SPEECH PROCESSING: AN EVOLVING TECHNOLOGY

Ronald E. Crochiere and James L. Flanagan

***Ronald E. Crochiere** is head of the Speech Processing Department and **James L. Flanagan** is director of the Information Principles Research Laboratory at AT&T Bell Laboratories in Murray Hill, New Jersey. Mr. Crochiere's department is responsible for the common technology, both subsystems and components, for speech processing. He has a B.S. from Milwaukee School of Engineering and an M.S. and Ph.D. from Massachusetts Institute of Technology (MIT), all in electrical engineering. Mr. Flanagan's laboratory is responsible for research in speech, signal processing, acoustics, robotics, vision, artificial intelligence, linguistics, and computer-aided instruction. He has a B.S. from Mississippi State University and an M.S. and Sc.D. from MIT, all in electrical engineering.*

As we enter the information age, speech processing is emerging as an important technology for making machines easier and more convenient for humans to use. It is both an old and a new technology—dating back to the invention of the telephone and forward, at least in aspirations, to the capabilities of HAL in *2001*. Explosive advances in microelectronics now make it possible to implement economical real-time hardware for sophisticated speech processing—processing that formerly could be demonstrated only in simulations on main-frame computers. As a result, fundamentally new product concepts—as well as new features and functions in existing products—are becoming possible and are being explored in the marketplace. As the introductory piece to this issue, we draw a brief perspective on the evolving field of speech processing and assess the technology in the three constituent sectors: speech coding, synthesis, and recognition.

Perspective

Speech processing technology traces its origins back fifty years, well beyond the electronics revolution triggered by the invention of the transistor. The origins lie in telephony.

Voice communication between continents of the world became a burning challenge soon after the first undersea telegraph cable linked North America and Great Britain. The year was 1858, and the cable had a usable bandwidth of 1.5 Hz! Subsequent progress toward transoceanic voice required the invention of the telephone (in 1876) and an understanding of the transmission bandwidth needed to support speech signals. By 1924, permalloy loading increased the usable bandwidth of long electrical cables to about 200 Hz—a significant advance, but not nearly enough for speech waveform transmission.

In 1927, low-frequency radio (at 60 kHz) provided the first transatlantic voice carrier, with high-frequency radio coming along in the next couple of years. But radio quality was variable, privacy was difficult to ensure, and the desire for voice over cable remained strong.

In the late 1930s, the vocoder (voice coder)—perhaps the first speech processing system to deserve the name (certainly, the first speech analysis-synthesis system)—was conceived as a way to compress speech bandwidth by a factor of about 10, making possible speech transmission over low-bandwidth telegraph cables. But new advances in radio and the marginal quality of vocoder speech thwarted wide application. Nevertheless, during World War II, the vocoder satisfied critical applications for voice privacy by providing full digital security on transatlantic radio channels.

Eventually the technology of submerged repeaters (amplifiers) advanced and, in 1956, the first transatlantic voice transmission was achieved over an undersea cable, called TAT-1. The cable provided 36 two-way channels at a cost of about $1M per channel.

Bandwidth conservation for this medium remained an interest, and the speech processing technique known as time assignment speech interpolation (TASI) was deployed in 1959. TASI doubled the capacity of undersea cable by exploiting the activity factor of conversational speech. Its electronic processing for speech detection was the forerunner of the technique that is now being applied in wideband packet technology.

Efforts in speech processing continued to be driven by efficient use of transmission capacity up to about 1970. Then, owing to the advent and proliferation of digital computers, a new dimension emerged; namely, that of making computers more useful for humans.

This interest centered on the use of natural voice for computer-human interaction. Speech synthesis gives computers a *mouth* to speak. Speech recognition provides computers an *ear* to listen. And, most of the technology for reducing speech bandwidth applies to speech synthesis and recognition.

But the motivation of transmission efficiency remains, even with the promise of enormous bandwidth from optical fiber. Specifically, efficient use of the mobile (cellular) radio spectrum and the benefits of wideband packet technology (for optimally loading digital networks with signals that may interleave voice, data, and image) requires much more than simple analog-to-digital conversion of speech waveforms. Instead, it requires algorithmic transformations that permit transmitting speech information using the smallest digital rates consistent with good quality. Further, security and privacy techniques—especially for radio—remain of strong interest, and they, too, require sophisticated speech processing for effective implementation.

Until about 1980, advances in the fundamental understanding of speech processing had largely outstripped our ability to implement this acquired knowledge in economical and reliable real-time hardware. Some of the more advanced processes for speech coding, synthesis, and recognition were much too complex to support. But the explosive advances in microelectronics over the last few years have substantially changed this picture.

Now, the algorithmic complexity that can be supported in small, economical electronics is rapidly depleting the backlog of fundamental understanding, and current research is sorely challenged to originate techniques that productively harness large computational capabilities. Simultaneously, this expanded hardware capability is creating opportunities for new products and applications, and for enhancing the features and functions of existing products. The market for these speech processing products is evolving in ways and directions that are only beginning to be understood.

