KEER2010, PARIS | MARCH 2-4 2010
INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2010

STUDY ON CHILDREN'S PREFERENCE FOR AUDITORY AND VISUAL SOCIAL CUES DURING INTERACTION WITH COMPUTERS

Fang-Wu Tung[*a], Yi-Shin Deng[b] and Kuen-Meau Chen[a]

[a] *National United University, Taiwan*
[b] *National Chiao Tung University, Taiwan*

ABSTRACT

This study aims at enhancing the sociability of instructional media by adding auditory and visual social cues to engage children's attention in e-learning environments. The social cues used in the study are speech and facial expressions, both of which are essential in interpersonal communication. A quasi experiment was conducted to explore how the rendering of these two social cues affects children's attitudes toward computers and their motivation as they participate in e-learning environments. A total of 40 girls and 46 boys participated in the study. Data were collected via a questionnaire that probed into the perceived social presence, similar attraction toward computers, and intrinsic motivation. The results demonstrate children's preference for speech in auditory modality over facial expressions in visual modality. It is found that speech is more capable of eliciting children's social perception and motivating them in learning activities. Speech feedback works better than facial-expression feedback in enabling children to feel stronger social presence and yield similar attraction to computers, as well as in stimulating their motivation. The findings may help designers in choosing between social cues of speech and facial expressions for creating more sociable interfaces for children.

Keywords: *Social Cue,Interface design, Speech, Facial Expression, Child*

Corresponding author: Postal address: 1, Lienda, Miaoli, Taiwan, 36003 email address: fwtung@gmail.com

1. INTRODUCTION

1.1. Enhancing social presence of instructional media

Computers have expanded the available learning media for children and provided them with a potential way to learn by themselves. In addition to the issue of usability, the design of instructional material needs to improve engagement and motivation for younger users [1]. Computer-mediated learning initially appealed to many learners because of the novelty of the learning experiences. As the excitement brought about by the novel features diminishes, it becomes more challenging to stimulate and sustain learners' motivation while they engage in computer learning activities. Since learning is a social activity, the learner's desire for social interaction should be taken into consideration while designing an e-learning environment. In fact, enriching social presence to create social contexts or facilitate social interaction among participants in e-learning environments is considered necessary for enhancing and improving effective instruction [2, 3]. As stated by Rifkind [4], a lack of social presence may lead to greater frustration and less effective learning. Thus, social presence could help counterbalance the isolated nature of e-learning and create a learning environment that is perceived as warm and approachable for all those involved [5] and is therefore supportive of learning objectives.

Most related research focuses mainly on the enhancement of the computer's social presence necessary for creating a successful learning experience involving learners and instructors in an online environment. In addition to supporting collaborative learning environments, computer used as learning technology can function as tutor, tool, and tutee for individual self-paced learning [6]. When learning in this way, the sense of social presence is required even more because the learner is disconnected from other peers and tutors. In the light of the CASA (computers as social actors) paradigm, computers that exhibit social cues can convey a sense of sociability and intimacy and thereby induce social responses from people, leading people to treat computers in the way ~~as~~ they treat other people [7]. Thus, computers may be perceived as a social entity for improving the learning experience while a single learner participates in a computer-based learning activity with no other participants involved. In other words, we can design e-learning environments to support simulated social interactions between a learner and the computer by employing social cues borrowed from social contexts [8].

1.2. Social cues: Speech vs. facial expressions

As concluded above, adding social cues to interface design can allow users to perceive computers as social entities and therefore deliver social responses to them. According to previous findings, social cues such as speech and human face have been widely employed to enrich social attributes of the media. This study aims at investigating the effects of two social cues, speech and facial expressions, in e-learning environments on children's feelings of social presence, similar attraction toward computer, and their motivation in learning. Speech plays a dominant role in interpersonal communication. As long as social groups continue to be an integral part of human life, human sensitivity to voice and language cues has a critical role to play in the interactions among people. On the other hand, the face, being the most important channel of emotional expression, does play a significant role in social communication.

Therefore, symbols of facial expressions have been applied in human-computer interfaces as a means of communication.

