*InterVoice*BRITE

Speech-Enabled Interactive Voice Response Systems

Definition

Serving as a bridge between people and computer databases, interactive voice response systems (IVRs) connect telephone users with the information they need, from anywhere at any time. These systems have been around for more than a decade, and today they are used to support stock trade transactions, make travel arrangements, and manage bank accounts. Most of today's IVR and transaction-processing applications employ a touch-tone or dual-tone multifrequency (DTMF) user interface. However, applications that allow callers to use their own voice rather than DTMF inputs to complete transactions are rapidly emerging as the latest innovation in telephony-based remote self-service.

Overview

The quickening pace of adoption of speech solutions in the IVR industry is currently driven by improvements in speech algorithms, natural language processing, vocabulary management, and language modeling. In addition to these technological advancements, a number of related market factors are combining to make speech the IVR user interface of choice for today and tomorrow. This tutorial explores the current state of speech-enabled IVR applications, with emphasis on phonetic speech recognition, features and benefits, and development and deployment strategies.

Topics

1. Introduction
2. Current State of the Art for Speech-Enabled IVR Applications
3. Phonetic Speech Recognition
4. Features and Benefits of IVR Speech Solutions
5. Current Commercial Applications
6. Development and Deployment Strategy

   Self-Test

   Correct Answers

   Acronym Guide

1. Introduction

Today's IVR and transaction-processing applications allow self-service access to automated bank, portfolio, and account information, airline schedules, and movie times. Callers may also place orders, track order status, or use a directory to contact a department or individual. Automated speech recognition enhances the flexibility and power of these IVR applications.

The functionalities that can be delivered by today's IVR speech solutions vary widely and range from recognition of spoken letters and numbers to more complex phrases and sentences. Some applications simply replace touch-tone interfaces with speech-enabled applications that recognize a very limited set of spoken letters and numbers that primarily represent the touch keypad. More advanced applications employ directed dialogue or system prompts that guide users to respond with fairly simple spoken words that can be accurately recognized. Today's most advanced natural language applications enable recognition of more complex phrases and sentences spoken in a conversational manner at a natural speed.

Simple touch-tone replacement is faster and less expensive to develop than directed dialogue or natural language recognition applications. However, touch-tone replacement is also much less sophisticated and does not capture the navigational flexibility and increased call automation potential promised by speech recognition technologies (see *Figure 1*).

Figure 1. Speech Recognition Spectrum
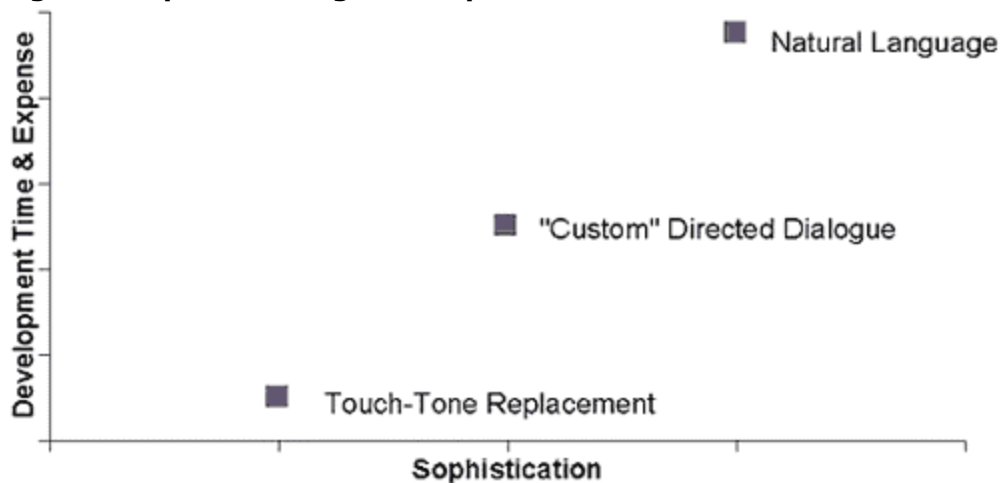
The following system prompts and caller responses illustrate the different speech-recognition capabilities in these types of applications.

