IX. INFORMATION THEORY

Prof. R. M. Fano
Prof. P. Elias
Dr. F. A. Muller
A. Adolph
J. Capon
A. Feinstein
R. M. Lerner
S. Muroga
J. C. Stoddard
W. A. Youngblood

A. A SIGNAL ANALYZER

The signal analyzer, first discussed in the Quarterly Progress Report, October 15, 1953, p. 30, was completed. It functions in accordance with expectations. However, as a result of its essentially linear power scale the analyzer has serious shortcomings for use as a tool for speech analysis.

A large number of stages and a high accuracy would be required to indicate the weaker frequency regions of speech sounds.

F. A. Muller

B. CONTROLLABLE FILTER

The controllable filter, described in the Quarterly Progress Report, October 15, 1953, has been generalized.

A filter is sought whose amplitude transfer characteristic can be controlled independently at frequencies $\omega_o = 0, \omega_1, \omega_2, \ldots, \omega_k, \ldots, \omega_{n-1}, \omega_n = \infty$. This property is obtained with the filter of Fig. IX-1, when the values of the elements obey the relations $L_k C_k \omega_k^2 = 1$. It is assumed that the losses in the coils and the resistances of the potentiometers can be neglected.

Furthermore, the transfer characteristic should be well behaved at frequencies between the controlled points. For instance, the amplitude response may be required to be entirely flat when all controlled values are equal. This requirement can be met by choosing the values of the elements as follows. Let

$$L_k = \frac{R}{2} \left(\frac{d\phi}{d\omega}\right)_{\omega=\omega_k} \qquad (k = 1, 2, \ldots, (n-1))$$

$$L_o = R \left(\frac{d\phi}{d\omega}\right)_{\omega=0}$$

$$C_k = \frac{2}{R\left(\frac{d\phi}{d\, 1/\omega}\right)_{\omega=\omega_k}} \qquad (k = 1, 2, \ldots, (n-1))$$

$$C_n = \frac{1}{R\left(\frac{d\phi}{d\, 1/\omega}\right)_{\omega=\infty}}$$

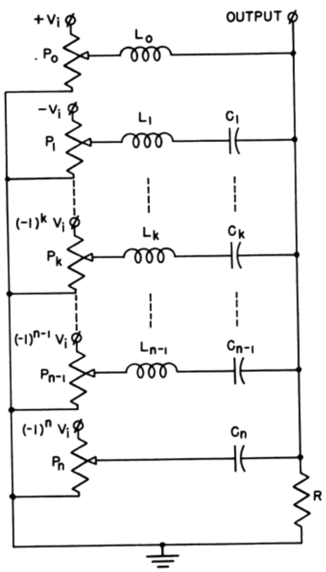

Fig. IX-1
Controllable filter.