Speech and facial expressions are powerful cues in human-human interaction. The two are conveyed via the auditory modality and visual modality, which are two important components in multimedia. Adding visual components like facial expressions or sound effects such as speech to instructional media does not only make them more appealing, but can render the media sociable and friendly, thus enabling young learners to feel warm and motivated in the learning process. In order to achieve such by utilizing the two social cues, the way children process visual and auditory information should be studied further because their modality preference may influence how they perceive while interacting within e-learning environments. Research reveals that young children are more likely to process auditory stimuli [9]. Young children have a preference for auditory information and the bias changes in the course of development [10]. A differential modality preference in children and adults was found; children aged below 4 rely mainly on auditory stimuli while adults rely mainly on visual inputs. However, school-age children's modality preference is not well understood and thus remains to be explored. Pupils are one of the main user groups of e-learning; hence, understanding their preference might be useful for instruction design.

Posner et al. [11] argued that visual stimuli are less likely to automatically engage attention than auditory stimuli, and people have to learn to direct their attention to visual information. Compared with texts and symbols in the visual channel, speech in the auditory channel is cognitively less demanding [12]. Moreover, speech is a dynamic process that can trigger people's focus on the content and keep the audience in awareness during its creation [13], thus exerting a high degree of social presence.

As seen from the above, the sound of speech is effective in gaining users' attention; nevertheless, Zaidel and Mehrabian [14] claimed that facial expressions were more important than voices in interpersonal interaction. Speech may transmit emotional messages by changing pitch, tempo as well as loudness, but research revealed that emotions transmitted by facial expressions were more precise than those conveyed by speech [15]. Burns and Beier [16] also pointed out that visual cues were also more influential and accurate than vocal cues in the designation of a mood state.

In short, both speech and facial expressions have their own unique characteristics in interpersonal communication. Speech plays a dominant role in interpersonal communication since humans are voice-activated and more sensitive to verbal sounds, and facial expressions are an important channel of nonverbal communication, which are regarded more trustworthy than words [17]. Thus, this leads us to hypothesize that the educational media equipped with feedbacks of speech and facial expressions may have similar effects on improving children's perception of computers and their motivation. If the auditory and visual social cues used in instructional media can equally trigger children's social schema and boost their motivation, then they serve as good alternative solution for each other. This can benefit instruction designers in allowing flexibility in applying the two modalities to interface design to meet some special requirements. For example, an auditory format may be useful for applications that require visual attention to be directed elsewhere and vice versa. This research focuses on the two social cues, namely speech and facial expressions within interpersonal

communication, and compares their possible influence exerted on children within e-learning environments. Experiments are conducted to understand the influence of speech and facial expressions on children's perception and learning motives on computers in e-learning context.

2. ASSUMPTIONS AND PREDICTIONS

Social presence: According to the idea that speech and facial expressions convey social traits and play essential roles in human-human interaction, it is expected that they both can be used as social cues for interface design and work similarly on children's feelings of social presence during their interaction with computers.

Similar attraction: The similarity-attraction hypothesis does not only apply to personality, but can also be expanded to relationships between human and computer [18]. If children can feel the sense of social presence from a computer itself, the computer may appear more as a social actor than a tool to them. It is predicted that computers with either of the two social cues can lead users to generate similar attraction.

Motivation: This study explores whether computers equipped with social cues could improve children's intrinsic motivation for learning. Learning is a social activity. A learning environment provides a child with social presence that may satisfy one's social desire and enhance intrinsic motivation while participating in learning activity alone. Thus, it is predicted that speech and facial expressions can work equally for bettering learning experiences and motivating learners.

3. METHOD

3.1. Participants

Eighty-six sixth graders (40 girls and 46 boys) from three classes of two elementary schools in Hsinchu participated in the experiment. The average age was 11.4 years. All subjects have used computers and have received computer instruction at least once a week in the schools since third grade. A total of 83 percent of the subjects reported that they often used computers at home for game playing and web browsing.

3.2. Experimental design

The design was within-subjects with one factor with two levels (speech and facial-expression). The dependent variables were the perceived social presence, children's similar attraction towards computers, and intrinsic motivation.

