Introduction

Telephone menu systems take advantage of nearly ubiquitous voice communication infrastructure to provide highly accessible services. Businesses from financial institutions to utility companies allow customers to check balances, order services, pay bills, and make other significant transactions through interaction with automated systems by telephone. Telephone lines at home, work, and nearly anywhere else people spend time; public telephones in many places; and—for the many who own them—cellular phones put these services frequently within arm’s reach.

In this report we evaluate the use of speech recognition in telephone menu systems. To get a sense for the advantages and disadvantages of this technique, we explored a system called 1-800-555-TELL that offers a number of sophisticated services such as stock quotes, news, shopping, and even driving directions using only voice input. (Keypad input can optionally substitute for voice in some cases.) We consider the tradeoffs involved in using speech recognition, both as an interaction technique unto itself and as compared to keypad input.

Advantages

Fewer Demands on the User

A conversational interfaces requiring only listening and speaking leaves the hands and eyes free for other tasks. Inadvisable though it may be, many people use cellular phones while driving automobiles. It’s bad enough when drivers incur the cognitive load of performing another task while controlling the vehicle, but it’s even worse for them to look away from the road to see the keypad and release the steering wheel to press buttons.

Even when users aren’t driving, they benefit from the conversational interaction style. Speaking in response to prompts is familiar and facile, requiring much less physical and mental effort than many other interfaces including voice menus with keypad entry. As a result, the experience is less fatiguing and more pleasant.
**Direct Mapping**

Speech recognition allows users to say what you want (such as “main menu”) instead of remembering an arbitrary key to press for that option. Not only does this make the gulf of execution smaller, but it saves some time the user would have to spend listening to explanation of the correspondence (as in “press zero for main menu”); the system can more quickly enumerate or describe the set of available options.

The set of options may be described instead of enumerated when it is large, such as names of cities or stock symbols. Free speech input easily allows many options in one menu, since with a succinct description the user can easily specify the desired item simply by naming it. Keypad entry can be used to spell out names, but this is awkward and error-prone.

**Error Prevention**

With voice input, users are less likely to give accidentally incorrect input, such as by bumping an unintended key or pressing a button in such a way it pulses more than once. This is particularly relevant on cellular phones, whose keypads become more constricted as the phones themselves get smaller.

Another kind of error users make is missing output from the system while the phone is away from their ears. Voice input allows the phone to stay in place, in the mode for which the phone was primarily designed.

**Accessibility**

Older rotary telephone systems are rare nowadays, so it’s a minor but nevertheless worthwhile point that speech recognition accommodates their use whereas systems that require keypad entry do not.

**Disadvantages**

**Speech Recognition Inaccuracy**

A number of factors can contribute to problems for the speech recognition engine. Speakers of different English dialects or those with foreign accents or otherwise unconventional pronunciation may have some difficulty. Background noise can interfere with the speech, and poor signal quality, especially with cellular phones, can be problematic as well. These difficulties may cause misrecognitions, especially when the number of possible utterances is large enough to include many that sound relatively similar.

**Privacy**

Unlike with keypad entry in telephone menu systems or most traditional input devices in other interfaces, speech input is inevitably available to anyone nearby. Depending
on the nature of the transaction, this might be a concern; certainly, it would be undesir-able to compromise access codes or personal identification numbers for accounts by speaking them. Of course, keypad entry can be used alongside speech input.

**Expert Navigation Speed**

Experts who know the exact sequence of keypresses to reach the option they want may be able to navigate more quickly than by speaking. Keypresses can be rapid and brief, and recognition of the tones on the system side takes less time than speech recognition. This is probably inevitable, but the problem can be mitigated by allowing speakers to jump more directly to the content they want—menu hierarchies can be made flatter by offering more options at each level, and complex utterances could even specify selections from multiple levels at once.

**Limitations Intrinsic to Voice Menus**

With or without speech recognition, voice interfaces are ultimately constrained as a linear modality. Users must listen to lists of options, and if they miss the one they want, they usually have to repeat again from the beginning—if that option is even available. Being able to interrupt long recordings is helpful; having to listen to something one definitely has no interest in is aggravating.

**Conclusions**

There are compelling reasons to use speech recognition in telephone menu systems. The interaction design is more flexible, and the resulting systems are easier to use, less distracting, and more enjoyable. With improvement in the audio transmission quality of cellular phones and probable advances in speech recognition technology, this technique will be even more desirable, but even now it is remarkably good. Therefore it is unsurprising that more and more automated telephone systems are incorporating speech recognition, and we expect to see this trend continue.