

**SPEECH RECOGNITION SOFTWARE:
ITS POSSIBLE IMPACT ON THE LANGUAGE LEARNING CLASSROOM**

by **Gina Mikel Petrie**

Washington State University,

Pullman, Washington, USA

<http://www.wsu.edu/~gmpetrie>

gina_wsu@yahoo.com

A little girl picks up a ringing telephone and says, "Hello?" Three businessmen are seen and heard on the other end of the phonenumber speaking Japanese. At the same time, the sounds coming out of the phone the girl is holding are recognizably English. The little girl leans away from the phone and asks her father in English the question that the men are asking. He yells the answer from another room, she relays it in English; her answer is heard by the men in Japanese. The men happily end the conversation and hang up.

Description

This scenario is based on a recent U.S. commercial for a communications company. The technology being demonstrated is speech recognition software and accompanying translation technology. Speech recognition is often confused with speech synthesis and voice recognition. *Speech recognition* allows people to talk to computers, and then the computers do something with the uttered speech. Either the computer types the utterance, carries out a command that was given with the utterance, or carries out an analysis of the utterance. *Speech synthesis*, on the other hand, allows computers to talk to people. *Voice recognition* allows computers to identify the identity of a speaker from their voice and then carry out a task such as allowing (or disallowing) entry into a building based on the clearance granted to that person.

Speech recognition technology works in the following way: the user speaks into a microphone, and a computer uses acoustic analysis to analyze the phonemes (individual sounds) uttered. The computer searches the available vocabulary database and then chooses the words that seem most likely to have been produced. Accuracy increases under the following circumstances: words are spoken slowly and individually, there is a small range of vocabulary possible, low background

noise exists, repetition exists, and/or the computer is familiar with the speaker's voice. Speech recognition accuracy can reach 99 % if these conditions exist; 87 % is the best that can be done without these aids (Ordinate, 2002)

History

Speech recognition technology has had an interesting history. According to Christensen, Maurer, Miranda and Vanlandingham (2002), the first speech recognition product that was ever offered on the commercial market was actually a toy dog. When the dog's name, "Rex," was uttered, the acoustic energy of the vowel sound broke an electromagnetic field and caused the dog to come out of his house. During the 1940's the U.S. Department of Defense searched for a way to automatically translate messages sent in Russian into English. Although the program was a failure, the government did go on to fund more successful research in speech recognition as a result. Bell Laboratories experienced early success with speech recognition technology, in 1952 producing a system that could recognize the numbers 0 through 9 and then in 1959 a system that could recognize English vowel sounds with 93% accuracy. Today's technology has progressed greatly as it has been possible to handle increasingly varied vocabularies, dialects and rates of speech - the keys to future progress (Kewley-Port, 1994). (For more specific information about the development of speech recognition technology, visit <http://cslu.cse.ogi.edu/HLTsurvey/ch1node4.html>) However, the technology needed to carry out the task in the opening scenario above has not yet been developed.

Social Context

In the consumer market, most of us have encountered speech recognition technologies on the telephone when utilizing directory assistance. Several telephone companies use a speech recognition server that recognizes the names of cities uttered by customers, and then connects those customers with the correct operator. (For an audio demonstration of this type of application, visit <http://www.nsc.co.il/>). Those working in the medical field utilize speech recognition software for medical dictation rather than relying on sending out tapes to transcriptionists, a process which can take days and several drafts to eliminate errors. Many people who are unable to use a keyboard due to disabilities are able to enter data or surf the Web with the assistance of speech recognition technology. This technology entered the military landscape recently when a hand-held device, the Phraselator, was used by U.S. troops in Afghanistan and then again in Iraq (Mieszkowski, 2003; Terry, 2002). The device allowed the

soldiers' spoken English to be heard as simple Arabic phrases. Two online articles report on this at <http://www.washingtonpost.com/ac2/wp-dyn?pagename=article&node=&contentId=A58740-2002Apr16¬Found=true> as well as at http://www.salon.com/tech/feature/2003/04/07/phraselator/index_np.html.

Educational Context

Since technologies usually find their way from the consumer market to the educational arena, it is worth noting any developing technology for its inevitable impact on education. Speech recognition technology most often shows up in schools as an assistive device for students with disabilities. Two commonly used programs are ViaVoice Pro USB Edition (2003) by IBM (<http://www-3.ibm.com/software/speech/>) and Naturally Speaking Preferred 7.0 (2003) by Dragon Systems (<http://www.1st-dragon.com/dragonspeak.html>). (For an evaluation of ViaVoice and Naturally Speaking, visit <http://www.webreference.com/new/991108.html>).

