VUI Visions

Speech for Mobility

Amy Neustein, Linguistic Technology Systems

In this guest column, we ask designers skilled in creating Voice User Interfaces to highlight a particular aspect of VUI design inspired by actual deployments. In this issue, Amy Neustein, Ph.D., Founder and CEO, ***Linguistic Technology Systems****, reviews the views on the role of speech technology in mobile devices of the contributors to a recent book she edited. Dr. Neustein is Editor-in-Chief of the International Journal of Speech Technology (Springer Verlag), and series editor of SpringerBriefs, Series in Speech Processing and Technology. She edited the recently published book "Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics" (Springer 2010). She serves as a quest columnist on speech processing for Womensenews. Linguistic Technology Systems is a NJ-based think tank for intelligent design of advanced language based emotion-detection software to improve human response in monitoring recorded conversations of terror suspects and helpline calls. Neustein is a graduate of Boston University (1981) where she received her Ph.D. in sociology; her specialty area is Conversation Analysis. Her articles appear in academic, industry and mass media publications. Her books have been cited in the Chronicles of Higher Education. Here work has won her several awards: a pro Humanitate Literary Award; a Humanitarian Award; the Woman of Valor Lifetime Achievement Award; and the Information Technology Next Generation Medical Informatics Award. She serves on the visiting faculty of the National Judicial College and as a plenary speaker, moderator, and panelist at academic and industry conferences. Dr. Neustein is a member of MIR (machine intelligence research) Labs, http://www.mirlabs.org/), which does advanced work in computer technology to assist underdeveloped countries in improving their ability to cope with famine, disease/illness, and political and social affliction. She is a founding member of the New York City Speech Processing Consortium, a newly formed group of NY-based companies, publishing houses, and researchers dedicated to advancing speech technology research and development.*

As Editor of *Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics* (Springer, October 2010), I was able to provide a forum for today's speech tech industry leaders to discuss the challenges, advances, and aspirations of voice technology. While the book's fourteen chapters were divided into three sections—mobile environments, call centers, and clinics—the practical ubiquity of mobile devices made this three-part division seem almost irrelevant at times. For example, Matt Yuschiks's chapter, which opened the call centers section, provides a vivid discussion of how to provide today's call centers with multimodal capabilities (to support text, graphic, voice, and touch) in self-service transactions**,** so that customers who contact the call center using their mobile phones (rather than a fixed line) can expect a sophisticated interface that lets them resolve their service issues in a way that uses the full capabilities of their handsets. On the flip side, Yuschik showed how call center agents themselves, who use mobile devices that support multimodality, can experience more efficient navigation and retrieval of information to complete a transaction for a caller. Similarly, James Rodger, professor of Management Information Systems and Decision Sciences at Indiana University of Pennsylvania, Eberly College of Business and Information Technology**—**with his co-author, James George, senior consultant at Sam, Inc.—in their chapter which appears in the "clinics" section of the book addressed a breakthrough mobile app for preventive healthcare in the military. Rodger and George demonstrated in their chapter how they tested and validated end-user acceptance of speech in the clinic setting aboard U.S. Navy ships, by focusing on measuring user satisfaction with a voice-activated medical tracking application that is run on a compact *mobile* device for a "hands-free" method of data entry in a clinical setting.

The book begins with an introduction to the role of speech technology in mobile applications written by Bill Meisel, President of TMA Associates, and editor of *Speech Strategy News* and co-chair (with AVIOS) of the annual Mobile Voice conference in northern California. Meisel opened his discussion by quoting the predictions published by the financial investment giant Morgan Stanley in its *Mobile Internet Report***,** issued near the end of 2009. Meisel pointed to Morgan Stanley's 694-page report in which Mobile Internet Computing was said to be "the technology driver of the next decade," following the Desktop Internet Computing of the 1990s, the Personal Computing of the 1980s, the Mini-Computing of the 1970s and**,** finally, the Mainframe Computing of the 1960s. In his chapter, fittingly titled "Life on the Go – The Role of Speech

Technology in Mobile Applications," Meisel pointed out that since "the mobile phone is becoming an indispensable personal communication assistant and multi-functional device . . . [such a] range of applications creates user interaction issues that can't be fully solved by extending the Graphical User Interface and keyboard to these small devices." Instead, Meisel envisioned "speech recognition, text-to-speech synthesis, and other speech technologies" as "part of the solution" by explaining that "unlike PCs, every mobile phone has a microphone and speech output."

