

# AT&T VOICE PROCESSING SYSTEM ARCHITECTURES

Robert J. Perdue, John M. Leikness, and John B. Sharp

**Robert J. Perdue, John M. Leikness, and John B. Sharp** are all associated with AT&T. Mr. Perdue is a supervisor in the Voice Transaction Systems Department at AT&T Bell Laboratories, Columbus, Ohio, where he is responsible for advanced speech technology features in voice response systems. He joined AT&T in 1973 with a B.S.E.E. and M.S.E.E. from the Massachusetts Institute of Technology, Cambridge. Mr. Leikness is district manager in the System Peripherals and Adjuncts department of AT&T Business Communications Systems, Bridgewater, New Jersey. He and his group are responsible for project management for the AUDIX and the AUDIX Voice Power product line. Mr. Leikness joined AT&T in 1969 with a B.S. in education from the University of North Dakota, Grand Forks. Mr. Sharp is a supervisor in the (continued on page 60)

The voice processing equipment market has expanded to include both large and small businesses, and the influence of voice processing—evidenced in voice mail messages and voice response systems—has become widespread. AT&T systems such as the Conversant® Voice Information System and the AUDIX voice messaging system have been designed to enhance both business and personal telecommunications. This paper explores the architectures of AT&T's Conversant and AUDIX systems in terms of hardware, software, and application design.

## Introduction

The market for voice processing equipment was almost \$1 billion in 1989, and is growing rapidly. Large and small businesses have discovered the benefits of applying voice processing technology to enhance communication among people, and between people and machines. Almost everyone has left a voice mail message at one time or another. Increasingly, some business operations can be handled by using the dual tone multifrequency (DTMF) pad on the telephone to access or change information in a computer. The spectrum of applications that enhance interpersonal communications are generally classified as *voice messaging*. The applications that allow people to communicate easily with computers are classified as *voice response*. AT&T plays a major role in both the voice messaging and voice response markets.<sup>1</sup>

The AT&T Conversant voice information system is a voice response system designed for telephone attendant automation.<sup>2</sup> As a programmable system that prompts calling customers to enter numbers (using DTMF or speech) to request or enter data, it can be used in many customer applications, such as account inquiry.<sup>3</sup> The AT&T employee savings plans, administered by American Transtech, are accessed via Conversant voice response systems.

AT&T's AUDIX (Audio Information Exchange) and AUDIX Voice Power voice messaging systems enhance business and personal telecommunications by helping users create, edit, store, send, or receive spoken messages via their telephones. These capabilities allow effective

**Panel 1. Terms and Acronyms in This Paper**

AGC	automatic gain control
AUDIX	Audio Information Exchange
CPU	central processing unit
DBP	database processor
DNIS	dialed number identification
DSP	digital signal processor
DTMF	dual-tone multi-frequency pad
FP	feature processor
IVP-4	Integrated Voice Power-4 port board
LAN	local area network
PRI	primary rate interface
PBX	private branch exchange
SBC	subband coding
SP	signal processing
SPC	signal processing companion
TD	time division bus
TDM	time division multiplex
VSP	voice session processor

information management using the voice format with which people are most accustomed. The AUDIX system provides high quality, highly functional voice messaging services to headquarters or regional locations ranging from 100 to thousands of users. AUDIX Voice Power provides similar cost-effective services to branch offices and small businesses with between 10 and 300 users.