The Underlying Technology

Work in speech processing traditionally divides into three related sectors:

- *Speech coding*—primarily concerns human-to-human voice communication (even with a machine intermediary).

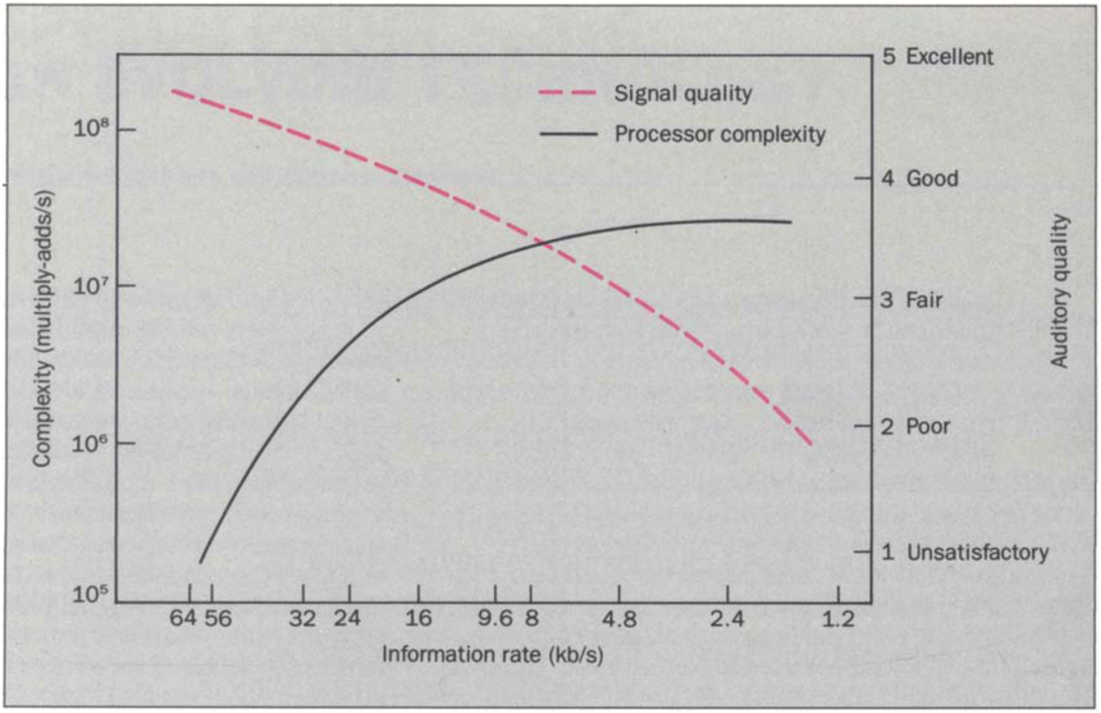

Figure 1. Representative relations among information transmission rate, computational complexity, and auditory quality of current speech coding methods.

- *Speech synthesis*—concerns machine-to-human communication.
- *Speech recognition*—relates mainly to human-to-machine communication.

The thrust of this issue is to reflect current applications of these technologies in the movement and management of information, summarize the state of the technology, and suggest projections for the next few years.

Speech Coding. The objective in speech coding is to analyze and represent (encode) speech as a digital signal that uses the smallest amount of transmission capacity required to recreate (decode) the speech at the receiver. Ideally, the resulting auditory quality should be comparable to the original. There are theoretical reasons to believe that a transmission capacity as small as about 2 kb/s could transmit speech with natural quality. But we are far from this capability today.

Figure 1 shows representative relations among information transmission rate, transmitted speech quality, and computational complexity of encoder-decoder equipment. The dashed line indicates the signal quality that is typical of current algorithms, while the solid line represents typical computational complexity needed to implement these algorithms. As a basis for comparison, current digital signal processor (DSP) chips have computational powers of about 2 to 5 million multiply-adds per second.

As Figure 1 indicates, high-quality transmission can be achieved with modest complexity from 64 kb/s—the traditional rate for pulse code modulation (PCM)—down to about 32 kb/s—the rate for the more recent, adaptive differential PCM (ADPCM).

To achieve good quality below 32 kb/s, codes must take increasing advantage of the constraints of speech production and perception. At transmission rates below 16 kb/s, quality diminishes significantly, requiring more of the (as yet, poorly known) properties of speech production and perception. Also, at the lower transmission rates, the computational complexity to implement the coding algorithm increases, while the ability to handle nonspeech-like sounds—such as music and voice-band data—diminishes. Typically, too, the encoding delay increases as the transmission bit rate decreases.

The primary challenge, then, is to develop new understanding that will significantly elevate the speech-quality curve for the lower bit rates, even with substantial (but acceptable) increase in complexity.

