INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Entropy Coding of Spectral Envelopes for Speech And Audio Coding Using Distribution Quantization

Srikanth Korse[1], *Tobias Jähnel*[2], *Tom Bäckström*[1,2]

[1]Fraunhofer IIS, Erlangen, Germany
[2]International Audio Laboratories, Friedrich-Alexander University (FAU), Erlangen, Germany
srikanth.korse@iis.fraunhofer.de

Abstract

Speech and audio codecs model the overall shape of the signal spectrum using envelope models. In speech coding the predominant approach is linear predictive coding, which offers high coding efficiency at the cost of computational complexity and a rigid systems design. Audio codecs are usually based on scale factor bands, whose calculation and coding is simple, but whose coding efficiency is lower than that of linear prediction. In the current work we propose an entropy coding approach for scale factor bands, with the objective of reaching the same coding efficiency as linear prediction, but simultaneously retaining a low computational complexity. The proposed method is based on quantizing the distribution of spectral mass using beta-distributions. Our experiments show that the perceptual quality achieved with the proposed method is similar to that of linear predictive models with the same bit rate, while the design simultaneously allows variable bit-rate coding and can easily be scaled to different sampling rates. The algorithmic complexity of the proposed method is less than one third of traditional multi-stage vector quantization of linear predictive envelopes.

Index Terms: spectral envelope, arithmetic coding, speech and audio coding

1. Introduction

Modern speech and audio codecs, such as 3GPP Enhanced Voice Services (EVS), MPEG-D Unified Speech and Audio Coding (USAC), ITU-T G.718 and Extended Adaptive Multi-Rate – Wideband (AMR-WB+) are generally hybrid codecs featuring separate coding modes for speech signals and generic audio signals [1–4]. While this separation into two modes has well-motivated origins, the trend is to unify coding tools to improve scalability and flexibility of the overall design. Speech codecs have traditionally excelled at low bit-rates such as 13 kb/s, but with increasing bit-rates their computational complexity usually grows exponentially, and they are applicable only on narrow- and wideband sampling rates, preventing flexible and efficient application. Audio codecs on the other hand are scalable in bit- and sampling-rate, but since their coding efficiency has been lower, their prime coding modes are usually above 64 kb/s. Our goal is to develop scalable coding methods which bridge over the intermediate bit-rates 20 kb/s to 60 kb/s.

A central task in development of scalable codecs are spectral envelope models. Speech codecs are traditionally based on linear predictive coding, which is essentially a polynomial model of the spectral envelope [5,6]. There have been many improvements to the basic linear predictive model, such as [7–9], but the basic approach for coding the parameters of linear predictive models, using vector codebooks and line spectral frequencies, has remained fairly unchanged [10, 11]. This approach gives a relatively high coding efficiency, but finding the line spectral frequencies and vector codebook searches are computationally expensive operations, whereby they scale poorly to higher bandwidths. Moreover, vector codebooks are generally designed for a fixed bit-rate, whereby scaling across bit-rates is also cumbersome. The latter problem can though be resolved by using beta- or Dirichlet-distributions to encode the line spectra instead of vector codebooks [12, 13].

Classic audio codecs do not explicitly encode the spectral envelope, but only the perceptual envelope [14, 15], though newer codecs can do both [4, 16]. The traditional approach for encoding envelope shapes in audio codecs is to model envelope magnitude in piece-wise constant steps, known as scale factor bands. These factors are then differentially encoded with an entropy coder. While this approach is simple to implement and algorithmic complexity is low, the coding efficiency has been clearly lower than for linear predictive envelopes.

In the current context, the most important application of envelope models is spectral envelope modeling. However, envelope models are also used in other areas of speech and audio coding. For example, temporal magnitude envelopes are often encoded with linear predictive models [17] and the results of the current study can be readily applied also there. Likewise, spatial audio objects and scenes can also be described using scale factors [18].

In this work we propose an entropy coder for scale factor bands based on modeling the distribution of spectral mass. We have already shown that such envelopes can be efficiently quantized by energy or magnitude distribution [19]. The central contribution of the current work is to apply beta-distributions in entropy coding of the scale factors. To be able to compare the scale factor band -representation with linear prediction, we further apply spline-smoothing on the scale factors. Since we can thus avoid the computationally complex computation of line spectral frequencies, our implementation is simpler than coders using linear prediction [13]. Moreover, since our model is based on a continuous, parametric probability model, it can be easily adopted to any accuracy, whereby the method is scalable and applicable to any speech and audio codec.

