Journal of Experimental Psychology: Applied
2001, Vol. 7, No. 3, 171–181

1076-898X/01/$5.00 DOI: 10.1037//1076-898X.7.3.171

Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction

Clifford Nass and Kwan Min Lee
Stanford University

Would people exhibit similarity-attraction and consistency-attraction toward unambiguously computer-generated speech even when personality is clearly not relevant? In Experiment 1, participants (extrovert or introvert) heard a synthesized voice (extrovert or introvert) on a book-buying Web site. Participants accurately recognized personality cues in text to speech and showed similarity-attraction in their evaluation of the computer voice, the book reviews, and the reviewer. Experiment 2, in a Web auction context, added personality of the text to the previous design. The results replicated Experiment 1 and demonstrated consistency (voice and text personality)-attraction. To maximize liking and trust, designers should set parameters, for example, words per minute or frequency range, that create a personality that is consistent with the user and the content being presented.

The vast majority of content on computers and on the World Wide Web is textual. E-mail, documents, spreadsheets, presentation outlines, e-commerce sites, news and information sites, bulletin boards and chat rooms, advertisements, and search engines, although enhanced by graphical content, are all dominated by text.

The prevalence of textual material is both a problem and an opportunity. The problem is that there are significant barriers to accessing textual content. Only half of U.S. homes have access to the Internet (The White House, 2000), and less than two-thirds have computers; these are the highest rates in the world (Samuelson, 1999, p. 52).

Text is not readily accessible to the 11 million people in the United States with significant visual impairment, nor the 1.5 million people who are blind (James, 1998); aging populations will lead to growing rates of visual impairment. There are also numerous "eyes-occupied" situations, such as driving a car, for which the Web and traditional computer content in general is inherently inaccessible (Sawhney & Schmandt, 1997). Devices that present text cannot be made arbitrarily small because of visibility constraints. Finally, nonliterate individuals, including young children, cannot consume textual content.

A key solution to these problems is the presentation of textual material via voice. Many industrialized nations have telephone penetration rates approximating 90% or greater. Cell phones allow for location-independent and eyes- and hands-free access to computer and Web content; in the past year, the number of cell phones sold in the United States and worldwide far exceeded the number of computers (see Väänänen-Vainio-Mattila & Ruuska, 1999, p. 24). Virtually all computers sold (including highly portable machines) provide at least low-fidelity audio output, and speakers can be made much smaller than screens. Of course, consumption of audio information does not require reading ability.

Because computer content is inherently dynamic and Web content is growing at an incredible rate, it is impractical to present textual content with recorded speech. Instead, one must rely on the computer to generate speech using so-called text-to-speech (TTS) systems, which can generate any arbitrary content in real time in a wide variety of languages (Olive, 1997).

Assessment Criteria for TTS

In general, TTS systems are evaluated according to two dimensions: (a) intelligibility and (b) naturalness of the resulting speech (Beutnagel, Conkie, Schroeter, Stylianou, & Syrdal, 1999; Kamm, Walker, & Rabiner, 1997). There is relatively little difference in intelligibility across systems (Beutnagel et al., 1999; Lai, Wood, & Considine, 2000); indeed, word intelligibility scores for the best TTS systems are close to 97%, approaching that of real human speech (Kamm et al., 1997).

Although TTS systems are intelligible, naturalness scores for even the best TTS systems are in the poor-to-fair range (Kamm et al., 1997) because they lack the quality and prosody of natural human speech. Even human speech that is recorded with poor fidelity or altered by packet transmissions on the Internet sounds very different than synthetic speech. TTS systems tend to have inexplicable pauses, misplaced accents and word emphases, discontinuities between phonemes and syllables, and inconsistent prosody. Although these disfluencies might be dismissed as mere technological constraints when exhibited by a computer, these same paralinguistic cues play a critical role in human–human interactions, manifesting personality, emotion, and other attributes (Apple, Streeter, & Krauss, 1979; Pittam, 1994; Tusing & Dillard, 2000). For example, the term *personality* comes from the Latin *personare*, to sound through, referring to the mouth opening in a mask worn by an actor (Giles & Powesland, 1975), and cues such as loudness, fundamental frequency, frequency range, and speech-rate range distinguish dominant from submissive individuals (Tusing & Dillard, 2000). However, when produced by computer-

Clifford Nass and Kwan Min Lee, Department of Communication, Stanford University.

Correspondence concerning this article should be addressed to Clifford Nass, Department of Communication, Stanford University, McClatchy Hall, Building 120, Stanford, California 94305-2050. Electronic mail may be sent to nass@stanford.edu.

synthesis systems, these cues should have no social meaning, especially when juxtaposed with the other paralinguistic cues that derive from technological limitations. Put another way, although TTS systems effectively communicate linguistic content, they seem to make paralinguistic cues meaningless (Olive, 1997).

Given the obvious deficiencies of computer-generated speech, is there any reason to think that individuals might be influenced by paralinguistic cues in TTS? One line of research, the Computers as Social Actors paradigm, suggests that individuals automatically apply a wide range of social responses to technology (see Nass & Moon, 2000, and Reeves & Nass, 1996, for reviews). For example, research performed under this paradigm has shown that even computer experts are polite to computers (Nass, Moon, & Carney, 1999), gender-stereotype computers (Lee, Nass, & Brave, 2000; Nass, Moon, & Green, 1997), and exhibit moral obligations toward computers (Fogg & Nass, 1997), among other social responses.

Of most relevance to the present studies is a series of studies that demonstrate that individuals attribute personality to computer interfaces on the basis of word choice. In these studies, participants worked with a text-based computer that displayed either a dominant or submissive personality style: The dominant computer used strong, assertive, and confident language during the task (e.g., "You should definitely do this"), whereas the submissive computer used more equivocal and less confident language (e.g., "Perhaps you should do this"). The first experiment (Nass, Moon, Fogg, Reeves, & Dryer, 1995) used a standard personality test (Wiggins, 1979) to categorize participants according to whether they had dominant or submissive personalities. Consistent with the principle of similarity-attraction (Blankenship, Hnat, Hess, & Brown, 1984; Byrne & Griffitt, 1969; Byrne, Griffitt, & Stefaniak, 1967; Duck, 1973; Jellison & Zeisset, 1969; Novak & Lerner, 1968), which posits that individuals are attracted to other people who are similar to themselves, dominant participants were more attracted to, assigned greater intelligence to, and conformed more with the dominant computer, compared with the submissive computer, whereas submissive participants preferred the submissive computer compared with the dominant computer, despite the essentially identical content (Nass et al., 1995). Later studies extended text-based similarity-attraction to consumer behavior (Moon, 1998a), reduction of the self-serving bias (Moon & Nass, 1998), and gain effects (Moon & Nass, 1996). Personality cues (in the introductory instructions) even influenced the perception of neutral content: When the personality of a computer-based *News and Entertainment Guide* matched users' personalities, users found the music, humor, and health advice (which was identical for all participants) to be significantly better (Moon, 1998a).

