



RLE

currents

Volume 1, Number 1 • December 1987

The Research Laboratory of Electronics at the Massachusetts Institute of Technology

OVERVIEW OF RLE SPEECH RESEARCH

What would it be like to dial a voice-activated telephone or to produce a report using a voice-operated typewriter? Imagine a handicapped person who can communicate effectively by using a typewriter to operate a human-like voice. Is it possible for automatic language translators to assist people who speak two different languages to communicate in a conversational context? Equipped with an input device that recognizes human speech, machines can accept verbal commands from a human operator. This unusual ability, combined with human-like speech outputs produced by the machine, makes direct verbal communication between man and machine possible.

Although the applications of these man-machine interactions might be considered a step into the future, it will be the advances in the fundamental scientific understanding that will make them possible.

What happens during the transmission of speech—from encoding the speaker's thoughts, to motor control of the articulators, to the vocalization of the speech sounds, to the vibration of sound waves at the acoustic level, to the stimulation of the listener's auditory mechanisms, to the nerve impulses that relay the encoded message, to the listener's brain which ultimately decodes the speaker's message. Speech communication is an exclusively human process that has several levels of activity, and is far more complex than one would imagine. At RLE,

the ongoing interdisciplinary research in this area attempts to piece together the many questions involved in solving the mysteries of the speech puzzle.

Since 1957, RLE's Speech Communication group has investigated a wide range of topics. The Speech Communication group began after the MIT Acoustics Laboratory, led by Leo Beranek, formally disbanded. Some of

the Acoustic Lab's work in speech was transferred to a newly formed group in RLE, under the direction of Professor Morris Halle and then-RLE Director Jerome Wiesner. Studies included the acoustics of speech production, the electrical synthesis of speech, and the development of systems for the bandwidth compression of speech. Previ-

(continued on pg. 2)

Introduction to *currents*

Because RLE is a large and diversified laboratory, it is often difficult to keep up-to-date on research directions and significant questions that are being addressed. In *currents*, we seek to describe the intellectual concerns and activities of RLE's investigators, in a way that is accessible to the scientifically trained reader who is not necessarily an expert in the specific field, or whose educational discipline is different from the investigator's.

We want to portray not only what is happening, but the motivation for the research and the significance of the results. Common themes will be elicited, and a sense of future directions will be developed.

Current activities are derived, of course, from past experience, and RLE's history provides a fascinating view into the evolution of science and

engineering. We will look back into the past to better view the present, and to speculate on the future. Profiles of RLE investigators will be featured, providing the human dimension to our research. These personal interviews will introduce the reader to the essential driving force of RLE research, its investigators, in a way that provides an insight into our roots and directions.

New research initiatives will be described, especially when they involve common concerns and facilities. A special focus will be placed on new, interdisciplinary research directions that involve the collaboration of several investigators.

We look forward to hearing from you, and eagerly enlist your advice and support to guide the growth and direction of RLE *currents*.

Jonathan Allen

OVERVIEW

(continued)

Director's Message

For centuries, speech communication has been the object of awe, invention, and research. It provides one of the best windows on human cognitive capability, and its study requires a wide range of disciplines and facilities. RLE has unique resources for the study of human speech production and perception, and the laboratory has made important contributions to both basic understanding and applications.

RLE's Speech Communication group can be seen in a larger context of language and cognition research at MIT. There are strong interactions with other RLE research groups in sensory communication (including auditory psychophysics), auditory physiology, digital signal processing, and VLSI design, as well as other MIT departments (Linguistics; Brain and Cognitive Sciences; and Electrical Engineering and Computer Science). These interactions offer a unique perspective and a deep interdisciplinary understanding that is unmatched elsewhere.

A continuing theme within speech communication research is acoustic phonetics, or the study of the acoustic production of phonetic events. But, these sound waves which emanate from a speaker arise from a cognitive planning process, and lead to the motion of the articulators (tongue, lips, glottis, jaw, etc.). These trajectories, or "gestures," are difficult to observe, and this problem has resulted in the construction of an innovative magnetic apparatus that senses motion.

The basic studies of speech production and perception have led to major research efforts in text-to-speech conversion and speech recognition. These applications come together and provide the incentive for many fundamental research projects in sentence



Professor Jonathan Allen, Director, Research Laboratory of Electronics. (Photo by John Cook)

structure, the nature of the lexicon and word formation, the inventory and description of distinctive speech sounds, pitch and timing, the spectral characterization of individual and concatenated speech segments, and the nature of the human peripheral auditory apparatus. Elaborate computational strategies are used to exploit this understanding in the application context as well as to provide a productive research environment. In addition to speech synthesis and recognition, an important concern has been the development of prosthetic aids for the deaf and visually impaired.

Fascinating, fundamental, and increasingly useful in its applications, speech research thrives as one of RLE's major efforts.

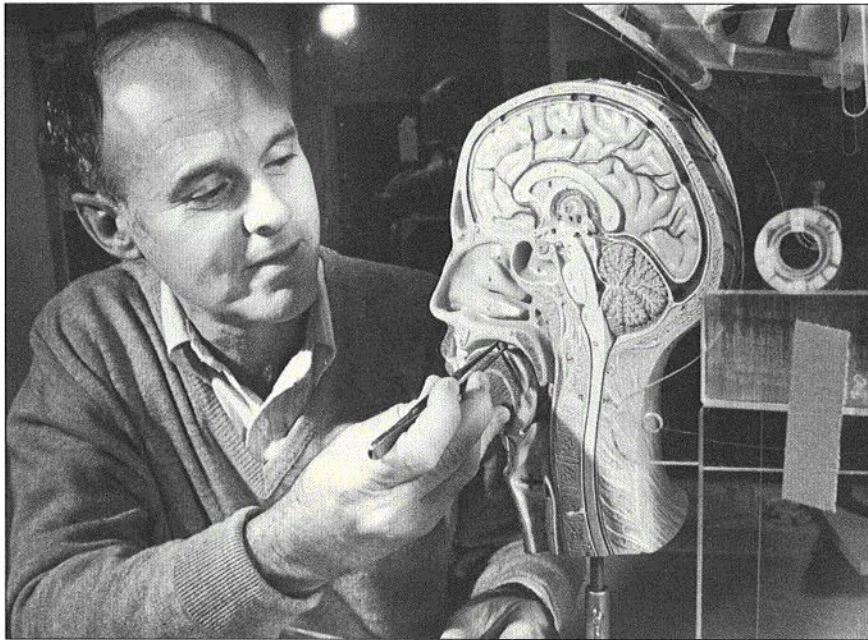
ously, there had been investigations into speech analysis at RLE, but this influx of new people brought a fresh perspective to this area of research. Linguists collaborated with acoustic engineers and phoneticians to study the relation between the acoustic, linguistic, and articulatory properties of speech.

