The Multiple Source Effect and Synthesized Speech
Doubly-Disembodied Language as a Conceptual Framework

KWAN MIN LEE
University of Southern California
CLIFFORD NASS
Stanford University

*Two experiments examine the effect of multiple synthetic voices in an e-commerce context. In Study 1, participants (*N = *40) heard five positive reviews about a book from five different synthetic voices or from a single synthetic voice. Consistent with the multiple source effect, results showed that participants hearing multiple synthetic voices evaluated the reviewed books more positively, predicted more favorable public reaction to the books, and felt greater social presence of virtual speakers. The effects were mediated by participants' feelings of social presence. The second experiment (*N = *40) showed that the observed effects persisted even when participants were shown the purely artificial nature of synthesized speech. These results support the idea that characteristics of synthetic voices in doubly disembodied language settings influence participants' imagination of virtual speakers, and that technological literacy does not hinder social responses to anthropomorphic technologies such as text-to-speech (TTS).*

Many forms of mediated communication such as newspapers, radio, films, TV, and computers include disembodied language, "language that is not being produced by an actual speaker at the moment it is being interpreted" (Clark, 1999, p. 1). Disembodied language is quite abundant in everyday life and most people understand it with no difficulties. Consequently, people tend to think that the interpretation of disembodied language is a trivial process. Clark (1996, 1999) argued, however, that the process for interpreting disembodied language is a remarkable one: Through imagination, people must visualize virtual speakers who have written or spoken sentences in disembodied language.

Kwan Min Lee (Ph.D., Stanford University) is an assistant professor at the Annenberg School for Communication, University of Southern California; *Clifford Nass*, (Ph.D., Sociology, Princeton University, 1986) is a professor in the Department of Communication, Stanford University. This study is based on the first author's doctoral dissertation, for which the second author was major advisor. Correspondence may be addressed to Kwan Min Lee, Annenberg School for Communication, 3502 Watt Way, University of Southern California, Los Angeles, CA 90089: email: kwanminl@usc.edu.

Human Communication Research, Vol. 30 No. 2, April 2004 182–207

In this sense, disembodied language is no more than a representation of embodied language produced by virtual speakers. It is understood in the same way that people understand natural embodied language (e.g., face-to-face conversation).

There are two forms of disembodied language: written language and prerecorded human speech (Clark, 1999).[1] The two forms are clearly distinct not only because their modalities are different (visual versus auditory) but also because they yield two psychologically different imagination mechanisms. For written language, readers' imagination of virtual speakers is based on internal visualization occurring when people read a text (Bleich, 1981). If a speaker is unknown to a reader, linguistic cues in the writing (e.g., content, writing style, word choice, verb–adjective ratio, certainty versus uncertainty; see Scherer, 1979) become the basis for imagining.[2] For pre-recorded speech, both vocal (e.g., loudness, pitch, and perceived speech rate) and linguistic cues of the speech affect listeners' attributions toward a virtual speaker. The judgment of vocal cues is evolutionarily hard-wired (Cooper & Aslin, 1990; Fernald, 1992; Nass & Gong, 2000), so these cues are usually assessed immediately (Debus, 1978) and become more salient and influential in attributions toward unknown speakers (see Lee & Nass, 2001 for empirical evidence).

In recent years, there has been increasing interest in and use of a new type of disembodied language—*computer-synthesized speech*. This is due to several factors: (a) a growing demand for Speech User Interfaces (SUIs) as an alternative/complement to Graphical User Interfaces (GUIs) (see Shneiderman, 1997; Sawhney, & Schmandt, 1997); (b) a need to provide better access to computers and the Internet to visually disabled people (James, 1998); (c) a need for a quick and easy transformation of time-sensitive textual information (e.g., news, traffic condition, weather) into audible information; and (d) a quest for intuitive human–computer interfaces utilizing natural language interaction (Cassells, Sullivan, Prevost, & Churchill, 2000). Computer-synthesized speech provides a practical solution to the various demands listed above.

When a computer synthesizes a voice and produces disembodied speech from a given text, the speech becomes a special case of disembodied language because the voice is clearly not from a human.[3] The current study calls this special type of disembodied language *doubly disembodied language*. Computer-synthesized speech is doubly disembodied because of the absence of an actual speaker at the moment of its interpretation—"first degree disembodiment"—and the broken association between the characteristics of the speech and its source (the speaker)—"second degree disembodiment."

From one viewpoint, the content of synthesized speech should be the sole basis for listeners' imaginations of a virtual speaker, because the

vocal characteristics of synthesized speech are clearly artificial, predetermined by computer algorithms, and have nothing to do with the actual source of the speech. In other words, a clearly nonhuman synthetic voice should make vocal aspects of synthesized speech irrelevant to listeners' imagination of virtual speakers. From another viewpoint, however, vocal characteristics of even synthetic voices will be relevant to listeners' imagination of virtual speakers, because human brains are not evolved to respond differently to synthetic voices as compared to real human voices (for a similar claim, see Nass & Gong, 2000). The present study thus provides a critical test of the effect of vocal cues of synthesized speech on listeners' judgments in the context of endorsements by multiple sources.

To understand listeners' attributions toward virtual speakers, the concept of social presence is operationalized and measured. "Social presence" is defined as the mental simulation of other intelligences (Biocca, 1997). In other words, social presence is a technology users' sense that other intelligent beings coexist and interact with them in a virtual or imagined environment (Biocca, 1997; Heeter, 1992). Technology users feel social presence either when they forget the technology-mediated nature of their social interaction with other humans (e.g., phone conversation) or when they do not notice the artificiality of experienced nonhuman social actors such as computers, robots, and software agents (see Lee, 2002 for a detailed explication of social presence). In short, social presence occurs when technology users successfully imagine or simulate intelligent social actors (whether human or nonhuman intelligences) when they use media or simulation technologies.

MULTIPLE SOURCES IN EMBODIED, DISEMBODIED, AND DOUBLY DISEMBODIED LANGUAGE SETTINGS

In persuasive communication, one of the most powerful strategies for increasing the persuasive potential of a message is to use multiple sources. Multiple sources are employed both in embodied (e.g., face-to-face communication) and disembodied communication situations. For example, it is very common in political rallies to use multiple speakers supporting a political issue or a candidate. Attorneys present as many witnesses as possible to persuade the judge and jury (Harkins & Petty, 1987). Similarly, advertisers provide an army of product endorsers to consumers, who constantly see, hear, and read a plethora of product endorsements from multiple sources whenever using media.

The effect of multiple sources on persuasion has been documented as the "multiple source effect" in social psychology (Harkins & Petty, 1981, 1983, 1987) and the "source magnification effect" in marketing (Moore &

Reardon, 1987; Moore, Mowen, & Reardon, 1994). Both research traditions show that multiple sources have more persuasive power than a single source when (a) each source provides a convincing argument; (b) each source provides a different argument; (c) each source is perceived as an independent source; and (d) target audiences are physically exposed to actual sources rather than merely knowing the existence of multiple sources and arguments.