Although the breadth and depth of social responses demonstrated in these studies is remarkable, a key limitation of the numerous studies in this area is that in all cases, the user is presented with a computer that does nothing to remind the user that it is not human. In these studies, the computers do not exhibit any behaviors uniquely associated with technology, such as error messages or unusual terminology (e.g., "Abort, retry, fail?"), crashing, or seemingly random responses. Thus, the content produced was as consistent with computer-mediated communication or other interactions with a person as it is with human–computer interaction.

Computer-synthesized speech thus represents a critical test of the tendency for individuals to apply and use social attributions toward computers. TTS presents consistent reminders that one is not interacting with a person: Even the most low-fidelity presentation of human speech, including by nonnative speakers, does not sound as artificial and bizarre as the best synthesized speech. Furthermore, because the critical differences between text and voice occur in the paralinguistic domain, individuals must make social attributions using the same categories of cues that indicate the inappropriateness of using social responses.

The current studies present individuals with TTS voices that manifest the paralinguistic cues that suggest personality in human–human interaction and determine whether these cues lead individuals to attribute personality to the computer-generated voices. Also examined is whether these assessments of personality in turn extend beyond mere labeling to influence responses to the content. Thus, the following is the core question: Given identical content, will the particular settings in a TTS engine lead individuals to identify and respond to the computer-generated voice as if it had a personality? If individuals do respond socially to TTS voices, despite the manifest nonhuman disfluencies, this will have implications for the design of TTS systems as well as broader issues of the human–technology relationship.

Personality Markers in Speech

Numerous dimensions of personality have been identified (Murray, 1990; Smith, Brown, Strong, & Rencher, 1975; Wiggins, 1979). The present experiment focuses on the extroversion–introversion dimension, both because it is strongly marked by vocal cues and because it has proved to be important in the human–computer interaction literature (Isbister & Nass, 2000; Nass & Moon, 2000).

Four vocal features—intensity (loudness), mean fundamental frequency (F0), frequency range (F0 variation), and speech rate—are associated with judgments of extroversion (Scherer, 1978, 1979; see also Apple et al., 1979; Pittam, 1994). Loudness is usually measured by mean amplitude, which is an average amount of sound energy expended to produce a vocalization: Loudness is positively associated with extroversion, dominance, or both (Buller & Burgoon, 1986; Pittam, 1994; Tusing & Dillard, 2000). Fundamental frequency is the average number of complete cycles of vibration per second made by the vocal cords: Extroverts–dominant people speak with higher fundamental frequency (Pittam, 1994; Tusing & Dillard, 2000). This higher fundamental frequency is relative to both gender variation (120 Hz for men vs. 225 Hz for women) and age (e.g., 265 Hz for children; see Fry, 1979). Frequency range is the extent to which a vocalization varies around its mean F0 value. Extrovert–dominant people speak with more frequency range than do introverts–submissive people (Aronovitch, 1976; Scherer, London, & Wolf, 1973). Speech rate refers to the number of words (or syllables) uttered in a given period of time: Faster speech rate is associated with extroversion–dominance (Aronovitch, 1976; Buller & Burgoon, 1986; Tusing & Dillard, 2000).

To ensure a successful manipulation, consistency with humans' paralinguistic behaviors, and ecological validity, we manipulated all four attributes simultaneously. Furthermore, studying vocal cues in isolation may produce unrealistic results because vocal cues may interact in complex ways (Tusing & Dillard, 2000).

To determine whether vocal 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. Second, we determined 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 because it requires users to think of the voice as the source of the content. Finally, as the most stringent test, we examined whether users would draw on social rules when responding to the voice.

Experiment 1: Similarity-Attraction

This experiment was executed in the context of a book-buying Web site that presented five different books, all on the same Web page.[1] The Web page 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 the book description in text, there was a link to an audio (.wav) file. Participants clicked on the link to play the review.

In addition to examining the identification of personality in voice and content, we focused on the principle of similarity-attraction. Specifically, we tested the following hypotheses:

Hypothesis 1: Users will recognize vocal cues indicating personality, even when they hear computer-synthesized speech.

Hypothesis 2: Users will be more attracted to a TTS voice that exhibits a personality similar to their own than a TTS voice that exhibits a dissimilar personality.

A number of studies in human–human interaction have demonstrated that vocal cues influence perception of the content that is presented (Kahan, 1962). Hence, we predict the following:

Hypothesis 3: Users will evaluate the content more positively if the review is narrated by a TTS voice that matches their own personality.

Hypothesis 4: Users will show more buying intention for a book whose review is narrated by a TTS voice that matches their own personality.

Although the correlation between people's voices and their personality is not very strong (see *Results* section 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 (Reeves & Nass, 1996, chap. 16) 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 on the basis of the most readily available cues. This leads to the following hypotheses:

Hypothesis 5: Users will infer the personality of a review writer based on the personality of the TTS voice they heard.

Method

Participants

Several weeks before the experiment, a Web-based personality survey was administered to students in a large introductory communication course. A short form of the Myers–Briggs Type Indicator (see Murray, 1990, for a review) and Wiggins (1979) personality tests were administered to maximize the likelihood of correctly assessing the personality of participants. From a total of approximately 150 students, 72 participants (age range = 18–23, M = 20.20, SD = 1.17) with the most extreme and consistent scores on the two scales—36 extrovert and 36 introvert participants—who spoke English as a first language were invited to participate in the experiment. 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) × 2 (participant personality: extrovert vs. introvert) balanced, between-subjects design, with the five book descriptions as a repeated factor. On arrival to the laboratory, each participant was assigned to a computer equipped with a pair of headphones and an Internet Explorer 4.0 browser. Participants were instructed to wear the headphones for the duration of the experiment and not adjust the volume level of either the headphone or the computer (to control volume). As part of the experimental instructions, we explicitly told each of the participants that they would be hearing computer-generated speech, and we chose a TTS engine that was unambiguously synthetic.

