LOW-BITRATE CODING OF SOUND AND IMPLICATIONS FOR HIGH-QUALITY DIGITAL AUDIO*

Adrianus J.M Houtsma

Aircrew Protection Division
U.S. Army Aeromedical Research Laboratory
Fort Rucker, AL 36362-0577, USA
Adrian.Houtsma@amedd.army.mil

ABSTRACT. The Compact Disc launched in the early 1980s has set a worldwide standard for high-quality digital audio. Its sampling rate of 44.1 kHz and 16-bit word size appeared to provide enough bandwidth and dynamic range to capture every sound feature perceivable by the human ear. Although digital technology today is capable of delivering an even larger dynamic and spectral range at relatively little cost, there has been an ever-increasing need to produce digital sound of CD-equivalent quality at much lower bitrates than the CD's 1411 kbit/s. Examples of instances requiring lower bitrates are digital audio broadcast, fixed-head digital tape recorders, recordable mini-disc, multi-channel cinema sound, and internet audio. The original bit-compression schemes of transform and subband coding developed in the late 1980s, both exploiting the known perceptual limitations of the human auditory system, have now mostly been merged into a number of standardized complex coding systems that combine the best elements of the original schemes. Other techniques have been developed such as model-based parametric coding where not the waveform but rather the controlling parameters of a waveform-generating model are encoded and transmitted. Synthesis of the intended sound wave is then accomplished at the decoder end. This technique leads to additional bit savings and gives the end-user control over playback conditions like the positioning of sound sources. This feature is not only potentially useful for music playback in the home, but also for live electronic concerts, teleconferencing, and multi-user voice communication systems.

1. INTRODUCTION

High-quality audio is a concept that is not exactly defined and not always properly understood. To some it refers directly to physical similarity between a real sound field and its electro-acoustical reproduction. Within this viewpoint, acoustical knowledge and electronic technology are the only limiting factors preventing audio quality from being perfect. To others, however, audio quality refers to the audible similarity between a real-life sound event and an electronic reproduction. Given this viewpoint, the human auditory system with all its limitations has become an essential factor determining audio quality.

If one measures the locus of all pure tone frequency–intensity combinations that a young and healthy listener is able to perceive without experiencing discomfort, one typically finds a frequency range extending from 20 Hz to 20 kHz and an intensity range from about 10 to 110 dB. Frequency spectra of musical instruments and the human voice, considered all together, largely cover this frequency range. The dynamic range of a full symphony orchestra is, depending on size and concert hall acoustics, somewhere between 90 and 100 dB, although dynamic ranges of individual instruments are considerably smaller [1]. Figure 1 shows the typical frequency-intensity area for human hearing and, inside that area, a conservative estimate of the area covered by conventional music.

For digital storage and reproduction of sound with a bandwidth of 20 kHz, one needs a sampling frequency of at least 40 kHz. To reproduce sounds with a dynamic range of 96 dB without audible quantization noise, one needs a sample word size of at least 16 bits. These considerations led to the Compact Disc format proposed by Philips and Sony in the early 1980s, and standardized in 1987 [2]. The CD uses 16-bit words and a sampling frequency of 44.1 kHz. The total bitrate in 2-channel CD audio is therefore a little over 1.4 Mbits/s. On the market, the CD replaced the LP ('vinyl') format in a few years, and quickly became an informal standard for audio quality across the world.

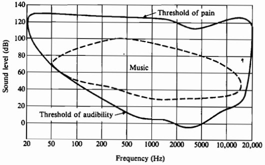

Figure 1: The auditory frequency and dynamic range. The 'music' range is represented by the dashed area. (From Rossing [42])

In subsequent years, storage and transmission of audio bits became progressively cheaper through advances in digital technology and, consequently, a mostly technology-driven

* Reprinted from the Proceedings of the International Congress on Acoustics ICA-2004, Kyoto, Japan

push towards an even higher digital audio quality standard ensued. In some studies, evidence was found that sound components well above 20 kHz can cause measurable effects on a-EEG activity [3], difference tone perception [4], and other perceptual features [5,6]. The Acoustic Renaissance for Audio (ARA) Foundation, residing at Meridian Audio Ltd in the UK, has been systematically pushing to establish stricter standards for high-quality audio on high-density CD carriers [7]. Critical evaluations of claims that sound components above 20 kHz are perceivable have, in some cases, uncovered artifacts like equipment distortion as the cause of the observed effects [8]. Therefore, it seems fair to conclude that psychophysical evidence in favor of a stricter and more bit-intensive CD standard is not overwhelming. It should be recognized, however, that for purposes of documenting and archiving sound recordings for later analysis, a more broadband format than the present CD standard would have its merits.

