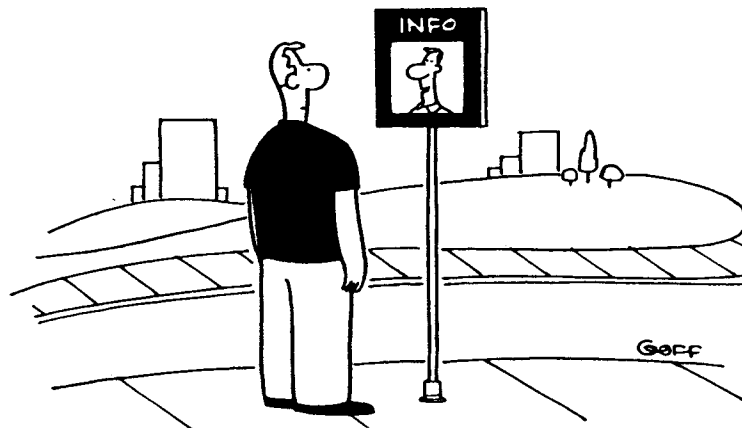


1996-2001
Effective Speech User Interfaces

Chapter

1

Basic Concepts



"Sorry, I don't know anything. We just thought it would be really cool to build this high-tech information kiosk."

“There is one thing stronger than all the armies in the world, and that is an idea whose time has come.”

Victor Hugo

This book is about speech interfaces. For this first chapter we will focus on what an interface is, types of interfaces, and what we mean by a speech interface. At the end of this chapter, we describe the individual parts (sections) and chapters of the book.

What Is an Interface?

An interface is the part of technology that people interact with. The interaction between a computer and a user is a two-way interaction. Sometimes the user is giving information to the computer, and sometimes the computer is giving information to the user. This interaction can include hardware components like keyboards, mice, or keypads, or software components such as a screen, window, page, sound, or talking voice. Any technology, computer, or machine also has elements the operator or user never sees or interacts with. In the case of software, code and programming are hidden elements. Since the user only interacts with the interface, the interface is in many ways the embodiment of the product. If a user comments that a particular product is easy to use or hard to use, they

are not necessarily commenting on the underlying technology, but usually on the interface.

Designing an interface is no small task. What you consider a “well-designed” interface depends on your perspective. To a programmer, a well-designed interface might mean that the interface works within the technical constraints of the project. To a product manager, a well-designed interface might mean that the interface can be easily changed over time as new versions of the product are developed. To a business stakeholder, a well-designed interface might mean that the interface has more bells and whistles than the competition’s product. To a usability engineer a well-designed interface might mean one that is designed with a particular user group in mind. And to the user of the product, a well-designed interface might mean that the product works the way they expect it to. All these perspectives are important, but because our background is usability engineering, we tend toward the last two definitions, and concentrate on them in this book.

As usability engineers we are interested in how to make technology easy to learn and use. We are concerned with how to design products so that people can be as productive as possible with the product, as quickly as possible. The biggest single impact we can have on products is in the area of the interface. Designing to optimize usability means paying specific attention to how the interface looks and acts. This includes

- ensuring that the interface matches the way people need or want to accomplish a task;
- using the appropriate modality (for example, visual or voice) at the appropriate time;
- spending adequate design time on the interface.

Designing to optimize usability means paying specific attention to how the interface looks and acts.

What Are the Types of Interfaces?

There are several types of interfaces. In the past (and still around to some extent) were character-based user interfaces (CHUI), then graphical user interfaces (GUI) became prevalent, and next came Web user interfaces (WUI). In this book we discuss speech user interfaces (SUI).

What Is a Speech Interface?

It seems as if this question, What is a Speech Interface?, would be simple to answer, but it is not. Because it is a relatively new idea to have our technology involve speech, this is a field that is just starting to grow. Like any new field, the definitions and terminologies are not standard. Here are some of the current terms in use, what they usually mean, and how we will apply them here.

Speech Interfaces

The term *speech interface* describes a software interface that employs either human speech or simulated human speech. You can further break down speech interfaces into *auditory user interfaces* and *graphical user interfaces with speech*.

Auditory User Interfaces (AUI)

An auditory user interface (AUI) is an interface which relies primarily or exclusively on audio for interaction, including speech and sounds. This means that commands issued by the machine or computer, as well as all commands issued

by the human to control the machine or computer, are executed primarily with speech and sounds. Although AUI may include a hardware component, such as a keypad or buttons, visual displays are not used for critical information. Examples of auditory user interfaces include

- medical transcription software that allows doctors to dictate medical notes while making rounds;
- automobile hands-free systems that allow drivers to access travel information and directions (systems without screens and maps);
- interactive voice response (IVR) systems in which users access information by speaking commands, such as menu numbers, to listen to information of their choice;
- products for the visually impaired that rely only on audio text and cues.

Graphical User Interfaces with Speech (S/GUI)

In this book we discuss AUI, where the user interacts with the software primarily via speech. We also discuss interfaces where speech is part of the interface, but is joined by other interface forms—usually visual—such as a GUI window or a Web page. We call these multi-modal interfaces *graphical user interfaces with speech*, or S/GUI (for Speech/GUI). Examples of S/GUI include

- a wordprocessor that allows users to dictate text instead of or in addition to typing it in;
- Web navigation software that allows users to navigate to and within Web sites by using voice;
- talking dictionaries that speak definitions.