The experimenter opened either the extroverted TTS or introverted TTS Web page containing the first book description. The descriptions were edited versions of the actual descriptions on the Amazon.com site. A posttest questionnaire confirmed that none of the participants were familiar with any of the books.

Below the icon for the audio file, there was a questionnaire regarding the book being reviewed and the review itself. Subsequent book descriptions and questionnaires were placed sequentially on the Web page. Except for the personality of the voice, the visual layout, textual information, and book description content were identical across conditions.

After hearing all five book descriptions, participants were presented with a final 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 this final questionnaire, we audiorecorded each participant giving their name and describing their experience in the experiment. This allowed us to code each participants' voice as extroverted or introverted (see *Results* section below). Finally, all participants were debriefed and thanked.

Manipulation

The Center for Spoken Language Understanding (CSLU) toolkit (a free download at http://cslu.cse.ogi.edu) was used to produce the extroverted and introverted synthesized voices. The four voice parameters discussed above were simultaneously manipulated to instantiate the personality of the voice. The extrovert voice had the maximum volume level possible with the toolkit, a fundamental frequency of 140 Hz, a frequency range of 40 Hz, and a speech rate of 216 words per minute. The introvert voice had the volume level set at 15% of the maximum; a fundamental frequency of 84 Hz, a frequency range of 16 Hz, and a speech rate of 184 words per minute. Pretests conducted with nonparticipating graduate students ensured that the voices manifested the appropriate personality and did not differ in other dimensions such as attractiveness, credibility, informativeness, or persuasiveness.

[1] The site can be found at www.stanford.edu/~nass/comm169/Kwan/1-1.fft.

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, 10-point Likert scale.

Questions concerning book quality 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. By asking the questions at the end, and in this order, we ensured that participants were not cued to think about voice personality. 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. All indices were analytically distinct and highly reliable.

Extrovertedness–introvertedness was an index composed of 10 Wiggins (1979) personality adjective items: cheerful, enthusiastic, extroverted, introverted (reverse coded), inward (reverse coded), jovial, outgoing, perky, shy (reverse coded), and vivacious. The index was used for personality assessments of both the TTS voice (Cronbach's $\alpha = .89$) and the reviewer ($\alpha = .91$). The higher the score, the more extroverted the voice or the reviewer.

Liking of the voice 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 Wheeless and Grotz's (1977) trust scale ($\alpha = .88$).

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 analysis of variance (ANOVA) with book as the repeated factor and computer voice personality and participant personality as the between-subjects factors. For the items that were only asked once, we used a full-factorial 2 × 2 ANOVA. Table 1 reports mean, standard deviations, F values, and effect sizes for all dependent variables. We also conducted a full-factorial repeated measure analysis of covariance (ANCOVA) and a full factorial 2 × 2 ANCOVA with a gender covariate for all hypotheses to check for gender effects. The results were substantively identical in all cases, indicating that the gender of the participant was not an influential factor in participants' responses to TTS voices. Hence, only the ANOVA results are reported here.

Consistent with Hypothesis 1, users applied vocal stereotypes to computer-synthesized voices: The extrovert computer voice was perceived as being much more extroverted ($M = 4.86$, $SD = 1.67$) than the introvert computer voice ($M = 3.34$, $SD = 0.91$; see Table 1). Neither a main effect of participant personality nor an interaction effect was found.

Consistent with the literature on similarity-attraction and its application to human–computer interaction (Hypothesis 2), there was a large and significant crossover interaction between computer voice personality and participant personality for voice attractiveness, such that introverts preferred the introvert voice and extroverts preferred the extrovert voice. Because attractive communicators are more credible (Eagly & Chaiken, 1975; Mills & Aronson, 1965), there was also a large similarity-attraction effect with respect to voice credibility, as indicated by the significant crossover interaction.

Table 1
Analysis of Variance Results From Experiment 1

	*M*s and *SD*s				*F* values and effect sizes		
	Introvert participant		Extrovert participant		Main effects		Interaction effects
Dependent variable	Introvert voice	Extrovert voice	Introvert voice	Extrovert voice	Participant (P) personality	Voice (V) personality	P × V
Voice extrovertedness	3.41	4.43	3.28	5.30	1.35	23.64***	2.54
	(1.02)	(1.73)	(0.81)	(1.53)	$\eta^2 = .02$	$\eta^2 = .26$	$\eta^2 = .04$
Liking of the voice	2.82	2.21	1.69	3.08	.25	2.16	14.10***
	(1.26)	(0.92)	(0.96)	(1.32)	$\eta^2 = .00$	$\eta^2 = .03$	$\eta^2 = .17$
Credibility of the voice	6.43	5.06	4.18	5.46	3.84†	.01	7.98**
	(1.46)	(2.55)	(1.60)	(1.18)	$\eta^2 = .05$	$\eta^2 = .00$	$\eta^2 = .11$
Quality of the review	5.44	4.96	4.06	4.79	6.00*	.14	3.64†
	(0.96)	(1.31)	(1.35)	(1.66)	$\eta^2 = .08$	$\eta^2 = .00$	$\eta^2 = .05$
Liking of the reviewer	5.06	4.74	2.89	5.24	3.30†	4.93*	8.47**
	(2.07)	(2.37)	(1.27)	(1.88)	$\eta^2 = .05$	$\eta^2 = .07$	$\eta^2 = .11$
Credibility of the reviewer	4.98	4.52	4.40	5.02	.06	.24	11.00***
	(0.55)	(0.89)	(0.70)	(0.57)	$\eta^2 = .00$	$\eta^2 = .00$	$\eta^2 = .14$
Reviewer extrovertedness	4.68	5.56	4.22	5.77	.12	12.47***	.90
	(1.39)	(1.37)	(1.66)	(1.42)	$\eta^2 = .00$	$\eta^2 = .16$	$\eta^2 = .01$
Buying intention	3.68	3.10	2.93	3.59	.24	.03	5.40*
	(0.98)	(1.24)	(1.08)	(1.17)	$\eta^2 = .00$	$\eta^2 = .00$	$\eta^2 = .07$

Note. Standard deviations are in parentheses. † $p < .10$ (marginally significant). * $p < .05$. ** $p < .01$. *** $p < .001$.

