SPEECH RECOGNITION: FROM THE LABORATORY TO THE REAL WORLD

Jay G. Wilpon, Rajendra P. Mikkilineni, David B. Roe, and Sedat Gokcen

Jay G. Wilpon *and* ***David B. Roe*** *are with AT&T Bell Laboratories at Murray Hill, New Jersey.* ***Rajendra P. Mikkilineni*** *and* ***Sedat Gokcen*** *are with AT&T Bell Laboratories at Columbus, Ohio. Mr. Wilpon is a member of technical staff in the Speech Research Department, where he works on word-spotting techniques, training procedures for speaker-dependent and speaker-independent word recognition systems, speech detection algorithms, and implementing speech recognition systems for general use in the telephone network. He has a B.S. in mathematics and A.B. in economics from Lafayette College, Easton, Pennsylvania, and an M.S. in electrical engineering/computer science from Stevens Institute of Technology, Hoboken, New Jersey. He joined AT&T in 1977. Mr. Mikkilineni is a distinguished member of technical* (continued on page 23)

The art and science of speech recognition research have advanced to the point where it is now possible to communicate reliably with a machine via telephone to carry out simple tasks. However, the creation of robust algorithms is only part of the overall process of making speech recognition technology a commercial reality. In this paper, we present a brief overview of speech recognition technology and discuss the implementation of the algorithms with digital signal processing chips. In addition, we will show how theory and practice come together in real-world conditions by describing a telephone network trial of automatic credit card verification using speech recognition technology.

Introduction

The ability to recognize speech automatically has long been sought by AT&T Bell Laboratories[1-3] as well as other laboratories.[4] The use of such technology—for example, for automating portions of operator-assisted calls or verifying credit card transactions—will greatly enhance current telephone network-based services. In addition, automatic speech recognition can be used to create a wide variety of new services, from telemarketing (voice response and audiotex services) to office automation (a voice calendar system).

Although automatic recognition of fluently spoken, natural speech remains elusive, a great deal of fundamental knowledge has been obtained. Three significant advances in algorithm development illustrate the rapid progress over the past five years:

- Speaker independence: recognizing any speaker or dialect without specific training
- Continuous speech recognition: understanding fluent sentences rather than distinctly spoken words
- Word spotting: focusing on the key words in casual speech.

These algorithms have performed exceptionally well under laboratory conditions, and are the basis for AT&T's leadership in speech recognition technology.

Panel 1. Acronyms in This Paper

ACTR	automatic call type recognition
CAMA	centralized automatic message accounting
DSP	digital signal processor
HMM	hidden Markov model
I/O	input/output
ONI	operator number intercept
OSPS	operator services position system
SP	speech processor
VIS	voice information system

In addition to the progress in speech recognition research, there have been many advances in digital signal processor (DSP) technology. AT&T is a leader in DSPs, with powerful floating-point signal processors such as the WE® DSP-32C and fast, cost-effective fixed-point processors like the WE DSP-16.[5] These fast, flexible DSP chips enable advanced research algorithms to be implemented on real-time, cost-effective hardware for integration into commercial applications.

With these successes in the research laboratory and in DSP design, AT&T has conducted several field trials to show practical uses of speech recognition technology in real-world applications. The most sophisticated of these is a fully automatic credit card verification system in which merchants wishing to validate a credit card number call the credit card company. Instead of speaking to a human attendant, merchants speak the credit card number and the dollar amount of the transaction to a speech recognizer and receive an authorization for the transaction.

This paper focuses on three main areas:

- Advances in speech recognition research, particularly recent research on speech algorithms at AT&T Bell Laboratories.
- DSP technology as the basis for real-time speech recognition systems, particularly the AT&T Conversant® voice information system (VIS).
- Field trials of speech recognition applications, both previous telephone network experiments and the recent trial of a credit card verification system.

These topics are interrelated because real-time hardware is needed to run field trials and because problems encountered in the field become the basis of research into new algorithms. The paper closes with some observations on future trends in speech recognition applications.

Overview of Research in Speech Recognition

The basic concept of speech recognition is shown in Figure 1. In the pattern-matching approach to speech recognition, features extracted from speech are compared to previously stored patterns, or models. The model may represent either a word or a smaller subword unit such as a phoneme. The model with the closest match to the features is chosen as the recognized result, subject to any constraints (such as grammar or semantics) of the postprocessing step.

