Subjective evaluation of four low-complexity audio coding schemes

Stella M. Joseph and Robert C. Maher[a)]
Department of Electrical Engineering and Center for Communication and Information Science, University of Nebraska-Lincoln, 209N WSEC 0511, Lincoln, Nebraska 68588-0511

(Received 31 August 1994; accepted for publication 28 January 1995)

In this study the subjective performance of four low-complexity audio data compression methods are compared, operating at nominal bit rates of 2, 3, 4, and 5 bits per sample, applied to four 20-kHz bandwidth, 16-bits per sample digitized musical signals. The simple compression schemes compared were elementary differential pulse-code modulation (DPCM), noise feedback coding DPCM (NFC-DPCM), adaptive quantizer DPCM (DPCM-AQB), and a recently proposed method known as recursively indexed quantizer DPCM (RIQ-DPCM). Pairs consisting of a reconstructed signal and a reference signal were presented in a two-interval preference experiment. The reference signals were processed for specified levels of modulated noise reference unit (MNRU) in order to estimate the equality threshold rating (ETR) of the reconstructed audio stimuli. The subjective MNRU values were found to increase by 2–5 dB for each increment in bits per sample. The DPCM-AQB scores were found to be 8–10 dB higher than for DPCM and NFC-DPCM. RIQ-DPCM was rated highest, exceeding the DPCM-AQB results by 2–5 dB in all tests. Objective measurements of segmental signal-to-noise ratio (SNRSEG) for the reconstructed signals predicted a performance level 2–5 dB lower than was actually found in the subjective results, particularly for SNRSEG values below 25 dB.

PACS numbers: 43.60.Dh, 43.60.Cg

INTRODUCTION

Digital data compression can play an important role in the storage and transmission of wideband (20 Hz–20 kHz) audio signals in many practical communications systems. Like all data compression methods, audio compression schemes are intended to reduce the inherent redundancy of the signal. Although lossless compression is obviously desirable from a fidelity standpoint, the ill-defined statistical character of wideband audio signals prevents average compression ratios any greater than perhaps 5:4 for most lossless techniques. Obtaining a useful degree of data compression, say, 5:1, requires *lossy* techniques in which the reconstructed data stream is not identical to the original. Therefore, an important performance consideration for lossy coders is to evaluate the subjective quality of the reconstructed signal. Formal subjective testing is required in general because simple objective measures do not always correlate well with human perceptual judgments.

This study involved a subjective test conducted to compare the perceptual quality of musical signals processed by four simple digital audio coders operating at four bit rates. All of the coders were based on the differential pulse-code modulation (DPCM) framework, and may be classified as low-complexity coders requiring minimal computation.

The equality threshold rating (ETR) was used to compare the subjective quality of each digital coder with a set of reference signals (Dimolitsas, 1991). The modulated noise reference unit (MNRU) was used as the reference scale (CCITT, 1989).

[a)]E-mail: rmaher@unl.edu

The results of the subjective test were also compared to an elementary objective performance measure, the segmental signal-to-noise ratio (SNRSEG). SNRSEG is defined for an arbitrary N-segment signal (Jayant and Noll, 1984):

$$\text{SNRSEG} = \frac{1}{N}\sum_{n=0}^{N-1} \text{SNR}_n . \tag{1}$$

SNR_n is the signal-to-noise ratio of segment n expressed in dB, viz.:

$$\text{SNR}_n = 10 \log\left[\frac{\sum_{m=nM}^{(n+1)M-1} x^2[m]}{\sum_{m=nM}^{(n+1)M-1} (x[m]-\hat{x}[m])^2}\right], \tag{2}$$

where n is the segment number, M is the length in samples of each block (2200 samples in this study), $x[m]$ is the original (uncoded) signal, and $\hat{x}[m]$ is the reconstructed signal.

I. THE CODING TECHNIQUES

In this section a brief description is given of the four low-complexity waveform coders employed in this experiment.