In many applications of digital audio other than compact disc or digital audio tape, there is an inherent bandwidth limitation that impedes or prevents straightforward application of the CD standard. Examples are digital audio broadcast (DAB), digital compact cassette (DCC), cinema sound (Sony Dynamic Digital Sound, Dolby ProLogic, Dolby Digital), and sounds transmitted over the internet. In some cases, the required audio bandwidth cannot be reached because of hardware limitations (DCC), and in other cases, there is bandwidth competition with other desirable features of the communication such as video quality or the number of available radio stations. In all of these cases, one ideally wants to reduce the audio bitrate of 1.4 Mb/s to something considerably lower, without affecting perceived sound quality. A straightforward approach of lowering the sampling rate would immediately be audible as a reduction of brightness due to the absence of high frequencies, whereas a reduction of the word size would result in (occasional) audibility of undesirable sound effects caused by quantization noise.

Bitrate reduction without audible sound degradation can be achieved by (a) reduction of statistical redundancy, and (b) minimization of perceptual irrelevance. Both processes are independent and cumulative. The first process removes redundancies in the code, is a strictly statistical mathematical operation, and results in *lossless* code (i.e., a code from which the original sound waveform is completely reconstructible). The second process exploits known limitations of our hearing system in order to remove inaudible sound elements and to selectively allow quantization noise that cannot be perceived. This results in a lossy, but at the same time, a *perceptually transparent* code. From such a code, the original sound waveform cannot be exactly reconstructed any more, but a listener will not be able to tell the difference between the original and the bitrate-reduced versions. Finally, if bitrates are reduced to a level where audible sound elements are removed or audible quantization noise is added, the result will be a *lossy* and *perceptually degraded* code.

2. TRANSPARENT CODING

A general scheme for perceptually transparent coding is shown in Figure 2. A running waveform is first subjected to a time—frequency analysis, shown in the top-left part of the figure, which can take several different forms depending on the coding strategy. Short time segments, comparable to the temporal resolution of our hearing system, are passed through a set of uniform or nonuniform. bandpass filters, as is done in *subband coding* [9]. The output parameters of this time-frequency analysis block in this case are the time responses of each subband to each input segment. *Transform coders* typically use some unitary transformation into the frequency domain, for instance a cosine transform [101. In this case, the output parameters are represented by a set of Fourier coefficients for each transformed time segment. Both subband and transform coders are purely waveform--based and do not assume any knowledge of how the waveform was produced.

Figure 2: General scheme for perceptual entropy coding. (From Painter and Spanias [15])

Another analysis strategy is to assume that the sound waveform is, in principle, made up of simple building blocks that can be modeled and adequately described by a limited set of parameters. *Parametric coders* try to describe the input sound segments in terms of sinusoidal parameters [11], combinations of sinusoids and noise [12], or frequency-modulated sinusoids [13]. Output parameters then take the form of sinusoidal amplitudes and frequencies, noise intensities and bandwidths, and modulation indices. Another example is the linear predictive coder [14] that has become very successful in speech but less successful in music coding. It is based on the idea of a source-filter model of sound, where the source is either a periodic pulse or noise and the filter is an n-pole linear filter. Such a sound model seems a reasonable representation for the human voice and, to some extent, for harmonic musical instruments. It appears less appropriate as a model for inharmonic tonal instruments such as bells, gongs or other percussion instruments.

It is not the purpose of this presentation to provide a complete list of all the coding strategies that have been proposed or implemented. A detailed and elaborately documented review of this material was recently provided by Painter and Spanias [15].