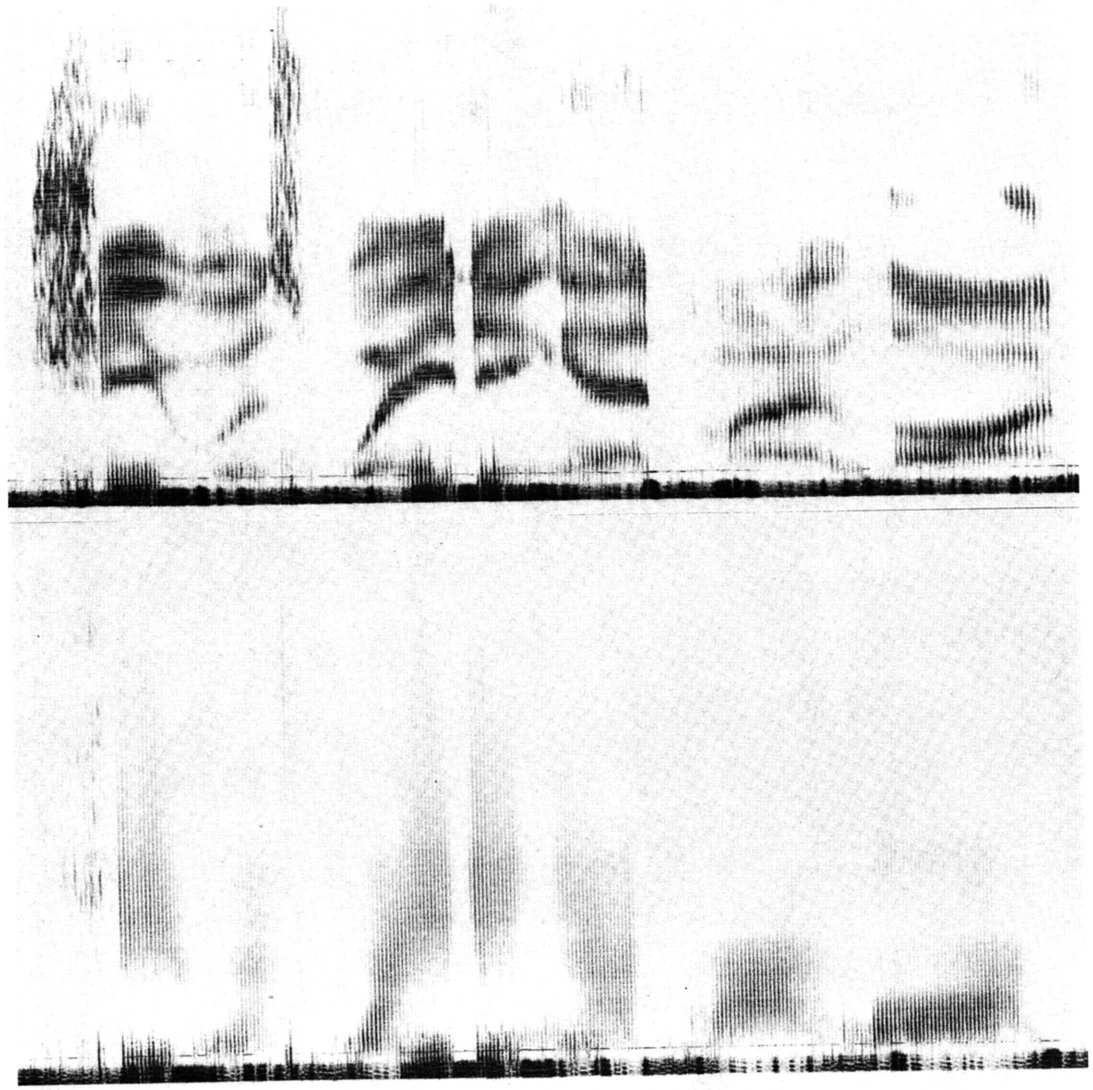

Fig. IX-2
Spectrogram of the sentence, "She was waiting at my lawn;" (top) natural speech, (bottom) output of the vocoder.

with

$$\phi = 2 \tan^{-1} \lambda\omega \frac{\left[1 - (\omega/\omega_2)^2\right]\left[1 - (\omega/\omega_4)^2\right]\left[1 - (\omega/\omega_6)^2\right]\dots}{\left[1 - (\omega/\omega_1)^2\right]\left[1 - (\omega/\omega_3)^2\right]\left[1 - (\omega/\omega_5)^2\right]\dots}$$

The phase of the network is given by ϕ. The parameter λ governs the relative importance of odd and even controlled points.

The filter described for the special case in the Quarterly Progress Report, October 15, 1953, has this property: when the amplitude response is considered as a function of the phase response, it does not contain more than n terms in the Fourier analysis. This property does not hold for the general case. When the controlled points are distributed very irregularly, some undesirable crossover effects may occur.

F. A. Muller

C. FIVE-CHANNEL VOCODER

A five-channel vocoder was built provisionally. This instrument differs from Dudley's vocoder (1) in the way the short-time power spectrum of speech is represented for transmission purposes.

A basic problem in the operation of a vocoder is the simplification of the shape of the spectrum, without losing many "information-bearing elements," in such a way that the simplified spectrum can be described by a small number of parameters.