The most popular and successful speech recognition algorithms are based on the theory of hidden Markov modeling. Because of its rich statistical properties, hidden Markov model (HMM) speech recognition is generally considered to be more powerful than the older template-based recognition techniques. For a tutorial review of HMMs, see Reference 3.

Most of the speech recognition algorithms at laboratories throughout the world follow the pattern-matching approach of Figure 1, though the details of the various components vary widely.[4] Bell Laboratories has had considerable success with a word-based HMM speech recognizer[2] that is the basis for several applications of connected-speech recognition, speaker-independent recognition, and word spotting. Now we present more details of this particular algorithm, following Figure 1:

- *Spectral analysis.* The digitized speech signal is converted to a set of cepstral vectors and cepstral-derivative vectors. (The ***cepstrum*** is a measure that models the frequency content of speech signals.) This parameterization of the speech signal has been shown

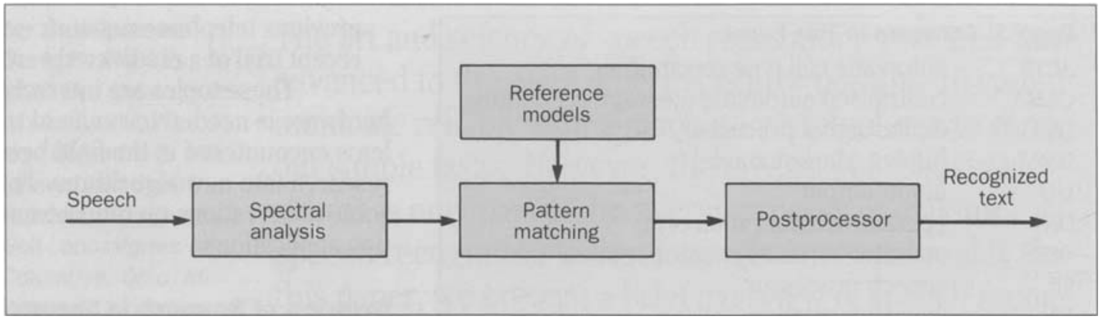


Figure 1. Basic concept of speech recognition. Features in speech are automatically extracted and compared with previously established reference patterns.

to yield high recognition performance.[2]

- *Reference models.* Each reference pattern represents a word as a linear sequence of Markov model states. Reference patterns are statistically derived from previously recorded speech data, and consist of means and variances of the speech features for each Markov state. For speaker-independent recognition, speech from many diverse dialects is used to generate one or more models. For word spotting, background noise and extraneous speech are also modeled.
- *Pattern matching.* The sequence of spectral vectors of the unknown speech signal is matched against the HMMs. A dynamic programming procedure known as the Viterbi algorithm aligns, in time, the input utterance with the set of reference models. The result of this computation is a probability that each of the models matches the incoming features.
- *Postprocessor.* The potential words from the pattern-matching process are subjected to further validity tests, such as comparison to duration rules, which are used to eliminate unreasonable candidates. In connected-word recognition, the postprocessor chooses the most likely sentence among the remaining candidates.

This algorithm has achieved speaker-independent recognition accuracies as high as 99.6 percent per word on phrases of fluently spoken digits.[2] The algorithm has had excellent success on an English sentence task with an 80-word vocabulary in the HuMaNet project.[6] It is also the basis of the more recent field trials described later in this paper.

A basic assumption for most speech recognition systems is that the speech to be recognized consists solely of words from a predefined vocabulary. For speech recognition applications in the telephone network, where any person can pick up any telephone at any time from anywhere, it is naive to assume that users will adhere strictly to this protocol. Studies on recognition of a limited set of isolated command phrases for automating operator-assisted calls[7] have confirmed that it is extremely difficult, if not impossible, to get customers to speak only the allowed words. For example, when customers are asked to say only *collect, calling-card, third-number, person,* or *operator*, they may instead say, "I want to make a *collect* call please." With these constraints, the burden of correct speech falls on the users.