Professors Morris Halle and Ken Stevens summarized the objectives of the speech group in a 1961 RLE Quarterly Progress Report as "to further the understanding of (1) the process whereby human listeners decode an acoustic speech signal into a sequence of discrete linguistic symbols, such as phonemes, and (2) the process whereby human talkers encode a sequence of discrete linguistic symbols into an acoustic signal."

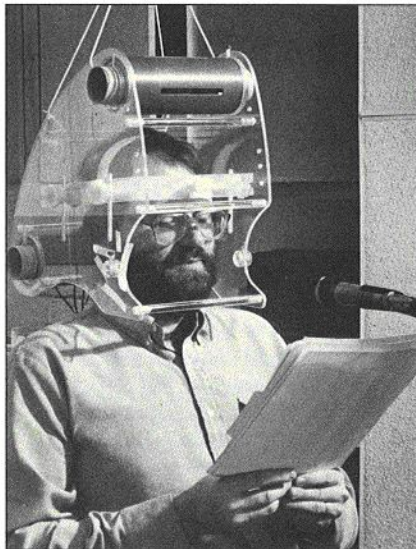
The next two decades heralded the extensive use of high-speed digital computers to achieve these objectives. Since that time, computers have proven indispensable to speech researchers in the collection and analysis of large bodies of data necessary to obtain statistically significant results. Computers have also been used to prove or modify hypotheses, thus eliminating the need to build complex, experimental equipment.

Current areas of research within the Speech Communication group are focused on automated speech recognition, human perception and production, and computer synthesis. One common challenge shared by these different approaches to the problems associated with speech communication is the achievement of naturalness and the use of the human model of speech as a prototype for experiments. The concept of naturalness in speech describes how humans communicate in a natural, continuous, and free-form style.

Professor Ken Stevens conducts research using synthetic speech and is concerned with developing a thorough understanding of the speech production process. This understanding is fundamental to the study of speech perception and in the development of new techniques for speech communication. In order to study what is known as acoustic variance in speech production, investigations are aimed at



Using a cutaway model of the human head, Principal Research Scientist Dr. Joseph S. Perkell demonstrates the placement of several transducers that measure the articulatory movements of a subject's tongue during an experiment of the alternating magnetic field movement transducer system. Post-doctoral Associate Dr. Thomas Carrell serves as a subject for one of these experiments. Two alternating magnetic fields are generated by large coils, and then picked up by the transducer. The induced signal is transformed into transducer coordinates in the mid-line plane of the subject. (Photos by John Cook)



the properties of what is common in the production of speech and speech-like sounds. The approach is to categorize these properties or features of the sounds, and to discover what those properties are. Specific research studies focus on how sounds are generated in the human vocal tract, and how the sound is ultimately encoded by the speaker and decoded by the listener.

In the study of speech production, a model of the vocal mechanism is used as the basis for speech synthesis. This work, conducted by Ken Stevens and Senior Research Scientist Dennis Klatt, is known as vocal tract modelling, and contributes to the understanding of how individual speech sounds are

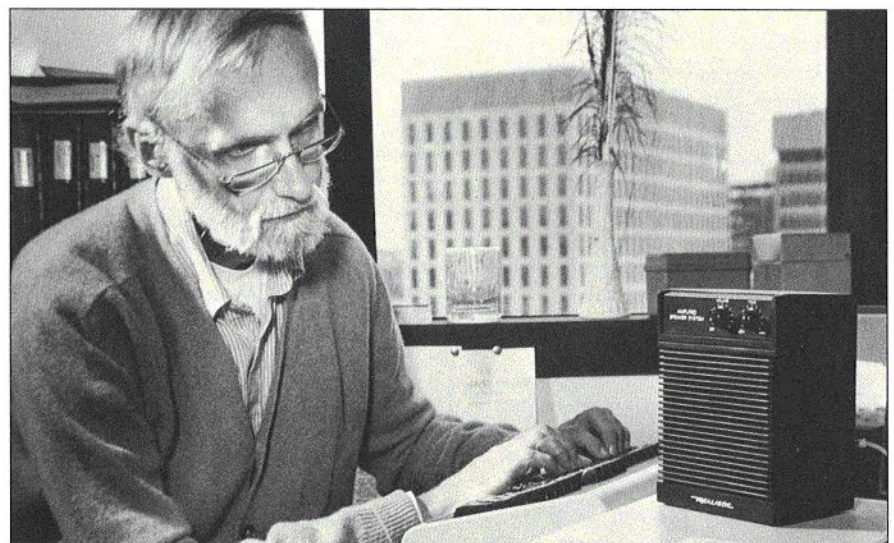
produced. Complex models of the vocal tract are developed in order to determine the properties of the vocal tract walls and the vocal cords.

A major problem has been the construction of techniques to convert unrestricted natural language text to an abstract linguistic description that is sufficient to drive a speech synthesizer. Techniques for analyzing words to reveal atomic chunks such as prefixes, roots, and suffixes have been developed by Professor Jonathan Allen, along with letter-to-sound rules and rules for assigning word-level stress developed by former RLE staff member Sheri Hunnicutt. Together with phrase-level parsing procedures, these text analysis tools provide the abstract linguistic specification that is common to both text and speech message realizations.

Dr. Klatt has developed a rule system for constructing the speech waveform from this linguistic basis. Currently, he is investigating the properties of sound generated by the female voice in order to achieve a natural sounding synthesis. Some degree of naturalness has already been achieved with the male voice, but the different speech properties of the female voice make this research more difficult. He also pursues research in speech perception, and the specification of significant acoustic properties for various phonetic distinctions.