3.3. Experimental materials

The instructional material was prepared as a math problem-solving practice program designed in Macromedia Flash. There are two versions of the program developed respectively with speech and facial-expression feedbacks during interaction. Feedbacks provided include a greeting at the beginning, feedback on correct or wrong answer, and giving comments on the subject's performance at the end. Speech for this study was created

from recordings made by a young female. Facial expressions were depicted in small yellow faces with different emotions that matched different situations.

3.4. Measurements

Social presence: The first set of questions adopted the four items proposed by Short, Williams, and Christie [19] for measuring social presence, namely sociable/unsociable, personal/impersonal, sensitive/insensitive, and warm/cold, using a semantic differential technique. The semantic differential technique was used with bipolar five-point scales. (Cronbach's □ = 0.85)

Similar attraction: Similar attraction is measured by social and intellectual attractions which are two important criteria in the similarity-attraction hypothesis of ~~the~~ interpersonal relationship [20]. The questions were adapted from Moon [21] for measuring subjects' perceptions of the computer with respect to similar attraction. The scale consists of six items: Do you find the computer friendly, competent, clever, and intelligent? Do you like this computer? Do you like doing math exercise with this computer? (Cronbach's □ = 0.87)

Intrinsic motivation: The questions were adapted from the Activity-Feeling Scales (AFS) developed by Reeve and Sickenius [22] for measuring subjects' intrinsic motivation. The 12-item measure was made up of four 3-item scales for assessing self-determination, competence, relatedness, and tension. The name and individual items for each index are as follows: self-determination-- offered choice what to do, I want to answer the questions, and my participation is voluntary; competence-- capable, competent, and achieving; relatedness-- imitate, involved with friends, and part of a team; and tension-- pressured, uptight, and easy (reverse). (Cronbach's □ = 0.78)

3.5. Procedure

The study took place in the computer lab of the two schools during one of their computer class sessions. Each computer was separated by some distance from one another to prevent interference. Each student was assigned to one computer in the lab. The speech and facial-expression conditions are counterbalanced. The experiment took around 100 minutes. Upon completing the experiment, the subjects were debriefed and rewarded with a toy.

4. RESULTS

4.1. Social presence

As shown in Table 1, participants rated the sense of social presence stronger in the speech condition (M=7.82, Sd=2.51) than in the facial-expression condition (M=6.79, Sd=2.33). A paired t-test was performed to compare the social presence the participants perceived between the two conditions. A statistically significant difference was observed ($t = 4.34$, $p < 0.001$). The results indicated that children perceive a higher degree of social presence while interacting with computers that provide speech feedback than with those that provide facial-expression feedback.

4.2. Similar attraction

The results (Table 1) show that the subjects exhibited higher similar attraction toward computers with speech feedback (M = 6.37, Sd = 1.87) than those with facial-expression feedback (M = 5.94, Sd = 1.77). A paired t-test was performed to compare the participants' similar attraction toward the computer between the two conditions. A statistically significant difference between the speech and facial-expression conditions was observed among subjects ($t = 2.53$, $p < 0.01$).

4.3. Intrinsic motivation

Table 1 demonstrates that participants reported higher intrinsic motivation in the speech condition (M = 5.82, Sd = 1.30) than in the facial-expression condition (M = 5.36, Sd = 1.33). A paired t-test was performed to compare the participants' intrinsic motivation between the two conditions. A statistically significant difference between the speech and facial-expression conditions was observed ($t = 4.49$, $p < 0.001$). This reveals that e-learning environments providing speech can motivate children more than those providing facial expressions.

Table 1: Results of the study

Dependent variable	Independent variable	Mean	SD	N	T value
Social presence	speech	7.82	2.51	86	4.34
	Facial-expression	6.79	2.33		
Similar attraction	speech	6.39	1.87	86	2.53
	Facial-expression	5.94	1.77		
Intrinsic motivation	speech	5.82	1.30	86	4.49
	Facial-expression	5.36	1.33		

5. DISCUSSION

This study investigated the application within e-learning environments respectively with speech and facial expressions acting as social cues to engage children in learning. As mentioned earlier, speech and facial expressions are essential in interpersonal communication. The two were expected to be used as social cues in e-learning environments and have similar effects on enhancing children's attitudes toward computers and boosting their motivation. The analysis of the data from the current study, however, indicates that the two social cues made a significant difference in the participants' perception and motivation. The participants receiving speech feedback tended to perceive computers on the social dimension and reported stronger motivation.

