

Automated Natural Spoken Dialog



The next generation of voice-based interface technology will enable easy-to-use automation of new and existing communication services, making human-machine interaction more natural.

Allen L.
Gorin
Alicia Abella
Tirso Alonso
Giuseppe
Riccardi
Jeremy H.
Wright
AT&T Labs—
Research

Engineers have long sought to design systems that understand and act upon spoken language. Extracting meaning from natural, unconstrained speech over the telephone is technically challenging, and quantifying semantic content is crucial for engineering and evaluating such systems.

Traditional menu-driven speech recognition systems force users to learn the machine's jargon, but many people are unwilling or unable to navigate such highly structured interactions. AT&T's "How May I Help You?" (HMIHY) technology (<http://www.research.att.com/~algor/hmihy>) shifts the burden to the machine by requiring it to adapt to human language and understand what people actually say rather than what a system designer expects them to say.

The intuition underlying our approach is that for a given task, some linguistic events are more *salient*—crucial to recognize and understand—than others. Researchers have already developed methods to automatically extract named entities such as phone and credit card numbers embedded in natural spoken language¹ and to translate utterances into Spanish and Japanese.² Building on these and other early laboratory experiments,³ we have developed algorithms^{4,5} that automatically learn the salient words, phrases, and grammar fragments for a given task far more reliably than other methods.

ACHIEVING NATURAL FUNCTIONALITY

Traditional telephony automation systems offer a list of menu options and prompt users to navigate

to the appropriate destination to obtain service or resolve a problem. Sometimes the machine itself provides the desired service, such as supplying an account balance or billing a call to a credit card, while in other instances only a specially trained person can do so.

Menu systems can be implemented using a touch-tone system ("Press 1 if you want *x*, press 2 if you want *y*, ..."), voice labels ("Please say collect, calling card, ..."), or a hybrid of the two ("Press or say 1 if you want *x*, ..."). Each can be useful when the list of options is short and well understood by customers, but for certain tasks designers must resort to unwieldy hierarchical menus that can bore and frustrate users. On the other hand, callers faced with succinct menu options may have difficulty deciding which of the proffered categories matches what they want. In both cases, users often either bail out of the system by pressing zero or do nothing in the hope of eventually being connected to a person.

Determining the number of options and amount of detail to include in menus can be difficult. In contrast, a human receptionist could simply handle this same routing task by asking, "How may I help you?" and let the caller describe the request or problem. The receptionist should know enough about the task to transfer the caller to the appropriate person or automated module.

Our goal in developing HMIHY was to engineer a call-routing system with this same natural functionality. A caller receives a greeting and makes a request as if talking to a person. The system's job is to recognize and understand what the user wants—