Guest Editorial

That Wondrous Signal: Speech

Andrew Sekey

AS AN ENGINEER, I have never ceased to be amazed by that marvel of nature: speech communications. Think first of the signal itself: it alternates between quasi-periodic vowel-like sounds and fricatives looking much like random noise, interspersed with high-energy plosives which are preceded by brief but important silences. In the time-frequency domain we often see a "formant structure" of prominent energy peaks [1], moving about with more regularity and consistency than the waveform itself. While these carry information about the shape of the vocal tract, the underlying rapid pulsation of energy reflects the "voice pitch," which allows us to vary intonation.

The robustness of the speech signal defies engineering comparison: you can lowpass or highpass filter it at 1800 Hz, yet, like the earthworm cut in two, each half "lives" and remains about 70% intelligible. You can also subject it to peak clipping, center clipping, nonlinear distortions, or periodic or random interruptions, and still make out most of what is said. With special devices you can speed it up 50-100% by cutting out repetitious parts, or slow it down by repeating them even more. Nature, it seems, has invented time-frequency diversity coding long before we did!

Where, then, is the information in the speech signal? Researchers have been inching their way towards the answer for years, and computers have speeded up the search [2]. But what do we mean by "the information" anyway? After all, speech tells us a great deal more than what remains after it is transcribed to text. From the spoken sound we can infer physiological and psychological data about the speaker: sex, approximate age, size of vocal tract, geographical origin of his speech habits, perhaps his identity as well as his mood, attitude to the subject matter and to listeners, and so on. Often, we can also tell something about the physical space occupied by the speaker, background sounds, and media of recording and transmission. All this is available from the speech signal, in addition to the actual semantic information content.

What makes speech bandwidth compression possible is the miraculously graceful way in which speech can be made to degrade as channel bandwidth is reduced. When going from the full 64 kb/s commercial telephony PCM to highly efficient vector-quantization vocoders needing but 1/100th of that rate, we first lose such frills as individuality, speaker characteristics, and environmental features, then accent and finer inflections, and only towards the end actual essential phonetic information. Which of us could design a signal of such robustness?

It is then even more humbling to realize that Nature gives us the transmitter for this ingenious signal practically for free, as a kind of "overlaid function," since no major organ in our body is uniquely dedicated to speech. The power source (lungs and diaphragm) is there for breathing; the main functions of the carrier generator (larynx) are swallowing, keeping the respiratory tract free of particles, and maintaining air pressure in the torso during bodily exertion. The major articulators (tongue and jaws) are primarily food-processing devices; even lips get rounded for suckling long before they do so to form vowels . . . It therefore behooves us to look at speech with respect and admiration, for as communications engineers we have much to learn from it.

The articles in this special issue highlight but a few of the active major areas where communications technology and speech science interact. (Two topics well covered by recent *IEEE Communications Magazine* and *IEEE Spectrum* articles, speech coding and automatic speech recognition [1-4], have been intentionally omitted. You may wish to read them with the foregoing thoughts in mind.)

The first three contributions address *bona fide* speech transmission issues. Mermelstein presents an overview of voice message systems, a digital technique barely five-years old but gaining ground rapidly. We all know the main drawback of telephones, compared to telegraphy or mail: information source and recipient must be available simultaneously, often leading to the game known as "telephone tag." Voice message systems exploit digital transmission, coding, and compression for treating segments of speech as informational packages, which can be stored and forwarded at different times. Their users can be not only busy executives, but also children, illiterates, or physically handicapped people, since speech—unlike writing—is a natural faculty shared by most of us.

Some of the potentially most annoying kinds of distortion are channel echoes. Fang's article shows us how recently developed integrated circuits can cancel them by adaptively modeling the echo path. Such devices introduce spurious noises of their own, but apparently the human ear can handle these better than competing speech sounds.

The paper by Gersho and Cuperman describes vector quantization, a promising upstart in the seemingly never-ending quest for good quality speech transmission by fewer bits at less cost. Inasmuch as the authors concentrate on waveform coding, their survey should also be of interest to readers involved in the compression of data other than speech, for example, images and geophysical or biomedical waves.

