Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction

Clifford Nass
Department of Communication
Stanford University
Stanford, CA 94305-2050
+1 650-723-5499
nass@stanford.edu

Kwan Min Lee
Department of Communication
Stanford University
Stanford, CA 94305-2050
+1 650-497-7357
kmlee@stanford.edu

ABSTRACT
This study examines whether people would interpret and respond to paralinguistic personality cues in computer-generated speech in the same way as they do human speech. Participants used a book-buying website and heard five book reviews in a 2 (synthesized voice personality: extrovert vs. introvert) by 2 (participant personality: extrovert vs. introvert) balanced, between-subjects experiment. Participants accurately recognized personality cues in TTS and showed strong similarity-attraction effects. Although the content was the same for all participants, when the personality of the computer voice matched their own personality: 1) participants regarded the computer voice as more attractive, credible, and informative; 2) the book review was evaluated more positively; 3) the reviewer was more attractive and credible; and 4) participants were more likely to buy the book. Match of user voice characteristics with TTS had no effect, confirming the social nature of the interaction. We discuss implications for HCI theory and design.

Keywords
TTS (Text-to-Speech), CASA (Computers are social actors), Speech User Interfaces, personality, similarity-attraction effect.

INTRODUCTION
In recent years, there has been increasing interest in and use of Text-to-Speech (TTS) technologies (see [13] for a review). This is due to several factors. First, there has been growing demand for Speech User Interfaces (SUIs). SUIs are needed as an alternative/complement to Graphical User Interfaces (GUI), because GUIs have several limitations (see [19] for a review). For example, GUIs are not appropriate for hands-occupied or eyes-occupied environments such as driving [18]. GUIs also cannot be effectively implemented on technologies such as telephones or PDAs, because of screen-size limitations [5].

Second, TTS systems are useful for presenting text (via synthesized speech) and voice on telecommunication devices, such as cellular phones [6].

Third, for some people, it is hard or even impossible to use GUI-based systems. For example, about 11 million people have some form of visual impairment and about 1.5 million are totally blind in the U.S. alone [5]. TTS technologies provide a new opportunity to use the power of the computer and the Internet for the disabled. In addition, TTS technologies also make it easier for the illiterate and pre-literate (e.g., children) to use computers.

Finally, TTS systems are increasingly used to develop more human-like interfaces. The basic assumption is that by incorporating anthropomorphic indicators such as speech, users will feel more comfortable with computers and perceive them as more intelligent [10, 11, 14, 17].

Assessment Criteria for TTS
In general, TTS systems are evaluated according to two dimensions: 1) intelligibility and 2) naturalness of the resulting speech [1, 6]. There is relatively little difference in intelligibility across systems [1]; indeed, word intelligibility scores for the best TTS systems are close to 97%, suggesting that the intelligibility of the best TTS systems approaches that of real human speech [6].

Naturalness scores for the best TTS systems are in the fair-to-good range, indicating that even the best TTS systems still do not match the quality and prosody of natural human speech [6]. Consequently, the speech generated by current TTS systems would never be confused with real human speech, even though it is clearly understandable [10, 13].

Given the un-naturalness of synthesized speech, it would seem absurd for users to respond to speech characteristics

CHI '2000 The Hague, Amsterdam

that suggest essentially human attributes. That is, although it would not be surprising for users to be influenced by the comprehensibility of speech (e.g., difficult-to-understand speech leading to greater frustration) or the naturalness of speech (e.g., more natural speech leading to greater feelings of comfort and social presence [10]), the obviously synthetic nature of TTS should make the non-verbal aspects that influence *social* assessment irrelevant to users' attitudes and behaviors. For example, odd hesitations in human speech are interpreted as attempts at deception [7]. These inappropriate pauses are a ubiquitous aspect of all synthesized speech engines, but it is unlikely that users would respond to all TTS engines as highly deceptive.

Even the literature that most dramatically demonstrates that individuals consistently apply social rules to computers (although users know that it is absurd to do so) — the Computers are Social Actors (CASA) paradigm [11,17] — cannot be directly applied to assignment of human attributes for systems employing text-to-speech.

