Metadata of the chapter that will be visualized online

Chapter Title	Multimodal Learning	
Copyright Year	2011	
Copyright Holder	Springer Science+Business Media, LLC	
Corresponding Author	Family Name	**Massaro**
	Particle	
	Given Name	**Dominic W.**
	Suffix	
	Division/Department	Perceptual Science Laboratory
	Organization/University	University of California- Santa Cruz Social Sciences II
	Postcode	95064
	City	Santa Cruz
	State	CA
	Country	USA
	Phone	831-459-2330
	Email	massaro@ucsc.edu

M

Multimodal Learning

Au1 DOMINIC W. MASSARO
Perceptual Science Laboratory, University of California-Santa Cruz Social Sciences II, Santa Cruz, CA, USA

Synonyms

Bimodal learning; Multisensory learning

Definition

Multimodal learning refers to an embodied learning situation which engages multiple sensory systems and action systems of the learner. This type of learning is traditionally emphasized for children with learning challenges, and can include a variety of visual inputs in addition to text. Some examples include pictures, art, film, video, and graphic organizers. Auditory inputs can include text-to-speech synthesizers, various forms of singing and musical instruments, rhyming, and spoken language games. One salient example is the use of the alphabet song to learn the alphabet. Tactile inputs are often manipulatives such as the use of an abacus for math learning, sculpting materials such as clay, paint, and paper for representing objects and ideas, and puzzles for fact learning such as learning the states and their capitals. Finally, kinesthetic engagement includes all forms of motor behavior and gesture such as jumping rope to memorize songs and hop scotch to practice school lessons. A recent trend is the change from fairly passive computer games such as Sudoku, Tetris, and Solitaire to much more active types of game activity such as the sports and fitness games for the Wii Nitendo (2010). Another trend with great promise is the creative integration of the physical engagement of traditional hands-on board games with miniaturization technology and methodology from wireless sensor networks, as in siftables (Sifteo 2010).

An interactive multimedia environment is ideally suited for multimodal learning. For example, incorporating text and visual images of the vocabulary to be learned along with the actual definitions and sound of the vocabulary facilitates learning and improves memory for the target vocabulary and grammar. At the same time, the learner is actively engaged by listening to the words, pronouncing the words, and if literate, reading and writing the words. In one typical application (Massaro 2006; Animated Speech Corporation 2010), a computer-animated agent guides the students to (1) observe the words being spoken by a realistic talking interlocutor, (2) experience the word as spoken as well as written, (3) see visual images of referents of the words, (4) click on or point to the referent or its spelling, (5) hear themselves say the word, followed by a correct pronunciation, (6) spell the word by typing, and (7) observe and respond to the word used in context. Although half of the exercises involve multiple choice testing, there is evidence that this experience boosts performance on later tests. The other half of the tests involve either spoken or written generation of the students' answers, which facilitates learning (Metcalfe and Kornell 2007). The test exercises can be viewed as learning exercises because testing has been demonstrated to increase learning and retention.

In a recent experimental test, children, whose native language was Spanish, were tutored and tested on English words they did not know. The research utilized a multiple baseline design to insure that any learning was due to the application itself rather than from outside of the lesson environment. The children learned the words when they were tutored but not words that were simply tested. This result replicates the previous studies carried out on hard of hearing and autistic children with Baldi as the animated conversational tutor. In other experiments, we have also observed that Baldi's unique characteristics allow a novel approach to training speech production to both children with hearing loss (Massaro 2004) and adults learning a new language.

Theoretical Background

Perhaps the most germane background for Multimodal Learning is Montessori's Principles of Educational Practice (Stoll-Lillard 2005). Montessori's Principle 1 claims that motor behavior and cognition are closely intertwined and that physical movement can enhance thinking and learning. At first glance, this principle seems the antithesis of direct computer-aided instruction with an animated

Norbert Seel (ed.), *Encyclopedia of the Sciences of Learning*, DOI 10.1007/978-1-4419-1428-6,

tutor. However, we have learned that our nervous systems appear to be wired in a way that observations of actions activate neural mechanisms involved with the actual performance of those actions. The so-called mirror neurons involved in performing an action are activated when that action is observed. One possibility, therefore, would be to implement lessons on Nintendo's Wii to allow the child to have larger physical movements. Another would be to have animated movies as well as pictures for learning.

