ERROR ROBUST LOW DELAY AUDIO CODING USING SPHERICAL LOGARITHMIC QUANTIZATION

Stephan Preihs, Timm Lamprecht and Jörn Ostermann

Leibniz Universität Hannover, Institut für Informationsverarbeitung
Appelstr. 9A, 30167 Hannover, Germany, Correspondence: preihs@tnt.uni-hannover.de

ABSTRACT

This paper reveals the potential gain in audio quality that can be achieved by combining Spherical Logarithmic Quantization (SLQ) with advanced broadband error robust low delay audio coding based on ADPCM.
We briefly summarize the basic properties and mechanisms of SLQ and the employed ADPCM scheme and show how they can be combined in a freely parameterizable coding algorithm. The resulting codec includes techniques for error robustness and a shaping of the coding noise. We present results of optimizing the codec parameters in our framework for global optimization based on psychoacoustic measures.
Our evaluation shows that by using SLQ instead of scalar quantization a PEAQ ODG-score improvement with a maximum of about 1 point and a mean of 0.2 can be achieved. An analysis of the bit-error behavior of the combined SLQ-ADPCM shows that a major improvement in bit-error performance results from our proposed efficient error detection and processing.

***Index Terms*—** Low delay audio coding, Spherical Logarithmic Quantization, ADPCM, global optimization, PEAQ

1. INTRODUCTION

When real time audio transmission systems have to deal with limited data rates, a demand for low bit rate low delay audio coding arises. If, in addition, a wireless audio transmission is supposed to take place in a live scenario with a high number of simultaneous channels, the constraints on delay and bandwidth become even more stringent. Furthermore, there are some applications that take place at the beginning of an "audio production chain" and therefore require a near transparent audio quality. This all together leads to a demand for low delay source coding algorithms with delays less than $1\,\mathrm{ms}$ and a very high audio quality.
Unfortunately the well known and established low delay codecs like AAC-ELD [1] and FhG ULD [2] or even the CELT (Opus) codec [3, 4] and the new 3GPP EVS codec [5] are usually unsuitable for application in these scenarios since they either lead to a higher delay or do not provide an adequate audio quality.
Therefore in the past we [6–8] and others [9, 10] did extensive research on how to modify the algorithms presented in [11] regarding error robustness while maintaining the near transparent audio quality. Although especially with [10] a major improvement regarding error resilience was reported the audio quality still suffers from the modifications needed for ensuring error robustness.
In this work we therefore target at improving the audio quality by replacing the scalar quantizer of the ADPCM scheme by SLQ. The basic ideas and theoretical work behind SLQ go back to publications like [12] but in [13] for the first time algorithms for a practical implementation are given, a comprehensive evaluation of the SNR gains is done and the combination with DPCM is investigated. Nevertheless Matschkal et al. claim to only do a proof of concept for the combination of SLQ and ADPCM, do not use advanced prediction schemes, an evaluation with broadband audio and regarding psychoacoustical quality or a global optimization of the parameters involved in the coding process.
Therefore within this paper we evaluate the potential gain in audio quality that can be achieved by combining SLQ with advanced broadband error robust low delay audio coding while deliberately accepting computational overheads.

2. SPHERICAL LOGARITHMIC QUANTIZATION

SLQ is a special kind of vector quantization that is based on representing a vector $\mathbf{x}$ formed by sequential data in spherical coordinates and searching for a quantization cell on a D-dimensional sphere while quantizing the corresponding radius with respect to a logarithmically spaced codebook.
While the basic concept of SLQ is rather intuitive and straightforward, the detailed algorithms for equal distribution of quantization cells on the D-dimensional sphere and the index search have an underlying mathematical formulation that goes beyond the scope of this paper. Therefore in this section we only give a brief introduction of the SLQ algorithms and introduce the terminology and parameters needed for being able to understand the following sections. For more details and formulas as well as a discussion about other methods for vector quantization and their relation to SLQ the reader is referred to [13] and the references given there.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