The research frontier in coding currently centers on ways to achieve good quality at transmission rates of 9.6 kb/s and below. Undoubtedly, increased computational

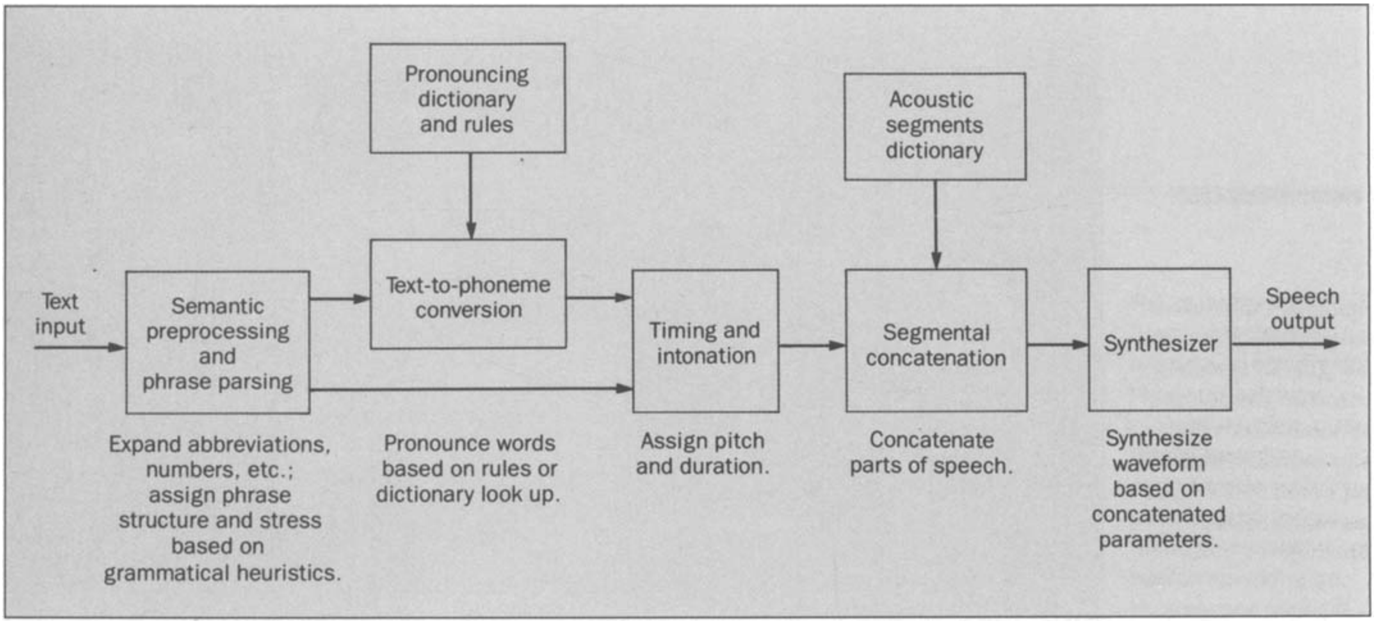
Pronouncing dictionary and rules
Acoustic segments dictionary
Text input
Semantic preprocessing and phrase parsing
Text-to-phoneme conversion
Timing and intonation
Segmental concatenation
Synthesizer
Speech output
Expand abbreviations, numbers, etc.; assign phrase structure and stress based on grammatical heuristics.
Pronounce words based on rules or dictionary look up.
Assign pitch and duration.
Concatenate parts of speech.
Synthesize waveform based on concatenated parameters.

Figure 2. The elements of text-to-speech synthesis.

complexity will be required to elevate the quality of low bit-rate codes, which must extensively use the known redundancies of speech production and perception. Breakthroughs will occur only when new properties of redundancy are found.

Speech Synthesis. The methods of speech synthesis permit a machine to speak instructions or information to a user. In its simplest form, the machine just plays back one of several stored messages. Broad applications already exist for such techniques, and are widely used in today's telecommunications network to provide a range of automated network announcements. Messages are recorded and stored in digital memory, using coding techniques similar to those described above. The tradeoffs between quality and bit rate (and, hence, required storage) apply here as well.

With modest added complexity, more versatile announcements can be composed from a library of message parts—sentences, phrases, or words. Message content is constrained by the size and flexibility of this library. Typically, several versions of a particular phrase or word, recorded with different (rising or falling) inflections, are required to achieve naturalness. The inflection selected depends on where the part occurs in the sentence and whether the sentence represents a question, statement, or exclamation. Separate libraries must be designed and constructed for each application and message scenario.

Ideally, a talking machine should be able to convert any arbitrary message text into its spoken form without the restrictions of a prestored library of encoded phrases. Also, the machine should pronounce the message with proper emphasis based on the text's punctuation and a knowledge of syntax and semantics. This unrestricted text synthesis (Figure 2) is an ambitious objective and the focus of intensive research.

Input text (which could be electronic mail) often includes abbreviations, dates, times, and punctuation marks. The machine must compute the proper pronunciation, stress, and timing and synthesize the speech output with good intelligibility and naturalness. To date, such machines achieve good intelligibility. The challenge is to make their naturalness comparable to stored announcement systems.

