

The Effect of Metaphoric Gestures on Schematic Understanding of Instruction Performed by a Pedagogical Conversational Agent

Dai Hasegawa^(✉), Shinichi Shirakawa, Naoya Shioiri, Toshiki Hanawa, Hiroshi Sakuta, and Kouzou Ohara

Aoyama Gakuin University, 5-10-1 Chuo-ku Fuchinobe,
Sagamihara-shi, Kanagawa, Japan
hasegawa@it.aoyama.ac.jp

Abstract. In this paper, we examine the impact of metaphoric gestures performed by Pedagogical Conversational Agent (PCA) on learners' memorization of technical terms, understanding of relationships between abstract concepts, learning experience, and perception of the PCA. The study employed a one-factor three-level between-participants design where we manipulated gesture factor (speech-gesture match vs. speech-gesture mismatch vs. no-gesture). The data of 97 students were acquired in on-line learning environment. As the results, while there was no effect found on memorization of technical terms, we found that students showed accurate schematic understanding of the relationship between abstract concepts when the PCA used metaphoric gestures matched to speech content than when used gestures mismatched, and no gesture. Contrary to the result, we also found that students judged the PCA useful, helpful, and felt the PCA looked like a teacher when performed mismatched gestures to speech content than when performed matched gesture.

Keywords: Pedagogical agent · Metaphoric gesture · Understanding · Reliability

1 Introduction

New education paradigm, as typified by the flipped classroom model, has changed the focus of education — from offering students knowledge to developing students' experience, from the teacher as a transmitter of knowledge to the teacher as a facilitator of learning. In such education process, students are often required to learn basic knowledge and skills by themselves before the classroom activity. And, the efficacy of the classroom activity depends on the initial knowledge students have. The development of effective self-education tools, therefore, becomes an important issue in modern education.

Among various computer-assisted self-education environments, a Pedagogical Conversational Agent (PCA) based environment suggests a distinct direction, introducing social relations between human and computer by utilizing

human-likely embodied features [15,18]. The previous studies in the field have confirmed preferable effects of PCA on students' understanding and motivation, compared to text-based/audio-based instruction [3,13,21]. Furthermore, to better design effective PCAs, a variety of findings in human-human communication were adopted. For example, the roles of PCA as a learning partner, and the impacts of stereotyped appearances, facial expression, audio expressions, and posture has been examined so far [4,11,12,23]. Likewise, the impact of gesture are naturally of great concern [18,24,28].

In general conversation, it is confirmed that speakers' gestures help listeners to comprehend their speech content [9,16,17]. Similarly, not only recent studies in human-human educational interaction have also pointed out that gestures in scientific explanation play an important role on students' understanding and collaborative learning [2,25], but also some empirical studies have confirmed the learning efficacy of the use of gestures in education [10,26,27].

However, little studies in the field of PCA research addressed the impact of gesture and the types of gesture. Buisine et al. reported that when PCA used redundant gestures, which convey the same information included in audio instruction, the performance of verbal recall test increased, compared to when PCA used complementary gestures, which convey new information not included in audio interaction [6,7]. Also, Buisine et al. showed that the redundant use of gestures increased the perception of quality of explanation, and likability and expressiveness of PCA.

Buisine et al. examined the redundancy and complementarity of the gesture use, but they mainly employed pointing gestures which refer a certain part of a learning material and gestures which present pictorial features of concrete objects. A gesture can convey abstract meanings as well. This aspect of gesture has not addressed yet, despite the importance of conveying abstract concepts in educational instruction.

To examine the impact of gesture which presents abstract meanings in PCA based learning system, in this paper, we conduct an on-line experiment in which we evaluate learners' ability to complete a vocabulary recall test, a figure selection test, and answers to a questionnaire, when the versions of the system use gestures matched with abstract speech content, gestures mismatched with speech content, or no gesture at all. The design allows us to examine the impact of the gesture on memorization of technical terms, understanding of the abstract concepts, learning experience, and perception of the PCA.

