THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 7, September, 1972
Printed in U.S.A.

Perceptual Evaluation of the Effects of Dither on Low Bit Rate PCM Systems

By L. R. RABINER and J. A. JOHNSON

(Manuscript received April 10, 1972)

It has previously been shown that by adding a pseudo-random "dither" noise to a signal to be quantized, and by subtracting an identical noise sequence from the quantizer output, it is possible to break up undesirable signal-dependent patterns in the quantization error sequence without increasing the variance of the error. The effect of the dither noise becomes significant when the number of bits per sample is less than about six. An experimental evaluation of the perceptual effects of dither on speech has shown:

(*i*) *strong preferences for dithered speech over straight PCM encoding at identical bit rates,*
(*ii*) *for low bit rates (2–4 bits/sample), a preference for dithered speech over PCM encoded speech even when the PCM speech had one more bit per sample than the dithered speech,*
(*iii*) *an increase in word intelligibility for dithered speech over PCM speech when 4 to 6 bits/sample were used,*
(*iv*) *a decrease in word intelligibility for dithered speech over PCM speech when 2 to 3 bits/sample were used.*

I. INTRODUCTION

When a signal, such as a speech waveform, is quantized, the quantization error waveform is usually correlated with the original signal. This correlation is virtually inperceptible when the quantization is quite fine–i.e., a large number of bits/sample. For crude quantizations, however, the correlation becomes quite large and the quantization error is easily perceived. As a result, it can become quite disturbing to listen to speech quantized to a low number of bits/sample for an extended period of time. In such cases, techniques that decorrelate the quantization error from the signal are attractive, even if they do not increase the signal-to-noise ratio of the system. Dithering is such a technique in which a pseudo-random "dither" noise is added to the speech before quantizing, and then the identical noise is subtracted producing a quantization error which is uncorrelated with the original speech waveform.[1] Figure 1 shows a comparison between straight PCM and a system in which dithering is used. In an earlier work, Jayant and Rabiner[2] discussed several theoretical issues involved with dithering and demonstrated its utility for the quantization of speech signals. In this paper, we present experimental results on the perceptual effects of dither on both the preference and intelligibility of PCM encoded speech.

II. PREFERENCE EVALUATION TEST

The purpose of this experiment was to determine the perceptibility of the decrease in correlation between the quantization error and the original speech, as a function of the number of signal bits.

The stimuli used in the preference test were a set of ten sentences chosen from a list of "everyday speech" sentences[3] compiled at the Central Institute for the Deaf. The sentences used are shown in Table I. These ten sentences were spoken by a General American speaker, digitized at a 10 kHz rate with 16 bits/sample, and stored on the disc of the DDP-516 computer.

In order to limit the number of stimuli to be used in the paired-comparisons preference test, the number of bits/sample was restricted to the range of 2 to 6 bits. Therefore, there were ten distinct stimuli in the test, i.e., (five possible values for the number of bits) × (two types of quantization–dither or straight PCM). For notational convenience, the stimuli were coded using a two-digit code. The first digit refers to the number of bits/sample (i.e., 2–6) and the second digit specifies the type of quantization. A 0 in the second digit means straight PCM encoding, whereas a 1 in the second digit means dithered speech. Thus stimulus 31 has 3 bits/sample and uses the dither noise, whereas stimulus 50 has 5 bits/sample and does not use dither noise.

Since there were ten distinct conditions to be evaluated, a complete

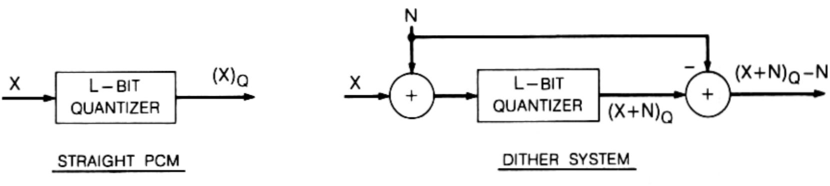

Fig. 1—Block diagrams of a straight PCM system and a dither system.