Principal Research Scientist Victor

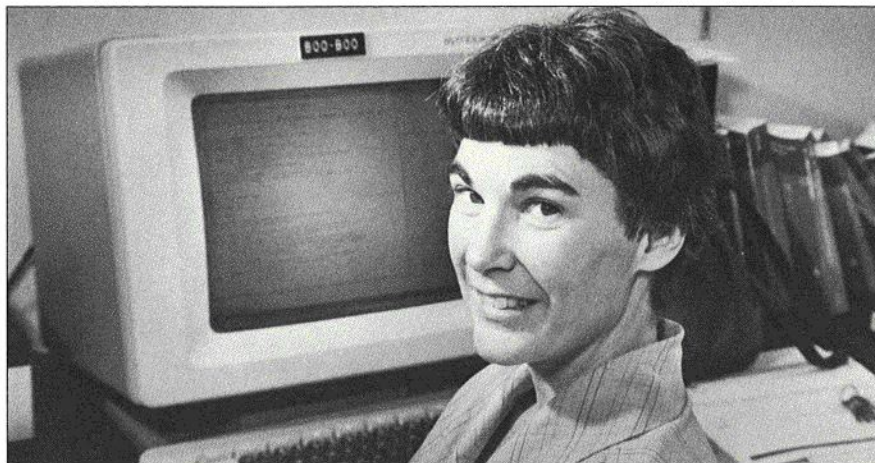
(continued on pg. 4)



The synthesized voice from the speaker (black box on the right) articulates the English text typed from a keyboard by Senior Research Scientist Dr. Dennis H. Klatt. The program, developed by Dr. Klatt at RLE, is available commercially. (Photo by John Cook)

Zue's investigations into automatic speech recognition involves the development of a sophisticated computer system that will be able to understand and respond to speech without limitations on an individual's vocabulary or speaking style. One of the goals that is sought in his research are what components of the speech signal are needed to produce intelligible and natural-sounding speech, and whether all parts of the speech signal are necessary to understand the speaker's message.

Dr. Zue's research also involves the collection and analysis of large amounts of data—then breaking it down to understand the linguistic information contained in the signal. Investigations into auditory modelling conducted by Research Associate Stefanie Seneff



Research Associate Dr. Stephanie Seneff's research on the modelling of the human auditory system contributes to the improved performance of automatic speech recognition systems. (Photo by John Cook)

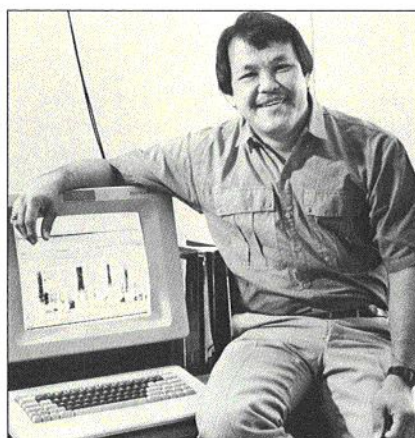
Focus on Speech Recognition

Dr. Victor Zue, Principal Research Scientist, conducts research in computer recognition of human speech, which focuses on the acquisition and application of acoustic/phonetic knowledge.

Dr. Zue commented on his group's research:

"Our long-term research goal is to develop a graceful environment for human/machine interaction. Because computers continue to play an ever-increasing role in our lives, it is essential to create a habitable environment so that relatively unsophisticated users can take advantage of the available computing power. We believe that interaction between a human operator and a computer can be made more natural by enabling the computer to communicate with humans via oral speech as an alternative to text."

"Research conducted over the last two decades has resulted in significant improvements in speech synthesis and encoding technology, so that good quality synthetic or vocoded speech can now be obtained. Speech recognition, on the other hand, is still in its infancy. While algorithms for some speech recognition tasks are well understood, procedures to achieve human-quality performance of continuous speech recognition, without restrictions on vocabulary and speakers, remain elusive."

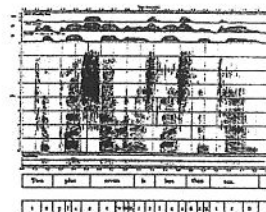


Principal Research Scientist Dr. Victor Zue displays a computer image of a sound spectrogram used in the study of speech recognition. (Photo by John Cook)

"Most available speech recognition devices utilize little or no speech-specific knowledge. Instead, they rely on general purpose pattern recognition techniques. We believe that this technology is limited, fragile, and not readily extendable to more complex tasks. In our opinion, the three major criteria of a successful system—speaker independence, large vocabulary, and continuous speech—cannot be met without many years of active research. We also believe that the solutions to these difficult problems lie in our understanding of the procedures of human

speech production and perception, as well as our ability to identify, quantify, and incorporate specific knowledge about how context affects the acoustic characteristics of speech sounds."

"However, we must keep in mind that there are powerful mathematical procedures which can help us organize and optimize the knowledge we do have. Our approach to speech recognition, therefore, emphasizes the combination of the two: expressing our knowledge in a coherent framework while allowing mathematical tools to make use of this knowledge in an optimal fashion."



The above spectrogram is a voice print-out that displays different features or attributes of the speech signal.

Ongoing projects in Victor Zue's group include the modelling of the human peripheral auditory system, the development of algorithms to recognize specific classes of speech sounds, and the building of a phonetic recognition front-end system.

contribute to the research in speech recognition. In recent years, modelling of the human auditory mechanisms has helped to improve the "robustness" or performance of speech recognition systems. Dr. Seneff is now working on a vowel identification model used in combination with a successful auditory model developed at RLE.

When Senior Research Scientist Joseph Perkell first came to work in RLE under Ken Stevens in the mid-1960s, he was assigned to hand-trace cineradiographs produced by Professor Stevens at Gunnar Fant's laboratory in Sweden. This laborious and time-consuming task involved tracing hundreds of frames from x-ray motion pictures depicting the motions of speech articulators (for example, the

tongue, palate, jaw, and lips). These investigations were aimed at developing new techniques to transduce the movements of the articulators (convert the input articulator movements into a speech waveform output), and to look at patterns of variability associated with speech production.

Today, Dr. Perkell is still involved with these studies, but the techniques used are drastically different and quite innovative. Various methods such as scanning x-ray microbeams and alternating magnetic fields have been developed to study these movements. Dr. Perkell is currently using pellet tracking techniques to investigate the various aspects of different articulators' movements, how these movements are related to different speech sounds, and

their acoustic consequences. Probe coils, or pellets, are attached to the articulators to track their movement. Each coil can generate three or four signals at varying rates that must be recorded with great precision. Finally, a computer digitizes the speech, enabling the researcher to sample different signals at different rates of speed.

The achievements and investigations of the Speech Communication group at RLE not only benefit other scientists involved with speech communication, they also provide researchers in other fields with stimulating insights and new methods of approach.