In these S/GUI applications, tasks can

- be completed using speech only, where users issue a speech command or listen to the software speak to them;
- rely on visual or manual GUI aspects, for example, viewing a graphic or clicking on a hyperlink;
- require or at least allow a combination of both a GUI aspect and speech, such as using speech commands to edit a document.

Non-Speech Audio

Some interface elements include audio, but not speech. These interface elements include music and sounds. Some non-speech audio is included in almost all interfaces of any type, including S/GUI, AUI, and GUIs. Examples of non-speech audio include:

- The computer beeps when the user makes an error.
- The user clicks on a map and hears a low tone to indicate that the water to be found at that site is deep in the ground or a high tone to indicate that the water is closer to the surface.

We do not focus on non-speech audio in this book.

Book Focus

In our research of books and articles on speech interfaces we were struck by two aspects of the literature. The first was the lack of information about speech interfaces overall. Second, the literature that does exist is technical in nature and assumes a thorough familiarity with speech technology. What happens, we wondered, when a

practitioner in software and interface design needs to work on a speech interface, but is new to the field? This is the point of view we adopted in this book.

We focus primarily on the AUI and S/GUI aspects of the interface, not the underlying programming techniques. Our bias is toward a usability point of view. We cover the basics of speech technology that an interface designer needs to understand in order to construct a usable interface, for instance, the basics of sound and language. We discuss guidelines for effective communication, prompting, wording, and error handling to ensure that the interface is easy to learn and use. We do not discuss writing code for speech recognizers.

Intended Audience

The intended audience for this book includes

- interface designers—those responsible for designing the interface portion of a product that includes a speech component or is primarily speech-based. We assume you are new to designing speech interfaces;
- usability engineers—those responsible for ensuring that a product that includes speech or is primarily speech-based is usable. We assume you are new to speech interfaces;
- developers—programmers and technical people on a speech project who are new to speech interfaces;
- product and marketing managers—those responsible for bringing a project to fruition. Product managers will be interested in what makes a speech interface work, and marketing managers will be interested in what makes a particular project usable and competitive.

We assume that you—our intended audience—are familiar with GUI interfaces, but have had little exposure to the design and development of speech interfaces. Interface design expertise will help you understand some of the underlying principles we discuss, but is not necessary to put the book to practical use. We have included information on human factors, interface design, and usability engineering for readers new to these fields.

While we assume you have a minimal familiarity with speech technology or acoustic areas, we have included summaries to provide critical information about these subjects.

Book Summary

Here is a summary of the book contents. Part One sets the stage with an introduction to the basic concepts of speech interface, sound, and language, and a discussion of how human factors impact these areas. Part Two covers available speech technologies and how to apply them. With these tools in hand, the boundaries and guidelines to using them in interface design are explained in Part Three. Part Four demonstrates best practices and techniques.

Part One

Part One: Introduction to Speech Interfaces contains information on speech technologies and human-computer interaction concepts, including

Chapter 1: Basic Concepts. Covers definitions of user interfaces, types of user interfaces, and definitions of speech interfaces;

Chapter 2: Human Factors in Speech Technology.

Gives an overview of the field of human factors, covers the basic concepts of human computer interaction, and summarizes the major human factors issues in speech interfaces. Discusses cost/benefit calculations for usability engineering of speech interfaces;

Chapter 3: The Nature of Sound. Explains sound terminology and technology;

Chapter 4: The Nature of Language. Summarizes what we know about language and communication, including the structure of language, grammar, syntax, and semantics.

Part Two

Part Two: Speech Technology explains the current state of the field of speech, including

Chapter 5: Speech Technologies. Defines and describes the main terms in the field, including speech recognition, speech synthesis, continuous versus discrete speech, and natural language;

Chapter 6: Computer Software. Surveys the current state of speech software on the market, both applications and development tools;

Chapter 7: Hardware. Covers hardware such as sound cards, microphones, voice cards, and modems that affect the usability of speech applications;

Chapter 8: Applications of Speech Technology. Describes how speech applications are being used today.

Part Three

Part Three: Laws and Guidelines for Speech Interface Design. Provides specific advice for how to improve the usability of speech applications, including

Chapter 9: Laws of Interface Design. Describes 20 laws of human factors that apply to speech interfaces. The discussion of each law includes the human factors research behind it;

Chapter 10: Speech Guidelines. Provides specific guidelines for designing speech interfaces. These guidelines are based on the laws from Chapter 9, and contain the dos and don'ts to follow when designing speech applications.

Part Four

Part Four: Best Practices. Covers best practices in interface design and usability engineering, including

Chapter 11: Usability Processes and Techniques. Describes a comprehensive method for designing applications to ensure they are usable, and the basic usability engineering techniques used in the process;

Chapter 12: Universal Design. Discusses designing for universal design and accessibility.

Throughout the book we include interviews with several industry experts on the topics of speech interfaces:

Kate Dobroth, American Institutes for Research
S. Ahmed Reza, Speech Interface Design

Sharon Oviatt, Center for Human-Computer
Communication, Oregon Graduate Institute of Science
and Technology

Candace Kamm, AT&T Labs

John Karat, IBM T.J. Watson Research Center

Nicole Yankelovich, Sun Microsystems

 WILEY

Designing Effective Speech Interfaces

Susan Weinschenk
Dean T. Barker