2. Distribution Quantization

Our objective is to develop an entropy coding method for envelopes, which explains as much as possible of the overall shape of the signal. Furthermore, to keep the algorithmic complexity of the method low, we want the parameters of the model to be independent and orthogonal to each other to the greatest possible extent. Orthogonal parameters have the useful property that they can be quantized and encoded independently, whereby computationally complex vector quantization methods can be

http://dx.doi.org/10.21437/Interspeech.2016-55

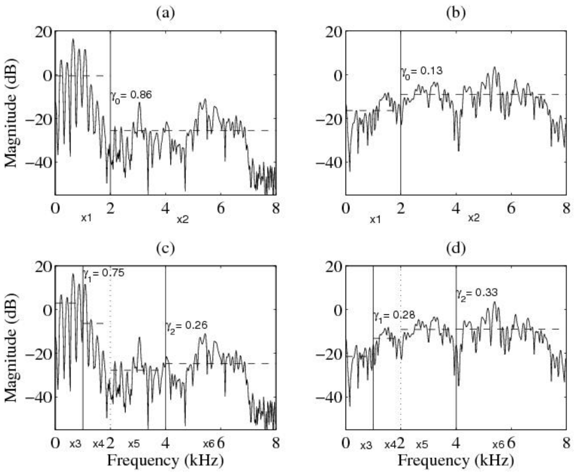

Figure 1: The p-norm ratios ($p = 1$) for the spectra of (a and c) voiced and (b and d) unvoiced speech signals, on the two first levels of the recursion. The first split point is here at 2 kHz and the second-level split points are at 1 kHz and 4 kHz, as indicated by the vertical lines. The mean level of each band is indicated with a dashed horizontal line.

avoided [20].

An envelope is a smooth shape, like an oversampled signal. At the highest level of oversampling, only the tilt of the envelope remains. It is therefore natural to use the tilt of the envelope as the highest level descriptor. To measure that tilt, we have chosen to use the ratio of p-norms of a signal x as

$$\gamma_0 = \frac{\|x_1\|_p^p}{\|x_1\|_p^p + \|x_2\|_p^p}, \qquad \text{where} \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad (1)$$

whereby $\gamma_0 \in [0, 1]$. That is, we split the $N \times 1$ vector x into two parts, x_1 and x_2, each of length N_1 and $N_2 = N - N_1$, respectively.

The p-norm ratio applied on speech signals is illustrated in Figure 1(a) and (b). We can see that a signal dominated by low-frequency components have a p-norm ratio close to one $\gamma \to 1$, while high-frequency signals have a low value $\gamma \to 0$.

We can then recursively split each sub-vector x_k again into sub-vectors x_{2k+1} and x_{2k+2} and determine their corresponding p-norm ratio as

$$\gamma_k = \frac{\|x_{2k+1}\|_p^p}{\|x_{2k+1}\|_p^p + \|x_{2k+2}\|_p^p}. \qquad (2)$$

The p-norm ratio γ_1 and γ_2 for speech signals are illustrated in Fig. 1(c) and (d). The recursion can then be continued to the desired depth.

Conversely, given the sequence of p-norm ratios γ_k we can determine the p-norm of each sub-vector $\|x_k\|_p^p$ by solving Equation 2. For example, we can for $k = 3$ solve $\|x_8\|_p^p = \|x\|_p^p \gamma_0 \gamma_1 (1 - \gamma_3)$, whereby we see that the norm of each subsegment depends on the norm of the original vector and a product of the ratios γ_k. In other words, the whole signal is split up into segments, whose p-norm is known. This representation is thus equivalent with the scale factor representation used in audio coders. The final scale factor representation of a speech signal is illustrated in Figure 2.

Finally, if we quantize the γ_k's to $\widehat{\gamma}_k$ such that $\gamma_k \approx \widehat{\gamma}_k$, then for the quantization we have, for example, $\widehat{\|x_8\|_p^p} = \|x\|_p^p \widehat{\gamma}_0 \widehat{\gamma}_1 (1 - \widehat{\gamma}_3)$. It is immediately clear that γ_0 has to be quantized by the highest accuracy, since its accuracy has an impact on all scale factors. The quantization of γ_1 and γ_2 then have impact on only half the spectrum, whereby its accuracy can be half that of γ_0. We can thus have the accuracy of each γ_k depend on the length N_k of the corresponding vector x_k. In the current work we have chosen to use a quantization accuracy $\Delta\gamma_k$ which is proportional to the inverse of the vector length $\Delta\gamma_k = \epsilon N_k^{-1}$, where ϵ is a scalar constant which determines the overall quantization accuracy for the whole frame. The quantization accuracy can be further improved by applying weighting on $\Delta\gamma_k$ based on perceptual criteria, for example such that the accuracy of scale factors are relative to the ERB- or Bark-scale.