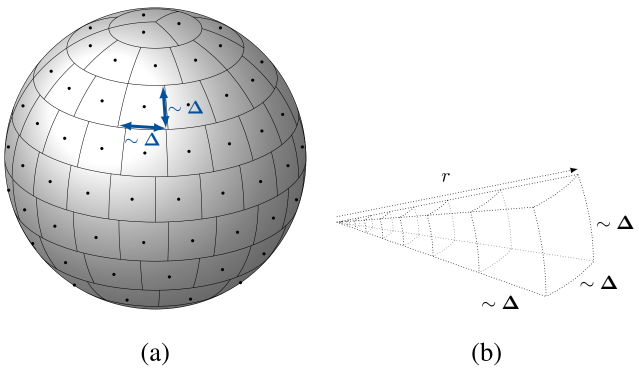

Fig. 1. (a) Distribution of cells and reconstruction vectors (•) on a 3-dimensional sphere for $R = 4$ bits/sample. (b) Logarithmic quantization of radius r for equal length of cube edges $\boldsymbol{\Delta}$. (both based on [13]).

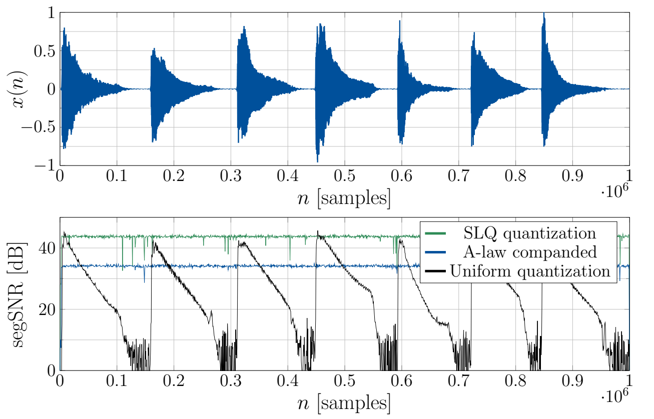

Fig. 2. Comparison of segmental SNRs of different quantization types with $R = 8$ bits/sample for a part (min. 01:02-01:23) of the normalized SQAM Track 39.

Figure 1 shows the basic concept of SLQ with the example of a three-dimensional sphere and a codeword length of four bits/sample. Part (a) shows the whole sphere with the distribution of cells on the surface and the reconstruction vectors in the middle of each quantization cell. Part (b) shows the logarithmic quantization of the radius r and how it enables for a nearly equal length of cube edges Δ.

The algorithms of Matschkal et al. provide rules and formulas for the computation of an optimal distribution of bits between the radius r and the angles φ which allows for evenly distributed quantization cells of cube edge length Δ on the surface of the sphere. For a plain SLQ of a given vector $\mathbf{x}$ it is therefore only necessary to do a transformation into spherical coordinates, search for the quantization cell the input vector is lying in and transmit a combined index N from the angular index N_φ and the radius index N_r. In general it is possible to do this with an infinite number of Dimensions D. In practice and with the precomputed surface integrals provided by Matschkal et al. a dimensionality up to 12 is manageable.

The reason why in this work we employ SLQ for an enhancement of ADPCM is that it provides a significant gain in SNR if compared to uniform scalar or logarithmic scalar quantization. Figure 2 shows a comparison of the segmental SNRs for a normalized part of the SQAM [14] track 39. The A-law companding is done with $A = 2229.7$ and the same A is used for the logarithmic quantization of the radius in the SLQ of dimensionality $D = 6$. The comparison shows that SLQ is able to provide a constant SNR over a wide amplitude range of the input signal that is close to the maximum in SNR reached by uniform quantization. The most important parameters of the plain SLQ (the first two will also be part of our parameter optimization in Section 5) are listed below:

- D: Vector length $\rightarrow$ number of dimensions used
- A: A-law parameter used for radius quantization
- R: Resulting bits/sample used for quantization

3. ERROR ROBUST LOW DELAY AUDIO CODING

The base ADPCM codec employed in this work is the one presented in [9] and includes a static noise shaping filter in the encoder. Figure 3 shows the flow graph of the encoder (here with scalar quantization that is later replaced by SLQ). The base ADPCM can be divided in the functional blocks prediction, envelope estimation, quantization and noise shaping which are briefly described in the following. The lattice prediction filter also is the one from [9] including leakage factors α and β for transmission error robustness.

