

Preface

Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship represents the culmination of ten years of research into the psychology and design of voice interfaces by Clifford Nass and his coworkers at Stanford University. Scott Brave joined the laboratory as a Ph.D. student in 1999 and, since graduating in June 2003, has continued as a postdoctoral scholar and integral member of the team.

Until 1996—when Cliff’s first book (coauthored with Byron Reeves), *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, appeared—the laboratory focused on textual and graphical user interfaces. Soon after, many companies and researchers turned their attention to interfaces that talked and listened. Voice-interface technology was not quite ready for prime time, but most agreed that it was “worth watching.” Computers were routinely shipping with speakers; using interactive voice systems was proving to be more cost-effective than hiring additional workers for call centers; and efficient and inexpensive technologies for digital signal processing enabled voice interfaces to appear in cars, toys, and appliances.

Unfortunately, because little theoretical research or practical guidance was available on creating successful voice interfaces, most voice interactions with technology were frustrating and ineffective. The juxtaposition of the clear need for better interfaces and the opportunity for fundamental research in an intriguing and unexplored area was irresistible to our lab. Furthermore, the lab was in a unique position to make significant contributions: its focus has always been the *social* aspects of human-technology interactions, and nothing is more social than speech.

Unfortunately, building voice interfaces with which to test our theories was difficult. For the most part, technologies that could talk and listen required a tremendous amount of nurturance from engineers and were expensive and difficult to use. Even if we could assemble the requisite staff and funding, we could not craft new experiments with the necessary speed. The impediments seemed insurmountable until 2000,

when the CSLU Toolkit appeared on the scene. This remarkable and free piece of software—available when voice extensible mark-up language (VXML) didn't yet exist—enabled virtually anyone to create synthetic speech files, perform voice recognition, produce lip-synch with a synthetic face, and even build a telephone call center.

Because we wanted to answer as many questions and train as many voice-interface researchers as quickly as we could, Cliff created an intensive, ten-week course that focused on design and execution of voice-interface experiments (Scott was a researcher in the class). The course was divided into nine research groups of three students each. Based on research concerning successful teams, we opted for maximum diversity: each group included at least one student from engineering, one student from the social sciences, one Ph.D. student, one undergraduate student, one male student, and one female student. At the start of the course, the groups were shown approximately thirty open questions in the research on voice interfaces that involved psychological as well as design issues. Students were told, "Meet early and often with your group, pick a question, design a relevant experiment, run it, analyze the results, and create a presentation that summarizes your findings. You have ten weeks." Cliff forgot to tell the groups and the teaching assistant, Eva Jettmar, that this was an impossible task, but (as Stanford students tend to do) they went beyond meeting the requirements of the course and produced research that was worthy of scientific and industry attention.

The abundance of compelling results prompted Cliff to invite industry and academic leaders to a two-hour presentation by the student groups, which was attended by seventy-five people from forty companies and universities from throughout the world. The significant and exciting discoveries concerning how people behave with and think about voice technology (many of which were subsequently published) coupled with the enthusiastic responses from industry encouraged researchers at the lab to continue with additional experiments that investigated human interactions with voice technologies.

At the same time that the laboratory was producing its insights, Cliff was independently pursuing the design of a number of voice interfaces for Microsoft (its AutoPC and CD-ROM products), BMW (a voice interface for the Five Series in the United States, Germany, and Japan), IBM (a corporate training system), Philips (a movie-selection system, *Voices in Your Hand*), Fidelity (a stock-ordering system in Japan), US West (a voice-mail system), Verizon (call centers), General Magic (the Portico Virtual Assistant), and OMRON (a factory automation system). This activity provided inspiration for basic research questions and sensitized the lab to the day-to-day issues that designers face.

Wired for Speech describes and synthesizes our research on voice interfaces in terms of how the human brain is activated by voice and how computers can best relate to us. This book is filled with general principles, specific examples, rigorous experiments, and “fascinating facts to know and tell.” *Wired for Speech* should enable anyone who is interested in the present and future of voice products and services to understand their promise and pitfalls, scientists to refine their research on the psychology of voices and voice-based technology, and designers and marketers to produce interfaces that engage and support users rather than frustrate them.