Similarity-attraction extended beyond the voice. There was a moderate interaction indicating similarity attraction with respect to the quality of the book review (Hypothesis 3). Consistent with Hypothesis 4, there was a large and significant crossover interaction with respect to purchase behavior, with participants more likely to buy the book when the voice personality matched their own.

The personality of the voice influenced perceptions of the personality of the reviewer, even though the content of the review was held constant (Hypothesis 5). Specifically, the reviewer was perceived as being clearly more extroverted when the descriptions were narrated by the extrovert computer voice (M = 5.67, SD = 1.38) than by the introvert computer voice (M = 4.45, SD = 1.52). Neither a main effect of participant personality nor an interaction effect was found. Consistent with this result, participants found the reviewer more attractive when the voice personality and participant personality were similar. In addition, participants trusted the reviewer more when the two personalities matched. Both of these effects demonstrate the power of the similarity-attraction principle.

Addressing an Alternative Explanation

One compelling alternative explanation to the above results is that although 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 acoustic approach to the same results. In addition to the theoretical implications of this explanation, there is also a practical one: Should one measure the user's personality with 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—two masked coders rated each participants' voice as either extroverted or introverted. Consistent with the literature (Pittam, 1994; Ramsay, 1966), there was extremely high agreement between the two coders (intercoder 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] The correlation between questionnaire personality and the coded voice personality was remarkably low (though significant), $r = .24$, $p < .05$. Thus, there are clear and possibly consequential differences between user personality and user voice.

To determine whether both personality and the user's voice were the basis for similarity-attraction, we repeated all of the statistical analyses with 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. Of course, it could be argued that the relevant similarity is between the voice that people hear inside their head when they are speaking and the TTS voice, but the former information is not readily accessible and hence beyond the scope of experimental research.

As further evidence of the fundamental aspects of actual (questionnaire) personality, we created a new variable that 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 auditory, aspects of similarity-attraction that are relevant to users.

Experiment 2: Consistency and Similarity-Attraction Experiment

In the previous experiment, we used content that was neutral with respect to personality. This method allowed us to focus on the influence of vocal cues on users' psychological responses to a computer voice. However, a great deal of computer and Web content—especially E-mails, personal narratives, and content associated with a particular brand—manifest a clear personality through text. What happens when the text personality (which is not readily changeable) is consistent with or inconsistent with the users' personality? Will the selection of the voice continue to influence the users' perceptions and attitudes?

In this experiment, we replicate and extend the ideas behind the first experiment by examining both similarity-attraction and consistency effects in a context in which both the linguistic cues manifested by the words to be spoken and the paralinguistic cues conveyed by TTS output provide personality cues. Thus, the second experiment examines the trade-off between similarity-attraction and the users' desire for consistency when the voice, the content, and the user all manifest personalities.

This second experiment was executed in the context of an on-line auction Web site (see www.stanford.edu/~nass/comm169/Kwan/2-1.fft). Each Web page had an identical visual interface based on e-Bay's auction item descriptions. Each page included the names and pictures of nine antique or collectible auction items (e.g., 1963 classic lamp, 1920s radio, 1968 Russian circus poster, etc.). Instead of having the item description in text, there was a link to an audio (.wav) file for each item. Participants clicked on the link to play the description of the item. Participants heard either extrovert or introvert descriptions (text) of the items with either introvert or extrovert TTS.

Hypotheses

In human speech, the verbal and vocal channels are so well-integrated that they seldom provide inconsistent information to listeners. However, in computer-synthesized speech, inconsistency is very common because the verbal features are determined by a sender and the vocal features are usually determined by system defaults or by programmers.

In social situations, people prefer to interact with individuals who behave consistently as compared with individuals who behave inconsistently because consistency in others lightens cognitive

[2] One participant declined to record his or her voice.

load and makes it easier to predict what will happen when they engage with others (Fiske & Taylor, 1991): This is called the consistency-attraction principle (Field, 1994; Thomas & Johnston, 1981). On the basis of the consistency-attraction principle, we propose the following hypotheses:

Hypothesis 6: Users will like a computer voice more if TTS personality and text personality are consistent with each other.

Hypothesis 7: Users will like the writer and the text content more if TTS personality and text personality are consistent with each other.

Given the power of voice personality and the tendency for voice to influence perception of content (Kahan, 1962), we also predict the following:

Hypothesis 8: Users will infer the personality of the text on the basis of the personality of the TTS voice they heard.

Finally, we predict that this experiment will replicate Hypotheses 1, 2, 3, and 5 of Experiment 1 (data relevant to Hypothesis 4 were not available).

Method

Participants

Identical to Experiment 1, a portion of the Myers–Briggs personality inventory and a portion of the Wiggins (1979) interpersonal adjective set were administered several weeks before the experiment. From two undergraduate introductory classes, a total of 80 participants—40 extrovert and 40 introvert participants (age range = 17–23, M = 20.20, SD = 1.18)—were invited to participate in the experiment. Gender was approximately balanced across conditions, and all participants were native English speakers. All participants signed informed consent forms, were debriefed at the end of the experiment session, and received class credit.

Procedure

All participants were randomly assigned to condition in a 2 (participant personality: extrovert vs. introvert) × 2 (computer voice personality: extrovert vs. introvert) × 2 (text personality: extrovert vs. introvert) balanced, between-subjects design.

Manipulation

Computer voice personality. Extroversion and introversion in the TTS voice were created exactly as in Experiment 1.

Text personality. Following previous studies (e.g., Isbister & Nass, 2000), the extroversion or introversion of the item description was operationalized by manipulating the phrasing and length of the description, as both word selection and number of words are markers of personality. (Because our focus is on the interaction between text personality and voice and participant personality rather than the main effect of text personality, the use of two different means of manifesting personality is appropriate.) The extrovert description was relatively lengthy and used strong and descriptive language expressed in the form of confident assertions. The introvert description was relatively short and used weaker language expressed in the form of suggestions. For example, the extroverted description of the lamp read as follows.

> This is a reproduction of one of the most famous of the Tiffany stained glass pieces. The colors are absolutely sensational! The first class hand-made copper-foiled stained glass shade is over six and one-half inches in diameter and over five inches tall. I am sure that this gorgeous lamp will accent any environment and bring a classic touch of the past to a stylish present. It is guaranteed to be in excellent condition! I would very highly recommend it.