**Voice Response Applications**

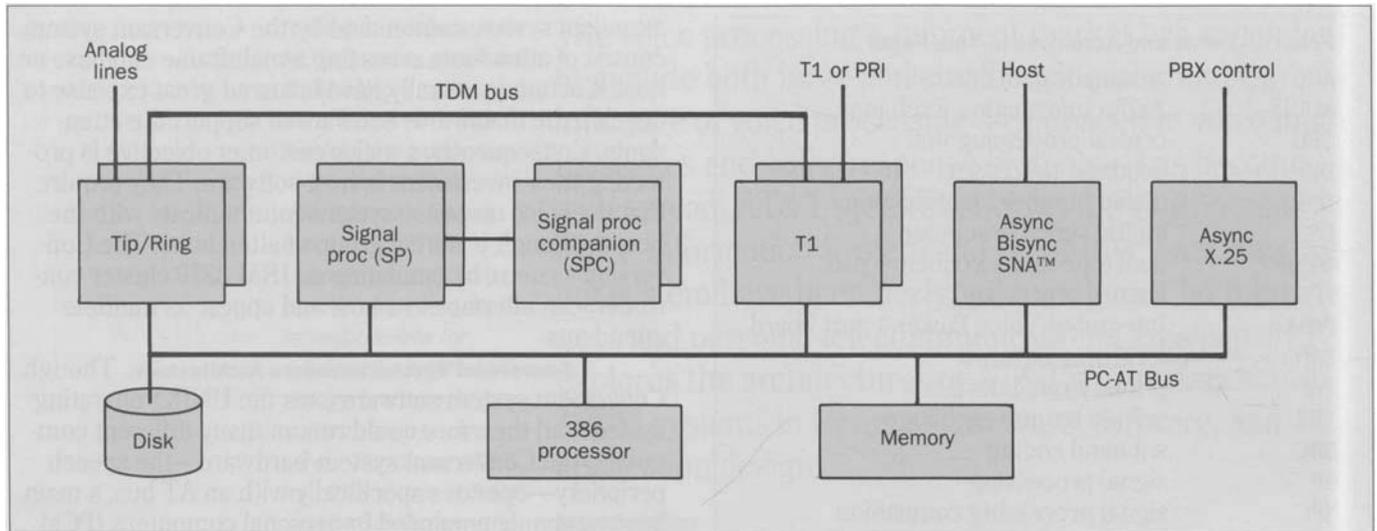
The range of voice response applications is large, and requires the Conversant system to be flexible in terms of line sizes, telephone interfaces, voice features, transaction and speech storage capacity, and database interfaces. Though the system is programmable, developing a customer's application program must be easy and straightforward to minimize total application costs. ScriptBuilder, the Conversant system's resident applications generator, permits customers to program their applications in a short time with little instruction. Most

attendant systems automated by the Conversant system consist of attendants accessing a mainframe database or host. Customers usually have incurred great expense to develop the mainframe software to support the attendants. Consequently, a major customer objective is protecting their investment in host software. They require that the voice response system communicate with the host as though it were a group of attendants. The Conversant system, by emulating an IBM 3270 cluster controller, can interface to a host and appear as multiple attendants.

**Conversant System Hardware Architecture.** Though Conversant system software uses the UNIX® operating system and therefore could run on many different computers, the Conversant system hardware—the speech periphery—operates specifically with an AT bus, a main bus commonly employed in personal computers (PCs) and workstations. (UNIX is a registered trademark of UNIX Systems Laboratories, Inc.) The Conversant system Intel 80386-based system controller uses the AT bus to transfer data among the memory, disks, host interface cards, a LAN (local area network) card, and the speech periphery cards illustrated in Figure 1. Communication with the customer's external host machine normally is handed by a variety of communications protocol cards supporting the 3270 SNA™ network, X.25, synchronous data link control (SDLC), and the asynchronous RS-232 interface. (SNA is a trademark of IBM.) A LAN connection to the host is also possible.

In installations with more than one Conversant system, centralized administration of multiple systems simplifies administration and maintenance, and collects performance reports. The Conversant Voice Management System (VMS) provides centralized administration for several Conversant voice information systems. A LAN interconnects multiple Conversant voice information systems, the VMS, and sometimes a host.

The system hardware uses two internal buses: an AT bus and a time division multiplex bus (TDM) (Figure 1). To reduce speech storage and data transfer



**Figure 1. Hardware architecture of the Conversant voice information system. The Intel 80386-based system controller uses the AT bus to transfer data among the memory, disks, host interface cards, a LAN card, and the speech periphery cards.**

requirements, the Conversant system encodes speech by using one of several user-selectable speech compression algorithms. Encoded voice prompts and captured speech are stored on the disk. They are retrieved by the system controller, and are passed to the speech periphery for decoding and playback over the telephone network. The speech periphery cards connect with each other and transport speech over the TDM bus. The TDM bus gives any card in the Conversant system bidirectional access to any conversation. With the TDM, functionality can be spread among different cards. Dynamic allocation and deallocation of resources—such as speech coding and speech recognition—can occur during a call.

The speech periphery cards for a voice response system must perform two tasks:

- They must handle network signaling protocols to

interface to the public switched network.

- They must process the audio signals received over the telephone channels.

A set of cards for the Conversant system spans the network signaling and signal processing tasks.

Analog lines are the simplest way to connect the Conversant system to a customer's private-branch exchange (PBX). For full voice response capability in a small line size and at low cost, a Conversant system 6-channel analog card incorporates both the functions of network interface, to a local telephone loop, and signal processing functions into one card. Besides interfacing to the telephone line, it performs DTMF generation and detection, speech encoding and decoding; and simple call progress tone classification (to distinguish a busy signal from a ring) for intelligent call transfer.