The next general operation is the quantization and encoding of the parameters yielded by the time—frequency analysis. The quantization process is adaptive, where the number of bits available for encoding parameters of each time frame is determined by the results a psychoacoustic analysis shown in the lower portion of Fig. 2. This analysis typically assesses (1) which parameters can safely be dropped because their effect is inaudible, and (2) how much quantization noise can be allowed for signal parameters within that time frame. The principal psychoacoustic effect exploited is simultaneous masking, where the perception threshold for quantization noise in any frequency band is raised by the simultaneously presence of a sound signal [16]. Forward and backward masking [17], in which a sound signal masks quantization noise in succeeding or preceding time frames, can, in principle, also be included. In two--channel stereo sound, masked thresholds for quantization noise may be lowered by binaural masking release [18], resulting in a somewhat higher bit allocation in comparison with mono sound. On the other hand, signal information in the different sound channels can, to some extent, be combined, lowering the over-all bitrate [19,20]. For multi-channel 3D sound presentation, sound channels may also be partially combined [20], and even transformed into a single monaural channel plus a stream of binaural parameters, as is done in *Binaural Cue Coding* [21]. This leads, on the one hand, to a very low bitrate/sound quality ratio and, on the other hand, provides more flexible control over source positioning at the output in comparison with conventional multi-channel coding and transmission techniques.

In the final stage of the general encoder of Fig. 2, all mathematical and statistical redundancies that are left in the signal code are removed. This stage of the processor is *lossless,* whereas the previous stages of the coder yielded a *lossy* code. Hence, the end result is a *lossy* code. Whether or not this code is perceptually transparent depends primarily on the available bitrate (i.e., the number of bits that can be allocated in each time frame). It also depends, however, on details of the design of the time/frequency analysis block and on the finesse and correctness of the psychoacoustical model that is used.

If available bitrates are too low in comparison with the degree of sophistication of signal preprocessing and psychoacoustic analysis, coding artifacts will occur that are likely to affect the perceived sound quality in a negative way. Subjectively, such artifacts may sound like brief irregular bursts of audible noise, distortion of timbre, or disturbances in perceived location of sound source images. A particularly well-known and frequently occurring artifact is *pre-echo.* This happens when time frames are chosen too long with respect to aural time resolution, and a strong music signal onset (i.e., a castanet or bell sound) falls in the later portion of a time frame. Because allowable quantization noise estimated by the psychoacoustic model is spread over the entire time frame, it is likely to be audible just before the onset. To estimate obtainable limits for transparent coding, Johnston [22] developed a general computation scheme for *Perceptual Entropy* based on quantitative psychoacoustic models for simultaneous and non-simultaneous masking. He predicted that, for a variety of coders, bitrates could be typically reduced to about 2.1 bits/sample (at sampling rates above 40 kHz) before perceptual transparency is lost. Early listening tests with an MPEG-1 layer I coder and simple fixed--difference or adaptive forced choice procedures between full 16-bit/sample and bit-compressed signals, yielded difference detection thresholds between 2 and 3 bits/sample, dependent on music fragment [23]. Subsequent technical developments of combining the best elements of different coding strategies, however, have yielded hybrid coders that are able to produce perceptually transparent codes at rates less than 1 bit/sample [24].

3. EXAMPLE OF A SUBBAND CODER

This section describes the principles of a specific subband coding scheme known as MUSICAM [25] that became the basis for the ISO/IEC MPEGI layers I and II standard [26]. To help understand how this coder works, a schematic representation of a 3-tone music signal against a quantization noise background typically produced by a PCM (i.e., CD-type) coder is shown in Fig. 3. It shows the masked threshold of our auditory system, given three masking tones at 250, 1000 and 4000 Hz and at a sound pressure level of approximately 70 dB. One can see that at the three tone frequencies, auditory threshold has been raised by about 50 dB compared with the threshold in quiet. Quantization noise for some arbitrary bitrate has been represented by a spectrally flat band of noise, which is an oversimplification because quantization noise, although

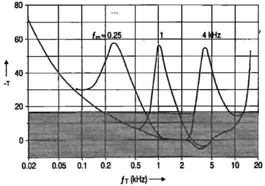

Figure 3: Threshold (LT) of a test tone in quiet and in the presence of three simultaneous masking tones. Shaded area represents arbitrary quantization noise.

spectrally broadband, is typically neither flat nor stationary. From the figure, it is clear that, for this situation, the quantization noise is audible in the valleys of the masking profile, at approximately 800, 2500, and 10000 Hz.

