Computers in Human Behavior 43 (2015) 76–84

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Carrot and stick 2.0: The benefits of natural and motivational prosody in computer-assisted learning

Susann Wolff, André Brechmann *

Leibniz Institute for Neurobiology, Magdeburg, Germany

ARTICLE INFO

Article history:
Available online 19 November 2014

Keywords:
Computer-supported learning
Speech-based communication
Feedback
Motivation
Praise and blame
Prosody

ABSTRACT

For acquiring new skills or knowledge, contemporary learners frequently rely on the help of educational technologies supplementing human teachers as a learning aid. In the interaction with such systems, speech-based communication between the human user and the technical system has increasingly gained importance. Since spoken computer output can take on a variety of forms depending on the method of speech generation and the employment of prosodic modulations, the effects of such auditory variations on the user's learning achievement require systematic investigation. The experiment reported here examined the specific effects of speech generation method and prosody of spoken system feedback in a computer-supported learning environment, and may serve as validational tool for future investigations of spoken computer feedback effects on learning. Learning performance in a basic cognitive task was compared between users receiving pre-recorded, naturally spoken system feedback with neutral prosody, pre-recorded feedback with motivating (praising or blaming) prosody, or computer-synthesized feedback. The observed results provide empirical evidence that users of technical tutoring systems benefit from pre-recorded, naturally spoken feedback, and do even more so from feedback with motivational prosodic modulations matching their performance success. Theoretical implications and considerations for future implementations of spoken feedback in computer-based educational systems are discussed.

1. Introduction

Today's learning is not anymore restricted to the classroom where human students are instructed by human teachers. Learners frequently employ educational technologies like e-learning platforms, smartphone applications or console games to improve their skills and further their knowledge. Tools like computer-assisted instruction and intelligent tutoring systems (e.g., Cognitive Tutor,[1] Andes,[2] or AutoTutor[3]) have been specifically developed to simulate human teachers' and tutors' behavior and support learners in reaching their study objectives by giving targeted assistance and adaptive feedback customized to their users' individual knowledge and performance (Larkin & Chabay, 1992; Anderson, Corbett, Koedinger, & Pelletier, 1995; Bayraktar, 2001; Shute & Towle, 2003; Woolf, 2010; Graesser, Conley, & Olney, 2012).

In the interaction with such educational technologies, like in any interaction between human users and technical systems, a multitude of dialogue forms can occur, ranging from simple commands to information retrieval dyads to elaborate 'conversations' (Allen et al., 2001). Besides the content of the system's contributions to such dialogues, an important issue to consider is the form in which the system output is generated. In principle, a tutoring system's contributions to the dialogue can be transported in a variety of ways. They can be presented through visual as well as auditory channels, and in both modalities verbal (written text or speech; e.g., AutoTutor) and non-verbal output (symbols, colors, or tones; e.g., Andes) are possible. In certain (non-tutorial) applications, even tactile feedback has been proven useful (Akamatsu, Mackenzie, & Hasbroucq, 1995). However, due to its closeness to the predominant inter-human dialogue form and based on the ever increasing demand for hands-free and eyes-free interfaces, there is a continuing trend towards speech-based system interfaces, regarding both user input and computer output (Cohen & Oviatt, 1995; Nass & Gong, 2000; Allen et al., 2001; Graesser, VanLehn, Rosé, Jordan, & Harter, 2001; Nass & Brave, 2005).

When considering this development, it becomes crucial to take a closer look at the potential effects of auditory variations in such 'spoken' system output. Especially for developers of computer-based

* Corresponding author at: Brenneckestr. 6, 39118 Magdeburg, Germany. Tel.: +49 391 6263 92161; fax: +49 391 6263 92589.

E-mail addresses: swolff@lin-magdeburg.de (S. Wolff), brechmann@lin-magdeburg.de (A. Brechmann).

[1] http://www.carnegielearning.com/secondary-solutions/cognitive-tutor/.
[2] http://www.andestutor.org/.
[3] http://www.autotutor.org/.

http://dx.doi.org/10.1016/j.chb.2014.10.015
0747-5632/

tutoring systems it is highly relevant to know not only with which words the technical tutor should respond to the learner's input, but also how the speaking system's voice should sound to optimally support the user's learning progress. Therefore, besides the content of the feedback given by the program, it is essential to empirically validate the effects of the way in which the feedback is spoken on users' learning performance.

1.1. Neutral vs. motivating prosody

One highly important aspect of speech is its prosody, i.e., the rhythm, stress, and intonation of the produced utterances. In human-to-human interaction, prosody is employed as a linguistic means serving a variety of purposes. Prosodic variations are employed by human speakers to separate the speech stream into structural units (syntactic prosody), but also to express emotions and intentions (emotional prosody). Human recipients then use these prosodic cues to analyze the syntactic structure of the sentence (parsing) and to assess the intentions and feelings of the speaker. In an educational context, teachers may employ emotional prosody to underline evaluative feedback and motivating comments which are part of the standard initiation–response–feedback (IRF) exchange (cf. Sinclair & Coulthard, 1975).