Three important dimensions are evident in speech synthesis: speech quality, message versatility, and complexity. The desired objective, of course, is high quality and versatility with low complexity.

In reality, simple systems that store recorded waveforms have high quality and low complexity, but they also have low versatility; i.e., the messages can be used only in the form recorded. At the other extreme, text-to-speech systems that are based on computational synthesis have great versatility but also have high complexity and, as

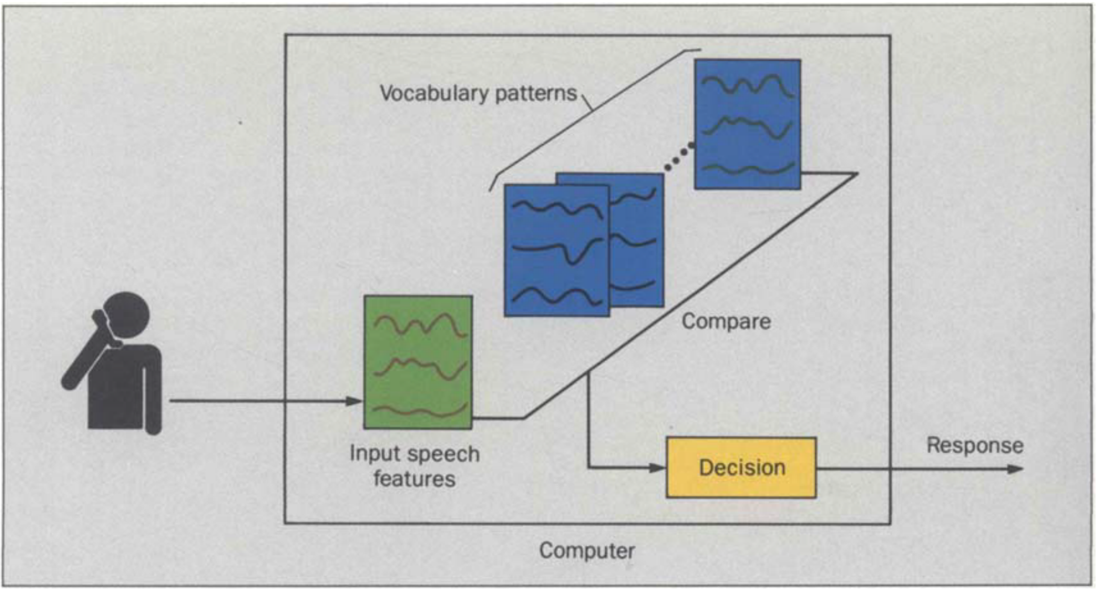

Figure 3. Constituent functions of a speech recognition system based on whole-word patterns. This system tries to match individual words of the input speech to stored patterns.

yet, limited quality.

Speech Recognition. Of the various approaches to automatic speech recognition, two techniques for recognizing isolated words have achieved good utility. One matches an unknown input utterance to stored spectral templates using dynamic time warping (DTW). The other statistically characterizes speech as a hidden Markov model (HMM).

Both techniques can be represented by the generic diagram in Figure 3. An unknown spoken command is analyzed, typically using filter banks or linear predictive coding (LPC), to extract its spectral features. The features can be classified into a finite set of patterns, using vector quantization. The patterns are then compared to stored sets of vocabulary patterns to determine the closest match.

The stored vocabulary patterns are predetermined from measurements on speech data. A system that uses a particular person's speech data is said to be *speaker dependent.* If the data are obtained from an appropriate large user population, the system is *speaker independent.* The unknown input is then identified as the closest-matching vocabulary entry. If the machine does not find a close enough match, it can announce this result using its synthetic voice.

In the DTW approach, the unknown feature set is matched against the stored patterns using a procedure that dynamically alters the time dimension to minimize the accumulated distance score for each pattern. This makes the recognition process insensitive to variation in talking rate.

In the HMM approach, training data create a statistical, finite-state Markov chain for each vocabulary word. In the classification phase for the unknown input, the probability of generating the state sequence for each vocabulary word is computed, and the word with the highest accumulated probability is selected as the correct identification. Time alignment is obtained indirectly through the sequence of states.

The costs of implementing the DTW and HMM techniques are comparable.

The recognition problem has at least three dimensions: vocabulary size, speaker identity, and fluency of input speech. Current understanding permits building practical systems that reliably recognize several hundred words spoken by a person who trained the system. Recognition for any or all speakers requires about ten times more computation than for individuals whose vocabulary patterns have been stored. Recognition of single words or short phrases—spoken in isolation—can be done reliably, even over dialed-up telephone channels. Recognition of connected words is under active development. Recognition of conversational, fluent speech is in fundamental research, and advances strongly depend on good computational models for syntax and semantics.

