

# Symbiotic Value of an Embodied Agent in Language Learning

Dominic W. Massaro  
Department of Psychology  
University of California, Santa Cruz  
Santa Cruz, CA 95060 U.S.A.  
<http://mambo.ucsc.edu/dwm>  
[Massaro@fuzzy.ucsc.edu](mailto:Massaro@fuzzy.ucsc.edu)

## Abstract

*Our perception and understanding are influenced by a speaker's face and accompanying gestures, as well as the actual sound of the speech. Given the value of face-to-face interaction, our persistent goal has been to develop, evaluate, and use animated agents to teach speech and language. Baldi® is an accurate three-dimensional agent appropriately aligned with either synthesized or natural speech. We describe our language-training program, which utilizes Baldi as a tutor, who guides students through a variety of exercises designed to teach vocabulary and grammar, to improve speech articulation, and to develop linguistic and phonological awareness.*

## 1. Introduction

Spoken language is a pervasive and effective means of communication among our species, and it is only natural that we should have the goal of speech as the primary communication medium to optimize human machine interaction. Notwithstanding the challenges of naturally sounding speech synthesis and accurate speech recognition and meaningful language understanding by machine, some progress has been made. Even so, we cannot expect to achieve this admirable goal in the near future. There are specific applications, however, in which humans can benefit from the speech and language produced by machine. We describe a Language Tutor centered on a computer-animated talking head for language and speech training.

### 1.1. Need for language tutoring

Language challenges are pervasive in today's world. There are millions of individuals who have language and speech disabilities, and these individuals require additional instruction in

language learning. Currently, however, these needs are not being met. The problem that the people with these disabilities face is that there are not enough skilled teachers and professionals to give them the one on one attention that they need. So they resort to other resources, such as books or other media, but the problems with these are that they are not easily personalized to the students' needs, they lack the engaging capability of a teacher, they are rather expensive, and they are relatively ineffective.

In addition to these individuals with specific disabilities, there are many other persons who must learn a new language. Given the highly mobile society, individuals of all walks of life find themselves in situations in which successful business and social interactions require use of a nonnative language. As an obvious example, English is becoming increasingly necessary and desirable, and the number of people in the world who are learning English is increasing at a rapid rate. Many call centers are now locating outside of the U.S., for example, and there is an immediate need to coach accent reduction for their personnel.

### 1.2. Value of talking heads

While the auditory signal alone is often adequate for communication, visual information from movements of the lips, tongue and jaws enhance intelligibility of the acoustic stimulus (particularly in noisy environments). Moreover, speech is enriched by the facial expressions, emotions and gestures produced by a speaker [1]. For individuals who hear well, these visible aspects of speech are especially important in noisy environments. The visual components of speech offer a lifeline to those with a substantial hearing loss: Understanding visible speech can make the difference in effectively communicating orally with others or a life of relative isolation from oral society [2].

Our persistent goal has been to develop, evaluate, and apply animated agents to produce accurate visible speech and to facilitate face-to-

face oral communication. These enhanced characters can also function effectively as language tutors, reading tutors, or personal agents in human machine interaction. Although traditionally speech has been viewed as solely an auditory phenomenon, speech as a multimodal phenomenon is supported by experiments indicating that our perception and understanding are influenced by a speaker's face and accompanying gestures, as well as the actual sound of the speech [1].

The number of words understood from a degraded auditory message can often be doubled by pairing the message with visible speech from the talker's face [3]. Furthermore, the strong influence of visible speech is not limited to situations with degraded auditory input, and a perceiver's recognition reflects the contribution of both sound and sight.

There are several reasons why the use of auditory and visual information together is so successful. These include a) robustness of visual speech, b) complementarity of auditory and visual speech, and c) optimal integration of these two sources of information. Speechreading, or the ability to obtain speech information from the face, is robust in that perceivers are fairly good at speechreading even when they are not looking directly at the talker's lips [1]. 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.

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 tends to be relatively ambiguous in the other modality. Two complementary sources of information make their combined use much more informative than would be the case if the two sources were non-complementary [1].

Perceivers combine or integrate the auditory and visual sources of information in an optimally efficient manner. Many different empirical results have been accurately predicted by a Fuzzy Logical Model of Perception (FLMP) that describes an optimally efficient process of combination [1].