We should shift the burden from the users to the speech recognition algorithms themselves. Recent AT&T systems are able to recognize *key words* from a vocabulary list spoken in an unconstrained fashion.[7] The novelty of this approach is that statistical models are created for both the vocabulary words and the extraneous speech, including background noises. As discussed under "Speech Recognition and Credit Card Verification," this word-spotting technique achieved good results either for correct, isolated speech or when customers ignored the instructions and spoke a sentence.

The current frontier of speech recognition research is recognizing speech from vocabularies of a thousand words or more. Having achieved success in the laboratory with systems that require isolated word input or systems trained to individual speakers, most researchers are aiming for recognition of fluent speech (*continuous-speech recognition*) by any talker.[4] Additional research topics include recognition of speech by neural networks, investigation of language modeling as an aid to speech recognition, studies of the changes in speech with background noise and variations in speaking style, and research on speaker verification and identification.

A key issue in developing large-vocabulary recognizers is the choice of the basic speech recognition unit. In speech recognition systems where the vocabulary size is relatively small (less than 100 words), the basic unit chosen is usually the whole word. But it seems intellectually appealing that we should model a more fundamental unit of speech sound, for example the phoneme, in order to recognize speech. Moreover, for large vocabularies, it is impractical to record enough speech to cover many people speaking each of 1000 or more words appearing in each possible phonetic context. Such a training set would be prohibitively large. Therefore, most large vocabulary recognition algorithms use *subword* models, such as phonemes, diphones, or syllables, as their basic recognition units. With this approach, 95 percent word recognition accuracy and 71 percent sentence accuracy have been reported in tests on a 1000-word task (speaker-independent continuous speech).[9]

DSP-Based Speech Recognition Systems

Without question, the commercialization of speech recognition has been made possible by the speed and power of low-cost digital signal processing chips. Speech recognition, like speech compression and synthesis, is computation-intensive, requiring tens of millions of computations per second even for modest speaker-independent tasks. In 1980, the fastest available digital signal processor chip was the AT&T WE DSP-10 chip, which executed 2.5 million instructions per second. Today the AT&T WE DSP-16 executes 30 million instructions per second. The result is more than an order-of-magnitude reduction in the cost of speech recognition systems.

In 1986 AT&T announced a voice-activated dialer for its mobile telephones. The DSP-based low-cost system used HMM technology with filter bank feature analysis.[10] This voice dialer would recognize up to 50 names in a personal directory and was marketed by AT&T before it withdrew from the cellular telephone business. The directory was programmed by an individual user by speaking each name two or three times and entering the phone number. Special care was needed because of the high noise levels in a moving automobile.

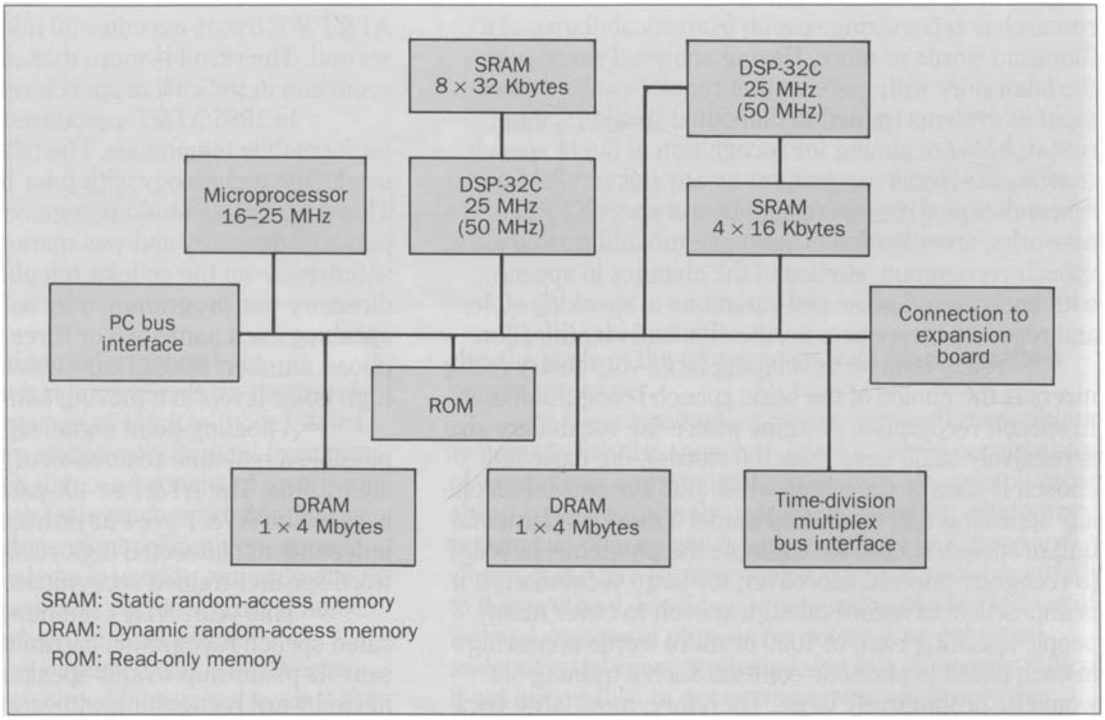
A floating-point digital signal processor made possible a real-time realization of high-accuracy HMM algorithms. The AT&T BT-100 parallel processor, which is based on AT&T's DSP-32 processor, can run speaker-independent connected digit recognition[11] as well as 1000 word speaker-trained recognition for complex tasks.[12]

