

An Exploratory Study Toward the Preferred Conversational Style for Compatible Virtual Agents

Ameneh Shamekhi¹✉, Mary Czerwinski², Gloria Mark³,
Margeigh Novotny², and Gregory A. Bennett⁴

¹ College of Computer and Information Science, Northeastern University, Boston, MA, USA
ameneh@ccs.neu.edu

² Microsoft Research One Microsoft Way Redmond, Redmond, WA, USA
{marycz, margeigh.novotny}@microsoft.com

³ Department of Informatics, University of California, Irvine, Irvine, CA, USA
gmark@uci.edu

⁴ Salesforce UX, San Francisco, CA, USA
gbennett@salesforce.com

Abstract. Designing virtual personal assistants that are able to engage users in an interaction have been a challenge for HCI researchers for the past 20 years. In this work we investigated how a set of vocal characteristics known as “conversational style” could play role in engaging users in an interaction with a virtual agent. We also examined whether the similarity attraction principle influences how people orient towards agents with different styles. Results of a within subject experiment on 102 subjects revealed that users exhibited similarity attraction toward computer agents, and preferred the agent whose conversational style matched their own. The study results contribute to our understanding of how the design of intelligent agents’ conversational style influences users’ engagement and perceptions of the agent, compared to known human-to-human interaction.

Keywords: Virtual agents · Conversational style · Human-computer interaction · Social psychology · Interpersonal attraction · Similarity attraction

1 Introduction

Virtual agents increasingly play a role in human-computer interaction, assisting users in various areas such as education, health care and behavior change, marketing, and simple daily tasks. These agents serve as proxies for human representatives and are sometimes designed to engage in meaningful conversations with their users. In virtual agent-related research, it has always been a goal to construct an engaging, long-term relationship with the user [2]. Among several potential aspects of such an interaction,

A. Shamekhi, M. Novotny and G.A. Bennett---This research was done while the authors were working at Microsoft.

we feel that investigating the best conversational styles for personal agents deserves in-depth attention, given its importance in human-human relationships [14].

Research revealing that people “orient towards computers as social actors” [11] sheds a new light on human-computer agent related research. Some theories from the social psychology of personal relationships and sociolinguistics have been shown that they can be applied to a human-virtual agent dyad. Nass and Moon revealed that people mindlessly apply social rules such as politeness norms, personality attraction and gender differences while interacting with computers [12].

On the other hand, similarity attraction (aka homophily), posits that individuals are more attracted to others who match them in values, behaviors and interests. According to studies by Berscheid and Walster [5], and Byrne [3], people are generally most attracted to others with whom they share similar attitudes and styles. With human-virtual agent interaction, Nass and Reeves also showed that users are more attracted to the TTS (TextToSpeech) voice that matches the user’s own personality (e.g., extroverted vs. introverted, and dominant vs. submissive [12]). Another study by Gratch et al. has shown that a virtual agent that exhibits contingent nonverbal responses to a human dyad can effectively create rapport [8].

Research in the social psychology of personal relationships suggests numerous factors influence the quality of a social interaction. The intentions people convey during a conversation not only rely on the semantic interpretation of the words, but also on many other factors, including nonverbal behaviors [9], emotions [10], conversation starters and personality traits [6]. This line of reasoning led us to design an experiment to investigate the role of an agent’s vocal characteristics and conversational style in human-agent interaction.

In this study, we examine the speaking style of a virtual agent, and aim to explore if the similarity attraction principle in conversational style applies to human-virtual agent interaction the same way it does in human-human interaction. We designed the speaking style based on the linguistic phenomenon known as conversational style [17]. We sought to find out if people feel more satisfied and engaged when they interact with a virtual agent whose speaking style matches their own. To evaluate this hypothesis, a within subject experiment was conducted to investigate users’ attitudes toward a virtual agent speaking in different styles.

The significant amount of attention and investment in the design of realistic, virtual agents offers a promising future in which virtual agents and robots will be able to provide a wide range of human-like behaviors in social interaction. Thus, improving vocal characteristics of conversational agents’ discourse will be critical in improving the overall quality of human-agent interaction. Also, engagement and satisfaction can be two influencing factors in the success of the human-agent interaction.