Conversely, the introverted description of the lamp read as follows.

> This is a reproduction of a Tiffany stained glass piece. The colors are quite rich. The hand-made copper-foiled stained glass shade is about six and one-half inches in diameter and five inches tall.

The pretest results confirmed that the manipulation was successful. We chose items whose monetary value would not be obvious to the participants.

Measures

All dependent measures were based on items from a paper-and-pencil questionnaire administrated after the experiment. Each question had an independent, 10-point Likert scale. The indices in Experiment 2 paralleled those in Experiment 1. All indices were highly reliable.

Extrovertedness–introvertedness was computed as in Experiment 1 and was used for personality assessments of the TTS voice (Cronbach's α = .92), the narrated text (α = .90), and the writer who wrote the text (α = .92).

Liking of the voice was computed as in Experiment 1 (α = .89).

Liking of the text was an index composed of the following three adjectives: enjoyable, likable, and satisfying (α = .89).

Liking of the writer (α = .83) and credibility of the writer (α = .92) were computed as in Experiment 1.

Results

All analyses were based on a full-factorial 2 × 2 × 2 ANOVA model. Consistent with Experiment 1, users recognized the personality cues in computer-synthesized voice (Hypothesis 1): The extrovert computer voice was perceived as being clearly more extroverted (M = 5.55, SD = 1.88) than the introvert computer voice (M = 4.18, SD = 1.27), $F(1, 72) = 21.41$, $p < .001$, $\eta^2 = .23$.[3] Similarly, the manipulation of text personality was successful: Extrovert text clearly was perceived as being more extroverted (M = 6.37, SD = 1.29) than introvert text (M = 4.56, SD = 1.09), $F(1, 72) = 58.00$, $p < .001$, $\eta^2 = .45$.

There was very strong evidence for consistency-attraction effects. When the voice personality and text personality matched, participants very clearly preferred the voice compared to voice–text mismatch participants (Hypothesis 6; see Tables 2 and 3). Similarly, voice–text match participants liked the text much more than did mismatch participants (Hypothesis 7). Matched participants found the writer to be clearly more credible and tended to like the writer more than mismatch participants. These latter results are particularly striking in that the text, which was a (presumably) direct indication of the writer (in contrast to Experiment 1), clearly manifested the writer personality.

Consistent with Experiment 1, participants inferred the personality of a writer on the basis of the personality of the computer voice they heard (Hypothesis 2): The writer was perceived as being much more extroverted if the writing was narrated by an extrovert voice (M = 6.32, SD = 1.48) as compared with an

[3] There were two unexpected interaction effects on the perception of voice personality: a Participant Personality × Voice Personality interaction effect, $F(1, 72) = 4.62$, $p < .05$, $\eta^2 = .06$, and a Voice Personality × Text Personality interaction effect, $F(1, 72) = 8.67$, $p < .01$, $\eta^2 = .11$.

Table 2
Means and Standard Deviations of Major Dependent Variables in Experiment 2

	Voice × Text Consistency		Voice × Participant Similarity		Text × Participant Similarity	
Dependent variable	Matched	Mismatched	Matched	Mismatched	Matched	Mismatched
Liking of the voice	3.90	2.94	3.78	3.06	3.80	3.04
	(1.77)	(1.20)	(1.92)	(1.06)	(1.78)	(1.27)
Liking of the text	4.44	3.75	4.10	4.10	4.22	3.97
	(1.42)	(1.49)	(1.50)	(1.49)	(1.59)	(1.38)
Liking of the writer	4.62	4.11	4.65	4.08	4.63	4.10
	(1.40)	(1.45)	(1.80)	(0.89)	(1.40)	(1.45)
Credibility of the writer	4.74	4.30	4.56	4.48	4.42	4.63
	(0.94)	(0.86)	(0.91)	(0.95)	(0.87)	(0.97)

Note. Standard deviations are in parentheses.

introvert voice (M = 5.17, SD = 1.57), $F(1, 72) = 24.60$, $p < .001$, $\eta^2 = .26$. Similarly, the computer voice personality influenced the perception of text personality (Hypothesis 8): Text narrated by an extrovert computer voice was perceived as being clearly more extroverted (M = 5.84, SD = 1.41) than text narrated by an introverted computer voice (M = 5.09, SD = 1.44), $F(1, 72) = 9.9$, $p < .01$, $\eta^2 = .12$. This latter result is particularly striking given that the text personality was clearly manifested.[4] It should be noted that the consistency effects were more accentuated for the extrovert text as compared with the introvert text, perhaps as a result of the lengthier content producing a stronger instantiation of the manipulation. (As noted earlier, extending the introvert text to match the length of the extrovert text was problematic because introverts use much fewer words than extroverts.)

Replicating Hypothesis 2, there was a large and significant similarity-attraction effect for voice, with voice–participant matching participants clearly preferring the voice as compared with nonmatching participants.

There was limited evidence for similarity-attraction extending beyond the voice, perhaps because text personality was strongly manifest and a confounding aspect in this experiment. Voice–participant matching participants did like the writer much more than did mismatching participants (Hypothesis 3). There was no significant interaction with respect to credibility of the writer nor for liking of the text, in contrast to Experiment 1.

In contrast to previous research, there was very little evidence of similarity-attraction with respect to the match between participant and text personality, perhaps because the effects were swamped by the voice effects (which in all cases were stronger than the text effects).

In contrast to Experiment 1, there were clear main effects for voice personality and text personality. Participants in the extroverted-voice condition liked the voice (M = 4.18, SD = 1.72), the text (M = 4.59, SD = 1.33), and the writer (M = 4.83, SD = 1.48) much more than did participants in the introverted-voice condition (M = 2.66, SD = 0.96; M = 3.60, SD = 1.48; M = 3.90, SD = 1.24, respectively), perhaps because the dynamism of the extroverted voice made the boring content of arcane auction items more palatable. The richer extrovert text may have similarly enlivened the content, leading to more positive perceptions of the content (M = 4.61, SD = 1.62) and the voice (M = 3.77, SD = 1.94) as compared with the flat introvert text (M = 3.58, SD = 1.14; M = 3.07, SD = 1.03, respectively). Conversely, the obvious persuasive attempt of the extrovert text (M = 3.90, SD = 0.70) might have reduced its credibility relative to the seemingly neutral introvert text (M = 5.14, SD = 0.66).