Figure 4 illustrates how one could ideally shape the quantization noise so that it is always below the masked threshold created by the signal components. Careful inspection reveals that the total quantization noise power in Fig. 4 is larger than that in Fig. 3, but nevertheless the noise shown in Fig. 4 will be inaudible. It is also clear that, because a music signal continuously changes with time, the quantization noise shaping should be adaptive and be updated very frequently.

In order to achieve this, the music signal is cut into short (typically 8-ms) time frames, and each time frame is decomposed into a set of responses of a bank of 32 constant-bandwidth (700-Hz) polyphase quadrature filters. Referring back to Fig. 1, this is done in the 'time/frequency analysis' block of the coder. The output 'parameters' are the (temporal) filter response waveforms. All responses are critically subsampled at 1400 Hz, so that at the output of the filter bank the total bitrate is still the same as at the input. Parallel to this filtering operation, the 'psychoacoustic analysis' block performs an FFT on the time frame and computes a masking profile from this spectral representation based on elementary psychoacoustic rules. Next, from the total number of bits available for that particular time frame, bits are allocated to each response signal in a manner that varies inversely with the

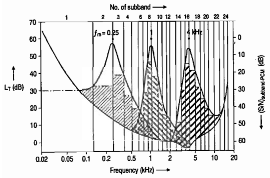

Figure 4: Same as Fig. 3, with quantization noise divided over 24 subbands. (From Wiese and Stoll [25])

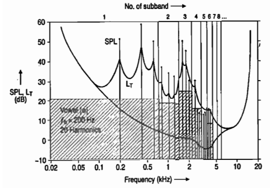

Figure 5: Spectrum of a vowel, masking pattern, and quantization noise as allocated in an MPEG 1 -layer II coder. (From Wiese and Stoll [25])

amount of masked threshold elevation at the center frequency of the passband for that signal. Since allocation of less bits implies generation of more quantization noise, the overall result will be that the spectral profile of the noise will generally conform to the masking profile of the signal, as is shown in Fig. 4.

Figure 5 illustrates the actual performance of the coder for a female-voice vowel segment comprised of 20 harmonics. Shown are (a) the spectral components of the vowel sound, (b) the masking profile of this vowel segment computed by the psychoacoustic analyzer, and (c) the spectrum of quantization noise allowed by the coder. One can observe that bitrate saving is achieved in two ways: (a) by allocating fewer bits in subbands where the masking profile is high, and (b) by allocating no bits at all where all signal components within a subband are inaudible, as is the case in subband 5. It is also clear that, if one wants to maintain perceptual transparency, any uncertainty about the correctness of the psychoacoustic model will require a larger safety margin between computed masked threshold and quantization noise levels. Therefore, the quality of the psychoacoustic model has a direct influence on the transparent bitrates that a coder can ultimately achieve. Therefore, continued efforts are made to further improve psychoacoustic masking models [27].

4. SUBJECTIVE EVALUATION OF SOUND QUALITY

Because perceptual entropy coders exploit limitations inherent to the human auditory system, classical measures of sound quality such as signal/noise ratio or total harmonic distortion are inadequate and irrelevant. Therefore, the performance of coding schemes is typically evaluated with the use of human listeners.

If one only wants to evaluate whether a sound from a particular coder at a certain bitrate is perceptually transparent, the simplest and most relevant method is a forced-choice discrimination test. The underlying thought here is that, if listeners cannot aurally detect the difference between a sound signal from an accepted quality standard such as a CD and the same sound signal from a lower-bitrate coder, then the two signals are perceptually equivalent in all respects and therefore the lower-bitrate code is perceptually transparent. The fastest form of such a discrimination test is an adaptive 2-down I-up comparison method [28] in which each trial consists of a presentation of a reference and a coded signal in random order. Listeners have to respond on each trial what the presentation order was, and typically will receive response feedback. A run will typically start with a low bitrate for the coded signal, so that differences between reference and coded signals are not difficult to perceive. After two successive correct responses, the bitrate of the coded signal is increased by a certain increment, making the discrimination task more difficult. This procedure is continued until an incorrect response is given, whereupon the bitrate of the coded signal is decreased by one increment. Somewhere the adaptive trace will end up oscillating between increments and decrements, at which point the adaptive run is ended. The bitrate at which oscillation of the trace occurs is considered the perceptual discrimination threshold and corresponds to a 71% correct discrimination level if fixed-difference runs were used.