Speech has unique properties which can be associated with humans. It is not only true that speech in the auditory channel draws directly people's attention, but it also originates from the evolution of human social interaction. Therefore, when people hear speech, the social presence linking them to other people is achieved. Clearly, speech can create stronger gains in social presence and elicit more attention from others in the process of interaction than text or symbols ~~can~~. Since the social responses delivered by instructional media are easy to manipulate, people's social awareness toward the media is triggered not only by fancy or sophisticated interface but also by voice-based interface. Thus, instruction designers could

promote human-computer interaction to the level of human-human interaction by adding speech feedback.

Compared with speech feedback, facial-expression feedback does not seem sufficient for inducing children's social responses and engaging them in e-learning activities. As to whether school-age children show differential preferences for auditory social cues and visual social cues while interacting with computers, the findings reveal that children prefer speech in auditory format to facial expressions in visual format. To elicit unconscious social responses, a computer still has to exhibit enough cues to make people feel the computer worthy of social responses [23]. In other words, it might mean that facial expressions are not as sufficient as speech in functioning as social cues for interface design. Nevertheless, human brain processes facial images differently from images of all other objects, just as the way it processes voices differently from all other sounds [24, 25]. In social contexts, humans depend highly on facial expressions during communication, which also means that they are important elements of the communication message. Given that, it is plausible that symbols of emotional face icons have a limitation and may fail to represent facial expressions.

Research has shown that emotional expressions by synthetic faces are recognized as well as emotions on natural faces [26, 27]. People can recognize emotions in facial expressions, even in low-quality photos or line drawings, but facial expressions are more than just static symbols. Ekman and Friesen [28] categorize three types of signals provided by the human face: static, slow, and rapid. They emphasize that emotion messages are transmitted by rapid facial signals, not by static facial signals; this finding may affect the implication of an emotion message. Rapid facial signals mean that facial muscles stretch and contract, and there are visible changes in the appearance of the face when these feelings occur. Indeed, a person's smile and or expression of worry never has a fixed pictorial image but features as a singular, transient and alternating message; static facial expression symbols are unnatural representations. Consequently, whenever designers used facial expressions as social cues for interface design, it was required to pay special attention to features in real life and observe insightfully the way they could be demonstrated on people's faces.

As stated above, facial displays of emotion are a dynamic phenomenon and thus a static symbol is an unnatural representation. The dynamic display of facial expressions provides unique temporal information about the expressions that is otherwise absent in static displays. Studies have examined the dynamic effect on the intensity of emotional expressions and suggested that dynamic characteristics are necessary for full extraction of emotional information from faces [29, 30]. In view that dynamics is important in facilitating the perception of facial expressions, we will conduct further research to investigate the effects of dynamic facial expressions versus speech on children to see whether children demonstrate different preference for the two social cues in visual and auditory modalities. The subsequent study may allow us to obtain the knowhow of utilizing social cues like speech and facial expressions for enriching social attributes of computer-based applications and at the same time deduce a comprehensive understanding about children's preference for social cues in auditory or visual modality.