The multiple source technique is one of the oldest persuasive strategies; nonetheless, it is still used heavily in new media and computers. In electronic commerce, programmers and designers heavily rely on multiple endorsements from lay people (rather than celebrities) in order to promote their products. As a result, it is now almost normative for an e-commerce site to provide multiple consumer reviews about a product. It is believed that consumers utilize the comments from other consumers when they evaluate an unknown product.

In GUI-based e-commerce sites, only linguistic cues are utilized to manifest multiple sources. That is, designers present multiple customer reviews by sorting each review under a real name ("Tom from New York") or anonymously ("A customer from Palo Alto, CA"). More fundamentally, users recognize multiple sources by noticing different linguistic cues (e.g., content, content styles, grammar, or word choice) in each review.

As technologies (e.g., speech synthesis, natural language recognition, and mobile phones) for SUIs become more mature and business environments become increasingly mobile, new media designers can manipulate both linguistic and vocal cues to maximize the multiple source effect. For example, SUI or voice portal designers can easily manipulate such parameters as fundamental frequency, frequency range, decibel level, and speech rate of a text-to-speech (TTS[4]) system. While multiple prerecorded human voices manifest multiple human sources and thereby naturally induce the multiple source effect, it is not known whether synthetic voices differentiated by vocal parameters affect listeners' imagination of multiple virtual speakers and thereby increase the persuasive impact of synthetic voice narration.

Following the idea that responses to synthetic voices are grounded in user's automatic application of social rules and heuristics (Nass & Gong, 2002; Nass & Moon, 2000; Reeves & Nass, 1996), the current study tests the mediating effect of social presence on other dependent variables measuring persuasive impact. If people have a greater feeling of social presence of multiple virtual speakers when they hear multiple synthetic voices (compared to hearing just a single synthetic voice), it would support the proposition that vocal cues of synthesized speech influence listeners' imagination of virtual speakers, and thereby mediate the effect of multiple synthetic voices on other dependent variables. If there is no mediating

effect, one can conclude that multiple synthetic voices affect listeners' attitudes directly.

To address these issues, the two current experiments involve multiple synthetic voices in a persuasive context. Both experiments use the context of a very realistic advertisement based on testimonials. Specifically, they examine whether narration of multiple testimonials via multiple synthetic voices has greater persuasive impact than narration of the same multiple testimonials via a single synthetic voice. As current text-to-speech is so obviously not similar to a human voice, it would be strong evidence for the influence of the vocal cues of synthesized speech in people's imagination of virtual speakers. Additionally, to provide convincing evidence for listeners' "social responses" (i.e., people's use of social rules and heuristics usually directed at other people; Reeves & Nass, 1996) to virtual speakers in doubly disembodied language situation, the mediating role of the social presence of virtual speakers on the multiple source effect process is measured and tested.

EXPERIMENT 1

Hypotheses

According to the "Computers Are Social Actors" paradigm, people will feel the existence of another human, or a human-like intelligence (i.e., social presence as defined by Biocca, 1997), even when they hear a synthetic voice. Since one voice equaled one human throughout human evolution, multiple voices have always indicated the existence of multiple humans or human-like intelligences to listeners. In the context of multiple customer testimonials pertaining to books on a book review website, the first hypothesis predicted that:

> H1: People will feel stronger social presence of multiple virtual speakers when they hear customer reviews of a book via multiple synthetic voices than when they hear the same reviews via one synthetic voice.

If the use of multiple synthetic voices positively affects listeners' imagination of multiple virtual speakers, results will show a stronger persuasion impact of multiple synthetic voices over a single synthetic voice. Based on the finding that multiple testimonial providers are more effective than a single testimonial provider (see the previous section), the next hypothesis predicted:

H2: People will evaluate a reviewed book more positively when they hear positive reviews of the book via multiple synthetic voices than when they hear the same reviews via one synthetic voice.

Persuasion can be realized not just by convincing a person that he or she likes the book, but also by leading individuals to believe that the public likes the book, because the sense that multiple people are endorsing a book, engendered by the multiple voices, could lead to a sense of public opinion. This necessitated a separate test of the multiple source effect on people's assessment of other people's judgment of a reviewed book:

H3: People will assess other people's evaluation of a reviewed book more positively when they hear positive reviews of the book via multiple synthetic voices than when they hear the same reviews via one synthetic voice.

Multiple voices apparently reflect multiple opinions which in turn reflect a broader cross-section, and hence more credible set, of viewpoints. A website providing multiple synthetic voices should therefore be regarded as more credible than one providing a single synthetic voice:

H4: People will judge the credibility of a website more highly when they hear reviews via multiple synthetic voices than when they hear the same reviews via one synthetic voice.

If listeners' social responses to synthesized speech are oriented toward imagined virtual speakers, the perceived social presence of virtual speakers would predictably mediate the effect of multiple synthetic voices on persuasion (as measured by the dependent variables associated with H2, H3, and H4—personal opinion of reviewed books; assessment of public opinion of reviewed books; website credibility).

H5: The effect of multiple synthetic voices on personal opinion of a reviewed book, assessment of public opinion of a reviewed book, and website credibility will be mediated by listeners' feelings of social presence of virtual speakers.

Method

The current experiment was executed in the context of a book-buying web site that presents customer reviews of a book. The experimental web site listed three different books, all on the same webpage. For each book,

five actual reviews from Amazon.com were selected and used. The webpage had a visual interface based on Amazon.com's book descriptions. The page included the titles, the author(s) (in text), and pictures of the books. Instead of having customer reviews of a book in text form, there was a link to an audio (.wav) file; clicking on the link would play the reviews. Participants heard five reviews for each book, either delivered via five different synthetic voices (one for each book review) or all via one of the synthetic voices (one voice for all five reviews).

Participants

Participants were 40 (22 women and 18 men) college undergraduates enrolled in a large introductory class. Following procedures recommended by Keppel (1982), power analyses were conducted to calculate an optimal sample size. In order to obtain a power of more than .80, with effect sizes (ranging from .07 to .17) reported in previous research on social responses to TTS (Nass & Lee, 2001), a minimum sample size between 41 and 107 was required. With the sample size of 40, the current design was thus a slightly conservative test. As compensation, indices were expected to be more reliable than in the earlier studies (which they were, as indicated below), thereby boosting confidence in our sample size.

Participants were randomly assigned to either the single synthetic voice condition or the multiple synthetic voices condition. Gender was almost balanced across conditions (9 men and 11 women for each condition). All participants signed informed consent forms and were debriefed at the end of the experiment session.

Procedure

The experiment was a two-group between-subjects design (20 participants per condition), with a set of three different books as a repeated factor. Participants logged on to the experimental website for their condition and provided their responses through mouse clicks, in exactly the same way as they would normally do in everyday Web surfing. All participants used the Internet Explorer 4.0 (or higher) browser in order to ensure the same graphic environment across conditions. As noted earlier, each book review page consisted of a picture of the book, a title, author names, and a .wav file of the reviews. The customer reviews were edited versions of actual customer reviews on the Amazon.com site (see Appendix 1 for the list of titles, author names, and actual review scripts used). The books and their authors were selected based on low sales so that the participants would not be familiar with the books. Lack of familiarity was verified by a question asked at the end of the experiment. All books were fiction to avoid bias based on users' general knowledge about various topics.