A digital interface card handles both T1.5 and PRI (primary rate interface) from the network. These digital interfaces are the most common methods used to access AT&T's Megacom® and MultiQuest® telecommunications services. In T1.5 mode, this card supports 24 digital channels, all using the dialed number identifi-

cation (DNIS). DNIS allows multiple called numbers to terminate within a 24 channel trunk group. The single data channel on the PRI can pass signals—such as network protocol signals and the caller's telephone number (automatic number identification)—for 23 to 47 bearer or conversation channels.

The signal processing (SP) card, which contains a general purpose processor and multiple digital signal processors (DSPs), performs signal processing tasks, such as speech encoding and decoding for large line size systems, public switched network call progress tone analysis, and advanced speech technology tasks. One SP provides 48 channels of speech coding and decoding. Additional SPs can be installed in a system for other features.

To supplement the SP on computationally intensive tasks such as speech recognition, a nest of SP companion (SPC) cards can be connected to an SP. Designed for the demanding computing requirements of multi-channel connected word-speech recognition, the SPC provides 12 DSPs per board. A 5-board nest of one SP and four SPCs can perform 1.5 billion floating point operations per second.

**Conversant System Software Architecture.** Both the Conversant voice information and voice management systems are based on the UNIX operating system. But the normal customer or programmer system interface is the powerful ScriptBuilder applications generator. ScriptBuilder is a simple menu-based tool that permits rapid development of an application script, i.e., the program flow that prompts the caller for input, queries a database based on the input, and handles error conditions. Figure 2 shows a sample ScriptBuilder program that answers a phone, prompts the caller for input, and then routes the call. It displays ScriptBuilder's simplicity, compactness, and ease of programming.

Within ScriptBuilder, the speech prompts associated with an application can be recorded. Also, during the ScriptBuilder session, the customer or programmer retrieves 3270 host screens for that application, and by pointing the cursor and naming the screen fields, defines

#### Standard Display

start:

1. Answer phone
2. Announce
3. Prompt & Collect
4. Quit

start:

#### Expanded Display

1. Answer phone
2. Announce

Speak with interrupt

Phrase: ''Thanks for calling our company''

3. Prompt & Collect

Prompt

Speak with interrupt

Phrase: ''Press 1 for sales''

Phrase: ''Press 2 for service''

Phrase: ''Press 3 for accounting''

Phrase: ''Press 4 to speak with the operator ''

Input

Max number of digits: 01

Checklist

Case: ''1''

Speak without interrupt

Phrase: ''Please hold for''

Phrase: ''sales''

Transfer to 1022

Case: ''2''

Speak without interrupt

Phrase: ''Please hold for''

Phrase: ''service''

Transfer to 1050

Case: ''3''

Speak without interrupt

Phrase: ''Please hold for''

Phrase: ''accounting''

Transfer to 1010

Case: ''0''

Speak without interrupt

Phrase: ''Please hold for''

Phrase: ''the operator''

Transfer to 1000

Case: ''Not on list''

Reprompt

Case: ''Initial timeout''

Reprompt

Case: ''Too few digits''

Reprompt

Case: ''No more tries''

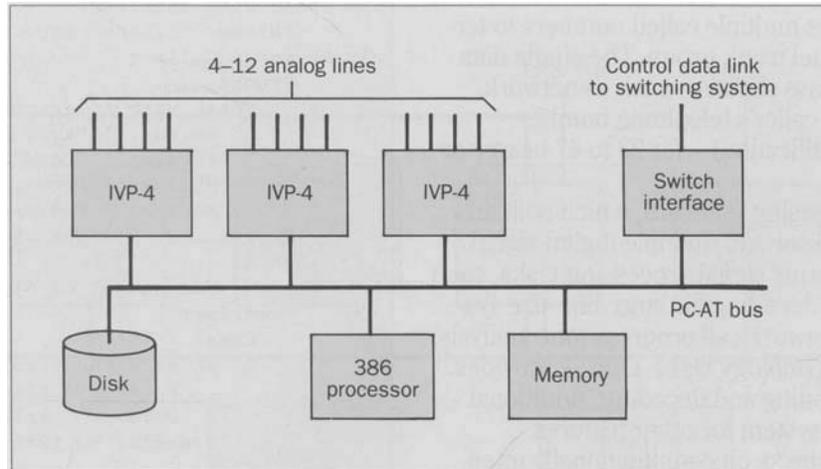
Quit

End Prompt & Collect

4. Quit

**Figure 2. A ScriptBuilder script for call prompting and routing. The sample program shows answering a phone, prompting the caller for input, and routing the call.**

