Autism

http://aut.sagepub.com

Read my lips: The importance of the face in a computer-animated tutor for vocabulary learning by children with autism

Dominic W. Massaro and Alexis Bosseler

Autism 2006; 10; 495
DOI: 10.1177/1362361306066599

The online version of this article can be found at:
http://aut.sagepub.com/cgi/content/abstract/10/5/495

Published by:
SAGE Publications
http://www.sagepublications.com

On behalf of:

The National Autistic Society

autism © 2006
SAGE Publications
and The National
Autistic Society
Vol 10(5) 495–510; 066599
1362-3613(200609)10:5

Read my lips

The importance of the face in a computer-animated tutor for vocabulary learning by children with autism

DOMINIC W. MASSARO *University of California, USA*

ALEXIS BOSSELER *University of California, USA*

ABSTRACT A computer-animated tutor, Baldi, has been successful in teaching vocabulary and grammar to children with autism and those with hearing problems. The present study assessed to what extent the face facilitated this learning process relative to the voice alone. Baldi was implemented in a Language Wizard/Tutor, which allows easy creation and presentation of a vocabulary lesson involving the association of pictures and spoken words. The lesson plan included both the receptive identification of pictures and the production of spoken words. A within-subject design with five children with autism followed an alternating treatment in which each child continuously learned to criterion sets of words with and without the face. The rate of learning was significantly faster and the retention was better with the face. The research indicates that at least some children with autism benefit from the face in learning new language within an automated program.

KEYWORDS *autism; computer animation; language learning; vocabulary tutor*

ADDRESS Correspondence should be addressed to: DR DOMINIC W. MASSARO, Department of Psychology, University of California, Santa Cruz, CA 95064, USA. email: massaro@fuzzy.ucsc.edu

Our studies of children with autism are motivated by research findings in behavioral science and the development of innovative technology. Research has shown that the human face affords visual information during speech production that improves effective communication (Massaro, 1998). While the voice alone is usually adequate (and can be used effectively in many remote interactions), observing the movements of the lips, tongue, face, and jaw enhances intelligibility of a degraded auditory message. Speech communication is further enriched by the speaker's facial expressions, emotions, and gestures (Massaro, 1998, Chapters 6, 7, 8). One goal of our research agenda has been to create embodied conversational agents that produce accurate auditory and visible speech, as well as realistic facial expressions, emotions, and gestures. Figure 1 shows a picture of Baldi on