## 2. Baldi®, an animated tutor

The value of visible speech in face-to-face communication was the primary motivation for the development of Baldi®, a 3-D computer-

animated talking head. Baldi provides realistic visible speech that is almost as accurate as a natural speaker [1,4]. The goal of the visible speech synthesis carried out in the Perceptual Science Laboratory (PSL) has been to develop a polygon (wireframe) model with realistic motions (but not to duplicate the musculature of the face). Baldi's visible speech can be appropriately aligned with either synthesized or natural auditory speech. Our software can generate a talking face in real time on a commodity PC, and Baldi is able to say anything at any time in our applications.

In our synthesis algorithm, each segment is specified with a target value for each facial control parameter. Coarticulation, defined as changes in the articulation of a speech segment due to the influence of neighboring segments, is based on a model of speech production using rules that describe the relative dominance of the characteristics of the speech segments.

A central and somewhat unique quality of our work is the empirical evaluation of the visible speech synthesis, which is carried out hand-in-hand with its development. The quality and intelligibility of Baldi's visible speech has been repeatedly modified and evaluated to accurately simulate naturally talking humans [1]. The gold standard we use is how well Baldi compares to a real person. Given that viewing a natural face improves speech perception, we determine the extent to which Baldi provides a similar improvement. We repeatedly modify the control values of Baldi in order to meet this criterion. We modify some of the control values by hand and also use data from measurements of real people talking [4,5].

### 2.1. Value of agent tutors

Several advantages of utilizing a computer-animated agent as a language tutor are clear, including the popularity of computers and embodied conversational agents. Computer-based instruction is emerging as a prevalent method to train and develop vocabulary knowledge for both native and second-language learners and individuals with special needs [6,7]. An incentive to employing computer-controlled applications for training is the ease in which automated practice, feedback, and branching can be programmed. Another valuable component is the potential to present multiple sources of information, such as text, sound, and images in parallel.

A second advantage is the availability of the

program. Instruction is always available to the child, 24 hours a day 365 days a year. Furthermore, instruction occurs in a one-on-one learning environment for the students. We have found that the students enjoy working with Baldi because he offers extreme patience, he doesn't become angry, tired, or bored, and he is in effect a perpetual teaching machine. Applications with animated tutors perceived as supportive and likeable will engage foreign language and ESL learners, reading impaired, autistic and other children with special needs in face-to-face computerized lessons. We now describe several different applications utilizing Baldi to carry out language tutoring.

### 3. Pedagogy of language learning

Vocabulary knowledge is critical for understanding the world and for language competence in both spoken language and in reading. There is empirical evidence that very young children more easily form conceptual categories when category labels are available than when they are not [8]. Even children experiencing language delays because of specific language impairment benefit once this level of word knowledge is obtained. It is also well-known that vocabulary knowledge is positively correlated with both listening and reading comprehension [9]. It follows that increasing the pervasiveness and effectiveness of vocabulary learning offers a timely opportunity for improving conceptual knowledge and language competence for all individuals, whether or not they are disadvantaged because of sensory limitations, learning disabilities, or social condition.

Learning and retention are positively correlated with the time spent learning. Our technology offers a platform for unlimited instruction, which can be initiated when and wherever the child and/or mentor chooses. Baldi and the accompanying lessons are perpetual. Take, for example, children with autism, who have irregular sleep patterns. A child could conceivably wake in the middle of the night and participate in language learning with Baldi as his or her friendly guide.

Instruction can be tailored exactly to the

student's need, which is best implemented in a one-on-one learning environment for the students. Other benefits of our program include the ability to seamlessly meld spoken and written language, and provide a semblance of a game-playing experience while actually learning. Given that education research has shown that children can be taught new word meanings by using drill and practice methods [10], we implement these basic features in an application to teach vocabulary and grammar.

### 4. Language Wizard/Player

Our early experience with the use of our initial software applications was disconcerting but highly informative. Our initial goal was to provide so-called easy-to-use tools to teachers to develop their own instructional applications across all of their teaching domains [11]. Much to our disappointment, however, we found that only one of the teachers who already was using other software products and writing his own applications was successful in creating lessons. We thus embarked on the development of a so-called Wizard that would be useful to everyone involved in the educational process. The Language Player, although relatively mellow by video game standards, has enough engaging interactive features to engage the student in mastering the lesson. The resulting lessons encompass and instantiate the developments in the pedagogy of how language is learned, remembered and used.