Below the icon for the audio file, there was a questionnaire regarding the book being reviewed and the review itself. The two subsequent book reviews and questionnaires were placed sequentially on the web page. In both conditions, there was only one audio file for each book. With the exception of the number of voices used in the audio file, the visual layout, textual information, and book review content were identical across conditions. After hearing the reviews for all three books, the participants were presented with a final set of questions with regard to their evaluation of the website and their experience. Finally, all participants were debriefed and thanked.

Manipulation

Our major manipulation goal was to create synthetic voices which were clearly different from one another. Previous research has documented the effects of the gender of synthetic voices (Lee, Nass, & Brave, 2000; Morishima, Nass, Bennett, & Lee, 2001); therefore, all voices were from one gender (male). Three distinct voices were created by manipulating preset TTS parameters (e.g., fundamental frequency, speech rate, and frequency range) provided by the CSLU Toolkit (the Toolkit is available as a free download[5]; see Appendix 2 for the specific settings of the three voices). The remaining two voices were the "big man" and "man" synthetic voices offered by Bell Lab's TTS engine demo website[6] (the website permitted one to create sound files simply by typing in text and selecting a voice to read the text). The key point is that particular vocal characteristics were unimportant; the experiment simply attempted to provide five synthetic voices that would be perceived as different from each other. Pre-tests ensured that each voice was distinctive enough to be readily distinguishable.[7]

One-voice participants (N = 20) heard a single voice read all five reviews for all three books. One-fifth of the one-voice participants heard each of the voices to control for possible voice effects. In the multiple voice condition (N = 20), each voice read one of the five reviews of each of book 1, book 2, and book 3. The order of presentation of the five voices was balanced via Latin squares to control for possible order effects.

Measures

All dependent measures were based on items from the Web-based, textual questionnaires. Participants used radio buttons (buttons in a survey web page that allow only one button to be selected at any time; selection of one button leads to deselection of a previously-selected button) to indicate their responses. Each question had an independent, 10-point scale.

Four questions concerning one's personal opinion of a reviewed book were asked for each of the three books tested: (a) How likely would you

be to recommend this book to your friends?; (b) How much would you enjoy reading this book?; (c) How would you judge the quality of this book?; and (d) How likely would you be to buy this book if you were going to buy a novel?

Four questions about one's assessment of public opinion of a reviewed book also were asked for each book: (a) How would the typical reader judge the quality of this book?; (b) How much would the typical reader enjoy reading this book?; (c) How likely would other people be to recommend this book to their friends?; and (d) How well will this book sell?

At the end of the complete hearing session, participants were asked to indicate their general impression about the tested site by clicking one of ten radio buttons beside a list of adjectives. The response scales were anchored by *Describes Very Poorly* (= 1) and *Describes Very Well* (=10). Four adjectives—credible, honest, reliable, trustworthy—were used to measure the credibility of a site.

Seven questions regarding the social presence of multiple sources were asked at the end of the questionnaire[8]: (a) While you were hearing each review, how much did you feel as if each reviewer was talking to you?; (b) While you were hearing each review, how vividly were you able to mentally imagine each reviewer?; (c) When moving from one review to another review, how easily were you able to distinguish one reviewer from another?; (d) After hearing all reviews for a book, how much did you feel as if two or more people had talked to you about the book?; (e) After hearing all reviews for a book, how vividly were you able to mentally imagine all reviewers?; (f) How much attention did you pay to the reviews?; and (g) How much did you feel involved with the reviews?

All indices were analytically distinct and highly reliable: personal opinion of a reviewed book (Cronbach's $\alpha = .93$, .93, and .90 for the three books, respectively), assessment of public opinion about a reviewed book ($\alpha = .93$, .93, and .91, respectively), website credibility ($\alpha = .89$), and feeling of social presence ($\alpha = .81$).

Results

Table 1 shows a full correlation matrix of the measured variables. Personal opinion and assessment of public opinion are highly correlated; however, they are conceptually distinct and therefore analyzed separately.

For the measures that were asked for each book, the experimenters used repeated measure ANOVAs with book as the repeated factor and the number of synthetic voices (one versus multiple) as the between-participants factor. For other items that were asked only once, one-way, between-participants ANOVAs were used. In addition, a path analysis was conducted to test the mediating effect of social presence.

TABLE 1
Correlation Matrix of Measured Variables in Experiment 1

Measured Variable	*1*	*2*	*3*	*4*
1. Social presence		.44**	.46**	.58***
2. Personal opinion (Average of three books)			.88***	.45**
3. Assessment of public opinion (Average of three books)				.39*
4. Website credibility				

NOTE: * $p < .05$, ** $p < .01$, *** $p < .001$ (2-tailed).

Consistent with H1, the differentiation of voices in synthesized speech influenced listeners' imagination of multiple sources. Specifically, users felt a stronger sense of social presence when the reviews were narrated by multiple synthetic voices than by a single synthetic voice (see Table 2).

User responses to multiple synthetic voices manifested the multiple source effect. Consistent with H2, participants evaluated the reviewed books more positively when the reviews were narrated by multiple synthetic voices (Book 1: $M = 5.24$, $SD = 2.25$; Book 2: $M = 6.34$, $SD = 2.26$; Book 3: $M = 5.30$, $SD = 1.67$) than by a single synthetic voice (Book 1: $M = 4.00$, $SD = 2.03$; Book 2: $M = 5.05$, $SD = 2.15$; Book 3: $M = 4.49$, $SD = 2.00$).

The multiple source effect influenced not only users' own personal opinion for reviewed books but also their assessment of other people's feelings about the reviewed books. Participants judged that other people would evaluate the reviewed books more positively when the reviews were

TABLE 2
Comparison of Single Voice vs. Multiple Voices: Experiment 1

Measured Variable	*One Voice Mean* (SD) (N = *20*)	*Multiple Voice Mean* (SD) (N = *20*)	F(*1, 37*)	η^2
Social Presence	2.83 (1.42)	4.26 (1.54)	9.24**	.20
Personal Opinion (Average of three books)	4.51 (1.78)	5.63 (1.71)	4.10*	.10
Assessment of Public Opinion (Average of three books)	4.99 (1.41)	5.93 (1.36)	4.65*	.11
Website credibility	4.29 (1.69)	5.19 (1.54)	3.07+	.08

NOTE: + $p < .10$, * $p < .05$, ** $p < .01$ (all 2-tailed).

