	86.087	66	1.304					
	CTMEAN	66.947	66	1.014					
Total	Interactions Initiated by Subject	11401.000	72						
	SPMEAN	1853.196	72						
	TPMEAN	1416.655	72						
	FLMEAN	2189.310	72						
	ETMEAN	1909.440	72						
	CTMEAN	2271.440	72						
Corrected Total	Interactions Initiated by Subject	2798.653	71						
	SPMEAN	43.225	71						
	TPMEAN	51.111	71						
	FLMEAN	40.909	71						
	ETMEAN	93.404	71						
	CTMEAN	75.804	71						

a. Computed using alpha = .05

b. R Squared = .132 (Adjusted R Squared = .066)

c. R Squared = .049 (Adjusted R Squared = -.023)

d. R Squared = .080 (Adjusted R Squared = .010)

e. R Squared = .142 (Adjusted R Squared = .077)

f. R Squared = .078 (Adjusted R Squared = .009)

g. R Squared = .117 (Adjusted R Squared = .050)

2.3 Descriptive Statistics and Univariate Results of Social Presence

Between-Subjects Factors			
		Value Label	
Avatar Control	0	No Avatar	36
	1	Avatar	36
Voice Control	0	Text Only	24
	1	Voice Only	24
	2	Text + Voice	24

Descriptive Statistics

Dependent Variable: SPMEAN

Avatar Control	Voice Control	Mean	Std. Deviation	N
No Avatar	Text Only	5.1548	.88370	12
	Voice Only	4.9167	.86245	12
	Text + Voice	4.9167	.79530	12
	Total	4.9960	.83123	36
Avatar	Text Only	4.7738	.78118	12
	Voice Only	5.2973	.71913	12
	Text + Voice	5.0238	.67236	12
	Total	5.0316	.73716	36
Total	Text Only	4.9643	.83857	24
	Voice Only	5.1070	.80054	24
	Text + Voice	4.9702	.72229	24
	Total	5.0138	.78026	72

Tests of Between-Subjects Effects

Dependent Variable: SPMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Mode	2.121[b]	5	.424	.681	.639	.049	3.406	.232
Intercept	1809.971	1	1809.971	2906.269	.000	.978	2906.269	1.000
AVATAR	2.281E-02	1	2.281E-02	.037	.849	.001	.037	.054
VOICE	.313	2	.156	.251	.779	.008	.502	.088
AVATAR * VOIC	1.786	2	.893	1.434	.246	.042	2.868	.297
Error	41.104	66	.623					
Total	1853.196	72						
Corrected Total	43.225	71						

a. Computed using alpha = .05

b. R Squared = .049 (Adjusted R Squared = -.023)

2.4 Descriptive Statistics and Univariate Results of Telepresence

Between-Subjects Factors			
		Value Label	
Avatar Control	0	No Avatar	36
	1	Avatar	36
Voice Control	0	Text Only	24
	1	Voice Only	24
	2	Text + Voice	24

Descriptive Statistics

Dependent Variable: TPMEAN

Avatar Control	Voice Control	Mean	Std. Deviation	N
No Avatar	Text Only	4.1574	.80745	12
	Voice Only	4.1019	1.07043	12
	Text + Voice	4.1389	.55480	12
	Total	4.1327	.81382	36
Avatar	Text Only	4.4444	.81236	12
	Voice Only	4.7318	1.07488	12
	Text + Voice	4.5556	.59553	12
	Total	4.5773	.83449	36
Total	Text Only	4.3009	.80556	24
	Voice Only	4.4168	1.09730	24
	Text + Voice	4.3472	.60176	24
	Total	4.3550	.84845	72

Tests of Between-Subjects Effects

Dependent Variable: TPMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Model	4.080[b]	5	.816	1.145	.346	.080	5.726	.382
Intercept	1365.544	1	1365.544	1916.311	.000	.967	1916.311	1.000
AVATAR	3.557	1	3.557	4.992	.029	.070	4.992	.595
VOICE	.163	2	8.165E-02	.115	.892	.003	.229	.067
AVATAR * VOIC	.360	2	.180	.252	.778	.008	.505	.088
Error	47.031	66	.713					
Total	1416.655	72						
Corrected Total	51.111	71						

a. Computed using alpha = .05

b. R Squared = .080 (Adjusted R Squared = .010)