The Language Wizard is an easy-to-use application that allows the coach to create a lesson that is tailored to the needs of the student. The design of this pedagogy is based on educational principles to optimize learning, which are not always intuitive. Other benefits of our program include the ability to seamlessly meld spoken and written language, provide a semblance of a game-playing experience while actually learning, and to lead the child along a growth path that always bridges his or her current "zone of proximal development." The Wizard allows the coach to exploit this zone with individualized lessons, and with lessons that can bypass repetitive training when student responses indicate that material is mastered.

Table 1. Description of the 8 exercises available in the Language Tutor. Each exercise is optional and the specifications for each exercise can be made independently of the other exercises. For each exercise, the items can be randomly presented in a block of trials. The number of trial blocks is also specified independently for each lesson. The dialog and feedback for each exercise are also chosen by the coach creating the lesson. The feedback can include emoticons showing a happy or sad face.

Exercise	Description
Pre-Test	Baldi instructs the student to “click on the (word)” and the student is required to drag the computer mouse over the item that was just presented and click on it.
Presentation	One image is highlighted and Baldi tells the student “this is the (word)”. Baldi then instructs the student to “show me the (word)” and the student is required to click on it. The student’s response shows that they knew which image was being described.
Recognition	Baldi instructs the student to “click on the (word).
Reading	The written text of each item is displayed in a separate area from the images. Baldi instructs the student to click on the written word corresponding to the highlighted image.
Spelling	One of the images is highlighted and Baldi asks the student to type the corresponding word.
Imitation	One of images is highlighted and Baldi names the item. The student is instructed to repeat the name Baldi had just said.
Elicitation	One of images is highlighted and Baldi asks the student to name it.
Post-Test	Baldi instructs the student to “click on the (word)”.

One of the principles of learning that we exploit most is the value of multiple sources of information in perception, recognition, learning, and retention. An interactive multimedia environment is ideally suited for learning [6]. 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. Many aspects of our lessons enhance and reinforce learning. For example, the existing program makes it possible for the students to 1) Observe the words being spoken by a realistic talking interlocutor (Baldi), 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, and 6) Spell the word by typing, and 7) Observe and respond to the word used in context.

#### 4.1. Effectiveness for hearing loss

It is well known that children with hearing loss have significant deficits in both spoken and written vocabulary knowledge. One reason is that these children tend not to overhear other conversations because of their limited hearing and are thus shut off from an opportunity to learn vocabulary. These children often do not have names for specific things and concepts and



Figure 1. A computer screen from a vocabulary lesson on fruits and vegetables, illustrating the format of the Language Player. Each lesson contains Baldi, the vocabulary items and written text (not present in this exercise), and “stickers”. For example, Baldi says “Click on the beet”. The student clicks on the appropriate region and feedback in the form of Baldi’s spoken reaction and stickers (e.g., happy and disgusted faces) are given for each response.

therefore communicate with phrases such as “the window in the front of the car,” “the big shelf where the sink is,” or “the step by the street” rather than “windshield,” “counter,” or “curb” [7].

The Language Wizard/Player has also been in use at the Tucker Maxon Oral School (TMOS) in Portland, Oregon. Both an investigative report carried out by ABC Primetime Live and Barker [7] provided some evidence that the application is effective for vocabulary learning across a fairly wide range of ages and language ability. Although these evaluations indicated that the children learned and retained a significant number of new words, it is possible the children were learning the words outside of the Language Player environment. Furthermore, it is valuable to assess both identification and production of the words given that only identification was measured in the previous study [7].

To address these issues, we carried out an experiment based on a within student multiple

baseline design where certain words were continuously being tested while other words were being tested and trained [12]. Although the student's instructors and speech therapists agreed not to teach or use these words during our investigation, it is still possible that the words could be learned outside of the Language Player environment. The single student multiple baseline design monitors this possibility by providing a continuous measure of the knowledge of words that are not being trained. Thus, any significant differences in performance on the trained words and untrained words can be attributed to the Language Player training program itself rather than some other factor.

Eight children with hearing loss, 2 male ages 6 and 7, 6 female ages 9 and 10, were recruited from The Jackson Hearing Center in Los Altos, California. The male students were in grade 1 and the female students in grade 4 respectively and all students needed help with their vocabulary building skills as suggested by their regular day teachers. One child had a cochlear implant and the seven other children had hearing aids in both ears except for one child with an aid in just a single ear. Using the Language Wizard, the experimenter developed a set of lessons with a collection of vocabulary items that was individually composed for each student. Each collection of items was comprised of 24 items, broken down into 3 categories of 8 items each. Three lessons with 8 items each were made for each child.