In traditional CASA experiments, the user is presented with a computer that does nothing to remind the user that it is *not* human or, more specifically, that it is an ersatz human. For example, the vast majority of studies in the CASA paradigm involve plain text presenting commonplace language and omitting any behavior uniquely associated with computers (e.g., error messages, "crashing,"). Thus, the content produced is as consistent with computer-mediated communication as it is with human-computer interaction. Even the use of the word "I" is eschewed in the experiments to ensure that the user is not reminded that there is an incongruity in a machine using the term for human identity.

In contrast to the failure to mark non-humanness in the previous CASA studies, TTS is a "liminal" case [21] which sits at the boundary of human and machine, although it is clearly on the machine side of the divide. A demonstration that users employ social cues even with highly-present reminders that the voice is not coming from a person would represent a critical test of the CASA framework, as well as having critical implications for HCI design.

PERSONALITY MARKERS IN SPEECH

Among humans, paralinguistic aspects of speech convey a wide range of trait information about the speaker, such as gender and age. In the present study, we attempt to determine whether the vocal characteristics of speech can convey "personality." That is, given identical content, will the particular settings in a TTS engine lead individuals to identify and respond to the voice as if it had a personality?

We chose personality for a number of reasons. First, personality is more subtle and complex than the more obvious characteristics of gender and age. Second, there is a large literature demonstrating that individuals will assess computers according to their personality [4,11,12,17], although this literature has focused primarily on the message's linguistic content [see 4 for an exception]. Third, the term "personality"[2] comes from the Latin *personare*, to *sound* through, referring to the mouth opening in a mask worn by an actor [2]; thus, it is natural to focus on the "sound" characteristics of personality. Fourth, the field of personality psychology provides a rich set of predictions concerning how individuals will respond to various personality types. Thus, we are not limited to merely determining whether individuals can *identify* the personality of a TTS voice; instead, we can also determine whether these personality markers will influence users' attitudes and behaviors. Fifth, personality is an important and easily-measured individual difference among users, so the parallels to computers, as well as the interaction between user personality and computer personality, become particularly interesting. Finally, and most importantly, there is a significant literature that outlines the specific linkages between the characteristics of the vocal channel and the assignments of personality that result.

Numerous dimensions of personality have been identified [3, 12, 17, 20, 23]. The present study focuses on the extroversion/introversion dimension, both because it is the dimension most strongly marked by paralinguistic cues and because it has proven to be important in the HCI literature [4, 12,17].

Four readily-manipulable aspects of TTS engines are associated with characteristics that individuals use to distinguish introversion from extroversion in human-human interaction:

Speech rate: Extroverts speak more rapidly than introverts [15, 20, 24].

Volume: Extroverts speak more loudly than introverts [15].

Pitch: Extroverts speak with higher pitch than introverts [15]

Pitch range: Extroverts speak with more pitch variation than do introverts [3].

To ensure a successful manipulation, consistency among the features, and consistency with humans, we manipulated all four attributes simultaneously.

THEORETICAL EFFECTS OF TTS PERSONALITY

To determine whether paralinguistic cues in TTS lead users to respond to the voice as if it manifests a "personality," we used three different approaches. First, we determined whether users could simply identify whether the voice was "introverted" or "extroverted." This assessment not only involved direct questions about the voice, but also whether the personality of the voice affected assessments of the *content* that the voice presented. This is a much stronger test than the direct test, in that it requires users to think of the voice as the *source* of the content.

A yet more stringent test is to determine whether the users draw on social rules to respond to the voice. In this study, we examined the principle of similarity-attraction.

According to this principle, people like others who possess a personality that is similar to their own [12, 17]. Similarity-attraction has proven to be a very robust finding when a computer's personality is manifest through verbal/textual cues [e.g., 4, 12, 17]. However, text-to-speech is so obviously *not similar* to a human voice that to be able to find second-order similarity effects would be remarkably strong evidence for the assignment of personality to TTS.