Advances in Speech Recognition—which was published at the beginning of this auspicious decade for mobile computing—examines the practical constraints of using voice in tandem with text. Following Meisel's comprehensive overview of the role of speech technology in mobile applications, Scott Taylor, Vice President of Mobile Marketing and Solutions at Nuance Communications, Inc., cautioned the reader about the need to "balance a variety of multimodal capabilities so as to optimally fit the user's needs at any given time." While there is "no doubt that speech technologies will continue to evolve and provide a richer user experience," argues Taylor, it is critical for experts to remember that "the key to success of these technologies will be thoughtful integration of these core technologies into mobile device platforms and operating systems, to enable creative and consistent use of these technologies within mobile applications." This is why speech developers, including Taylor himself, view speech capabilities on mobile devices not as a single entity but rather as part of an *entire* mobile ecosystem that must strive to maintain homeostasis so that consumers (as well as carriers and manufacturers) will get the best service from a given mobile application.

To achieve that goal, Mike Phillips, Chief Technology Officer at Boston-based Vlingo, together with members of the company have been at pains to design more effective and satisfying multimodal interfaces for mobile devices. In the chapter following Taylor's, titled "Why Tap When You Can Talk – Designing Multimodal Interfaces for Mobile Devices that Are Effective, Adaptive and Satisfying to the User" (excerpts of this chapter appeared in Speech Technology Magazine, September/October 2010), Phillips and his co-authors presented the findings from over 600 usability tests in addition to results from large-scale commercial deployments to augment their discussion of the opportunities and challenges presented in the mobile environment. Phillips and his co-writers stressed how important it is to strive for user-satisfaction: "It is becoming clear that as mobile devices become more capable, the user interface is the last remaining barrier to the scope of applications and services that can be made available to the users of these devices. It is equally clear that speech has an important role to play in removing these user interface barriers."

Johan Schalkwyk, Senior Staff Engineer at Google, along with some of his colleagues greatly enhanced the mobile environments section by reporting on their case study on voice search. In their chapter, titled "Your Word is my Command –Google Search by Voice: A Case Study," Schalkwyk and his co-authors illuminated the technology employed by Google "to make search by voice a reality" – and follow this with a fascinating exploration of the user interface side of the problem, which includes detailed descriptions and analyses of the specifically-tailored user studies that have been based on Google's deployed applications.

In painstaking detail, Schalkwyk and his colleagues demystified the complicated technology behind 800-GOOG-411 (an automated system that uses speech recognition and web search to help people find and call businesses), GMM (Google Maps for Mobile) which – unlike GOOG-411 – applies a multimodal speech application (making use of graphics), and finally the Google Mobile application for the iPhone, which includes a search by voice feature. The coda to the chapter is its discussion of user studies based on analyses of live data, and how such studies reveal important facts about user behavior, facts that impact Google's "decisions about the technology and user interfaces." Here are the essential questions addressed in those user studies: "What are people actually looking for when they are mobile? What factors influence them to choose to search by voice or type? What factors contribute to user satisfaction? How do we maintain or grow our user base? How can speech make information access easier?"

The mobile environments section concludes with the study findings on a new speech-enabled framework that aims at providing a rich interactive experience for smartphone users – particularly in those mobile environments that can benefit from hands-free and/or eyes-free operations. Canadian professor Sid-Ahmed Selouani (Université de Moncton, Shippagan Campus) presents study findings on a new speech-enabled framework that aims at providing a rich interactive experience for smartphone users – particularly in those mobile environments that can benefit from hands-free and/or eyes-free operations. Selouani introduces this framework by arguing that it is based on a conceptualization that divides the mapping between the speech acoustical microstructure and the spoken implicit macrostructure into two distinct levels, namely the signal level and linguistic level. Selouani utilized the Carnegie-Mellon Pocket Sphinx engine for speech recognition and the Artificial Intelligence Markup Language (AIML) for pattern matching. The evaluation results showed that including both the Genetic Algorithms (GA)-based front-end processing and the AIML-based conversational agents led to significant improvements in the effectiveness and performance of an interactive spoken dialog system in a mobile setting.

The call center section that follows contains interesting case studies on evaluating the performance of automated customer care contact centers, including those that deliver multimodal customer assistance; helping system designers create a voice app that delivers a satisfying user experience that meets or exceeds customer expectations; designing a voice app that can robustly detect angry user turns, including the accretion of anger that builds up from prior conversational speaking turns; and demonstrating the nuts and bolts of phonetic-based search and indexing to deliver the best possible speech analytic solutions to enterprises who need to keep track of customer interactions with call centers in real time. The clinic section rounds off the book with empirical findings on the benefits of incorporating speech recognition as part of the electronic medical record; a study of user acceptance of a voice-activated medical tracking application; a review of studies on the use of computational approaches (based on language "cues" that consist of acoustic signal, lexical and semantic features) to model speaker (emotional) state to gauge illness and recovery; and finally, the use of spectrographic analysis to assess neonatal health status from an infant's cry so that physicians can perform early intervention in the diagnosis and treatment of pulmonary dysfunction and other abnormalities.