Images of the vocabulary items were presented on the screen next to Baldi as he spoke, as illustrated in Figure 1. Assessment was carried out on all of the items at the beginning of each lesson. It included identifying and producing the vocabulary item without feedback (the Pre-Test and Elicitation exercise in Table 1). Training on the appropriate word set followed this testing.

Figure 2 gives the results of identification and production for one of the eight students. These results are typical because the results were highly consistent across the eight students. As expected, identification accuracy was always higher than production accuracy. This result is expected because a student would have to know the name of an item to pronounce it correctly. There was little knowledge of the test items without training, even though these items were repeatedly tested for many days. Once training began on a set of items, performance improved fairly quickly until asymptotic knowledge was obtained. This knowledge did not degrade after training on these words ended and training on

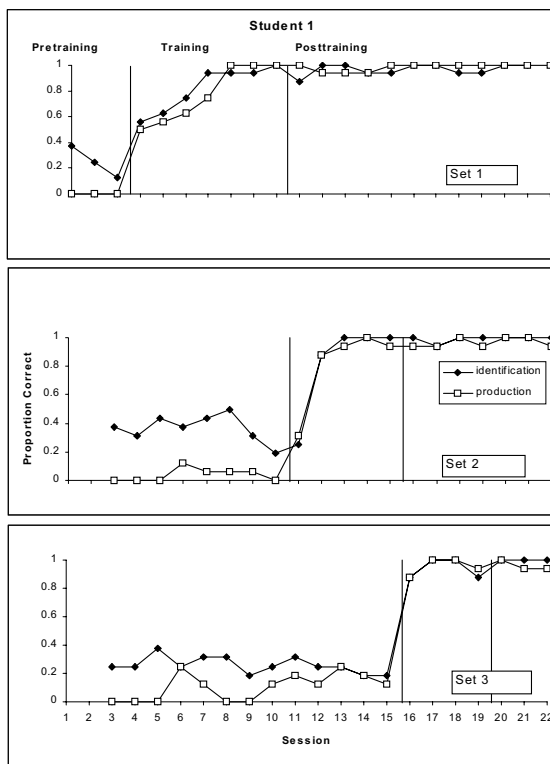


Figure 2. Proportion of correctly identified (solid black diamonds) and correctly produced (empty white squares) items across the testing sessions for student 1. The training occurred between the two vertical bars. The figure illustrates that once training was implemented identification performance increased dramatically, and remained accurate without further training.

other words took place. In addition, a reassessment test given about 4 weeks after completion of the experiment revealed that the students retained the items that were learned.

It should be noted that this experimental evaluation of effectiveness challenged the students much more than what would be typically be involved in the use of the software. The students were repeatedly tested on many words before actually given an opportunity to learn the words, and they were continually tested on words after they were mastered. Although there was a large diversity among the 8 students, all of them were successful in learning and retaining new vocabulary.

#### 4.2. Effectiveness for autism

The Language Wizard/Player has also been used in evaluating vocabulary acquisition, retention and generalization in children with autism [13]. Autism is a pervasive developmental disorder, which apparently has increased from affecting approximately 1 in every 500 children to 1 in 300 [14]. Although the etiology of autism is not known, individuals diagnosed with autism must exhibit a) delayed or deviant language and communication, b) impaired social and reciprocal social interactions, and 3) restricted interests and repetitive behaviors [15]. The language and communicative deficits are particularly salient, with large individual variations in the degree to which autistic children develop the fundamental lexical, semantic, syntactic, phonological, and pragmatic components of language. For the roughly one-half of the autistic population who develop some form of functional language [15], the onset and rate at which the children pass through linguistic milestones are often delayed (e.g. no single words by age 2 years, no communicative phrases by age 3. Given their limitations in language processing, a better understanding of their speech perception should be particularly valuable.