The performance of speech recognizers depends on the design parameters selected, vocabulary nature and size, and acoustic environment. Typically, a conventional DTW design (without vector quantization) does slightly better than today's HMM designs (which use vector quantization).

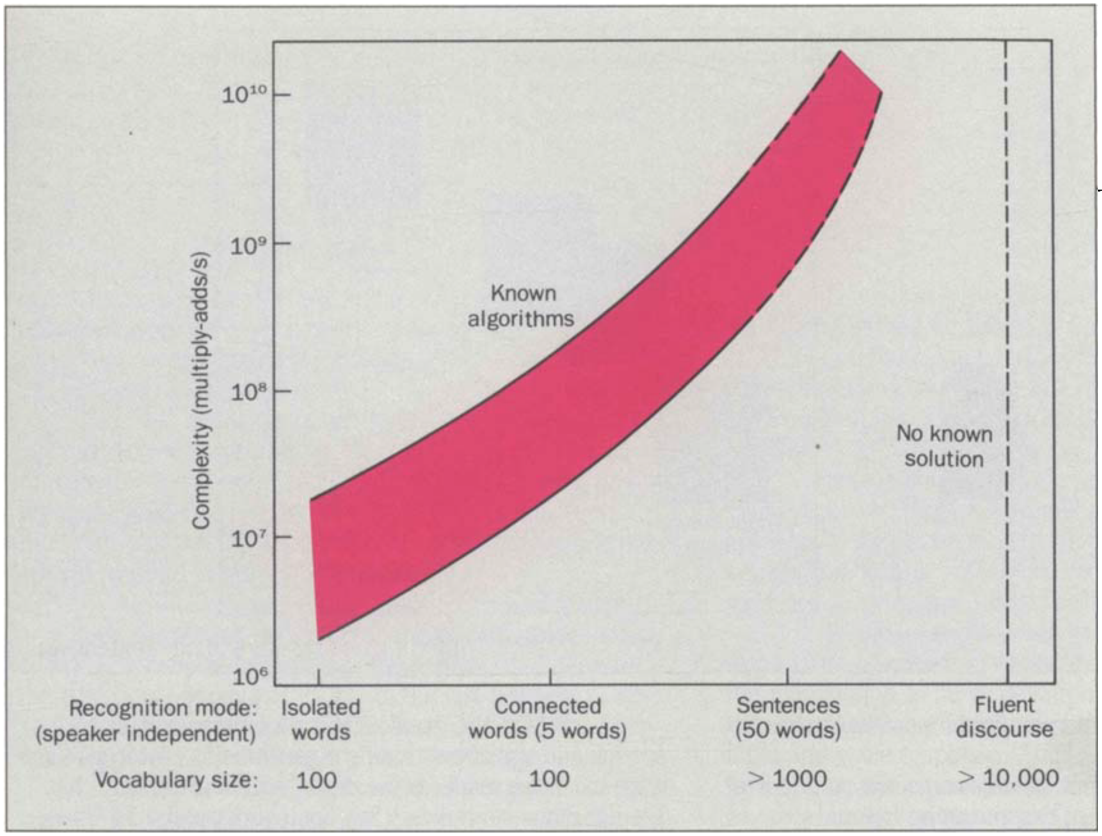

Figure 4. Computational complexity required to implement known speech-recognition algorithms based on stored whole-word patterns and no syntactic processing. The person speaking may be someone who did not train the system.

The size and nature of the recognition vocabulary significantly influence performance. The spoken digits—a distinct, small vocabulary—can be recognized with high accuracy even in speaker-independent operation. (Typically, the recognition accuracy of a DTW design in a well-controlled acoustic environment exceeds 99 percent for a speaker-dependent vocabulary and 98 percent for a speaker-independent vocabulary.) When spoken digits and alphabet (39 alpha digits) are combined, word recognition performance typically deteriorates (below 80 percent for speaker-dependent operation). This happens because the spoken alphabet symbols—such as *b, c, d, e, t, v*—have similar acoustic features. Here, syntactic constraints must be used to achieve acceptable recognition.

Performance also tends to decrease as vocabulary size increases. Again, the constraints of context and imposed syntax are a major resource. Task performance can improve considerably if syntactic redundancy can be built into the application. Computational requirements increase dramatically as one moves from isolated to connected word recognition. Figure 4 gives an impression of current algorithmic complexity as a function of the recognition task for systems that store whole-word patterns and do not model syntax.

Speaker Recognition. A technology that is closely related to speech recognition is speaker recognition, or automatic recognition of a talker from measurements of individual characteristics in the voice signal. Two tasks are relevant: absolute *identification* from the universe of possible talkers, and talker *verification* that evaluates an identity claim. Identification is the more difficult task, and performance typically decreases as the talker population increases. Verification is a more tractable task, and performance does not depend strongly on the size of the user population. Verification is usually more relevant in business applications, where a user may seek access to restricted or privileged information.