This year, AT&T announced its most sophisticated speech recognition capability to date. The Conversant VIS platform provides speaker-independent connected word recognition with word-spotting capability.

The Conversant Voice Information System. The Conversant VIS platform is based on AT&T 6386 computers running the UNIX operating system. (UNIX is a registered trademark of UNIX System Laboratories Inc.) The basic system supports 48 simultaneous voice-response applications. In the system's more sophisticated form, many of these computers may be networked together for applications requiring more than 10,000 simultaneous users. Most applications of this successful product have used Touch-Tone input and recorded voice response, as discussed in the paper by D. R. Fischell, S. S. Kanwal, and D. S. Furman.[13]

This platform has several processor boards, including analog and digital network interface boards, and a general-purpose speech processor (SP) board. The SP board consists of a control processor, 256 digital voice-data input-output (I/O) channels, and two DSP-32C

Figure 2. Conversant system signal processing board for speech. The versatile board can run a wide variety of speech algorithms—speech recognition as well as voice response and speech synthesis.



processors (Figure 2). (See the paper by R. J. Perdue, J. Leikness, and J. B. Sharp for a more detailed discussion of the architecture.[14]) This board, with its flexible architecture, is capable of running many different speech algorithms including speech coding, text-to-speech synthesis, speaker verification, and speech recognition. The speech-recognition algorithms include both speaker-trained and speaker-independent routines, codes for isolated-word and continuous speech with grammars, and word spotting. The board can run speech recognition simultaneously on two channels of input speech.

To handle large numbers of channels (simultaneous users) and/or large recognition tasks, more processing power is needed. The companion board, consisting of 12 DSP-32Cs and a common pool of dynamic memory of up to 16 megabytes (MB), was built to augment the capabilities of the SP board. A cluster of up to eight companion boards, providing up to 2.4 gigaFLOPS (2.4×10^9 floating-point operations per second) of processing power, can be connected to an SP board. For the credit card authorization application, a Conversant VIS with one companion board handles six simultaneous transactions, while, for the automated operator application, up to 30 simultaneous users can be accommodated.

Telephone Network Trials of Speech Recognition

Some speech recognition systems are notorious for degrading in field conditions, especially when they are designed and trained on speech recorded in the laboratory. For demonstrations in the laboratory, users may be given instructions on how to speak and on limitations of the recognition system. But conditions in the

Table I. AT&T Speech Recognition Trials in the Telephone Network

Trial	Date	Technology	Vocabulary	Accuracy
CAMA/ONI telephone number input	1982	Template-based, isolated words	11 words (0-9, oh)	93% per word[15]
Amtrak call routing	1986	↓	4 words (1-4)	95% per word[16]
Stock price quotes	1985-86	↓	11 words	93% per string[17]
OSPS/ACTR	1985-87	↓	5 words	94% per word (isolated inputs)[7]
automated operator services		HMM with	5 words	98.3% per word (*all* inputs)[8]
Credit card verification	1989-90	word spotting	11 words	97% per 10- and 15-digit string

telephone network are far less favorable than these ideal laboratory conditions. Background noise, extraneous speech, different handset types, and variations in speech level can cause a dramatic increase in speech recognition errors. Table I lists speech recognition trials that have been run in the AT&T telephone network. In this section, we will briefly review several of these trials.

