Multimedia Technologies and Solutions for Educational Applications:

Opportunities, Trends and Challenges

Patti Price
PPRICE Speech and Language Technology
Menlo Park, CA USA
pjp@pprice.com

Abstract— **This report aims to provide an overview of multimedia technologies in education, particularly language technologies, and particularly applications aimed at children. Emphasis is on those aspects that may not necessarily be familiar to engineers, such as linguistics and language pedagogy.**

Keywords—education, linguistics, language pedagogy

Topic area—Emerging Applications

I. Introduction

Signal processing engineers are likely to know much about signals and about processing them. They also are likely to be good at finding new signals to process and to combine in innovative ways. Many of these innovations are likely to find uses in educational applications. Success, however, can be limited by the fact that there is often a negative correlation between those who crave high tech and those who crave high touch (typical of pedagogical experts).

People are spending more time in educational activities. An increasing percentage of the U.S. population is finishing high school (the percentage has more than tripled in 50 years) and going on to college over the last 50 years [1]. UNESCO reports similar trends worldwide, as measured, *e.g*., by rising literacy rates [2]. In an age where today's teenagers will likely have several jobs, none of which exists today, the trend of continuing education is likely to continue to increase. Educational demands outside the classroom are also increasing. In the sense of sharing knowledge or information, much of what we do is education, whether we are in school, online, or reading labels at the grocery store.

Our primary method of exchanging information is still language, which may account for some bias toward language technologies among multimedia technologies in education. About 40,000 years ago or so, our species underwent what Jared Diamond (*e.g.*, [3]) has characterized a 'Great Leap Forward', when human culture seemed to change at a much greater speed. Many believe this coincided with the spread of human language. Oral language, as the first information technology, enabled us to communicate ideas and experiences to others *without* requiring them to have direct experience themselves. Another leap, written language, enabled communication with others distant in time and place. Let's imagine the possibility that current multimedia technology is poised to create another great leap, one in which that original, pre-language method of direct experience, together with language, can be recreated to improve education.

II. Overview of Multimedia in Education

Let's begin by giving an idea of how multimedia is used in education by giving a sample of some current applications:

- MossTalk.com provides training for those with language impairments such as aphasia, typically acquired through stroke or head injury. The software provides exercises and tracks results in the ability to name or describe pictures. Our aging population will face higher incidents of stroke and other age-related changes in senses and cognitive skills. For a summary of some of the more general issues of interfaces for aging populations, see [4].
- Several companies are now offering multimedia technology to learn a new language. EnglishXchange.com has an innovative approach using computer games. A study involving 121 sixth grade Chinese students learning English compared its effectiveness relative to traditional classroom methods using native English teachers. After 8 weeks, students using the software achieved significantly higher oral and written English scores than did those taught by native English teachers [5].
- Although using olfaction is less well studied than other sensory channels, it is a sense that has been around for a very long time and offers some interesting interactive possibilities now beginning to be used. See, for example, the use of aroma as a notification method less disruptive than sight or sound [6].
- A set of case studies from web3d.org describes 3d immersive educational applications, *e.g,*: visualization of statistics, Iraqi checkpoint training in cultural gestures, a 3d atlas, water systems treatment visualization and training, training and mission rehearsal for oil and gas operations, lunar orbit and landing site visualization, a GPS device to monitor position and speed on outdoor

Partial support from NSF Grant No. 0326214 is gratefully acknowledged. Opinions are those of the author and not of NSF.

fitness trails, NASA's interactive 3D guide to the galaxy, and a high-fidelity diamond simulation for sales and marketing. For an overview, see [7].

- WorldHaptics2007.org describes haptic technology used in educational applications, including, *e.g.*, a self-defense training system, a skill transfer system using a robotic training simulation, rehabilitation of fine motor skills, and an interface for the blind to access visual information.
- The Nintendo Wii has made headlines for simulating sports activities. Although, these could be used as sports or physical training, their prime use is pure entertainment. This is not to say that entertainment cannot be useful in learning. In fact, [8] argues that video games are precisely so addictive because they appeal to and enhance our craving and ability to learn. Research is starting to appear indicating that playing action video games can help in cases of amblyopia, *e.g.,* [9] and a simple search at MedGadget.com for 'video games' reveals emerging uses for games in other rehabilitation training.