www.sagepublications.com
DOI: 10.1177/1362361306066599

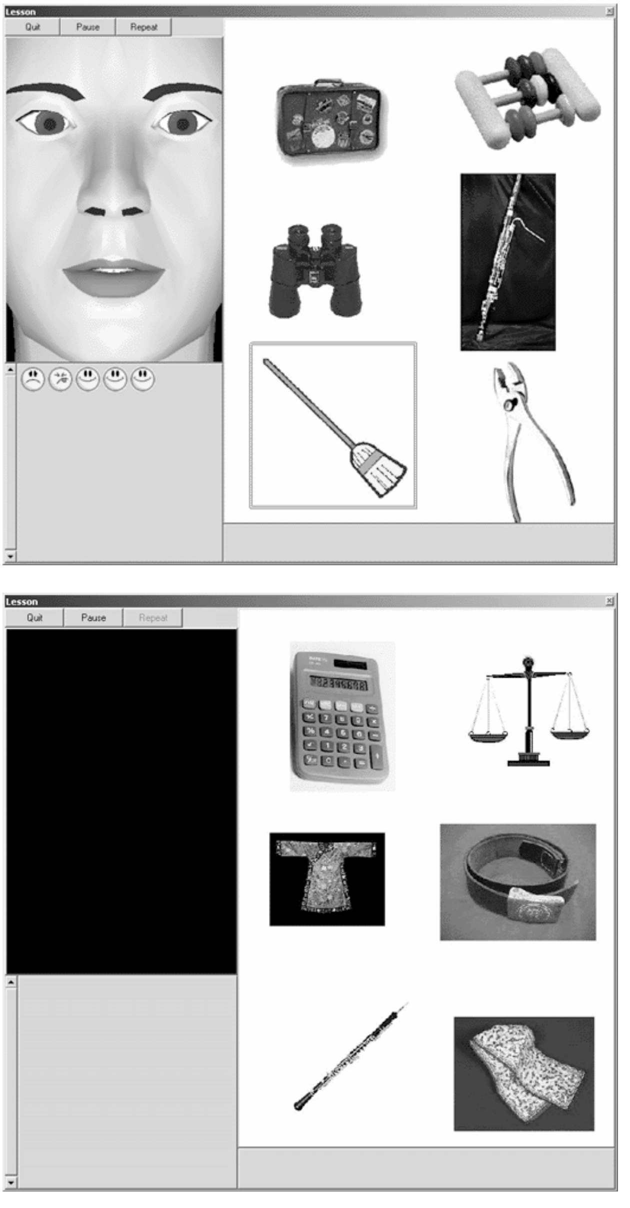

Figure 1 **A prototypical computer screen from the Language Wizard/Tutor. The top panel illustrates the screen of the lessons with the face and the bottom panel illustrates the screen without the face. In the recognition exercise, for example, the student is prompted to identify each item to demonstrate word recognition. In the top panel, for example, Baldi says 'You sweep with this.' The student clicks on the appropriate region and feedback stickers (the happy and sad faces) are given for each response. The highlighted region indicates the student's selection**

a prototypical computer screen from a tutoring lesson; Baldi has been successful at teaching children new vocabulary and language.[1]

The limited ability to produce and comprehend spoken language is the most common factor leading to a diagnosis of autism (American Psychiatric Association, 1994). These language and communicative deficits extend across all components of language (Tager-Flusberg, 1999). Computer-based instruction can be ideal for individuals with special needs (Barker, 2003; Heimann et al., 1995; Moore and Calvert, 2000). Automated practice, feedback and branching can be programmed as well as the presentation of multiple modalities, such as text, sound, and images (Chun and Plass, 1996; Dubois and Vial, 2000). 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. Computer-based instruction also makes it possible to include embodied conversational agents rather than simply disembodied voices in lessons. We have found that Baldi has contributed to vocabulary learning for both children with hearing difficulties (Massaro and Light, 2004) and children with autism (Bosseler and Massaro, 2003).

Baldi is featured in a language-tutorial application to train and develop vocabulary, language and listening skills. This Language Wizard/Tutor allows easy creation and presentation of a language lesson involving the association of pictures and spoken words. The lesson plan includes both the identification of pictures and the production of spoken words. The Language Wizard/Tutor has been used to train vocabulary and grammar to children with autism (Bosseler and Massaro, 2003). This study consisted of two phases. Phase 1 measured vocabulary acquisition and retention. Phase 2 tested whether vocabulary acquisition was due to the Language Wizard/Tutor or outside sources and whether the acquired words could be generalized to new images and outside of the Language Wizard/Tutor environment. Vocabulary lessons were constructed, consisting of over 550 vocabulary items selected from the curriculum of two schools. The participants were eight children diagnosed with autism, ranging in age from 7 to 11 years. All of the students exhibit delays in all areas of academic functioning, particularly in language and adaptive functioning. The results indicated that these children learned and retained many new words, grammatical constructions and concepts. Although all of the children demonstrated learning, it is possible that the children were learning the words elsewhere (for example, from speech therapists or in their school curriculum). To determine that the program was actually responsible for the learning, the children were tested and trained in phase 2 using a within-student multiple baseline design (Baer et al., 1968) in which two categories of words were continuously being tested while a third category was being tested and trained. The students identified significantly more words following implementation of training compared to pre-training performance, showing that the program was responsible for learning. Learning also generalized to new images in random locations, and to interactions outside of the Language Wizard/Tutor. This learning and retention of new vocabulary, grammar, and language use is a significant accomplishment for autistic children.

The embodied conversation agent Baldi provides both facial and vocal support for learning. The purpose of the current investigation is to evaluate whether Baldi facilitates learning relative to presenting the audible speech alone. Children with autism would appear to benefit least from the presence of the face because they tend to be impaired in both their face processing (Dawson et al., 2002; Rogers, 1999; Williams et al., 2001) and their ability to integrate information across modalities (e.g. Bryson, 1970; de Gelder et al., 1996; Lelord et al., 1973; Martineau et al., 1987; Waterhouse et al., 1996). Children with autism tend to avoid the face-to-face contact with others required by shared attention (Happé, 1996) and therefore would naturally have less experience with visual information from the face.

Our experimental test was carried out using two within-subject experimental conditions: training with Baldi and the voice, and training with the voice only. We evaluated whether the face would increase the rate of learning for receptive measures. To accomplish our goal we compared the two conditions according to an alternating treatment design (Barlow and Hayes, 1979) in which each student received each of the two learning conditions concurrently, the order of presentation of the two conditions counterbalanced across days. This alternating treatment design eliminates inter-subject variability and permits a direct observation of the two treatment conditions (Barlow and Hayes, 1979). If children with autism do not extract meaningful information from the face, then we would expect to see no difference in learning between the two conditions.