Montessori's Principle 2 states that choice and perceived control promote children's concentration and contentment in the learning process. As is currently exists, direct instruction does not appear to allow much choice. On the other hand, the child can be given a library of lessons and she can choose the lesson to study. A precocious child might even be able to create a lesson of her choosing.

Principle 3 assumes that personal interest enhances learning in a context where interests build on prior knowledge and the children's own questions. For example, a deaf French child used the Lesson Creator to document her travel and holiday pictures in a set of English vocabulary lessons. Thus, learning a new language was facilitated by involving her direct experience and interests with a normally tedious task.

Principle 4 indicates that extrinsic rewards negatively impact long-term motivation and learning. Rewards and feedback can be controlled exactly in computer-assisted learning. Directed feedback can allow errorless learning without focusing on rewarding the child for correct answers and punishing the child for incorrect answers.

According to Principle 5, collaborative (child–child) arrangements are conducive to learning. Although most automated instruction is one-on-one and precludes collaborative learning, this principle can be instantiated in several different ways. First, the animated agent can be a child who works along with the child. Second, children can work together on a lesson or on creating lessons, and can even distribute the required learning and thereby achieve the benefits of the Jigsaw Classroom.

Principle 6 assumes that learning situated in and connected to meaningful contexts is more effective than learning in abstracted contexts. Although most automated instruction can be considered relatively unsituated and not connected to a meaningful context, the Lesson Creator allows the immediate creation of lessons on subjects that are currently taught: Just-in-time learning. Thus, the child sees the value and appropriate context of the lesson when it is connected to her appropriate interest and cognitive level.

Principle 7 claims that sensitive and responsive (nurturing) teaching is associated with more optimal outcomes. Tutors can be created and programmed to be highly nurturing. For example, the difficulty of the lessons can be controlled to meet the child's preferred difficulty level, and errorless feedback can be provided.

Principle 8 assumes that order in the environment promotes and establishes mental order and is beneficial to the child. Direct instruction is highly orderly in its functioning, which adheres to this principle.

Another relevant background source is the empirical and theoretical literature on multimedia learning (Mayer 2005). This research, for example, gives principles for the ideal placement of illustrations in science texts. It is a challenge to have both illustrations and written text appropriately placed. Usually this requires that the text is placed near the referent. Gestalt principles of organization could be used to insure that the text and the appropriate aspect of the illustration are perceived as near one another. Spoken language during the lesson is not easily localized because of our perceptual limits in perceiving small differences in the localization of sound. In this case, the appropriate part of the illustration can be highlighted while it is being discussed. More generally, it is important to make it easy for the learner to hold pictorial and verbal representations in working memory at the same time. Finally, when illustrating a sequence of events, successive or causal links in the sequence should be presented near one another.

A theory that serves important background for Multimodal Learning is the Fuzzy Logical Model of Perception (FLMP) According to this model, multiple sensory influences are combined before categorization and perceptual experience. In face-to-face speech perception, for example, the FLMP assumes that the visible and audible speech signals are integrated. Before integration, however, each source is evaluated (independently of the other source) to determine how much that source supports various alternatives. The integration process combines these support values to determine how much their combination supports the various alternatives. The perceptual outcome for the perceiver will be a function of the relative degree of support among the competing alternatives. Across a range of studies comparing specific mathematical predictions, the FLMP has been more successful than other competitor models in accounting for the experimental data (Massaro 1998).

