BRITISH ACOUSTICAL SOCIETY

"SPRING MEETING" at Chelsea College, London, S.W.3 on Wednesday 25th April / Friday 27th April, 1973

SPEECH AND HEARING: Session 'B': Speech Analysis and Transmission.

Paper No:

73SHB3

TRENDS IN DIGITAL CODING OF SPEECH WAVEFORMS

by

David J. Goodman
Bell Laboratories
Murray Hill, New Jersey 07974

INTRODUCTION. The choice of a specific means of representing speech by a binary sequence requires a compromise among cost and performance objectives. The most efficient schemes extract perceptually relevant features from the speech waveform and then encode these features in a binary format. Although more efficient than waveform encoders, the feature extraction approach to coding is far more costly and to date has found application only in special situations. This paper considers binary representations of the speech waveform itself.

New applications of digitally-coded speech, combined with new technological opportunities, have stimulated a search for new types of compromise among cost and performance objectives and today new methods of digital speech encoding proliferate. In this paper we refer to elementary coding methods in order to demonstrate how the goals of economy, manifested by inexpensive encoding and decoding equipment, and efficiency, evidenced by a low bit rate, lead to different types of codes. We then describe three ways in which the elementary coding methods are modified in order to create new combinations of cost and performance.

COST AND CODE CHARACTERISTICS. Time sampling and amplitude quantization are fundamental aspects of waveform encoding and in generalizing about cost, we observe that it is often economical to increase circuit operating speed if the increase admits a reduction in quantizer size. This observation follows from the fact that there is a wide range of operating speeds over which hardware costs are constant; while, on the other hand, cost is generally quite sensitive to analog tolerances.

PERFORMANCE AND CODE CHARACTERISTICS. The required accuracy of the binary representation of a speech waveform has an important influence on coding cost and efficiency. Usually, the application imposes a minimum band of signal frequencies to be represented, W, and a maximum mean square difference between the original waveform and the waveform reconstructed from the binary sequence. In pulse code modulation (PCM), with uniformly-spaced quantization levels, N_Q, the quantizing noise, varies with M, the number of quantizer bits, and f_s, the sampling frequency according to

$$N_Q = C_1(2^M-1)^{-2}\ (f_s/2W)^{-1}. \qquad (1)$$

The number of binary symbols per second is $f_B = Mf_s$ and Fig. 1 shows the relationship of N_Q to $f_B/2W$, where 2W

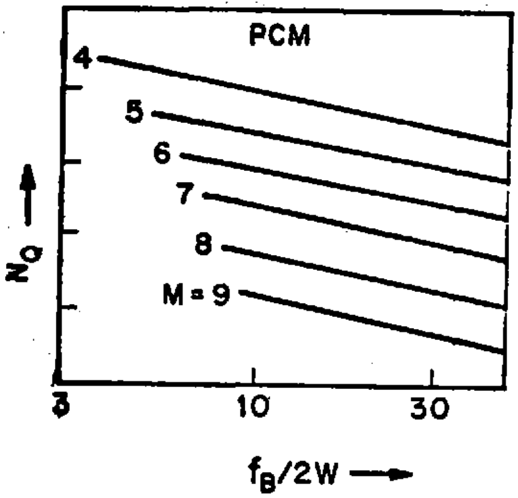

Fig. 1

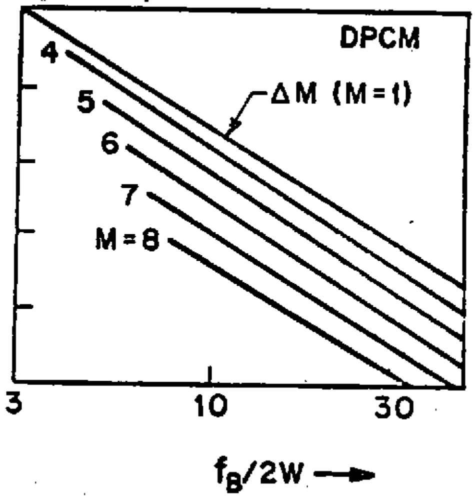

Fig. 2

is the minimum sampling rate for coding the W Hz signal band. It is clear that in meeting a given fidelity requirement, the most efficient PCM code (lowest f_B) has a low sampling rate and a multibit quantizer.

Differential pulse code modulation (DPCM) encodes the difference between a speech sample and an estimate of the sample obtained from previously generated binary symbols. Increasing the sampling rate reduces the dynamic range of the difference between speech sample and estimate and thereby admits a finer quantizer, even though the number of quantization levels is fixed. Compared with PCM, therefore, DPCM exhibits a more rapid decrease in N_Q as f_s increases.

$$N_Q = C_2(2^M-1)^{-2}\ (f_s/2W)^{-3};\ M \geq 2$$

$$N_Q = \frac{1}{2}\ C_2(f_s/2W)^{-3} \qquad ;\ M = 1. \qquad (2)$$

One-bit DPCM, delta modulation (ΔM), is especially economical, with a "quantizer" that responds only to the polarity of a voltage rather than to the precise level. Although ΔM is the most economical coding scheme, it is extremely inefficient in low-noise coding applications, a property apparent in Fig. 2.

DISPARITY BETWEEN ECONOMY AND EFFICIENCY. Although Eq. (1) and (2) and Figs. 1 and 2 describe a limited class of coding schemes - characterized by uniform, time-invariant quantizers - they exhibit properties that apply to a much more extensive set of code formats: distortion decreasing exponentially with quantizer size and algebraically with sampling rate. Thus it is efficient to maintain a low sampling rate and control distortion with quantizer complexity. On the other hand, we have pointed out that economy is achieved with a simple quantizer even though the sampling rate may be quite high.