In addition, some schools are beginning to use speech recognizers to assist students as they read aloud. Videos describing Carnegie Mellon University's Project LISTEN (Literacy Innovation that Speech Technology Enables) are available at <http://www-2.cs.cmu.edu/~listen/mm.html>. Problems encountered by schools adopting speech recognition software include inadequate hardware and a lack of staff training (British Educational Communication and Technology Agency, 2001). To read more about these problems and one company's answer to them, visit <http://www.becta.org.uk/technology/speechrecog/information/software2.html>. The CALL (Communication Aids for Language and Learning) Centre in Scotland maintains a website with training materials, curriculum ideas and useful links at http://callcentre.education.ed.ac.uk/SEN/5-14/Special_Acc_FFA/Speech_Recog_FFB/speech_recog_ffb.html#Resources.

Across student populations, speech recognition technology that may hold the most promise for those learning or needing to communicate between languages. This promise makes itself evident, for example, in the recent television commercial described earlier. How might this technology affect the language learning classroom?

Language Learning Context

Speech recognition software has already begun to make an impact on language learning. One example is that of language testing or grading. In an intersection between psychology and

linguistics, Ordinate carried out research on how native speakers of English rate the understandability of non-native speakers of English and then utilized speech recognition software to create a test in which a non-native speaker of English places a phone call to the Ordinate testing number, listens to prompts in English, answers the questions in English and receives a rating from the software on fluency, listening, vocabulary and pronunciation. (A demonstration of this test is available at <http://www.ordinate.com>.) Interestingly, Ordinate claims to have higher accuracy at judging non-natives' speaking abilities than that arrived at by human raters (Ordinate, 2002).

Another educational application is that of pronunciation training for the profoundly deaf. Projects such as the Tucker-Mason Project, which is supported by a National Science Foundation grant, involve the creation of software that allows deaf users to give oral commands to the computer (Center for Spoken Language Understanding, 2002). For a description of these speech recognition applications, visit <http://cslu.cse.ogi.edu/asr/>. If a minimum level of understandability is not reached, the computer will not carry out a command. It is worth noting that rather than being focused on accuracy of language use, such applications appear to hold communicative competence as their goal.

Currently, a few educational software packages for English language learners take advantage of speech recognition technology. DynEd has produced *New Dynamic English* (2001) for adult learners (<http://www.101language.com/dyned-nde.html>) and *Let's Go* (2001) for child learners (<http://www.esl.net/dyned-lgfeatures.html>). The children's version allows the user to orally produce a single word at a time, while the adult version allows the user to produce either a single word or an entire sentence in response to video or graphic cues and then receive feedback on the pronunciation of the user's production. If a minimum level of understandability is not reached, the program encourages the user to try again. One current drawback of *New Dynamic English* is that if the uttered sentence is very close in sound to the intended answer, the program may not catch an error. For example, if the learner uttered a sentence with "is" instead of "isn't" - a serious difference in meaning - the learner may not be alerted of the difference. Auralog has also developed programs utilizing speech recognition: *TeLL me More Pro* (2000) for adults (<http://www.multilingualbooks.com/aura-tellp.html>) and *TeLL me More Kids* (2000) for children (<http://multilingualbooks.com/aura-tellk.html>). The minimum level of understandability can be adjusted for each student with these programs. In addition, *TeLL me More Pro* allows the user to

view the acoustic patterns of an utterance. However, there are two problems with offering learners acoustic patterns as evidence of their pronunciation ability. First, most language learners are not linguists, and a linguistic background is practically necessary in order to understand these wave forms. Second, even native speakers have difficulty reproducing the exact wave forms produced by the speakers on the software.

One possible application of speech recognition software for beginning language learners is that of a scaffolding device for building literacy. If learners are able to produce spoken English much more readily than they are able to produce written English, it might be useful for them to bridge into writing by, for example, telling stories to the computer and then seeing their own stories in print. The problem with this scenario is that the usefulness of such a tool would probably be shortlived in terms of the learners' need for this literacy assistance, yet a program such as ViaVoice, which takes only minutes for a native speaker to train it to his or her voice, might take many hours to adapt to the non-native speakers' voices and thus accurately type the words spoken. This would most likely put an added burden on the teacher, as well, whose efforts might be better spent on other literacy-building activities.