In the following section, we will explain the specific type of gesture we aim to examine in this paper.

2 Conduit Metaphoric Gesture

Gestures, usually movements of the hands and arms, which spontaneously occur accompanying with speech are classified into four types, iconic, metaphoric, beat, deictic [19].

Iconic gestures present images of concrete objects or actions, and deictic gestures are used to indicate objects around the speaker. Iconic and deictic may be

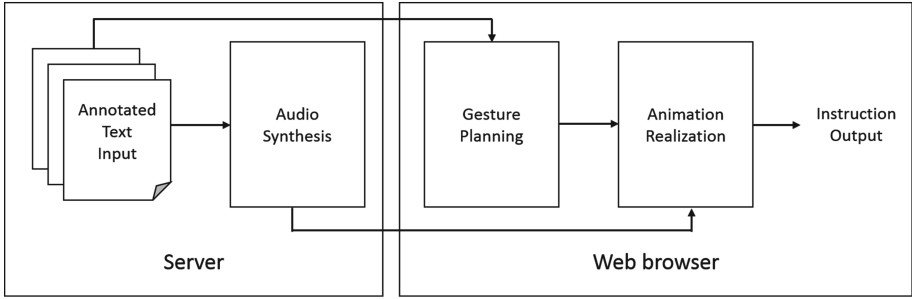


Fig. 1. System overview

called gesture of concrete. Beats are the movements with the rhythmical pulsations of speech which index the accompanied word of phrase as being significant.

Contrary to those three types, in metaphoric gestures, abstract meaning is presented as if it had form by utilizing space. For example, a speaker appears to be holding a object, meaning an abstract object, such as idea, memory, and etc. This is most often seen metaphoric gesture and it is a gestural version of the ‘conduit’ metaphor. The appearance of conduit metaphoric gestures are typically shown as a cup-shaped hand or hands being holding something, where the hand/hands represent as a container and word/phrase as a substance to be transferred to the listeners [8, 20]. In addition, the conduit metaphoric gestures are often used to show the relationship between several concepts by being put in space.

We will examine the impact of the metaphoric gesture performed by a PCA. As the first step of our investigation, we will focus on the three types of conduit metaphoric gestures and its combinations utilizing space as described below.

Conduit Gesture: The appearance of holding an object with the both hands, indicating a single abstract concept.

Two-conduit Gesture: The movements of the object-holding hands from the right side of the speaker’s upper body to the left side, synchronized with the word/phrase it accompanies in speech, indicating a schematic relationship between two abstract concepts by utilizing space.

Three-conduit Gesture: The movements of the object-holding hands from the right side of the speaker’s upper body to the center, then, the left side, synchronized with the word/phrase it accompanies in speech, indicating a schematic relationship between three abstract concepts.

3 Method

In this section, we will describe our PCA based on-line learning environment, learning material, and the details of the study design.

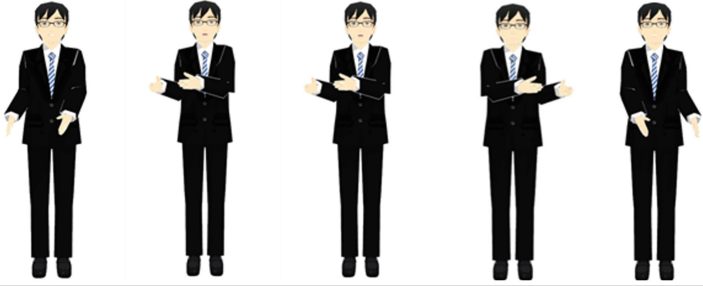
Gestures realized in a web-browser					
Gesture phase	pre-stroke	stroke	hold	stroke	post-stroke
Script Annotation	ウェブページを/ [サーバに/ 保存して]おき、 / [ユーザに]/ 提示する <i>CR</i> <i>CL</i>				
(translated from Japanese)	... web-page-wo/ server-ni/ hozon-shite-oki,/ user-ni/ teiji-suru ... (... stores a web-page [in a server], then submits it [to a user] ...)				

