Speech Research at Bell Telephone Laboratories

M. R. SCHROEDER, J. L. FLANAGAN and P. B. DENES*

On the occasion of the first anniversary of the Research Institute of Logopedics and Phoniatrics of the University of Tokyo, best wishes and congratulations are extended from speech colleagues at the Bell Telephone Laboratories. The research objectives of the new institute are welcomed as timely and relevant. They complement well the aims of other groups in the speech community, especially those of co-workers at Bell Laboratories.

Just as at the new institute in Tokyo, much of the speech research at Bell Laboratories is concerned with obtaining more information about vocal tract dynamics and about linguistic factors that influence the speech process. It is strongly felt that only a substantially increased understanding of these articulatory and linguistic factors, combined with our increasing knowledge of speech acoustics will give a satisfactory explanation of the speech process.

Vocal tract area functions are being computed at Bell Laboratories in two ways: from measured speech spectra and from values of the vocal tract impedance measured at the lips. It is expected that by combining such information about area functions with cine X-ray and myographic data of vocal organ movements, it will be possible to derive rules for the dynamics of the tongue, lips and other articulators. It is thought that parameters specifying the movement of the principal vocal organs may well have a significantly simpler relation to the phonemes and words of language than the raw vocal tract area functions. An example of this approach is the synthesis of speech from a five-parameter model of the vocal tract. This latter work has enjoyed the participation of Professor O. Fujimura while he was at Bell Laboratories.

Projected work of the new institute in speech synthesis-by-rule also relates closely to interests at Bell Laboratories. Several projects on synthesizing speech by rule from an input of discrete phonemic symbols are in progress. They aim to investigate the effects of sentence stress, sentence structure as expressed by punctuation, and word and phrase juncture on the fundamental frequency, duration, amplitude, and formant structure of the corresponding speech sound waves. One of these speech synthesis projects uses a large digital computer in which the entire process of synthesis by rule and of specifying the resulting acoustic wave is simulated by program. Diagnostic listening tests on syllables and sentences are presently being performed to evaluate the quality and intelligibility of the synthetic results. Other projects use a smaller digital computer in an on-line manner. In this case, the computer program uses linguistic and articulatory rules

* Bell Telephone Laboratories, Inc., Murray Hill, New Jersey, U.S.A.

to calculate formant frequencies and other acoustic features; the computer then uses these values to control an electronic formant synthesizer. The control parameters sent to the formant generator are displayed on a computer controlled cathode ray tube and can be adjusted by the experimenter in real time. This synthesis program will eventually be changed to compute vocal tract area functions and to control a vocal tract analog.

The on-line computing facility is also being used for a wide variety of other psychophysical experiments. Some of the new graphical methods being developed for these other uses will also benefit speech research. For example, new methods for computer controlled stereoscopic displays of three or more dimensional functions may greatly enhance the value of the multidimensional scaling techniques that have already been used effectively in speech research.

In other studies, the physical correlates of speech quality are being considered. Synthetic speech produced from systematically-abstracted excitation information is providing insight into significant features of the glottal wave. Further work in speech recognition and speaker recognition has led to a high-precision method of pitch extraction which is capable of revealing temporal detail and characteristic forms in the pitch function.

The engineering fruits of basic speech work lie in the efficient encoding and transmission of voice signals. Toward utilizing this knowledge, computer simulations are being used to investigate a wide range of speech coding techniques. Included are channel vocoders, voice-excited vocoders, phase vocoders, formant vocoders and analytic-signal representations of speech information. These computer investigations embrace full digital encoding of speech signals.

Again, on this anniversary occasion, we join our colleagues in saluting the new Research Institute of Logopedics and Phoniatrics, University of Tokyo. We bid welcome to this new member of the research community, and look forward to mutually productive interchanges with this respected peer.