The first telephone network speech recognition trial with AT&T telephone customers was in 1982. In the trial, customers were asked to speak their telephone numbers as sequences of isolated digits.[15] The accuracy achieved in this trial was about 93 percent per word. Later, in 1986, customers of Amtrak were able access train timetables and general information by speaking a single digit.[16] Depending on the user's response, the call would be routed to the appropriate operator. About 95 percent of the utterances were recognized correctly. Both of these trials used isolated-word, template-based speech recognition technology.

In 1985 a trial was performed with an investment firm to automate a stock quote service. In this trial, each stock was assigned a 9-digit code specially designed to correct recognition errors (based on Reed-Solomon coding techniques). The recognition system used a hybrid between isolated-word and connected word technology and yielded a string accuracy of about 93 percent.[17]

Also in 1985, a series of trials was initiated to evaluate speech recognition performance and user protocols of a feature called *automatic call type recognition* (ACTR). This feature used speech recognition to give telephone customers the option of verbally identifying the type of call they wished to make. The task required that users speak, in isolation, one of five defined vocabulary words (*collect, calling-card, third-number, person,* or *operator*). However, observations of customer responses during the trials showed that about 20 percent of the utterances had the desired vocabulary word along with extraneous input that ranged from nonspeech sounds to groups of nonvocabulary words (e.g., "I want to make a *collect* call please"). This led to the development of the word-spotting technique described above. In formal testing using the ACTR trial database, word recognition accuracies were 99.3 percent on purely isolated speech (i.e., only vocabulary items and background noise were present), and 95.1 percent when the vocabulary word was embedded in unconstrained speech.[8]

In the next section, we will discuss in depth the credit card authorization trial.

Speech Recognition and Credit Card Verification

The connected-word speech recognition algorithm implemented on the Conversant VIS was tested in a field trial of a credit authorization service for a major credit card company.

This credit authorization service is currently provided by the credit card company to authorize retail sales

Table II. Customer Responses with Extraneous Speech

Response Type	Percentage	
	Human-Human	Machine-Human
Merchant	37	10
Card	29	6
Dollar amount	n.a.	7

n.a.: not available

transactions involving its credit card. A merchant calls the credit authorization service to get an authorization code for a transaction. A typical transaction between a merchant and an attendant is:

- *Attendant:* Merchant number please.
- *Merchant:* 1234567899.
- *Attendant:* Card number?
- *Merchant:* 344121213456789.
- *Attendant:* Amount?
- *Merchant:* Two hundred and forty five dollars.

The attendant types the merchant number, card number, and the dollar amount into the computer. The computer processes this information and completes the call by "speaking" the authorization code to the merchant. Thus, the attendant's only function is to listen to the merchant and type the data into the computer. A previous attempt by the credit card company to automate this service with a Touch-Tone input voice response system was unsuccessful, in part because merchants were accustomed to the current system, in which they can speak.

The Conversant VIS with speech recognition automates the function of the attendant in this transaction. Though most of the spoken transaction consists of connected digits, the dollar amount involves natural numbers (ten, eleven, hundred, etc.). The recognition system must be able to recognize these words for the dollar amount portion of the transaction.

Transaction Design. Speech recognition technology performs better if the input speech contains no extraneous speech. In order to take full advantage of speech recognition technology, the users of the system should be instructed to speak only the speech recognition vocabulary words during the recording interval. In this trial, it was not possible to train the merchants or to send them written instructions on the use of the system. Therefore, the speech responses contained background noise and/or extraneous speech (e.g., "My card number is *1234*"). Consequently, an early phase of the trial was dedicated to designing a protocol, consisting of a set of speech prompts, which minimizes the extraneous words spoken by the merchants.[18]

The quality of the speech responses was defined as the percentage of responses with extraneous speech. First the extraneous speech was measured as a human attendant handled the calls. Then a protocol using recorded prompts was designed and implemented on the trial system. The quality of the responses was measured again for this human-machine protocol. The results are given in Table II. The number of responses with extraneous speech was reduced from 37 percent to 10 percent for the merchant number responses and from 29 percent to 6 percent for the card number responses. The improved transaction design helps reduce the extraneous speech but does not eliminate it. Therefore, the word-spotting technique was incorporated into the Conversant VIS platform and used during the trial.