This study consisted of two phases. Phase 1 measured vocabulary acquisition and retention. Phase 2 tested whether vocabulary acquisition was due to the Language Player or outside sources and whether the acquired words could be generalized to new images and outside of the Language Player environment. Vocabulary lessons were constructed, consisting of over 84 unique lessons with vocabulary items selected from the curriculum of two schools [13]. The participants were eight children diagnosed with

autism, ranging in age from 7-11 years. All of the students exhibit delays in all areas of academics, particularly in the areas of language and adaptive functioning. Seven of the eight children were capable of speech.

The average results indicated that the children learned many new words, grammatical constructions and concepts, proving that the Language Player provided a valuable learning environment for these children. In addition, a delayed test given more than 30 days after the learning sessions took place showed that the children retained over 85% of the words that they learned. This learning and retention of new vocabulary, grammar, and language use is a significant accomplishment for autistic children.

Seven of the eight students appeared to enjoy working with Baldi. The children made statements like "Hi Baldi" and "I love you Baldi". The stickers generated for correct (happy face) and incorrect (sad face) responses proved to be an effective way to provide feedback for the children. Some students displayed frustration when they received more than one sad face and responded positively to the happy faces, saying "Look", pointing, or laughing when a happy face appeared. We also observed the students providing themselves the same verbal praise given by Baldi such as "Good job", or prompting the experimenter to say "Good job". One student did not enjoy the program, even though he did demonstrate learning during both the training sessions and the reassessment. His refusal to cooperate led to his withdrawal after the experiment was completed by parent request.

Although all of the children demonstrated learning from initial assessment to final reassessment, it is possible that the children were learning the words outside of our learning program (for example, from speech therapists or in their school curriculum). Furthermore, it is important to know whether the vocabulary knowledge would generalize to new pictorial instances of the words. To address these questions, a second investigation used the single subject multiple probe design, as was done in [12]. Once a student achieved 100% correct, generalization tests and training were carried out with novel images. The placement of the images relative to one another was also random in each lesson. Assessment and training continued until the student was able to accurately identify at least 5 out of 6 vocabulary items across four unique sets of images.

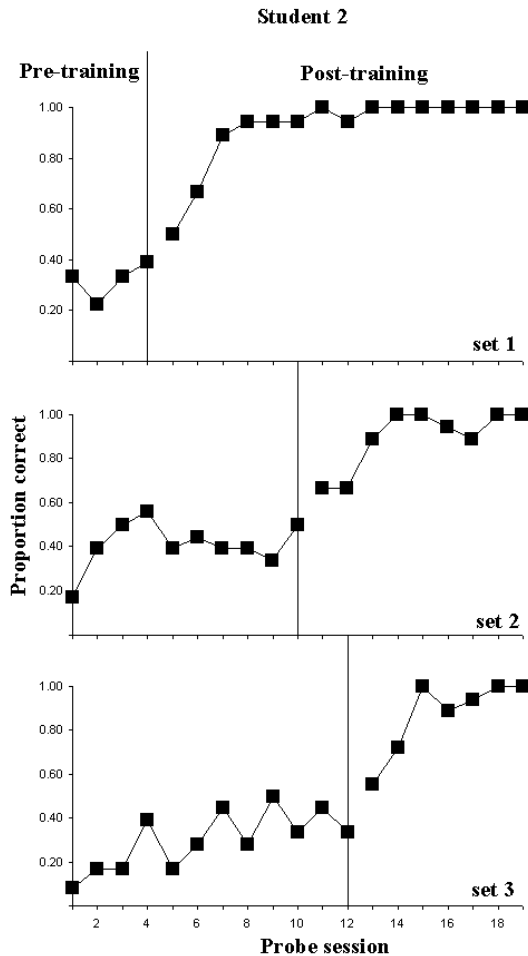


Figure 3. Mean proportion correct during pre-training and post-training probes for each of the three word sets for one of the six students. The vertical lines separate the pre-training and post-training. The figure illustrates that once training was implemented identification performance increased dramatically.

Although performance varied dramatically across the children and across the word sets during the pre-training sessions, training was effective for all words sets for all children. Figure 3 displays the proportion of correct responses for one of the students during the test sessions conducted at pre-training and post-training for each of the three word sets. The vertical lines in each of the three panels indicates the last pre-training session before the onset of training for that word set. Some of the words were known prior to training, and were even learned to some degree without training. Given training, all of the students attained our criterion for identification accuracy for each word set and were also able to generalize accurate identification to four instances of untrained

images. 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 Player. These results show that our learning program is effective for children with autism, as it is for children with hearing loss.