- **touch-tone replacement—**
  System Prompt: "For checking information, press or say one."
  Caller Response: "One."

- **directed dialogue—**
  System Prompt: "Would you like checking account information or rate information?"
  Caller Response: "Checking", or "checking account," or "rates."

- **natural language—**
  SystemPrompt: "What transaction would you like to perform?"
  Caller Response: "Transfer $500 from checking to savings."

2. Current State of the Art for Speech-Enabled IVR Applications

The goal of speech-enabled applications has always been to allow callers to obtain information and perform transactions simply by speaking naturally. Recognition of free-form conversation is not yet a reality, and speech recognition has sometimes been over-hyped in the past. However, speech-recognition technology is now proving itself commercially viable in a number of customer service applications. Every day, thousands of callers are using their own voice to perform millions of business transactions.

Since the late 1980s, the U.S. government's Advanced Research Projects Agency has supported significant research in computer speech recognition and understanding systems. This research has improved recognition accuracy and increased capacity to recognize larger vocabularies.

The earliest speech-recognition engines provided only *discrete number recognition* and did not understand a caller's speech unless each word was separately and distinctly enunciated. Improvements in the recognition of continuous speech—words spoken in an unbroken string—now allow users to talk normally without artificial pauses between each word. *Continuous speech recognition* requires greater processing capacity and may require artificial intelligence (AI) or other associated technologies to ensure a high degree of accuracy.

Early speaker-dependent dictation systems had to be trained to understand the speech of one specific user. Now, speaker-independent recognition technologies allow IVR systems to interpret the speech of many users. Today's speech-enabled IVR applications serve large numbers of unknown callers without prior training of the system.

3. Phonetic Speech Recognition

Phonetic recognition systems now break down spoken words into small fundamental sound units called phonemes. When compared to word-based recognition systems, phonetic recognizers enable increased accuracy in understanding larger vocabularies. *Language modeling* technology is also used to heighten accuracy by comparing recognized sounds to a list of usage rules or constraints to determine the probability of one sound following another (see *Figure 2*).

Figure 2. Phonetic Speech Recognition
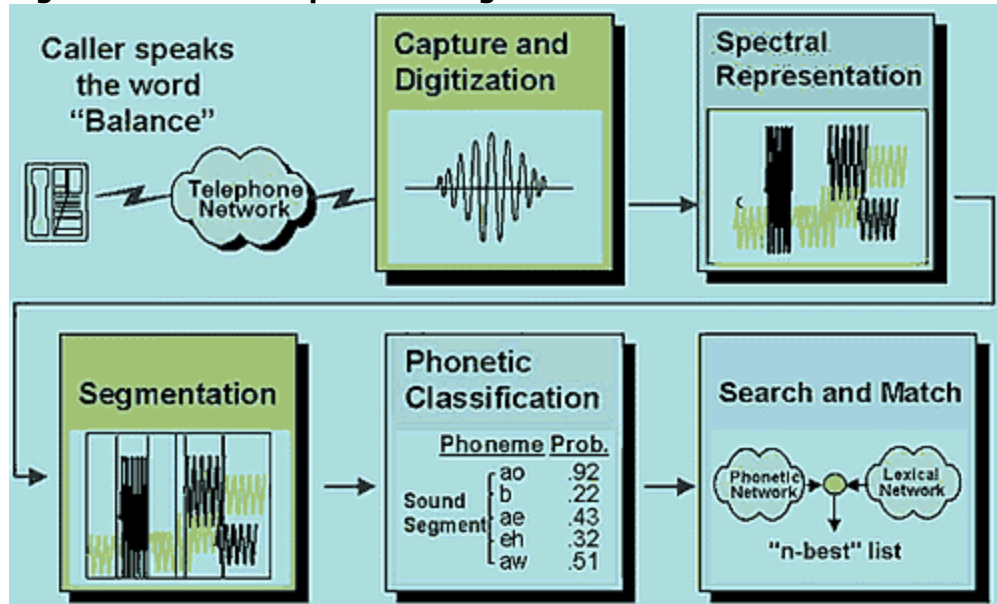
Speech solutions are now enabling the development of IVR applications that go beyond rigid touch-tone interface models to exploit the navigational flexibility offered by natural language processing. To aid recognition and reduce complexity, an "n-best list" of the most likely word or phrase matches to the spoken utterance and associated confidence scores are then created and interpreted by the system. Natural language recognition and advanced user interfaces that conduct interactive dialogues with users in order to complete transactions are driving the creation of the most versatile and robust applications ever developed for the IVR industry.