SIMILARITY-ATTRACTION EXPERIMENT

This study was executed in the context of a book-buying website site (an example can be found at www.stanford.edu/class/comm169/Kwan/1-1.fft). Each web page had an identical visual interface based on Amazon's book description. Each page included the titles, the authors (in text) and the pictures of five books. Instead of having the book description in text, there was a link to an audio (.wav) files; clicking on the link would play the review

Hypotheses

Based on the idea that paralinguistic cues will be used to assess personality even in TTS, we can draw the following conclusions.

H1: Users will recognize vocal cues indicating personality even when they hear computer-synthesized speech. TTS that has a speech rate, volume, pitch, and pitch rate associated with extroversion or introversion in humans will be perceived to be extroverted or introverted, respectively.

Although the correlation between people's voices and their personality is not very strong [see below], individuals nonetheless use voice characteristics to assess personality, especially when other cues are absent. Despite the fact that all of the individuals in our experiment understood that the book descriptions were written by actual people, the tendency toward proximate source orientation [see 17, chap. 16 for a review] may lead users to assign the characteristics of the voice to the *writer* of the review, especially when the text is not strongly revealing of the author's personality. That is, absent other cues, individuals decide on the attributes of the distant source based on the most-readily available cues [17, see Ch. 16]. This leads to the following hypothesis:

H2: Users will infer the personality of a review writer based on the personality of the TTS voice they heard.

The principle of similarity-attraction is not only powerful; it is generalized from mere attraction to a variety of other positive attributes. Thus, we can derive the following hypotheses:

H3: Users will be more attracted to a TTS voice that exhibits similar personality to their own than a TTS voice that exhibits a dissimilar personality.

H4: Users will regard a TTS voice that exhibits similar personality to their own as more credible than a TTS voice that exhibits a dissimilar personality.

If we combine the literature on similarity-attraction and proximate source orientation, then the match between TTS personality and user personality should affect the reactions to the reviewer and the review itself, rather than merely affecting direct assessments of the voice. This leads to the following hypotheses:

H5: Users will evaluate a book review more positively if the review is narrated by a TTS voice that matches their own personality.

H6: Users will like a review writer more if the writer's review is narrated by a TTS voice that matches their own personality.

H7: Users will regard a review writer as more credible if the writer's review is narrated by a TTS voice that matches their own personality.

Finally, from a web design perspective, one would like to influence behavior as well as attitude. Given the above, we can predict:

H8: Users will show more buying intention for a book whose review is narrated by a TTS voice that matches their own personality.

Method

Participants

Several weeks prior to the study, a web-based personality survey was administered to students who registered for a large introductory communication course. Both Myers-Briggs (see [9] for review) and Wiggins [23] personality tests were administered to maximize the likelihood of correctly assessing the personality of students. From a total of approximately 150 undergraduate students, a total of 72 participants—36 extrovert and 36 introvert students—who had English as a first language were invited to participate in the study. Participants were randomly assigned to condition, with gender approximately balanced across conditions. All participants signed informed consent forms, were debriefed at the end of the experiment session, and received class credit for their participation.

Procedure

The experiment was a 2 (computer voice personality: extrovert vs. introvert) by 2 (participant personality: extrovert vs. introvert) balanced, between-subjects design, with five book descriptions as a repeated factor. Upon arrival to the lab, each participant was assigned to a computer equipped with a headphone and an Internet Explorer 4.0 browser. Participants were instructed to use the headphones during the whole experiment and not to adjust the volume level of either the headphone or the computer.

The experimenter opened either the extroverted TTS or introverted TTS web page containing the first book description. As noted earlier, each book description page

consisted of a picture of the book, a title and author, and a wave file description of the book. The descriptions were edited version of the actual descriptions on the Amazon.com site. The books were chosen to have the following characteristics: 1) the books and their authors had not sold very well (so that our participants would not be familiar with them; this was confirmed in debriefing); 2) the descriptions had a single main character (so that we could unambiguously ask questions about the characters in the book), and 3) all books were fiction (so that affective criteria would be more important).

Below the icon for the audio file, there was an eight-question questionnaire. Below the questionnaire was a button that allowed the user to progress to the next book description page (which employed the same voice). Except for the personality of the voice, the visual layout, textual information, and book description content were identical across conditions.