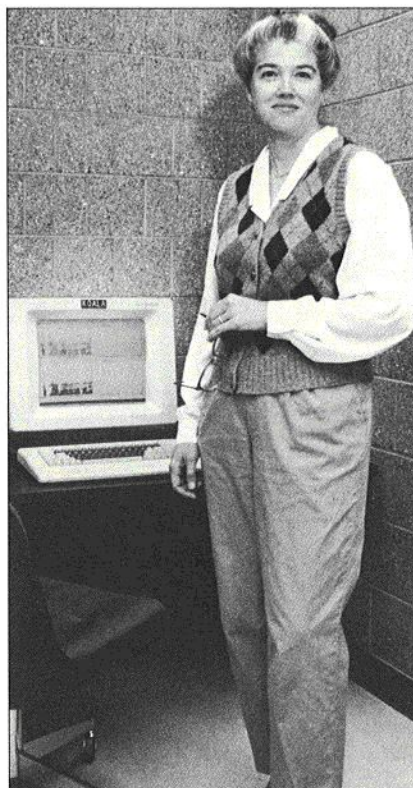


Pult of Cursonality

Unless your tar was just cowed away, or you think some things are noth wort knowing, you may find that spoonerisms have a function other than just being amusing.

The Reverend Spooner, an Oxford don in the late 1800s, gave his name to slips of the tongue like, "You have hissed all my mistory lectures" instead of "missed all my history lectures." Modern day equivalents such as, "We had a cone phall" for "phone call," and "Are the boon duggies available?" for "dune buggies" illustrate the fact that speakers don't always say what they intend to say. What do these malfunctions in the speech production planning system tell us about the process that goes on in our heads when we speak?

First, the sentences that we speak are represented not only as strings of words, but also in the form of individual sounds called phonemes (for example, the /d/ and the /b/ of "dune buggies"). Sometimes, these individual pieces can break loose from their moorings and move to another place in an utterance. Second, there is a framework, or scaffolding, in which sounds can move around. Only sounds that belong to similar positions in this framework can exchange with each other. For example, in "phone call," we might get "cone phall" or "pholl



Research Associate Dr. Stefanie Shattuck-Hufnagel is a psycholinguist who studies speech production. She uses data from spoonerisms and tongue twisters, as well as prosodic patterns, to build models of the speech planning process. (Photo by John Cook)

cawn," but not "lone cawf." Finally, basic word forms are represented separately from their endings, or affixes, which are attached to them in this framework. This can be seen in errors like, "You have to use rop rokes" for "rock ropes," where the /s/ ending is ignored.

Spoonerisms like these, and other slips of the tongue, can tell us about the planning steps that go on in our heads when we speak. Tongue twisters are another source of data. Why is, "She sells sea shells" difficult to say five times quickly, but, "She bells bee shells" easy to say? Is it because /s/ and /sh/ are very similar, while /s/ and /b/ are not?

As we learn more about the way we plan our speech, we will be better equipped to find ways to help others who have trouble speaking because of a stroke, head wound, or developmental problem. It is difficult to develop methods to help others when we know so little about the way the normal speech process works, and what can go wrong. Meanwhile, spoonerisms happen every day, right in front of our very ears—so listen carefully!

Contributed by
Stefanie Shattuck-Hufnagel,
Research Associate,
RLE Speech Communication Group

FACULTY PROFILE:

Kenneth N. Stevens

Professor Ken Stevens, a Toronto native, came to MIT in 1948 as a Teaching Assistant in the Electrical Engineering Department after receiving his Master's in Engineering Physics at the University of Toronto. Since joining the RLE faculty in 1958, he has been central to the development of speech communication research at the laboratory.

• *What was the focus of your graduate research at MIT?*

At first, I worked in the MIT Acoustics Lab. The speech work in the Acoustics Lab started in 1948 with Leo Beranek, who had an Air Force contract to study problems related to the intelligibility of processed speech. He worked on that project with some students over the years, and I became involved in that work as a graduate student in 1951.

At that time, (about 1949 or 1950), Gunnar Fant visited MIT to study the acoustics of speech production. I became interested in the perception side of speech and worked with Leo Beranek and J.C.R. Licklider on the perception of speech-like sounds. I wrote my doctoral thesis on the perception of sounds that had speech-like characteristics. Beranek's work, combined with Gunnar Fant's studies on the acoustics of speech production and my research on the perception of speech-like sounds, and some additional work on the intelligibility of speech, formed the beginning of speech work at the Acoustics Lab.

• *Did you have a mentor?*

I would say that it was Leo Beranek, who was one of the directors of the Acoustics Lab. He taught courses in acoustics, and one of his interests was speech. When I first came to MIT, I hadn't thought about going into acoustics, but Beranek needed teaching assistants in his acoustics course. Originally,

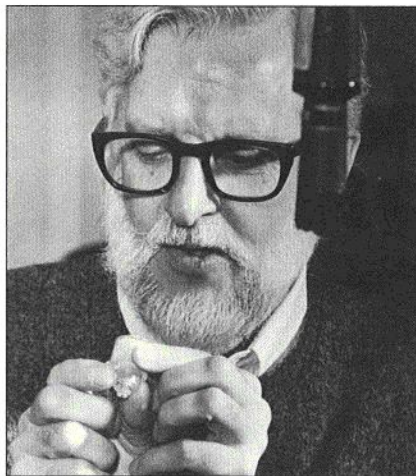
my background was engineering physics, but not so much in acoustics.

• *Why did you choose to teach?*

I really liked doing research with the graduate students here at MIT, and so the teaching fit in with that. It was a good place to do research, and so I did some teaching.

• *What was the nature of your research as a Guggenheim Fellow from 1962–1963?*

I worked in Gunnar Fant's laboratory at the Royal Institute of Technology in Stockholm. One of the things that I studied was speech movements with cineradiographic (x-ray) motion pictures. Recently, we haven't collaborated, but he did visit here in 1982, and we do keep in touch with each other.



Professor Kenneth N. Stevens investigates a device used in speech experiments for the mechanical constriction of airflow. (Photo by John Cook)

• *In the early days of RLE's speech communication research, what was the focus of its investigations?*

Some of the work in speech at the Acoustics Lab became part of RLE. The Acoustics Lab had disbanded, and I remember talking to Professor Wiesner at the time about the possibility of this small group of researchers working in speech coming under the umbrella of RLE. He thought that it was in line with the other communications work that was already going on at RLE. There

was already some speech work being conducted at RLE, and a group of people met regularly to talk about the problems of speech. One of the individuals in this group was the Director of the Modern Languages Department, William Locke. Bob Fano was also part of this group.

