THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 10, December, 1971
Printed in U.S.A.

The Preference of Slope Overload to Granularity in the Delta Modulation of Speech

By N. S. JAYANT and A. E. ROSENBERG

(Manuscript received June 18, 1971)

A preference study was made to assess the relative annoyance values of slope-overload distortion and granular noise in delta-modulated speech. A recently described adaptive delta modulator was simulated at frequencies of 20 and 40 kHz, and controlled amounts of the two types of degradation were introduced into samples of a 2-second utterance. Rankings were obtained for these samples on the basis of preference judgments of nine listeners, each of whom assessed the samples, pairwise, in a tournament-type strategy. Results indicate that the speech sample exhibiting the minimum degradation on an objective, overall-noise-power basis is not subjectively the most preferred sample. Furthermore, the subjectively optimum delta modulator exhibits greater overload and lesser granularity than the objectively optimum device.

I. INTRODUCTION

The principle of delta modulation[1] has been widely described in the literature. Briefly, delta modulation is a digital encoding strategy which uses a simple feedback mechanism to produce a "staircase" approximation to an input signal. A block diagram of the simplest form of delta modulation appears in Fig. 1. The input sequence $\{X_r\}$ is usually band-limited and suitably oversampled. The "staircase" sequence Y_r is generated according to the equations

$$C_r = \text{sgn}\,(X_r - Y_{r-1}) \tag{1}$$

$$Y_r - Y_{r-1} = m_r = \Delta_r \cdot C_r\,. \tag{2}$$

The step-size Δ_r is assumed to be a constant in conventional (linear) delta modulation. "Adaptive" delta modulation, on the other hand, allows for modifications of Δ_r in accordance with the changing slope

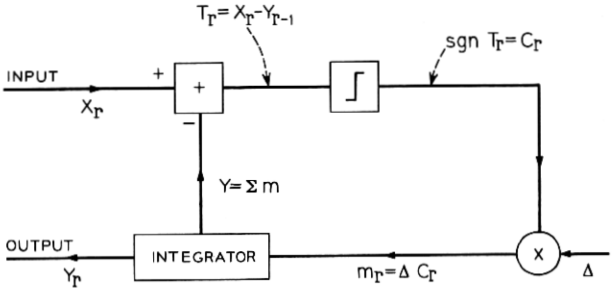

Fig. 1—Schematic diagram of a linear delta modulator.

characteristics of the input signal. Such adaptation results in better encoding, and several types of adaptive delta modulation have been described in the literature.[2,3,4]

Figure 2 illustrates the mechanism of an adaptive delta modulator and demonstrates how suitable increases and decreases of step size facilitate better encoding during steep and flat regions of the input signal waveform. Such adaptations can be effected by observations on a "recent" segment of the binary sequence $\{C_r\}$; this is illustrated by equation (5) in the sequel.

Figure 2 also brings out the distinction between two types of encoding error in delta modulation, viz., "granular noise" and "slope-overload"

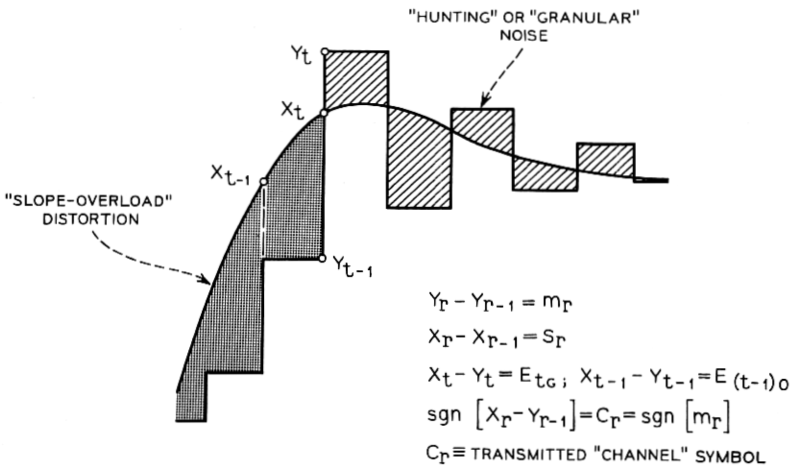

Fig. 2—Illustration of adaptive delta modulation.

distortion. A given error sample

$$E_r = X_r - Y_r \tag{3}$$

can be defined to fall into the granular or slope-overload category, depending on whether the corresponding step m_r crosses the input waveform or not. Thus, in Fig. 2, there is a 'granular' error E_{tG} at the sampling instant t, and an 'overload' error $E_{(t-1)O}$ at the sampling instant $(t - 1)$. As a matter of definition, we will note that $E_{tO} = E_{(t-1)G} = 0$.