PRACTICAL COMPROMISES. Existing and proposed coding methods exhibit a variety of compromises between economy and efficiency. We identify three approaches: a) time-sharing of expensive coding elements among several speech channels, b) adaptive ΔM and other time-varying schemes that quantize to a resolution of only a few bits per sample and c) code conversion, using digital circuitry to transform the inefficient code obtained from an economical speech-to-binary converter to a more efficient format, and vice versa.

Time-sharing. In digital telephone trunking, overall cost is determined by the efficiency with which the time-multiplexed transmission line is utilized. The speech signals to be transmitted on a digital line are encoded at the same place and the multibit quantizer that produces an efficient PCM code is shared among many signals. The standard code format on digital trunks is PCM with quantization levels unequally spaced for enhanced efficiency relative to the uniform PCM. The coder in the terminal originally produced by the Bell System (D1) is shared among 24 speech channels. The D2 channel bank[1], which has recently entered production, is shared among 96 channels.

Time-varying schemes. Although time-shared, efficient coders are appropriate to telephone trunk transmission, there are many newer applications that call for cheaper, though perhaps less efficient devices. In many of these applications, opportunities for time sharing are limited and the cost of coding is a significant fraction of total system cost. The emphasis on economy, rather than efficiency, in such applications has led to renewed interest in ΔM.

Considerable effort has been devoted to the discovery of more efficient formats which sacrifice some of the economy of the simplest form of ΔM but retain the cost advantages inherent in two-level quantization. The variable-step-size approach to ΔM has brought the greatest success to this effort, resulting in efficiencies comparable to that of non-uniform PCM. At present two time-varying ΔM codes are in use over time-multiplexed transmission lines. In the subscriber loop multiplex system developed at Bell Laboratories,[2] the step size is allowed to change by a factor of 2 at 125 μsec intervals. On the other hand, the step size of the modulator in the Δ P.A. 60 system,[3] developed by T.R.T. (in France), varies with a time constant of a few msec.

Code conversion. Another type of compromise between the goals of economy and efficiency is motivated by the observation that standardized digital circuit elements are becoming increasingly economical as the scale of circuit integration increases. It follows that the combination of a simple, inefficient modulator and a digital code converter is, in some cases, less expensive than an efficient encoder. Using simple ΔM for speech-to-binary and binary-to-speech conversion, this approach has been pursued with time varying ΔM[4] and with PCM as the efficient code format. A ΔM/PCM encoder is currently employed in an experimental system[5] for converting among baseband, frequency multiplex and time multiplex representations of a set of 12 speech signals.

A ΔM/PCM encoder uses a finite-impulse-response digital filter to suppress high-frequency components of the ΔM quantizing noise. A recent theoretical study of this technique[6] demonstrates that uniform filters, with all coefficients unity, are particularly appropriate for this purpose. With uniform filtering, code conversion is realized by an up/down counter and a resettable accumulator that obtains the sum of N successive counter levels. Figure 3 shows that many combinations of N, the

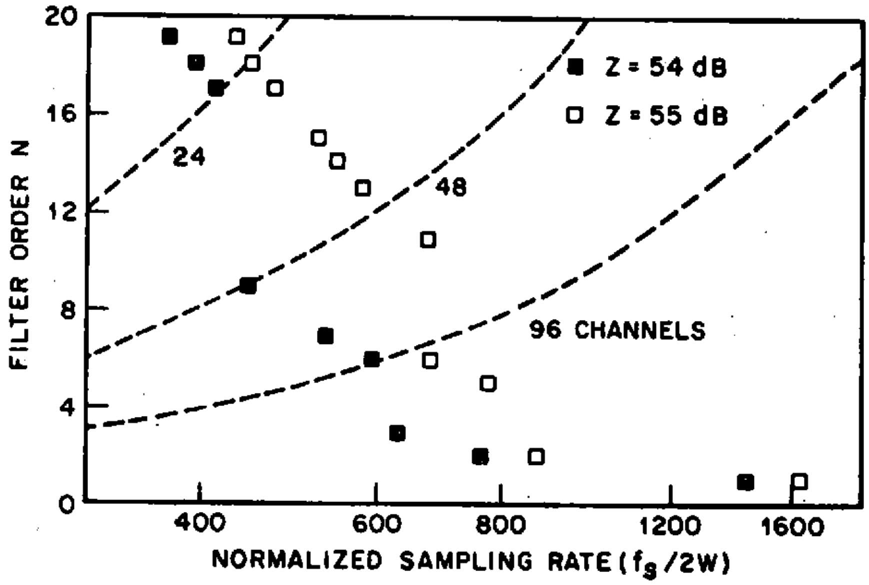

Fig. 3

number of terms in the sum, and f_s, the ΔM sampling rate, lead to the same level of quantizing distortion. Therefore, a designer has considerable flexibility in satisfying economic and technological constraints. If the code conversion for many speech signals is to be performed at a single point, the accumulator may be shared among several channels. For example, in Fig. 3, the designs that admit sharing by 24, 48, and 96 speech channels are to the right of the respectively labeled broken lines.

REFERENCES

1. Dammann, C. L., et al., B.S.T.J., 51, 1675-1700, October 1972.
2. McNair, I. M., Jr., IEEE Trans. Com. Tech., COM-19, 523-527, August 1971.
3. Gaquere, C. G. P., Conf. Rec., 1971 IEEE Int. Conf. Comm., 22.16-22.22, June 1971.
4. Goodman, D. J. and Flanagan, J. L., Conf. Rec., 1971 IEEE Int. Conf. Comm., 1.33-1.36, June 1971.
5. Freeny, S. L., et al., IEEE Trans. Com. Tech., COM-19, 1050-1059, December 1971.
6. Goodman, D. J. and Greenstein, L. J., B.S.T.J., 52, 183-204, February 1973.