If you look back at early RLE reports, you might find a section of linguistics with Noam Chomsky, and Morris Halle was there too. We've always had interaction with Morris Halle, and I guess our work could be characterized by trying to find or quantify more closely the relations between the acoustic and articulatory events in speech and the linguistic descriptions that underlie speech events. Morris Halle had a strong influence on the early directions of the speech group, although his interests centered on the phonological aspects of speech. Morris Halle has always had a strong influence on my own thinking, and Gunnar Fant.

Even in those early days, we were interested in speech synthesis. So, apart from understanding the fundamental aspects of speech production and perception (which we are still doing), the application of speech synthesis was an early activity, even when Gunnar Fant visited in the early 1950s. That developed even further with Jonathan Allen's and Dennis Klatt's interests in speech synthesis. Allen and Klatt, together with the RLE students, brought the speech synthesis work to a culmination with some practical results. Then, within the last five years, there has been an increasing interest in speech recognition and the application of speech to computers. So, this brings to bear much of the basic information that has accumulated in various places over the years to the practical problem of speech recognition.

As people here work on the problems of synthesis and recognition, we realize that there are still some basic aspects of speech production and perception that we still don't understand. An example is the recent work of Dennis Klatt. He found that although he could get reasonable naturalness in the synthesis of male voices, it was a problem to achieve good naturalness for female voices. So, it was necessary

to go back and study in greater detail the properties of sounds that are generated by females. Then, that basic information could be used to improve the synthesis of female voices. Similar things have happened in speech recognition.

Also, as the speech recognition work continues, we realize that we must rely heavily on what the linguists are able to come up with—phonological representations of speech that bring to light, in a natural way, some of the modifications that occur in speech when we speak in a conversation. What happens when you put speech into context, and other kinds of modifications that are made in the sounds when speech occurs in a natural context.

• How would you characterize your research in the acoustical aspects of speech production in contrast to other RLE research groups (auditory physiology, sensory communication, and digital signal processing)? What is the nature of your interaction with these different groups?

We began to look at how sounds were generated in the vocal tract and the actual acoustic mechanisms of sound production, and in fact, we are still continuing that work. We are interested in the link between what happens in the sounds and what are the underlying linguistic descriptions in terms of phonemes and features. Our goal has been to join the understanding of the sound and the linguistic description. One of the big influences over the years in this area has been the people in linguistics, particularly Morris Halle and Jay Keyser.

In relation to auditory physiology, we are interested in the stages in processing of the sound, leading ultimately to a linguistic description. One of the stages through which sounds must pass is the ear, obviously. The shaping of sounds in the auditory periphery could form an initial step in the chain of processes that produce a description in categorical terms. Our concern with auditory physiology is to keep in touch with what the investigators are doing, and, where possible, to incorporate their research into our models.

In terms of digital signal process-

ing, the speech signal has to be processed initially by digital methods. In fact, when Alan Oppenheim started on the faculty, he was in the speech communications group. Then, he branched out into digital signal processing, and it became an important field in its own right.

• How would you characterize the diverse background of investigators who are attracted to the field of acoustic phonetics?

Many linguists are not concerned with the actual details of sound. Phonologists think of speech as being a

sequence of sounds, and do not go beyond this characterization. They address the different kinds of regularities and constraints on patterns of sound; how a language is described in terms of constraints on the sequences of speech sounds that are allowed; and, how these sequences are changed when you place the words into context.

But, more recently, there is a group of phonologists who are becoming interested in phonetics. They are trying to explain some of these phonological regularities in terms of constraints on either the listener or the speaker, and the constraints of the

(continued on pg. 8)

Tools of the Speech Trade

Research in speech communication requires its investigators not only to collect large bodies of data, but also to access and categorize smaller pieces of scientific information. The development of interactive software "tools" to collect and analyze this information provides a convenient yet necessary environment to determine significant results.

SPIRE allows researchers to digitize and transcribe speech, and then to compute and display the parameters of a speech waveform or signal. These parameters include energy in different frequency bands, zero-crossing rate, and the fundamental frequency of voicing. SPIRE also permits examination of more than one signal at a time. The system has evolved over the last four or five years, and many people both inside and outside of MIT have found it tremendously useful. It is one of the best-selling programs offered by MIT's Technology Licensing office.

SEARCH facilitates the interactive and exploratory analysis of speech data by categorizing data into a set of user-specified criteria. In addition, it is a convenient statistical package that displays data

in a variety of ways. Histograms and scatter plots are just two of the methods that can be used to analyze the statistical behavior for different classes of sounds.

ALEXIS provides many options to computer and display the distributional constraints of lexicons. For example, an investigator can determine the frequency of the occurrence of sounds patterns, or study the phonotactic constraints imposed by the language, or test the effectiveness of phonetic and phonological rules.

SYNTH produces user-controlled synthetic speech with a cascade/parallel speech synthesizer developed by Senior Research Scientist Dr. Dennis Klatt at RLE.

LAMINAR synthesizes speech from various vocal tract configurations. The system computes responses from static vocal tract configurations. It is hoped that with further development, the system will provide a model for many time-varying articulatory movements.

Courtesy of "Summary of Research in Speech Recognition" by the RLE Speech Communication Group

actual mechanics of how these sounds are generated.

For example, certain sounds influence others. A classic example is “did” followed by “you” becomes “didju.” Phonologists would simply say that there’s a rule that says /d/ plus /y/ will change to /j/. Now, people are trying to explain these changes in terms of the mechanics of the ear and the vocal tract. So, there has been a coming together of people who work on the speech area and those individuals who work in that part of linguistics.

• Your ongoing research involves acoustic variability and invariance in speech production. Can you explain the nature of this investigation?

When different people say a particular sound, or when one individual says the same sound in different words or sentences, it appears as though the sound undergoes a lot of change from one person to another, and from one context to another. We are interested in exploring what is common between all those productions of the sound. In spite of the variability, there are some attributes that remain invariant. That’s what we pick up on when we listen to each other. It doesn’t matter who says the sound, it doesn’t matter what word the sound appears in, we still hear the same sound.

Our approach is to categorize these sounds by certain properties or features, and to discover what those properties are. We believe there is an inventory of properties or features that is an integral part of the human speech production and processing system. Different combinations of properties are used in different languages, but there is a fixed inventory of properties.