**Figure 3. AUDIX Voice Power architecture, based on the Conversant system platform. The platform enables cost-effective implementation for voice messaging functions with the addition of the IVP-4 (Integrated Voice Power-4 port) boards into the AT&T 6386WGS computer.**



those variables for ScriptBuilder. If local database information is needed, ScriptBuilder interfaces easily to a commercially available local database.

Once the transaction script is written, it may be scheduled for activation on a particular channel. More often, the scripts are assigned to particular incoming DNIS. Thus, the number the customer called starts the transaction.

An efficient software state machine simultaneously operates 48 distinct scripts. At any time, 48 incoming callers can work with 48 separate transactions. In addition, any of 500 active scripts could be available for startup with the next incoming call.

The Conversant system, when coded at high quality 32 kilobit (kbit) ADPCM,<sup>4</sup> can be configured with enough disk storage in the voice database for almost 70 hours of stored speech. All phrases are stored on disc until first needed, but main memory caching of the most recently spoken phrases minimizes disk I/O (input/output). Playing prompts and coding on 48 simultaneous channels (corresponding to 200 kilobytes per second [Kb/s] speech I/O) requires a streamlined and efficient voice database. The Conversant system

concatenates speech phrases into one prompt at run time. Slight delays within a prompt are egregiously obvious to the listener. Therefore the voice database is carefully designed for no delays. Similarly, for voice capture, the system must accommodate the coding and disk storage of incoming speech as it enters the system because delays in resource availability or buffer exhaustion result in lost speech.

#### **Voice Mail Applications**

The AUDIX and AUDIX Voice Power systems are rich in features that allow voice messages to be easily created, sent, or accessed. A user can create a voice message and send it to an individual or list of people—even an entire work group or department—by pressing only a few buttons. The message can be sent immediately, or can be scheduled for delivery up to a year in advance. Message recipients can be designated either by telephone numbers or by using the letters on the telephone pad to spell their names. AUDIX will verify to the sender that the correct address was input by responding with the name of the recipient.

The AUDIX systems are designed to be

---

cooperative partners with the user's PBX or central office switching system. A data link between AUDIX and the switching system provides a rich signaling mechanism between them. For example, AUDIX will automatically answer incoming calls and, based on information passed over the data link, will play a personal greeting appropriate to the circumstance (e.g., busy, on vacation), then record a message. Whenever messages are deposited in a subscriber's mailbox, AUDIX will instruct the switching system to turn on the message-waiting indicator for that subscriber. AUDIX can tell the switching system to transfer calls to different destinations. This allows various call control functions to be implemented. For example, the automated attendant application—a popular feature—gives a caller a verbal menu of options for routing a call (e.g., *Press 1 for Engineering, press 2 for Accounting*). It then directs the call to the requested destination.

Many businesses require close communication among employees, with customers, or with vendors who are not served by the same voice mail system. AUDIX's digital networking offers voice messaging features that allow users to communicate with individuals on other AUDIX systems. These systems can be located in the same building or elsewhere. The same high-quality voice recording and playback, together with a rich set of addressing and message control features, are available to subscribers, whether they are communicating with people in their office or across the world.

**AUDIX Voice Power Architecture.** AUDIX Voice Power architecture is based on the Conversant system hardware and software platform. This platform provides a cost-effective implementation for voice messaging functions with the addition of one or more IVP-4 (Integrated Voice Power-4 port) boards into the AT&T 6386WGS computer shown in Figure 3. Each IVP-4 provides telephone interfaces, digital signal processing, and low-level control processing for four voice mail access ports. An AUDIX Voice Power system can be configured with up to 3 IVP-4 boards to allow 12 service ports.

AUDIX Voice Power's basis in the UNIX operating system and ScriptBuilder allows the voice mail application to run concurrently with call accounting and other workstation tasks. This concurrency and multiple application feature enables smaller locations to easily justify the cost of the processor.