2.5 Descriptive Statistics and Univariate Results of Flow

Between-Subjects Factors			
		Value Label	
Avatar Control	0	No Avatar	36
	1	Avatar	36
Voice Control	0	Text Only	24
	1	Voice Only	24
	2	Text + Voice	24

Descriptive Statistics

Dependent Variable: FLMEAN

Avatar Control	Voice Control	Mean	Std. Deviation	N
No Avatar	Text Only	5.3833	.83430	12
	Voice Only	5.5417	.71409	12
	Text + Voice	5.0750	.81478	12
	Total	5.3333	.79138	36
Avatar	Text Only	5.3083	.50715	12
	Voice Only	6.0083	.82952	12
	Text + Voice	5.4583	.61268	12
	Total	5.5917	.71289	36
Total	Text Only	5.3458	.67629	24
	Voice Only	5.7750	.79359	24
	Text + Voice	5.2667	.73168	24
	Total	5.4625	.75906	72

Tests of Between-Subjects Effects

Dependent Variable: FLMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Model	5.813[b]	5	1.163	2.186	.066	.142	10.932	.681
Intercept	2148.401	1	2148.401	4040.208	.000	.984	4040.208	1.000
AVATAR	1.201	1	1.201	2.259	.138	.033	2.259	.316
VOICE	3.591	2	1.795	3.376	.040	.093	6.753	.617
AVATAR * VOIC	1.021	2	.510	.960	.388	.028	1.920	.210
Error	35.096	66	.532					
Total	2189.310	72						
Corrected Total	40.909	71						

a. Computed using alpha = .05

b. R Squared = .142 (Adjusted R Squared = .077)

2.6 Descriptive Statistics and Univariate Results of Emotional Trust

Between-Subjects Factors			
		Value Label	
Avatar Control	0	No Avatar	36
	1	Avatar	36
Voice Control	0	Text Only	24
	1	Voice Only	24
	2	Text + Voice	24

Descriptive Statistics

Dependent Variable: ETMEAN

Avatar Control	Voice Control	Mean	Std. Deviation	N
No Avatar	Text Only	4.8000	1.08209	12
	Voice Only	5.4000	.73855	12
	Text + Voice	4.6667	1.05945	12
	Total	4.9556	.99841	36
Avatar	Text Only	4.9167	1.32242	12
	Voice Only	5.5167	1.22833	12
	Text + Voice	4.8333	1.31518	12
	Total	5.0889	1.28947	36
Total	Text Only	4.8583	1.18319	24
	Voice Only	5.4583	.99298	24
	Text + Voice	4.7500	1.17103	24
	Total	5.0222	1.14698	72

Tests of Between-Subjects Effects

Dependent Variable: ETMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Model	7.318[b]	5	1.464	1.122	.357	.078	5.610	.375
Intercept	1816.036	1	1816.036	1392.299	.000	.955	1392.299	1.000
AVATAR	.320	1	.320	.245	.622	.004	.245	.078
VOICE	6.988	2	3.494	2.679	.076	.075	5.357	.514
AVATAR * VOICE	1.000E-02	2	5.000E-03	.004	.996	.000	.008	.051
Error	86.087	66	1.304					
Total	1909.440	72						
Corrected Total	93.404	71						

a. Computed using alpha = .05

b. R Squared = .078 (Adjusted R Squared = .009)

2.7 Descriptive Statistics and Univariate Results of Cognitive Trust

Between-Subjects Factors			
		Value Label	
Avatar Control	0	No Avatar	36
	1	Avatar	36
Voice Control	0	Text Only	24
	1	Voice Only	24
	2	Text + Voice	24