Discussion

A potential criticism of previous experimental tests of the Computers Are Social Actors paradigm (Reeves & Nass, 1996) is that they were too liberal. In those studies, there were no strong cues to remind individuals that they were working with a computer (other than the monitor). Thus, participants might have simply conceptualized the interaction, implicitly or explicitly, as chat or E-mail, because there was no difference between how one would interact with a person via computer and the computer itself (cf. Morkes, Kernal, & Nass, 2000).

Unlike these previous studies, participants in the present experiments interacted with a computer with an obviously nonhuman voice that constantly reminded them of the nonsocial nature of the interaction. Nonhumanness of the voice was established not only by our explicit instruction but also by the obviously nonnatural characteristics of the voice. (The low personality scores of both extrovert and introvert voices—in terms of absolute scores—provide additional evidence of the nonhumanness of the voice.) The participants nonetheless assigned a fundamental human property—personality—to the voice and were strongly influenced by that human characteristic. Thus, these experiments demonstrate that the Computers Are Social Actors paradigm is very robust: Despite knowing that computers do not have personality in any human sense, and despite being confronted with constant reminders that the voice was not human, individuals exhibited attitudes associated with similarity-attraction and consistency-attraction.

These studies provide a number of contributions to the literature beyond the demonstration of the strength of the Computers Are Social Actors paradigm. First, these studies demonstrate that paralinguistic cues are relevant to individuals' responses to synthetic voices; previous research has only focused on identification (e.g.,

[4] There was also a significant Participant Personality × Text Personality interaction effect on perception of text personality, $F(1, 72) = 5.94$, $p < .05$, $\eta^2 = .08$.

Table 3
Analysis of Variance Results (F Values and Effect Sizes) From Experiment 2: Consistency-Attraction and Similarity-Attraction

Dependent variable	Voice × Text Consistency	Voice × Participant Similarity	Text × Participant Similarity	Three-way interaction	Voice personality	Text personality	Participant personality
Liking of the voice	13.52* (η^2 = .16)	7.50 (η^2 = .09)	8.36** (η^2 = .10)	1.52 (η^2 = .02)	34.28*** (η^2 = .32)	7.26** (η^2 = .09)	.82 (η^2 = .01)
Liking of the text	5.57* (η^2 = .07)	.00 (η^2 = .00)	.72 (η^2 = .01)	.00 (η^2 = .00)	11.81** (η^2 = .14)	12.77** (η^2 = .15)	.95 (η^2 = .01)
Liking of the writer	3.10† (η^2 = .04)	4.00* (η^2 = .05)	3.39† (η^2 = .04)	.14 (η^2 = .00)	10.57** (η^2 = .13)	1.45 (η^2 = .02)	5.27* (η^2 = .07)
Credibility of the writer	9.01** (η^2 = .11)	.24 (η^2 = .00)	2.06 (η^2 = .03)	.32 (η^2 = .00)	1.01 (η^2 = .30)	72.40*** (η^2 = .50)	.96 (η^2 = .01)

† $p < .10$ (marginally significant). * $p < .05$. ** $p < .01$. *** $p < .001$.

Mullenix, Johnson, Topcu-Durgun, & Farnsworth, 1995) or assessment of naturalness. Second, these studies demonstrate that an arbitrarily assigned synthetic voice can influence perceptions beyond the voice to influence perception of the content producer (even when the content has a marked personality) and anticipated buying.

The first experiment also demonstrated that these paralinguistic responses do not reflect a simple matching of synthesized speech parameters and the participants' own speech characteristics; instead, participants must be imputing personality to the synthesized voices and comparing the personality ascribed to the voice with their own personality; previous research with text personality did not distinguish between the mere matching (i.e., the text was similar or different to type of text that the person produces) and social explanations.

The second experiment examines the ways in which paralinguistic cues could influence the perception of the personality of content, even when the personality was clearly marked in the text. Furthermore, it demonstrates that individuals are influenced by cross-modality consistency-attraction.

Designing With Synthetic Voices

These findings have key implications for designing interactions with synthetic voices. Modern TTS systems have generally solved the intelligibility problem, although there are a large number of very difficult problems to solve (e.g., natural language understanding, emotional understanding, better voice models) before naturalness can be achieved. The research presented here suggests that at present, the "low-hanging fruit" for design is the social and personality aspects of speech manifested in vocal cues. These cues are powerful and readily manipulable in real time in virtually all synthetic speech systems.

Leveraging Similarity-Attraction

For example, virtually all content producers want to increase the liking of the material they present. Customizing the content to match the user is generally impractical, as maintaining multiple versions is not manageable and automating the process with natural-language production software is not possible. Conversely, because TTS operates in real time, a simple change of the speech parameters can be used to immediately and effectively change the voice personality to match the users' personality, thereby increasing the desirability of the entire site. Of course, this strategy would only be clearly effective when the content is personality neutral; otherwise, one might introduce inconsistency effects.

Similarly, producers of medical content, educational software, and advertising, among others, are concerned with increasing the credibility of their content (Fogg & Hsiang, 1999). The present research suggests that a simple manipulation of synthetic speech characteristics to match the personality of the user, as long as the personality of the content is neutral, can dramatically increase the persuasiveness of content at essentially no cost.

Another domain in which strategic use of the personality of TTS would be helpful is surveys. Recent studies have examined how modality, particularly the differences between text administration and voice administration, affect the honesty of responses (e.g., Moon, 1998b). The present results suggest that speech output should not be conceptualized as a single option. Instead, voice characteristics likely exert a strong influence on level of disclosure. For example, the increased liking and credibility associated with a voice that is similar to the user should dramatically increase the interviewee's honesty relative to a poorly chosen voice.

The similarity-attraction principle is so powerful that interface designers can increase the positive evaluation of a product and even of the company that makes the product by simply matching the personality of the voice and the user.

User Personality Versus User Voice

The fact that questionnaires rather than voice characteristics are the critical determinant of user personality with respect to similarity-attraction is a critical finding for interface designers. Answers to just a few questions can rapidly classify users as extroverts or introverts and guide the setting of the appropriate TTS parameters. Conversely, many computing systems are not equipped with microphones or have microphones with insufficient resolution to determine voice parameters, and cell phones often do not provide the fidelity to analyze voices. Hence, if designers had to rely on categorization of user voices rather than questionnaire assessments, the applicability of this research would be limited.