Figure 5 illustrates the machine functions in a verification task. The claimant's voice pattern is analyzed into a set of spectral features. Unlike speech recognition features, these measured features reflect individual differences and are compared against a set of patterns that the claimant stored when he or she enrolled. A DTW pattern match is made, as in speech recognition, and the

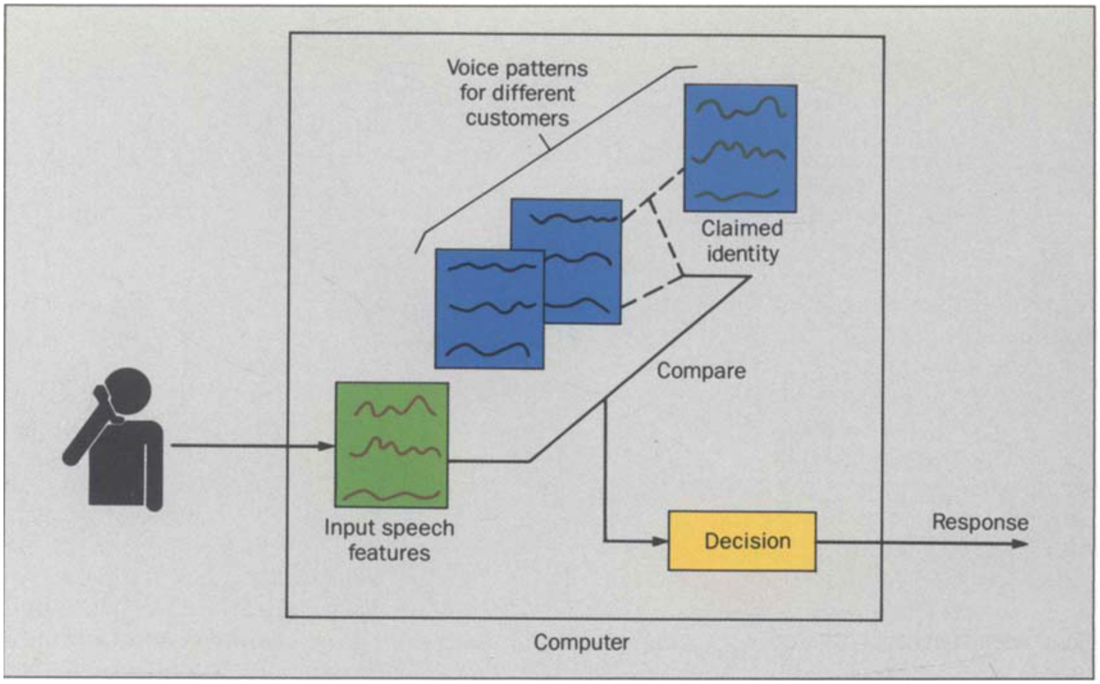

Figure 5. Signal processing operations for speaker verification in a pattern-recognition framework. This system matches speech patterns to check a speaker's claimed identity.

accumulated score is compared against a predetermined threshold.

Two types of errors can occur: accept an imposter or reject a valid user. These errors can be traded by adjusting the decision threshold. For sentence-length utterances, total error rates of about 4 percent (equally divided between imposter acceptance and valid user rejection) are typical from a single sentence spoken over the telephone. Using multiple utterances for independent verifications or combining speech verification with other security methods can reduce error rates to a low value.

Speech Processing Hardware. Over the last few years, VLSI (very-large-scale integration) has dramatically transformed speech processing technology from laboratory demonstrations into economical hardware. Algorithmic techniques that once required lengthy computer simulations now run in real time on single-board or single-chip hardware.

Today, single DIP (dual in-line package) synthesizers can provide good-quality voice announcements for voice response applications. Coprocessors can provide the computation needed for sophisticated DTW pattern matching for medium-size vocabularies (e.g., 256 templates). Speech coding techniques—such as ADPCM—can be realized in single chips that provide single or multichannel coding and decoding.

By far, the greatest impact on transforming speech and signal processing algorithms to prototype hardware has been made by the digital signal processor chip. The DSP has served as a low cost, programmable vehicle to transform algorithmic concepts rapidly into real-time hardware that is economically suitable for many practical applications. Many advantages of this approach are reflected by Boddie et al.[1] in this issue.

As the technology evolves, DSPs can become the *operational amplifier* of the digital signal processing world and serve as a basic building block in signal processing designs. Current DSPs largely favor algorithms that are based on *sums of products* (e.g., digital filtering, correlation processing, and frequency-time domain transforms). But new types of programmable DSPs whose designs are optimized for other generic classes of signal processing algorithms—such as parsing, graph searches, and vector quantization—can be expected.

Application of the Technology

Potential uses for speech processing permeate society. Applications range from communications networks, business offices, consumer electronics, to government and military systems.

In some applications, speech processing is a necessity—not a mere convenience. Examples include hands-free, eyes-free dialing for mobile radio telephone; enhanced telephone services from rotary-dial station equipment; and communication aids for speech, hearing, and motor-handicapped people.