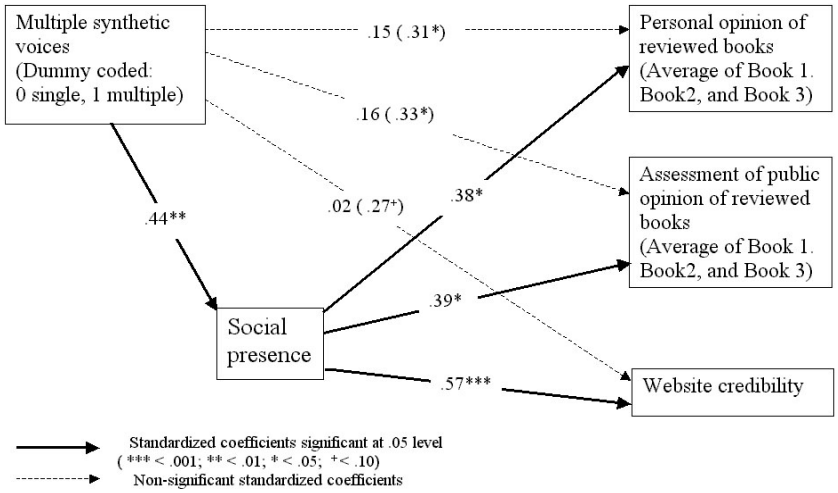

Figure 1. Path Model for Experiment 1.
NOTE: Numbers inside arrows are standardized coefficients for each regression.

narrated by multiple synthetic voices (Book 1: $M = 5.43$, $SD = 1.55$; Book 2: $M = 6.39$, $SD = 1.86$; Book 3: $M = 5.99$, $SD = 1.53$) than by a single synthetic voice (Book 1: $M = 4.25$, $SD = 1.56$; Book 2: $M = 5.56$, $SD = 1.45$; Book 3: $M = 5.16$, $SD = 2.03$).

H4 was marginally supported: Multiple synthetic voice participants perceived the website as marginally more credible than single voice participants.

A path analysis was conducted to test H5, which predicted the mediating effect of social presence on other dependent variables (see Figure 1). In general, there are four criteria necessary to demonstrate mediation (Baron & Kenny, 1986, p. 1177). First, the independent variable (the use of multiple synthetic voices) must have a significant effect on the mediating variable (the feeling of social presence). Second, the mediating variable must have a significant effect on the dependent variables. Third, when the dependent variables were regressed on the independent variable alone, the independent variable must have a significant effect. Finally, when the dependent variables are regressed on both the mediating variable and the independent variable, the effect of the mediating variable on the dependent variables must be significant, while the effect of the independent variable on the dependent variables should decline.

From the path diagram (see Figure 1), one can readily check the first two conditions for mediation. To confirm the third condition, a series of

simple linear regressions were conducted. As can be inferred from the result of the previous analyses, the independent variable (the use of multiple synthetic voices) had significant positive effects on the average of personal opinion of reviewed books ($\beta = .31$, $p < .05$), the average of assessed public opinion of reviewed books ($\beta = .33$, $p < .05$), and website credibility ($\beta = .27$, $p < .09$). These analyses indicated that with the minor exception of website credibility, the data support the third condition for mediation. The final condition for mediation can be confirmed by examining the standardized coefficients reported in the path diagram (see Figure 1). The effect of social presence on the dependent variables was consistently significant, whereas the effect of the independent variable on the dependent variables significantly dropped, even losing significance. In conclusion, the current path analysis provides evidence for the meditating role of social presence in people's social responses to synthetic voices.

Conclusion

A number of conclusions can be drawn from the present results. First, multiple synthetic voices yield higher social presence of virtual speakers than does a single synthetic voice (H1). This result implies that vocal cues of doubly disembodied language influence listeners' imagination of virtual speakers. Synthetic voices thus manifest individuality in the same way as human voices do.

Second, multiple synthetic voices are more persuasive than a single synthetic voice (H2, H3, and H4). These results imply that people socially respond to vocal cues of doubly disembodied language. As a result, a complicated social rule such as the multiple source effect applies even when people hear purely synthetic voices.

Third, the social presence of virtual speakers is the key mediating variable for the multiple source effect in a doubly disembodied language setting (H5). This result implies that people's social responses (i.e., the multiple source effect) to doubly disembodied language (i.e., synthesized speech) are oriented toward imagined virtual speakers.

EXPERIMENT 2

Overview

One compelling possible explanation for the results reported in Experiment 1 is that participants incorrectly believed that the synthetic voices they heard were actually prerecorded human voices or, more convincingly,

distorted versions of prerecorded human voices. This erroneous belief would arise from people's ignorance of the potential of current technologies.

According to this alternative explanation, there is nothing unique about people's social responses to synthetic voices, because people do not interpret the synthetic voices as artificial. Indeed, this type of explanation—technological ignorance—is one of the most popular explanations of people's social responses to artifacts[9] and is frequently used to criticize the Computers Are Social Actors paradigm.

This alternative explanation questions the robustness of social responses to synthetic voices, because it predicts that as soon as people realize the ontological status of synthesized speech, they will stop exhibiting social responses. That is, the more people know about the technological details of an artifact, the less they will apply social rules to that artifact.

To test the validity of this argument, Experiment 2 was conducted, adding a learning manipulation to Experiment 1. The learning manipulation was added in order to make it clear that the voices were synthesized and that the creation of different voices was readily performed. Other than the learning manipulation, Experiment 2 was identical to Experiment 1. Experiment 2, thus, directly answers the question: Can social responses to synthesized speech be eliminated by the later learning of the voice synthesis mechanism?

Hypotheses

We set five hypotheses (H1a, H2a, H3a, H4a, and H5a) by adding the following phrase—"*even when people explicitly know that it is simple and straightforward to create multiple synthetic voices"*— to each of the previously stated hypotheses (H1, H2, H3, H4, and H5) in Experiment 1. For example, H1a reads "People will feel stronger social presence of multiple sources when they hear testimonials of a book via multiple synthetic voices than when they hear the same testimonials via a single synthetic voice, even when people explicitly know that it is simple and straightforward to create multiple synthetic voices." If the learning manipulation eliminates participants' social responses, these new hypotheses would not be supported.

Method

Learning Manipulation

Two procedures were used to maximize the impact of the learning manipulation. First, participants read the following brief description of how TTS works before they started the experiment:

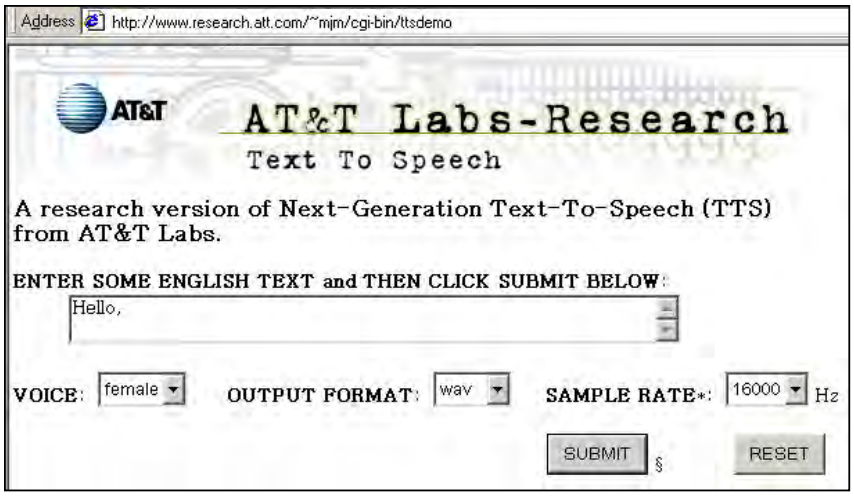

Figure 2. TTS Demo Site.