We were gratified to learn that the same application could be used successfully with both autistic children and children with hearing loss [16]. It has also been used successfully with normally developing children [7]. Specific interactions can be easily modified with the Language Wizard to accommodate group and individual differences. For example, autistic children are much more disrupted by negative feedback, and the lesson can be easily designed to instantiate errorless learning.

It is important to know to what extent the face facilitated relative to the voice alone facilitated this learning process. To address this question, Baldi was implemented in the Language Wizard/Player, in which each child continuously learned to criterion two sets of words with and without the face [17]. The rate of learning was significantly faster and the retention was better with than without the face. Thus, we can conclude that Baldi adds a significant component to the learning process.

### 5. Speech production training

Although many of the subtle distinctions among segments are not visible on the outside of the face, the skin of our talking head can be made transparent so that the inside of the vocal tract is visible, or we can present a cutaway view of the head along the sagittal plane. Baldi has a tongue, hard palate and three-dimensional teeth and his internal articulatory movements have been trained with electropalatography and ultrasound data from natural speech [18]. These internal structures can be used to pedagogically illustrate correct articulation. The goal is to instruct the child by revealing the appropriate movements of the tongue relative to the hard palate and teeth.

As an example, a unique view of Baldi's internal articulators can be presented by rotating the exposed head and vocal tract to be oriented away from the student. It is possible that this back-of-head view would be much more conducive to learning language production. The

tongue in this view moves away from and towards the student in the same way as the student's own tongue would move. This correspondence between views of the target and the student's articulators might facilitate speech production learning. One analogy is the way one might use a map. We often orient the map in the direction we are headed to make it easier to follow (e.g. turning right on the map is equivalent to turning right in reality).

Another characteristic of the training is to provide additional cues for visible speech perception. Baldi can illustrate the articulatory movements, and he can be made even more informative by embellishing of the visible speech with added features. Distinguishing phonemes that have similar visible articulations, such as the difference between voiced and voiceless segments, can be indicated by vibrating the neck. Nasal sounds can be marked by making the nasal opening red, and turbulent airflow can be characterized by lines emanating from the mouth during articulation. These embellished speech cues could make the face more informative than it normally is.

**5.1. Effectiveness for hearing loss**

Children with hearing loss require guided instruction in speech perception and production. Some of the distinctions in spoken language

cannot be heard with degraded hearing--even when the hearing loss has been compensated by hearing aids or cochlear implants. To overcome this limitation, we use visible speech when providing our stimuli. Based on reading research, we expected that visible cues would allow for heightened awareness of the articulation of these segments and assist in the training process

Seven students with hearing loss (2 male and 5 female), from the Jackson Hearing Center and JLS Middle School in Los Altos, California participated in the study [19]. The students ranged in age from 8 to 13. Their unaided hearing varied to some extent, but all children had a severe hearing loss in at least one ear. The students were trained to discriminate minimal pairs of words bimodally (auditorily and visually), and were also trained to produce various speech segments by visual information about how the inside oral articulators work during speech production. The articulators were displayed from different vantage points so that the subtleties of articulation could be optimally visualized. The speech was also slowed down significantly to emphasize and elongate the target phonemes, allowing for clearer understanding of how the target segment is produced in isolation or with other segments.

Figure 4 shows that the students' ability to perceive and produce words involving the trained segments improved from pre-test to post-