Why Are Many Customer Service Organizations Incorporating Speech Recognition into Their IVR Systems?

Many factors are driving the emergence of speech as the IVR user interface of choice for today and tomorrow. The first is spiraling labor costs. The cost of employing live customer service agents is rising at the same time that organizations are facing increased pressure to reduce the cost of serving customers. When an automated call-processing solution must be employed, a speech-enabled IVR application increases caller acceptance because it provides the friendliest and fastest self-service alternative to speaking with a customer service agent. Speech solutions also create new opportunities to automate transactions that are too cumbersome to complete using a DTMF interface, such as bill payment or stock trading.

Higher call volumes make the addition of speech recognition more cost-effective. In this era of mergers and consolidations, customer service organizations are often large enough to realize significant cost savings and a quick return on their investment in creating and deploying speech-enabled IVR applications.

Recognition of larger vocabularies is faster and more accurate with today's increased processing power, and hardware and software costs are now lower. Today's IVR systems are more robust and have the processing power required to support the latest speech recognition technologies. In addition, speech solutions for IVRs can be less costly and faster to implement than Web-based solutions.

The Gartner Group reports that an added advantage of voice recognition is the ease of adding verbal security steps to reduce risk with minimal customer impact. Speech recognition eliminates the need for touch-tone keypad entry and makes it easier to ask and answer traditional security questions, such as social security number or personal identification number (PIN), so that customers are not as bothered by them as they are on an IVR.

Speaker verification technologies also promote enhanced security by testing vocal characteristics of spoken utterances to verify the identity of the speaker. *Voice print* technology is now available that seamlessly verifies the identity of the speaker while it concurrently recognizes conversational language. Applications for this technology are especially attractive for providers of financial services and telecommunications because the information exchanged to complete their customer transactions is often valuable and therefore subject to fraud.

4. Features and Benefits of IVR Speech Solutions

What Benefits do IVR Speech Solutions Provide for Customer Service Organizations?

Speech recognition can deliver more than a simple upgrade or enhancement of touch-tone IVR applications. The value proposition for integration of speech recognition capabilities into an IVR environment is derived from the automation of transactions that are too tedious or impractical to perform using a touch-tone-only interface. A speech interface gives callers more flexible navigation options that are less complex and more intuitive than rigidly hierarchical touch-tone menu options. This provides faster, easier self-service transactions for customers and promotes increased customer satisfaction and retention.

However, the primary benefit that is motivating organizations to embrace speech solutions is the potential for dramatic reductions in operational costs. Speech solutions improve the productivity of customer service personnel because a higher percentage of customer calls can be fully or partially automated. Increased automation frees the customer service agent from many routine administrative tasks and reduces costs related to customer service staffing, as fewer agents are able to serve more customers. Thus, profitability improves when savings in overhead expenses are realized.

Flatter decision-tree menus and more direct access to the information or service that callers need will reduce the duration of an average call over time. This benefit is particularly important for organizations whose callers typically use toll-free access lines. Even a minor reduction in the length of an average call can produce significant savings in an organization's operational cost for telecommunications service.

The addition of speech recognition capabilities helps IVR system owners derive more benefit from their investment in existing IVR resources. Speech solutions typically run on the same platforms used for touch-tone applications and require minimal changes.

Resources that have been developed to support an Internet presence can support an IVR speech solution as well.

> Because most companies have or will build Web self-service applications, they've already got the business databases that can serve double duty for the speech recognition application. All they need is the speech interface, the vocabulary, and a connection to the business database.
> —*CIO Magazine*, June 15, 1999

Thus, organizations can use some of the same data modules built for an IVR speech-enabled application for their intranets. This could deliver a high degree of code reuse.