2 Conversational Style in Human-Virtual Agent Interaction

Human communication through conversation is not limited to the semantic content of the words expressed by participating interlocutors. Rather, linguists and behavioral scientists have shown that interpersonal communication is also achieved through

paralinguistic features such as silence [1] and embodied features such as facial expressions [4]. These features can be utilized in various ways to construct one's own culturally-informed conversational style (CS). Deborah Tannen, a sociolinguist who pioneered this work, offers that CS is the means by which people convey meaning in daily conversations beyond solely the semantic content of the words they use. Speakers convey CS by employing variations of features such as tone of voice, pausing, and rate of speech to signal intention and relation in talk [15]. Tannen offers that CS is not a rarity; rather, "anything that is said must be said in some way, and that way is style" [16].

"Style" plays a role in determining one's personality in speech, according to Sapir [17]. He defined style as "*an everyday facet of speech that characterizes both the social group and the individual*" [14]. Conversational style is the manner in which we perform any conversational task or interaction. From the words we use to express ourselves in an interaction, to the tone, pitch, intonation, pauses, etc., all of these features constitute our conversational style. People convey meta-messages in a conversation containing information about their relationship and attitudes toward the others involved in that conversation. Tannen has stated that those meta-messages determine a speaker's style and are culturally specific [19]. In her research, Tannen observed five features of conversational style: conversation topic, genre, pace, expressive paralinguistics (e.g. relative loudness), and humor [16]. Based on the ways in which her study participants employed these features, she illuminated two main conversational styles for human interactions: "High Involvement" (**HI**) and "High Considerateness" (**HC**). For HI speakers, some practice "cooperative overlapping" in conversations. They associate interruptions with enthusiasm and *expect* the topic to change abruptly. These people usually talk a little more quickly, keep pauses short, are (a little) more verbose, are animated, show more emotion, exaggerate their feelings, express and expect overlapping talk to show enthusiasm, reciprocate (e.g., in chit chat, ask the same question the user asked in return), and provide explanations using a story about oneself. "High Considerateness" (**HC**) speakers, on the other hand, might maintain that overlapping talk is an interruption rather than display of enthusiasm, and thus, may pause slightly before responding, match the user's rate of speech. HC speakers are comparatively more succinct, yet still personable, and show emotion without exaggeration. They adjust their range/pitch/intonation to match their partner's style as well. Tannen has shown that people prefer interacting conversationally with interlocutors who match one's own style [16]. In other words, HI speakers prefer talking to those who are also HI and vice versa for HCs. It is important to note that conversational style should not be treated as a personality trait, or a permanent behavioral attribute, as it is not 100 % determined for each individual (e.g., speakers might adopt HI or HC speaking characteristics based on a number of factors). However, based on Tannen's definition, people could be more HI or HC in general in their conversations, and we explore this issue in human-agent interaction. Our initial thought was that dialog with an intelligent agent should reflect the way humans prefer to interact with other humans, but there is no research we know of that has examined this question. Our hypotheses are:

Hypothesis1: Participants will prefer to interact with a virtual agent whose conversational style matches their own conversational style.

Hypothesis2: Participants will be more engaged/interested in a virtual agent interaction when the agent’s conversational style matches their own conversational style.

3 Experiment Design

We used a within-subjects study: each participant experienced two short conversations with a virtual agent who asked about the user’s daily tasks, and tried to promote a healthy life style. Each task took about 4–5 min. The main sentences and structure in both conversations were comparably the same, and pre-scripted. The conditions are as follows:

Condition A: (High Involvement Agent) The agent talks a little more quickly, keeps pauses short, is (a little) more verbose, shows more emotion, exaggerates, and provides a story about herself using humor.

Condition B: (High Considerateness Agent) The agent pauses slightly before responding, speaks clearly and not too quickly, shows emotion without exaggeration (adjusts prosodic contour, pitch, rate, volume, and content appropriately).

Each conversation contains 5 small sub-conversations (Good morning, after lunch, after work, review for tomorrow, and pillow talk). One conversational round is with the HI agent and the other is with an HC agent. The order of conditions was assigned randomly to counterbalance any order effects.

Microsoft’s text to speech toolkit¹ was used to generate the agent’s voice. The tool allowed us to manipulate the prosody, pitch, speech rate and pause length to generate HI and HC voices for the agent. The conversation scripts were the same in both conditions, except for that in the HI condition, a few sentences were added to make the agent chattier, and humorous. Since previous research showed that the female voice is more acceptable by both female and male users, a female voice was used for the agent [13]. Examples of two scenarios for the HI and HC agent script are provided in Table 1.