Fig. 2. Metaphoric gestures representing the relationship between two abstract concepts ‘server’ and ‘user.’ ([: stroke and hold)

3.1 E-Learning System and Materials

We developed a web-based learning environment where a PCA gives a speech instruction accompanying with gestures. The animation is processed and depicted on a web browser by using a program written in HTML5, JavaScript and WebGL (a graphic library written in JavaScript) so that participants can take part in the experiment without any special preparation of their computer. Figure 1 shows a system overview. The system processes text inputs with gesture annotation sentence by sentence. The sentence is divided into phrases (in Japanese, a phrase is consist of a content word and function words). The phrases are annotated by using the following schema: *CC* (a conduit gesture appears holding something with both hands in front of the speaker), *CL* (a conduit gesture appears holding something with both hands at the left side of the speaker), *CR* (a conduit gesture appears holding something with both hands at the right side of the speaker). The annotation is interpreted on a web-browser, then gesture animations are realized, synchronizing with audio. In order to synchronize gesture and speech, we estimated the timing of each gesture by indexing the order of the first character of each phrase from the top of the sentence, given that the synthesized audio’s speed is constant. Lip-sync animations are also realized, synchronizing with audio.

We prepared an instruction script (4-minutes-long speech when audio synthesized) to be performed. The learning subject was the basics of web application development. The script explains the basic concepts of web application, the components of web application, and the programming languages for web application developments. A part of the instruction performed by the system is shown in Fig. 2.

3.2 Independent Variables

There is one independent variable, gesture, in the study, and we have three treatment conditions described as below.

- C1. Speech-gesture Match:** Two of authors who were familiar with web application development annotated the script with the schema: *CC*, *CL*, *CR*. The realized instruction included twenty-seven *conduit gestures*, twelve *two-conduit gestures*, and three *three-conduit gestures*. And overall, the number of gesture strokes was sixty.
- C2. Speech-gesture Mismatch:** Based on the annotation data described above, we randomly changed the tags to one of others, resulting the realized animation included forty two *conduit gestures*, fifteen *two-conduit gestures*, and five *three-conduit gestures*. The number of strokes was eighty nine. The percentages of speech-gesture mismatches in the realized animation were 25 % for *conduit gestures*, 65 % for *two-conduit gestures*, 60 % for *three-conduit gestures*.
- C3. No Gesture:** No gesture is performed and only audio instruction is presented with lip-sync animations.

3.3 Dependent Variables

Dependent variables included memorization of technical terms, abstract concepts understanding, learning experience, and perception of the PCA. To evaluate those dependent variables, three types of data are used: performance of a vocabulary recall test, performance of a figure selection test, and answers to a forty-six item questionnaire.

Vocabulary Recall Test: Participants were asked to answer the technical terms explained in learning material, such as “Answer three programming languages often used in server-side.” In total, they were required to answer fifteen technical terms by six questions.

Figure Selection Test: Participants were asked to choose all appropriate figures which describe the relationship between abstract concepts: the interaction between components of web application (user, client, server, and database), the relationship between kinds of application program (general application program, desktop application program, web application program), and the relationship between kinds of programming languages (programming languages for general purpose, server-side development, and client-side development). Nine appropriate figures and eleven inappropriate figures were presented, we counted the number of figures correctly being selected and correctly not being selected.

Questionnaire: The forty-six item questionnaire (with a seven-point Likert scale anchored by “Strongly Disagree” and “Strongly Agree”) was grouped into eight categories: learning motivation, self-efficacy, usefulness of the PCA, reliability of the PCA, human-likeness of the PCA, likability of the PCA, self-reflection of learning, and concentration on learning.