A speech-enabled interface can serve as the foundation for providing convenient new IVR services for customers as well as reduced operational costs, improved customer satisfaction and retention, increased return on investment, and a stronger market presence for the IVR service provider.

The latest speech solutions for IVRs feature the following:

- speaker-independent phonetic recognition of large vocabularies

- natural language and number recognition

- continuous alpha-numeric recognition

- building-block approach that reduces the complexity of application development

- barge-in capability

- configuration flexibility (client-server or integrated with telephony host)

- speaker-trained voice verification

- support for multiple languages

- full integration with other IVR features

5. Current Commercial Applications

How Commercially Viable Are Current Speech Technologies for IVRs?

Analysts and technology experts are increasingly vocal about the commercial potential of today's speech-recognition technologies. In June 1998, the Gartner Group announced that "speech-recognition technology finally works and is viable for customer service organizations. It is . . . an emerging 'self-serve' technology that will enhance customer service while reducing personnel costs."

Speech recognition technology has matured to the point that IVR systems can reliably and accurately recognize spoken responses more than 90 percent of the

time. "Speech recognition is accurate for all but the most risk averse environments," says Donna Fluss, an analyst with Gartner Group, Inc.

For IVR suppliers, the addition of speech recognition capabilities provides opportunities for incremental revenues, increases the variety of automated service delivered over the telephone, and increases the demand for telephony equipment.

Just as interactive touch-tone applications fueled the development of new businesses, speech recognition is creating new opportunities for imaginative entrepreneurs.

Businesses are currently using speech-enabled IVRs in the following ways:

- package tracking
- stock quotes and trading
- insurance claims
- travel booking
- pharmacy prescription refills
- restaurant reservation information
- banking
- social services administration and delivery by government agencies
- directory assistance

What are Industry Experts Predicting about the Adoption of Speech Recognition in the IVR Systems of the Future?

"Speech recognition has gone from the bleeding edge to the leading edge," says Brian Bischoff, general manager for AT&T Solutions Business Development. The top 25 percent of large companies are implementing their first-level speech applications, he says, because in the near term, speech recognition provides a bigger gain than even the Web. "That's how customers contact and interact with businesses" (*Information Week*, February 22, 1999).

William S. Miesel, a leading speech industry consultant and president of TMA Associates, noted in 1998 that the availability of speech recognition on the

platform is already becoming an important customer criterion for the purchase of an IVR or computer telephony system. Fluss estimates that by 2003, 30 percent of the new automated lines in call centers will respond to customers' speech.

> Often, when people begin to interact with conversational [IVR] systems, they tend to be conversational in return. Consequently, a good speech interface is one that will steer them into providing only responses that the system can recognize.
> —*Speech Technology Magazine*, May/April, 1999

These systems use directed dialogue to constrain caller responses to those that will be easily and accurately understood. Fluss says that a directed speech recognition application will pay for itself within nine to eighteen months in a call center with more than fifty agents.

In his *Telephony Voice User Interface* report, Miesel predicted that total worldwide revenues from advanced speech technology products and services in telephony (including speech recognition, speaker verification, and text-to-speech) will exceed $38 billion by 2003.

6. Development and Deployment Strategy

Design: Project Scope and System Architecture, Speech-User Interface Design

First, explore your business objectives and goals, and then evaluate how callers will use your system and what type of transactions it must support. Using this insight, create a list of the development tasks required to meet your objectives. Summarize the scope of your project and draft a high-level system architecture plan. Next, determine what applications will need to be developed and create specifications for development of your speech-enabled application, including prompt and playback messages and interfaces with host information systems. Plan what will be required to design a user interface that ensures a smooth, gradual transition to your speech-enabled environment. When your plan is complete, assign development responsibilities and obtain budgetary quotes for all planned project activities.

Develop: Prototype Application, Production Application

Develop a prototype application that will help you demonstrate and evaluate how users react to your new system. Observe and assess the experiences of a select group of testers as they access sample data and a subset of your system's ultimate functionality. After the testing of your prototype is completed, develop all of the

features your production application will offer, including connectivity with data, telephony, and information systems in your environment.