One issue that instructors of adult English language learners often grapple with is that of the special spelling problems of students who speak either Arabic or Hebrew as a first language. Since neither of these languages usually includes vowel sounds in writing, students often face seemingly insurmountable spelling issues in English; words are often written with such unusual spellings that even spell checkers cannot locate the correct words. Speech recognition software would allow these students to sidestep this serious writing issue. Once again, the time that it takes the technology to adapt to a non-native user's voice is an issue here, although less so than with a child learner. Also, this technology might actually step in the way of a learner ultimately improving spelling problems; rather than utilizing the tool as a scaffolding device, a learner could become dependent upon the tool.

Speech recognition software shows promise for assisting language learners with pronunciation issues. Pronunciation is an area that few language teachers have expertise in, yet many learners need or demand assistance with in order to gain communicative competence. Although quality pronunciation training following from the most recent research would be optimal, software utilizing this technology may be able to help learners understand when they have reached a level

of general understandability, especially as this technology continues to improve in its ability to respond to learners' utterances.

Referring back to the example at the beginning of this paper, although it is most likely far into the future, speech recognition software with accompanying translation technology might allow those with little or no speaking ability in a foreign language to carry on conversations via telephone with speakers of that language. For example, a middle-school EFL class in Hong Kong could brainstorm questions that they have about some aspect of British culture, arrange for a phone conference with a native of England, plan out what they are able to say in English, and then let the translation software pick up where the learners' abilities to speak and understand English break down.

Deeper Issues

In Fabos' (2001) study, "Media in the Classroom: An Alternative History," Fabos stated that although all new technologies in the classroom over the last century have been greeted with the same initial enthusiasm and hope that the technology would be able to solve administrative problems and enhance the teaching process, these technologies have eventually been rejected to some degree by teachers. Fabos suggested that the problem has often been the content that consciously or unconsciously enters the classroom along with the medium. Whenever a technology is brought into a learning environment, it always creates a *slightly different* learning environment, although the differences may be difficult to discern at first (Postman, 1992). So, how might our utilizing of speech recognition software with language learners influence our classrooms? What would we (possibly unknowingly) be teaching our learners about the world, about language and about communication with others?

The use of speech recognition technology in combination with software that includes role plays based on authentic situations would teach our students that oral interactions with others is the goal of language learning and that pronunciation is one aspect of communicative competence. The use of this technology to assist those who have problems with writing would teach that we are able to access our strengths in language learning to assist with our weaknesses. It might, however, also teach learners that they can rely on their strengths without having to improve the areas that most challenge them.

By using the technology as a translating device, we would be giving many messages to our students: that language learning is not essential and that communication is simply a matter of

translating vocabulary items and grammar. Monke (2001) asked in response to educational choices such as this one:

Just how small do we want our children to believe the world to be? How much of the illusion of next-dooriness do we want to give a student who hasn't traveled much beyond the borders of his or her state, or city for that matter? What kinds of misunderstandings about the world does this kind of undifferentiated communication give a young person? (Monke, 2001: 66)

Mastering a second or foreign language is a huge task; successfully negotiating meaning with native speakers is an enormous accomplishment. By utilizing speech recognition technology in ways such as this, we may be obscuring this reality from our students.

In addition, if technology reaches a point at which we no longer need to learn a second or foreign language in order to communicate with others, we need to rethink our reasons for acquiring another language. Research has pointed towards a link between language learning and cognitive development. Although some researchers caution against drawing strong conclusions about a causal link, there does seem to be a positive relationship between bilingualism and linguistic, metalinguistic and cognitive abilities which reach far into other areas of the language learners' lives (Diaz, 1985; Hakuta, Ferdman & Diaz, 1986). Any such gains from language learning could be lost, however, if the government no longer sees a need to fund programs for foreign language teaching or for language minority students due to advanced speech recognition and translation technology.

Since this technology is fairly inexpensive and could potentially be adopted by many intensive English programs as pronunciation aids, for example, the use of this tool may hinder a recently-improved aspect of M.A. TESOL programs. In the early and mid-nineties, few M.A. TESOL programs trained pre-service teachers in pronunciation issues. However, in the last five years, such preparation has become more widespread. Although software can never replace the role of the teacher in pronunciation training, it may be viewed as capable of this. Once again this aspect of communicative competence may no longer be covered for Master's degree students.