Participants read the instruction on the web page and heard five book descriptions. After hearing each book review, they provided answers for questions regarding the book reviewed, main character(s) in that book, and the review itself. After hearing all five book descriptions, participants were presented with a web-based questionnaire. Participants were then told that all five book descriptions had been written by the same person (although they were actually written by different people) and were asked questions about the reviewer. They were also asked questions about the voice they had heard. After participants filled out the web-based questionnaire, we made an audio-recording of them giving their name and describing their experience in the experiment; this allowed us to code each participants' voice as extroverted or introverted (see below). Finally, all participants were debriefed and thanked.

Manipulation

The CSLU Toolkit was used to produce and manipulate the parameters for the two synthesized voices. The four voice parameters discussed above were simultaneously manipulated to instantiate the personality of the voice. The extrovert voice had a speech rate of 216 words per minute, the maximum volume level possible with the Toolkit; a fundamental frequency of 140Hz; and a pitch range of 40Hz. The introvert voice had a speech rate of 184 words per minute; the volume level set at 15% of the maximum; a fundamental frequency of 84 Hz; and a pitch range of 16 Hz. Pre-tests ensured that the voices manifested the appropriate personality and did not differ in intelligence (e.g., too slow speech was classified as both introverted and very unintelligent).

Measures

All dependent measures were based on items from the web-based, textual questionnaires. Participants used radio buttons to indicate their responses. Each question had an independent, ten-point Likert scale.

Questions concerning book quality, the main character, and review quality were asked for each of the different book descriptions. General questions regarding the reviewer and voice were asked once at the end of the complete hearing session. One set of these questions asked, "How well do each of these adjectives describe the reviewer" (identical questions were asked about the voice itself), followed by a list of adjectives. The response scales were anchored by "Describes Very Poorly" (1) and "Describes Very Well" (10). Other sets of questions were based on standard scales.

A number of indices were created, based on theory and factor analysis. All indices were highly reliable.

Introvertedness was an index composed of seven Wiggins [23] introvert adjectives items: Bashful, Introverted, Inward, Shy, Undemonstrative, Unrevealing, and Unsparkling. It was used for assessments of both the voice (Cronbach's alpha=.80) and the reviewer (alpha = .83).

Extrovertedness was an index composed of seven Wiggins [23] extrovert adjective items: Cheerful, Enthusiastic, Extroverted, Jovial, Outgoing, Perky, and Vivacious. It also was used for assessments of both the voice (alpha = .95) and the reviewer (alpha = .94).

Voice attractiveness was an index composed of the items, "How much did you enjoy hearing the computer voice?," "How likely would you be to have the voice read you other descriptions?," and the following adjectives: enjoyable, likable, and satisfying (alpha = .89).

Voice credibility was an index composed of three adjectives: credible, reliable, and trustworthy (alpha = .89).

Quality of the review was an index composed of three items: "What was the quality of the review that you just heard?," "How much did you like the review?," and "How trustworthy was the review?" (alpha ranged from .75 to .91, with a mean of .86).

Liking of the reviewer was an index composed of three adjectives: enjoyable, likable, and satisfying (alpha = .92).

Credibility of the reviewer was measured by a standardized trust scale [22] (alpha = .88).

Users' buying intention was measured by a single item: "How likely would you be to buy this book?"

Results

For the measures that were asked for each book, we used a full-factorial repeated measure ANCOVA with book as the repeated factor and computer voice personality and subject personality as the between-subjects factors. For the items that were only asked once, we used a full-factorial 2x2 ANCOVA. Gender was used as a covariate for all analyses.[1]

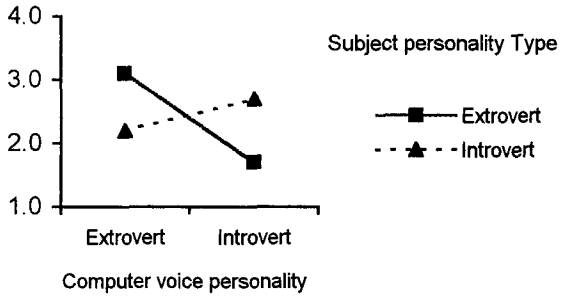
[1] We also conducted a full-factorial repeated measure ANOVA and a full factorial 2x2 ANOVA for all hypotheses. The results were substantively identical in all cases, indicating that the