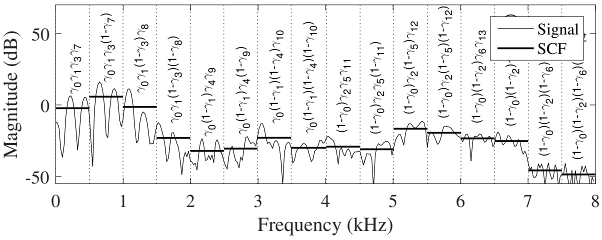

Figure 2: Illustration of scale factors used to model the spectral envelope. The relative level of each scale factor band can be computed recursively from Eq. 2, whose solution gives the product indicated in each scale factor band.

3. Coding of p-Norm Ratios

The parameters γ_k give a unique description of the spectral envelope. Moreover, due to the recursive structure, each γ_k is more or less independent of the other γ_k's. This property can be observed in Figure 1 as follows; First, note that x_1 and x_2 are non-overlapping, whereby γ_1 and γ_2 are independent. Furthermore, since each γ_k is the *normalized* energy of the left subvector, it follows that it is independent of the p-norm of its parent. The p-norm ratios γ_k are thus uncorrelated with each other and we can quantize and encode their values independently.

To encode γ_k's efficiently, we next have to determine their probability distribution. If we assume that x follows the multivariate normal distribution, then $\|x\|_2^2$ will follow the Chi-squared distribution. Likewise, if x is Laplacian, then $\|x\|_1^1$ follows the Chi-squared distribution. Moreover, it is well-known that the ratio $\frac{A}{A+B}$ follows the beta-distribution if the two variables A and B follow the Chi-squared distribution [21], whereby γ_k's will also follow the beta-distribution.

Plotting the histogram of γ_0 in Figure 3 over a speech signal however reveals that the distribution has several peaks. Informal experiments have shown that the two main peaks correspond to voiced and unvoiced phonemes. Moreover, transitions and complex phonemes such as voiced fricatives can have a γ_0 which lies in between the two peaks. Our interpretation of this result is that each peak corresponds to a unique source, whereby each source can be modelled with a beta distribution. We have therefore chosen to use a beta mixture model to represent the probability distributions of γ_k where each beta distribution in the mixture model represents a class of phonemes i.e. voiced or unvoiced. The parameters of such models can be estimated using the methods in [12, 13].

The remaining step is to choose how to best split each vec-

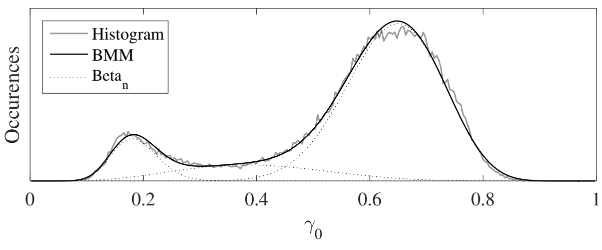

Figure 3: Histogram of γ_0 over the NTT-AT database, as well as beta mixture model (BMM) fitted to the data, where we used a mixture of three beta probability distributions. The individual beta pdf's are indicated by dotted lines.

tor x_k into its sub-parts x_{2k+1} and x_{2k+2}, that is, how to determine the length N_{2k+1} of vector x_{2k+1}. The objective is to choose this splitting points such that the corresponding p-norm ratio γ_k then describes a maximal amount of the signal variance. Conversely, the best split is that where γ_k reaches highest variance. Specifically, we will determine the variance of γ_k for all possible split points $N_{2k+1} \in (0, N_k)$ and choose the highest-variance $N^*_{2k+1} = \arg\max_{N_{2k+1}} \mathrm{var}[\gamma_k]$ as the best split point. Starting from γ_0 we can then recursively determine all split points N_k.

Using the beta mixture models, we can then use arithmetic coding to encode each parameter γ_k, to obtain near-optimal coding efficiency of the envelope shape [22].