> In this experiment, you will hear *machine* voices synthesized by a TTS (Text-to-Speech) system. TTS is the creation of audible speech from computer readable text. Many TTS systems are currently available. Most of them can generate more than one type of voices from any given text by altering speech rate, fundamental frequency (F0), loudness, and other vocal parameters.

Also before the main experiment, participants logged on the TTS demo site of AT&T Research Lab (see Figure 2), typed a sentence, and listened to the possible variations of synthesized speech for that sentence. This was done so that they could understand that one can easily produce many different synthetic voices without using prerecorded human speech.

Participants

Forty college undergraduate (20 women and 20 men) enrolled in a communication class participated in Experiment 2. Participants were randomly assigned to one of the conditions. Gender was balanced across conditions. All participants signed informed consent forms and were debriefed at the end of the experiment session.

Procedure

As explained above, participants were instructed to read a brief description of the TTS technology and to access a TTS demo site before starting the experiment. After experimenting with the possible variations in

TABLE 3
Correlation Matrix of Measured Variables in Experiment 2

Measured Variable	*1*	*2*	*3*	*4*
1. Social presence		.56***	.61***	.34*
2. Personal opinion (Average of three books)			.75***	.40*
3. Assessment of public opinion (Average of three books)				.46**
4. Website credibility				

NOTE: * $p < .05$, ** $p < .01$, *** $p < .001$ (2-tailed).

TTS (by experiencing the ability to create multiple voices), participants logged on to the same websites used in Experiment 1 and followed the same procedures as in Experiment 1.

Measures

The experiment used the same measures as in the first experiment. Again, all indices were analytically distinct and highly reliable—personal opinion of a reviewed book (Cronbach's α =.94, .90, and .88 for the three books, respectively), assessment of public opinion about a reviewed book (α =.90, .90, and .94, respectively), website credibility (α = .91), and the feeling of social presence (α = .79).

Results

Experiment 2 used the same analysis strategy as in Experiment 1. Table 3 shows a full correlation matrix of the measured variables.

For the measures that were asked for each book, repeated measure ANOVAs were served with book as the repeated factor and the number of synthetic voices (one versus multiple) as the between-participants factor. For other items that were asked only once, one-way, between-participants ANOVAs were used. In addition, a path analysis was conducted to test the mediating effect of social presence (see Table 4).

Consistent with H1a, multiple synthetic voice participants felt more social presence than single voice participants, even when they explicitly knew the artificial nature of synthetic voice(s) they heard.

Confirming H2a, participants evaluated the reviewed books more positively when the reviews were narrated by multiple synthetic voices (Book 1: M = 5.16, SD = 2.30; Book 2: M = 6.21, SD = 2.02; Book 3: M = 5.54, SD = 1.64) than by one single synthetic voice (Book 1: M = 4.24, SD = 1.54; Book 2: M = 5.12, SD = 1.73; Book 3: M = 4.63, SD = 1.78), even after they explicitly learn the artificiality of synthetic voice.

TABLE 4
Comparison of Single Voice vs. Multiple Voices: Experiment 2

Measured Variable	*One Voice Mean* (SD) (N = *20)*	*Multiple Voice Mean* (SD) (N = *20)*	F*(1, 37)*	η^2
Social Presence	2.69 (1.16)	4.27 (1.43)	14.86**	.28**
Personal Opinion (Average of three books)	4.66 (1.23)	5.64 (1.76)	4.14*	.10*
Assessment of Public Opinion (Average of three books)	5.30 (1.03)	6.08 (1.26)	4.57*	.11*
Website credibility	4.98 (2.13)	4.71 (2.07)	0.16	.00

NOTE: $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$ (all 2-tailed).

Participants judged that other people would evaluate the reviewed books more positively when the reviews were narrated by multiple synthetic voices (Book 1: $M = 5.69$, $SD = 1.70$; Book 2: $M = 6.56$, $SD = 1.68$; Book 3: $M = 5.98$, $SD = 1.42$) than by one single synthetic voice (Book 1: $M = 4.39$, $SD = 1.95$; Book 2: $M = 6.00$, $SD = 1.23$; Book 3: $M = 5.53$, $SD = 1.46$). This result confirmed H3a.

H4a was not supported. There was no significant difference between the multiple synthetic voices after learning condition and the single voice condition with regard to website credibility.

Similar to Experiment 1, a path analysis was conducted to test H5a. Figure 3 illustrates the results. This path diagram (see Figure 3) shows the mediating role of social presence in people's social responses to synthesized speech, even after the learning manipulation. The path model and the results of regression analyses confirm H5a.

The use of multiple synthetic voices was a significant predictor for social presence ($\beta = .53$, $p < .01$), personal opinion of reviewed books ($\beta = .31$, $p < .05$), and assessment of public opinion of reviewed books ($\beta = .33$, $p < .05$), when it was the only predictor entered into a series of regression equations. The use of multiple synthetic voices, however, was not a significant predictor for website credibility ($\beta = -.06$, *n.s.*). Social presence was a significant predictor for all other dependent variables when it was the only predictor (personal opinion of reviewed books [$\beta = .57$, $p < .001$]; assessment of public opinion of reviewed books [$\beta = .61$, $p < .001$]; and website credibility [$\beta = .34$, $p < .05$]).

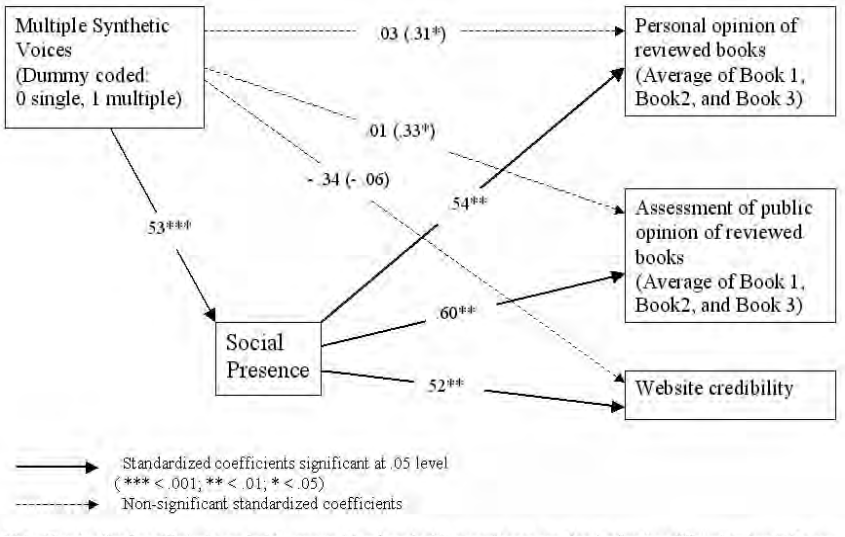

Figure 3. Path Model for Experiment 2.
NOTE: Numbers inside arrows are standardized coefficients for each regression.