The signal output $\{Z_r\}$ of the delta modulator is actually obtained by filtering the staircase sequence $\{Y_r\}$ to the input signal band. Let $\{X_r^F\}$ be the result of passing $\{X_r\}$ through the same lowpass filter. A perceptually relevant measure of signal degradation is accordingly defined by the encoding error

$$e_r = X_r^F - Z_r\,. \tag{4}$$

As with the quantity E_r in (3), one can distinguish samples of granularity and slope overload, e_{rG} and e_{rO}, in the error sequence $\{e_r\}$. Referring to Fig. 2 once more it can be seen that a physical distinction between the two types of error is suggested. Granularity can be described as a "signal-uncorrelated" random noise-type of phenomenon. It is characterized by alternation of signs and tends to be independent of signal amplitude. Slope overload, on the other hand, can be described as a "signal-correlated" distortion, since its sign and magnitude are related to the slope of the signal. This physical difference between slope overload and granularity suggests a corresponding perceptual distinction and raises the question of the relative annoyance values of the two forms of signal degradation in delta modulation. The present paper describes a study of the above question as referred to the delta modulation of a speech signal.

Earlier work in this subject is in the form of a perceptual experiment[5] in which H. Levitt, et al., characterized the perceptibility of slope-overload distortion as such. As mentioned earlier, our paper will seek to answer the complementary question of the relative perceptibilities of slope overload and granularity when they occur simultaneously in delta-modulated speech, as they usually do.

The approach we used was to vary the relative amount of slope overload and granularity introduced into samples of a test utterance, and to evaluate these samples on the basis of both objective and perceptual criteria; and then to interpret these evaluations with specific reference to the overload-granularity dichotomy.

Section II summarizes the salient features of a computer-simulated adaptive delta modulator that was utilized in the present study. This adaptive encoder has been recently described and shown to provide toll-quality speech reproduction at bit rates of practical importance.[4]

Section III defines the objective measures of speech quality used in our study, while Section IV defines a subjective measure of preference and describes an underlying perceptual experiment.

II. DESCRIPTION OF THE DELTA MODULATOR

Figure 3 is a schematic block diagram of the adaptive delta modulator utilized in the present study. This encoder is defined by the basic equations (1) and (2), and by the adaptation rule

$$\left.\begin{aligned}\Delta_r &= P\cdot\Delta_{r-1} \quad \text{if} \quad C_r = C_{r-1}\\ &= \frac{1}{P}\cdot\Delta_{r-1} \quad \text{if} \quad C_r \neq C_{r-1}\end{aligned}\right\} ; \qquad P \geqq 1. \tag{5}$$

Notice that a conventional (linear) delta modulator corresponds to the special case of $P = 1$. In our study the value of P was a variable parameter; different (delta-modulated) speech samples corresponded to different suitably spaced values of P, and thereby to different mixtures of slope-overload and granularity.

The original speech sample X was a 2-second male utterance of "Have you seen Bill?" that had been band-limited to 3.3 kHz. The delta modulation was performed at sampling rates of 20 and 40 kHz. The latter frequency provides speech reproduction that approaches telephone

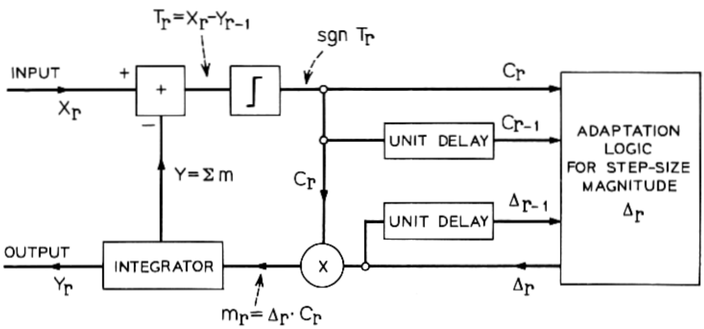

Fig. 3—Schematic diagram of an adaptive delta modulator.

quality.[4] The lower sampling rate was included to provide a better demonstration of the annoyance properties of delta-modulated speech.

III. OBJECTIVE MEASURES OF SPEECH QUALITY

We recall the encoding error e_r, (4), and define the following measures of delta-modulator performance. (Summations are over the entire length of the speech utterance, and a nonzero granularity error at $r = t$ implies zero overload error, and vice versa.)