4. Experiments

To test the proposed entropy coding method of spectral envelopes, we used the CELP-portion of a proprietary implementation of the MPEG-H standard [23]. The objective is to compare entropy coding of distribution quantization to conventional linear predictive coding with vector quantization as used in MPEG-H. To estimate the envelope parameters, we therefore applied the same windowing function on the input signal as is used when estimating the conventional predictor parameters.

For our experiments, we used the NTT-AT database, which consists of 3941 male and female speech samples in American and British English, German, Chinese, French and Japanese, each of 2 s length [24]. The signals were resampled to 16 kHz and down-mixed to mono. The window length was 32 ms with 50% overlap. The split points were chosen to maximize the variance of the p-norm as described in section 2. To calculate the variance, we used all sentences in the database except the test set (described below). The value of 0.5 was chosen as the p value in the Equation 2.

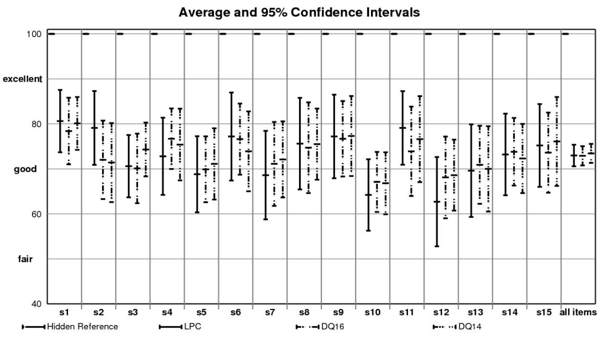

Figure 4: Average absolute MUSHRA scores for 15 speech items with 10 listeners using 95% confidence intervals of Student's t-distribution.

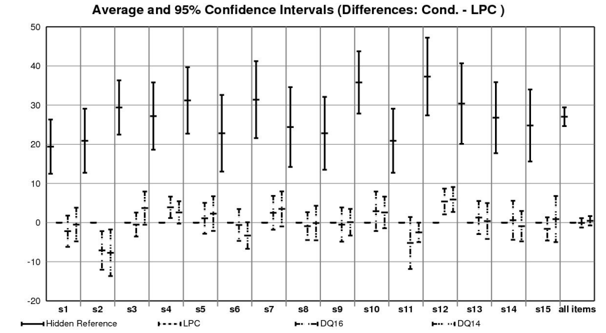

Figure 5: Difference MUSHRA scores for 15 speech items with 10 listeners using 95% confidence intervals of Student's t-distribution. The difference is computed with the LPC.

To use CELP, the spectral envelope needs to be converted to a linear predictive filter. For that purpose, we created a smooth envelope by spline interpolation using the same configuration as in [19]. We then applied the inverse Fourier transform on the power spectrum of the envelope to obtain an estimate of the autocorrelation, and calculated an order $m = 64$ predictor from that using Levinson-Durbin recursion [5, 25]. The predictor is naturally only approximating the estimated envelope, whereby a comparison with the linear prediction can unfairly penalize the proposed method. This configuration was however still retained, since it was the only one which we know of, which allows direct comparison of the envelope coding methods. Since the approximation can only degrade the quality of proposed method, it can not cause a false positive, where we erroneously would conclude that the proposed method is better than conventional linear prediction. To encode the p-norm ratios using arithmetic coding, a simple beta probability distribution was used.

The testing phase involved 15 files also from NTT database [24] which were not included in the training phase. These files were encoded and decoded using the MPEG-H encoder and decoder at a bit-rate of 24 kb/s. To evaluate the proposed method for entropy coding, we measured the average bit consumption, conducted a subjective listening test using MUSHRA [26] and performed an objective analysis of perceptual signal to noise ratio. Since we are comparing only the envelope models, all the three variants were constrained to use the same CELP based core coder. The initial quantization level (i.e. the quantization level with which the p-norm of the first split point γ_0 is quantized) for both variants of distribution quantization were set to 100, which coincides with the number of bins used in the LSF search [27].

The subjective listening test using MUSHRA consisted of 10 expert listeners and 15 speech items. The expert listeners were asked to evaluate four conditions: hidden reference, conventional linear prediction (LPC), distribution quantization with 16 split points (DQ16) and distribution quantization with 14 split points (DQ14). The results of the subjective listening test were analyzed using Student's t-distribution with 95% confidence intervals.