• Can you describe the research that you and Dennis Klatt have conducted on vocal tract modelling?

There are two sides to vocal tract modelling. One question that we are trying to answer is: by developing complex models of the vocal tract itself (including the properties of the vocal tract walls and properties of the vocal cords)—can we further understand the mechanisms of the generation of indi-

vidual speech sounds?

Then, there is the broader aspect of speech modelling (you might call it speech synthesis). How can we build a device that will take the printed words as an input, and put the words into speech? Not only do you have to know how to produce the individual sounds, but you also have to put these sounds together with the right sense of timing and intonation. That’s a problem that Jonathan Allen and Dennis Klatt have worked on for the last twenty years with some success.

We are now currently interested in moving toward more natural types of speech, looking at similar properties to understand this whole process of how sounds become modified within natural speech.

• Does your research also include the study of speaker verification and recognition?

It automatically comes out of some of our work. If you’re looking for the invariants, you’re also studying variability when you examine how one speaker differs from another. Over the years, I’ve had one or two thesis students in this area, but I haven’t delved into it very much. This whole business of speaker verification using spectrograms, or by some other method, is a difficult area, and I’m not certain these methods will lead to reliable identification of speakers.

• Data collection in the field of speech processing is a tremendously labor-intensive and time-consuming task. What are some of the scientific tools that help you to collect and analyze this large body of data?

With the capability to store large amounts of data in computers, it has been possible to record a database with large numbers of talkers and lots of sentences, and then label all of the

sounds in that database. As a result, it is possible to access that database, request a specific sound, and perform some statistical analysis of the properties of that particular sound. Victor Zue and his group have assembled a large database for that purpose.

SPIRE is a basic tool that enables us to look at individual speech sounds in many ways—spectra and spectrograms, for example. The SEARCH program is an extension of SPIRE. It allows us to search a large database and plot distributions of different acoustic properties for speech sounds in different phonetic contexts.

• Does your research involve speech aids for the handicapped?

I have worked on speech training aids for the handicapped, especially for children who must learn to speak, but cannot hear. One approach is to provide them with some type of feedback of their speech patterns by abstracting and displaying information from the spectrogram so that they can see when they speak properly. In my consultancy with BBN, we didn’t use spectrograms because the technology wasn’t available to generate it fast enough at the time. So, we displayed simpler patterns like the pitch and timing of speech.

• What is the nature of your consultancy at Bolt, Beranek, and Newman?

My more recent work with BBN was to develop methods for measuring people’s hearing at very high frequencies, far beyond what is needed for speech. It is important to be able to do this because there are some invasive things that influence hearing. High-intensity noise or certain drugs, like aspirin, can influence certain people’s hearing if large doses are taken. In some cases, it influences hearing first at the very high frequencies. Then, it gradually spreads down into the lower frequencies. So, it is important to be able to measure those effects on hearing at very high frequencies.

• Are you excited about a current project that you’re working on?

In the past, we have tried to examine

speech sounds and their properties as they occur in simple utterances (consonants, vowels, syllables, etc.). We are now currently interested in moving toward more natural types of speech, looking at similar properties to understand this whole process of how sounds become modified within natural speech. That's the thrust now, both in recognition and synthesis.

I'm enthusiastic about "rounding off" our previous work. We've learned a lot about individual speech sounds and how they are produced and perceived. There are still many loose ends to pull together before we move on to the next stage. At this moment, I'm interested in pulling together those loose ends and putting them in a book. Then, I would like to move on to the study of speech in a conversational context.

• How do you measure success in terms of testing and developing your ideas?

One measure of success is whether the applications in speech synthesis or speech recognition can actually work and be used by people. In the case of synthesis, there has been some reasonable success. In speech recognition, perhaps not so much. Another mea-

We are getting to the point where we have exhausted the study of individual speech sounds or simple utterances. We now want to move into more conversational speech, where the sounds that we generate and the ones that we hear in normal conversation have been modified quite a bit.

sure of success is that you understand the concept of how this whole speech process works, and you fill in the gaps of your knowledge of the process; gradually piecing together this jig-saw

puzzle. Whether or not it leads to an application is not the point, but rather, if all pieces of the puzzle fit together.

So, you could say that one measure of success is if all of these different pieces of information—whether they be from speech physiology, speech acoustics, speech perception or phonology—fit together into a coherent picture. Obviously, we are still trying to build that picture, and I believe that it's beginning to fit together. To some extent, we are happy about that, and to some extent we are frustrated because there's still so much to learn.

• What has been the most challenging project that you've worked on?

One of the most challenging things is to try to uncover the basic invariant properties from the speech signal, in spite of all of its variability. Particularly for some sounds, it's been a real challenge. For example, what distinguishes a /p/ from a /t/ from a /k/? It's the kind of question that we still don't have a good answer for.

• During your professional career, what do you consider to be the major breakthroughs or milestones that have significantly contributed to or changed the field of acoustic phonetic research?

There is no question that the ability to use the computer to look at data conveniently and quickly, and to perform signal processing, is a major breakthrough. The computers give us access to larger databases, and allow us to test hypotheses with a much faster turnaround time. The disadvantage is that it is too easy to test ideas, and we don't spend enough time thinking about them before they are implemented, because they are so easy to implement.

More broadly, I would say that Gunner Fant's work on acoustics and the insights of Roman Jakobson into the linguistic description of sounds have represented major milestones. In the past decades, researchers have been trying to build on these ideas.

• What do you see as the direction of future research in acoustic phonetics, or speech processing in general?

In the next decade or so, we will have to understand more about these phonological/phonetic changes that occur when we speak in conversational speech. We are getting to the point where we have exhausted the study of individual speech sounds or simple utterances. We now want to move into more conversational speech, where the

Up until now, acoustics and signal processing people have been the major components in speech research. To proceed further, we have to involve people from other disciplines more than we have in the past.

sounds that we generate and the ones that we hear in normal conversation have been modified quite a bit. In other words, the listener perceives only a fragment of the original sounds that occur rapidly. What the listener picks up on is only some of the sound because of redundancy, and because the listener might know something about the topic that is being discussed. It is this area that we will have to work on. Up until now, acoustics and signal processing people have been the major components in speech research. To proceed further, we have to involve people from other disciplines more than we have in the past.

• What do you like most about RLE?