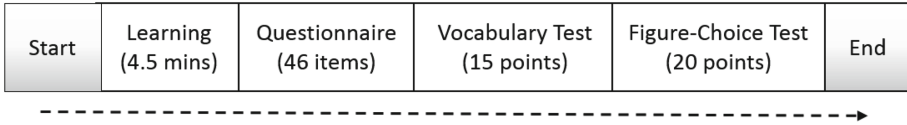


Fig. 3. Procedure of experiment

3.4 Participants and Procedure

The participants were 120 undergraduate students (male 84%, female 16%) in a web application development course held in a large private university located at Kanagawa prefecture in Japan. Students were required to participate in the experiment as a part of course activity, but they were explained that the results will not be treated as a part of their grade.

The three versions of the e-learning systems were deployed on a server, and the URL of the introduction page was announced to the students, explaining they have to prepare the latest version of FireFox and speakers/a headphone, and the URL was going to be expired in a week. The students were randomly assigned to experimental conditions when transferred from the introduction page.

Figure 3 shows the flow of learning and testing. The participants firstly watched 4.5 min long instruction, then took the questionnaire, and the tests. The questionnaire should work as a distraction task for the following tests at the same time.

3.5 Design and Hypothesis

The study employed a one-factor three-level between-participants design. The data were analyzed by one-way analysis of variance (ANOVA) with Tukey's HSD post-hoc comparisons.

We hypothesized that *conduit gestures* draw learners attentions to technical terms and help the learners remember the words, resulting in more correct vocabulary recall test results, and that *two-conduit gestures* and *three-conduit gestures* schematically convey the relationship between concepts by utilizing space, resulting in more correct figure selection test results, when the instruction speech was given with appropriate gestures than when speech was given with mismatched gestures or no gestures. In addition, we also hypothesized that learners judge learning experience and the perception of the PCA higher when the PCA performed metaphoric gestures matched with speech.

4 Results and Discussion

Of 120 students, nine did not access the URL, three did not completed the test, and one reported a web-browser error. And, the data of ten students who took longer than sixty minutes to complete the study were eliminated. Eventually, the data of 97 students were acquired. The numbers of participants assigned to each conditions were 32 for C1, 32 for C2, and 33 for C3.

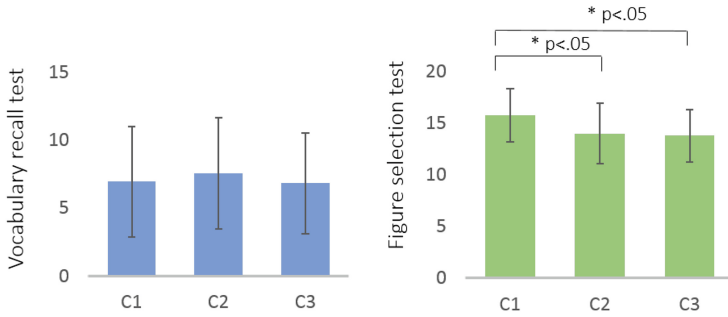


Fig. 4. Results of the vocabulary recall test and the figure selection test

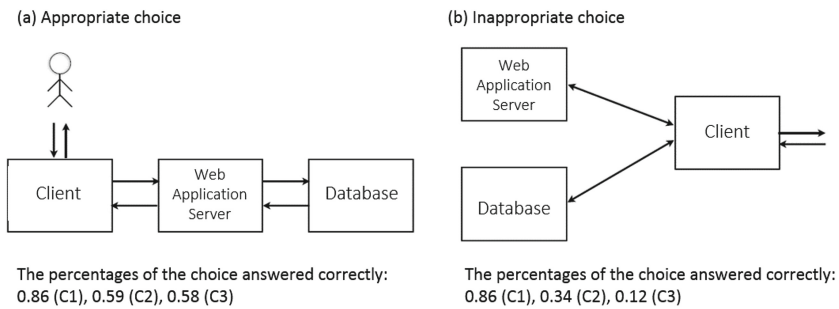


Fig. 5. Examples of appropriate choice and inappropriate choice that the participants in speech-gesture-mismatch group answered wrongly (Q. Select all appropriate figures which describe the interaction between components of web applications).