When both the independent (the use of multiple synthetic voices) and the mediating (social presence) were used as predictors for a series of multiple regression analyses, the same patterns as was observed in Experiment 1 emerged. First, the effect of social presence remained significant for all three dependent variables (personal opinion of reviewed books [$\beta = .54$, $p < .01$]; assessment of public opinion of reviewed books [$\beta = .60$, $p < .01$]; website credibility [$\beta = .52$, $p < .01$]). Second, the effect of the independent variable on the personal opinion ($\beta = .03$, n.s.) and assessment of public opinion variables ($\beta = .01$, n.s.) were no longer significant. For website credibility ($\beta = -.34$, n.s.), the effect remained nonsignificant.

In conclusion, these results provide evidence for the meditating role of social presence in people's social responses to synthetic voice even when people explicitly know the purely artificial nature of synthetic voice through learning and self-trials.

Conclusion

Experiment 2 successfully replicated the results of Experiment 1 in a situation in which participants explicitly learned the artificiality of the stimuli just before engaging in the main experiment. The key finding of

Experiment 2 is that technological literacy does not preclude social responses to synthetic voices. That is, even when people clearly know the artificial nature of synthetic voices, they nonetheless respond socially to synthetic voices. Experiment 2, thus, effectively weakens the general alternative explanation that social responses to media simply come from people's ignorance of the artificial nature of technologies that they are using.

GENERAL CONCLUSIONS AND DISCUSSION

Based on the results of the two experiments, the following general conclusions emerge:

1. Synthetic voice influences listeners' imagination of virtual speaker.
2. People respond socially to synthetic voice.
3. The use of multiple synthetic voices to represent multiple sources increases the persuasive power of synthesized speech.
4. The effects of multiple voice presentations on persuasion are mediated by feelings of social presence.
5. Technological literacy does not eliminate social responses to synthesized speech.

The following implications of the above findings reflect four different areas—SUI design, multiple source effect research, media equation theory, and ethics.

Implications for Speech User Interface Design

The fact that the multiple source effect applies to synthetic voices has direct applicability to the design of voice-based commercial systems, especially when these systems provide multiple reviews of a product/service. For example, designers of a SUI system can increase the persuasive power of communications by simply providing various synthetic voice types rather than choosing a default synthetic voice selected by a TTS vendor. This is a very cost-effective way to increase persuasion, because producing different synthetic voices with a TTS engine is very easy and inexpensive. An even better approach would be to use more than one TTS engine, thereby producing very different voices because of differences in the synthesis algorithms. The cost for licensing more than one TTS engine would be nominal in comparison to the persuasive power gained.

The effect sizes are relatively weak for website credibility; however, the findings of the current study indicate that a website which portrays testimonial reviews might be more trusted if it used multiple TTS voices to portray different sources. In addition to e-commerce sites, political

campaign sites, religious sites, and any sites in which trustworthiness of the sites is critical would become more credible if multiple voices were used to present various opinions.

Another possible implication of the current study for speech user interface design is that TTS designers should be careful when converting text manifesting distinctive personal characteristics. It may be preferable to match linguistic cues in text with the vocal cues in synthesized speech, because both cues affect listeners' imagination of virtual speakers. People do not discount the influence of vocal cues on the mental image of an actual source simply because the voice is artificial and is disconnected from the actual source who wrote the text (Nass & Gong, 2000). This is true even when people understand and are reminded of the technologies for producing these synthetic voices. Casting of synthetic voices should therefore be done as carefully as casting of real human voices. In fact, previous work has empirically shown that if the language and vocal cues are mismatched, people will be confused, less convinced, and have negative feelings about the synthesized speech they heard (Nass & Lee, 2001).

Implications for Multiple Source Effect Research

The current study provides empirical evidence that the multiple source effect can be induced by manipulating vocal cues. Previous literature (Harkins & Petty, 1981, 1983, 1987; Moore & Reardon, 1987; Moore et al., 1994) on the multiple source effect relied heavily on the manipulation of linguistic cues (i.e., textual content), probably because they were easy to manipulate. By providing evidence that the manipulation of vocal cues can also induce the multiple source effect, the current study expands the domain of modality (from text to audio) to which the multiple source effect applies.

Another contribution of the current research to the study of the multiple source effect is that it discovers a key mediating variable—social presence—of the multiple source effect in doubly disembodied language situations. It would be interesting to determine if social presence also plays a mediating role in human-human interaction situations.

Implications for the Media Equation Theory

The current study has three major implications for media equation theory (Reeves & Nass, 1996). First, the current study provides the most direct evidence supporting the Computers Are Social Actors paradigm (Nass & Moon, 2000). Even though previous Computers Are Social Actors studies showed that people consistently apply social rules to computers (or other simulation technologies), they did not directly test their fundamental assumption that computers or other simulation technologies

can manifest individuality. The results of the current study clearly indicate that people assign individuality to a synthetic voice in the same way as they would assign individuality to a real human voice, despite the fact that a synthetic voice is merely a computer-generated artifact, which cannot have individuality.

Second, the current study provides counterevidence to the technological ignorance argument which has been a strong challenge to media equation theory, and especially the Computers Are Social Actors paradigm (Nass & Moon, 2000). This argument criticizes media equation theory by saying that participants in a typical experiment might not know or remember the artificiality of their interaction and incorrectly believe that they are interacting with other humans. The current study directly tests the argument by administrating a condition in which participants explicitly learn and experience voice synthesis technology before the main experiment. The finding that technological literacy does not eliminate social responses to synthetic voices thus provides strong evidence for the robustness of the media equation phenomenon.

Another contribution of the current study is that it shows that social responses to synthesized speech are oriented toward imagined virtual actors rather than a computer which provides the speech. This is confirmed by the result that social presence of virtual speakers is the key mediating variable for the social response tested. This result confirms the conclusion of Reeves and Nass (1996) that computers are as many social actors as there are voices, especially when computers provide a playground for multiple voices. This result, however, does not necessarily mean that social responses to media and technology are always oriented toward imagined virtual speakers. It might be the case that social responses are oriented toward a technology itself, especially when the technology takes either an actual or a virtual form (e.g., artificial agents, or robots). In fact, most previous Computers Are Social Actors studies imply (though do not directly test) that social responses to a technology having a physical form are oriented toward the technology itself. A reasonable conclusion is that social responses to a technology are oriented toward imagined virtual actors when the language used by the technology is disembodied.

Ethical Implications

The human tendency to respond socially to an artifact possessing human-like characteristics can be either good or bad. Thanks to this tendency, people can enjoy movies, even though they clearly know the artificiality of filmed objects after more than 100 years of exposure to film technologies. By the same token, however, people are at the risk of being

unknowingly influenced by technology even when they are clearly aware of its artificiality (see Experiment 2).

The current study poses a potential ethical issue with regard to the use of synthetic voices. For example, it might be possible to deceive consumers into believing that a product has received many positive reviews from different sources by using multiple synthetic voices, while in fact a single reviewer wrote all the reviews. As the naturalness of synthetic voice improves, it will become even more difficult to differentiate an authentic human voice from a synthetic version of it. Also, it will become possible to create pseudocelebrity endorsements by using a synthetic imitation of a celebrity voice. Clearly, the ethical issues of using synthetic voices (or any other artifacts) to increase persuasive impacts of messages should be considered and discussed.