Salaberry (2001) suggested that we express cautious and reflective interest in new technologies rather than an overly enthusiastic attitude. Many of the issues raised above point towards the need for much consideration of the impact that speech recognition technology might have on the language learning classroom. Readers are encouraged to critically explore the possibilities and implications of speech recognition themselves by downloading some examples of current

technologies. Several examples can be found at
<http://www.speechtechnology.com/free/links.html>.

References

- British Educational Communications and Technology Agency. (2001). *Speech recognition: Information and advice*. Retrieved June 9, 2003, from <http://www.becta.org.uk/technology/speechrecog/information/software2.html>
- Carnegie Mellon University. (2003). *Project LISTEN videos*. Retrieved June 7, 2003, from Project Listen (Literacy Innovation that Speech Technology Enables), <http://www-2.cs.cmu.edu/~listen/mm.html>.
- Communication Aids for Language and Learning (CALL) Centre. (2001) *Resources to view or download*. Retrieved June 7, 2003, from http://callcentre.education.ed.ac.uk/SEN/5-14/Special_Acc_FFA/Speech_Recog_FFB/speech_recog_ffb.html#Resources
- Center for Spoken Language Understanding. (2002). *Automatic speech recognition at CSLU*. Retrieved June 9, 2003, from <http://cslu.cse.ogi.edu/asr/>.
- Christensen, B., Maurer, J., Miranda, N., Vanlandingham, E. (2002). *Accessing the internet via the human voice*. Retrieved January 16, 2003, from <http://www.stanford.edu/~jmaurer/homepage.htm>
- Diaz, R.M. (1985). "The intellectual power of bilingualism." ERIC Document Reproduction Service No. ED 283368.
- Fabos, B. (2001). "Media in the classroom: An alternative history." *Proceedings of the American Educators: Research Association*. Seattle, WA.
- Hakuta, K., Ferdman, B.M., Diaz, R.M. (1986). *Bilingualism and cognitive development: Three perspectives and methodological implications*. Los Angeles: Center for Language Education and Research.
- Kewley-Port, D. (1994). "Speech recognition". In A. Syrdal, R. Bennet & S. Greenspan (Eds.), *Applied Speech Technology*. Ann Arbor, MI: CRC Press.
- Mieszkowski, K. (2003). How do you say "regime change" in Arabic? *Salon.com*. Retrieved June 7, 2003, from http://www.salon.com/tech/feature/2003/04/07/phraselator/index_np.html.
- Monke, L. Burniske, R.W. (2001). *Breaking down the digital walls: Teaching in a post-modem world*. Albany: State University of New York.
- Natural Speech Communication. (2002). *NSC speech recognition demo*. Retrieved June 6, 2003, from NSC Website: <http://www.nsc.co.il/>
- New World Creations. (2002). *Free voice recognition software*. Retrieved June 6, 2003, from SpeechTechnology.com Website: <http://www.speechtechnology.com/free/links.html>
- Ordinate. (2002). *Set 10 Demo Test*. Retrieved June 9, 2003, from Ordinate Website: <http://www.ordinate.com>
- Postman, N. (1992). *Technopoly*. New York: Random House.
- Salaberry, M. R. (2001). "The use of technology for second language learning and teaching: A retrospective." *Modern Language Journal*, 85(1), 39-56.

Terry, R. (2002, April 16). The Phraselator: Translation system put to the test in Afghanistan. *Washington Post*.

Retrieved June 7, 2003, from <http://www.washingtonpost.com/ac2/wp-dyn?pagename=article&node=&contentId=A58740-2002Apr16¬Found=true>.

Zue, V., Cole, R., & Ward, W. (1996). *1.2: Speech recognition*. Retrieved June 6, 2003, from the Survey of the State of the Art in Human Language Technology Website: <http://cslu.cse.ogi.edu/HLTsurvey/ch1node4.html>.

Software references

Let's Go: English Language Learning. (2001). Burlingame, CA: DynEd International, Inc.

Naturally Speaking Preferred 4.0. (2001). Hereford, UK: Dragon Systems.

New Dynamic English. (2001). Burlingame, CA: DynEd International, Inc.

TeLL me More Kids. (2000). Tempe, AZ: Auralog, Inc.

TeLL me More Pro. (2000). Tempe, AZ: Auralog, Inc.

ViaVoice Pro Millenium Edition. (2001). Kansas City, MO: IBM.