A. What signals to process?

The usual 'signals' in educational applications have been text, graphics, animations, audio and video, and sometimes speech synthesis and/or speech recognition. However, as seen in the sample applications above, many other signals could be used for input and/or output, *e.g.,* haptics, natural language generation and understanding, aroma, gesture, G.P.S. information, pronunciation scoring, face synthesis/analysis, and detection of mood, stress, or health indices.

Engineers will be aware of many existing and emerging technologies. In principal any of our senses (sight, hearing, touch, taste, smell) could be used for input or output, though of course some are used more than others. Engineers may also be aware of how signals can be combined to supplement impaired senses, to augment normal senses, to collaborate in new ways, or to have fun. Pedagogy experts will likely know about the constraints of human abilities, preferences and limitations. Linguists will know about language and dialect issues. There may be very few people who know all of these areas well. Perhaps our technology can help us to increase our collective intelligence through collaboration.

B. Language and dialect issues

Language is a signaling system since it conveys information by conventions agreed on by the community using it. The mapping is between sound (for spoken language) or symbol (for written language) and meaning. Those who speak the same dialect share more signaling conventions than those who speak different dialects or languages.

We tend to think of a language as a rather discrete entity. A linguist, however, will say that a language is just 'a dialect with its own army and navy' – pointing out that two 'dialects' of 'Chinese' may not be mutually intelligible to the speakers, though speakers of some pairs of different 'languages' can understand each other. Serbian and Croatian form an interesting example: the oral varieties are very similar, though they use different writing systems.

The simple observation of how different, though not entirely foreign, Chaucer's English is from modern English should illustrate how languages can change over the course of 650 years or so: "In Flaundres whilom was a compaignye of yonge folk that haunteden folye, as riot, hasard, stywes, and tavernes, … which they doon the devel sacrifise withinne that develes temple in cursed wise by superfluytee abhomynable." [10]. Changes over time of pronunciation can also be viewed in the often frustrating English spelling system, which tends to preserve earlier pronunciations.

Language also varies with geographical distance. However, especially in this age of communication, the distance is better described in terms of culture: those who speak frequently to each other tend to speak more alike. Thus, the difference between neighboring villages (or among members of a group of teenagers who hang out together) is typically less than that between groups who communicate less. Linguistic variability with time, geographic location, and cultural group, makes language boundaries very slippery.

Some dialect differences are in the vocabulary, *e.g*., British '*bonnet'* and American '*hood'* for the piece of a car covering the engine. But many of the differences are also in the way the words are pronounced. Linguists capture these pronunciation variations by using two levels in the sound system. The more concrete level of letter-sized pieces of sounds, the **phonetic** level, is determined by human speech production and perception constraints. Languages of the world can be transcribed at this level using, for example, the International Phonetic Alphabet (IPA) [11]. The more abstract unit of a letter-sized piece of sound, the **phonemic** level, is where a difference in sound marks a difference in meaning. This is the representation level used by most dictionaries. It ignores the fact that in casual speech *'probably'* might lose one or two syllables. It ignores the fact that '*bin'* as spoken in one part of the U.S. might rhyme with the word '*bean'* in another part.

Current speech technology typically does not represent dialect very well. A frequently used method is to gather data representing the demographics of the speakers who will use the system and statistically train the system on that data. This is suboptimal from a technical point of view because it tends to lose precision: the representation of the /iy/ in *'bean'*, for example, is fuzzier than it needs to be because it includes phonetically diverse realizations. This modeling may work well on average, but can be frustrating for those whose dialects may not have been the majority in the training data. A better technique from a linguistic point of view would be to model the linguistic system. That is, for a given speaker of English, /iy/ and /ih/ should contrast, and /iy/ should be produced higher and fronter in the mouth than /ih/, even though at the phonetic level [iy] could be phonemically either /iy/ or /ih/, depending on the speaker. For more on dialects of North American English, see [12].