Method

Participants

Five children diagnosed with autism, one female and four males, ranging in age from 8 to 13 years, were recruited from a school program for children with autism in Santa Cruz County. Prior to the start of our investigation, we received parent permission. These students manifest a wide range of symptoms and abilities. Table 1 gives a brief description of each student's academic and adaptive functioning abilities as well as the diagnostic information provided by the school, parent, and teacher reports. All of the students exhibit delays in all areas, particularly in the areas of

Table 1 **Diagnostic information for the five children: additional diagnoses (primary diagnosis is autism), chronological age (CA), non-verbal IQ or cognitive functioning (CF), level of adaptive functioning**

Student	*Additional diagnoses*	*CA*	*IQ or CF**	*Adaptive* functioning*
1	Mentally retarded	11:6	n.a.	n.a.
2	–	10:11	n.a.	n.a.
3	–	13:1	57[a]	32[b]
4	–	9:7	38[b]	52[b]
5	Moderate mental retardation	8:2	n.a.	n.a.

[a] Wechsler Intelligence Scale for Children–Third Edition (Wechsler, 1991).
[b] Vineland Adaptive Behavior Scales (Sparrow et al., 1984).
* Not assessed.

language and adaptive functioning. Four of the students (1–4) participated in several earlier studies (Bosseler and Massaro, 2003) and in a study in which they practiced lipreading the letters B, D, V, and Z (Massaro and Bosseler, 2003). All five children were capable of speech.

Vocabulary items

Each student was given a unique set of items appropriate to his or her vocabulary knowledge and abilities. Twenty-four different vocabulary items were selected for each child, which were separated into four sets of six items each for each of the five students. There were two lessons per condition and each lesson had six unique vocabulary items. The lessons were matched in pairs for word length measured by the number of syllables, syntactic class, and word frequency (Kucera and Francis, 1967), so that each of the two word sets contained vocabulary items of equal difficulty (full details from the authors). Each lesson had six vocabulary items and was randomly assigned to one of two conditions: with Baldi or with just the auditory speech. The instructors and speech therapists agreed that they would not teach or use these words during our investigation.

Lessons

A Vocabulary Wizard was used to create the vocabulary lessons (Bosseler and Massaro, 2003; Massaro, 2004). This software incorporates an animated talking head, text-to-speech synthesis (TTS), and images of the vocabulary items. The visual images were imported by the experimenter, who determined which parts of the visual image were associated with the spoken words or phrases.

Each lesson consisted of five exercises in the same order: pre-test, presentation, recognition, elicitation, and post-test (see Table 2). The pre-test was

Table 2 **Training schedule following an alternating treatment design counterbalanced across the five students for the four lessons; two with the face (F1 and F2) and two without the face (N1 and N2)**

Session	*Student*				
	1	*2*	*3*	*4*	*5*
1	F1 N1 F2 N2	N1 F1 N2 F2	F1 F2 N1 N2	N1 N2 F1 F2	F1 N1 F2 N2
2	N1 F2 N2 F1	F1 N2 F2 N1	F2 N1 N2 F1	N2 F1 F2 N1	N1 F2 N2 F1
3	F2 N2 F1 N1	N2 F2 N1 F1	N1 N2 F1 F2	F1 F2 N1 N2	F2 N2 F1 N1
4	N2 F1 N1 F2	F2 N1 F1 N2	N2 F1 F2 N1	F2 N1 N2 F1	N2 F1 N1 F2

used to assess the student's performance. The student was prompted to identify each vocabulary item by clicking on the appropriate image and the order of the item presentations was randomized. No feedback was given for correct or incorrect responses. In the presentation exercise, each item was named, the outline of the respective region was highlighted in orange, and the student was asked to click on the appropriate part of the image. No feedback was given for responses, but the child had to click on the highlighted region for the exercise to continue. In the recognition exercise, the student was verbally prompted to identify each item by clicking on the corresponding part of the image. Praise was given for correct responses (e.g. 'Good job') and a happy or sad face was shown for correct and incorrect responses, respectively (see Figure 1). Each item was presented just once unless it was misidentified. The students were given two additional attempts to answer each item correctly. After two incorrect identifications, the correct answer was highlighted and the student was prompted to click on it. In the elicitation exercise, the image was highlighted and the student was asked to name it (e.g. 'What is this?'). The post-test module was identical to the pre-test module, except that the student received feedback for correct and incorrect responses, as in the recognition exercise. The locations of the items were fixed throughout a lesson, but there were six versions of each lesson with different item locations to ensure the children were learning to identify the images rather than simply their locations.