Descriptive Statistics

Dependent Variable: CTMEAN

Avatar Control	Voice Control	Mean	Std. Deviation	N
No Avatar	Text Only	5.4167	.94372	12
	Voice Only	5.6500	.82737	12
	Text + Voice	5.3167	1.31414	12
	Total	5.4611	1.02853	36
Avatar	Text Only	5.1000	1.35244	12
	Voice Only	6.2167	.55569	12
	Text + Voice	5.4333	.80378	12
	Total	5.5833	1.04895	36
Total	Text Only	5.2583	1.15190	24
	Voice Only	5.9333	.74756	24
	Text + Voice	5.3750	1.06699	24
	Total	5.5222	1.03328	72

Tests of Between-Subjects Effects

Dependent Variable: CTMEAN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Model	8.858[b]	5	1.772	1.747	.136	.117	8.733	.567
Intercept	2195.636	1	2195.636	2164.588	.000	.970	2164.588	1.000
AVATAR	.269	1	.269	.265	.608	.004	.265	.080
VOICE	6.248	2	3.124	3.080	.053	.085	6.159	.575
AVATAR * VOIC	2.341	2	1.171	1.154	.322	.034	2.308	.245
Error	66.947	66	1.014					
Total	2271.440	72						
Corrected Total	75.804	71						

a. Computed using alpha = .05

b. R Squared = .117 (Adjusted R Squared = .050)

2.8 Descriptive Statistics and Univariate Results of Interaction Frequency

Between-Subjects Factors			
		Value Label	
Avatar Control	0	No Avatar	36
	1	Avatar	36
Voice Control	0	Text Only	24
	1	Voice Only	24
	2	Text + Voice	24

Descriptive Statistics

Dependent Variable: Interactions Initiated by Subject

Avatar Control	Voice Control	Mean	Std. Deviation	N
No Avatar	Text Only	7.83	5.890	12
	Voice Only	10.33	5.549	12
	Text + Voice	8.58	3.397	12
	Total	8.92	5.033	36
Avatar	Text Only	13.83	7.309	12
	Voice Only	13.50	7.562	12
	Text + Voice	11.50	5.760	12
	Total	12.94	6.803	36
Total	Text Only	10.83	7.179	24
	Voice Only	11.92	6.685	24
	Text + Voice	10.04	4.859	24
	Total	10.93	6.278	72

Tests of Between-Subjects Effects

Dependent Variable: Interactions Initiated by Subject

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Model	369.736[b]	5	73.947	2.009	.089	.132	10.047	.637
Intercept	8602.347	1	8602.347	233.748	.000	.780	233.748	1.000
AVATAR	292.014	1	292.014	7.935	.006	.107	7.935	.793
VOICE	42.528	2	21.264	.578	.564	.017	1.156	.142
AVATAR * VOIC	35.194	2	17.597	.478	.622	.014	.956	.125
Error	2428.917	66	36.802					
Total	11401.000	72						
Corrected Total	2798.653	71						

a. Computed using alpha = .05

b. R Squared = .132 (Adjusted R Squared = .066)

2.9 Correlations between Dependent Variables

Correlations

		SPMEAN	TPMEAN	FLMEAN	ETMEAN	CTMEAN
SPMEAN	Pearson Correlation	1	.468**	.573**	.521**	.548**
	Sig. (2-tailed)	.	.000	.000	.000	.000
	N	72	72	72	72	72
TPMEAN	Pearson Correlation	.468**	1	.540**	.406**	.315**
	Sig. (2-tailed)	.000	.	.000	.000	.007
	N	72	72	72	72	72
FLMEAN	Pearson Correlation	.573**	.540**	1	.479**	.576**
	Sig. (2-tailed)	.000	.000	.	.000	.000
	N	72	72	72	72	72
ETMEAN	Pearson Correlation	.521**	.406**	.479**	1	.754**
	Sig. (2-tailed)	.000	.000	.000	.	.000
	N	72	72	72	72	72
CTMEAN	Pearson Correlation	.548**	.315**	.576**	.754**	1
	Sig. (2-tailed)	.000	.007	.000	.000	.
	N	72	72	72	72	72

**. Correlation is significant at the 0.01 level (2-tailed).

2.10 Scheffe Test on Different Treatment of TTS Voice

FLMEAN

Scheffe[a]

Voice Control	N	Subset for alpha = .05
		1
Text + Voice	24	5.2667
Text Only	24	5.3458
Voice Only	24	5.7750
Sig.		.064

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 24.000.