Although forced-choice discrimination methods are straightforward in classical psychophysics, their application in the present case is not without problems. Firstly, meaningful comparison of music fragments requires some minimum length of these fragments, which interferes with short-term memory and often causes listener fatigue. Secondly, coding artifacts have many different perceptual forms and can occur at random times, so that the subject is in constant uncertainty about *when* to listen, *what* to listen *for* and, in cases of spatial sound, *where* to listen. To make matters worse, discrimination cues will not only quantitatively change when the bitrate is changed, but may also change qualitatively. This makes adaptive discrimination runs difficult to perform, and often causes unstable non-converging traces and significant learning effects.

A more popular test method is the double-blind A-B-C hidden reference paradigm, standardized in ITU-R Recommendation BS. 1116 [29], which is used for the evaluation of both perceptually transparent and degraded codes. On each trial, expert listeners hear three identical music fragments A, B, and C, where A is always the high-quality reference and B and C are, in random order, the reference and the coded signal. Listeners must (a) identify which of B and C is the hidden reference, and (b) rate the difference between B and C on a 41-point scale. A variant on this method is ITU-T Recommendation P.800/P.830, which uses a 7-point scale and is intended for more heavily degraded signals. Incorrect identification of the hidden reference automatically assigns the highest rating ('perceptually equivalent') to the coded signal for that trial. Average scale values assigned to perceived differences provide an indication of the amount of perceptual degradation of coded signals. This method of rating is important for situations where the highest possible quality of audio is desired, but perceptual transparency cannot be achieved because of technical limitations.

5. OBJECTIVE MEASURES OF SOUND QUALITY

Given the many problems encountered with subjective quality testing, including the finding that 'expert' listeners tend to be systematically biased in their sensitivity to particular types of perceptual degradation [30], it is not surprising that several attempts have been made to develop objective measures for perceived sound quality [31–33]. A fundamental problem underlying all these attempts is the fact that, since perceptual entropy coding is based on imperfections of the human ear, an accurate and complete model of human hearing is required to build a reliable and objective evaluation algorithm. It is obvious that a complete model of human hearing does not exist and will not exist for some time to come. Moreover, there is always a hidden danger that an 'objective' evaluation algorithm adopts the same or a similar hearing model as was used in a particular coding algorithm, since all perceptual entropy coders use some kind of psychoacoustic model. If this is the case, the evaluation algorithm is likely to favor the coder that uses the same hearing model, and will therefore be everything but 'objective'.

Despite the inherent pitfalls and potential shortcomings of objective evaluation methods, an international standard for objective sound quality testing has been developed and agreed upon between 1998 and 2001 by the International Telecommunications Union as Recommendation ITU-R BS.1387 [34]. This standard contains a psychoacoustic model yielding quantitative measures for the internal representation of signal features, and a cognitive model for describing feature extraction and combination. It incorporates six earlier proposed perception and cognition models, specifically the Disturbance IndeX (DIX) model [35], the Noise-to-Masked Ratio (NMR) model [32], the Objective Audio Signal Evaluation (OASE) model [36], the Perceptual Audio Quality Measure (PAQM) model [37], the PERCeptual EVALuation (PERCEVAL) model [38], and the Perceptual Objective Measurement (POM) model [39]. The essence of the evaluation model is that it always compares a degraded signal with a high-quality reference, and estimates specific loudness differences in each critical band following a classical method outlined by Zwicker and Feldtkeller in 1967 [16]. These quantities are then weighted and combined, and ultimately yield a single quantitative measure representing the perceptual distance between the reference and the coded signal.

