

---

## Defining voice search

### *Editor's Notes*

The term “voice search” is being used for a wide range of applications. How should we understand the term?

Applications most obviously related to the term “voice search” include using speech recognition and often text-to-speech synthesis to automate:

- Directory assistance (including local business directories), a search for a company by name or category or for a listing for an individual;
- A search for information such as news, stocks, traffic, driving directions, weather, or movie information, often by navigating a multi-level voice menu;
- A Web search analogous to what one gets by typing in a search term, but obtained by speaking the search term (sometimes resulting in a text list on a mobile phone); and
- The choice of options from a long list available in a mobile service, a portable device, or an automotive system, such as ring tones, songs, or radio channels.

All of these are a form of search initiated by a voice request, with confirmation or information delivery possibly using text-to-speech synthesis. A distinguishing characteristic of these core applications is quick access to information, minimizing complex navigation. The motivation for using voice is often because the device being used is a wireless phone, where text input is inconvenient or dangerously distracting and where it is difficult to scroll through long lists because of the small screen. The business model for the applications can span free services to maintain loyalty or build brands, subscription-fee services, ad-supported services, usage-based-fee services, and services resulting in a sales transaction.

Another category of applications that could be considered part of voice search is dialing by speaking a name—“voice dialing”—searching for a contact in a directory. Directory assistance is a form of dialing by name if the system allows completing the call once a listing is confirmed. Directory assistance uses a very large name list, while voice dialing on mobile phones typically uses a shorter list, usually a personal directory. Voice-dialing features in unified communications applications use a corporate directory, usually much larger than a personal directory. Since the distinction between these applications is simply directory size, they should all be considered a form of voice search. In a “Dialtone 2.0” world (SSN, tbd 2006, p. tbd), these voice-dialing options might be combined into a single layered application that started with a personal directory and worked its way up the complexity chain.

Another category of voice search starts with a text request rather than a voice request. Sometimes called “audio search,” this category searches audio (or the audio track of video) for speech and transforms it into searchable content. Applications include finding where in a video or podcast a specific topic or company is mentioned; enterprises can convert archived audio/video content into more usable information, and broadcasters can more easily use their archives. This subcategory covers the search of voice files for content using speech technology. When applied to call center recording or similar business audio, the term often used for these applications is “speech analytics,” which goes beyond the basic speech recognition to do some statistical and natural-language processing on the resulting text or phonetic representation. These applications often allow searching for specific audio files that have requested attributes (such as customer service calls that left the caller angry).

Conversion of voice files to text can go beyond searching for specific phrases or analytics. For example, some services are available today that convert voicemail to text for easier search and scanning, as well as to allow delivery of voice mail (or the “gist” of the voice mail) as an email or text message. Similarly, services that allow taking voice notes by phone and converting them to text are available. One of the primary objectives of such services is to make the voice content more easily archived and searched.

Searching for information is often only a step toward a transaction with a company. If one is searching for a book at a web site or a song on a music site, it is usually with intent to buy the book or song. If searching for information such as a stock price is voice search, is search for a bank account balance substantially different? The answer today is probably yes, it is different, given the specific requirements of contact centers to authenticate callers and other specialized functionality. Callers usually have to navigate a series of hurdles to get to the information they want, while the spirit of voice search is quick access to information or services. Voice search will get callers to a contact center faster and in larger volume. If a call center uses natural-language call routing, that part of the process fits the voice search objective of quick access, but is usually followed by a more structured interaction. Today the main impact of voice search on call centers is the likely increase in call volume and a change in the nature of calls when they are spawned by voice search applications. (See last month’s Editor’s Notes.)

“Voice search” is *not* defined by a specific technical approach other than the general category of speech recognition. Directory assistance requires dealing with a long item list and a large vocabulary that must be created automatically, while information access, for example, can be an expertly-tuned interactive dialog design designed to minimize the number of steps required to get to specific content. Searching audio given a text search term requires pre-processing that audio with a statistical language model to get a rough transcription, or a pre-processing technology that translates the audio into a phonetic-like representation.

Voice search can apply to all these categories, defined roughly as using speech recognition to achieve quick access to information or transactions. The maturing of the underlying speech technology has made the applications feasible on a commercial scale.

There is an alternative way to view “voice search” as an opportunity. Web applications such as text-based search and email owe much of their success to the ease with which text can be scanned quickly by the user—much faster than listening to a voice file—and later searched to find particular content or files. Voice search can be viewed as the effort to use speech technology to reduce or eliminate the difference between the way we use speech and the way we use text.

The equivalence of voice and text is becoming increasingly important as we attempt to use our mobile devices—which certainly started out as voice devices—for more than just making phone calls. When speech-recognition and text-to-speech applications are in the network, they are available to any phone, and work the same on any phone, with no need for the user to learn a new interface when buying a new wireless phone or calling from a landline phone. Text-based solutions will by necessity be dependent on the device’s characteristics. (Steve Jobs was recently quoted in the *Wall Street Journal* as saying that it takes a couple of weeks to get comfortable typing on the iPhone’s touch screen, after which it is superior to alternatives.) Unless someone gets a dominant market share, such as is the case in PC operating systems and web browsers, creating applications other than voice will tend to be device-specific efforts.

“Voice search” identifies a market opportunity in providing rapid access to information, particularly on mobile devices. It transfers some of the advantages of text to speech, while retaining the advantages of speech.