(*i*) The overload-noise energy in Z:

$$N_O = \sum e_{rO}^2 \tag{6}$$

(*ii*) The granular-noise energy in Z:

$$N_G = \sum e_{rG}^2 \tag{7}$$

(*iii*) The signal-to-noise ratio:

$$\text{SNR} = \frac{\sum X_r^2}{\sum e_{rO}^2 + \sum e_{rG}^2} \tag{8}$$

(*iv*) The signal-to-granular (overload)-noise ratio:

$$\text{SNR}_{G(O)} = \frac{\sum X_r^2}{\sum e_{rG(O)}^2} \tag{9}$$

IV. A SUBJECTIVE MEASURE OF SPEECH QUALITY

The perceptual evaluations of this paper are based on the pooled* judgments of nine listeners each of whom assessed speech stimuli† in six runs of a perceptual experiment. Each of these 54 experiments was a double-elimination tournament‡ (with a different, random, starting line-up). Matches in each tournament were between contending stimuli, playing two at a time. The result of each match was in the form of a binary preference judgment by the listener, while the result of a tournament was a set of scores awarded to each of the contesting speech stimuli on the basis of its record in the tournament. The actual scoring rule§ was one which, together with the double-elimination

* Intralistener variations were found to be less significant than the intrastimulus differences.

† The number of contending speech stimuli was also nine, at each sampling rate.

‡ The tournament ended when every losing contestant had lost twice.

§ In the course of each tournament a contestant accumulated a score as follows. No score was earned for a match that was lost; while, after every match that was won, the contestant's score was the sum of the accumulated scores, before the match, of the contestant and of the loser, plus one.

strategy, provided a useful alternative—as concluded from a separate simulation—to the more comprehensive testing procedure where every contending stimulus would be pitted against every other.

It was recognized, however, that both the scoring rule and the double-elimination strategy were empirical procedures. This was more so because they were applied to what was apparently a probabilistic environment: the binary preference-response of a listener to a given pair of contending stimuli can well be random, especially when the stimuli are not obviously different. It was, therefore, decided not to emphasize the actual scores obtained in the perceptual test. They were only used, instead, to extract a crude ranking information that would be less sensitive to the testing and scoring procedures.

Consequently, the following subjective preference value Q was assigned to each of M contesting speech stimuli:

$$Q = \frac{M - R}{M - 1}; \qquad R = 1, 2, \cdots, M \tag{10}$$

where R is the rank assigned to a stimulus on the basis of its accumulated score in the 54 runs of the perceptual test.

V. SUMMARY OF RESULTS

Figure 4 displays normalized values of the objective measures of quality SNR, SNR_G, and SNR_O, as well as the subjective preference function Q, as functions of the adaptation parameter P. The following observations emerge:

(*i*) The speech sample representing the minimum overall-noise-energy is not subjectively the most preferred sample. In fact, at both 20 and 40 Hz, the objective and subjective optima can be characterized by

$$P_{\text{OPT}}^{\text{SUBJ}} = 1.2 \tag{11}$$

$$P_{\text{OPT}}^{\text{OBJ}} = 1.5. \tag{12}$$

(*ii*) The approximate coincidence of the SNR and SNR_O curves indicates, by virtue of equations (6) through (9), that

$$N_O \gg N_G \tag{13}$$

for all considered values of P.

(*iii*) The relative disposition of the SNR_O, SNR_G, and Q curves—and of their maxima—demonstrates that, in spite of the preponderance (13) of overload in the overall-noise-energy, the granularity in a speech

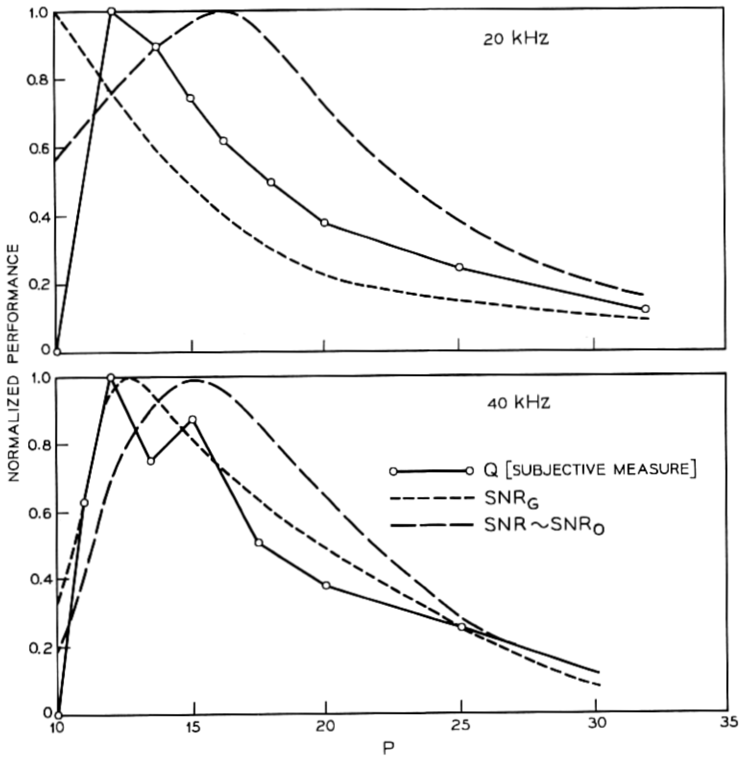

Fig. 4—Evaluations of delta-modulator performance.

stimulus has a strong influence on its subjective preference value.