4.1 Memorization and Understanding

The results of the vocabulary recall test and the figure selection test are shown in Fig. 4. Although there was no statistically significant difference found in the results of recall test, we found a significant main effect of the gesture factor in the figure selection test ($F(2,93) = 3.09$, $p = .007$). The students in speech-gesture match condition scored significantly higher ($M = 15.7$, $SD = 2.59$) than the students in speech-gesture mismatch condition ($M = 13.9$, $SD = 2.93$, $t(62) = 2.61$, $p = .011$), and the students in no-gesture condition ($M = 13.8$, $SD = 2.52$, $t(63) = 2.94$, $p = .028$). The standardized effect sizes for these differences were Cohen's $d = 0.63$ and 0.76 , which indicate between medium and large effects.

Figure 5(a) shows an example of appropriate figures and Fig. 5(b) shows an example of inappropriate figures in a question, "Select all appropriate figures which describe the interaction between components of web applications." The percentages of the choice answered correctly in the speech-gesture mismatch condition decreased more than 20% against the percentage in speech-gesture matched group, which was over 80%. Figure 6 shows the annotated script in both

Annotated Script (Speech-gesture match)	... a user makes a request through [a client program] to [a web application server, then <div style="text-align: center;"><i>CR</i> <i>CC</i></div> the web application server dynamically produce a web page], cooperating with [a database] <div style="text-align: right;"><i>CL</i></div> , then sends it to [a client program.] <div style="text-align: center;"><i>CR</i></div>
Annotated Script (Speech-gesture mismatch)	... a user makes a request through [a client program] to [a web application server, then <div style="text-align: center;"><i>CC</i> <i>CR</i></div> the web application server dynamically produce a web page, cooperating with a database] , then sends it to [a client program.] <div style="text-align: center;"><i>CL</i></div>

Fig. 6. A part of the script annotated differently in speech-gesture match condition and speech-gesture mismatch condition, describing the interaction between three components of web application “client program,” “web application server,” and “database” ([]: stroke and hold, translated from Japanese).

conditions. As seen in Fig. 6, in speech-gesture match condition, conduit gestures present the appropriate layouts of abstract components of web application as described in Fig. 5 (a). In contrast, in speech-gesture mismatched group, conduit gestures were meaninglessly put in space.

These results support our hypothesis stating that spatial use of conduit metaphoric gestures help listeners’ schematic understanding of the relationship between abstract concepts. However, a single conduit gesture which presents an abstract concept did not affect learners memorization of the name of the concept.

4.2 Learning Experience and Perception of the PCA

We did not find many significant differences between conditions from the results of questionnaire, but the results revealed a significant main effect of the gesture factor on perception of reliability of the PCA ($F(2,94) = 3.57, p = .032$).

Interestingly, despite the incoherence between speech content and gesture representation, the multiple test results showed that the students in speech-gesture mismatch condition answered the PCA was more reliable ($M = 4.45, SD = 0.97$) than the students in speech-gesture match condition ($M = 3.74, SD = 1.21, t(62) = 2.59, p = .027$). The standardized effect size for these differences were Cohen’s $d = 0.65$ which indicates between medium and large effects. The results of each item in the category are shown in Table 1.