Figure 4 shows the average absolute MUSHRA scores. The absolute scores for all the three conditions (LPC, DQ16 and DQ14) lie in the "good" and "excellent" range except for speech items s12, s10 and s7, for which the absolute scores for the condition LPC lie in the "fair" and "good" range. Averaged over all speech items, there is no significant statistical difference between the three conditions that were compared. The difference MUSHRA scores is depicted in Figure 5. The difference is computed with respect to the condition LPC. From this difference figure, we can observe that LPC performs better than the two versions of DQ for the speech items s2 and s11. However, for speech items s4 and s12, the two versions of DQ perform better than LPC. For 9 speech items, there is no statistical difference between the three different conditions. For the speech item s6, the condition DQ14 performs worse compared to LPC although performance of DQ16 is same as LPC. For the speech item s3, the condition DQ14 performs significantly better compared to LPC although the performance of DQ16 is the same as that of LPC. Averaged over all items, there is no statistically significant difference between all the three conditions that were compared. Hence, we can safely conclude that subjective quality of the two variants of distribution quantization is as good as the conventional linear prediction technique.

To gain further insight into the performance of the proposed method, we performed the following objective measurements. First, we determined the average bit consumption per frame of the three conditions LPC, DQ16 and DQ14 as 44.98 bits, 43.58 bits and 40.86 bits respectively. Since the two variants of the distribution quantization yield the same subjective quality as the conventional linear prediction with almost the same bit consumption, it can be concluded that distribution quantization as an envelope modeling technique can replace linear prediction as an envelope modeling technique. In addition, distribution quantization is computationally less complex compared to conventional linear prediction as it does not involve computationally complex techniques such as estimation of line spectral frequencies and vector quantization. In addition, distribution quantization can be implemented, as here, as a variable bit rate technique since the bit consumption is directly dependent on the initial quantization level. Depending on the availability of bits, the value of initial quantization level can be increased or decreased. Distribution quantization can therefore also be implemented as a fixed bit-rate codec using a rate-loop.

Second, we measured the perceptual SNR for the outputs of each method. Figure 6 shows the comparison of perceptual signal to noise ratio for a 2.5 s segment of a speech signal. We observe that the perceptual SNR for all the three conditions are similar. During the speech phase, the perceptual SNR values for LPC tends to be slightly higher than both the variants of distribution quantization. The mean perceptual SNR values are 8.11 dB, 7.62 dB, 7.78 dB for LPC, D16 and DQ14 respectively. Though the mean perceptual SNR of the conventional linear prediction is better than the two variants of the distribution quantization, the difference is less than 0.4 dB. However, since the MUSHRA test did not reveal a difference in perceptual quality between methods, we conclude that the difference in perceptual SNR is either too small to be perceived or that the measure itself is not an accurate measure of perceptual quality. Linear prediction and the proposed method have, either way, perceptually equivalent quality.

Finally, the complexity of the conventional linear prediction is compared with that of our novel approach using the Big-O notation. The complexity of DQ is $\mathcal{O}(N \log N + QM)$ and the complexity of LPC using multi-stage vector quantization is $\mathcal{O}((\sum_k 2^{B_k} M) + Q \log Q)$ where M is the model order, N is the FFT length, Q is the number of quantization bins and B_k is the bit consumption by the vector codebook at stage k. For a typical set of constants ($N = 256$, $Q = 100$, $M = 16$, $B_k = \{8, 8, 7, 7, 7, 7\}$), the proposed method thus has a complexity of only one third of traditional linear predictive coding.

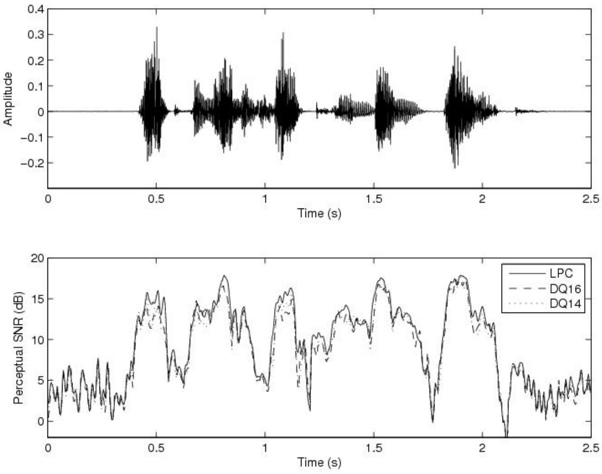

Figure 6: Perceptual signal to noise ratio comparison for a 2.5 s segment from the input file. The waveform corresponding to the measurement is depicted at the top followed by the actual measurement of perceptual SNR in dB.