Recognizing the identity of the speaker, rather than the speech itself, is a new and intriguing branch of computer-mediated speech research. Adults, children, and even pets do it in a matter of seconds, but making a machine do it, even in slow motion, is still more an art than a science. Foil and Johnson explain some of the reasons why, and outline currently popular strategies.

The extensive article by O'Shaughnessy describes automatic speech synthesis, which can be thought of as the creation of a complex, redundant, interference-resisting spread-spectrum signal from low-rate basic data. This signal must *sound* like speech, even if it does not always *look* like speech; good designs hence take account of the properties of the receiver, the human ear. Synthetic speech is sometimes resisted because it is, for the time being, rather "machine-like." This resistance, like that against typewritten letters or machine-washed dishes, will probably diminish as quality improves and the freedom of movement associated with listening, as opposed to reading, becomes exploited. It is thus not difficult to foresee the time when synthesis combined with voice message switching will permit low-cost information retrieval by telephone: before going to bed you may ask a central database for a "brief" on, say, a new software package; the next morning you will be able to listen, in your car, to the voice on a cassette, transmitted at low overnight rates, reading off specifications, applications, and commentaries.

Our final paper, by Woodard and Cupples, treats some military applications of speech recognition: voice commands, word spotting, and low-rate transmission. The military establishment's interest in speech transmission has much benefited the field; many important studies, most notably the ARPA-SUR (Advanced Research Project Agency—Speech Understanding Project) were funded by defense agencies. This is especially true for secure voice communications, which has recently also become of commercial importance.

In less than a century of commercial telephone use, speech communications has radically changed our business and personal lives. Combined with computer technology, it will no doubt continue doing so at an even faster rate.

References

[1] G. R. Doddington and T. B. Schalk, "Speech recognition: turning theory to practice," *IEEE Spectrum*, vol. 18, no. 9, pp. 26-32, September 1981.

[2] R. E. Crochiere and J. L. Flanagan, "Current perspectives in digital speech," *IEEE Communications Magazine*, vol. 21, no. 1, pp. 32-40, January 1983.

[3] J. E. Holmgren, "Applying automatic speech recognition to telephone services," *IEEE Communications Magazine*, vol. 20, no. 6, pp. 31-34, November, 1982.

[4] G. Kaplan, R. Reddy, and Y. Kato, "Words into action," *IEEE Spectrum*, vol. 17, no. 6, pp. 22-29, June 1980.

Andrew Sekey began his studies at the Technical University of Budapest, and obtained his Ph.D. degree in Electrical Engineering (Psychoacoustics) from Imperial College, University of London. He first worked at the Post Office Research Station, England, and at Bell Laboratories, NJ; he then was a full-time teacher at City College, New York, NY, and at Tel-Aviv University.

Since 1976 he has been with the University of California, Santa Barbara. In 1982, he joined Speech Technology Laboratory, Santa Barbara, continuing part-time affiliation with UCSB. He has published over 30 papers and was editor of the recent book *Electroacoustic Analysis and Enhancement of Alaryngeal Speech*. After serving for three years as Book Reviews editor for *IEEE Communications Magazine*, he became a technical associate editor in 1981. ■

IEEE COMMUNICATIONS MAGAZINE is published monthly by the Institute of Electrical and Electronics Engineers, Inc. Headquarters address: IEEE, 345 East 47 Street, New York, NY 10017; telephone (212) 705-7018; Cable ITRIPLEE.
Responsibility for the contents rests upon authors of signed articles and not the IEEE or its members. Unless otherwise specified, the IEEE neither endorses nor sanctions any positions or actions espoused in IEEE Communications Magazine.

ANNUAL SUBSCRIPTION: Sent automatically and without additional cost to members of the Communications Society. Annual subscription rate for nonmembers: $36.00. Single copy: $3.00 for members and $6.00 for nonmembers.

CHANGE OF ADDRESS: Please use MOVING? card in front of magazine. Change of address must be received one month before issue is mailed.

EDITORIAL CORRESPONDENCE: Address to Editor, IEEE Communications Magazine, c/o Carol Lof at Headquarters address.

SUBSCRIPTIONS, orders, address changes—IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854.

ADVERTISING: Advertising is accepted at the discretion of the publisher. Address correspondence to IEEE Communications Magazine, 345 East 47 Street, New York, NY 10017.

Printed in U.S.A. Second-class postage paid at New York, NY and at additional mailing offices.