Previous psychology studies revealed that non-verbal behaviors, included gestural styles, are linked to personality [1, 5]. The argument was also partially confirmed in the studies of interaction between human and animated character [14]. Neff et al. reported that the perception of extroversion increased when PCA’s gesture rate was high, and when movements were produced fast [22]. In speech-gesture mismatch condition, the number of gesture strokes were larger than speech-gesture match condition, despite the length of audio speech was identical. And, strokes performed rapidly were often seen in the realized animation. This indicates that the difference of the speed of gestures caused the

Table 1. Questions asking reliability of the PCA (Cronbach's $\alpha = 0.84$)

Questions	C1 (<i>SD</i>)	C2 (<i>SD</i>)	C3 (<i>SD</i>)
I felt the PCA has a considerable knowledge	4.38 (1.45)	4.53 (1.19)	4.27 (1.46)
I felt the PCA was intellectual	3.88 (1.62)	4.47 (1.41)	4.15 (1.46)
I felt the PCA was useful	3.53 (1.54)	4.63 (1.31)	4.47 (1.27)
I felt the PCA was helpful	3.63 (1.48)	4.47 (1.32)	4.27 (1.13)
I felt the PCA was like a teacher	3.28 (1.63)	4.16 (1.25)	4.03 (1.31)
Overall	3.74 (1.21)	4.45 (0.97)	4.21 (1.06)

difference of the perception of extroversion of the PCA, and that caused the difference of the perception of reliability.

5 Conclusion

We examined the impact of metaphoric gestures performed by Pedagogical Conversational Agent (PCA) on learners' memorization of technical terms, understanding of relationships between abstract concepts, learning experience, and perception of the PCA. The study employed a one-factor three-level between-participants design where we manipulated gesture factor (speech-gesture match vs. speech-gesture mismatch vs. no-gesture). The data of 97 students were acquired in on-line learning environment. As the results, while there was no effect found on memorization of technical terms, we found that students showed accurate schematic understanding of the relationship between abstract concepts when the PCA used metaphoric gestures matched to speech content than when used gestures mismatched, and no gesture. Contrary to the result, we also found that students judged the PCA useful, helpful, and felt the PCA looked like a teacher when performed mismatched gestures to speech content than when performed matched gesture.

Acknowledgments. This work was partially supported by JSPS Grant-in-Aid for Young Scientists (B) Grant Number 25870698, JSPS Grant-in-Aid for Challenging Exploratory Research Grant Number 26540185, and Council for Science, Technology and Innovation(CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), Structural Materials for Innovation (SM4I) (Funding agency:JST).

References

1. Argyle, M.: Bodily communication. Methuen, London (1975)
2. Arzarello, F., Paola, D., Robutti, O., Sabena, C.: Gestures as semiotic resources in the mathematics classroom. *Educ. Stud. Math.* **70**(2), 97–109 (2009)
3. Atkinson, R.K.: Optimizing learning from examples using animated pedagogical agents. *J. Educ. Psychol.* **94**(2), 416–427 (2002)