TABLE I—SENTENCES USED IN PREFERENCE TEST

1. Walking's my favorite exercise.
2. Here's a nice quiet place to rest.
3. Our janitor sweeps the floor every night.
4. It would be much easier if everyone would help.
5. Good morning.
6. Open your windows before you go to bed.
7. Do you think she should stay out so late.
8. How do you feel about changing the time when we begin work.
9. Here we go!
10. Move out of the way.

paired-comparison preference test involved 100 pairs. These 100 pairs were randomly generated by a DDP-516 program which randomly accessed each of the ten stimulus sentences ten times in the course of the experiment. Each of the 100 stimulus pairs was recorded on magnetic tape for offline running of the experiment.

Ten subjects participated in the experiment. Each subject was given the following instructions:

> "In this test you will be listening to pairs of sentences. Each of the two sentences (first is called A, second B) was processed by some type of speech transmission system. After you hear both sentences, there is a five-second interval in which you are to write down the sentence, A or B, you prefer, i.e., the type of transmission system you would prefer listening to for an extended period of time. You *must* choose either A or B–even if you have no preference."

The preference test required two 15-minute listening sessions per subject and was run on two separate days.

III. RESULTS OF PREFERENCE TEST

For each of the ten subjects, a matrix of preferences was determined in which a 1 in a particular cell of the matrix denoted that stimulus B is preferred to stimulus A, and a 0 indicated the reverse condition. Table II shows the matrix obtained by summing the matrices for the ten subjects. Careful inspection of this matrix shows a strong preference for dithered speech over straight PCM encoding at a fixed number of bits/sample, and, in many cases, a preference for dithered speech at L bits/sample ($L = 2$–4) over straight PCM encoded speech at $(L + 1)$ bits/sample.

To verify these preference results, the data was analyzed using a multidimensional preference program of Carroll.[4] The program indicated

TABLE II—MATRIX OF SUM OF PREFERENCES FOR PAIRED COMPRESSION PREFERENCE TEST

Stimulus A \ Stimulus B	20	21	30	31	40	41	50	51	60	61
20	3	8	8	10	10	10	10	10	10	10
21	0	5	5	9	7	10	10	10	10	9
30	0	5	0	9	9	10	10	10	10	10
31	0	1	0	4	2	10	10	10	10	10
40	0	1	0	8	4	9	9	10	10	10
41	0	0	0	0	2	7	4	8	9	10
50	0	0	2	4	1	6	4	10	10	10
51	0	0	0	1	1	4	0	7	8	7
60	0	0	0	0	0	1	0	7	2	3
61	0	0	0	0	1	0	1	3	1	4

Number of preferences of B over A in 10 trials

that the preferences were essentially one-dimensional (over 95 percent of the variance was accounted for by one dimension), and produced a graphical interpretation of the overall preferences which is shown in Fig. 2. Since the preference judgments were one-dimensional, all the conditions lie on a line. The direction of preference goes from left to right in terms of decreasing preference. Figure 2 clearly shows:

(*i*) For a fixed number of bits/sample the dithered speech samples are always preferred to straight PCM encoding,

(*ii*) For 2–4 bits/sample, dithered speech is preferred to straight PCM encodings even with one extra bit/sample, i.e., condition 41 is preferred to condition 50, condition 31 is preferred to condition 40, and condition 21 is preferred to condition 30.

Thus in some perceptual sense, dithered PCM speech has a one-bit advantage over straight PCM encoding under certain conditions. This,

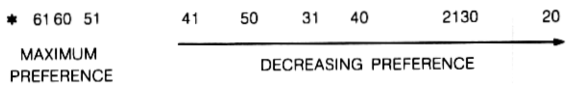

Fig. 2—Ordering of the stimuli in terms of preference.

of course, is not correct in terms of physical measures such as signal-to-noise ratio, or, as we will see, word intelligibility.