Your application should include prompts that automatically provide callers with tips on how to get more direct access to a specific transaction the next time. Your callers will appreciate faster service, and you will save on customer service costs as the duration of an average call falls over time.

Test: Prototype Evaluation for Usability, Production Pilot Testing

Reduce risk by testing the usability of your prototype speech application before it is deployed throughout your enterprise. To prepare your system for introduction to targeted end users, make your prototype application available to selected callers involved in your industry or to customers participating in focus groups. Evaluate the experience of these minimally trained users in realistic, observed settings. Solicit caller feedback using surveys or other tools and plan usability improvements based on their suggestions.

Drawing on the results of your system's usability and prototype tests, implement a preliminary, limited deployment of your application. The pilot test serves as the foundation for the full-scale production of your speech solution. It verifies that your prototype system is accurately and reliably accessing back-end telephone, data, or information hosts as needed. It also aids in the assessment of the full functionality of your application to ensure that it is operating at peak efficiency.

For the first time, allow live, untrained callers to interact with your system as it actually operates in your business. Then evaluate call activity during the initial period of use. The pilot test is useful in identifying desired improvements for speech vocabulary libraries, prompts, recognition contexts designed to support transactions required by your customers, and other application functions.

Deploy: Production Deployment, User Interface Tuning

After your speech-enabled IVR system is fully deployed, there is still important work to be done. Your system must be monitored, and the user interface and vocabulary must be tuned to the production environment to ensure that your speech solution is consistently satisfying callers and meeting your initial business objectives.

Whenever you deploy significant system upgrades or functionality enhancements, maintaining system reliability can be a challenge. Evaluate how

you can support fully functional speech-enabled IVR applications with minimal impact on system reliability.

Monitor: Post-Production Evaluation, Usability Monitoring

To monitor and evaluate the operation of your speech solution, track critical usability indicators such as recognition accuracy and transaction completion rates. Do these demonstrate solid application design and an effective user interface? To answer this question, create usage reports and detailed call statistics. Design evaluation tools such as customer satisfaction surveys. Continually strive to improve the usability of your speech-enabled application and fine-tune your system for maximum performance (see *Figure 3*).

Figure 3. Continuous Improvement Cycle
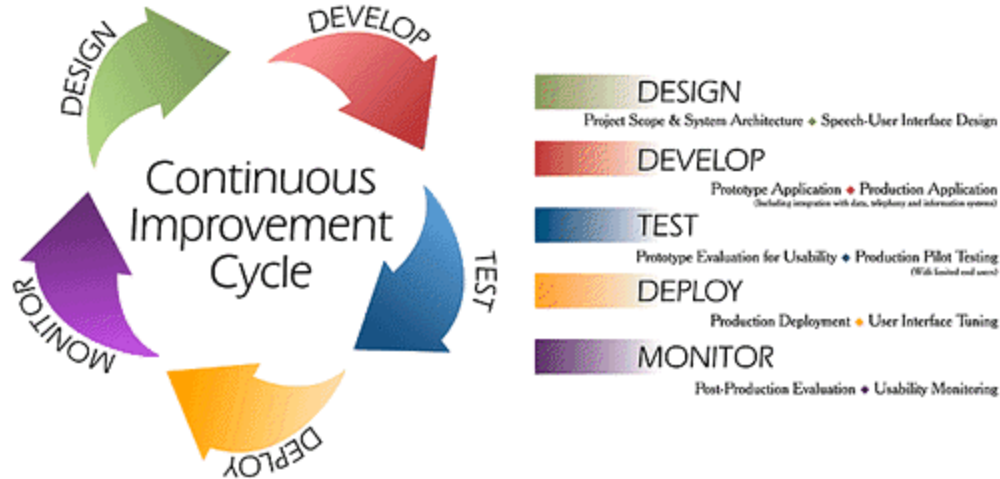
If I must outsource a part or all of the development and deployment of the speech solution for my IVR, how should I select a reliable partner?