The coefficients k_m of the lattice prediction filter are adapted with a modified version of the GAL algorithm [9] which we refer to as "leaky GAL" and is formulated as

$$k_m(n+1) = k_m(n) + \frac{\tilde{\mu}}{\sigma_m^2(n) + \sigma_{min}^2} \cdot \\ (f_m(n) \cdot \alpha\beta b_{m-1}(n-1) + b_m(n) \cdot f_{m-1}(n)) \quad (1)$$

where

$$\sigma_m^2(n) = (1-\tilde{\mu})\sigma_m^2(n-1) + \tilde{\mu}(f_{m-1}^2(n) + b_{m-1}^2(n-1)). \quad (2)$$

The envelope estimation is done with a first order recursive filter

$$v^2(n) = (1-\lambda) \cdot v^{2\gamma}(n-1) + \lambda \cdot \tilde{e}^2(n-1) \quad \text{where} \quad (3)$$

$$\lambda = \begin{cases} \lambda_{\text{AT}} & \text{if } \tilde{e}^2(n-1) > v^2(n-1) \\ \lambda_{\text{RT}} & \text{else} \end{cases}$$

with the two signal adaptive attack- and release-constants $\lambda_{\text{AT}}, \lambda_{\text{RT}}$ (where $\lambda_{\text{AT}} > \lambda_{\text{RT}}$) and a modification by introducing a leakage factor γ for ensuring recovery after transmission errors. Additionally $v(n)$ has a lower limit of v_{min} for preventing a normalization and therefore division by zero.

Furthermore at the encoder the reconstruction error $e_{\text{ADPCM}}(n)$

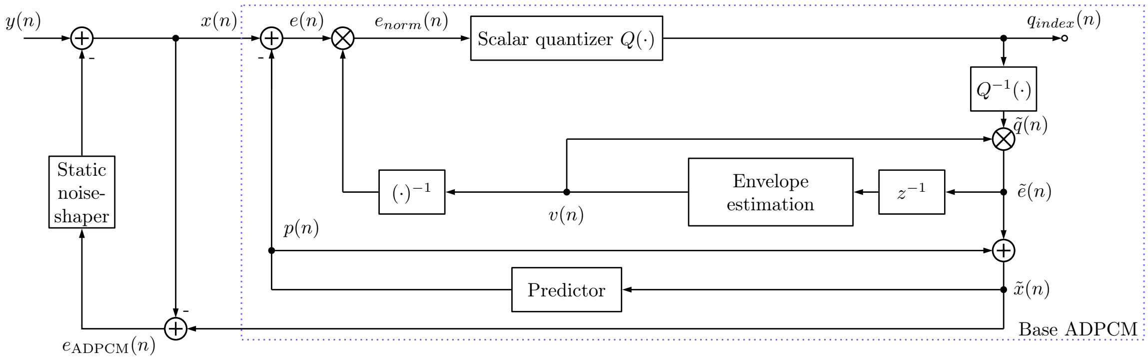

Fig. 3. Flow graph of the base ADPCM codec (dotted blue box, here with scalar quantization) with static noise shaping.

$= \tilde{x}(n) - x(n)$ is filtered by a fourth order static noise shaping filter [11] and fed to the input of the ADPCM scheme.
Although reconstruction signal driven leaky GAL does not guarantee a bit exact synchronization of the reconstruction signal between encoder and decoder we choose this as baseline codec since it is well studied and documented and therefore allows for a fair comparison with the state of the art.

4. COMBINATION OF ADPCM AND SLQ

The combination of vector quantization and ADPCM-based low delay audio coding can be realized in several ways depending on the kind of ADPCM scheme that is used. If subband coding is employed, the input vector for the SLQ can be formed directly from the subband signals. Since in this work we are mainly interested in the influence of SLQ on the audio quality of advanced prediction schemes without bias of a potential subband gain, we stick with the broadband ADPCM.
Combining a sample based broadband ADPCM processing and a vector quantization seems not possible at a first glance since the update of the predictor relies on the reconstruction signal $\tilde{x}(n)$ and this can only be computed from the quantized prediction error, which in general, is only available if a full block of samples has been processed by the SLQ. For solving this conflict Matschkal et al. propose an algorithm that iteratively searches for the best radius quantization with the lowest reconstruction error and incorporates a successive index search.
Algorithm 1 summarizes the procedure of the combined SLQ and ADPCM processing in the encoder. Here $\mathbf{x}$ is a vector of the normalized prediction error signal $e_{norm}(n)$ since we use the prediction error normalization employed for scalar quantization also for the combined SLQ-ADPCM. The algorithm starts with an estimation of the radius r. For this the ADPCM-state for the last sample of the preceding block is used and the ADPCM coding is done for the current block of D samples without using any quantization. Then this estimate r is employed for a search for the best quantized r named $\hat{r}$ by doing a quantization of the radius within certain bounds and using the different radii for a SLQ-ADPCM processing of the whole block. The block with the lowest Euclidean distance in the reconstructed normalized prediction error is used for computing the final SLQ index that is transmitted to the decoder.

