R & D Connections

Can Speech Technology Improve Assessment and Learning? New Capabilities May Facilitate Assessment Innovations

Isaac I. Bejar

The need for a highly skilled and literate population is increasing the demand for tools to monitor and measure students' learning progress at all age levels. Can speech technology help to meet this unprecedented demand?

A confluence of developments in education worldwide suggests that *speech technology* could help to address challenges such as the development of literacy, especially reading proficiency, and the acquisition of communicative competence in English. In this article, I discuss the use of computers to recognize and synthesize speech (i.e., speech technology), their educational applications, how they may help with pressing assessment and instructional needs, and, finally, describe ETS research in this area.

Technology's Role

To see examples of how speech technology can enable solutions to broader educational problems, it is useful to consider successful applications of other technologies for educational purposes. Many such examples exist, including adaptive testing and automated scoring (Wainer, 2000; Williamson, Mislevy, & Bejar, 2006). Demonstrations that technology can promote learning on a large scale are more difficult to find, and skeptics abound (Cuban, 2001). Nevertheless, the hope remains that technology, in concert with the application of knowledge about how students learn, is bound to inevitably succeed in markedly improving student achievement at some point (Bennett, 2002; Quellmalz & Pellegrino, 2009).

Writing assessment is a good case in point. There is reason to believe — and empirical evidence to support this belief — that students can become better writers when they use the computer for writing (Goldberg, Russell, & Cook, 2003). How can technology have such an effect? By creating more frequent opportunities for students to learn, perhaps. Having student writing in digital form makes it possible to analyze writing quality in more detail, grade the writing by automated means, and provide immediate feedback to both the student and the teacher about how well the student performed (Miller, 2009).

A digital writing environment also can provide students with tools or *scaffolds* (Deane, Quinlan, & Kostin, in press) that can facilitate writing. The feasibility of detailed writing analysis also makes it possible to study the development of writing skills and to chart their development on a meaningful scale (Attali & Powers, 2008).

Editor's note: Isaac Bejar is a Principal Research Scientist in the Research Applications & Development area of ETS's Research & Development division.

> *"Speech technology could help to address challenges such as the development of literacy, especially reading proficiency, and the acquisition of communicative competence in English."*

"Technology could have a positive effect by creating more frequent opportunities for students to learn."

"The example provided by the application of technology to writing highlights one aspect of using technology well in an educational setting: Technology can increase efficiency, and therefore optimize results for a given level of resources without necessarily compromising educational objectives."

Finally, the availability of student work in digital form can liberate teacher time that otherwise would be devoted to scoring (Burstein, Chodorow, & Leacock, 2004), and perhaps focus teachers' time more strategically on students who need additional feedback, which Meyers (2003) argues could enhance teacher professionalization. In short, the example provided by the application of technology to writing highlights one aspect of using technology well in an educational setting: Technology can increase efficiency, and therefore optimize results for a given level of resources without necessarily compromising educational objectives.

Speech technology also may have a role in addressing the challenges facing education this century, among them the challenge of literacy described in the ETS Policy Information Report *America's Perfect Storm* (Kirsch, Braun, Yamamoto, & Sum, 2007). The report notes, for example, that a growing portion of the U.S. population is lacking English-language literacy skills. Moreover, statistical trends suggest that this problem will grow and affect economic competitiveness unless remedial action is taken.

Speech technology may have a distinct role in addressing the acquisition of reading skills. Analysis of the learning-to-read process in children suggests that the more children read, the better readers they become. The reasons are many and are very complex, but according to Perfetti (2003, p. 17), they start with being able to recognize that "the printed form corresponds to words in the spoken language." That process takes place through practice because it presents "… opportunities to map spoken language to print and then to practice this mapping through reading" (Perfetti, 2003, p. 19). The inability to recognize the correspondence between spoken and printed language disrupts the reading process, which affects comprehension and, as a result, also disrupts the acquisition of new vocabulary, which further delays the improvement of reading comprehension.