3.1 The Difference Between HI and HC Conditions

We applied several manipulations to the agent’s voice to cover the five aspects of conversational style stated by Tannen. The HI agent spoke with a higher speech rate (average HI speech rate = 1.16), kept the pauses shorter, had a higher average pitch, prosody, and conveyed more emotion by using stronger adjectives (e.g., “great” instead of “good”). We considered several changes in words and phrases to distinguish the topic and genre (e.g. the HI agent tells stories from her experiences/feelings and changes the topic more often). The HI agent also conveyed a sense of humor by saying human-like phrases [see Table 1-row 2]. Alternatively, the HC agent spoke a little slower (average HC speech rate = 0.92), kept pauses longer between sentences, and showed emotion without exaggeration. In order to manage the turn-taking between the user and the agent, we designed an automated process in which the agent took the turn after a fixed number of seconds (3 s in HI condition, 5 s in HC condition). Upon the completion of each

¹ <https://www.microsoft.com/en-us/download/details.aspx?id=10121>.

Table 1. Examples of the virtual agent’s script in HI and HC conditions. The script differences over two conditions are limited. The main difference is on the speaking rate, prosody and pitch.

Agent’s CS	Sample utterance	Agent’s CS	Sample utterance
HI-after lunch talk	Welcome back Pat! I saw you went to Purple Café for lunch today. I really like there. You should try the chicken soup it’s amazing! A good lunch also fuels the rest of your day. By the way, how would you rate the cafe? Good, average, or bad?	HC-after lunch talk	Welcome back Pat! I saw you went to Purple Café for lunch today. A good lunch also fuels the rest of your day. By the way, how would you rate the cafe? Good, average, or bad?
HI-After works 1	Hey Pat, I just noticed that traffic is a bit heavy to your next appointment, Charity meeting at 6 pm. I think you should either go earlier, or pick another route. Do you want an alternative rout suggestion?	HC-After work 1	Hey Pat, Traffic is a bit heavy to your next appointment, Charity meeting at 6 pm. Do you want an alternative rout suggestion?
HI-After work 2	... I have a fantastic idea. Do you want a magic solution to relieve your stress of these hard work days? I know many people who overcome their stress just by exercising. Physical activity like jogging, going to a gym, or a swimming pool is an important part of a healthy life style. um, let’s see. The forecast for tomorrow shows, it’s a hot day. How about going swimming tomorrow at 6:30. I don’t know if it’s open tomorrow. I’m looking for that. Yay, found it. There’s one that is open tomorrow and has good reviews. I’ll add it to your calendar, sounds good?	HC-After work 2	... Many people overcome their stress just by exercising. Physical activity like jogging, going to a gym, or a swimming pool is an important part of a healthy life style. Let’s see. The forecast for tomorrow shows, it’s a hot day. How about going swimming tomorrow at 6:30. I don’t know if it’s open tomorrow. I’m looking for that. Here it is. There’s one that is open tomorrow and has good reviews. I will add it to your calendar, sounds good?

utterance by the agent, the user was shown a button to start recording his/her utterance, and then after a fixed amount of time the agent took the turn and started her next utterance. However, since the HI conversational style included many “bargue in” behaviors, the HI agent waited only three seconds after each utterance, consequently, the user was interrupted by the agent at several points during the conversation. This was by design. The length of the pauses between utterances in the HC condition was longer (five seconds), and designed so that participants never experienced an interruption from the HC agent.

4 Method

4.1 Participants

102 Amazon Mechanical Turk (mTurk) workers participated in this experiment (age range = 19 – 66, $M = 33.2$, $SD = 10.3$), 33 % were female. We had to eliminate the data for two of the participants due to technical issues. Three requirements had been set for the workers: approval rate had to be greater than or equal to 97 %, approved hits had to be greater than or equal to 1000 HITs, and they had to be from either Canada or the U.S. The

task took ~15–20 min to complete, and participants could reject completing the task at any time. Upon the completion of the experiment, participants were paid \$4 US.

4.2 Procedure

The experiment was a randomized, counterbalanced, within-subjects design with two conditions: the agent’s conversational style was either high involvement (HI) or high considerate (HC). The study objective was to have the participant interact with an agent, via a scripted dialog. After interacting with each conversational style of the agent, users were asked questions about their enjoyment and satisfaction with that agent. According to Tannen’s theory, users should prefer interacting with a person whose conversational style matches their own.