The FLMP has proven to be a universal principle of pattern recognition. In multisensory texture perception, for example, there appears to be no fixed sensory dominance by vision or haptics, and the bimodal presentation

Comp. by: MANIKANDAN R Stage: Galleys Chapter No.: 273 Title Name: ESL
Page Number: 0 Date:28/5/11 Time:10:52:21

yields higher accuracy than either of the unimodal conditions. Preschool as well as school children integrate auditory and visual speech to produce a multimodal benefit of having two sources of information relative to just one. In addition, both hard of hearing children and autistic children appear to integrate information from the face and the voice. These results from typically developing children as well as deaf and hard of hearing and autistic children indicate that multisensory environments should be ideal for speech and language learning.

Important Scientific Research and Open Questions

There are, of course, many remaining research and theoretical questions to be addressed in future research. For example, one might question why perceivers integrate several sources of information when just one of them might be sufficient. Most of us do reasonably well in communicating over the telephone, for example. Part of the answer might be grounded in our ontogeny. Integration might be so natural for adults even when information from just one sense would be sufficient because, during development, there was much less information from each sense and therefore integration was all the more critical for accurate performance.

A natural question concerns the neural mechanism underlying the integration algorithm specified in the FLMP. An important set of observations from single cell recordings in the cat's brain could be interpreted in terms of integration of the form specified by the FLMP. A single hissing sound or a light spot can activate neurons in the superior colliculus. A much more vigorous response is produced, however, when both signals are simultaneously presented from the same location. The FLMP is mathematically equivalent to Bayes' theorem, which is an optimal method for combining two sources of evidence to test among hypotheses. The brain can implement an analogous computation so that the response of a neuron is proportional to the posterior probability that a target is present in its receptive fields, given its sensory input. Therefore, the target-present posterior probability computed from the impulses from the auditory and visual neurons is higher given sensory inputs of two modalities than it is given input of only one modality, analogous to the synergistic outcome of the FLMP. This type of research informs questions about the neural underpinings of Multimodal Learning.

Multimodal Learning situations are often implemented in virtual rather than real worlds. It is feasible that limiting the students' experience to the two-dimensional world of computer monitors would constrain learning relative to a live teacher. The success of two-dimensional media such as the television and the Internet, however, is a real-world experimental proof of the sufficiency of two dimensions for learning. To date, tutoring on two-dimensional surfaces appears to be as effective as live tutoring, although additional research is still required on this question. However, with the exploding popularity of three-dimensional (3D) movies such as *Up* and *Avatar*, and the increasing availability of 3D projection systems, TVs, and computer monitors, learners will more often find themselves in more realistic simulated 3D worlds.

Cross-References

References

Animated Speech Corporation. (2010). http://www.animatedspeech.com/. Accessed 19 Jan 2010.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Massaro, D. W. (2004). Symbiotic value of an embodied agent in language learning. In R.H. Sprague, Jr. (Ed.), *Proceedings of 37th annual hawaii international conference on system sciences, (HICCS'04), (CD-ROM,10 pages)*. Best paper in Emerging Technologies. IEEE Computer Society Press, Los Alimitos

Massaro, D. W. (2006). A computer-animated tutor for language learning: Research and applications. In P. E. Spencer & M. Marshark (Eds.), *Advances in the spoken language development of deaf and hard-of-hearing children* (pp. 212–243). New York: Oxford University Press.

Mayer, R. E. (Ed.). (2005). *The cambridge handbook of multimedia learning*. New York: Cambridge University Press.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review, 14*, 225–229.

Nitendo (2010). http://www.nintendo.com/wii. Accessed 19 Jan 2010.

Sifteo (2010). http://sifteo.com/. Accessed 19 Jan 2010.

Stoll-Lillard, A. (2005). *Montessori: The science behind the genius*. Oxford: Oxford University Press.

Author Query Form

Encyclopedia of the Sciences of Learning
Chapter No: 273

Query Refs.	Details Required	Author's response
AU1	Please check if author affiliation is okay.	
AU2	Please provide content under the section "Cross-References", if applicable.	
AU3	Please provide article title/report title for the references "Animated Speech Corporation (2010)", "Nitendo (2010)", "Sifteo (2010)", if feasible.	