6. CONCLUDING REMARKS

One should always keep in mind that lossy, perceptually transparent codes are meant to be listened to only. They should never be used for documentation purposes in which sounds are recorded and stored for physical analysis at a later time. The fact that such codes are transparent to our ears does not imply that they are transparent to acoustical or signal processing analysis procedures.

The original waveform-based coding techniques basically allow only playback of a music signal as it was recorded. Modem model-based parametric techniques allow, besides considerable bit savings, a great deal of flexibility and control on the playback site [40,41].

ACKNOWLEDGEMENT

The author is indebted to Armin Kohlrausch, Steven van de Par and Jeroen Breebaart for helpful hints and suggestions during the preparation of this paper.

DISCLAIMER

Opinions, interpretations and conclusions contained in this letter are those of the author and are not necessarily endorsed by the U.S. Army and/or the Department of Defense.

REFERENCES

[1] Patterson, B., "Musical Dynamics", *Sci. Am.* **231**(5), 78-95, 1974.

[2] International Electrotechnical Commission IEC 908: "Compact Disc Digital Audio System", Geneva, Switzerland, 1987.

[3] Oohashi, T., Nishina, E., Kawai, N. and Fuwamoto, Y., "Hypersonic effect by high frequency sound above the audible range", *Proc. Int. Symp. on Mus. Acoust.,* Tokyo, 119-122, 1992.

[4] Nakamura, N, Toraichi, K. and Yamaura, L, "Audible tone on a human tympanic membrane evoked by ultrasonic components", *Proc. Int. Synip. on Mus. Acoust.,* Tokyo, 123-126, 1992.

[5] Nakamura, N., Toraichi, K., Kamada, M. and lwaki, M., "Contribution of ultrasound to timbre", *Proc. Int. Symp. on Mus. Acoust.,* Tokyo, 127-130, 1992.

[6] Sato, H., Yoshida, M. and Yasushi, M., "Evaluation of sensitivity of the frequency range higher than the audible frequency range", *Pioneer R&D,* 7, 10-16, 1997.

[7] Stuart, J.R., "A proposal for high-quality application of high-density CD carriers", *Stereophile,* August 1995.

[8] Ashiliara, K. and Kiryu, S., "Audibility of components above 22

kHz in a harmonic complex tone", *Acta Acustica,* **89**, 504-546, 2003.

[9] Stoll, G., Link, M. and Theile, G., "Masking-pattern adapted subband coding: Use of the dynamic bit-rate margin", *Proc. 84h Conv. Aud. Eng. Soc.,* 1988, preprint 2585.

[10] Brandenburg, K, "High-quality sound coding at 2.5 bits/sample", *Proc. 84th Conv. Aud. Eng. Soc.,* 1988, preprint 2582.

[11] Serra, X. and Smith, LO.III, "Spectral modeling and synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition", *Comp. Mus. J.* 12-24,1990.

[12] Edler, B. "Technical description of the MPEG-4 audio coding proposal from University of Hannover and Deutsche Bundespost Telekorn", ISO/IEC *MPEG96/M0632 1996.*

[13] Chowning, J. "The synthesis of complex audio spectra by means of frequency modulation", *J. Aud. Eng. Soc.* 526-529, 1973.

[14] Makhoul, J. "Linear prediction: a tutorial review", *Proc. JEEE,* **63,** 561-580,1975.

[15] Painter, T. and Spanias, A. "Perceptual coding of digital audio", *Proc. IEEE,* **88,** 451-513, 2000.

[16] Zwicker, E. and Feldtkeller, R. *Das Ohr als Nachrichtenempfanger.* Hirzel Verlag, Stuttgart, 1967.

[17] Elliott, L.L. "Backward and forward masking of probe tones of different frequencies", *J Acoust. Soc. Am.* **34**, 1116-1117, 1962.

[18] Houtsma, A.J.M., Trahiotis, C., Veldhuis, R.N.J. and van der Waal, R. "Bit rate reduction and binaural masking release in digital coding of stereo sound", *Acustica,* **82,** 908-909, 1996.