Under ideal circumstances, parents or patient reading tutors are the ones who provide the supervised reading practice and instant feedback that seems essential to developing literacy. While not a substitute for this ideal teaching and learning situation, speech technology may help create additional teaching and learning opportunities by emulating the reading-aloud loop, where the student reads and gets feedback and help along the way. Just a few years ago, this possibility was not even mentioned in a review of the applications of technology to literacy (Rosen, 2000).

In the meantime, several computer-based reading tutors relying on speech recognition and oriented to promote reading in children have been developed (Gruenenberg, Katriel, Lai, & Feng, 2008; Williams, 2002). A recently completed evaluation of several off-the-shelf reading programs that rely on speech technology (Campuzano, Dynarsky, Agodini, & Rall, 2009) showed that for at least one such program, statistically significant effects with respect to standardized test scores were found. By far, the most extensive research effort is Project Listen[1] at Carnegie Mellon University. Among the project's recent findings are improvements in comprehension over a four-month period (Mostow et al., 2008), and improvements in reading fluency for elementary students whose first language is not English (Poulsen, Wiemer-Hastings, & Allbritton, 2007).

[1] http://www.cs.cmu.edu/~listen/

"Under ideal circumstances, parents or patient reading tutors are the ones who provide the supervised reading practice and instant feedback that seems essential to developing literacy. While not a substitute for this ideal teaching and learning situation, speech technology may help create additional teaching and learning opportunities."

In short, the available research suggests computer-based reading tutors that use speech technology, specifically speech recognition, can lead to increases in reading performance over a relatively short period of time. Such results must, of course, be evaluated carefully. Would they generalize to cases where researchers were not actively involved and assisting in the application of the technology? In other words, what is required in terms of technical support, teacher professional development, and, of course, support by school, district, and state leaders (Fishman, Gomez, & Soloway, 2001; Stites, 2004)?

English as a Second and Foreign Language

Speech technology also has important applications in the acquisition and evaluation of English as a foreign language. For most of the last century, the TOEFL® test and other tests for assessing English-language readiness did not include spoken skills on a large scale, and focused instead on grammar and vocabulary. The assessment of speaking skills was addressed, typically, by a separate test, such as the *Test of Spoken English*™, and was scored by human graders. However, the growing acceptance of *communicative competence* (Canale & Swain, 1980) as the goal of language instruction has had a significant impact on language-proficiency tests. The fact that spoken skills were not emphasized in language instruction and assessment was not a reflection of their lesser importance, but likely because they require one-on-one instruction and lots of practice. Those conditions are difficult to satisfy in a conventional classroom. Nevertheless, without well-developed speaking skills, it is hard to argue that a student has reached communicative competence in English.

When the revised TOEFL test appeared in 2005, it tested overall communicative competence and, as a result, incorporated a spoken component in addition to reading, listening, and writing (Chapelle, Enright, & Jamieson, 2007, p. 41). Adding the spoken component was feasible because the test was administered on computer, and therefore it was feasible to capture speaking samples economically, so that they could be scored by human graders.

The inclusion of speaking proficiency in the TOEFL test has led to a positive instance of the so-called *washback effect* (Messick, 1996), or the repercussions that testing speaking proficiency may have in the classroom (Chapelle et al., 2007, p. 42). In this case, the hope is that classroom instruction will emphasize speaking skills more, rather than exclusively emphasizing grammar and vocabulary, as has been typical in the past. There is some evidence that such washback is beginning to take place (Wall & Horák, 2008) now that the TOEFL test has been operational for a while. If the test does succeed in increasing the focus on oral skills as intended, technology-based tools could help instructors to meet the increased demand on classroom resources.

In short, the move to communicative competence and the use of computers for delivering the test made it possible to assess speaking proficiency. As a result, we can now expect that teachers will want to pay increased attention to developing students' speaking skills, which should create a demand for technology-based tools for accelerating the acquisition of speaking proficiency. Speech technology, of course, has a potential role both in the scoring of speaking samples (Xi, Higgins, Zechner, & Williamson, 2008; Zechner, Higgins, Xi, & Williamson, 2009) and in the development of speaking proficiency by means of instructional software.