Algorithm 1 SLQ-ADPCM encoder

Estimate r from processing vector $\mathbf{x}$ without quantization
for all $\tilde{r}$ within $[r + \Delta, r - \Delta]$ {
  Quantize radius $\tilde{r} \rightarrow \hat{r}$
  Do SLQ-ADPCM processing $\rightarrow \hat{\mathbf{x}}$
  Compute Euclidean distance $d^2(\mathbf{x}, \hat{\mathbf{x}})$
}
Search for r with lowest $d^2(\mathbf{x}, \hat{\mathbf{x}})$
Compute SLQ index N from N_r and N_φ and transmit

The number of iterations that is used for finding the best $\hat{r}$ is named η and will be part of our global parameter optimization presented in Section 5. Of course this way of repeated SLQ-ADPCM processing in the encoder with a number of iterations η in the range of $10 < \eta < 100$ is computationally intensive but at the same time allows for a heavily parallelized calculation.
Due to the nature of the ADPCM of not introducing any delay to the signal, the only algorithmic delay resulting from the SLQ-ADPCM is coming from the pooling of samples in blocks of size D. Nevertheless in a real world system with a sequential transmission of coded SLQ indices and a limited processing power another D to $2D$ samples have to be considered. With D being 12 at maximum in our implementation, the resulting overall delay still clearly is below 1 ms.

Enhanced error robustness

By combining broadband ADPCM with SLQ another problem arises if it comes to erroneous transmission between encoder and decoder. While with scalar quantization only one sample of $\tilde{q}(n)$ is affected from a single bit-error in case of SLQ the whole block of D samples is corrupt. Without any further procedure for error treatment therefore the SLQ-ADPCM produces heavily distorted audio for common bit-error rates but still recovers after periods of erroneous transmission.

In consequence we looked at efficient ways of improving the bit-error performance of the SLQ-ADPCM without having to increase the data rate. As a first approach lowering the codeword length to $D \cdot R - 1$ bits and simply adding a parity bit at the end of each block has proven to drastically improve the performance. In the case of a detection of a flipped bit zeros are inserted as reconstructed prediction error values. Of course the bit-error performance can be further enhanced by the use of techniques like the ones presented in [10].

5. GLOBAL OPTIMIZATION AND EVALUATION

The SLQ as well as the ADPCM coding involve several partially interacting parameters that are not amenable to manual tuning. Therefore we used our PEAQ-based (Perceptual Evaluation of Audio Quality [15]) framework [16] which we already employed in [7] and [8] for global optimization.

The core of this framework is a genetic algorithm that controls the parallel encoding and decoding of test set items and uses a PEAQ-based evaluation of the achieved audio quality for a given parameter vector χ. Like proposed in [16] the PEAQ-results ODG_i (Objective Difference Grade) for a given χ are mapped to the cost function value $C(\chi)$ by computing

$$C(\chi) = \sum_{i=1}^{N} (\mathrm{ODG}_i(\chi))^4 \qquad (4)$$

where N is the number of test set items. This puts a stronger emphasis on signals with a worse audio quality without completely ignoring the results of better ones. We use the PEAQ C-code provided with [17] for calculation speedup.

Due to the computational overhead caused by the combined SLQ-ADPCM processing only a limited number of optimization runs with different starting points, borders and general settings were possible. The results presented in the next section therefore still offer possibilities for further improvement.