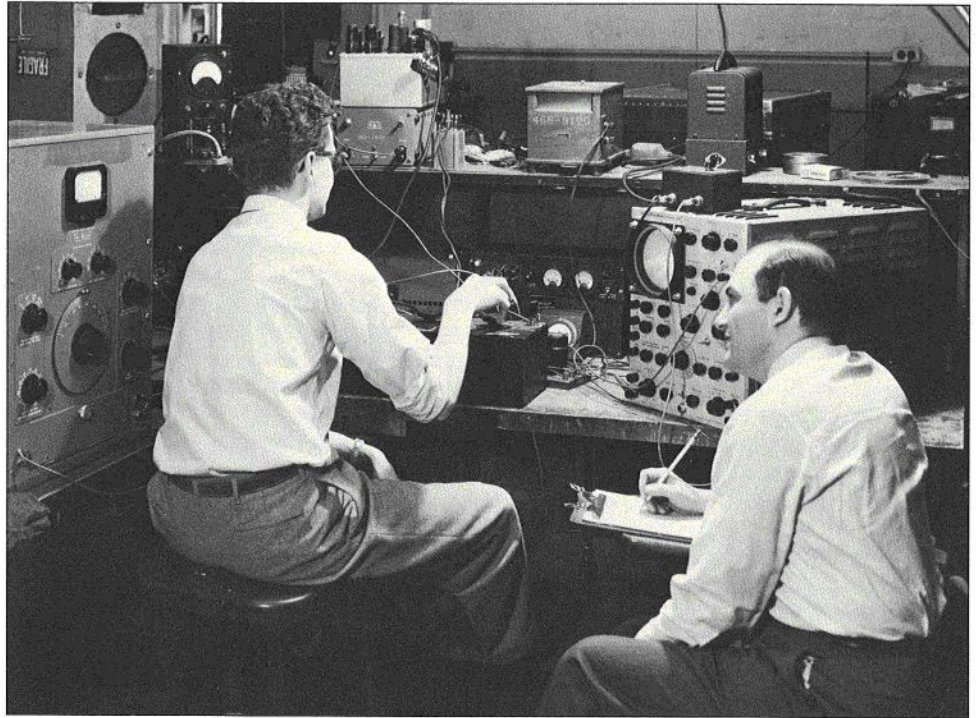
The thing I like most is the proximity of colleagues who are in fields related to mine, and who are among the very best in the world—people who really understand hearing, people who understand linguistics and acoustics—and to interact with those people and with such very good students. That's what makes the place exciting.



History of Speech at RLE

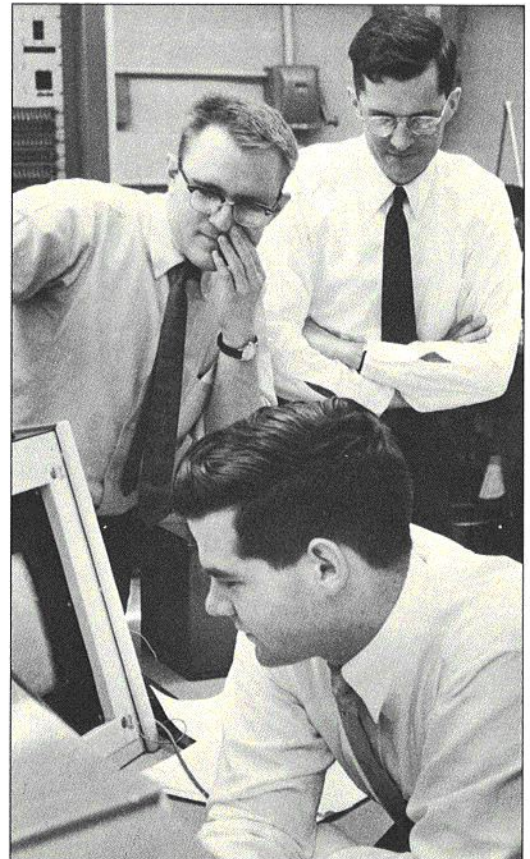
1957

G. W. Hughes (left) and Professor Morris Halle (then Associate Professor of Modern Languages) measure the power spectrum of a speech sound recorded on a loop of tape. (Photo by Benjamin Diver)



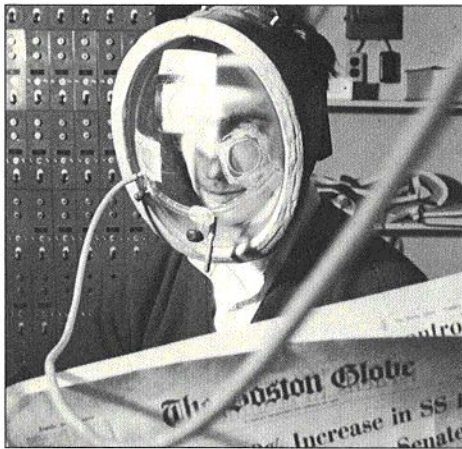
1958

Ken Stevens (left) points out one of the unique features of a tape recorder used for speech research purposes. John Heinz (center) and Leo L. Beranek (right) look on. (MIT Historical Collections)



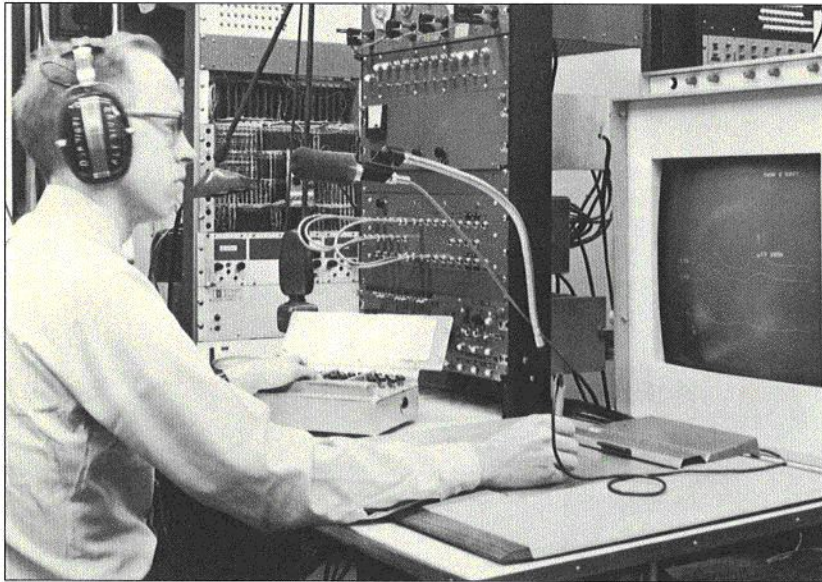
1960

Pete Brady (front), Gordon Bell (left), and Ken Stevens (right) use RLE's TX-0 computer to analyze speech sounds. (MIT Historical Collections)