AT&T's traditional strengths—systems engineering and systems integration; networking; VLSI; and coordinated R&D, manufacturing, marketing, sales, maintenance, and service—position it well to contribute in these applications. Several areas of successful application are exemplified in this issue's papers.

Transmission Network Standards. Techniques have long been sought for high-quality coding and transmission of digital speech at information rates substantially below the traditional 64 kb/s of PCM. In the last few years, 32-kb/s transmission equipment that uses ADPCM has been deployed in the telecommunications networks.

An international standard has been promulgated to cover the traditional demands of the network environment, including voice and nonvoice signals and tandem connections of analog-to-digital conversion. In this issue, the paper by Benvenuto et al.[2] gives details of this new transmission standard. It also describes applications in the BCM 32000 transcoder, SLC® Series 5 carrier, and D4 channel bank. The ADPCM technique effectively doubles the capacity of existing digital facilities.

Voice Store-and-Forward Standards. Voice store-and-forward systems now on the market have used a variety of coding methods. Their coding rates typically range from PCM down to data speeds, with the attendant tradeoff between quality and storage requirements.

Several AT&T product developments have recently focused on the use of subband coding at rates of 16 and 24 kb/s. Because of this interest, AT&T has adopted an internal standard that addresses not only the coding method, but also the format for embedding the coded signal in a voice file. This permits adding useful features—such as silence compression, automatic gain control, insertion of application-dependent data in the coded file, synchronization, and other signal modifications—in the voice store-and-forward context.

The paper by Josenhans et al.[3] discusses this standard and its effect on the compatibility and interoperability of AT&T products and services for voice store-and-forward. Papers by Perdue and Rissanen,[4] and Ackenhusen et al.[5] (about speech for workstations) discuss systems applications of the standard.

PBX Enhancements. Voice mail services in PBX (private branch exchange) systems—such as System 85, System 75, and Enhanced Dimension® system—represent a major application of the new voice store-and-forward standard. Papers in a future issue will cover this subject.

Automated Attendant and Database Access Systems. Automated attendant systems and audiotex systems permit information to move between humans and machines through a combination of touch-tone (or voice recognition) input and voice response. Applications of this technology lie in information services for commercial operations, such as stock market quotation and order entry. This area exhibits considerable market activity and is one that AT&T has entered through its new-venture company, Conversant Systems.

In this issue, Perdue and Rissanen[4] discuss the activities of this venture and the hardware architecture of the Conversant® I voice system. To provide voice response, the system may use 9.6-kb/s multipulse LPC speech synthesis or the subband standard described above, depending on storage and real-time requirements. Using a unique example, the paper also describes how to design embedded syntax into voice recognition applications to improve the system performance of a speech recognizer.

Voice Control. Today, a large amount of vendor hardware is available for isolated word recognition to serve terminals, user computers, workstations, and consumer applications. Targeted applications typically require word vocabularies that range from several hundred, speaker-trained isolated words and phrases to a dozen or so speaker-independent commands. In this issue, Ackenhusen et al.[6] describe a representative hardware system, a single-board speech recognizer.

Unlike the markets for coding in transmission and store-and-forward, the speech recognition markets are only beginning to evolve. They will require careful analysis before the value of the technology can be firmly established. At present, growth in applications of speech recognition is not limited by cost-effective hardware or system accuracy and performance. Instead, it is conditioned more by the design of the human-machine interaction environment and user acceptance in a particular task.

Recently, much interest has centered on expanding recognition capabilities to large vocabularies that are speaker independent and can be spoken in connected fashion. As this understanding evolves, greater use of speech recognition embedded in communications systems—for functions such as control and switching—can be anticipated. More near-time use in robotics and future use in machine translation can also be expected.

Speech in Terminals and Workstations. With the rapid advances in the computer industry, the cost of computing has dropped dramatically, which has led to a broad range of desk-top and portable computers, workstations, and terminals. Speech processing is often viewed as a tempting adjunct to such systems—to improve their versatility, functionality, and human-machine interface.

The cost of speech processing has similarly dropped to where significant, multifunctional speech processing capabilities (e.g., voice response, voice store-and-forward, speech recognition, and text-to-speech synthesis) can be economically integrated into such equipment, often as a low-cost, plug-in board. The paper by Ackenhusen et al.[5] describes such an implementation for AT&T's UNIX® PC. It also discusses potential applications to small businesses and office automation.

Voice Password Security Systems. With the increased ability to access information and databases through more flexible human-machine interfaces, concern arises about the need for privacy and protection of privileged information. Examples include access to information in personal bank accounts, credit-card accounts, and production inventories; electronic transfer of funds; as well as entry into computer rooms and buildings.

Speaker verification offers the ability to screen users and prevent imposters from illegally accessing restricted information and services. In this issue, Birnbaum et al.[7] describe a speech-verification password system intended for such applications.

Future Outlook

The preceding discussion has already suggested some areas of future progress. But what are the fundamental limitations to progress in the technology and its application in the marketplace?