While in the last few decades a substantial amount of research has been conducted with regard to the production and processing of syntactic as well as emotional prosody in the interaction between humans (see, e.g., Frick, 1985; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Baum & Pell, 1999; Scherer, 2003; Friederici & Alter, 2004; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Wolff, Schlesewsky, Hirotani, & Bornkessel-Schlesewsky, 2008), little effort has been directed towards enlightening the role of prosody in the context of human–computer interaction, especially regarding the effects of emotional prosody produced by the computer. Available data from this field rather focused on the recognition and classification of human prosody by technical systems (cf. Cowie et al., 2001; Schuller, Rigoll, & Lang, 2003; El Ayadi, Kamel, & Karray, 2011) or on the problem of how to simulate prosodic variations in synthesized speech (e.g., Murray & Arnott, 1993; Schröder, 2001, 2009; Burkhardt & Stegmann, 2009), but in how far prosodic variations employed by a technical system might influence a human user attempting to interact with the system remains a largely open question. Even though several basic emotions are now prosodically implemented in various speech synthesizing programs (for a continuously updated overview, see Burkhardt, 2014), thereby rendering an examination of their impact feasible, controlled empirical studies in this regard are still pending.

Computer-assisted learning can be expected to be particularly susceptible to the impact of prosodic variations in the system output, since it constitutes an environment in which feedback given by the technical system plays an important role for the user's learning progress and the user is thus effectively dependent on the system feedback. Therefore, it is especially interesting to examine whether prosodic modulations in system feedback can enhance users' learning success in the task they are trying to complete with the system's help.

In the context of learning, it is opportune to select prosodic variations that are known to have a *motivating* effect in human learning, since—besides other factors like intelligence, task-specific skills, or attention—the learner's motivation has long been known to have a substantial effect on learning success (Stipek, 2001; Pintrich, 2003), especially in situations where learning is self-regulated (Zimmerman & Schunk, 2012, chap. 1) and computer-assisted (Song & Keller, 2001).

An individual's motivation for a particular behavior is partially based on dispositional sources which are relatively stable and unaffected by external factors, e.g., the individual's need for achievement (Murray, 1938; McClelland, Atkinson, & Clark, 1953; McClelland, 1985; Thrash, Elliot, & Schultheiss, 2007), esteem needs (Maslow, 1954), need for competence (Deci & Ryan, 1985, 2002), causal attributions for success and failure (Weiner, 1979, 1985, 2005, chap. 5), or goal orientation (Dweck, 1986; Ames & Archer, 1988; Elliot, 2005, chap. 4), and partially stems from the intrinsic interest and pleasure in the behavior itself (e.g., Csikszentmihalyi, 1975; Renninger, 2000). However, the desired behavior can also be extrinsically motivated by the prospect of desirable outcomes (Ryan & Deci, 2000a, 2000b).

In educational initiation–response–feedback exchanges, praise and blame are frequently consulted for this purpose: Here, positive feedback (i.e., informative feedback following a correct response) is often accompanied by praising comments, while negative feedback (i.e., feedback following an incorrect response) often entails blame or criticism in addition to its informational content. It is noteworthy that there may also be paradoxical effects of praise and blame on students' motivation; however, these appear to be limited to specific circumstances, like excessive praise for success in a task perceived as very easy, or in a task that other students did not receive praise for (e.g., Meyer, 1992; Miller & Hom, 1996; Kaspar & Stelz, 2013). Similarly, negative effects of praise on students' achievement have been reported in cases where praise has been administered unsystematically, i.e., independent of the correctness of the students' responses (Brophy, 1981). If, on the other hand, praise and blame are employed in a contingent fashion focusing on the student's individual mastery of the task, they can be considered effective tools of operant conditioning (cf. Skinner, 1953; see also O'Leary & O'Leary, 1977; Brophy, 1981; Henderlong & Lepper, 2002; Pintrich, 2003; Hattie & Timperley, 2007). Couched in this framework, praise (or more generally, positive reinforcement) is employed to enhance a desired behavior (i.e., correct responses), while blame (or more generally, punishment) serves to reduce the occurrence of undesired (i.e., incorrect) behavioral responses.

At this point, it is useful to consider how the assumed effects of praise and blame may translate to feedback given by a technical system. Since there is ample evidence that users tend to treat computers like human beings – attributing them with emotions and intentions even though they are fully aware that computers are not human (i.e., the CASA—Computers Are Social Actors—Paradigm; Reeves & Nass, 1996; Nass & Moon, 2000; Lee & Nass, 2010, chap. 1), it is reasonable to assume that such conditioning mechanisms may have similar effects if employed by a technical system. Following the effective usage of praise and blame in human education, it therefore appears suitable to examine the implementation of praising and blaming prosody into the feedback given by a technical tutoring system. If computer-assisted learning is in fact susceptible to the usage of prosodic praise and blame, we should thus be able to observe an improved learning performance with prosodically motivational feedback in comparison to prosodically neutral feedback.

On a semantic level (i.e., regarding the textual content of the utterances, not their prosody), at least the effects of computer-generated praise have been examined before (Fogg & Nass, 1997). While the authors did not report any effects on actual task performance, they showed that (written) praise given by a computer can have beneficial effects on the users' subjectively perceived performance, their mood, and their evaluation of the computer. Since this was the case with "sincere" praise (i.e., praise described to participants as contingent upon correct responses) as well as with "flattery" (i.e., praise described as independent of response correctness), the authors suggested that "computers should praise people frequently—even when there may be little basis for the evaluation" (Fogg & Nass, 1997, p. 559). Similarly, Mumm and Mutlu (2011) observed an increase in self-reported motivation and willingness to continue with the task when written praise was given irrespective