Speech Recognition Results. The trial system prompted the user for the transaction information, collected the raw speech data, performed speech recognition, and recorded the digit strings typed by the attendant. An attendant monitored the transaction, transcribed the speech responses, typed the data into the computer, and sent a copy of the data to the Conversant VIS. An estimate of the recognition accuracy was made by comparing the result of the speech recognition system with the attendant's result. We present the results of our analysis of 657 transactions handled by the system during the last week of February 1990.

Recognition of Merchant and Card Numbers. The merchant and card numbers have encoded in them a security

Table III. Measured Speech Accuracy—Trial System

Number	First prompt Success (%)	First prompt Fail (%)	First prompt Reject (%)	With reprompt Success (%)
10-digit merchant number	81.6	4.1	14.3	90.7
15-digit card number	82.7	5.1	12.2	89.4
Dollar amount with natural numbers	70.9	29.1	0.0	85.4

check digit. This was used by the recognizer to determine the validity of the digit sequence. The system reprompted the merchant if the recognized digit string did not satisfy the security check digit test (called a *checksum* test).

If (after reprompting) the recognizer's output did not satisfy the checksum, the recognition outcome was counted as a reject. If the output string satisfied the checksum, it was compared with the digit string keyed in by the attendant. The recognition outcome was counted as a success if these two strings matched and as a failure if there was a mismatch between these strings.

The recognition rates for the merchant and card numbers are presented in Table III. On the first attempt, the system had a success rate of 81.6 percent and a failure rate of 4.1 percent on the merchant numbers. The remaining portion, 14.3 percent, was rejected by the system. Of these rejections, 64.0 percent were corrected by reprompting the merchant and performing the speech recognition on the new response. The overall accuracy of the system with a reprompt was 90.7 percent for the merchant numbers. For the card numbers, the system achieved an accuracy of 82.7 percent on the first attempt and an overall accuracy of 89.4 percent.

Subsequent analysis of the field data indicates that the recognition accuracy could be improved significantly (to 97 percent accuracy on the first attempt for both merchant and credit card inputs) by using improved algorithms still in the research laboratory. These algorithms will be implemented on the production system, which will use the new companion/SP system described in the previous section.

Recognition of Dollar Amount. The merchants were not given any instructions in the specification of the dollar amount. Consequently, it was spoken using natural numbers like *one hundred and twenty five dollars and twenty five cents.* The credit card company was interested only in the dollar amount, in this case, *one hundred and twenty five dollars.* Consequently, the system truncated the recognition result containing dollars and cents to dollars only, then compared the result to the dollar amount recorded by the attendant. The results of this comparison are also shown in Table III. The system had a success rate of 70.9 percent on the dollar amount on the first try. The system prompted the merchant to verify its result by responding with "yes" or "no." If the response from the merchant was "no," the system reprompted the merchant for the dollar amount. With reprompting, the system achieved an overall accuracy of 85.4 percent on the dollar amount.

The Future of Speech Recognition Technology

With the recent improvements in speech recognition accuracy, we are confident that speech recognition will find several significant applications within the next few years. Examples of applications that are within reach are banking by phone, operator services, and stock-market inquiries. The growth of voice response systems[13] can naturally be extended to allow speech input instead of Touch-Tone signals.