Dudley's solution for this problem is to quantize the frequency scale of the spectrum.

In the system presented here, the power and frequency scales are changed by a nonlinear transformation in such a way that they resemble the physiological "importance" scales. The simplification is then obtained by deleting the high components of the Fourier analysis of the transformed spectrum and retaining only the lowest components; in this special case, five of them.

It is felt that this procedure might result in a better conservation of transitional effects in speech than the method of frequency-scale quantization.

The simplified spectrum can be transmitted by transmitting its Fourier coefficients, or by transmitting the same number of sample points. This latter method is used because it has some practical advantages. The receiver calculates more points of the spectrum and then synthesizes a signal having this spectrum with the help of the controllable filter described in the Quarterly Progress Report of October 15, 1953, p. 32, and in section B above.

It is a typical property of this system that both the analyzer and the synthesizer describe the spectrum in more points than are transmitted. The spectrum is, in this special case, measured logarithmically in nine points, at 140, 430, 750, 1100, 1500, 2100, 2900, 4300, and 7500 cps. The transmitted sample points are located at 140, 750, 1500, 2900, and 7500 cps. The receiver interpolates the points at 430, 1100, 2100, and 4300 cps before synthesizing the spectrum with the help of exponential devices. Figure IX-2 shows the influence of the vocoder on the spectrogram of a sample of speech.

The word articulation, measured with PB (phonetically balanced) lists, is around 60 percent after the listener is used to the type of sound for a few minutes.

Preliminary tests indicate that most of the errors are caused by the stop sounds, especially by the unvoiced stops, which are often confused. The vowels, fricatives, and diphthongs caused fewer errors. This correlation of the duration of a phoneme with the quality of the transmission seems to indicate that more attention has to be paid to the time filtering of the transmitted sample points. An improvement of the naturalness of the synthesized speech has to be expected from an increase in the number of channels. A new model of this type of vocoder has been designed. It uses 25 points of the frequency scale in the analyzer and synthesizer, and can be operated with either 5, 7, or 9 channels.

A. R. Adolph, F. A. Muller

References

1. H. Dudley, The vocoder, Bell Labs. Record 18 (1939).

D. ERROR FREE CODING

A simple constructive procedure has been found for transmitting information at a nonzero rate over a discrete channel in the presence of noise, with a probability of error that decreases with the duration of transmission, approaching zero in the limit. The approach used here was suggested by a comment of Dr. Victor Yngve, of this Laboratory, on the fact that redundancy in language was added at many different levels; a point that he discusses in section XV.

The procedure relies on using single error correction codes for sequences of symbols that are selected from larger and larger alphabets. Consider a channel that transmits binary symbols with the two values 0 and 1, with a symmetrical error probability p_o that a transmitted symbol will be received with the incorrect value. Hamming (1) has given a coding procedure that will correct single errors in blocks of N_1 binary symbols by making use of $C_1 = [\log_2 N_1 + 1]$ of the symbols as check digits, leaving $N_1 - C_1$ to be freely selected as information-bearing digits. (The brackets denote the largest integer which is less than or equal to the number they enclose.) If p_o is small, N_1 may be chosen small enough so that the probability p_1 of two or more errors in the block of N_1 binary symbols is less than p_o. That is, if $N_1 p_o$ is less than 1, we have

$$\left.\begin{aligned} p_1 &= \text{Prob}\,\{\text{two or more errors in } N_1 \text{ symbols}\} \\ &= 1 - (1 - p_o)^{N_1} - N_1 p_o (1 - p_o)^{N_1 - 1} \\ &= 1 - \left(1 - N_1 p_o + \frac{N_1(N_1 - 1)}{2} p_o^2 - \dots\right) - N_1 p_o (1 - (N_1 - 1) p_o + \dots) \\ &< \frac{N_1(N_1 - 1)}{2} p_o \\ &< \frac{N_1^2}{2} p_o \end{aligned}\right\} \tag{1}$$