C. Language and meaning

Language is not logical. We had to invent logic because natural language would not do. Natural language needs to be expressive. We need to convey not just concrete ideas, but more abstract ideas, and our feelings about those objects and ideas. Further, we need to adapt our language to our changing world and create new expressions to convey new meanings. Language often also needs to be vague or ambiguous (especially in diplomacy and for language play and poetry), but this property makes language frequently vague and ambiguous in normal use as well.

Language is also highly related to culture. We use language to mark what group we belong to, and language embodies the habit of thoughts of cultural groups. Concepts important in a given language may have many different words or expressions, with many fine distinctions that may not be easily conveyed in another language. This means that it is difficult or impossible to convey in a translation all and only what was conveyed in the original. This makes assessment of translation, or the representation or understanding of meaning more generally, an especially difficult task.

D. Children's speech

Speech produced by young children shows considerable variability. Children are, of course, smaller than adults, but their heads are relatively larger, compared to their height, and they have relatively larger mouth size compared to throat size. These differences have acoustic consequences, *e.g.*, [13]. Because children are still growing, their articulators, like their arms and legs, may be harder to coordinate. In many cases, parents will understand their child better than others because they are aware of the child's pronunciation and usage, and of changes as the child learns and grows. Parents may also often have a better sense of what the child is likely to say.

Children still learning to perceive and produce differences important in their language may have individual pronunciation difficulties, such as a lisp, or inability to properly make the 'r' sound. They typically lose their front teeth at about age 6, which also has acoustic consequences. In sum, child speech is typically more variable than adult speech because children are learning language rapidly, growing rapidly, learning fine motor skills, and becoming more familiar with and competent in producing the sounds, words and grammatical constructions of languages they are learning.

E. Written vs.spoken language

The linguistic level represented by a writing system varies across languages. Some use symbols to represent words or **morphemes** (units of meaning that might be smaller than a word, such as *–ful* or *hope* in the word *hopeful*). In these languages, there are many thousands of symbols to learn. Some languages use symbols that represent syllables. There are far fewer of these symbols to learn, but far more than for those languages that use the phonemic level of representation, as in English and other alphabetic systems.

The difficulties learners might face in learning these various conventions vary consequently from language to language. English, for example, has far more complicated letter to sound rules than, *e.g.*, Spanish. Spanish for the most part has a nearly one to one mapping of letters to sound: if you can pronounce it, you can spell it. French is more complicated, but mostly, once you learn the rules, you will know how to pronounce it given the spelling. The reverse, however, is far from true, and French is one of the richest languages in homophones (words that sound the same but are spelled differently), *e.g*., 'fois', 'foie', 'foi' all pronounced /f w aa/.

English presents difficulties both in letter to sound and in sound to letter rules. One sound can be represented by more than one letter, or sets of letters, and a letter or set of letters can be pronounced in more than one way. English has about 40 phonemes, but uses only 26 letters. Some phonemes are represented by a combination of letters, *e.g.*, 'th', 'sh', 'ee'. English uses the single letter 'x' to represent a sequence of sounds /k s/ ('fox' rhymes with 'socks'). English can also embed some morphological information in the spelling, *e.g.,* the morpheme for past tense '-ed' is spelled the same despite its different pronunciations as /t/, /d/ and /ax d/ respectively in 'ached', 'aimed' and 'painted'.

English is one of the most complex spelling systems, since the move to the written conventionalized forms happened long ago, there have been many borrowings from many languages (preserving some aspects of the spellings from those languages of origins), changes in pronunciations without changing the spelling, and there have not been any recent major spelling reforms. English is far from a one-to-one letter to sound system. Of particular note is that the writing system tends to better represent some dialects than others, making learning to read somewhat more difficult for speakers of less prestigious dialects, a phenomenon usually complicated further by socioeconomic differences. For example, a child who pronounces 'old' /ow l d/ will find the word easier to learn than a child who pronounces it /ow l/.