Note that there is a double peak in the subjective preference curve, Q, for the 40-kHz case. This curve unambiguously ranks each of the experimental stimuli according to equation (10). However, the actual scores underlying this ranking show only a small difference between the stimulus with the secondary peak and the one immediately preceding it. What is probably indicated is a general broadening of the peak of the preference function for values of P between 1.2 and 1.5.

Table I lists, for the optimal characterizations (11) and (12), values of $N_{O(G)}$ (given as fractions of signal energy), SNR, and Q. Notice that the subjectively optimum delta modulator displays lesser granularity (N_G) and greater overload (N_O) than the objectively optimum modulator. It is again obvious that in perception, overload and granularity are not weighed in proportion to the respective noise energies N_O and N_G ;

TABLE I—CHARACTERISTIC OF OPTIMAL ADAPTIVE DELTA MODULATION (N_O and N_G are entered as fractions of signal energy)

Sampling Frequency	P	N_O	N_G	SNR	Q
20 kHz	$P^{SUBJ}_{OPT} = 1.2$	0.0216	0.0003	43	1
	$P^{OBJ}_{OPT} = 1.5$	0.0158	0.0004	58	0.81
40 kHz	$P^{SUBJ}_{OPT} = 1.2$	0.0022	0.00003	450	1
	$P^{OBJ}_{OPT} = 1.5$	0.0016	0.00004	640	0.91

in fact, the perceptual preference of a speech sample seems to be determined very strongly by the extent of granularity in it, although the latter represents a very small fraction of the total noise energy.

Finally, Table I indicates that distinctions between objective and subjective assessments of speech quality appear to be less significant at the higher sampling rate of 40 kHz; thus, for example, the objectively best delta modulator has a greater value of subjective perference Q at 40 kHz than at 20 kHz.

VI. CONCLUSION

We have shown that in delta modulation, a speech sample exhibiting the minimum degradation on an objective, overall-noise-energy basis is not equivalent, in general, to the perceptually most preferred sample. We have also indicated that this distinction may be less significant in higher quality delta modulation than in a low-bit-rate encoder.

The subjectively optimum delta-encoder displays a greater overload N_O and lesser granularity N_G than the objectively best encoder. This feature, together with the fact that $N_O \gg N_G$ in either case, suggests the strong influence of granular noise on the perceptual assessment of a speech sample; equivalently, a lesser "annoyance value" is to be associated with slope-overload distortion.* A possible explanation of

* Companding in PCM exploits a similar but *not identical* subjective phenomenon, viz., the greater tolerance to encoding errors in regions of high input amplitude. (Notice however, that in delta modulation, slope overload is not confined to high-amplitude regions, nor is granularity associated only with low input amplitude.)

this observation would be the fact that granularity is explicitly perceivable by a listener as an "additive background noise," while slope-overload distortion exists only in relation to an original signal which is not known to the listener.

Finally, our observation that slope overload is "less annoying" than granularity is to be invoked with caution. Broadly speaking, we believe that our conclusion would apply very well to speech that achieves or approaches telephone quality. In extremely low-quality delta modulation (such as may be used in special applications), on the other hand, the intelligibility of speech will be a critical criterion; and in such a situation, depending on other factors like ambient noise at a transmitter, slope overload may very well become a more important perceptual attribute.

REFERENCES

1. de Jager, F., "Delta Modulation, A Method of PCM Transmission Using 1-Unit Code," Phillips Research Reports, *7*, No. 6 (December 1952), pp. 442–446.
2. Winkler, M. K., "High Information Delta Modulation," IEEE International Convention Record, pt. 8, 1963, pp. 260–265.
3. Greefkes, J. A., and de Jager, F., "Continuous Delta Modulation," Phillips Research Reports, *23*, No. 2 (April 1968), pp. 233–246.
4. Jayant, N. S., "Adaptive Delta Modulation with a One-Bit Memory," B.S.T.J., *49*, No. 3 (March 1970), pp. 321–342.
5. Levitt, H., McGonegal, C. A., and Cherry, L. L., "Perception of Slope-Overload Distortion in Delta-Modulated Speech Signals," IEEE Transactions on Audio and Electro-Acoustics, *AU-18*, No. 3 (September 1970), pp. 240–247.