Leveraging Consistency-Attraction

Voice attractiveness, text attractiveness, writer attractiveness, and writer credibility were significantly determined by the consistency of voice and text personalities. Of course, this is not news in

traditional media: The idea of casting actors, including voice actors, to match their role is well-established in Hollywood. The present research demonstrates that this same principle surprisingly applies even when users are constantly reminded that they are not interacting with a person.

Web sites and other content sources with strong brand identities tend to have clear and uniform linguistic personalities. For these sites, it is critically important to select a synthesized voice that is consistent with the personality of the language. This will increase liking and credibility of the content and, by extension, the brand reflected by the content.

Other content providers, including newsgroups, magazine and news sites, and auction sites, include a diversity of content. In these cases, the sites have three options: (a) assign the parameters according to the content of each message, (b) assign the parameters according to the characteristics of the user, or (c) select a single voice for the site. Opting for consistency-attraction (Option a) has the advantage that consistency seemed to have the most positive and strong outcomes, but it can be very difficult to automatically analyze the textual content of a message and assign it a personality. Hence, the selection of a voice would have to be performed by the content creator or an editor. Also, the use of multiple voices on a single site can lead to a variety of unanticipated arousal and other effects (Reeves & Nass, 1996). Matching the user and the voice (Option b) solves the content analysis problem, but the selection cannot be made when a visitor first accesses the content and consistency-attraction is a priori abandoned. Finally, the uniform voice (Option c) has the advantage of providing a brandable voice for the site, but the disadvantages likely outweigh the advantages of this approach.

Desire for TTS voice consistency may extend from text content to the role. For example, sites associated with extroverted activities, such as party planning or public speaking, may be better presented by an extroverted voice. Conversely, stereotypically introverted roles, such as librarian, should be represented by an introverted voice.

These design recommendations are grounded in the assumption that increasing likability, credibility, and buying intention are a priori desirable outcomes; this is a producer perspective. However, in any design decision, there are a variety of stakeholders whose interests should be evaluated when implementing any interfaces. Friedman (1997) has provided an excellent introduction to the issues relevant to value-centered design.

Open Issues

Although these experiments provide compelling evidence that synthetic voices can manifest personality, a number of open issues should be noted. First, we do not have a detailed understanding of the underlying processes that encourage social responses. As discussed at length elsewhere (Nass & Moon, 2000), we can rule out anthropomorphism because in debriefing, all participants indicated that they knew that the voices were not human and did not indicate an underlying personality of the computer.

The most compelling explanation, although not without its difficulties, comes from the psychology of speech processing. The argument is as follows (see Nass & Gong, 2000, for an extended discussion): (a) Speech sounds, including speech in foreign languages, nonsense syllables, and speech played backward, are processed differently than other sounds in the environment (Slobin, 1979); (b) this processing includes the assessment of social characteristics, such as gender and personality, from the analysis of paralinguistic cues (Denes, Caldognetto, Semenza, Vagges, & Zettin, 1984; Mullenix et al., 1995); (c) these social characteristics automatically influence human attitudes and behaviors; (d) because only humans produced speech throughout virtually all of human evolution, it requires significant awareness and effort to prevent Steps a–c from occurring; and (e) thus, individuals will apply a wide range of social rules and expectations to synthesized voices.

Although we tend to believe this account, there are two limitations. First, it is virtually impossible to demonstrate Steps c–e, although they could certainly be rejected. Second, different explanations, such as mindlessness (see Langer, 1989; Nass & Moon, 2000), can explain the same phenomena; there are no critical tests to distinguish this evolutionary explanation from other contenders.

A second open issue is the low correlation between personality as measured by questionnaire and personality as measured by a person's voice. Given the high levels of agreement across people on the vocal markers of personality and its effects (even when the speech is synthetic), the inefficiency of voice as a predictor of personality is very striking. Future work should systematically explore this result.

Two issues might limit the generalizability of the results presented here. First, participants in both experiments had limited exposure to the synthetic speech, on the order of 15 min, and none of the participants previously had spent a great deal of time with synthetic speech systems. We do not know whether the attributions and influences of personality would continue over an extended interaction or with extended exposure. One reason we might expect these attributions to remain relevant is that the responses seem to be automatic: As there are no obvious links between the settings of the TTS engine and individuals' attitudes toward the content or content provider, it is unlikely that individuals will scrutinize their reactions as influenced by the personality of the voice. Of course, future research should address this question.

The second caveat is the participant population. The present participants (college students) tend to be very technologically literate and frequent computer and audio users; hence, they have likely had exposure to a wide range of audio fidelity. On the one hand, it is possible that those individuals with limited exposure to technology may be so distracted by the inadequacies that they would ignore the paralinguistic cues. On the other hand, the greater cognitive load experienced by these novice users might actually lead to more automatic reliance on markers associated with the more common human interactions (Reeves & Nass, 1996, chap. 1). On the other extreme, blind users with extremely heavy use of synthetic speech may be a valuable population to examine.

Extensions

The present research demonstrates that despite the clear disfluencies associated with TTS, individuals nonetheless identify and are influenced by voice cues suggesting personality. Of course, voice manifests other social cues in addition to personality. For example, gender is a powerful voice cue that is encoded categorically through a detailed and complex auditory psychophysical process involving F0, formant frequencies, breathiness, and so

forth (Mullenix et al., 1995). It is likely, therefore, that the perceived gender of synthesized speech would lead to gender stereotyping, social identification, and other gender effects (Lee et al., 2000). Other demographic characteristics marked by vocal cues, such as age and ethnicity, might also prove to be powerful.

Beyond demographics, there is some evidence that synthesized speech can manifest emotion (Cahn, 1990). For example, although it is clear that computers do not possess emotions (cf. Picard, 1997), users might expect emotional consistency between content and vocal cues of synthesized speech (the bad actor problem). Emotional synthesized speech also may provide opportunities for arousal maintenance (Zillmann, 1991) and similarity-attraction.

In sum, the present research suggests that TTS is not merely a convenience to present textual content through audio. Instead, synthetic speech is a richly social modality that must be tuned to the user and the content being presented.

References

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37,* 715–727.