1. Has your potential speech partner established relationships with the best-of-breed speech technology developers? Are their experts qualified to help you select the best speech recognition technology for your specific application?

2. Does your potential speech partner focus primarily on the development and deployment of speech applications for IVR environments? Is automated speech recognition fully integrated into the platform?

3. Does your potential speech partner have the skills and experience required to prevent risk to your mission-critical systems during development and deployment of your speech solution?

4. Does your potential speech partner have extensive expertise in developing applications for the IVR platform of your choice?

5. Does your potential speech partner offer comprehensive design, development, and deployment services that span the entire life cycle of your solution?

Self–Test

1. IVR systems are a brand-new technology, developed and deployed within the last five years.

   a. true

   b. false

2. The most advanced interface currently available for IVR applications is based on natural language recognition.

   a. true

   b. false

3. Touch-tone replacement is slower and more expensive to develop than directed dialogue or natural language recognition applications.

   a. true

   b. false

4. Directed dialogue and natural language applications are more sophisticated than touch-tone replacement.

   a. true

   b. false

5. Which of the following is *not* a reason that service organizations are incorporating speech recognition into their IVR systems?

   a. spiraling labor costs

   b. lower call volumes

c. automated transactions

d. increased processing power

6. Speech solutions for IVRs are accessible to a larger number of customers than Web-based solutions.

    a. true

    b. false

7. The earliest, least-sophisticated form of speech recognition is _____.

    a. natural language modeling

    b. barge-in capability

    c. discrete number recognition

    d. support for multiple languages

8. Which of the following is the primary benefit of IVR speech solutions for customer service organizations?

    a. automation of transactions

    b. reduction in average call time

    c. minimal change to existing platforms

    d. reduction in operational costs

9. Speech recognition has matured to the point that IVR systems can reliably and accurately recognize spoken responses more than _____ percent of the time.

    a. 50

    b. 60

    c. 70

    d. 90

10. What tasks in what order are necessary for speech-enabled IVR applications?

    a. design, test, develop, deploy, monitor

b. design, develop, deploy, test, monitor

c. design, develop, test, deploy, monitor

d. design, develop, deploy, monitor, test

Correct Answers

1. IVR systems are a brand-new technology, developed and deployed within the last five years.

   a. true

   b. false

   See Definition and Overview.

2. The most advanced interface currently available for IVR applications is based on natural language recognition.

   a. true

   b. false

   See Topic 1.

3. Touch-tone replacement is slower and more expensive to develop than directed dialogue or natural language recognition applications.

   a. true

   b. false

   See Topic 1.

4. Directed dialogue and natural language applications are more sophisticated than touch-tone replacement.

   a. true

   b. false

   See Topic 1.

5. Which of the following is *not* a reason that service organizations are incorporating speech recognition into their IVR systems?

    a. spiraling labor costs

    b. lower call volumes

    c. automated transactions

    d. increased processing power

    See Topic 3.

6. Speech solutions for IVRs are accessible to a larger number of customers than Web-based solutions.

    a. true

    b. false

    See Topic 3.

7. The earliest, least-sophisticated form of speech recognition is_____.

    a. natural language modeling

    b. barge-in capability

    c. discrete number recognition

    d. support for multiple languages

    See Topic 1.

8. Which of the following is the primary benefit of IVR speech solutions for customer service organizations?

    a. automation of transactions

    b. reduction in average call time

    c. minimal change to existing platforms

    d. reduction in operational costs

    See Topic 4.

9. Speech recognition has matured to the point that IVR systems can reliably and accurately recognize spoken responses more than _____ percent of the time.

    a. 50

    b. 60

    c. 70

    d. 90

    See Topic 5.

10. What tasks in what order are necessary for speech-enabled IVR applications?

    a. design, test, develop, deploy, monitor

    b. design, develop, deploy, test, monitor

    c. design, develop, test, deploy, monitor

    d. design, develop, deploy, monitor, test

    See Topic 6.

Acronym Guide

ASR
automated speech recognition

DTMF
dual-tone multifrequency

IVR
interactive voice response