Knowledge of the vocabulary items was based on receptive identification of the object name or receptive identification based on a specific feature or function of the item. Receptive identification was assessed according to the student's ability to identify the object and click on it when prompted. There were three possible prompts: 'Show me the X', 'Click on the X', and 'Where is the X?', where X refers to the respective vocabulary item. Prompts were selected randomly on each trial. Receptive identification by feature or function was prompted as, 'You use this to X', 'You wear this on your X', or 'You X with this'. The prompt for a given item was the same on each trial. The prompt for the elicitation exercise for naming was 'What is this?' For naming by feature or function, the prompts for the elicitation exercise were 'You use this to?' or 'You wear this on your?' In all cases, the students were required to supply the appropriate verbal response. All students were familiar with this procedure prior to the onset of the study.

Pre-training sessions

Three counterbalanced assessment tests for each lesson were conducted across three consecutive sessions to establish pre-training performance. The students were required to identify each vocabulary item receptively and productively. To assess receptive identification, the program prompted the student to click on the respective image. To assess productive naming, the respective image was highlighted and the program prompted the student to name the specified item. The items were repeated in random order for a total of four times each, providing 24 independent observations (three receptive followed by one productive) for each of the six vocabulary items.

Training sessions

Training followed an alternating treatment design in which the order of lesson presentation was counterbalanced across days and across children. The children were administered each lesson once within a given session. Training continued until a student was able accurately to identify all 24 vocabulary items in the post-test exercise across two consecutive training sessions.

The students were trained in 30 minute sessions, on 3 days per week. During the training sessions, the students were administered the lessons following an alternating treatment design in which the order of presentation of the lessons and the two test conditions was counterbalanced across days and across students. Thus, if on day 1 student 1 received the training lessons in an F1 N1 F2 N2 sequence and student 2 in an N1 F1 N2 F2 sequence, then on day 2 the order of presentation for student 1 would follow an N1 F1 N2 F2 presentation whereas the order of presentation for student 2 would follow an F1 N1 F2 N2 sequence.

Post-training sessions

Following completion of training, three post-training sessions followed the exact procedure of the pre-training.

Procedure

The lessons were constructed and run on a Toshiba Satellite 5005-S504 laptop which has a 1 GHz Pentium III with 512 MB memory and NVIDIA

GeForce2Go graphics running Microsoft Windows 2000 Professional. Baldi was shown on the computer screen as in Figure 1 and the auditory speech was delivered via a Plantronics PC Headset model SR1. All sessions were conducted individually at the student's desk. The sessions lasted approximately 30 minutes. The students completed as many lessons as they could during each session, averaging about two lessons per session (mean = 2.3, SD = 6.1), minimum one and maximum four.

During all pre-training, training, and post-training sessions the students were required to work independently. The experimenter sat behind the student to observe performance and did not intervene during the task. The students knew what was expected of them and were able to complete the task without intervention.

Results

Lesson difficulty

Our evaluation determines the degree to which the face contributes to vocabulary acquisition and retention. Thus, it is important to insure that the lessons were equally difficult in the face (F) and no face (NF) conditions. The difficulty of the lessons can be assessed during the three pre-training blocks prior to the onset of training. Identification performance averaged 0.20 and 0.18 for the lessons with and without the face, respectively. To test whether the lessons were of equal difficulty, an analysis of variance was performed with the proportion of correct identifications as the dependent variable and the lesson type as the independent variable. The analysis revealed no difference in performance among lessons, $F(3, 12) = 0.318$, $p = 0.814$. A second analysis was carried out comparing the face and no face conditions with performance pooled across the two lessons within each condition. There was also no difference between the F and NF performance, $F(1, 4) = 0.128$, $p = 0.735$. Thus, any differences in learning and retention in the F and NF conditions can be safely attributed to the presence of the face rather than differences in lesson difficulty.

Training

All of the students were eventually able to learn correctly to identify their vocabulary items across two consecutive training blocks, although the number of training blocks required to reach this criterion varied. Student 4 reached performance criterion in 14 training blocks whereas the other students required 19, 23, 22, and 15 blocks. In order to carry out an analysis of variance with training block as a factor, a student's adjacent pair of training blocks was pooled when necessary to give 14 blocks for each student.