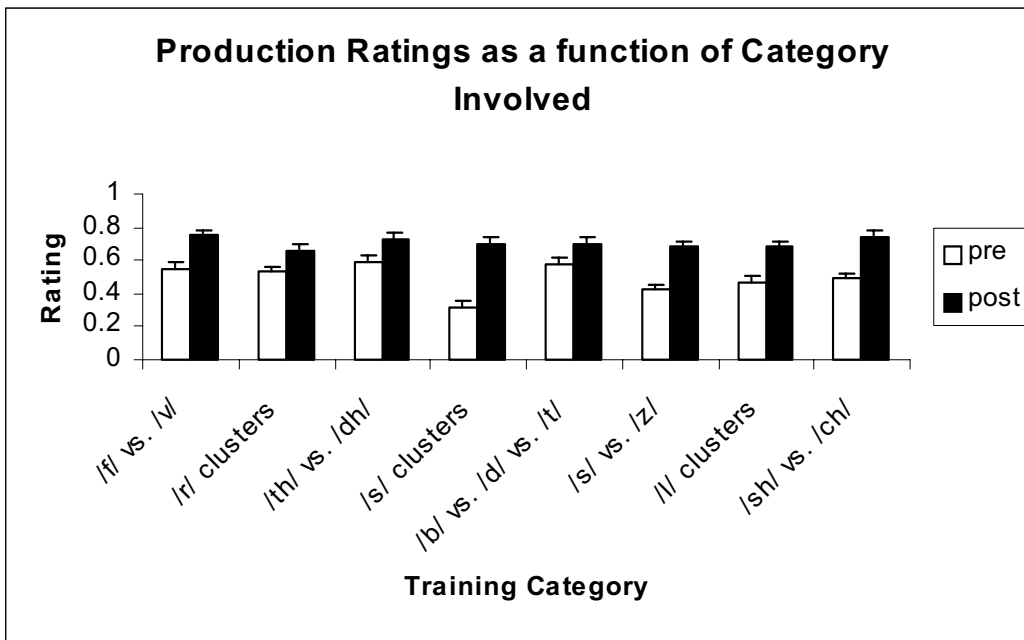


Figure 4. Intelligibility ratings of the pre-test and post-test word productions (and standard error bars) for each of the eight training categories.

test. Intelligibility ratings of the post-test productions were significantly higher than pre-test productions, indicating significant learning. It is always possible that some of this learning occurred independently of our program or was simply based on routine practice. To test this possibility, we assessed the students' productions six weeks after training was completed. Although these productions were still rated as more intelligible than the pre-test productions, they were significantly lower than post-test ratings, indicating some decrement due to lack of continued use. This is evidence that at least some of the improvement must be due to our program. Future studies can now focus on which specific training regimens for which contrasts are most effective.

Although there were individual differences in aided hearing thresholds, attitude, and cognitive level, the training program helped all of the children. Student 1 was cooperative but did not like working with Baldi, and she appeared to have had physiological difficulty producing some of the sounds. Student 2 was often not as cooperative as the others during training but he gained certain important skills from Baldi (e.g. voicing instead of nasalizing certain sounds). For the sounds that he knew he had learned, he was confident and impressive. Student 3 loved working with Baldi and was consistently involved and motivated. He is not a very social child and doesn't usually cooperate in class but Baldi was his favorite part of the day. Student 4 was apprehensive at first but she became more comfortable. Although her speech was already quite intelligible, she could hear and feel an improvement in her own speech. Student 5 was very cooperative and thought Baldi was funny. She recently received a cochlear implant and had gained a lot of confidence in her speech. She thought the program was too easy and didn't think she needed so much practice. By the end, she knew that she improved but perhaps felt it was only marginal. Student 6 was a very attentive and cooperative student. She was always asking questions and wanted to learn as much as she could. She indicated that the program was a great teaching tool and she quickly noticed the benefits of the program. She definitely could hear/feel an improvement in her own speech and would have liked to continue with this program. Student 7 was not as interested in working with Baldi as she was in talking to the experimenter, who often had to redirect her focus and use rewards to keep her motivated. She was receptive to these

instructions and rewards and knew that it was good practice for her.

The present findings suggest that Baldi is an effective tutor for speech training students with hearing loss. There are other advantages of Baldi that were not exploited in the present study. Baldi can be accessed at any time, used as frequently as wished and modified to suit individual needs. Baldi also proved beneficial even though students in this study were continually receiving speech training with their regular and speech teachers before, during and after this study took place. Baldi appears to offer unique features that can be added to the arsenal of Speech-language pathologists.

## 5.2 Second language learning

This study investigated the effectiveness of Baldi for teaching non-native phonetic contrasts, by comparing instruction illustrating the internal articulatory processes of the oral cavity versus instruction providing just the normal view of the tutor's face. Eleven Japanese speakers of English as a second language were bimodally trained under both instruction methods to identify and produce American English /r/ and /l/ in a within-subject design [20]. We expected that both perception and production of these segments could be improved with bimodal speech training in which movement of the internal articulators is illustrated.

Both the perception and production of words by the Japanese trainees generally improved from one day of training to the next. In addition, all of the trainees in this study were highly enthusiastic and motivated to improve their speech production of English. They were disappointed to see the experiment end because they realized they had not yet mastered all of the training items. They also became more confident in producing the easier training items. We might speculate that visible speech and our technique of revealing the inside articulators during training would show an advantage over simple auditory training.