One of the simplest schemes used for data compression of digital audio signals is differential pulse-code modulation (Jayant and Noll, 1984). Most audio signals are oversampled and exhibit a long-term low-pass spectral characteristic that results in a significant correlation between successive samples of the input signal. The DPCM coder exploits this feature essentially by transmitting the quantized difference, $\hat{e}[m]$, between the current input sample, $x[m]$, and a linearly *predicted* sample value, $\hat{x}[m]$. The structure of the basic DPCM coder is shown in Fig. 1.

 0001-4966/95/97(6)/3657/6/$6.00

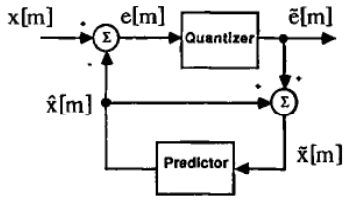

FIG. 1. DPCM encoder. The elementary DPCM encoder transmits the quantized difference between the current input value, $x[m]$, and a predicted value, $\hat{x}[m]$, calculated from knowledge of the previous coder output values.

The DPCM decoder recovers the quantized approximation to the input signal by summing the quantized adjacent sample differences. The structure of the DPCM decoder is shown in Fig. 2.

The basic DPCM scheme, though efficient, exhibits the fundamental tradeoff between low bit rate (requiring a *small* number of *large* quantization steps) and low reconstruction error (requiring a *large* number of *small* quantization steps).

The second coder used in the test incorporated a feedback structure for the quantization noise (Jayant and Noll, 1984). The noise feedback coding (or "noise shaping") was utilized to reduce the perceived reconstruction error by redistributing (filtering) the quantization noise in such a way that the noise spectrum was reduced in the frequency range where the human hearing apparatus is most sensitive (Wannamaker, 1992). Hence, the reconstructed signal may be *perceived* as being less noisy even though the total quantization noise power may be unchanged or possibly even greater than in the unshaped case. The NFC structure with DPCM is depicted in Fig. 3.

The third coder involved the basic DPCM structure of Fig. 1, but incorporated an adapative quantizer with variable step size, resulting in adaptive quantizer DPCM. The particular strategy used here is a simple single-step backward-adaptive procedure referred to as DPCM-AQB (Jayant and Noll, 1984). The step size multipliers used for the adaptive quantizer are given in Table I.

The fourth coder replaced the quantizer in the basic DPCM structure with a recursively indexed quantizer (RIQ) (Sayood and Na, 1993). The recursively indexed quantizer, as its name implies, uses a recursive algorithm to eliminate quantizer overload distortion. If the input signal to the quantizer is larger than the base quantization range, the instantaneous bit rate is increased to accommodate the large input without quantizer clipping. If the probability of quantizer overload is low the average bit rate is not adversely affected (Sayood and Na, 1992).

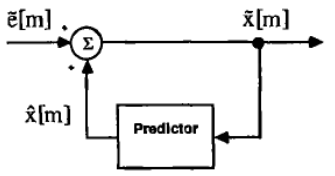

FIG. 2. DPCM decoder. The elementary DPCM decoder uses the same prediction filter as the encoder, and accumulates (sums) the input sequence.

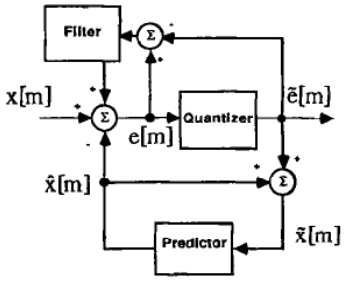

FIG. 3. NFC-DPCM encoder. In conceptual form, the noise feedback coding (NFC) scheme uses a feedback filter for the quantizer error in order to shape the spectrum of the quantization noise for reduced audibility.

II. METHODS

This study investigated the relative degradation in perceived quality of four wideband audio signals processed by the four low-complexity, lossy waveform coders, each operating at four nominal bit rates (2, 3, 4, and 5 bits per sample). The set of reference signals for the test consisted of the same set of four audio signals processed with a range of modulation noise reference unit values, or Q's according to the CCITT performance specifications (CCITT, 1989).