Normal spoken vocabularies are not nearly as rich as the vocabulary needed to understand most texts. The grammar for the very linear and ephemeral spoken form is also typically much simpler than for the written form. In reading one can see the whole sentence as a piece, and go back and reread it easily. This is part of why a technical paper that is read from the written form can be so difficult to understand. Another reason for the difficulty, which is also another difference between written and spoken language, involves prosody. Prosody consists of those parts of oral language that go beyond the words spoken, that is, not what you say, but how you say it. Part of prosody's job is to group together or separate sets of words, a role approximated by punctuation. Prosody also is used to emphasize some words relative to others, approximated in text by underlining or capitalization. Prosody is also used to convey moods, such as sarcasm and other attitudes, which in casual forms of written language can be approximated by 'emoticons' *e.g*., ☺

F. Reading pedagogy

Although there are many differences between written and spoken language, once reading is learned well the concrete image of the written word becomes so ingrained that it is hard to imagine the situation we all faced in our preliterate days with only ephemeral spoken words to hang meaning onto. Many people are shocked to learn that the acoustic record has no boundaries between words that correspond to the spaces between words in the written form.

For most people, learning to read means learning the conventions of the mapping between an oral version of a language and a written one. The reading process involves language skills (meaning-based skills such as vocabulary and grammar, both also important in spoken forms) and decoding skills (learning the mapping between a set of symbols and a set of sounds). Vocabulary and grammar become reading issues because of the differences between written forms and oral forms described earlier.

Learning to read a language whose written form represents morphemes or words means that each new word must be learned essentially one at a time. Presumably if the word is known, the pronunciation is known, though this is not necessarily true. In languages where the symbols represent sounds rather than meaning (as in syllabaries and alphabets), finding the meaning of the written word can be mediated by sounding out the word in order to recognize it. Sounding out words works better, of course, if the letter to sound rules map one letter to one sound and if the language has no homophones. This is rarely true.

The politics and practice of reading pedagogy in the United States has tended to emphasize cyclically either decoding ('phonics') or meaning ('whole words'). In fact, of course, English learners need to learn both. Because of the complexity of English spelling, particularly for high frequency words like 'the' and 'of', learning to see the word as itself a unit of meaning works better than trying to sound it out --- letter-to-sound rules used for these words would apply to few, if any, other words. However, learning phonics well can be extremely important when encountering new words.

Learning to read a language while learning it poses different problems for children compared to literate adults. For adults, the written form makes things more concrete and separates words. Further, resemblances to languages the adult may already know can be made more apparent in this form. However, this focus on the written form can make it difficult to learn the pronunciation of the new language since previously known letter to sound rules are often used rather than those for the new language.

Children not literate in any language face a different set of issues when learning to read a new language. They must learn the concept of literacy, which may be difficult if many of the words used to illustrate a particular letter to sound rule are not known by the child. This can induce a tendency to learn reading as only decoding, and particular care must be taken to insure that meanings are also attached to the words learned by the child. Otherwise, a child might look like a good reader on many tests, by reading fluently, though run into severe problems later when more complex understanding is reaquired. It appears that, at least for native speakers of Spanish learning English, it is on average better to first learn to read in Spanish before learning to read English, *e.g.*, [14].

Vocabulary and grammar development arise from seeing words and constructs in context. Children who experience less oral and written language will lag behind those who experience more. Socioeconomic factors seem to be correlated with how much a child is spoken to, is read to, and reads, as well as with how close the native dialect may be to the written form. This can mean a huge difference in vocabulary, grammar, and general experience with the conventions of the written form well before pre-school. Because of these factors and because of language learning issues, cognitive levels and age-related interests may differ greatly from the reading level. An adult language learner, for example, is probably much less interested in reading "See Puff run" than a child of 6 might be. Key in learning content, grammar and vocabulary from what we read is finding material of interest at the right level; if it is too easy nothing may be learned because it is already known, and if it is too hard, nothing may be learned because there is not enough already known to build on. High interest content means that the reader is more likely to spend more time 'exercising the reading muscle.' For these reasons, language technologies to sort documents into reading levels by topic are very important. Work by Peterson and Ostendorf, for example, is showing encouraging results in this area, *e.g.*, [15].