The parameters we optimize are (with $\alpha = 0.98, \beta = 0.91$):

- D, η and A of the SLQ
- λ_{AT}, λ_{RT}, v_{min} and $\gamma < 1$ of the envelope estimation
- $\tilde{\mu}$ and σ_{min} of the GAL algorithm

and a constant factor that is used for scaling the normalized prediction error before SLQ and therefore finding the tradeoff between overloading and an insufficient level.

In general the combined SLQ-ADPCM with 3 bits/sample shows a significant improvement of more than one point in the PEAQ-score for synthetic signals like the SQAM-tracks 1–7. This is mainly the consequence of a higher segmental SNR resulting from the SLQ.

For comparison to the state of the art we used the upsampled versions of the left channel from the SQAM-tracks for the coding and PEAQ-evaluation at 48 kHz for having the lowest possible bias of the pre- or postprocessing to the evaluation results. The length was adjusted to the borders proposed in [10] for reduction of optimization runtime and in order to make the comparison as fair as possible.

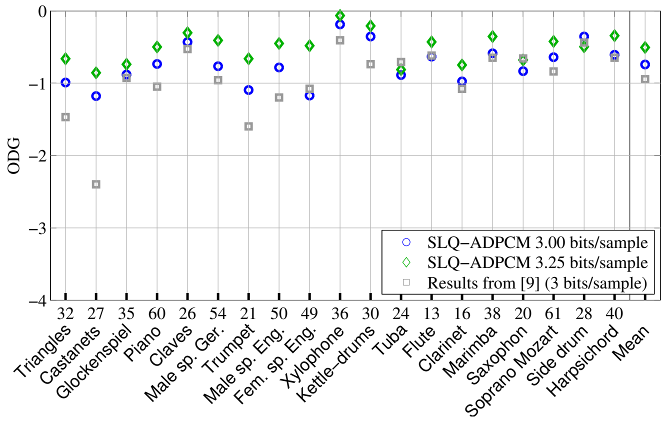

Fig. 4. PEAQ ODG-scores of the optimized SLQ-ADPCM and results from [10] (no overhead) for the error free case.

Figure 4 shows a comparison of ODG-values for natural audio. Especially for the critical signals “Castanets”, “Triangles” and “Trumpet” a significant improvement in the PEAQ ODG can be achieved by the SLQ-ADPCM. Although for some signals a small degradation of audio quality arises the mean PEAQ ODG-improvement is of about 0.2 points.

Figure 5 shows the evaluation of the bit-error behavior of the SLQ-ADPCM as the mean over 1000 realizations of a random single bit-error pattern with a bit-error rate of 10^{-4} in comparison to the results presented in [10]. The colored bars between the PEAQ ODG-values point out the improvement or degradation with respect to the reference. The results show that with our efficient error detection and processing scheme without additional overhead a comparable bit-error performance to a broadband ADPCM with scalar quantization can be achieved.

One big advantage of SLQ and vector quantization in general is that it allows for a rational number of bits/sample. Therefore we also did an optimization with a slightly higher bitrate resulting from 3.25 bits/sample. The results are also shown in Figure 4 and demonstrate that only a small increase in the datarate allows the SLQ-ADPCM to become near transparent in audio quality since all of the results are clearly above the $\mathrm{ODG} = -1$ line.

The results presented were validated by informal listening experiments and additional objective measurements like segmental SNRs. Of course they should be verified by a formal listening test like [18] in future studies.

6. CONCLUSION

In this paper we investigate a combination of advanced broadband error robust ADPCM and Spherical Logarithmic Quantization and show how they can be integrated in a freely parameterizable low delay audio codec. Together with our global parameter optimization it allows for an improved audio quality with a PEAQ ODG-score improvement of 1 point at max-

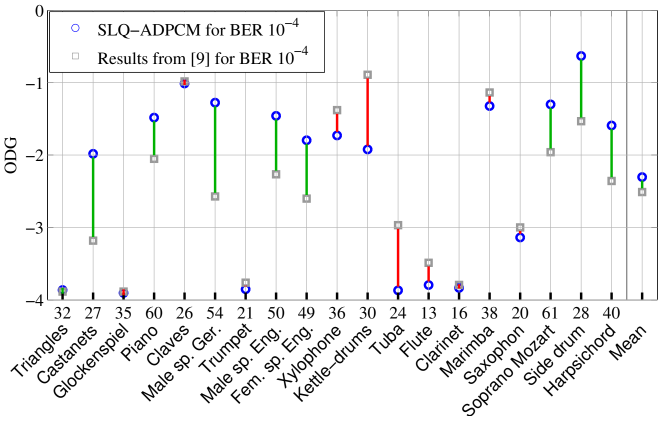

Fig. 5. Comparison of the bit-error performance of SLQ-ADPCM at BER $= 10^{-4}$ with results from [10] (no overhead).

