British Journal of Educational Technology *Vol 44 No 1 2013* E5–E9
doi:10.1111/j.1467-8535.2012.01311.x

Colloquium

Embodying gesture-based multimedia to improve learning

Chun-Yen Chang, Yu-Ta Chien, Cheng-Yu Chiang, Ming-Chao Lin and Hsin-Chih Lai

Address for correspondence: Prof Chun-Yen Chang, Professor and Director, Science Education Center and Graduate Institute of Science Education and Department of Earth Sciences, National Taiwan Normal University, No. 88, Section 4, Ting-Chou Road, Taipei 11677, Taiwan. Email: changcy@ntnu.edu.tw

Introduction

As revealed by the recent Horizon Report (Johnson, Smith, Willis, Levine & Haywood, 2011), the creation of gesture-based interfaces (eg, Microsoft Kinect, Nintendo Wii and Apple iPhone/iPad) create promising opportunities for educators to offer students easier and more intuitive ways to interact with the content in multimedia learning environments than ever before. For instance, Kinect, a motion-sensing input device developed by Microsoft for Xbox360 (Redmond, WA, USA), enables students to use their body motions, such as swiping, jumping and moving, to interact with the content on the screen. Growing communities (eg, KinectEDucation; http://www.kinecteducation.com/) advocate integrating Kinect applications into classroom teaching to make students' learning experiences more active and joyful. Few studies investigate the impact of gesture-based multimedia learning on students' cognitive learning outcomes and/or its theoretical underpinnings. This paper briefly discusses the theoretical underpinnings for adopting gesture-based multimedia learning, then describes how we used Kinect to embody the most common type of multimedia learning in classroom (ie, PowerPoint presentations) and finally details a preliminary study exploring the impact of gesture-based multimedia learning on students' cognitive learning outcomes.

Leveraging on human movements to facilitate multimedia learning

The role human movements play in terms of learning is gaining increasing attention in the field of cognitive science. The traditional assumption that knowledge representations in cognition are composed of amodal symbols and reside in a modular semantic system is being challenged. Currently, grounded cognition rejects the traditional view that knowledge representations in cognition are independent of the modal systems for perception and action; instead, it proposes that knowledge representations in cognition are grounded in multiple ways, including simulations, situated action and bodily states (Barsalou, 2010). Accumulating behavioral and neural studies also suggest that bodily states can cause cognitive states as well as be effects of these states (see reviews of Barsalou, 2008; de Koning & Tabbers, 2011). The implication of grounded cognition for education is that using human movements to embody instructional materials, either demonstrations by instructors or executions by learners themselves, may enrich learners' knowledge representations and assist them in acquiring abstract concepts.

A recent meta-analysis conducted by Höffler and Leutner (2007) clearly showed that the instructional efficiency of the multimedia presentations which involved human movements is generally higher than those which did not involve human movements. Several new studies also indicated that multimedia materials would be effective to assist learners in acquiring knowledge especially if they contained human movements, such as folding papers (Wong *et al*, 2009), tying knots

(Ayres, Marcus, Chan & Qian, 2009; Schwan & Riempp, 2004), solving earring and quadring puzzles (Ayres *et al*, 2009), and manipulating topographical measures (Chien & Chang, 2012). The success of this type of instructional multimedia materials may be ascribed to the function of the mirror neuron system in the human brain; while observing/listening someone else performing/explaining a human movement, the neurons in the observer/listener's brain responsible for executing that movement would be activated (Rizzolatti & Craighero, 2004; Tettamanti *et al*, 2005). This biological advantage could ease the consumption of cognitive resources while a learner is processing the instructional multimedia materials which contain human movements (Ayres *et al*, 2009; van Gog, Paas, Marcus, Ayres & Sweller, 2009; Wong *et al*, 2009), and thus enhance learning outcomes.