How It Works

It is useful to consider the process of developing and refining speech-recognition applications. Far more than just a matter of writing software to program computers, the application development process itself leads to advances in understanding how to define proficiency in spoken language and factors that contribute to proficiency. Figure 1 schematically shows the process of applying a speech recognizer to measuring English speaking proficiency. However, as noted in Figure 1, the results also can be used to further refine the recognition engine. For example, as more data is collected, it becomes possible to refine the acoustic and language models to better reflect the target construct.

Figure 1: The Development and Use of Automated Speech Scoring

In assessment and learning research, technology development is an iterative process: Tools are based on foundational knowledge, but they also yield data that enhance that foundational knowledge.
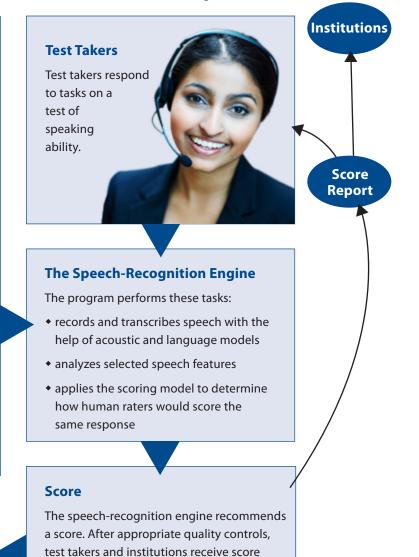
Foundational Research

Scientists compile a *speech corpus* — a database of recordings and speech transcripts that form the basis for the speech-recognition engine's components and its evaluation criteria. As part of this foundational work, scientists develop:

- *a construct definition* — an understanding of what it means for someone to be *proficient* in spoken language

- *a list of important speech features* — variables that are, according to the construct definition, strong indicators of proficiency, such as rate of speech, pronunciation quality, and number of fluency breaks

- *an acoustic model* — a software component that allows the engine to recognize words despite natural variations in the ways different people say them

- *a language model* — a software component that allows the engine to distinguish between similar sounding words (e.g., to know that the speaker meant "the red cup" and not "the read cup")

- *a scoring model* — a set of evaluation criteria based on scores that expert human raters assigned to responses similar to those that the speech engine will evaluate

Test Takers

Test takers respond to tasks on a test of speaking ability.

Institutions

Score Report

The Speech-Recognition Engine

The program performs these tasks:

- records and transcribes speech with the help of acoustic and language models

- analyzes selected speech features

- applies the scoring model to determine how human raters would score the same response

Score

The speech-recognition engine recommends a score. After appropriate quality controls, test takers and institutions receive score reports. Scoring data also informs further development of the engine.

Refinements

Over time, scientists use the data they gather from test takers to improve the foundational research behind the speech-recognition engine.

"One lesson we have learned over the many attempts to leverage technology in the classroom is that technology by itself is not the solution to educational problems. That is, technology should be seen as an enabler *and used in concert with knowledge about how students learn to read, how they acquire a second language, and, in general, how they learn."*

One lesson we have learned over the many attempts to leverage technology in the classroom is that technology by itself is not the solution to educational problems. That is, technology should be seen as an *enabler* and used in concert with knowledge about how students learn to read, how they acquire a second language, and, in general, how they learn. Equally important, just as teachers and other actors in the educational enterprises are key ingredients of any educational reform (Spillane, 2004), the acceptance of technology in the classroom cannot be expected without the cooperation and understanding of all the relevant players.