Choosing any $\alpha_1 < 1$, from Eq. 1 we have

$$p_1 < \alpha_1 p_o \qquad \text{if} \qquad N_1 \leq \left(\frac{2\alpha_1}{p_o}\right)^{1/2} \tag{2}$$

The message, after this first step, is a sequence of symbols selected, not from a binary alphabet but from an M_1-ary alphabet, where

$$M_1 = 2^{(N_1 - C_1)}$$

The probability that any one of these M_1-ary digits will be incorrectly received is $p_1 < \alpha_1 p_o$. It is now possible to construct a single error correcting code for a sequence of N_2 such M_1-ary digits, with a probability of error (that is, of two or more M_1-ary digits misreceived out of a block of N_2 such digits) of $p_2 < \alpha_2 p_1$. Such a procedure will require a number C_2 of M_1-ary digits to be used as checks, leaving $N_2 - C_2$ digits as information carriers.

The blocks of N_2 M_1-ary symbols may now be considered as symbols to the base M_2, where

$$M_2 = M_1^{(N_2 - C_2)} = 2^{(N_1 - C_1)(N_2 - C_2)}$$

and sequences of N_3 M_2-ary digits may be encoded in a single error correcting code, and so forth. At the k-th stage of this process we have

$$\left.\begin{aligned} p_k &= \alpha_k p_{k-1} = \left(\prod_{j=1}^{k} \alpha_j\right) p_o \\ N_k &= \left(\frac{2\alpha_k}{p_{k-1}}\right)^{1/2} = \left[\frac{2\alpha_k}{\left(\prod_{j=1}^{k-1} \alpha_j\right) p_o}\right]^{1/2} = \alpha_k\left(\frac{2}{p_k}\right)^{1/2} \\ M_k &= M_{k-1}^{(N_k - C_k)} = 2^{\prod_{j=1}^{k}(N_j - C_j)} \end{aligned}\right\} \tag{3}$$

The fraction of channel capacity used for information digits up to and including the k-th level is

$$F_k = \prod_{j=1}^{k} (1 - f_j), \qquad f_j = \frac{C_j}{N_j} \tag{4}$$

and F_∞, the limit of F_k as $k \to \infty$, is the rate of information transmission.

In order to select the values α_j which maximize F_k it is necessary to know C_j; that is, how many check digits are required for single error correction of N_j digits to the base M_{j-1}. Hamming (1) has answered this question for M = 2, and Golay (2) has extended this result to other bases which are prime numbers. Here, M is a power of 2, and thus, certainly, not a prime. An exact solution to this problem is difficult. However, it is easy to find M-ary coding procedures which are good enough to make F_∞ nonzero. This can be done with an inefficient but simple kind of single error correction which gives $C_j = [\log_2 N_j + 1]$, and with α_j fixed, $\alpha_j = 1/4$ for example. This gives, from Eqs. 3 and 4

$$\left.\begin{aligned} p_k &= \frac{p_o}{2^{2k}} \\ N_k &\doteq 2^{k-1} \cdot N_1 \doteq 2^{k-1}\left(\frac{1}{2p_o}\right)^{1/2} \\ C_k &= [\log_2 N_k + 1] = [k + \log N_1] \\ F_k &= \prod_{j=1}^{\infty}\left(1 - \frac{[j + \log N_1]}{2^{j-1} \cdot N_1}\right) \end{aligned}\right\} \tag{5}$$

The infinite product in Eq. 5 converges for $N_1 \geq 3$, corresponding to $p_o \leq 1/12$, to a positive value, as can be seen by considering the corresponding sum. Actually it is possible to do much better than this; less simple single error correction codes, in which $C_j \sim \log_{M_{j-1}} N_j$, rather than $C_j \sim \log_2 N_j$, can be constructed. It is also possible to iterate, in the same fashion, other error correcting and detecting codes. Work along both lines is proceding, together with a study of the optimization of the choice of values for a_k.

P. Elias

References

1. R. W. Hamming, Error correcting and error detecting codes, Bell System Tech. J. 29, 147-160 (1950).
2. Marcel J. E. Golay, Notes on digital coding, Proc. I. R. E. 37, 657 (1949).