Prior to starting, each participant filled out three questionnaires, which included demographic questions, followed by a ten item personality scale (TIPI) the short form of the Big Five questionnaire [7]. Participants were then provided instructions on how to converse with the agent, using a headset and microphone. Upon completion of the questionnaires, each participant was randomly assigned to a condition, in which they conversed with an HI or HC agent. Each conversation contained five short dialogs simulating five different times of the day: morning, lunch, leaving work, reviewing one’s appointments for tomorrow toward the end of the day, and “pillow talk” (simulating before they went to bed). Each short conversation was initiated by the agent and had a scripted dialog for the user to read aloud. Users could read the script exactly or they could ad lib (They were shown this message: *You can use the text displayed to you. Try not to read the text out loud, you can first read the text and then *say it in your way**). All users’ conversations were recorded. Though the agent’s responses were all pre-scripted, the users were told that the agent was listening to them and that they should respond accordingly. Participants were asked to imagine a situation at the beginning of each sub-conversation: (e.g., “*Imagine that you just woke up and want to review your day’s schedule with your virtual assistant*”). After the first set of five conversations, participants were asked to fill out a questionnaire about their experience, and then started the second set of five conversations with the agent of the opposite conversational style. When they finished the second conversational set, they were asked to fill out a questionnaire on their recent conversations. Next, they were asked to fill out a 12-item questionnaire for measuring their own conversational style, followed by an open-ended question for the same purpose. The reason we did not ask them to fill out a CS questionnaire before the experiment was to avoid biasing the participants about their interaction. Finally, users were asked to indicate which agent they preferred and why (Fig. 1).

5 Analysis

5.1 User Conversational Style Extrapolation (CSE)

Since the scope of this study did not permit in-person observation, and no validated questionnaire exists in the literature to tease out conversational style, a rudimentary questionnaire was developed by one of the authors, a sociolinguist who studied under

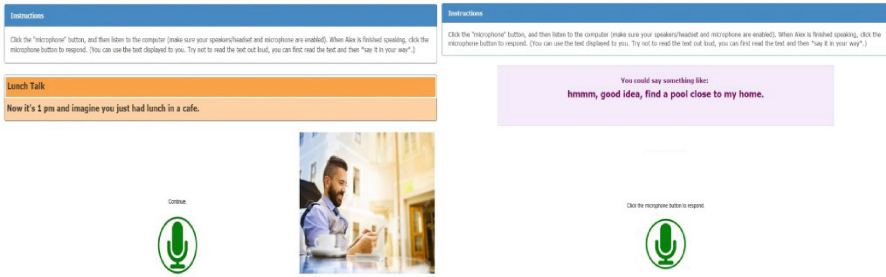


Fig. 1. Screenshots from the interaction system. When a participant clicked on “continue”, the agent started talking, and the users were provided with a microphone button to press in order to respond.

Professor Deborah Tannen, and given to participants. This questionnaire contains 12 items, asking indirectly about the users’ attitudes during a conversation. We face-validated the questionnaire by asking several experts to assess whether the questionnaire covers the conversational style concept well. We also criterion-validated the questionnaire by asking participants to elaborate on how they feel when they encounter overlapping speech during a conversation in real life. We asked each user to remember the last time (s)he had a conversation and describe the feeling when the user and the interlocutors spoke over each other. The answer to this question is a criterion variable that can help assess our measure. Three judges then extrapolated each participant’s CS based on their responses, and each participant was assigned to HC or HI accordingly. The results of the judges’ evaluation (criterion variable) significantly correlated with the questionnaire result ($X^2(1, N = 67) = 26.79, p < .0001$).

5.2 Summary of Measures Used for Dependent Variables

- 7 questions to assess the overall quality of interaction. (treated as single questions using a 5 point Likert scale: 1 = not at all to 5 = very much): would you continue interacting with the agent, how engaging, how focused, how emotional, how bored, do you trust the agent and how likeable was the agent?
- The 12 item CSE questionnaire extrapolates subjects’ conversational style (composite measure: 5 Likert scale items from 1 = strongly disagree to 5 = strongly agree). The reliability of both HC and HI indexes have been assessed. (HC: Cronbach’s alpha = 0.75, HI: Cronbach’s alpha = 0.62).
- One open-ended question about conversational behavior: Can you remember the last time you were talking with somebody and both you and that other person were talking at the same time (i.e., both of your voices could be heard saying things simultaneously in conversation)? How did this make you feel?.
- One final, open-ended question asking about the preferred agent, and the overall impression of the study.