Consistent with Hypothesis 1, users applied vocal stereotypes to computer-synthesized voices. Specifically, the extrovert computer voice was perceived as being more extroverted (M = 4.3) than the introvert computer voice (M = 2.5), $F(1, 67) = 18.01$, $p < .001$, $\eta^2 = .21$, and the introvert computer voice was perceived as more introverted (M=6.1) than the extrovert computer voice (M = 5.2), $F(1, 67) = 7.56$, $p < .01$, $\eta^2 = .10$. There was also a cross-over interaction for extroversion, $F(1,67) = 6.60$, $p < .05$, $\eta^2 = .10$.

Consistent with Hypothesis 2, the personality of the voice influenced perceptions of the personality of the reviewer, even though the content of the review was held constant. Specifically, the reviewer was perceived as being more extroverted when the descriptions were narrated by the extrovert computer voice (M = 5.7) than by the introvert computer voice (M = 4.3), $F(1, 67) = 8.7$, $p < .01$, $\eta^2 = .16$. Conversely, the reviewer was perceived as being more introverted when the descriptions were narrated by the introvert computer voice (M = 6.2) than the extrovert computer voice (M = 5.1), $F(1, 67) = 7.7$, $p < .01$, $\eta^2 = .10$.

Consistent with the literature on similarity-attraction and its application to human-computer interaction (Hypothesis 3), there was a significant cross-over interaction between computer voice personality and subject personality for voice attractiveness, such that introverts preferred the introvert voice and extroverts preferred the extrovert voice, $F(1, 67) = 14.6$, $p < .001$, $\eta^2 = .18$ (see Figure 1).

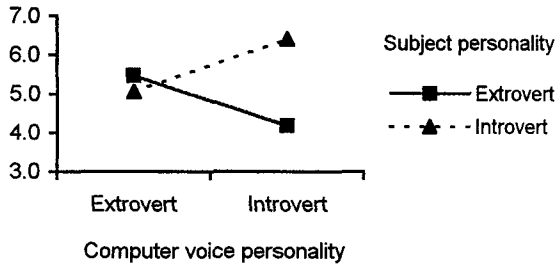
Figure 1. Voice attractiveness



Similarly, extroverts found the extroverted voice more credible than the introverted voice, while introverts felt the opposite (Hypothesis 4), as indicated by the significant computer voice personality by subject personality interaction, $F(1, 67) = 7.86$, $p < .01$, $\eta^2 = .11$ (see Figure 2).

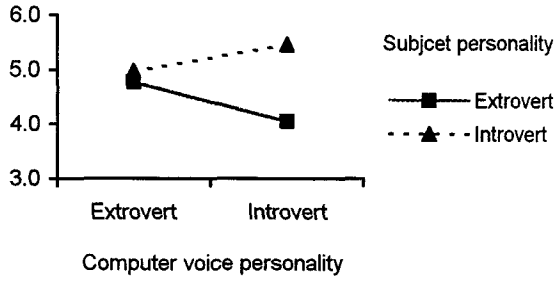
gender of subject was not an influential factor in participants' responses to TTS voices.

Figure 2. Voice credibility



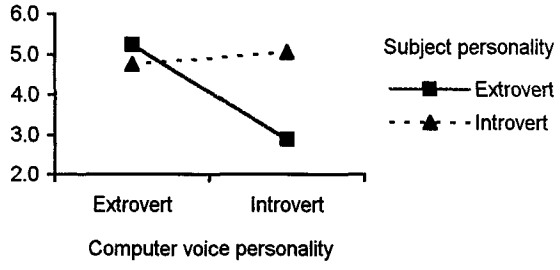
Similarity-attraction extended beyond the voice. There was a significant crossover interaction, with introverts preferring the review read by the introverted voice, and extroverts, the extroverted voice, $F(1, 67)=3.62$, $p<.06$, $\eta^2 = .05$ (Hypothesis 5; see Figure 3). Introverted subjects evaluated the book descriptions more positively in general than did extrovert subjects, $F(1, 67) = 6.31$, $p < .05$, $\eta^2 = .09$.

