

Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics. Amy Neustein, ed. New York: Springer Science and Business Media, 2010.

Jessica Zimmer
jessicazimmer@hotmail.com

In this edited volume, Amy Neustein brings together a collection of intriguing studies of the development and use of speech recognition (SR) technology in different workplaces. Cultural, linguistic, and medical anthropologists who study communication at work will be intrigued but perhaps overwhelmed by the ways engineers, linguists, and information technology specialists are increasingly placing computers between people in their speech acts with one another. *Advances in Speech Recognition* is more readable than many other collections about SR technology. The majority of the tables and figures are simple. Most importantly, each essay contains a clear introduction and conclusion that can be tied to research in social science. Anthropologists can use the book to teach themselves and their students about the world of SR technology research and the variety of SR technology researchers. It explains common abbreviations, such as GUI (graphic user interface), AIML (artificial intelligence markup language), and IVR (interactive voice response [systems]). The book also explains abbreviations for less well-known phrases, such as WIMP: “windows, icon, mouse, and pointer (WIMP)-based machine” (Selouani 2010:109).

The chapters are divided into three groups — the first concerning the use of mobile technology, the second concerning the creation of automated customer service in call centers, and the third concerning the use of SR technology in medical settings, particularly with regard to building electronic medical records. Some representative topics are a history of the development of Google Search by Voice, the evolution of studies to create satisfying experiences with automated verbal responses, and a discussion of how computers analyze the communication patterns of individuals with medical conditions such as autism spectrum disorders.

The research reported in this book emphasizes advantages of SR technology, pointing out, for instance, that computers equipped in this way cost less than human workers and also noting that it saves human workers from having to engage in repetitive and challenging tasks, such as data entry in medical transcription. Regrettably, the accounts do not quantify the number of jobs lost due to the use of SR

technology and rarely touch on how human communication patterns of customers and workers are changing in response to SR technology.

Anthropologists can use this collection to understand how SR technology is affecting different populations. It can help them design studies to collect data about the changes in language, culture, and personality that occur when people come into contact with SR technology. Two of the essays in particular showcase research that could benefit from future analysis. “‘For Heaven’s Sake, Gimme a Live Person!’: Designing Emotion-Detection Customer Care Voice Applications in Automated Call Centers” concerns customers’ frustration with automated responses. “‘You’re as Sick as You Sound’: Using Computational Approaches for Modeling Speaker State to Gauge Illness and Recovery” explains how individuals with different disorders sound to people without these disorders.

As an example, researchers in the study in “‘Gimme a Live Person!’” determined that certain qualities of human speech, including pitch and loudness, indicate anger. They attempted to measure changes in these qualities to understand whether customers were becoming angry when interacting with automated response systems. It would be relevant for future investigations to determine the meaning of such qualities in human speech across cultures. Anthropologists might be able to discover whether people from different national and ethnic groups changed these qualities in their speech, and to the same degree, when frustrated with call center systems. Such work might help industry professionals design culturally sensitive software, which would be responsive to data suggesting that an American customer’s change in pitch tends to mean something different than a Chinese customer’s change in pitch.

Although the field may seem highly technical and specialized, anthropologists studying workplaces should be aware of the devices and coded programs that are replacing speech acts with exchanges of data. Advances in cloud computing and near field communication (NFC) will soon allow individuals and groups to send each other data packets that reduce the need for verbal communication. Cellular phones enabled with NFC technology are already able to read data on skin patches and blister packages. Software programs in the phones can determine whether a patient has taken a prescribed medication and how the patient’s body is processing that medication. Patients can send the data directly to a medical professional, eliminating the need to go to the hospital to be assessed or even to call and update a medical professional about their condition.

These essays may also become critical in understand-

ing how individuals and groups come to perceive speech acts as stressful. A passage from Stephen Springer's "Great Expectations': Making Use of Callers' Experiences from Everyday Life to Design a Satisfying Speech-only Interface for the Call Center" illustrates a frustrated customer's attempt to escape the stress he was experiencing. Responding to a verbal command in a software program, the caller exclaimed: "Lady, if you can understand English, you can understand what I'm sayin' — I want to speak to a living body, please" (182).

In "Life on-the-Go': The Role of Speech Technology in Mobile Applications," William Meisel states that speaking on a "voice channel" may become an important way for people to use services on smartphones, adding that the number of "voice sites" supported by advertising is likely to increase. If individuals are asked to listen and speak extensively to automated systems as well as real people, they may become fatigued and reluctant to communicate.

The research reported here should help anthropologists design studies that will chronicle the development and impacts of different SR applications and document the changing significance of speech acts in the workplace and beyond. James Larson remarks in the epilogue that speech and text may come to be seen as equal and interchangeable. Imagining an Internet and communications network on which user comments on websites can be heard, with unique, identifiable voices, reveals that the anonymity of text may soon disappear. If speech and text are conflated, privacy is likely to be defined far more narrowly than it is today. As Stephen Swigart (2008:93) reminds us in reference to the insights of Pierre Bourdieu, "every linguistic exchange, no matter how seemingly insignificant, carries within it traces of the social structure it both expresses and helps to reproduce."

References

Selouani, Sid-Ahmed. 2010. "Well Adjusted": Using Robust and Flexible Speech Recognition Capabilities in Clean to Noisy Mobile Environments. *In* *Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics*. Amy Neustein, ed. Pp. 91-114. New York: Springer Science + Business Media.

Swigart, Leigh. 2008. The Limits of Legitimacy: Language Ideology and Shift in Contemporary Senegal. *Journal of Linguistic Anthropology* 10(1): 90-130