1968

Dennis Klatt wears apparatus used to measure the air flow during speech production. (Photo by John Cook)

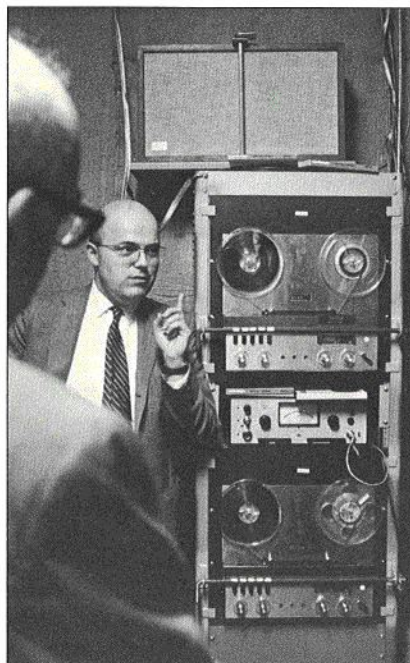


1970

William Henke uses a PDP-9 computer to measure the acoustic parameters of speech. (Photo by Richard Geraigery)

1970

Professor Jonathan Allen briefs visitors from the Joint Services Electronics Program on his work in text-to-speech conversion.



Publications

The RLE Communications Group will oversee publication activities of the new RLE Collegium. Recent publications that are currently available are the annual *Progress Report*, *Technical Report Abstracts*, the *RLE Brochure*, and *Collegium Prospectus*. *Progress Report #129* describes research programs at RLE for the period January 1, 1986 through December 31, 1986. This report contains both a statement of research objectives and a summary of research efforts for each of the research groups within RLE. Faculty, staff, and students who participated in these projects are identified at the beginning of each chapter, along with sources of funding. *Technical Reports Abstracts* lists abstracts of technical reports published by RLE for the period 1983 to 1987. The *RLE Brochure* describes RLE's forty-year history from the Rad Lab days to the present, with a focus on current research programs. The *Collegium Prospectus* provides information on collegium membership and benefits.

The Communications Group also oversees the RLE Document Room. This library maintains an important and historic collection of Radiation Laboratory Reports, RLE Technical Reports and theses, reprints by RLE authors, contract reports, and computational notebooks from RLE research. In addition, the Document Room library maintains a significant collection of books, journals, and reports from other universities, government agencies, and industry, all of which are useful to RLE investigators and students in their research. Much of the material found in the Document Room is rare and not readily available elsewhere. Documents in the collection may be made available to non-RLE staff on a "reference only" basis. The table-of-contents service provides access to the more than 100 journals, newsletters, and newspapers received regularly.

The RLE Communications Group welcomes inquiries regarding RLE research and publications.

Barbara J. Passero
Communications Officer
Research Laboratory of Electronics
36-412
Massachusetts Institute of Technology
Cambridge, MA 02139
(617) 253-2566

New Collegium at The Research Laboratory of Electronics

The Research Laboratory of Electronics (RLE) has established a new Collegium to promote innovative relationships between the Laboratory and business organizations. The goal of RLE's Collegium is to increase communication between RLE research staff and industrial professionals in electronics and related fields.

Collegium members have the opportunity to develop close affiliations with the Laboratory's research staff, and can quickly access emerging results and scientific directions. This kind of increased professional interaction provides RLE Collegium members with the most up-to-date technical information, often in areas not fully addressed by business and industry.

A wide range of publications will be available to RLE Collegium members. Members also have access to educational video programs and RLE patent disclosures. The Collegium's newsletter and on-line calendar/bulletin board will keep members up-to-date on seminars and other events sponsored by the Collegium.

Semi-annual seminars will offer Collegium members state-of-the-art information in the high-technology areas essential to contemporary business practice and planning. RLE will extend an open invitation to all Collegium members who wish to visit the laboratory's facilities and meet with its staff. Arrangements can be made for brief discussions with faculty members, or for long-term visits to acquire new knowledge and techniques through collaboration with RLE's scientists.

The RLE Collegium membership fee is \$20,000 annually. Members of MIT's Industrial Liaison Program can elect to transfer 25% of their ILP membership fee to the RLE Collegium. After an initial one-year membership, a three-year commitment will be required. Membership benefits are supported by the Collegium fee. In addition, these funds will encourage new research initiatives and build new laboratory facilities within RLE. These new facili-

ties will seek to promote interdisciplinary cooperation, the hallmark of all RLE research.

For more information on the RLE Collegium, please contact RLE Headquarters or the Industrial Liaison Program at MIT.

SYMPOSIUM: "Speech Communication and Processing"

The Research Laboratory of Electronics and the Industrial Liaison Program are hosting a symposium on Monday, December 14, 1987 at MIT's Kresge Auditorium. The topic of the symposium is "Speech Communication and Processing."

The need for natural and reliable man-machine communication is vital because of the extensive use of computers in all segments of our society. The development and increased availability of audio response devices, text-to-speech capabilities, and speech recognition systems meet this prevailing need.

Signal processing, coding theory, human speech production and perception, acoustics, phonetics, linguistics, experimental computer systems, and VLSI technology are among the many scientific areas that must work together in order to design successful systems. Advances in these fields have made possible improved systems to communicate with hearing-impaired individuals, as well as the efficient coding and enhancement of speech signals.

RLE's symposium on "Speech Communication and Processing" features several RLE investigators who describe their research, assess current speech technologies, and probe future directions in this broad-ranging area. Ongoing research projects are demonstrated, and RLE investigators are available to discuss contemporary topics.

RLE *currents*

RLE *currents* is a biannual publication of the Research Laboratory of Electronics at the Massachusetts Institute of Technology.

Jonathan Allen Editor-in-Chief

John F. Cook Photography

Everett Design Design

Dorothy A. Fleischer Staff Writer
and Editor

Barbara J. Passero Production
and Circulation

Donna Maria Ticchi Managing
Editor

Henry J. Zimmermann Advice and
Perspective

Inquires may be addressed to RLE *currents*, 36-412, Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.



RLE

© 1987. Massachusetts Institute of Technology.
All rights reserved worldwide.