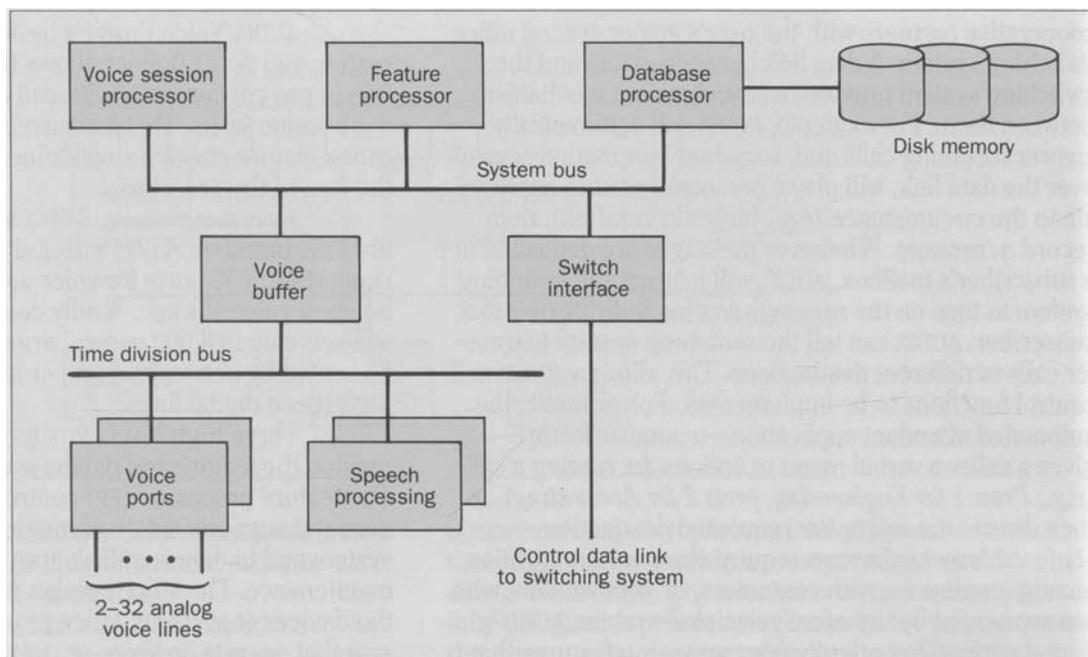
**AUDIX Architecture.** AUDIX's architecture (Figure 4) is based on AT&T's Definity® System 75 PBX. It supports 2 to 32 ports for voice access, and up to 470 hours of voice storage. A fully configured 32-port system will serve up to 4,000 users. Larger user populations can be served by networking two or more machines with high speed digital links.

Three high level CPUs (central processing units) provide the feature and database control for the system. The feature processor (FP) controls AUDIX feature operation, and supports data communications to the switching system and to data terminals used for administration and maintenance. The voice session processor (VSP) controls the devices (e.g., ports, voice processing, and voice buffers) that operate on voice or DTMF signals exchanged with AUDIX users. The FP and VSP functions are combined in small system configurations (i.e., less than 16 ports). The database processor (DBP) controls voice and data storage and retrieval in disk memory and digital networking interfaces.

Two primary busses in the system provide communication paths among the functional elements of the machine. The time division (TD) bus allows flexible interconnection of the voice ports with the speech processing cards and voice buffers. The system bus provides a multiprocessor control and data transfer capability.

When a call is directed to AUDIX, it terminates at a voice port, where the analog voice signal is digitized. A path is set up from the voice port to a speech processing port, and from there—via the TD bus—to a voice buffer. The SP port performs the many complex signal processing functions described later. The SP port then sends the voice information to the voice buffer, where it is tem-

**Figure 4. AUDIX architecture, based on AT&T's Definity System 75 PBX. Three CPUs provide the system's feature and database control. The feature processor (FP) controls feature operations, and supports communications to the switching system and data terminals. The VSP controls I/O devices. The DBP controls voice and data storage and retrieval in disc memory.**



58

porarily stored before being transferred into the DBP for storage in disk memory.

**AUDIX Speech Processing Technologies.** At the heart of the AUDIX and AUDIX Voice Power systems is a speech processing subsystem that provides rich functions to support the voice mail application. The important speech processing technologies are:

- Digital speech encoding, decoding and compression
- Silence compression
- Speed and volume control
- Automatic gain control
- Robust DTMF detection.

**Encoding, Decoding and Compression.** The AUDIX voice messaging system is based on voice store-and-forward processing. That is, voice is digitized, encoded, stored, and later decoded and played out. To ensure maximum density of voice information recorded on disk drives, the digitized encoded signal is compressed to the

lowest rate consistent with good output signal quality during playback decoding. It is essential for effective communication that the important nuances of speech (e.g., inflection, tone, and speaker identification) are maintained during this operation.