G. *Multimedia and pedagogy*

Multimedia technologies offer pedagogical advantages, see, *e.g.*, [7, 8, 16]. In particular:

- Mulitmedia can encourage active direct, as opposed to passive, learning, which tends to lead to more learning.
- The redundancy of the multi- in multimedia means that the same information can be presented in more than one way, which can accommodate more learning styles.
- Accessibility with differing abilities and preferences is afforded; *e.g.,* those who are blind or using their eyes for another task can use a medium other than visual.

A generation increasingly experienced in computer interaction and in mobility is rapidly demanding, and creating, more of the same. Universities are offering courses in game design, including educational games [16]. While it is true that computers and technology alone won't improve education, this is also true of books and of teachers. Some are better than others at what they do. However, multimedia can be replicated and distributed more widely than can a good teacher, and can provide some direct experience of the world through simulations that books cannot.

III. Examples from the TBALL project

Some of the reading, language and pedagogy issues discussed above can be illustrated by a multimedia research project: TBALL (Technology Based Assessment of Language and Literacy) [17].

A. *Dialect, idiolect and. reading issues*

As mentioned, dialects can vary considerably with geography, socioeconomic factors, native language of the speaker, culture, age, *etc.* Other sources of pronunciation variability are more personal, such as a child's lisp or substitution of /w/ for /r/ in English. If we are to assess whether a word has been read accurately or not, we cannot go by the acoustic record alone.

If a child says 'cow' when reading 'car', we don't know if it is a misreading or if that's the way the child pronounces the word. Similarly, because in some dialects 'ask' is pronounced

/ae k s/, when a child reads it that way, we do not know if it is diagnostic of a letter reversal problem, or simply the way the child's community pronounces the word. As another example, a child could read the word 'yes' as /jh eh s/ by applying Spanish letter to sound rules to English, or because of hearing English predominantly in a Hispanic accented form. A teacher may or may not want to count any or all of these as 'incorrect'. If the test is whether or not the child can read a word correctly in the school standard dialect, they may be all counted as incorrect. If the test is whether the child has correctly found the word indicated by the letters, they may all be counted as correct. Whatever the test, for incorrect responses, teachers are likely to want to diagnose the error.

To try to provide such diagnostic information by separating dialect and idiolect issues from reading errors, we have used several techniques. A large set of data was collected from over 250 children in kindergarten through the 4th grade from the Los Angeles Unified School District. The data represents children come from diverse socio-economic backgrounds and include many language learners. Included are picture naming tasks in which pictures were selected to represent objects known by most children and instances of the phonemes of English. These data with demographic information can be used to calibrate/approximate the child's dialect and changes in it as the child develops.

Knowing the phonology and phonetics of Spanish compared to English can be used to augment standard pronunciation dictionaries with pronunciation variants using various 'tags' such as 'Hispanic accented', 'misarticulated' or 'Hispanic letter to sound rules' [18]. A child who has only a few pronunciations matching best with one of these tag-sets may simply be making reading errors. However, a child whose pronunciations match most often with one of the tag-sets may be diagnosed with that tagged issue, Hispanic accent, lisp, *etc.*

B. Assessment of comprehension: overview

Automation of comprehension assessment is particularly difficult, as noted earlier concerning the representation of meaning. Many comprehension assessments use very open ended questions, such as "What happened in what you just read?" Not only are such assessments difficult to automate, they confound comprehension with other factors: a child may parrot without understanding many of the phrases just read, and a child who understood well may not remember it all or may give just the gist of the passage rather than as many details as can be remembered.

More directed questions, such as "What did they do after school?" may have as a scoring protocol something like; "They went home, or they went to bed, or any other justifiable answer." This leaves the scoring quite open to interpretation and affected by the rapport between the tester and the child, how long the child is given to respond, the encouragement given, *etc.* A further common problem is that for incorrect answers, it is not clear if it was the target material not understood or the question about it.

Even yes/no questions are often not strictly answered in the text, *e.g.*, "Will Kim write a song?" requires hypothesizing about what will happen next. A child could have good comprehension and get this "wrong." The scientist as a child may think that because it was not said what they will do, and the children were only discussing writing a song, it does not mean that they will do it. A child might not be confident enough to say "We don't know" and may just guess[1].

