Crossroads
The ACM Student Magazine

Superhuman Speech by 2010: An Interview with Jennifer Lai

by *Paula Bach*

At IBM Research, Jennifer Lai (jlai@us.ibm.com) has been a key player in speech technology research, holding thirteen patents, publishing chapters for several books, and having over thirty papers appear in peer-reviewed journals and conferences. Jennifer took time out of her busy schedule to answer some questions about speech technologies for Crossroads*. In the interview, Jennifer shares her expertise on designing speech interfaces, her history at IBM, and advice for undergraduate and graduate students who are interested in working with speech technologies.*

Please tell us about your experience with speech technologies. I started working with speech technologies when I joined the speech group at IBM Research in November 1987—hard to believe that it was that long ago. I was fortunate enough to join a team with some of the great names in speech research, like Fred Jelinek, Bob Mercer, Peter Brown, and the Della Pietra brothers (twins actually). We had so much fun researching and developing these technologies that it still amazes me that we actually got paid to work there. We were very focused on the Hansard texts, working on an automatic language-to-language translation system, since the Hansard texts were available in both French and English. I wrote my very first patent with those guys — it was a bear of a patent—some 100 pages or so, and I think it is one of the cornerstone patents in language translation. I worked on that team for about eight years and then moved out of the algorithm side of the business and closer to the end-users, which is really where my heart lies. Since then, I've used speech technologies in many of the solutions and prototypes that I have created but have not been involved in the actual creation of new versions of the IBM speech engine.

What are some principles of a good speech interface? Nicole Yankelovich and I have co-authored many publications on the design of speech interfaces (papers, book chapters, and encyclopedia articles), and they are a good place to start when faced with the task of designing a spoken interface. Many of the general principles of good user interface design apply: understand your user population, the context of use, and the goals of the task at hand. Of course, where speech technology, or any recognition technology for that matter, differs fundamentally from other technologies is in their accuracy. If a user strikes the 'k' key, a keyboard will always produce a 'k.' The only errors that a designer has to account for with keyboard input are user errors. With a spoken interface, a designer also has to account for errors that the technology introduces, so error detection and prevention plays a more important role in the design of SUIs (speech user interfaces) than it does with GUIs. Transparency of the interface (what can a user say) and efficiency of the interaction are two other key areas in SUI design. Lastly, it is important to set user expectations correctly. Nicole and I have defined a suite of guidelines for SUI designers, for each of these areas, that can be found in our publications.

From your experience as a leader in speech technology at IBM and a member of the editorial board of the *International Journal of Speech Technology,* please tell us about what you think the three greatest challenges in speech technology are. Well, I'm tempted to follow the old real estate maxim, "location, location, location," and say, "accuracy, accuracy, accuracy."

What is the biggest challenge you have faced when designing speech technologies? The accuracy issue previously mentioned aside, the greatest challenge in the design of a usable spoken interface is keeping the interaction efficient. People can read faster than they can listen, so presenting information in a spoken interface tends to slow the interaction down. Also, every time the interface has to confirm something or the user has to correct a recognition error, it slows the interaction down as well. Bearing in mind that for most applications the user's goal is usually to complete a task as efficiently as possible, a SUI designer has to keep the interaction moving forward. Unfortunately, it seems that for a lot of telephony speech applications, the first shortcut that users learn is to try the term agent, or representative. It should be every SUI designer's primary goal to get the user engaged fast enough that they are not looking for a way to opt out of the speech system.

Which speech technologies have you designed interfaces for? Do the different

architectures affect the way you design interfaces? Absolutely! When we create a solution that runs on a client machine with a dedicated speech engine and a regular (large) computer screen to interact with the user, we design it [much] differently than a solution that is going to run on a cell phone or one that will be embedded in the dashboard of a car.

Which speech interface project was your favorite? Why? There are lots of speech projects that I've really enjoyed being involved with, but I have to say that my favorite project is the one that we recently launched to improve literacy skills of kids and adults around the world. Reading Companion (**http://www.readingcompanion.org**) is a web-based reading tutor application that is freely available to non-profit organizations, such as elementary schools and adult literacy centers. An animated tutor character (we use a panda for the children and stick figure for adults) uses speech technologies to "listen" and "speak" when interacting with the reader. There were many really interesting problems to tackle with this project, including technical issues, such as the packaging and delivery of real-time, highly accurate speech recognition over the web, and UI issues, such as creating an engaging user experience for both the adults and the children. We are getting very positive feedback from the current user base of 14,000 users, but I want to know what you think. Check out the overview on the website (login is restricted to registered users) and send me your thoughts.

Do you still consider yourself to be a specialist in speech technologies? Actually, I don't. I am sort of technology-agnostic at this point. My team here at Watson is focused on creating innovative solutions that address real-world problems, and the technologies that we use as part of those solutions are incorporated based on what is required by the user population and the context of use. A key focus for us is prototyping systems that support informal learning in the workplace, and we are also very interested in understanding the transfer of skill from master to apprentice. Speech recognition may very well have a role to play in that prototype—it's still too soon to say.

What advice can you give to undergraduate students interested in pursuing a career in speech technology? Hard to say since the area of focus within speech technology really depends on your interests. For example, would you rather work on developing the algorithms for decoding speech, or do you want to create solutions that use speech and drive the requirements definition for the engine? I may be revealing my own particular bias when I encourage students in computer science to take HCI [human-computer interation] classes and complete field and lab work to help them

gain an understanding of fundamental principles of good HCI design.

What advice can you give to graduate students interested in researching speech technologies? Come do a summer internship at IBM Research. We love to be around your enthusiasm and energy, and summers are really pleasant in New York. As you may have heard or read, there is a big effort underway at Watson to develop superhuman speech by 2010 (**http://www.research.ibm.com/superhuman/superhuman.htm**), which has the goal of reaching, or exceeding, human performance in speech. I can't wait, nor should you. Come and be a part of the experience.

Thank you, Jennifer.