4. Baylor, A.L., Kim, Y.: Pedagogical agent design: the impact of agent realism, gender, ethnicity, and instructional role. In: Lester, J.C., Vicari, R.M., Paraguaçu, F. (eds.) ITS 2004. LNCS, vol. 3220, pp. 592–603. Springer, Heidelberg (2004)
5. Brebner, J.: Personality theory and movement. In: Kirkcaldy, B.D. (ed.) *Individual Differences in Movement*, pp. 27–41. Springer, The Netherlands (1985)
6. Buisine, S., Abrilian, S., Martin, J.C.: Evaluation of multimodal behaviour of embodied agents. In: Ruttkay, Z., Pelachaud, C. (eds.) *From Brows to Trust*, pp. 217–238. Springer, The Netherlands (2005)
7. Buisine, S., Martin, J.C.: The effects of speech-gesture cooperation in animated agents' behavior in multimedia presentations. *Interact. Comput.* **19**(4), 484–493 (2007)
8. Cassell, J., McNeill, D.: Gesture and the poetics of prose. *Poetics Today* **12**, 375–404 (1991)
9. Cassell, J., McNeill, D., McCullough, K.E.: Speech-gesture mismatches: evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics Cogn.* **7**(1), 1–34 (1999)
10. Cook, S.W., Duffy, R.G., Fenn, K.M.: Consolidation and transfer of learning after observing hand gesture. *Child Dev.* **84**(6), 1863–1871 (2013)
11. Hasegawa, D., Ugurlu, Y., Sakuta, H.: A human-like embodied agent learning tour guide for e-learning systems. In: 2014 IEEE Global Engineering Education Conference (EDUCON), pp. 50–53, April 2014
12. Hayashi, Y.: Togetherness: multiple pedagogical conversational agents as companions in collaborative learning. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 114–123. Springer, Heidelberg (2014)
13. Heidig, S., Clarebout, G.: Do pedagogical agents make a difference to student motivation and learning? *Educ. Res. Rev.* **6**(1), 27–54 (2011)
14. Isbister, K., Nass, C.: Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *Int. J. Hum. Comput. Stud.* **53**(2), 251–267 (2000)
15. Johnson, W.L., Rickel, J.: Steve: an animated pedagogical agent for procedural training in virtual environments. *SIGART Bull.* **8**(1–4), 16–21 (1997)
16. Kellerman, S.: 'I see what you mean': the role of kinesic behaviour in listening and implications for foreign and second language learning. *Appl. Linguist.* **13**(3), 239–258 (1992)
17. Kelly, S.D., Barr, D.J., Church, R.B., Lynch, K.: Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. *J. Mem. Lang.* **40**(4), 577–592 (1999)
18. Lester, J.C., Voerman, J.L., Towns, S.G., Callaway, C.B.: Cosmo: A life-like animated pedagogical agent with deictic believability. In: *Working Notes of the IJCAI 1997 Workshop on Animated Interface Agents: Making Them Intelligent*, pp. 61–69. Citeseer, August 1997
19. McNeil, D.: *Hand and Mind*. University of Chicago Press, Chicago (1992)
20. McNeill, D.: Gesture: a psycholinguistic approach. In: Brown, E., Anderson, A. (eds.) *The Encyclopedia of Language and Linguistics*, pp. 58–66. Elsevier, Amsterdam (2006)
21. Moreno, R., Mayer, R.E., Spires, H.A., Lester, J.C.: The case for social agency in computer-based teaching: do students learn more deeply when they interact with animated pedagogical agents? *Cogn. Instr.* **19**(2), 177–213 (2001)

22. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In: Allbeck, J., Badler, N., Bickermore, T., Pelachcaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 222–235. Springer, Heidelberg (2010)
23. Ogan, A., Finkelstein, S., Mayfield, E., D’Adamo, C., Matsuda, N., Cassell, J.: ”Oh dear stacy!”: social interaction, elaboration, and learning with teachable agents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2012, pp. 39–48. ACM, New York, NY, USA (2012)
24. Popescu, V., Adamo-Villani, N., Wu, M.L., Rajasekaran, S.D., Alibali, M.W., Nathan, M., Cook, S.W.: Animation killed the video star. In: Proceedings of Gesture-based Interaction Design: Communication and Cognition, 2014 CHI Workshop, pp. 55–59 (2014)
25. Reynolds, F.J., Reeve, R.A.: Gesture in collaborative mathematics problem-solving. *J. Math. Behav.* **20**(4), 447–460 (2001)
26. Singer, M.A., Goldin-Meadow, S.: Children learn when their teacher’s gestures and speech differ. *Psychol. Sci.* **16**(2), 85–89 (2005)
27. Valenzeno, L., Alibali, M.W., Klatzky, R.: Teachers’ gestures facilitate students’ learning: a lesson in symmetry. *Contemp. Educ. Psychol.* **28**(2), 187–204 (2003)
28. Voerman, J.L., FitzGerald, P.J.: Deictic and emotive communication in animated pedagogical agents. In: Cassell, J., Sullican, J., Prevost, S., Churchill, E. (eds.) *Embodied Conversational Agents*, pp. 123–154. The MIT Press, Cambridge (2000)