Figure 3. Quality of the review



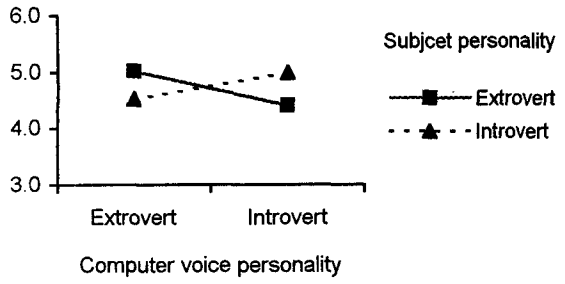
Consistent with proximate source orientation, there was a significant crossover interaction for reviewer attractiveness: Introverts found the reviewer represented by the introverted voice to be more attractive, while extroverts preferred the reviewer when presented with an extrovert voice, $F(1, 67) = 8.35$, $p < .01$, $\eta^2 = .11$ (Hypothesis 6; see Figure 4). Introverts liked the reviewers more in general, $F(1, 67) = 4.87$, $p < .05$, $\eta^2 = .07$.

Figure 4. Liking of the reviewer



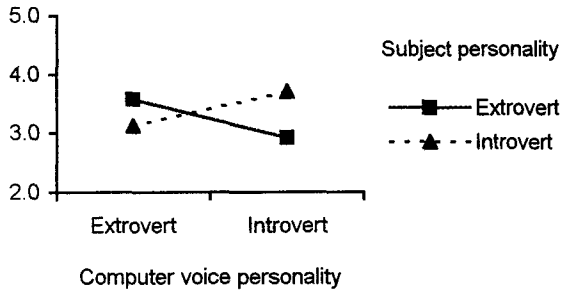
There was a significant cross-over interaction in the expected direction with respect to trust, $F(1, 67) = 10.88$, $p < .01$, $\eta^2 = .14$. (Hypothesis 7 and Figure 5).

Figure 5. Credibility of the reviewer



Finally, consistent with Hypothesis 8, there was a significant computer voice personality X subject personality crossover interaction in the predicted direction, $F(1, 67) = 5.45$, $p < .05$, $\eta^2 = .08$ (see Figure 6), such that introverts were more willing to buy the book when it was presented with an introverted voice, while extroverts were more willing to buy the book from an extroverted voice.

Figure 6. Buying intention



Addressing an alternative explanation

One compelling alternative explanation to the above results is that while similarity-attraction does occur, the similarity is actually between the TTS voice parameters and the *user's* speech parameters. That is, rather than a personality-based explanation for these phenomena, this alternative explanation implies a pure cognitive approach. In addition to the theoretical implications of this explanation, there is also a practical one: Should one measure the user's personality via questionnaires and adapt to personality, or should one instead measure voice characteristics of the user and set the TTS parameters to match?

To ensure that this question is worth asking, that is, to determine whether there are differences between personality as determined by voice characteristics and actual (questionnaire) personality, we had two blind coders rate each participant's voice as either extroverted or introverted. Consistent with the literature [15,16], there was extremely high agreement between the two coders (inter-coder reliability = .90; disagreements were resolved by discussion). Approximately 41 of the participants were labeled as having an extroverted voice, and 30 were labeled as having an introverted voice.[2] However, the correlation between the questionnaire personality and the coded voice personality was a remarkably low (though significant) $r = .24$, $p < .05$. Thus, there are clear and possibly consequential differences between user personality and user voice.