The encoding algorithm used by AUDIX and AUDIX Voice Power is subband coding (SBC) operating at a predetermined digital rate of 16 Kb/s. SBC is a waveform coding technique that takes advantage of speech properties as perceived by the human ear. The input signal is digitized, then compressed into the SBC signal.

Certain regions of the signal spectrum have been found to be more important than others when evaluating the perceived quality of encoded and decoded speech. For example, the lower frequency components of vowel sounds are more important than the higher frequency fricatives. SBC takes advantage of this property by more robustly encoding the spectral components that are

---

most important.

In the AUDIX implementation, the speech band is partitioned into subbands by bandpass filters in the speech processing subsystem. Each subband is sampled at its Nyquist rate, then digitally encoded using adaptive PCM encoding. The encoding rate for each subband is determined by the perceptual criteria for that band. When the signal is reconstructed at playback, each subband is decoded and reconstituted. The signals are then combined to closely replicate the original speech signal.

**Silence Compression.** Another AUDIX speech processing technique that improves voice information storage density is *silence compression*. This technique takes advantage of the silence that exists between words and sentences in normal speech. During recording, AUDIX determines whether voice is present by examining the short term energy of the input signal. If the energy of the input signal falls below a threshold, the input is considered silence. The periods of silence are compressed into a compact code that includes the length-of-silence interval. For good performance, the threshold is dynamically adjusted as a function of the background noise on the line. Hysteresis is applied to the threshold decision to prevent toggling of the silence detector when the energy of the speech signal is close to the threshold.

During playback, the silence code is converted into silence played out for the duration of the input silence interval. This technique provides up to 20 percent better stored voice density. That is, the speech encoding rate can be as low as 13 Kb/s.

**Speed and Volume Control.** Users can control the speed and volume of the playback signals via appropriate DTMF commands to AUDIX. Speed is controlled during playback simply by increasing or decreasing the silence intervals between words and sentences. This technique takes advantage of the silence compression algorithm previously described.

Volume control is also implemented during playback by changing the amplitude of the decoded signal before converting it back to analog.

**Automatic Gain Control.** The subband code, like most speech coders, works best when the input speech amplitude is within a certain range. But because calls can originate from different places and from people with different speech characteristics, the amplitude of incoming speech can vary dramatically. To improve performance, an automatic gain control (AGC) function is used to normalize the amplitude of the incoming signals to a standard level. This function also provides a standard output level for the listener. This means the volume control feature needs to be used less often, thereby simplifying the user interface to AUDIX.

**DTMF Detection.** The principal method to send commands to AUDIX is via the telephone DTMF tone pad. This critical control function should operate reliably under different conditions. DTMF signals that originate from overseas, or from acoustically coupled DTMF generators (commonly used with dial-pulse telephones to access voice mail systems), often have low amplitudes, high noise levels, or distorted frequency components of the dual tone signals.

To operate reliably under these conditions, a DTMF detector must be extremely sensitive and selective in detecting the control signals. If the detector either is too sensitive or not selective enough, then normal speech can emulate DTMF signals. If this occurs during recording, the recording function might be terminated. During playback, analog signals can be reflected back into the DTMF detector, and cause playback to cease.

## Conclusions

AT&T's historical leadership in providing voice systems is continuing today. The Conversant voice information system, AUDIX, and AUDIX Voice Power Systems provide flexible, high performance voice processing applications based on current developments in speech processing technologies.

## References

1. J. L. Flanagan and C. J. Del Riesgo, "Voice Processing—A Perspective on the Science and Its Applications," *AT&T Technical Journal*,

- 
- Vol. 69, No. 5, September/October 1990, pp. 2-13.
2. R. J. Perdue and E. L. Rissanen, "Conversant Voice System: Architecture and Applications," *AT&T Technical Journal*, Vol. 65, No. 5, September/October 1986, pp. 34-47.
  3. 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.
  4. N. S. Jayant et al. "Coding of Speech and Wideboard Audio," *AT&T Technical Journal*, Vol. 69, No. 5, September/October 1990, pp. 25-41.

Biographies (continued)

*System Terminals and Adjuncts Planning Department at AT&T Bell Laboratories in Denver, Colorado, where he is responsible for defining system requirements and architectures for voice messaging systems. He joined AT&T in 1977 with a B.A. and M.E.E. from Rice University, Houston, Texas.*