By subjectively comparing the quality of a coding system to a set of MNRU reference signals it is possible to estimate the Q value which equals the perceived degradation of the system under test (equality threshold rating). The method described here leads to a threshold of equality defined as the 50% preference level between a particular MNRU-processed signal and the digital system, i.e., the Q value where approximately half of the subject responses indicate a preference for the MNRU reference and the other half prefer the waveform coder under test. This method is

TABLE I. Stepsize multipliers for adaptive quantizer used in DPCM-AQB. The multiplier value is applied to the quantizer step size according to the previous quantizer output level. Numbers less than 1 decrease the step size, while numbers greater than one increase it.

Quantizer output level	Bits			
	2	3	4	5
1	0.8	0.9	0.9	0.9
2	1.6	0.95	0.9	0.9
3		1.5	0.95	0.9
4		2.0	0.95	0.9
5			1.2	0.95
6			1.4	0.95
7			1.8	0.95
8			2.3	0.95
9				1.2
10				1.5
11				1.8
12				2.1
13				2.4
14				2.7
15				3.0
16				3.3

expected to give stable and precise results even for high quality digital processes (CCITT, 1989; Dimolitsas, 1991; Rosenberger, 1989).

A. Subjects

The 15 volunteer subjects (four women and eleven men) were between the ages of 20 and 32 years. Two of the subjects were electrical engineering professors familiar with audio coding techniques, while the remaining subjects were students in various departments at the University of Nebraska-Lincoln. None of the subjects had any prior information about the test parameters and all of the tests were administered in the presence of one of the investigators (SJ). All of the subjects expressed at least an informal interest in recorded music, and many of the subjects had previously had some sort of musical training.

Since the subjects were only required to indicate a preference for one of the signals in a signal pair, no specific training trials were used. All subjects were required to read and sign an institutional adult consent form before the start of the experiment.

B. Screening

A simple absolute threshold test was conducted to screen the subjects for any substantial hearing defect. The subjects were presented with a series of pure tones at randomly increasing and decreasing loudness steps, or "staircases." Each staircase contained between 5 and 10 steps. After each staircase the subject responded with the number of steps heard in the sequence. The procedure was repeated for the frequencies 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. To be considered suitable for the test the subject's threshold was required to be within 5 dB of the average free-field threshold (ISO, 1961).

A screening test was also conducted to determine whether each subject could reliably distinguish a pure tone from a noisy tone (Johnston, 1988). The screening test compared a clean (>90 dB signal-to-quantizing noise ratio, 44.1-kHz sample rate) 1-kHz sinewave to a sinewave with critical band noise added. The subjects were asked to indicate the pure tone of each pair. None of the subjects was excluded on this basis.

C. Stimuli

Four source recordings were used for the subjective test. All were music signals, digitally transferred to a computer workstation from high quality compact disc recordings (16-bit, 44.1-kHz sample rate). The two channels of the original stereo signals were digitally summed to a monophonic signal prior to processing.

Source 1: **Violin** solo with orchestral accompaniment: excerpt from Mozart Violin Concerto #5; 5.4 s (Denon, 1985).

Source 2: Solo **castanets:** rhythmic pattern; 6.15 s (EBU, 1988).

Source 3: Solo **soprano** singer: phrase sung in Latin; 4.85 s (EBU, 1988).

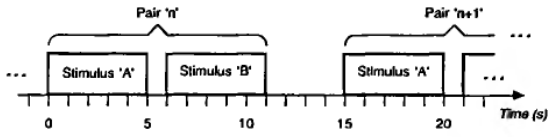

FIG. 4. Stimulus presentation timing diagram.

Source 4: Rock-and-roll piece excerpt, Steve **Winwood**: keyboards, drums, and vocals; 6.75 s (Winwood, 1988).

The audio stimuli were edited, processed, and assembled into sets with recorded announcements and instructions. The entire stimuli sets were then digitally transferred to digital audio tape (DAT) for use during the actual trials.