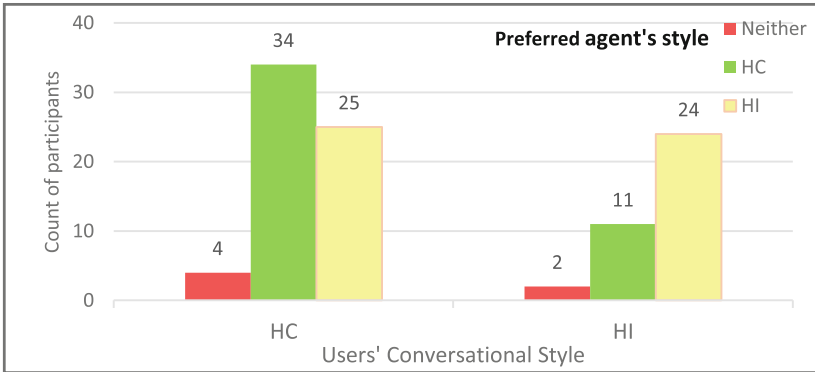


Fig. 2. Distribution of participants by their preferred agent's CS. The results show that 65 % HI users preferred the HI agent, and 54 % HC users preferred the HC agent.

6 Results

We used the CSE questionnaire results to determine the participants' conversational style: 37 participants were extrapolated as HI, and 63 participants as HC.

A chi-square test of independence was performed to determine whether people preferred the virtual agent whose conversational style matched their own or not. The chi-square test showed a significant correlation between users' style and their preferred agent. $X^2(2, N = 100) = 6.19, p < .05$. This result can be seen in Fig. 2.

We conducted a MANOVA test using a multivariate general linear model (GLM) in SPSS to account for our multiple dependent measures. The between-subject factors were users' CS (HI/HC) and Gender. Since the agent had a female voice, we entered the user's gender as a variable as we felt that gender might make a difference in the interaction assessment. The within-subjects factor was the Agent CS (HI/HC). We looked at all 2-way interactions. The results showed a significant effect of Gender ($F(7, 90) = 3.14, p < .005$) and a significant Gender by User CS interaction ($F(7, 90) = 3.00, p < .007$). User CS was not significant. For the within-subject effects, we found a significant Agent CS x User CS interaction: ($F(7, 90) = 2.61, p < .02$). The Agent CS showed a trend ($F(7, 90) = 1.93, p < .07$), and the interaction of Agent CS x Gender was not significant.

We also performed a series of paired sample t-tests to examine whether participants' responses to the 7 questions measuring the quality of the users' recent interactions with the agent differed significantly across matched (the condition in which the agent's CS was matched to the user's CS) or mismatched (Table 2). Using the Holm-Bonferroni sequential correction for multiple t-tests, the difference between the means of the two groups was significantly different only for the second question: *How engaging was your interaction with the agent?*

Since personality type of a text to speech (TTS) voice has been examined with human users [11], we examined the correlation between conversational style and personality

type. We performed a Multiple Logistic Regression, and found none of the five personality types (Extraverted, Agreeableness, Conscientiousness, Emotional Stability, and Openness to experiences) to be significantly predictive of extrapolated CS.

Table 2. Results of paired sample t-tests on the agent-interaction ratings (1 = not at all to 5 = very much). The Holm-Bonferroni correction showed only Q2 to be significant.

Question	Matched group means (SD)	Mismatched group means (SD)	p-val
Q 1(like to continue)	3.12(1.38)	2.9(1.40)	.151
Q 2(engaging interaction)	3.79(1.11)	3.36(1.13)	.005
Q 3 (focused)	4.5(0.84)	4.5(0.96)	1.000
Q 4(real feeling)	2.98(1.3)	2.83(1.34)	.218
Q 5-R(bored)	3.50(1.3)	3.48(1.36)	.90
Q 6(trust)	3.64(1.11)	3.4(1.23)	.035
Q 7(like)	3.63 (1.19)	3.33(1.29)	.032

Lastly, some feedback and open comments from participants were quite indicative of how much they preferred the agent’s style to match their own:

- P1 (HC): “The second one (HC) was much better because the speech pattern was slower and more realistic.”
- P2 (HI): “I preferred the first (HI), because she talked at a human-like pace instead of like a robot.”
- P3(HC): “I liked the first (HC). It was less stilted and the speech sounded more natural. Also, the second one was “chattier” and that got a bit annoying.”
- P4 (HI): “First agent (HI) was great. I felt as if I was talking with a real person and not a computer. The speech was clear. And the rate of speech was that of a human’s rate. Easy to have a fluid conversation and that agent to understand all of the things I asked of it.”
- P5 (HC): I preferred “First (HC) as the tone was more natural. The second seemed to be rolling pretty quickly but you could hang with her so the first was best for me.”