[19] Houtsma, A.J.M., Trahiotis, C., Veldhuis, R.N.J. and van der Waal, R. "Further bit rate reduction through binaural processing", *Acustica,* **82,** 909, 1996.

[20] Van de Par, S.L.J.D.E., ten Kate, W.R.T., Kohrausch, A. and Houtsma, A.J.M. "Bit-rate saving in multichannel sound: Using a band-limited channel to transmit the center signal", *J Aud. Eng. Soc.* **42,** 555-564, 1994.

[21] Faller, C. and Baunigarte, F. "Binaural cue coding: A novel and efficient representation of spatial sound", *Proc. ICASSP,* 1841-1844, 2002.

[22] Johnston, J. "Estimation of perceptual entropy using noise masking criteria", *Proc. ICASSP,* 2524-2527,1988.

[23] Houtsma, A.J.M. "Psychophysics and modern digital audio technology", *Philips J Res.,* **47,** 3-14, 1992.

[24] Sinha, D. and Johnston, J. "Audio compression at low bit rates using a signal adaptive switched filterbank", *Proc. ICASSP,* 1053-1056, 1996.

[25] Wiese, D. and Stoll, G. "Bitrate reduction of high quality audio signals by modeling the ear's masking thresholds", *Proc. 80 Conv. Aud. Eng. Soc.,* 1990, preprint 2970.

[26] "Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s- IS **11** 172-3 (audio)", 1SO/IEC, JTC1/SC29, 1992.

[27] Van de Par, S., Kohlrausch, A., Charestan, G. and Heusdens, R. "A new psychoacoustical masking model for audio coding applications", *Proc. ICASSP,* 11, 1805-1808, 2002.

[28] Levitt, H. "Transformed up-down methods in psychoacoustics", *J. Acoust. Soc. Am.* **49,** 467-477, 1971.

[29] "Methods for subjective assessment of small impairments in audio systems including multichannel sound systems", *ITU-R* Rec. BS 1116, 1994.

[30] Milne, A. "New test methods for digital audio data compression algorithms", *Proc. 11th Conf. Aud. Eng. Soc.,* 210-215, 1992.

[31] Karjaleinen, M. "A new auditory model for the evaluation of sound quality of audio systems", *Proc. ICASSP, 608-611,* 1985.

[32] Brandenburg, K. "Evaluation of quality for audio encoding at low bit rates", *Proc. 82nd Conv. Aud. Eng. Soc.,* 1987, preprint 2433.

[33] Beerends, J. and Stemerdink, J. "Measuring the quality of speech and music codecs: An integrated psychoacoustic approach", *Proc, 98th Conv. Aud. Eng. Soc.,* 1995, preprint 3945.

[34] "Method for objective measurements of perceived audio quality", *ITU-R* Rec. B S 1387, 2001.

[35] Thiede, T. and Kabot, E. "A new perceptual quality measure for bit rate reduced audio", *Proc. 100th Conv. Aud. Eng. Soc.,* 1996, preprint 4280.

[36] Sporer, T. "Objective signal evaluation - applied psychoacoustics for modeling the perceived quality of digital audio", *Proc. 103rd Conv. Aud. Eng. Soc.,* 1997, preprint 4512.

[37] Beerends, J. and Stemerdink, J. "A perceptual audio quality measure based on a psychoacoustic sound representation", *J Aud. Eng. Soc.,* **40,** 963-978, 1992.

[38] Paillard, B., Mabilleau, P., Morisette, S. and Sournagne, J. "Perceval: perceptual evaluation of the quality of audio signals", *J. Aud. Eng. Soc.,* 40, 21-31, 1992.

[39] Colomes, L., Lever, M., Rault, LB. and Dehery, Y.F. "A perceptual model applied to audio bit-rate reduction", *J Aud. Eng. Soc.,* **43,** 233-240, 1995.

[40] Levine, S. *Audio representations for data compression and compressed domain processing.* PhD Thesis, Dept. of EE, Stanford University, 1998 (http://www-crma.Stanford.edu/~scottl/thesis.html).

[41] Purnhagen, H. "Advances in parametric audio coding", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* 17-20, 1999.

[42] Rossing, T.D. *The Science of Sound,* Addison Wesley, Reading, MA, 1990.