REFERENCES

1. Druin, A. & Inkpen, K. When personal technologies are for children. *Personal Technologies*, Vol. 5, No. 3, pp.191-194, 2001
2. Aragon, S.R. Creating social presence in online environments, *Facilitating learning in online environments* (*eds Aragon, S.R.*), San Francisco: Jossey-Bass, pp. 57-68, 2003
3. Gunawardena, C.N., Social presence theory and implications for interaction and collaborative learning in computer conferences, *International Journal of Educational Telecommunications*, Vol. 1, No. 2, pp.147-166, 1995
4. Rifkind, L. J. Immediacy as a predictor of teacher effectiveness in the instructional television. *Journal of Interactive Television*, Vol. 1, No. 1, pp.31-38, 1992
5. Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. Assessing social presence in asynchronous, text-based computer conferencing. *Journal of Distance Education*, Vol. 14, No. 3, pp.51-70. 1999
6. Tylor, R. P. *The computer in the school: Tutor, tool, tutee*. New York: Teacher's College Press. 1980
7. Nass, C , Steuer, J., Tauber, E. Computers are social actors, *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, Boston, Massachusetts. 1994
8. Tung, F. W., & Deng, Y. S. Designing Social Presence in E-Learning Environments: Testing the Effect of Interactivity on Children, *Interactive Learning Environments,* Vol. 14, No. 3, pp.251-264. 2006
9. Sloutsky, V. M, & Napolitano, A.C. Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, Vol. 74, No. 3, pp. 822-833. 2003
10. Robinson, C. W. and Sloutsky, V. M. Auditory dominance and its change in the course of development. *Child Development*, No. 75, pp. 1387-1401. 2004
11. Posner, MI, Nissen, MJ, & Klein, RM. Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, No. 83, pp. 157–171. 1976
12. Kroll, B.M. Cognitive egocentrism and the problem of audience awareness in written discourse. *Research in the Teaching of English*, No.12, pp. 269-281. 1978
13. Chalfonte, B. L., Fish, R.S., & Kraut, R.E. Expressive Richness: A Comparison of Speech and Text as Media for Revision, In: Robertson, Scott P., Olson, Gary M., Olson, Judith S. (ed.): *Proceedings of the ACM CHI 91 Human Factors in Computing Systems Conference* (pp.21-26). New Orleans, Louisiana. 1991
14. Zaidel, S. F., & Mehrabian, A. The ability to communicate and infer positive and negative attitudes facially and vocally. *Journal of Experimental Research in Personality*, No. 3, pp. 233-241. 1969
15. Ekman, P. Facial expression and emotion, *American Psychologist*, No. 48, pp. 384-392. 1993
16. Burns, K.L., & Beier, E.G. Significance of vocal and visual channels in the decoding of emotional meaning. *Journal of Communication*, No. 23, pp.118–130. 1973
17. Ekman, P. & Friesen, W. V. *Unmasking the face. A guide to recognizing emotions from facial clues*. Englewood Cliffs, New Jersey: Prentice-Hall. 1975
18. Moon, Y. & Nass, C. How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research*, Vol. 23, No. 6, pp. 651-674. 1996
19. Short, J., Williams, E. & Christie, B. *The social psychology of telecommunications*. London: John Willey & Sons Ltd. 1976

20. Singh, R. & Ho S.Y. Attitudes and attraction: A new test of the attraction, repulsion and similarity-dissimilarity asymmetry hypotheses, *British Journal of Social Psychology*, Vol. 39, No. 2, pp. 197-211. 2000
21. Moon, Y. *Similarity effects in human-computer interaction: Effects of user personality, computer personality, and user control on attraction and attributions of responsibility*, Unpublished doctorial dissertation, Stanford University, Stanford, CA. 1996
22. Reeve, J., & Sickenius, B. Development and validation of brief measure of the three psychological needs underlying intrinsic motivation: The AFS scales. Educational Psychological measurement, No. 54, pp. 506-515. 1994
23. Nass, C., Moon, Y. Machines and Mindlessness: Social Responses to Computers. Journal of Social Issues, Vol. 56, No. 1, pp. 81-103. 2000
24. Allison T., McCarthy G., Nobre A., Puce A., Belger A. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex*, No. 5, pp. 544-554. 1994
25. Nass, C. & Brave, S. *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, MA: MIT Press. 2005
26. Katsikitis, M. Classification of facial expressions of emotions: a multidimensional scaling approach. *Perception,* No. 26, pp. 613–626. 1997
27. Bartneck, C. How convincing is Mr. Data's smile: Affective Expressions of Machines, *User Modeling and User-Adapted Interaction*, Vol. 11, No. 4, pp. 279-295. 2001
28. Ekman, P. & Friesen, W. V. *Unmasking the face. A guide to recognizing emotions from facial clues*. Englewood Cliffs, New Jersey: Prentice-Hall. Pp. 10-20. 1975
29. Ambadar Z, Schooler JW, Cohn JF. Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions, *Psychological Science,* No.16, pp. 403–410. 2005
30. Biele, C., & Grabowska, A. Sex differences in perception of emotion intensity in dynamic and static facial expressions, *Exp Brain Res*., Vol. 171, No. 1, pp. 1-6. 2006