C. Assessment of Comprehension: the BARLA

For both reading and oral comprehension, we want to separate the understanding of the assessment material from the understanding of the questions asked about it. Most young native speakers can understand materials at a higher level than they can read because they are still learning basic reading skill. A contributing factor is that the prosody can assist in understanding and is only poorly represented in orthography. Finally, it may be that more 'cognitive cycles' are available to assemble a meaning if they are not needed for decoding. The advantage of listening over reading comprehension may not always hold, however, as many adult language learners have experienced: written words are concrete and there is less of a memory load since the words don't go away as they do once spoken. Therefore, some language learners may find the written form easier than the oral form.

To tease out these factors, we sought an assessment that was 'comparable' across the oral and written versions and found none that specifically did this. We therefore developed the BARLA (Bay Area Reading and Language Assessment) to try to avoid common pitfalls and to separate language understanding (*e.g.,* vocabulary, syntax, morphology, needed in language understanding generally) from decoding skills (required for written but not language).

The comprehension assessment that came closest to meeting our needs was the PIAT-R (Peabody Individual Achievement Test-Revised [19]. The PIAT-R protocol has all images as responses, which helps separate the understanding of the response from that of the test material. Although the test is not intended to be an oral test, it lends itself easily to this use: The child points to the image that best represents the sentence heard rather than read. Analogous to the pronunciation tags, we can hang diagnostic tags on the 3 incorrect responses used as foils, *e.g.*, 'rhyme', 'same initial letters/sounds', 'semantically related'. Any one error may be by chance, but a pattern of errors may be useful diagnostically.

There are enough materials per grade (about 10) that alternate sentences can be used for oral *vs.* written comprehension in a counter-balanced design. The images and sentences have extensive validation and reliability studies associated with them. Since we received no response to our request for permission to use the materials as proposed, we created test sentences and images ourselves. An advantage of creating these materials ourselves is that we can more easily select the diagnostic foils and create materials better suited for younger children. We created a new test by creating:

---

[1] Thanks to M. Callahan for this example. Thanks to the rest of the BARLA team at UCB (P. D. Pearson, M. Callahan, and T. Duong) for our joint work developing the BARLA.

- New test sentences with syntax similar to the PIAT-R substituting words in the same frequency of occurrence range based on TASA's Word Frequency Guide [20].
- New images to illustrate the new test sentences and foils.
- Oral readings for each of the target sentences, for the instructions and all prompts.
- Diagnostic tags for the 'foils'.
- An online automated version of the assessment that gets content from a database, keeps track of which child has performed which part (which stimuli and condition, listening or reading), calculates scores and diagnostics and writes them out to the database.

The computerized version makes the test easier and much more consistent (each child receives exactly the same instructions and timing protocols). It is also much easier to analyze. Output displays include both summary data and drill-downs to show more detail (item level responses, time spent on the item, and diagnostic tags for any incorrect answers). Pilot testing is under way.

IV. Trends, challenge, opportunities

An increasingly mobile population, coupled with trends in the size and weight of components is an opportunity for applications and devices to make information ubiquitously available and easily found. These same factors together with an increasing pace of change predict opportunities in ongoing just-in-time education. Standards for multimedia technologies are becoming increasing important, as seen, for example, in the activities in web3d, voicexml, vuid, nl, and w3c. Activity in w3c includes working groups on synchronization of multimedia, voice browsers, and the semantic web [21].

Although the bulk of this report has addressed multimedia in children's education, a growing trend in the U.S. and world wide is an aging population, arising from lower birthrates and longer life spans. The increased proportion of the aged in our population results in an increase in problems associated with aging: *e.g.* in strokes, reaction time losses, vision and hearing loss, memory degradation. These trends are also, of course, opportunities for multi-media technologies to address, as illustrated a bit in the overview of sample applications.

Although much of the developed world is aging on average, much of the developing world is increasing in population, a situation likely to lead to increased immigration as people in developing nations seek jobs and those in aging populations seek workers. Immigration most often means new languages in contact. This is an opportunity for multimedia technologies, particularly multimedia language technologies. Technology can help in teaching and testing language skills, collaboration among those who don't speak the same language, selection or creation of texts of most interest and best leveled for particular individuals. Multimedia technology can provide many of the advantages of direct experience together with the advantages of ease in reproducing, retrieving and sharing information. Perhaps it can lead to another great leap by increasing our individual and collective potential.