7 Discussion

One of the primary motivations of this study was to investigate whether users could simply identify the differences in an agent’s conversational style, and if they preferred that style to match their own. As the above results show, users distinguished the changes in the agent’s conversational style quite well (Fig. 2) (only 6 % stated either no difference or a subtle one). The strong preferences of most users for the agent with a similar CS suggests that designing an agent’s style to match the user has great potential for improving human-agent interaction.

The gender findings may indicate that the effect of CS matching is stronger for our male participants. The gender by CS interaction may indicate further that HI males are

more sensitive to the agent's CS. As this is one of the first studies of its kind on this topic, we find it provocative and interesting to explore further.

Limitations and Challenges We used a novel questionnaire which was not validated in past studies to extrapolate the participants' CS. We showed the acceptable reliability and validity for the questionnaire, however we had hoped to investigate a third measure of participants' style by analyzing the recorded verbal open-ended description of talking with a friend, but this will be part of future research. So, we cannot be completely sure that our conversational style classification was entirely accurate, though we did try to use converging lines of evidence. Our analysis of the open-ended audio tracks remains as future work. Additionally, the authors acknowledge Tannen's key finding that CS is fluid and culturally-informed in nature—there could be far more paralinguistic features to examine beyond the five we explored herein—perhaps the audio tracks will reveal more in this realm.

The participants' conversation with the agent was scripted; we were afraid that this might influence the overall flow of the conversation and subsequently, ratings of the agent. We have addressed this concern in subsequent research. The agent did not adapt to the user's style automatically, which would be an obvious goal for the future.

Despite all of these concerns, we feel we did obtain some promising results that show that users do respond positively to an agent with a conversational style that matches their own extrapolated style. In future work, we are exploring style matching further with a more naturalistic dialog interaction and a more organic, targeted process of teasing out user CS.

References

1. Basso, K.H.: To give up on words: silence in western apache culture. *Southwest. J. Anthropol.* **26**(3), 213–230 (1970)
2. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. *ACM TOCHI.* **12**(2), 293–327 (2005)
3. Byrne, D., Griffitt, W.: Interpersonal attraction. *Annu. Rev. Psychol.* **24**(1), 317–336 (1973)
4. Frith, C., Griffitt, W.: Role of facial expressions in social interactions. *Philos. Trans. R. Soc. B Biol. Sci.* **364**(1535), 3453–3458 (2009)
5. Gilbert, D.T., et al. (eds.): *The Handbook of Social Psychology*. McGraw-Hill, Boston (1998). Distributed exclusively by Oxford University Press
6. Goldman, J.A., et al.: Effect of similarity of ego identity status on interpersonal attraction. *J. Youth Adolesc.* **9**(2), 153–162 (1980)
7. Gosling, S.D., et al.: A very brief measure of the Big-Five personality domains. *J. Res. Personal.* **37**(6), 504–528 (2003)
8. Gratch, J. et al.: Creating rapport with virtual agents. In: Pelachaud, C. et al. (eds.) *Intelligent Virtual Agents*, pp. 125–138 (2007)
9. Hartzler, A.L., et al.: Real-time feedback on nonverbal clinical communication: theoretical framework and clinician acceptance of ambient visual design. *Methods Inf. Med.* **53**(5), 389–405 (2014)
10. Lopes, P.N., et al.: Emotion regulation abilities and the quality of social interaction. *Emotion* **5**(1), 113–118 (2005)

11. Nass, C., Lee, K.M.: Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *J. Exp. Psychol. Appl.* **7**(3), 171–181 (2001)
12. Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. *J. Soc. Issues* **56**(1), 81–103 (2000)
13. Payne, J. et al.: Gendering the machine: preferred virtual assistant gender and realism in self-service. In: Aylett, R. et al. (eds.) *Intelligent Virtual Agents*, pp. 106–115 (2013)
14. Sapir, E., et al.: *Ethnology*. Mouton de Gruyter, Berlin (1994)
15. Schiffrin, D., et al. (eds.): *The Handbook of Discourse Analysis*. Blackwell, Malden (2010)
16. Tannen, D.: *Conversational Style: Analyzing Talk Among Friends*. Oxford University Press, New York (2005)
17. Tannen, D. (ed.): *Framing in Discourse*. Oxford University Press, New York (1993)