D. Apparatus

The single channel audio stimuli were presented diotically to the subjects in a quiet, windowless room from a Panasonic SV-3500 DAT machine via a pair of high quality electrostatic headphones (STAX Lambda Pro) driven by a STAX model SRM-1/MK-2 amplifier. The peak acoustic signal level was adjusted to 70 dB SPL (*re:*20 μPa) using a Bruel & Kjaer model 2235 sound level meter and a simple aluminum flat-plate coupler prior to the arrival of each subject. The sound level meter itself was calibrated prior to each measurement using a Bruel & Kjaer model 4230 calibrator.

E. Procedure

Only one subject was tested at a time. The four source signals were processed by the four coders running at four bit rates, resulting in a total of 64 coded signals. For each of the 64 test signals, four corresponding MNRU reference signals were generated. The Q values for the four reference signals in each group were selected in advance so that the reference signal with the highest Q was judged by the investigators to be noticeably better than the particular coded signal and the reference signal with the lowest Q was noticeably worse than the coded signal. The two remaining MNRU reference signals in each group were chosen to be in between the extreme values so that the likely 50% preference level was spanned.

The coded-signal + reference-signal pairs were presented in both orders, resulting in (64 coded signals)×(4 reference signals)×(2 presentation orders)=512 pairs presented in a random sequence. A 4-s gap between the pairs was provided to allow the subject some time to mark the response sheet (forced choice). A timing diagram of the presentation is shown in Fig. 4.

The total stimulus time for the test was approximately 2 h and 30 min, divided into three 50-min sessions conducted over a two day period. On the first day the absolute threshold and screening tests were conducted, followed by the first 50-min session. Sessions two and three were conducted on the second day. Each session had 12 4-min long sets consisting of 14 or 15 signal pairs. A 1-min rest interval followed each set and a 10-min rest period separated each session on the second day. The test subjects were provided with a sequence of instructions, both written and oral, during the test.

III. RESULTS AND DISCUSSION

A. Processing of raw responses

The 512 responses from each subject were collected and combined for a one-sample (binary) nonparametric statistical analysis (Siegel and Castellan, 1988). The chance probability, P, of obtaining preference responses as extreme or more extreme than the observed value is determined by

$$P[Y \geq k] = \sum_{i=k}^{n} \binom{n}{i} p^i q^{n-i}, \tag{3}$$

where

$$\binom{n}{i} = \frac{n!}{i!(n-i)!} \text{ is the binomial coefficient.}$$

In this experiment $N=30$ (15 subjects providing responses for both A-B and B-A order), $p=q=0.5$, and k is the larger of the number of responses preferring A and preferring B for that comparison. A 90% level of significance was arbitrarily chosen, meaning that if the chance probability of occurrence was more than 0.1 then it could not be claimed that either A or B was significantly preferred in that pair. A 90% level of significance for $N=30$ requires $k \geq 19$. This simple statistical comparison was deemed sufficient for the goals of this study.

The Q value (subjective SNR) of the coder was assigned to be the single lowest MNRU value among the four reference signals for which the chance probability was greater than 0.1, i.e., where there was no significant preference for the coded signal or the MNRU signal. If more than one of the four reference signals resulted in a calculated probability greater than 0.1, the average value of the transition point (switch from majority preferred A to majority preferred B) was used to designate the Q value of the coder. In cases where none of the four paired comparisons resulted in a significant preference, the extreme MNRU value of the four was chosen as the subjective equivalent SNR.

B. Subjective and objective coder performance

The calculated Q values for each of the coders are shown in Fig. 5, obtained by a simple average over the four source signals. Two expected general trends can be identified: (i) higher bit rates result in higher Q values for each of the coders, and (ii) the coders with adaptive quantizers (DPCM-AQB and RIQ-DPCM) result in 10–15 dB higher Q values at each bit rate than the lower complexity, nonadaptive coders (DPCM and NFC). For example, it is seen that DPCM-AQB with a bit rate of 2 performs as well as DPCM with a bit rate of 4, and RIQ at a bit rate of 2 performs better on average than DPCM with a bit rate of 5. Thus a modest increase in computational complexity provides a gain of 2 to 3 bits per sample compared to basic DPCM when coding these musical selections.