Aronovitch, C. D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *Journal of Social Psychology, 99,* 207–220.

Aronson, E. (1995). *The social animal* (7th ed.). New York: Freeman.

Beutnagel, M., Conkie, A., Schroeter, J., Stylianou, Y., & Syrdal, A. (1999, March). *The AT&T Next-Gen TTS system* [On-line]. Paper presented at the joint meeting of ASA, EAA, and DAGA, Berlin, Germany. Available at http://www.research.att.com/projects/tts/papers/1999_ASA_Berlin/nextgen.pdf.

Blankenship, V., Hnat, S. M., Hess, T. G., & Brown, D. R. (1984). Reciprocal interaction and similarity of personality attributes. *Journal of Social and Personal Relationships, 1,* 415–432.

Buller, D. B., & Burgoon, J. K. (1986). The effects of vocalics and nonverbal sensitivity on compliance: A replication and extension. *Human Communication Research, 14,* 548–568.

Byrne, D., & Griffitt, W. (1969). Similarity and awareness of similarity of personality characteristic determinants of attraction. *Journal of Experimental Research in Personality, 3,* 179–186.

Byrne, D., Griffitt, W., & Stefaniak, D. (1967). Attraction and similarity of personality characteristics. *Journal of Personality and Social Psychology, 5,* 82–90.

Cahn, J. E. (1990). The generation of affect in synthesized speech. *Journal of the American Voice I/O Society, 8,* 1–19.

Denes, G., Caldognetto, E. M., Semenza, C., Vagges, K., & Zettin, M. (1984). Discrimination and identification of emotions in human voice by brain-damaged subjects. *Acta Neurologica Scandinavia, 69,* 154–162.

Duck, S. W. (1973). Personality similarity and friendship choice: Similarity of what, when? *Journal of Personality, 41,* 543–558.

Eagly, A., & Chaiken, S. (1975). An attribution analysis of the effect of communicator characteristics on opinion change: The case of communicator attractiveness. *Journal of Personality and Social Psychology, 32,* 136–144.

Field, S. (1994). *Screenplay: The foundations of screenwriting.* New York: Bantam Doubleday Dell.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition.* New York: McGraw-Hill.

Fogg, B. J., & Hsiang, T. (1999). The elements of computer credibility. *Proceedings of the CHI 1999 Conference,* 80–87.

Fogg, B. J., & Nass, C. (1997, March). *Do users reciprocate to computers?* Paper presented at the annual meeting of CHI, Atlanta, GA.

Friedman, B. (Ed.). (1997). *Human values and the design of computer technology.* Stanford, CA: CSLI Press.

Fry, D. B. (1979). *The physics of speech.* Cambridge, England: Cambridge University Press.

Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation.* London: Academic Press.

Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Interaction, 53*(2), 251–267.

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.

Jellison, J. M., & Zeisset, P. T. (1969). Attraction as a function of the commonality and desirability of a trait shared by others. *Journal of Personality and Social Psychology, 11,* 115–120.

Kahan, S. (1962). *Introduction to acting.* New York: Harcourt, Brace & World.

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

Lai, J., Wood, D., & Considine, M. (2000). The effects of task conditions on the comprehensibility of synthetic speech. *Proceedings of the CHI 2000 Conference* 321–328.

Langer, E. (1989). *Mindfulness.* Reading, MA: Addison-Wesley.

Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotypes. *CHI 2000 Conference Companion,* Amsterdam, the Netherlands.

Mills, J., & Aronson, E. (1965). Opinion change as a function of communicator's attractiveness and desire to influence. *Journal of Personality and Social Psychology, 1,* 173–177.

Moon, Y. (1998a). *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.

Moon, Y. (1998b). Impression management in computer-based interviews: The effects of input modality, output modality, and distance. *Public Opinion Quarterly, 62*(4), 610–622.

Moon, Y., & Nass, C. (1996). How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research, 23*(6), 651–674.

Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human-computer interaction. *International Journal of Human-Computer Interaction, 49*(1), 79–94.

Morkes, J., Kernal, H., & Nass, C. (2000). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction, 14*(4), 395–435.

Mullenix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. W. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America, 98*(6), 3080–3095.

Murray, J. B. (1990). Review of research on the Myers-Briggs type indicator. *Perceptual and Motor Skills, 70,* 1187–1202.

Nass, C., & Gong, L. (2000). Social aspects of speech interfaces from an evolutionary perspective: Experimental research and design implications. *Communications of the ACM, 43*(9), 36–43.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.

Nass, C., Moon, Y., & Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology, 29*(5), 1093–1110.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can

computer personalities be human personalities? *International Journal of Human-Computer Studies, 43,* 223–239.

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology, 27*(10), 864–876.

Novak, D. W., & Lerner, M. J. (1968). Rejection as a consequence of perceived similarity. *Journal of Personality and Social Psychology, 9,* 147–152.

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

Picard, R. (1997). *Affective computing.* Cambridge, MA: MIT Press.

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

Ramsay, R. W. (1966). Personality and speech. *Journal of Personality and Social Psychology, 4,* 116–118.

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

Samuelson, R. J. (1999, April 5). The PC boom—And now bust? *Newsweek, 133*(14), 52.

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

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology, 8,* 467–487.

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

Scherer, K. R., London, H., & Wolf, J. J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality, 7,* 31–44.

Slobin, D. I. (1979). *Psycholinguistics* (2nd ed.). Glenview, IL: Scott, Foresman and Co.

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

Thomas, F., & Johnston, O. (1981). *The illusion of life: Disney animation.* New York: Hyperion.

Tusing, K. J., & Dillard, J. P. (2000). The sounds of dominance: Vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research, 26*(1), 148–171.

Väänänen-Vainio-Mattila, K., & Ruuska, S. (1999). Design: Designing mobile phones and communicators for consumer needs at Nokia. *Interactions, 6*(5), 23–26.

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

The White House (2000, January 21). *Information technology research and development: Information technology for the 21st century* [On-line]. Available at http://www.whitehouse.gov/WH/New/html/20000121_2.html.

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

Zillmann, D. (1991). Television viewing and physiological arousal. In J. Bryant & D. Zillmann (Eds.), *Responding to the screen: Reception and reaction processes* (pp. 103–133). Hillsdale, NJ: Erlbaum.

Received June 15, 2000
Revision received March 28, 2001
Accepted April 3, 2001 ■