## 6. Conclusion

We have found that the Language Wizard/Player allows easy creation and presentation of a language lessons, and is effective in teaching vocabulary and grammar. We have also observed that Baldi's unique characteristics allow a novel approach to training speech production to both children with hearing loss and adults learning a

new language. The science and technology of Baldi holds great promise in language learning, dialog, human-machine interaction, and education.

## 7. Acknowledgement

The research and writing of the paper were supported by the National Science Foundation (Grant No. CDA-9726363, Grant No. BCS-9905176, Grant No. IIS-0086107), Public Health Service (Grant No. PHS R01 DC00236), a Cure Autism Now Foundation Innovative Technology Award, and the University of California, Santa Cruz.

## 8. References

[1] Massaro, D. W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. MIT Press: Cambridge, MA.

[2] Trychin S (1997) *Guidelines for providing mental health services to people who are hard of hearing*. Washington D.C.: Gallaudet University.

[3] Jesse, A., Vrignaud, N., & Massaro, D. W. (2000/01). The processing of information from multiple sources in simultaneous interpreting. *Interpreting*, 5, 95-115.

[4] Cohen, M.M. and Massaro, D.W., and Clark R. (2002) Training a talking head. In *Proceedings of ICM'02, IEEE Fourth International Conference on Multimodal Interfaces*. October 14-16, Pittsburgh, Pennsylvania.

[5] Ouni, S. Massaro, D.W., Cohen, M.M., Young, K. & Jesse, A. (2003). Internationalization of a Talking Head. 15th International Congress of Phonetic Sciences, August 3-9, Barcelona.

[6] Wood, J. (2001). Can software support children's vocabulary development? *Language Learning & Technology*, 5, 166-201.

[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, 187-198.

[8] Waxman, S. R. (2002). Early word-learning and conceptual development: Everything had a name, and each name gave birth to a new thought. In U. Goswami (Ed.) *Blackwell Handbook of childhood cognitive development* (pp. 102-126). Malden, MA: Blackwell publishing.

[9] Anderson, R. C., & Freebody, P. (1981).

Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research perspectives* (pp. 71-117). Newark, DE: International Reading Association.

[10] Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust Vocabulary Instruction*. New York: The Guilford Press.

[11] Massaro, D.W., Cohen, M. M., & Beskow, J. (2000). Developing and evaluating conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.) *Embodied conversational agents*. Cambridge, MA: MIT Press, p. 286-318.

[12] Massaro, D.W., & Light, J. (in press). Improving the vocabulary of children with hearing loss. *Volta Review*, in press.

[13] Bosseler, A. and Massaro, D.W. (in press). Development and evaluation of a computer-animated tutor for vocabulary and language learning for children with autism. *Journal of Autism and Developmental Disorders*.

[14] M.I.N.D. Institute, University of California, Davis, [http://www.dds.ca.gov/Autism/Autism\\_main.cfm](http://www.dds.ca.gov/Autism/Autism_main.cfm).

[15] American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders, DSM-IV (4<sup>th</sup> ed.)*. Washington, DC.

[16] Massaro, D.W., Bosseler, A. and Light, J. (2003, August). Development and evaluation of a computer-animated tutor for language and vocabulary learning. 15th International Congress of Phonetic Sciences (ICPhS '03), Barcelona, Spain.

[17] Massaro, D.W., Bosseler, A. (submitted). Read my Lips: The Importance of the Face in a Computer-Animated Tutor for Autistic Children Learning Language. *Autism: The International Journal of Research and Practice*, submitted.

[18] Cohen, M. M., Beskow, J. & Massaro, D. W. (1998). Recent developments in facial animation: An inside view. *Proceedings of Auditory Visual Speech Perception '98*. (pp. 201-206). Terrigal-Sydney Australia, December, 1998.

[19] Massaro, D.W., & Light, J. (in press). Using visible speech for training perception and production of speech for hard of hearing individuals. *Journal of Speech, Language, and Hearing Research*, in press.

[20] Massaro, D. W., & Light (2003, September). Read my tongue movements: bimodal learning to perceive and produce non-native speech /r/ and /l/. *Eurospeech 2003-Switzerland (Interspeech)*. 8th European Conference on Speech Communication and Technology, Geneva, Switzerland.