The recently proposed RIQ-DPCM technique was rated highest in the subjective tests, exceeding the DPCM-AQB results by between 2 and 5 dB in all comparisons. This improvement was nearly 1 bit per sample, i.e., RIQ at bit rate R was generally very close to DPCM-AQB at bit rate $R+1$. Thus the ability of RIQ-DPCM to eliminate overload distortion in the DPCM framework with minimal increase in computational complexity is worthy of additional study and development.

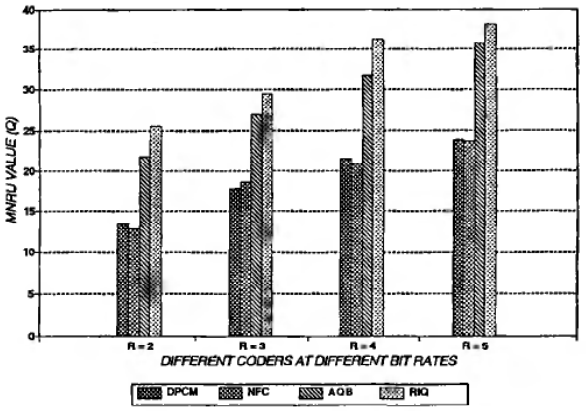

FIG. 5. Q values from subjective test. The calculated Q values in this figure were obtained by a simple average over the four source signals. As expected, increasing the bit rate generally improves the subjective quality, as does the use of more sophisticated coders such as RIQ-DPCM.

Since SNRSEG is often used as a subjectively meaningful objective measure, it is interesting to compare a quality ranking based on SNRSEG to a ranking based on the subjective responses (Table II). A ranking difference occurs for DPCM and NFC at low bit rates, where the noise feedback seems to improve the subjective quality of the quantization noise compared to the straight DPCM systems. At higher bit rates the ranking based on SNRSEG is the same as the ranking based on the subjective test results. This observation is useful because it is often convenient to perform initial evaluation testing on a new coder using simple objective quality

TABLE II. Subjective and objective ranking of coder performance. Ranking from lowest to highest quality. DPCM: differential pulse code modulation (elementary). NFC: noise feedback coding DPCM. AQB: adaptive quantizer DPCM. RIQ: recursively indexed quantizer DPCM. Number indicates bits per sample.

Subjective	Objective (SNRSEG)
NFC 2	NFC 2
DPCM 2	NFC 3
DPCM 3	DPCM 2
NFC 3	NFC 4
NFC 4	DPCM 3
DPCM 4	NFC 5
AQB 2	DPCM 4
NFC 5	AQB 2
DPCM 5	RIQ 2
RIQ 2	DPCM 5
AQB 3	AQB 3
RIQ 3	RIQ 3
AQB 4	AQB 4
AQB 5	AQB 5
RIQ 4	RIQ 4
RIQ 5	RIQ 5

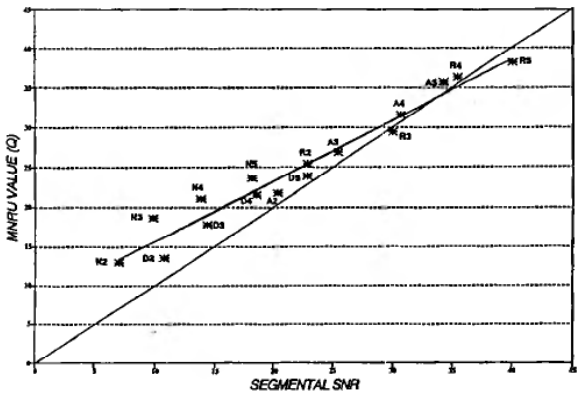

FIG. 6. Comparison of subjective (MNRU) and objective (SNRSEG) performance measures. The 45° diagonal line indicates the locus of points that would be occupied if the subjective and objective results were in perfect agreement. The least-squares fit to the data is also shown. The objective results underestimate the subjective results, particularly for segmental SNR levels below 25 dB.