Although the mechanism of how the activation of mirror neuron system contributes to human cognition processes has yet to be well understood, it seems that using human movements, such as gestures and body metaphors (de Koning & Tabbers, 2011), to embody the content presented in instructional multimedia materials could enrich learners' knowledge representations without extra consumption of their cognitive resources. This means using human movements thus makes learning from multimedia materials more effective. Therefore, we put forward the design of eight gestures to metaphorically embody the taxonomy of Gardner's theory of multiple intelligences (Gardner, 1993) for conducting a multimedia presentation; while introducing Gardner's theory of multiple intelligences to students, once the teacher performs one of the predefined gestures (eg, playing a violin), the Kinect device will recognize it and automatically switch the PowerPoint to the corresponding slide (eg, musical intelligence). In the next section, we briefly share a convenient and economic approach to fulfill the proposed gesture-based multimedia presentation.

Making a gesture-based multimedia presentation available with Kinect

As shown in Figure 1, we designed eight gestures to metaphorically embody the taxonomy of Gardner's theory of multiple intelligences. The predefined gestures were set as the *wireless hotkeys* to be recognized by Kinect for further controlling a PowerPoint presentation. Kinect was connected with a PC by a USB cable, and the PrimeSense Package, including OpenNI, NITE, and PrimeSensor Module (detailed information may be found on an open source website for developing natural interaction applications, http://www.openni.org/) was installed in the PC to make Kinect compatible with Windows x86 environments. The Flexible Action and Articulated Skeleton Toolkit (FAAST), which is developed by a research team from the University of Southern California (Suma, Lange, Rizzo, Krum & Bolas, 2012), was applied to track the user's gestures recognized by Kinect and then to emulate keyboard input to call corresponding slides of the PowerPoint presentation. We wrote eight commands for FAAST to emulate keyboard input. For instance, the following command will press the "7" key (to switch the PowerPoint to slide 7, musical intelligence) when the teachers' left hand extends more than 9 inches sideways distance from the shoulder (the gesture of playing a violin). Please refer the FAAST website for more detailed information about the input emulator usage.

Impact of gesture-based presentation on students' cognitive learning outcomes

A senior professor of science education implemented the aforementioned gesture-based multimedia presentation to teach Gardner's theory of multiple intelligences in a university classroom setting. Sixteen college students received the 20-minute presentation together. At the beginning of the presentation, the professor asked all students to brainstorm "what abilities can be counted as intelligence?" Once an idea about intelligence was proposed by students, the professor would use his bodily motions to metaphorically embody that idea as much as possible, whether that idea matched Gardner's theory or not. If students' ideas matched one of the taxonomy of Gardner's theory, the professor would perform the predefined gesture and then the gesture-based multimedia presentation would automatically switch to the corresponding slide to lecture. As a pretest, all

Taxonomy/ Body metaphor	Gesture	Taxonomy/ Body metaphor	Gesture
Kinesthetic/ Dribble		Musical/ Playing a violin	
Interpersonal/ Greeting		Intrapersonal/ Namaskara mudra (Buddhist gesture)	
Naturalistic/ Surveying		Linguistic/ Yelling out	
Logical-mathematical/ Calculating		Spatial/ Column	

Figure 1: Gestures to metaphorically embody the taxonomy of multiple intelligences

students were asked to write down the taxonomy of multiple intelligences before the intervention. The same test was administered to the students immediately after the intervention and 4 weeks later than the intervention as post- and delay-tests, respectively. The rubric to score the pre-, post- and delay-test was the same; if the student answered a correct category of the taxonomy, he or she would get one score (the maximum score was eight).