A complete analysis of variance was performed on the preference data and the results of this analysis are shown in Table III. The three factors and the number of levels of each are:

(*i*) number of bits/sample (5)
(*ii*) type of quantization (2)
(*iii*) subjects (10)

The analysis reconfirms the conclusions already discussed in that the most significant effects (significance $\gg$ 0.999 level) were number of bits/sample, and type of quantization. Subjects were significant at the 0.95 level, and the interaction between bits and dither was also significant at this level.

IV. WORD INTELLIGIBILITY TEST

The purpose of the intelligibility test was to determine the effects of dithering on the intelligibility of isolated monosyllables. As discussed earlier, the effect of dither is to make the quantization noise act like an additive wideband uncorrelated noise. Earlier studies[5] have indicated that such a noise tends to mask consonants, thereby lowering intelligibility. The effect of the correlated quantization noise on straight PCM encoding on word intelligibility was also measured.

In this experiment, 200 PB words[6] (Lists 2, 4, 5 and 6 in Ref. 5) were recorded, digitized, and stored on the disc of the DDP-516. The words were accessed at random, in groups of 50 (i.e., an entire list was processed before a new list was used), by one of the ten systems used in the

TABLE III—ANALYSIS OF VARIANCE OF PREFERENCE DATA

Factor	Degrees of Freedom	Mean Square	F-ratio	Significance Level
Subjects (s)	9	1.8	2.3	0.95
Type of quantization (TQ)	1	53.3	67.7	$\gg$0.999
Number bits per sample (NB)	4	142.9	181.5	$\gg$0.999
S × TQ	9	0.8	1.1	N.S.*
NB × S	36	1.0	1.2	N.S.
NB × TQ	4	2.3	2.9	0.95
Residual	36	0.8		

* N.S. ⇒ not significant above 0.90 level.

preference test. The 200 words were divided into two tests of 100 words, each test containing 10 versions of each stimulus condition. The same ten subjects were used in the intelligibility test as in the preference test. The two tests were given on separate days to all ten subjects.

V. RESULTS OF INTELLIGIBILITY TEST

Table IV shows the average error scores as a function of the number of bits/sample, and the type of quantization. (The notation of the previous section is used again here.) These data are averaged over subjects and tests. This table shows that at 2 bits/sample, the PCM system has an error rate of 59.5 percent as opposed to 76 percent for the dither system, i.e., a decrease of 16.5 percent in word intelligibility due to consonant masking. At 3 bits/sample, the PCM system has an error rate of 34.5 percent whereas the dither system has an error rate of 46.5 percent. Thus even at 3 bits/sample, the masking of the dither noise reduces word intelligibility by about 12 percent. At 4-6 bits/sample, the dither system has lower error rates than the PCM system–the differences being 10 percent at 6 bits/sample, 1.5 percent at 5 bits/sample and 0.5 percent at 4 bits/sample. Thus only at 6 bits/sample is the error rate difference significant. The data of Table IV are plotted in Fig. 3 to show how the error rate varies with the number of bits/sample for the two systems.

A complete analysis of variance was performed on the raw data of the intelligibility test. The four factors used in the analysis (and the number of levels of each factor) were

(*i*) number of bits/sample (5)
(*ii*) type of quantization (2)
(*iii*) subjects (10)
(*iv*) repetitions (2)

Table IV—Word Error Scores Averaged Over Subjects and Repetitions

Number of Bits per Sample	Error Rate		
	PCM	Dither	Difference
2	59.5%	76%	−16.5%
3	34.5%	46.5%	−12%
4	29.5%	29%	0.5%
5	25%	23.5%	1.5%
6	16.5%	6.5%	10%

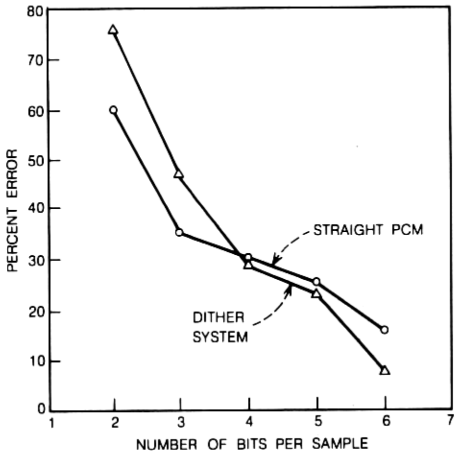

Fig. 3—The percentage error for word intelligibility as a function of the number of bits/sample for straight PCM and dither systems.