We first analyzed the training results by carrying out an analysis of variance on the proportion of receptive identifications with face versus no face, pre-test versus post-test, and the 14 training blocks as factors. Figure 2 gives the overall average proportion of correct receptive responses (pooled across both lessons) for the pre-test and post-test conditions with and without the face as a function of training block for the pre-training, training, and post-training conditions. As illustrated in Figure 2, accuracy increased from the initial training block (mean = 0.34, SD = 0.21) to the final training block (mean = 0.88, SD = 0.16), $F(13, 52) = 25.00$, $p < 0.01$. Post-test performance was better than pre-test performance, averaging 0.74 and 0.56, respectively, $F(1, 4) = 26.07$, $p < 0.01$. Overall accuracy was greater when training consisted of the face and the voice (mean = 0.70, SD = 0.250 relative to simply the voice (mean = 0.60, SD = 0.28), but this difference was not statistically significant, $F(1, 4) = 5.63$, $p = 0.08$. Most importantly, however, the face facilitated performance more for the post-test than for the pre-test, showing a statistically significant interaction, $F(1, 4) = 41.81$, $p < 0.01$. Since the post-test/pre-test difference reflects the amount learned in a given lesson, we can conclude that more learning occurred with the face than without it.

The three-way interaction of F versus NF, block, and test, $F(13, 52) = 1.93$, $p < 0.05$, and Figure 2 indicate that the advantage of the face relative to the no face was larger in the post-test than in the pre-test, and that this

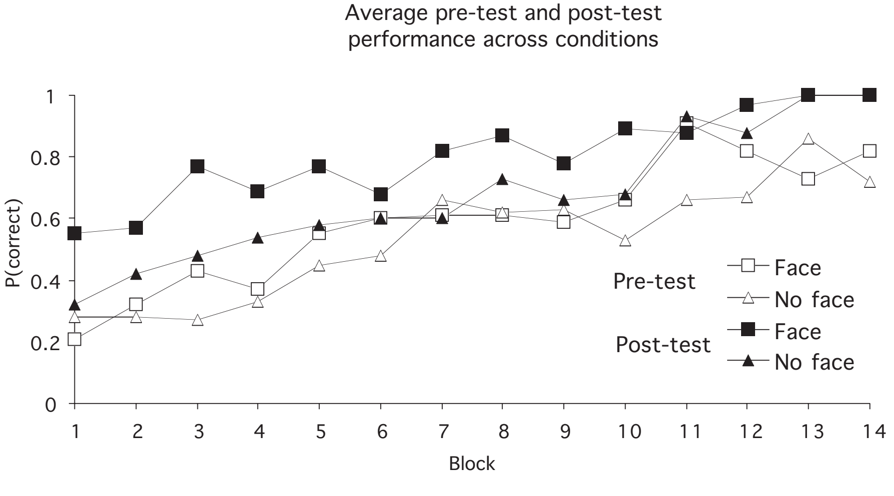

Figure 2 **The overall average proportion of correct receptive responses (pooled across both lessons) for the pre-test and post-test conditions with and without the face as a function of training block**

difference diminished with learning. As expected, the children eventually learned without the face but this learning occurred at a slower rate.

The results show a statistically significant advantage of learning with the face relative to just the voice. This positive result occurred even though the design of the experiment was biased against finding a strong advantage for the face over the no face condition. The recognition exercise always repeated the presentation of an item that the child did not identify correctly. Thus, the child received more practice when an item was missed than when it was correct. Another measure of the effectiveness of the face, therefore, is to determine if there were significantly more practice presentations in the no face relative to the face condition. We computed the total number of presentations of the test words during each block in the recognition exercise. There were more presentations in the no face (mean = 10.64) than the face condition (mean = 9.44), but this result was not statistically significant, $F(1, 4) = 3.736$, $p = 0.125$.

Pre-training versus post-training

Students were trained to criterion in both face and no face conditions. To assess the effects of retention of this training, an analysis of variance was performed on accuracy, with pre-training and post-training and face and no face as independent variables. Figure 3 shows the proportion of correct identifications as a function of the face versus no face conditions for pre-training and post-training performance for each of the five students. As shown in the figure, performance improved dramatically from pre-training (mean = 0.19, SD = 0.09) to post-training (mean = 0.86, SD = 0.11), $F(1, 4) = 273.61$, $p < 0.001$. The interaction between F versus NF conditions and the pre-training versus post-training conditions, $F(1, 4) = 7.46$, $p < 0.05$, indicates that there was better learning with the face than without the face. Specific comparisons were carried out to determine the influence of the face in the pre-training and post-training conditions. There was no significant effect of the face in pre-training, $F(1, 4) = 0.128$, $p > 0.05$, but a significant effect in post-training, $F(1, 4) = 15.046$, $p < 0.01$, averaging 0.92 and 0.80 with and without the face, respectively. This result is consistent with the training analyses and provides evidence that more learning occurred in the face relative to the no face condition.