References

[1] U.S. Census Bureau, "A half-century of learning: historical statistics on educational attainment in the United States, 1940 to 2000" PHC-T-41, 2002 census.gov/population/www/socdemo/education/phct41.html

[2] UNESCO, portal.unesco.org/education/en/ev.php-URL_ID=32635&URL_DO=DO_TOPIC&URL_SECTION=201.html

[3] J.Diamond, *The Third Chimpanzee: The Evolution and Future of the Human Animal*, New York: Harper Perennial, 1992.

[4] S. Basson, P. Fairweather, V. Hanson, "Speech recognition and alternative interfaces for older users", *Trans. ACM Interactins: Special Issue on Technology and Aging*, in press, July 2007.

[5] D. Roberts and D. Topolewksi, "Evaluation of the English Xchange spoken English teaching program," unpublished.

[6] Bodnar, A., Corbett, R., Nekrasovski, D., "AROMA: Ambient awaReness through Olfaction in a Messaging Application", *Proc. Int. Conf. on Multimodal Interfaces*, 2004. cs.ubc.ca/~rcorbett/p174-Bodnar.pdf

[7] L. Chittaro, R. Ranon, "Web3D Technologies in Learning, Education and Training: Motivations, Issues, Opportunities," *Computers & Education Journal*, 49:1, 3-18, August 2007.

hcilab.uniud.it/publications/2007-03.html

[8] J. Gee, *What Video Games Have to Teach Us about Learning and Literacy,* Palgrave Macmillan, 2003.

[9] R. Eastgate, G. Griffiths, P. Waddingham, A. Moody, T. Butler, S. Cobb, I. Comaish, S. Haworth, R. Gregson, I. Ash and S. Brown, "Modified virtual reality technology for treatment of amblyopia," *Eye* 20, 370–374. 2006 doi:10.1038/sj.eye.6701882 nature.com/eye/journal/v20/n3/abs/6701882a.html

[10] This is the first sentence from Chaucer's "The Pardoner's Tale". www.courses.fas.harvard.edu/~chaucer/teachslf/pard-par.htm#Introduction

[11] www.arts.gla.ac.uk/IPA/

[12] W. Labov, S.Ash, and C. Boberg, *Atlas of North American English: Phonetics, Phonology and Sound Change*, Mouton/de Gruyter, 2006. Also useful, and by the same authors, is some earlier data: www.ling.upenn.edu/phono_atlas/NationalMap/NationalMap.html

[13] U. Goldstein, *An Articulatory Model for the Vocal Tracts of Growing Children*, MIT PhD dissertation, 1980.

[14] E. Hiebert, P. D. Pearson, V. Richardson, and S. Paris, *Every Child a Reader: Applying Reading Research to the Classroom,* Center for the Improvement of Early Reading Development, Ann Arbor, MI, U. Mich. School of Education, 1998.

[15] S. Peterson, M. Ostendorf, "Assessing the reading level of web pages," Proc. Interspeech, 833-836, 2006.

[16] C. Aldrich, "Simulation philosophies: Researching, teaching and learning in a post-linear world," *Conduit*, Brown University, 8 – 15, Fall/Winter 2006.

www.cs.brown.edu/publications/conduit/conduit_v15n2.pdf

[17] A. Alwan, Y. Bai, M. Black, L. Casey, M. Gerosa, M. Heritage, M. Iseli, B. Jones, A. Kazemzadeh, S. Lee, S. Narayanan, P. Price, J. Tepperman, and S. Wan, "A system for technology-based assessment of language and literacy in young children: the role of multiple information sources," Proc. Interspeech, this volume, 2007.

[18] S. Wang, A. Alwan, and P. Price, "Automatic evaluation of children's performance on English syllable blending task," *SLATE,* in press.

[19] ags.pearsonassessments.com/assessments/bibliography/piat.asp

[20] tasaliteracy.com/wfg/wfg-main.html

[21] www.w3.org/Consortium/activities.html