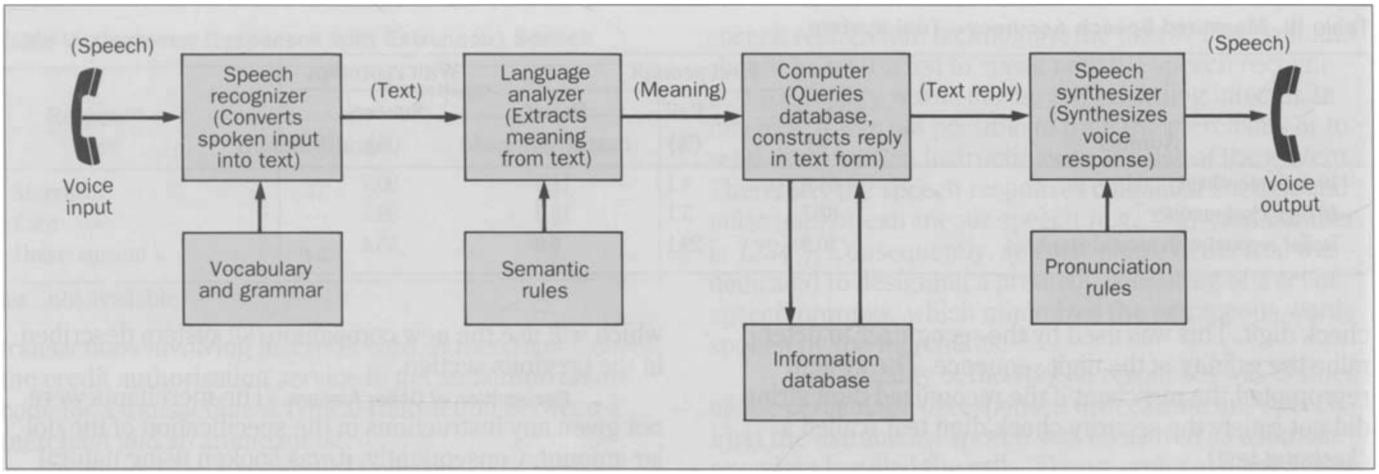


Figure 3. Voice conversation with a computer. Technology for fluent conversations—in real time—will lead to a wide variety of information-access applications.

It is important to bear in mind that speech recognition, notwithstanding advances in reliability, remains error-prone. For this reason, the most successful products and services will be those that have the following characteristics:

- *Simplicity:* Successful services will be natural to use—for instance, they may be menu-driven with small vocabularies.
- *Evolutionary growth:* The first systems are likely to be extensions of existing systems—for example, voice response systems.
- *Tolerance of errors:* Given that any speech recognizer will make occasional errors, the inconvenience to the user should be minimized.

Looking further into the future, we can make some guesses about new capabilities, based on ongoing research, that will open up more ambitious applications. Of course, no one can predict if or when a breakthrough will occur.

- *Subword units:* It will be possible to build a dictionary of models composed of constituent phonetic models, first for small, easily distinguishable vocabularies, and later for large vocabularies. Thus, the effort and expense of gathering speech from many talkers for each vocabulary word will be eliminated.
- *Noise immunity:* Better speech-enhancement algorithms will make speech recognizers more accurate in noisy environments.
- *Language understanding:* The ability to spot key words in a phrase is the first step toward understanding the essence of a sentence, even if some words are not recognized.
- *Speaker adaptation:* People can adapt quickly to dialects and accents in speech. Machines could respond more accurately if they had a similar learning capability.

The ultimate long-range goal is to converse fluently with a computer. Figure 3 is a block diagram of a conversational computer comprising a speech recognizer, a language analyzer, a computer with a database, and a speech synthesizer. A voice interaction with a computer has two advantages: First, speech is the most natural means of human communication. Second, *remote* access to computers over the telephone is possible because no keyboard or display is needed. Indeed, the desire for remote access to information is the driving force behind the success of voice response systems using Touch-Tone input. The concept of a fully conversational computer is indeed powerful because of its general applicability to a number of information-access applications.

Conversing with a computer is commonplace in science fiction—for instance, Captain Kirk of Star Trek* talking to the computer of the Starship Enterprise*. (*Trademarks of Paramount Pictures.) But practical systems for conversing with computers about elementary tasks are no longer just science fiction; they are an emerging reality. As speech recognition technology progresses, we will see the evolution of systems—simple at first but becoming more powerful—that let us talk to machines as easily as we converse with other people.

Acknowledgments

The authors are grateful to Larry Rabiner and Bob Perdue for their insightful comments during preparation of this paper. In addition, we thank our colleagues Roy Grubbe, Maureen McGee, and Padma Ramesh for their many contributions to the credit card authorization trial.