The results of the analysis are shown in Table V. The most significant factor was, of course, the number of bits/sample. The next most significant factors were subjects, repetitions, bits/sample × type of quantization, and bits/sample × repetitions. These results indicate a fairly large amount of learning between repetitions 1 and 2, as well as a lack of consistency between the intelligibility scores of the different subjects.

VI. CONCLUSIONS

The results of the preference test were quite encouraging in that subjects uniformly showed strong preferences for dithered speech over straight PCM encoding at all bit rates employed in the experiment. At the lower bit rates, the preference for dithered speech over higher bit rate PCM encoded speech presents strong evidence for the perceptibility and annoyance of highly correlated quantization noise.

The word intelligibility tests showed that the wideband uncorrelated dither noise tended to mask the consonants more than the correlated PCM noise thereby reducing word intelligibility by about 14 percent at low bit rates. At the higher bit rates used in the experiment, there was no decrease in word intelligibility for the dither system, and, in fact, at 6 bits/sample, the dithered words were 10 percent more intelligible than the straight PCM encoded words. Since the average percentage correct for the PCM system was 83.5 percent, an increase of 10 percent is a significant increase in intelligibility.

Overall, these experiments indicate that the use of dither noise in the range of 4–6 bits per sample has many beneficial effects.

TABLE V—ANALYSIS OF VARIANCE OF INTELLIGIBILITY DATA

Factor	Degrees of Freedom	Mean Square	F-ratio	Significance Level
Repetitions (R)	1	29.6	30.5	>0.999
Subjects (S)	9	5.7	5.8	>0.999
Type of quantization (TQ)	1	5.4	5.6	0.975
Number bits per sample (NB)	4	180.3	185.6	≫0.999
R × S	9	2.1	2.2	0.95
TQ × R	1	0.05	0.05	N.S.*
TQ × S	9	1.9	1.9	0.90
NB × R	4	7.1	7.3	>0.999
NB × S	36	1.7	1.7	0.90
NB × TQ	4	11.6	11.9	>0.999
TQ × R × S	9	0.5	0.5	N.S.
NB × R × S	36	1.1	1.1	N.S.
NB × TQ × R	4	0.3	0.3	N.S.
NB × TQ × S	36	0.8	0.8	N.S.
Residual	36	1.0		

* N.S. ⇒ not significant above 0.90 level.

VII. ACKNOWLEDGMENTS

The authors would like to thank J. L. Flanagan for suggesting the idea of applying dither noise to PCM encoding, and Professor H. Levitt and S. Pruzansky for assistance in analyzing the data and suggestions on the experimental procedures.

REFERENCES

1. Roberts, L. G., "Picture Coding Using Pseudo Random Noise," IRE Trans. Inform. Theory, *IT-8*, No. 2 (February 1962), pp. 145–154.
2. Jayant, N. S., and Rabiner, L. R., "The Application of Dither to the Quantization of Speech Signals," B.S.T.J., *51*, No. 6 (July–August 1972), pp. 1293–1304.
3. Davis, H., and Silverman, S. R., Appendix in *Hearing and Deafness*, Holt, Rinehart and Winston, 1970.
4. Carroll, J. D., "Non-parametric Multidimensional Analysis of Paired Comparisons Data," unpublished work.
5. Miller, G. A., and Nicely, P. E., "Analysis of Perceptual Confusions Among Some English Consonants," Jour. Acoust. Soc. Amer., *27*, No. 2 (March 1955), pp. 338–352.
6. Egan, J. E., "Articulation Testing Methods," Laryngoscope, *58*, No. 9 (September 1948), pp. 955–991.