Individual performance

Three students showed a substantial advantage of having the face during learning. The advantages for the face were more moderate for the other two, one of whom did not have the earlier experience with Baldi in previous experiments (Bosseler and Massaro, 2003: Massaro and Bosseler, 2003) that the other children had. Given the individual differences and that

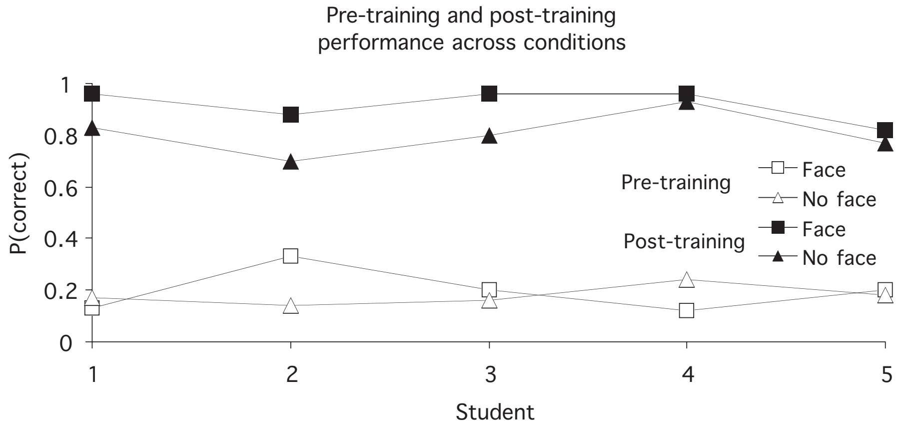

Figure 3 **The average proportion of correct receptive responses (pooled across both lessons) for each of the five children for the pre-training and post-training conditions with and without the face**

only five children were tested, we must be somewhat cautious about the overall effectiveness of the face in vocabulary learning by children with autism.

Discussion

Our results indicate that at least some children with autism can benefit from the presence of the face in vocabulary learning. This result might be surprising to some because children with autism are supposedly less interested in and less capable of understanding and being rewarded by faces (Dawson et al., 2002; Rogers, 1999; Williams et al., 2001). We believe that the face could be made even more effective in our Language Wizard/Tutor. In the present study, the images of the items and Baldi were on the screen simultaneously and the child might not have been looking at Baldi as much as he should have been for optimal learning. We have now included an option to modify the presentation and screen layouts to maximize the influence of the face. In this optional layout, Baldi is first presented alone during the presentation and/or questioning. For example, Baldi comes on and says, 'Look at me. Dandelion.' Then the dandelion is shown and highlighted and Baldi says, 'This is a dandelion. Click on the dandelion.' In the recognition exercise, Baldi comes on and says, 'Look at me. Dandelion'. The images are then shown and Baldi says, 'Click on the dandelion.'

There are several reasons why the use of auditory and visual information together is successful. These include (1) robustness of visual speech, (2) complementarity of auditory and visual speech, and (3) optimal integration of these two sources of information. Speech reading, or the ability to obtain speech information from the face, is robust in that perceivers are fairly good at speech reading even when they are not looking directly at the talker's lips. Furthermore, accuracy is not dramatically reduced when the facial image is blurred (because of poor vision, for example), when the face is viewed from above, below, or in profile, or when there is a large distance between the talker and the viewer (Massaro, 1998, Chapter 14).

Complementarity of auditory and visual information simply means that one of the sources is strong when the other is weak. A distinction between two segments robustly conveyed in one modality is relatively ambiguous in the other modality. For example, the place difference between /ba/ and /da/ is easy to see but relatively difficult to hear. On the other hand, the voicing difference between /ba/ and /pa/ is relatively easy to hear but very difficult to discriminate visually. Two complementary sources of information make their combined use much more informative than would be the case if the two sources were non-complementary, or redundant (Massaro, 1998, pp. 424–7).

The final reason is that perceivers combine or integrate the auditory and visual sources of information in an optimally efficient manner. There are many possible ways to treat two sources of information: use only the most informative source, average the two sources together, or integrate them in such a fashion in which both sources are used but that the least ambiguous source has the most influence. Perceivers in fact integrate the information available from each modality to perform as efficiently as possible. Many different empirical results have been accurately predicted by a fuzzy logical model of perception (FLMP) that describes an optimally efficient process of combination (Massaro, 1998).