imum and 0.2 in the mean if compared to state of the art algorithms employing scalar quantization. In addition with the introduced efficient error detection and processing scheme by parity bit addition and zero codeword insertion, a comparable bit-error performance to broadband ADPCM using scalar quantization can be achieved.

While the combined SLQ-ADPCM coding allows for an improved audio quality it results in a tremendous computational overhead. Therefore future work should target at improving the algorithm for radius search by reducing the number of iterations needed for finding a sufficient radius.

REFERENCES

[1] M. Schnell, M. Schmidt, M. Jander, T. Albert, R. Geiger, V. Ruoppila, P. Ekstrand, M. Lutzky, and B. Grill, "MPEG-4 Enhanced Low Delay AAC - A New Standard for High Quality Communication," Oct. 2008, 125th AES Convention.

[2] U. Krämer, G. Schuller, S. Klier Wabnik, and J. Hirschfeld, "Ultra Low Delay Audio Coding with Constant Bit Rate," Oct. 2004, 117th AES Convention.

[3] J.-M. Valin, T.B. Terriberry, C. Montgomery, and G. Maxwell, "A High-Quality Speech and Audio Codec With Less Than 10-ms Delay," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 1, pp. 58–67, 2010.

[4] J. M. Valin, G. Maxwell, T. B. Terriberry, and K. Vos, "High-Quality, Low-Delay Music Coding in the Opus Codec," Oct. 2013, 135th AES Convention.

[5] G. Fuchs, C.R. Helmrich, G. Markovic, M. Neusinger, E. Ravelli, and T. Moriya, "Low Delay LPC and MDCT-based Audio Coding in the EVS Codec," Apr. 2015, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[6] S. Preihs and J. Ostermann, "Error Robust Low Delay Audio Coding based on Subband ADPCM," Oct. 2011, 131st AES Convention.

[7] S. Preihs and J. Ostermann, "Globally Optimized Dynamic Bit-Allocation Strategy for Subband ADPCM-Based Low Delay Audio Coding," Apr. 2015, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[8] S. Preihs, C. Wacker, and J. Ostermann, "Adaptive Pre- and Post-Filtering for a Subband ADPCM-based Low Delay Audio Codec ," Oct. 2015, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).

[9] G. Simkus, M. Holters, and U. Zolzer, "Error Robust Delay-free Lossy Audio Coding Based on ADPCM," Sept. 2013, 16th Int. Conference on Digital Audio Effects (DAFx).

[10] G. Simkus, M. Holters, and U. Zolzer, "Error Resilience Enhancement for a Robust ADPCM Audio Coding Scheme," May 2014, pp. 3685–3689, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[11] M. Holters, R. Helmrich, and U. Zölzer, "Delay-free Audio Coding based on ADPCM and Error Feedback," Sept. 2008, 11th Int. Conference on Digital Audio Effects (DAFx).

[12] P.F. Swaszek, "Uniform Spherical Coordinate Quantization of Spherically Symmetric Sources," *IEEE Transactions on Communications*, vol. 33, no. 6, pp. 518–521, Jun. 1985.

[13] B. Matschkal and J. B. Huber, "Spherical Logarithmic Quantization," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 1, pp. 126–140, Jan. 2010.

[14] EBU Tech. 3253-E, "Sound Quality Assessment Material. Recordings for Subjective Tests," Apr. 1988.

[15] ITU-R Recommendation BS.1287-1, "Method for Objective Measurements of Perceived Audio Quality," Nov. 2001.

[16] M. Holters and U. Zölzer, "Automatic Parameter Optimization for a Perceptual Audio Codec," Apr. 2009, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[17] P. Kabal, "An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality," Dec. 2003, Department of Electrical & Computer Engineering, McCill University, Montreal.

[18] ITU-R Recommendation BS.1534-1, "Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems," Jan. 2003.