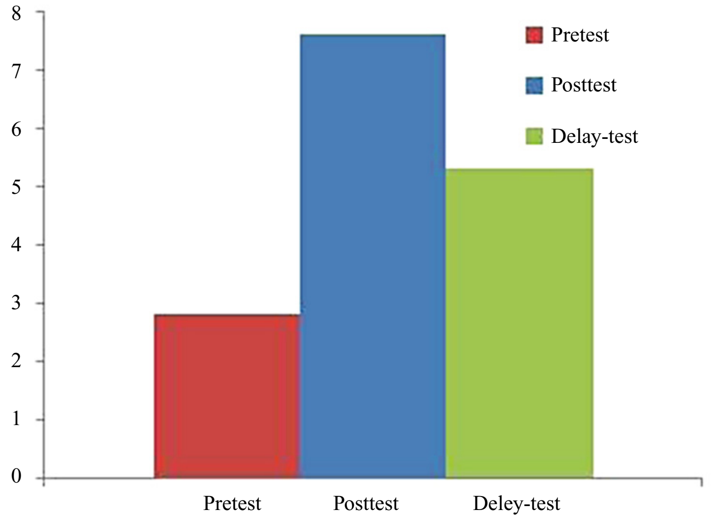

Figure 2: Students' mean scores on pre-, post- and delay-tests

As shown in Figure 2, the mean level of students' prior domain knowledge, in terms of the taxonomy of multiple intelligences, was low ($M = 2.8$, $SD = 1.2$). Immediately after receiving the intervention, almost all students could recall all categories of multiple intelligences taught in the class ($M = 7.6$, $SD = 1.0$). Four weeks later than the intervention, students in average could recall five categories of multiple intelligences on the delay-test ($M = 5.3$, $SD = 1.3$). Paired t-test analysis on students' pre-, post-, delay-test scores indicates that the positive impacts of the gesture-based multimedia presentation on students' retention performances, both immediately after the intervention and 4 weeks later than the intervention, had practical significance (pretest vs. posttest: $p < 0.001$, $d = 3.0$, a very large effect size; pretest vs. delay-test: $p < 0.001$, $d = 1.1$, a large effect size). The preliminary results suggest that the embodied approach proposed in this paper is very promising to facilitate students' cognitive learning outcomes in multimedia learning environments. Future studies are needed to clarify the comparative instructional effectiveness of embodied and nonembodied multimedia presentations. A larger sample size is also needed to further verify the advantage or disadvantage of embodied multimedia presentations in different perspectives. Moreover, research to explore what type of materials is suitable to be embodied in multimedia learning environments is recommended.

Acknowledgements

The authors wish to thank the Aim for the Top University (ATU) project of the National Taiwan Normal University (NTNU) for financing this study.

References

Ayres, P., Marcus, N., Chan, C. & Qian, N. X. (2009). Learning hand manipulative tasks: when instructional animations are superior to equivalent static representations. *Computers in Human Behavior*, *25*, 2, 348–353.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 1, 617–645.

Barsalou, L. W. (2010). Grounded cognition: past, present, and future. *Topics in Cognitive Science*, *2*, 4, 716–724.

Chien, Y. T. & Chang, C. Y. (2012). Comparison of different instructional multimedia designs for improving student science-process skill learning. *Journal of Science Education and Technology*, *21*, 1, 106–113.

de Koning, B. B. & Tabbers, H. K. (2011). Facilitating understanding of movements in dynamic visualizations: an embodied perspective. *Educational Psychology Review*, *23*, 4, 501–521.

Gardner, H. (1993). *Multiple intelligences: the theory in practice*. New York: Basic Books.
Höffler, T. N. & Leutner, D. (2007). Instructional animation versus static pictures: a meta-analysis. *Learning and Instruction*, *17*, 6, 722–738.
Johnson, L., Smith, R., Willis, H., Levine, A. & Haywood, K. (2011). *The 2011 horizon report*. Austin, TX: The New Media Consortium.
Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.
Schwan, S. & Riempp, R. (2004). The cognitive benefits of interactive videos: learning to tie nautical knots. *Learning and Instruction*, *14*, 293–305.
Suma, E. A., Lange, B., Rizzo, S., Krum, D. & Bolas, M. (2012). The Flexible Action and Articulated Skeleton Toolkit (FAAST). Retrieved January 16, 2012, from http://projects.ict.usc.edu/mxr/faast/
Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P. *et al* (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–281.
van Gog, T., Paas, F., Marcus, N., Ayres, P. & Sweller, J. (2009). The mirror neuron system and observational learning: implications for the effectiveness of dynamic visualizations. *Educational Psychology Review*, *21*, 1, 21–30.
Wong, A., Marcus, N., Ayres, P., Smith, L., Cooper, G. A., Paas, F. *et al* (2009). Instructional animations can be superior to statics when learning human motor skills. *Computers in Human Behavior*, *25*, 2, 339–347.