References

1. L. R. Rabiner and S. E. Levinson, "Isolated and Connected Word Recognition—Theory and Selected Applications," *IEEE Transactions on Communications*, Vol. COM-29, No. 5, May 1981, pp. 621-659.
2. L. R. Rabiner, J. G. Wilpon, and F. K. Soong, "High Performance Connected Digit Recognition Using Hidden Markov Models," *Conference Record, IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, New York, April 1988, pp. 119-122.
3. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, Vol. 77, No. 2, February 1989, pp. 257-286.
4. J. Mariani, "Recent Advances in Speech Recognition," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, May 1989, pp. 429-440.
5. M. L. Fuccio et al., "The DSP32C: AT&T's Second-Generation Floating-Point Digital Signal Processor," *IEEE Micro*, December 1988, pp. 31-48.
6. D. A. Berkeley and J. L. Flanagan, "HuMaNet: An Experimental Human/Machine Communication Network Based on ISDN," *AT&T Technical Journal*, Vol. 69, No. 5, September/October 1990, pp. 87-99.
7. J. G. Wilpon et al., "Isolated Word Recognition over the DDD Telephone Network—Results of Two Extensive Field Studies," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, April 1988, pp. 55-57.
8. J. G. Wilpon et al., "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, November 1990.
9. C. H. Lee, "Acoustic Modeling of Subword Units for Speech Recognition," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, New Mexico, April 1990, pp. 721-724.
10. P. K. White, W. P. Halas, and J. H. Davis, "Converging Technologies: Automotive, Cellular, and Voice I/O," *Proceedings of International Congress on Transportation Electronics*, Society of Automotive Engineers, Dearborn, Michigan, October 1986, pp. 261-264.
11. A. L. Gorin and D. B. Roe, "Parallel Level-Building on a Tree Machine," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, April 1988, pp. 295-298.
12. D. B. Roe, A. L. Gorin, and P. Ramesh, "Incorporating Syntax into the Level-Building Algorithm on a Tree-Structured Parallel Computer," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, May 1989, pp. 778-781.
13. D. R. Fischell, S. S. Kanwal, and D. S. Furman, "Interactive Voice Technology Applications," *AT&T Technical Journal*, Vol. 69, No. 5, September/October 1990, pp. 61-76.
14. R. J. Perdue, J. Leikness, and J. B. Sharp, "AT&T's Voice Processing System Architectures," AT&T Technical Journal, Vol. 69, No. 5, September/October 1990, pp. 52-60.
15. J. G. Wilpon, "A Study of the Ability to Automatically Recognize Telephone Quality Speech from Large Customer Population," *AT&T Technical Journal*, Vol. 64, No. 2, February 1985, pp. 423-451.
16. R. Thanawala, B. H. Fetz, and R. J. Pierth, "Automatic Speech Recognition in the Public Switch Network," *Proceedings of the Fifth World Telecom Forum*, Vol. 1, Part 2, 1986, pp. 235-238.
17. D. Schinke, "S-I recognition Applied to Telephone Access Information System," *Proceedings of Speech Technology Conference*, April 1986.
18. R. P. Mikkilineni and R. J. Perdue, "Experiences with Implementing Automatic Speech Recognition in the Public Telephone Network," *Proceedings of the Voice I/O Systems Applications Conference*, September 1989.

Biographies (continued)

staff in the Voice Transaction Systems Department. He works on the development and application of speech recognition *algorithms. He has a B.Tech. in aeronautical engineering from the Indian Institute of Technology, Madras, India, and an M.S. and Ph.D. in aerospace engineering from the University of Tennessee at Knoxville. He joined AT&T in 1978. Mr. Roe is supervisor of the Speech Systems Research Group. He does research to support applications of speech recognition, including work on parallel processing algorithms and on digital signal processors. He has a B.A. in mathematics and physics from St. Olaf College, Northfield, Minnesota, and a Ph.D. in physics from Duke University, Durham, North Carolina. He joined AT&T in 1977. Mr. Gokcen is a member of technical staff in the Voice Transaction Systems Department. He works on technology transfer and algorithm development in speech signal technologies. He has a B.S. in electrical engineering and an M.S. in cognitive sciences and artificial intelligence from Ohio University, Athens. He joined AT&T in 1987.*

(Manuscript received June 18, 1990)