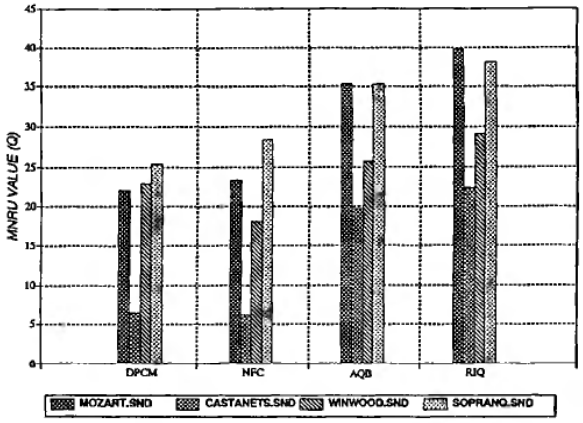

FIG. 7. Subjective performance for different musical signals. The Q values for each musical signal averaged over the four bit rates for each of the four different coders as shown. The subjective performance of each of the coders varies by as much as 20 dB over the four different musical signals, indicating the need for a wide range of musical styles in subjective quality testing.

measures (like SNRSEG), prior to implementing a formal subjective test procedure.

Another useful observation can be made by simultaneously plotting the subjective and SNRSEG values, as shown in Fig. 6. The diagonal line with slope of 1 indicates the locus of points that would be expected if SNRSEG and the subjective results were in perfect agreement. A least-squares fit for the data (using SNRSEG as the abscissa) is also shown in the figure. Note that for SNRSEG levels below approximately 25 dB, SNRSEG underestimates the subjective values for the stimuli considered in this study by between 2 and 8 dB. The bias is less apparent at the higher SNRSEG values.

The difference in subjective and objective results is perhaps most noticeable when comparing the NFC cases (N2–N5) in Fig. 6 to the corresponding DPCM cases (D2–D5). The SNRSEG results actually show a decrease of approximately 5 dB (horizontal axis) when switching from basic DPCM to NFC, while the subjective results show little difference in preference (vertical axis). Thus, the increase in total quantization noise level for NFC has a much smaller perceptual impact than predicted by SNRSEG. Note, however, that the simple NFC structure provides negligible perceptual improvement compared to DPCM at the bit rates examined in this study.

C. Performance variations due to specific signal characteristics

In Fig. 7 a plot is shown of the subjective values of each musical signal averaged over the four bit rates for each of the four different coders. It can be seen that the subjective performance of each of the coders varies by as much as 20 dB over the four different musical signals.

The castanets signal resulted in relatively poor subjective performance by all of the coders. The castanets signal consists of a series of high amplitude rhythmic clicks separated by nearly silent periods. It was observed informally that quantizer overload in the DPCM, NFC, and DPCM-AQB coders caused the characteristically pointed signal quality at the onset of the clicks to change into a more thumplike timbre. Thus the poor overall performance for the castanets signal was not unexpected because the predictive coders used in this study were ill suited to the abrupt transients and noise present in the signal and the nearly silent passages between the castanets clicks.

At the other extreme, the soprano signal is a solo legato piece with a few consonants and little reverberation. Similarly, the Mozart signal is an excerpt with a rhythmically smooth orchestral texture. Both of these signals exhibit spectral levels declining by 6–12 dB per octave above 1 kHz. Predictive coders are expected to handle such waveforms with minimal overload error, and this is borne out in the subjective results.

The subjective results for the rock-and-roll example, Winwood, appear between the relatively poor results for the castanets signal and the better results for the soprano and Mozart sources. The Winwood example contains a significant proportion of high-frequency transients (drums and cymbals) that are difficult for the predictive coders to handle. On the other hand, the complex sonic texture seems to reduce the detectability of the coding distortion, resulting in a higher subjective score than for the castanets signal.