These benefits from face-to-face communication may be less valuable to children with autism than others. Bryson (1970) found that the performance of children with autism was poorer when required to match auditory-to-visual and visual-to-vocal events compared to auditory-to-vocal and visual-to-visual events. Other evidence, however, suggests that those with autism are able to perceive information cross-modally (Boucher et al., 1998; 2000; Walker-Andrews et al., 1994). De Gelder et al. (1996) compared the bimodal speech perception of autistic children to normally developing children matched for mental age. It appeared that the children with autism showed less influence of the visual speech than the controls. Massaro and Bosseler (2003), however, provided some evidence that children with autism are influenced to some extent by speech information in the face, can be taught to improve their sensitivity to visible speech, and do integrate cross-modally in speech perception. They tested an integration model (the FLMP) against a non-integration model (the SCM) against the identification results from an expanded-factorial design in which the auditory and visual speech were presented alone or together. In addition, the children were taught to speech read visible speech between two replications of the expanded-factorial design. The FLMP gave a significantly better fit than the SCM across the two replications of the expanded-factorial design. Furthermore, they were successful in improving their speech reading, which resulted in a larger impact of visible speech in bimodal speech perception.

We expect that the face can facilitate performance in a variety of other educational settings and with other populations of students. Hard-of-hearing children, for example, would benefit even more from the face since it improves intelligibility of poorly heard auditory messages. Similarly, students learning a second language could benefit from seeing how articulation of unlearned words occurs. We expect that listening comprehension would also be better when a passage is read with a visible face than when it is presented with a disembodied voice. We hope that investigators will employ talking heads in their educational research and help delineate how the face contributes to perception, understanding, learning, and retention.

Acknowledgements

The research and writing of the article were supported by the National Science Foundation (grant nos CDA-9726363, BCS-9905176, IIS-0086107), Public Health Service (PHS R01 DC00236), a Cure Autism Now Foundation Innovative Technology Award, the National Alliance for Autism Research, and the University of California, Santa Cruz.

Appendix 1 **Description of the subsets of the Language Wizard/Tutor exercises used in this study.**

Exercise	*Purpose*	*Description*
Pre-test	To determine pre-existing knowledge of the vocabulary items prior to tutoring.	Student was prompted to select their name from a list. Baldi asked the student to find the named vocabulary item by clicking on the visual image on the screen. Each response was recorded in a log file.
Presentation	To teach the associated visual images of the word with the vocabulary item.	Baldi began by saying, 'Okay (student), now I will name these for you.' As each item was named, the associated region was highlighted. Baldi prompted the student to identify the item by clicking on the highlighted region. The student was required to click on the highlighted region to help form the association of the visual image and the vocabulary item.

Appendix 1 **continued**

Exercise	*Purpose*	*Description*
Recognition	To provide the student with an opportunity to practice	Baldi began by saying 'Now let's practice, (student)!' Baldi then prompted the student to identify each item by clicking on the visual image. After the student responded, another item was presented until all items had been practiced. The practice trials continued until the student was able to identify all items correctly or when all of the items had been presented twice. Oral praise and stickers were given for correct responses. Only stickers were given for incorrect responses. Each response was recorded in the log file.
Elicitation	To provide the student with an opportunity to practice saying the name of each vocabulary item.	Baldi began by saying 'Now let's practice saying these words.' Each vocabulary item was highlighted individually and Baldi prompted the student to name the vocabulary item by saying, for example, 'What is this?' Each response was recorded and stored in a .wav files for later evaluation.
Post-test	Instructor can compare this performance with the pre-test to determine the number of vocabulary items learned.	Baldi said, 'Now you're ready for the final test!' Baldi then asked the student to click on the specified vocabulary item. Feedback was given by Baldi and in the form of stickers.

Note

1 Baldi is a registered trademark of Dominic W. Massaro.

References

AMERICAN PSYCHIATRIC ASSOCIATION (1994) *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn (DSM-IV). Washington, DC: APA.

BAER, D.M., WOLF, M.M. & RISLEY, T.R. (1968) 'Some Current Dimensions of Applied Behavior Analysis', *Journal of Applied Behavior Analysis* 1: 91–7.

BARKER, L.J. (2003) 'Computer-Assisted Vocabulary Acquisition: The CSLU Vocabulary Tutor in Oral-Deaf Education', *Journal of Deaf Studies and Deaf Education* 8: 187–98.

BARLOW, D.H. & HAYES, S.C. (1979) 'Alternating Treatments Design: One Strategy Comparing the Effects of Two Treatments in a Single Subject', *Journal of Applied Behavior Analysis* 12: 199–210.

BOSSELER, A. & MASSARO, D.W. (2003) 'Development and Evaluation of a Computer-Animated Tutor for Vocabulary and Language Learning for Children with Autism', *Journal of Autism and Developmental Disorders* 33: 653–72.

BOUCHER, J., LEWIS, V. & COLLIS, G. (1998) 'Familiar Face and Voice Matching and Recognition in Children with Autism', *Journal of Child Psychology and Psychiatry* 39: 171–81.