Limitations. The technical limitations are not so much our ability to conceive—unfettered—new techniques and principles, or implement the new and daring principles in sophisticated special-purpose hardware. They lie in how quickly we can acquire knowledge through experiments with new algorithms and establish the feasibility and value of new models.

A fundamental factor in the rate of knowledge acquisition is the general-purpose computing power available in the laboratory. The speech recognition issues described in Figure 4, for example, illustrate this point, although the same limitations exist to a lesser extent in coding and synthesis.

The algorithms of current research interest (e.g., for whole, connected sentences) require about 100 times more computing speed for real-time operation than do isolated-word recognizers. Therefore, each small change in an experimental model or principle can be subject to significant delay in laboratory evaluation. Happily, this limitation is softening with the expanding availability of massively parallel, general-purpose computers.

The evolution and application of speech technology in the marketplace is governed by a different set of limitations and considerations. Success or failure of a product depends strongly on its perceived value to the end user. This marketing process requires experimentation—not in the laboratory, but in the field.

More conventional functions—such as transmission, voice response, or store-and-forward—are well-understood concepts and have relatively predictable markets. Functions such as speech recognition, speaker verification, and text-to-speech synthesis are not broadly experienced by the general user population. Therefore, the evolution of the technology will be controlled to some extent by the ability of systems designers to adapt these new concepts to the customers' needs and by the evolution of host products (e.g., PBXs, cellular terminals, personal computers) on which speech processing features and functions ride.

Projections. In concluding, we can speculate about where speech processing technology might advance over the next few years.

In speech coding, 32-kb/s technology is now deployed and firmly established for applications in the public switched network. The fundamental understanding for high-quality voice coding at 16 kb/s is in hand.

Over the next five years, prospects are good for establishing coding algorithms for 9.6 to 4.8 kb/s. Success here would contribute significantly to future needs in cellular radio and store-and-forward applications. Application of these low bit rates in the public switched network is less likely because of the increasingly available bandwidth from optical fiber.

Over the longer range, possibly ten years, good speech quality and robustness at vocoder rates as low as 2.4 kb/s may be possible. These techniques will still be of interest for applications where bandwidth is limited—for example, where full digital security is desired on high-frequency radio voice channels. Progress here will depend on significant new understanding of ways to model the speech signal.

In speech synthesis, first systems for unrestricted text-to-speech conversion are producing useful, intelligible synthetic speech but of limited naturalness.

Over the next five years, work already in progress aims to produce high-quality synthesis from text, where different voice qualities (such as man, woman, child) might be specified. Also, synthesis from text might be realized for languages, such as Chinese, that are quite different from Western languages. Over the long term, detailed understanding may permit specifying individual voice characteristics, dialects, and accents.

In speech recognition, systems for reliable recognition of isolated words are well established and beginning to prove their value. The near term will see speaker-independent recognition of connected digits established and applied.

Over a five-year period, the technology will advance to whole, connected sentences, using limited vocabularies and finite grammars. Over the longer term, understanding of programmed parsers and natural language analysis will allow the leverage of syntax, semantics and, eventually, even pragmatics to expand a machine's conversational ability. Ultimately, practical spoken language translation may be possible.

While we may not achieve the facility of HAL, the conversational computer of *2001*, we will come close—and about on schedule!

References

1. J. R. Boddie et al., "The DSP32 Digital Signal Processor and its Application Development Tools," *AT&T Technical Journal*, Vol. 65, No. 5, September/October 1986, pp. 89-104.
2. N. Benvenuto et al., "The 32-kb/s ADPCM Coding Standard," *AT&T Technical Journal*, Vol. 65, No. 5, September/October 1986, pp. 12-22.
3. J. G. Josenhans et al., "Speech Processing Applications Standards," *AT&T Technical Journal*, Vol. 65, No. 5, September/October 1986, pp. 23-33.
4. R. J. Perdue and E. L. Rissanen, "CONVERSANT® I Voice System: Architecture and Applications," *AT&T Technical Journal*, Vol. 65, No. 5, September/October 1986, pp. 34-47.
5. J. G. Ackenhusen et al., "Speech Processing for AT&T Workstations," *AT&T Technical Journal*, Vol. 65, No. 5, September/October 1986, pp. 60-67.
6. J. G. Ackenhusen et al., "Single-Board General-Purpose Speech Recognition System," *AT&T Technical Journal*, Vol. 65, No. 5, September/October 1986, pp. 48-59.
7. M. B. Birnbaum, L. A. Cohen, and F. X. Welsh, "A Voice Password System for Access Security," *AT&T Technical Journal*, Vol. 65, No. 5, September/October 1986, pp. 68-74.
8. B. S. Atal and L. R. Rabiner, "Speech Research Directions," *AT&T Technical Journal*, Vol. 65, No. 5, September/October 1986, pp. 75-88.

(Manuscript received August 18, 1986)