At ETS, researchers are actively pursuing the application of speech technology in the context of assessing and promoting reading and speaking skills. The most recent National Assessment of Adult Literacy Study (Baer, Kutner, & Sabatini, 2009) included reading fluency measures obtained through the application of speech technology. Researchers at ETS and elsewhere are also pursuing applications of speech technology to the process of learning to read (Zechner, Sabatini, & Lei, 2009). Similarly, researchers at ETS and elsewhere are actively pursuing the application of speech technology to the scoring of speaking samples (Xi et al., 2008) from English language learners. The speaking construct is very rich and includes fluency, which the current state of the art in speech technology handles well; it also includes the finer points of pronunciation, intonation, and stress, which currently are beyond the state of the art but are being actively researched (Lei, Zechner, & Xi, 2009) to gradually improve the assessment and acquisition of English speaking proficiency. Further components of the speaking construct are related to vocabulary use, grammatical accuracy and complexity, and aspects of content. However, in order to validly incorporate those components into the scoring of spoken responses, speech recognizers need to exhibit higher recognition accuracy, which is a significant challenge given the nature of the speech being produced by non-native speakers of potentially many different native languages and proficiency levels. Nevertheless, research at ETS is targeting such challenges (Zechner et al., 2009).

Conclusion

The development of speech technology during the 20th century was a major scientific achievement. The technology has reached a level of maturity that suggests the time may be right to apply it to the acquisition and assessment of reading and speaking skills, including assisting students and adults in developing literacy, and as a tool for the acquisition of English. However, technology should seldom be viewed as a solution to educational challenges in its own right. Technology can at best support and enable learning. A deep understanding of how students learn and a supportive learning environment are essential for technology to be an asset in the learning process.

Further Reading

The following sources provide overviews or examples of creative applications of technology to education. Quellmalz and Pellegrino (2009) provide an up-to-date review of the possibilities of utilizing technology to improve assessment and learning. A recent report by Tucker (2009) discusses technology and assessment and provides a sampling of projects. Discussion of leading-edge projects on applying speech technology to

education can be found in Litman and Silliman (2004), as well as Schwitter and Tawhidul Islam (2003). Poulsen, Wiemer-Hastings, and Allbritton (2007) discuss applications of speech technology to tutoring bilingual students. Holland and Fisher (2008) and Eskenazi (2009) discuss more general applications to education.

References

Attali, Y., & Powers, D. (2008). *A developmental writing scale* (ETS Research Report No. RR-08-19). Princeton, NJ: Educational Testing Service.

Baer, J., Kutner, M., & Sabatini, J. (2009). *The basic reading skills of America's adults: Results from the 2003 National Assessment of Adult Literacy* (U.S. Department of Education). Washington, DC: Educational Testing Service.

Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment*, *1*(1). Retrieved from http://www.jtla.org

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation. *AI Magazine*, *25*(3), 27–36.

Campuzano, L., Dynarsky, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts*. Retrieved from http://ies.ed.gov/ncee/pdf/20074005.pdf

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*, 1–47.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2007). *Building a validity argument for the Test of English as a Foreign Language™*. New York: Routledge.

Cuban, L. (2001). *Oversold and underused: Computers in classrooms*. Cambridge, MA: Harvard University Press.

Deane, P., Quinlan, T., & Kostin, I. (in press). *Automated scoring within a developmental, cognitive model of writing proficiency*. Princeton, NJ: Educational Testing Service.

Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication*, *51*, 832–844.

Fishman, B. J., Gomez, L. M., & Soloway, E. (2001). *New technologies and the challenge for school leadership*. Retrieved from http://www-personal.umich.edu/~fishman/downloads/fishman_b_gomez_l_soloway_e.pdf

Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment*, *2*(1). Retrieved from http://www.jtla.org

Gruenenberg, K., Katriel, A., Lai, J., & Feng, J. (2008). Reading companion: A interactive web-based tutor for increasing literacy skills. In C. Baranauskas, P. Palanque, J. Abascal, & S. D. J. Barbosa (Eds.), *Human-computer interaction – INTERACT 2007* (pp. 345–348). Berlin: Springer.

Holland, V. M., & Fisher, F. P. (2008). *The path of speech technologies in computer assisted language learning: From research toward practice*. New York: Routledge.

Kirsch, I., Braun, H., Yamamoto, K., & Sum, A. (2007). *America's perfect storm: Three forces changing our nation's future*. Princeton, NJ: Educational Testing Service.