The RIQ-DPCM performance exceeds the other three coders on each of the four signals by 2–5 dB. This again indicates a need for further investigation of the RIQ-DPCM procedure for audio data compression. The effect of NFC compared to DPCM is less impressive in this study, with NFC showing a slight perceptual improvement for the Mozart and soprano signals, little change for the castanets signal, and a decrease by nearly 5 dB for the Winwood signal.

In summary, a coder that performs well for a particular musical example is not guaranteed to perform equally well for some another example. These results indicate that a wide range of musical recordings and styles must be employed

when evaluating waveform coders for audio data compression.

IV. CONCLUSION

Three conclusions can be drawn from this study.

First, it is seen that the performance predictions of simple objective measurements for lossy coders (SNRSEG in this case) must be compared to subjective results. From this study, segmental SNR values below 25 dB tend to underestimate the actual perceptual quality by 2–5 dB.

Second, the results of this study reinforce the general expectation that different musical signals processed by the same coder will typically have different subjective quality levels for the reconstructed signals. Therefore, it is important to consider a worst-case range of audio styles when evaluating audio data compression quality.

Finally, the subjective performance of the RIQ-DPCM technique for lossy data compression exceeds the performance of the DPCM-AQB scheme of similar computational complexity by between 2 and 5 dB on average. This result is sufficiently encouraging to consider further development of RIQ-DPCM for audio data compression situations in which a low-complexity coder is required for economic or other practical reasons.

ACKNOWLEDGMENTS

This paper is based on a thesis submitted by Stella Joseph to the Department of Electrical Engineering, University of Nebraska-Lincoln, in partial fulfillment of the requirements for the Master of Science degree. The authors would like to acknowledge the support and assistance of the Nebraska Center for Communication and Information Science (Stuart Margolis, Director), and to offer special thanks to Khalid Sayood and Michael Hoffman for their insightful comments and encouragement, and to Diane Van Werden for her assistance with the experiment.

CCITT (International Telegraph and Telephone Consultative Committee) (**1989**). Blue Book **5**, 198–203, 341–358.

Denon (**1985**). Digital audio check CD, audio compact disc (Nippon Columbia Co., Ltd., Japan), No. C39-7441, track 17.

Dimolitsas, S. (**1991**). "Subjective quality quantification of digital voice communication systems," IEE Proceedings-I (Communications) **138**, 585–595.

European Broadcast Union (EBU) (**1988**). *Sound Quality Assessment Material (SQAM)*, audio compact disc (Polygram, Hanover, FRG), Cat. No. 422 204-2, tracks 27, 44.

International Standards Organization (**1961**). "Normal equal-loudness contours for pure tones and normal threshold of hearing under free field listening conditions" ISO-R-226, (ISO, New York).

Jayant, N. S., and Noll, P. (**1984**). *Digital Coding of Waveforms* (Prentice-Hall, Englewood Cliffs, NJ).

Johnston, J. D. (**1988**). "Transform coding of audio signals using perceptual noise criteria," IEEE J. Selected Areas Commun. **6**, 314–323.

Rosenberger, J. R. (**1988**). "Quality assessment methods for speech coding," Telecomm. J. **55**, 820–825.

Sayood, K., and Na, S. (**1992**). " Recursively indexed quantization of memoryless sources," IEEE Trans. Inf. Theory **38**, 1602–1609.

Sayood, K., and Na, S. (**1993**). "Recursively indexed differential pulse code modulation," DIMACS Series in Discrete Mathematics and Theoretical Computer Science **14**, 253–263.

Siegel, S., and Castellan, N. J., Jr. (**1988**). *Nonparametric Statistics* (McGraw-Hill, New York), 2nd ed.

Wannamaker, R. A. (**1992**). "Psychoacoustically optimal noise shaping," J. Audio Eng. Soc. **40**. 611–620.

Winwood, S. (**1988**). *Roll With It*, audio compact disc (Virgin Records Ltd., Beverly Hills, CA), No. 2-90946, track 4.