LIMITATIONS OF THE CURRENT STUDY AND SUGGESTIONS FOR FUTURE RESEARCH

There are key limitations in the present research. First, both experiments in the current study measure only attitudinal responses. Focusing on attitudinal responses is an efficient way to detect possible effects of social responses; however, it would have been useful to include actual buying behaviors by, for example, distributing real money or credit to participants in order to measure their actual buying behaviors. The use of this paradigm would increase the external validity of the current study.

Second, the current study did not measure participants' memory of narrated contents. One of the key theoretical claims of the multiple source effect literature is that people process information from a different source in a more diligent way than information from the same source. If this is valid, then the diligent cognitive processing would lead to better memory of information provided by multiple (thus, different) sources than information provided by a single source. The incorporation of memory measures to the current study, therefore, would make it possible to test empirically the key theoretical claim of the multiple source effect literature. In addition, the incorporation of memory measures would make it possible to empirically investigate the relationship between social presence and memory.

Third, the current study did not separately analyze the effect of each vocal cue on listeners' imagination of virtual speakers. As a first step to the inquiry into human responses to doubly disembodied language, the priority was to test the effects of vocal cues on social presence and social responses. Consequently, the experimenters manipulated various vocal parameters at the same time to maximize the differences among the

multiple synthetic voices. As a result, the current study does not clarify relative impacts of each vocal cue and possible dynamics between and among them, nor does it identify the relative differentiability of various vocal cues. Future studies should address this issue.

Finally, in both experiments, only college students were recruited as participants. The convenience sample of college students weakens the generalizability of the current study. Future studies should replicate the current study with participants recruited from the general public.

FINAL REMARKS

As TTS technologies develop, it may not be possible to differentiate real human voices from synthetic ones. Will the conceptualization of doubly disembodied language be still valid even in that situation? The answer is yes. In a situation when synthetic voices and real voices are perceptually the same, the media will be very likely to develop a cognitive system (e.g., labeling and disclaimer) to emphasize the artificial nature of perfectly natural synthetic voices. Both first and second degree disembodiment will therefore still exist even when we hear perfectly natural synthetic voices.

Is there any value in theorizing about human responses to doubly disembodied language even when no perceptual difference exists between real and synthetic voices? Again, the answer is yes. Without a theoretical understanding of doubly disembodied (or other new types of) languages, designers, advertisers, policy makers, educators, and consumers of mediated messages will not be able to make informed decisions on how to produce or consume doubly disembodied messages delivered by perfectly natural synthetic voices. Theorizing about doubly-disembodied speech, then, will continue to be important for our understanding of new media and advanced simulation technologies.

APPENDIX 1

Book Titles, Author Names, and Review Scripts Used in Experiment 1 and Experiment 2

Book 1:
Plainsong (Kent Haruf)
Customer Reviews:
Customer A: The pace of the story mimics that of the small town it takes place in . The characters are richly drawn, but not caricatures. Not a lot happens, but I believe that's the point. There is no urgency to the story, and I liked it that way. The author, like James Lee

Burke, has an affection for beauty of the land. Reading this book is like taking a stroll down the main street of Holt County, in which the story takes place. Highly recommended.

Customer B: I loved this book. I stayed up all night reading it because I cared so much about the people, I could not put it down. I haven't done that since I was a child. After a lifetime of reading, I'm aware how rare this kind of experience is. And the reason? Kent Haruf's honesty, skill, and compassion as a writer.

Customer C: I really enjoyed this book. It starts slowly with disconnected stories. As it builds character development and speed, the stories become intertwined. I like stories like this anyway, but this is better than most. The descriptions of the plains, weather, and other sensations are real. The characters are people you want to take home with you. A good read.

Customer D: I enjoyed this book more than expected. The story was a little slow in the beginning, but the characters begin to come together fairly quickly. Although there are no quotations marks, the book is an easy read. It also holds your attention throughout and leaves you feeling good. I would recommend it.

Customer E: This story had me engrossed the minute I started it. It has complex characters and it has simple characters. Loved the book very much! Don't miss the chance to read this one.

Book 2:

The Family Orchard (Nomi Eve)

Customer Reviews:

Customer A: How did a thirty two year old manage to mature enough to write a masterpiece? This book is a wonder to read, never boring and just breathtakingly good.

Customer B: It is hard to believe that this is a debut novel. Eve paints vivid characters that you can reach out and touch. The characters experience the glory of everyday life from the mundane to the extraordinary, from love and sex to death and sorrow. Each generation in this family "tree" captures a different facet of this experience we call life. This is a "big" book, Ms. Eve has masterfully succeeded in creating a new way to tell a story that everyone can relate to. Tell your family and friends that they will be rewarded many times over when they read The Family Orchard.

Customer C: This book held my interest from beginning to end. Ms. Eve makes everything come to life with a remarkable command of the language. She imparts feelings in such a way that you can really immerse yourself in the joys and sorrows of her characters. She made me laugh and she made me cry. The historical background was also beautifully done and the illustrations appropriate and fascinating.

Customer D: Once in a while there are books that change your life. I don't want to give away too much of the story—the regular review does a good job summarizing it—but I do want to say that I almost cried when I finished it. I wanted it to go on and on. The people Ms. Eve mentions in her book have become part of me. I hope to hear from them again.

Customer E: What an enjoyable read! Now I want to go out and find those hidden stories about my own family, although I doubt they could be as interesting as these ancestors. No wonder family tree research has become such a popular pastime. I just hope we see more from this talented author.

Book 3:

The Sheltering Sky (Paul Bowles)

Customer Reviews:

Customer A: I spent two days consumed by this book. It is magical, spiritual, depressing, and enlightening. It is relatively simple story of three Americans, and the physical and psychological trauma that befalls them. The ending rather shocked me, and ended on a painful note, but I thoroughly enjoyed it.

Customer B: This is one of the best books I have ever read. The characters are completely flushed out. Bowles doesn't miss a single emotion that the characters are experiencing. With a backdrop of a stifling hot desert we are taken on a dizzying journey of emotional deconstruction.

Customer C: To my mind, the main character of this outstanding novel is the North African postwar desert. Three American "travelers" (not "tourists") roam the desert whose starkness makes the psychological travail all the more dramatic. You can taste the sand. Great book.

Customer D: Bowles takes the reader into the deep desert and psyche of his characters. From the first incredible page, his images and characters are rendered in flawless prose. One of the most poignant, memorable books I've read. Highly recommended.

Customer E: In this book of three American travelers who journey through North Africa, Bowles shows us, with gripping yet subtle tones, how rigid is our comprehension of foreign culture, and how incomplete is our knowledge of ourselves. It is a novel for the mind. As the journeyers separate, first from each other then from their own sanity, we understand how delicate our grip on reality is, especially when faced with the awesome spectacle of untouched nature. Pick it up and begin a journey into yourself you will never forget.