To determine which is the basis for similarity-attraction, we repeated all of the statistical analyses using user voice personality rather than questionnaire personality. (Because there were approximately equal numbers of extroverted and introverted participants, this was a viable strategy). In contrast to the extremely strong results for questionnaire personality, there were *no* similarity-attraction effects, as indicated by *no* interactions for user voice personality by TTS voice personality on any of the dependent measures.[3]

As further evidence of the fundamental aspects of actual (questionnaire) personality, we created a new variable which indicated the (mis)match between the personality of the participants' voice and the personality of the computer voice. We used this variable as a covariate in all of the original ANCOVAs. The covariate had *no* effect on the results; that is, all of the original similarity-attraction effects were obtained. This provides additional evidence that it is the personality, not the cognitive, aspects of similarity-attraction that are relevant to users.

DISCUSSION

One can critique previous experimental tests of the Computers are Social Actors paradigm by arguing that they are too liberal. In previous studies, there were no strong cues to remind individuals that they were working with a computer (other than the monitor). Thus, participants may have simply conceptualized the interaction, implicitly or explicitly, as chat or email, because there was no difference between how one would interact with a person *via* computer and the computer itself. Unlike these previous studies, participants in the present study interacted with a computer via an obviously non-human voice that constantly reminded them of the non-social nature of the interaction. Surprisingly, then, the participants nonetheless assigned a fundamental human property (personality) to the voice and were strongly influenced by that human characteristic. Thus, this study demonstrates that the CASA paradigm is very robust.

The result that even a TTS voice can have a personality has critical implications for the design of TTS systems. Developers of TTS systems have been focused on those aspects of speech that increase intelligibility and naturalness. Now, however, intelligibility is not a serious

[2] One subject refused to record his/her voice.

[3] Of course, it could be argued that the relevant similarity is between the voice that people hear inside their head and the TTS voice, but this information is not readily accessible and hence beyond the scope of experimental research.

problem, and there are a large number of very difficult problems to solve (natural language understanding, emotional understanding, better voice models) before naturalness is achieved. The present study demonstrates that at present, the maximum leverage can be achieved by focusing on the social and personality aspects of speech. Virtually all TTS systems provide socially-relevant parameters that are readily manipulable and remarkably powerful.

The user assignment of personality (and other social attributes) to TTS voices suggests that TTS presents a "casting" problem. For example, when a voice is combined with a character, the personality of the two must be consistent: Pictorial representations of extroverted characters (e.g., open postures; expressive faces; see [4]) should be combined with an extroverted voice, for example.

The proximate source orientation results suggest that companies or individuals should be very careful in choosing their representing voice. For example, if a company or the product of the company is oriented toward extroversion (e.g., a match-making service), the company should employ an extrovert voice, even though the voice would seem to have no relationship with the product. The good news is that it is surprisingly easy to crate a perception of personality in computer voice. With the simple manipulation of accessible vocal dimensions, a designer can easily generate a clear and strong personality. The result that users are attracted to voices that are similar to themselves (similarity-attraction) means that customization of a computer voice according to the users' personality will increase the attractiveness, credibility and informativeness of computer speech output. Previous research [12, 17] has argued that one should match the textual content of an interface to the personality of the user. However, because of the limitations of natural-language production software, this has required designers to build two or more versions of the site, a generally impractical task. Because TTS operates in real-time, a simple change of the speech parameters immediately changes the personality of the entire site. The similarity-attraction principle is so powerful that interface designers can increase the positive evaluation of a product and even the company that makes the product by simply matching the personality of the voice and user.

The fact that questionnaires rather than voice characteristics are the critical determinant of user personality with respect to similarity-attraction is good news for interface designers, as the former information is often easier to obtain than the latter. Many systems are not equipped with microphones or have microphones with insufficient resolution to determine voice parameters.

A key limitation of the current study is that we used personality-neutral content to control the influence of linguistic cues on users' psychological responses to computer voice. However, text frequently manifests a clear and consistent personality. This is a common problem in email readers, branded products, various categories of tasks, etc.. Future studies should examine the tradeoff between similarity-attraction and users' desire for consistency when the voice, the content, and the user all have personality characteristics.

In sum, the results from the present study provide very strong evidence that despite its failure to seem human, individuals nonetheless respond as if text-to-speech has a personality. This fact is both an opportunity and a problem for interface designers.