ETMEAN

Scheffe[a]

Voice Control	N	Subset for alpha = .05
		1
Text + Voice	24	4.7500
Text Only	24	4.8583
Voice Only	24	5.4583
Sig.		.098

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 24.000.

CTMEAN

Scheffe[a]

Voice Control	N	Subset for alpha = .05
		1
Text Only	24	5.2583
Text + Voice	24	5.3750
Voice Only	24	5.9333
Sig.		.073

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 24.000.

2.11 Scheffe Test on Different Treatment of TTS Voice (Cont.)

Multiple Comparisons

Scheffe

Dependent Vari	(I) Voice Con	(J) Voice Cor	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interva	
						Lower Bound	Upper Boun
Interactions Initi by Subject	Text Only	Voice Only	-1.08	1.824	.839	-5.65	3.48
		Text + Voice	.79	1.824	.910	-3.77	5.36
	Voice Only	Text Only	1.08	1.824	.839	-3.48	5.65
		Text + Voice	1.88	1.824	.592	-2.69	6.44
	Text + Voice	Text Only	-.79	1.824	.910	-5.36	3.77
		Voice Only	-1.88	1.824	.592	-6.44	2.69
SPMEAN	Text Only	Voice Only	-.1427	.22765	.822	-.7122	.4269
		Text + Voice	-.0060	.22765	1.000	-.5755	.5636
	Voice Only	Text Only	.1427	.22765	.822	-.4269	.7122
		Text + Voice	.1367	.22765	.835	-.4328	.7063
	Text + Voice	Text Only	.0060	.22765	1.000	-.5636	.5755
		Voice Only	-.1367	.22765	.835	-.7063	.4328
TPMEAN	Text Only	Voice Only	-.1159	.24805	.897	-.7365	.5047
		Text + Voice	-.0463	.24805	.983	-.6669	.5743
	Voice Only	Text Only	.1159	.24805	.897	-.5047	.7365
		Text + Voice	.0696	.24805	.961	-.5510	.6902
	Text + Voice	Text Only	.0463	.24805	.983	-.5743	.6669
		Voice Only	-.0696	.24805	.961	-.6902	.5510
FLMEAN	Text Only	Voice Only	-.4292	.21230	.137	-.9603	.1020
		Text + Voice	.0792	.21230	.933	-.4520	.6103
	Voice Only	Text Only	.4292	.21230	.137	-.1020	.9603
		Text + Voice	.5083	.21230	.064	-.0228	1.0395
	Text + Voice	Text Only	-.0792	.21230	.933	-.6103	.4520
		Voice Only	-.5083	.21230	.064	-1.0395	.0228
ETMEAN	Text Only	Voice Only	-.6000	.32306	.186	-1.4083	.2083
		Text + Voice	.1083	.32306	.945	-.6999	.9166
	Voice Only	Text Only	.6000	.32306	.186	-.2083	1.4083
		Text + Voice	.7083	.32306	.098	-.0999	1.5166
	Text + Voice	Text Only	-.1083	.32306	.945	-.9166	.6999
		Voice Only	-.7083	.32306	.098	-1.5166	.0999
CTMEAN	Text Only	Voice Only	-.6750	.28984	.073	-1.4001	.0501
		Text + Voice	-.1167	.28984	.922	-.8418	.6085
	Voice Only	Text Only	.6750	.28984	.073	-.0501	1.4001
		Text + Voice	.5583	.28984	.164	-.1668	1.2835
	Text + Voice	Text Only	.1167	.28984	.922	-.6085	.8418
		Voice Only	-.5583	.28984	.164	-1.2835	.1668