APPENDIX 2

CSLU Toolkit TTS Parameters Chosen for Manipulation

	Name	*Pitch*	*Pitch Range*	*Speech Rate*
Type A	mwm	115 Hz	20 Hz	0.95
Type B	rab	105 Hz	14 Hz	0.94
Type C	ked	106 Hz	15 Hz	0.91

NOTE. This is a speech rate determined by the CSLU toolkit interface. "1.0" is 120 words per minute. Therefore, ".95" means 114 (=.95 * 120) words per minute, ".94" means 113 words per minute, and ".91" is 109 words per minute.

NOTES

1. In his paper, Clark used the term "mechanized speech" to refer to pre-recorded television shows, recorded telephone messages, books on tape, and pre-recorded fire alarms. That is, he used the term "mechanized" in the sense that real human speech is *recorded*. To eliminate possible confusion between his term "mechanized speech" and this study's term "computer-synthesized speech," the latter used the term "prerecorded human speech" instead of Clark's "mechanized speech."

2. The imagination of known speakers is a different topic. In the case of a known speaker, people automatically imagine the previously-established image of the speaker.

3. Listeners almost automatically recognize the artificiality of a synthesized speech, because even the best TTS systems have not yet achieved the quality and prosody of natural human speech (Kamm, Walker, & Rabiner, 1997). In fact, the quality of TTS is so low that most speech user interface deployments use pre-recorded human speech in their systems whenever possible, despite the additional cost of recording real human speech.

4. "TTS," standing for "text to speech," is the traditional abbreviation for "synthesized speech."

5. The toolkit is made by the Center for Spoken Language Understanding at Oregon Graduate Institute. It is freely downloadable at http://www.cslu.ogi.edu/.

6. The Bell Labs TTS engine demo site is no longer available due to the termination of speech synthesis research at Lucent Technology in November 2002.

7. Visit http://www-rcf.usc.edu/~kwanminl/research/multiTTS/ to hear each of the voices and a sample file of the multiple synthetic voice presentation.

8. There are many previous studies measuring social presence (e.g., Biocca, Burgoon, Harms, & Stoner, 2001; Nowak, 2000; Short, Williams, & Christie, 1976; Whitmer, & Singer, 1998). With the exception of the scale used in Lee and Nass (2001), however, most self-report social presence scales used in previous studies cannot be properly applied to the context of speech user interface. A new social presence scale was constructed based on Lee and Nass (2001), which is more applicable to speech user interfaces.

9. Nass and Moon (2000) categorized this argument as an "anthropomorphism"-based explanation of people's social responses to artifacts.

REFERENCES

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research, *Journal of Personality and Social Psychology, 51,* 1173–1182.

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

Biocca, F., Burgoon, J., Harms, C., & Stoner, M. (2001, May). *Criteria and scope conditions for a theory and measure of social presence*. Paper presented at Presence 2001, Philadelphia, PA.

Bleich, D. (1981). The individual language system. *ADE Bulletin, 68*, 5–8.

Cassells, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational agents*. Cambridge, MA: MIT Press.

Clark, H. H. (1996). *Using language.* New York: Cambridge University Press.

Clark, H. H. (1999). How do real people communicate with virtual partners? Proceedings of 1999 AAAI (American Association for Artificial Intelligence) Fall Symposium: Psychological Models of Communication in Collaborative Systems, 43–47.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595.

Debus, G. (1978). *Über Wirkungen akustischer Reize mit unterschiedlicher emotionaler Valenz*. Meisenheim, Germany: Hain.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. Barkow, L. Cosmides, J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). New York: Oxford University Press.

Harkins, S. G., & Petty, R. E. (1981). The effects of source magnification on cognitive effort and attitudes: An information processing view. *Journal of Personality and Social Psychology, 40*, 401–413.

Harkins, S. G., & Petty, R. E. (1983). Social context effects in persuasion: The effects of multiple sources and multiple targets. In P. Paulus (Ed.), *Advances in group psychology* (pp. 149–175). New York: Springer/ Verlag.

Harkins, S. G., & Petty, R. E. (1987). Information utility and the multiple source effect in persuasion. *Journal of Personality and Social Psychology, 52*, 260–268.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence, 1*, 262–271.

James, F. (1998). *Representing structured information in audio interfaces: A framework for selecting audio marking techniques to represent document structures.* Unpublished doctoral dissertation, Stanford University, Palo Alto, California.

Kamm, C., Walker, M. & Rabiner, L. (1997, February). *The role of speech processing in human-computer intelligent communication*. Paper presented at NSF Workshop on human-centered systems: Information, interactivity, and intelligence. Retrieved February 22, 2002, from http://www.ifp.uiuc.edu/nsfhcs/talks/rabiner.html

Keppel, B. (1982). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.

Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotypes. *CHI 2000 Extended Abstracts,* 289–290.

Lee, K. M. (2002). *Social responses to synthesized speech: Theory and application*. Unpublished doctoral dissertation, Stanford University, Palo Alto, California.

Lee, K.M., & Nass, C. (2001, May). *Social presence of social actors: Creating social presence with machine-generated Voices*. Paper presented at Presence 2001: 4th Annual International Workshop, Philadelphia, PA.

Moore, D. J., & Reardon, R. (1987). Source magnification: The role of multiple sources in the processing of advertising appeals. *Journal of Marketing Research, 24*, 412–417.

Moore, D. G., Mowen, J.C., & Reardon, R. (1994). Multiple sources in advertising appeals: When product endorsers are paid by the advertising sponsor. *Journal of the Academy of Marketing Sciences*, *22*, 234–243.

Morishima, Y., Nass, C., Bennett, C., & Lee, K. M. (2001). Effects of "Gender" of Computer-Generated Speech on Credibility. *Technical Report of IEICE TL2001–16, 31*, 557–562.

Nass, C., & Gong, L. (2000). Social aspects of speech interfaces from an evolutionary perspective: Experimental research and design implications. *Communications of the ACM, 43*, 36–43.

Nass, C., & Lee, K. M. (2001). Does computer-generated speech manifest personality?: Experimental test of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology, Applied, 7*, 171–181.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*, 81–103.

Nowak, K. (2000). *The influence of anthropomorphism on mental models of agents and avatars in social virtual environments.* Unpublished doctoral dissertation, Michigan State University, East Lansing.

Perloff, R. M. (1993). Third-person effect research, 1983–1992: A review and synthesis. *International Journal of Public Opinion Research, 5,* 167–184.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* New York, NY: Cambridge University Press.

Sawhney, N., & Schmandt, C. (1997, November). *Design of spatialized audio in nomadic environments*. Paper presented at the International Conference on Auditory Display (ICAD'97). Retrieved January 11, 2002, from http://www.media.mit.edu/~nitin/NomadicRadio/ICAD97/ICAD97_paper/ICAD97.html

Scherer, K. R. (1979). Personality markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 147–209). Cambridge, UK: Cambridge University Press.

Shneiderman, B. (1997). *Designing the user Interface: Strategies for effective HCI (3rd ed.)*., Reading, MA: Addison-Wesley.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* London: Wiley.

Whitmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environment, 7*, 225–240.