Reference

1. Beutnagel, M., Conkie, A., Schroeter, J., Stylianou, Y. & Syrdal, A. *The AT&T Next-Gen TTS system.* Paper presented at joint meeting of ASA, EAA, and DAGA, Berlin, Germany, March , 1999. [Online] Available at http://www.research.att.com/projects/tts/papers/1999_ASA_Berlin/nextgen.pdf
2. Giles, H. & Powesland, P. F. *Speech style and social evaluation.* Academic Press, London. 1975.
3. Hall, J. A. *Nonverbal sex differences.* Johns Hopkins University Press, Baltimore, MD, 1980.
4. Isbister, K. & Nass, C. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Interaction,* in press.
5. James, F. *Representing structured information in audio interfaces: A framework for selecting audio marking techniques to represent document structures.* Ph. D. Dissertation, Stanford University, 1998.
6. Kamm, C., Walker, M. and Rabiner, L. *The role of speech processing in human-computer intelligent communication.* Paper presented at NSF Workshop on human-centered systems: Information, interactivity, and intelligence, February, 1997. [Online] Available at http://www.ifp.uiuc.edu/nsfhcs/talks/rabiner.html
7. Massaro, D. M. *Perceiving talking faces: From speech perception to a behavioral principle.* MIT Press, Cambridge, MA, 1997
8. Moon, Y. When the computer is the "salesperson": Computer responses to computer "personalities" in interactive marketing situations. *Working paper No. 99-041,* Harvard Business School. Boston, MA, 1998.
9. Murray, J. B. Review of research on the Myers-Briggs type indicator. *Perceptual and Motor Skills, 70.* 1187-1202, 1990.
10. Nass, C. & Gong, L. *Maximized modality or constrained consistence?* Paper presented at audio-visual speech processing conference, August, 1999.
11. Nass, C. & Moon, Y. Machines and mindlessness: Social responses to computers. *Journal of Social Issues,* in press.

12. Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. Can computer personalities be human personalities? *International Journal of Human-Computer Studies, 43*, 223-239, 1995.

13. Olive, J. P. The talking computer: Text to speech synthesis. in D. Stork (ed.) *Hal's legacy: 2001's computer as dream and reality.* 101-130. The MIT Press, Cambridge, MA, 1997.

14. Oren, T., Salomon, G., Kreitman, K., & Don, A. Guides: Characterizing the interface. In B. Laurel (Ed.) *The art of human-computer interface design.* Addison-Wesley, Reading, MA, 1990.

15. Pittam, J. *Voice in social interaction: An interdisciplinary approach.* Sage, Thousand Oaks, CA, 1994.

16. Ramsay, R. W. Personality and speech. *Journal of Personality and Social Psychology, 4*, 116-118. 1966.

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

18. Sawhney, N. and Schmandt, C. *Design of spatialized audio in nomadic environments.* Paper presented at the international conference on auditory display (ICAD'97), November, 1997. [Online] Available at http://www.media.mit.edu/~nitin/NomadicRadio/ICAD97/ICAD97_paper/ICAD97.html

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

20. Smith, B. L., Brown, B.L., Strong, W.J., & Rencher, A.C. Effects of speech rate on personality perceptions. *Language and Speech, 18*, 145-152, 1975.

21. Turkle, S., *The second self: Computers and the human spirit.* Simon and Schuster, New York, NY, 1984.

22. Wheeless, L. R. & Grotz, J. The measurement of trust and its relationship to self-disclosure. *Human Communication Research, 3(3), 250-257*, 1977.

23. Wiggins, J. S. A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37(3), 395-412, 1979.

24. Woodall, W. G., & Burgoon, J.K. Talking fast and changing attitudes: A critique and clarification. *Journal of Nonverbal Behavior, 8*, 126-142, 1983

The research was supported in part by NSF CARE and Challenge grants awarded to the Oregon Graduate Institute (R. Cole, PI) and a grant from the Center for the Study of Language and Information (CSLI) at Stanford University. The views expressed in this article do not necessarily represent the views of the National Science Foundation or CSLI.