Lei, C., Zechner, K., & Xi, X. (2009, June). *Improved pronunciation features for construct-driven assessment of non-native spontaneous speech*. Paper presented at the annual meeting of the North American Chapter of the American Association for Computational Linguistics – Human Language Technologies, Boulder, CO.

Litman, D. J., & Silliman, S. (2004). ITSPOKE: An intelligent tutoring spoken dialogue system. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*. Retrieved from http://acl.Ldc.Upenn.Edu/n/n04/n04-3002.pdf

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241–256.

Miller, G. (2009). Computers as writing instructors. *Science*, *323*, 59–60.

Mostow, J., Aist, G., Huang, C., Junker, B., Kennedy, R., Lan, H., et al. (2008). 4-month evaluation of a learner-controlled reading tutor that listens. In V. M. Holland & F. P. Fisher (Eds.), *The path of speech technologies in computer assisted language learning: From research toward practice* (pp. 201–219). New York: Routledge.

Myers, M. (2003). What can computers and AES contribute to a K-12 writing program? In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading*, *7*(1), 3–24.

Poulsen, R., Wiemer-Hastings, P., & Allbritton, D. (2007). Tutoring bilingual students with an automated reading tutor that listens. *Journal of Educational Computing Research*, *36*(2), 191–221.

Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and testing. *Science*, *323*, 75–79.

Rosen, D. J. (2000). Using electronic technology in adult literacy education. In J. Comings, B. Garner, & C. Smith (Eds.), *Annual review of adult learning and literacy* (Vol. 1, pp. 304–315). San Francisco: Jossey-Bass.

Schwitter, R., & Tawhidul Islam, M. (2003). S-tutor: A speech-based tutoring system. In U. Hoppe, M. F. Verdejo, & J. Kay (Eds.), *Artificial intelligence in education: Shaping the future of learning through intelligent technologies* (pp. 503–505). Burke, VA: IOS Press.

Spillane, J. P. (2004). *Standards deviation: How schools misunderstand education policy*. Cambridge, MA: Harvard University Press.

Stites, R. (2004). Implications of new learning technologies for adult literacy and learning. In J. Comings, B. Garner, & C. Smith (Eds.), *Review of adult learning and literacy* (Vol. 4, Ch. 4). Mahwah, NJ: Lawrence Erlbaum Associates.

Tucker, B. (2009, February 17). *Beyond the bubble: Technology and the future of student assessment*. Retrieved from http://www.educationsector.org/

Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wall, D., & Horák, T. (2008, April). *The TOEFL impact study*. Paper presented at the Association of Language Testers in Europe (ALTE), Cambridge, U.K.

Williams, S. M. (2002). *Speech recognition technology and the assessment of beginning readers*. Paper presented at the National Research Council Workshop on Technology and Assessment: Thinking Ahead, Washington, DC.

Williamson, D. M., Mislevy, R. J., & Bejar, I. I. (2006). *Automated scoring of complex tasks in computer-based testing*. Mahwah, NJ: Lawrence Erlbaum Associates.

Xi, X., Higgins, D., Zechner, K., & Williamson, D. M. (2008). *Automated scoring of spontaneous speech using Speechrater v1.0* (ETS Research Report No. RR-07-02). Retrieved from http://www.ets.org/research/researcher/RR-08-62.html

Zechner, K., Higgins, D., Lawless, R., Futagi, Y., Ohls, S., & Ivanov, G. (2009, September). *Adapting the acoustic model of a speech recognizer for varied proficiency non-native spontaneous speech using read speech with language-specific pronunciation difficulty*. Paper presented at the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009), Brighton, U.K.

Zechner, K., Higgins, D., Xi, X., & Williamson, D. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication*, *51*, 883–895.

Zechner, K., Sabatini, J., & Lei, C. (2009, June). *Automatic scoring of children's read-aloud text passages and word lists*. Paper presented at the NAACL-HLT Workshop on Innovative Use of NLP for Building Educational Applications, Boulder, CO.