BOUCHER, J., LEWIS, V.A. & COLLIS, G.M. (2000) 'Voice Processing Abilities in Children with Autism, Children with Specific Language Impairments and Young Typically Developing Children', *Journal of Child Psychology and Psychiatry* 41 (7): 847–57.

BRYSON, C.Q. (1970) 'Systematic Identification of Perceptual Disabilities in Autistic Children', *Perceptual and Motor Skill* 31: 239–46.

CHUN, D.M. & PLASS, J.L. (1996) 'Effects of Multimedia Annotations on Vocabulary Acquisition', *Modern Language Journal* 80: 183–98.

DAWSON, G., CARVER, L., MELTZOFF, A.N., PANAGIOTIDES, H., MCPARTLAND, J. & WEBB, S.J. (2002) 'Neural Correlates of Face and Object Recognition in Young Children with Autism Spectrum Disorder, Developmental Delay, and Typical Development', *Child Development* 73: 700–17.

DE GELDER, B., VROOMAN, J. & VAN DER HEIDE, L. (1996) 'Face Recognition and Lip-Reading in Autism', *European Journal of Cognitive Psychology* 3: 69–86.

DUBOIS, M. & VIAL, I. (2000) 'Multimedia Design: The Effects of Relating Multimodal Information', *Journal of Computer Assisted Learning* 16: 157–65.

HAPPÉ, F.G.E. (1996) 'Studying Weak Central Coherence at Low Levels: Children with Autism Do Not Succumb to Visual Illusions. A Research Note', *Journal of Child Psychology and Psychiatry* 37 (7): 873–7.

HEIMANN, M., NELSON, K., TJUS, T. & GILBERG, C. (1995) 'Increasing Reading and Communication Skills in Children with Autism through an Interactive Multimedia Computer Program', *Journal of Autism and Developmental Disorders* 25: 459–80.

KUCERA, H. & FRANCIS, W.N. (1967) *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.

LELORD, G., LAFFONT, F., JUSSEAUME, P. & STEPHANT, J.L. (1973) 'Comparative Study of Conditioning of Averaged Evoked Responses by Coupling Sound and Light in Normal and Autistic Children', *Psychophysiology* 10: 415–25.

MARTINEAU, J., GARREAU, B., ROUX, S. & LELORD, G. (1987) 'Auditory Evoked Responses and Their Modifications during Conditioning Paradigm in Autistic Children', *Journal of Autism and Developmental Disorders* 17: 525–39.

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 *Proceedings of 37th Annual Hawaii International Conference on System Sciences (CD/ROM), Computer Society Press*, 2004, 8 pp. (best paper in 'Emerging Technologies').

MASSARO, D.W. & BOSSELER, A. (2003) 'Perceiving Speech by Ear and Eye: Multimodal Integration by Children with Autism', *Journal of Developmental and Learning Disorders* 7: 111–44.

MASSARO, D.W. & LIGHT, J. (2004) 'Improving the Vocabulary of Children with Hearing Loss', *Volta Review* 104 (3): 141–74.

MOORE, M. & CALVERT, S. (2000) 'Brief Report: Vocabulary Acquisition for Children with Autism: Teacher or Computer Instruction', *Journal of Autism and Developmental Disorders* 30: 359–62.

ROGERS, S.J. (1999) 'An Examination of the Imitation Deficit in Autism', in J. NADEL & G. BUTTERWORTH (eds) *Imitation in Infancy*, pp. 254–84. Cambridge, MA: Cambridge University Press.

SPARROW, S., BALLA, D. & CICCHETTI, D. (1984) *Vineland Adaptive Behaviour Scales.* Windsor: NFER-Nelson.

TAGER-FLUSBERG, H. (1999) 'A Psychological Approach to Understanding the Social and Language Impairments in Autism', *International Review of Psychiatry* 11: 325–34.

WALKER-ANDREWS, A.S., HAVILAND, J.M., HUFFMAN, L. & TOCI, L. (1994) 'Brief Report: Preferential Looking in Intermodal Perception by Children with Autism', *Journal of Autism and Developmental Disorders* 24: 99–107.

WATERHOUSE, L., FEIN, D. & MODAHL, C. (1996) 'Neurofunctional Mechanisms in Autism', *Psychological Review* 103: 457–89.

WECHSLER, D. (1991) *Wechsler Intelligence Scale for Children–Third Edition*. San Antonio, TX: Psychological Corporation.

WILLIAMS, J.H., WHITEN, A., SUDDENDORF, T. & PERRETT, D.I. (2001) 'Imitation, Mirror Neurons and Autism', *Neuroscience and Biobehavior Review* 25: 287–95.