5. Conclusions

We have presented an entropy coder for envelope models, with the objective of reaching the same coding efficiency and perceptual quality as offered by linear predictive models, while retaining the computational simplicity and flexibility of scale factor bands. The proposed distribution quantization approach is based on a recursive parametrization, where 1) the signal is split into two segments, 2) the magnitude distribution of the two segments is quantified with a ratio of their p-norms, and 3) the sub-segments are recursively processed until the desired accuracy is reached. We have shown that the distribution of parameters thus obtained follow a beta mixture model, which serves as the basis of our entropy coder.

In comparison to linear prediction, the proposed parametrization is computationally simple to determine, since it does not require estimation of line spectral frequencies. Moreover, since our parameters are approximately independent, we can encode them separately, whereby computationally complex vector quantizers can be avoided. It is thus clear that the proposed parametrization has lower algorithmic complexity than linear predictive coding. The approach is also flexible since the design does not depend on the bit-rate, and allows straightforward application of perceptual criteria in its design. Furthermore, our experiments demonstrate that both the objective and subjective quality of the proposed approach is equivalent or better than linear predictive coding.

The proposed method for encoding spectral envelopes using entropy coding of the distribution quantization parameters is therefore a competitive choice for encoding spectral envelopes in speech and audio codecs. In addition, the same approach is applicable for the coding of any other envelopes such as temporal magnitude envelopes or spatial sound fields.

6. References

[1] *TS 26.190, Adaptive Multi-Rate (AMR-WB) speech codec*, 3GPP, 2007.

[2] ITU-T Recommendation G.718, "Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8–32 kbit/s," 2008.

[3] M. Neuendorf, M. Multrus, N. Rettelbach, G. Fuchs, J. Robilliard, J. Lecomte, S. Wilde, S. Bayer, S. Disch, C. Helmrich, R. Lefebvre, P. Gournay, B. Bessette, J. Lapierre, K. Kjörling, H. Purnhagen, L. Villemoes, W. Oomen, E. Schuijers, K. Kikuiri, T. Chinen, T. Norimatsu, K. S. Chong, E. Oh, M. Kim, S. Quackenbush, and B. Grill, "The ISO/MPEG unified speech and audio coding standard – consistent high quality for all content types and at all bit rates," *Journal of the AES*, vol. 61, no. 12, pp. 956–977, 2013.

[4] *TS 26.445, EVS Codec Detailed Algorithmic Description; 3GPP Technical Specification (Release 12)*, 3GPP, 2014.

[5] P. P. Vaidyanathan, "The theory of linear prediction," in *Synthesis Lectures on Signal Processing*. Morgan & Claypool publishers, 2007, vol. 2, pp. 1–184.

[6] M. Schroeder and B. Atal, "Code-excited linear prediction (CELP): High-quality speech at very low bit rates," in *Proc. ICASSP*. IEEE, 1985, pp. 937–940.

[7] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 411–423, 1991.

[8] C. Magi, J. Pohjalainen, T. Bäckström, and P. Alku, "Stabilised weighted linear prediction," *Speech Communication*, vol. 51, no. 5, pp. 401–411, 2009.

[9] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Sparse linear prediction and its applications to speech processing," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 5, pp. 1644–1657, 2012.

[10] F. Itakura, "Line spectrum representation of linear predictor coefficients of speech signals," *The Journal of the Acoustical Society of America*, vol. 57, p. S35, 1975.

[11] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech Audio Process.*, vol. 1, no. 1, pp. 3–14, 1993.

[12] Z. Ma and A. Leijon, "Bayesian estimation of beta mixture models with variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2160–2173, 2011.

[13] Z. Ma, A. Leijon, and W. B. Kleijn, "Vector quantization of LSF parameters with a mixture of Dirichlet distributions," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 9, pp. 1777–1790, 2013.

[14] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2003.

[15] K. Brandenburg and M. Bosi, "Overview of MPEG audio: Current and future standards for low bit-rate audio coding," *J. Audio Eng. Soc*, vol. 45, no. 1/2, pp. 4–21, 1997. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7871

[16] T. Bäckström and C. R. Helmrich, "Arithmetic coding of speech and audio spectra using TCX based on linear predictive spectral envelopes," in *Proc. ICASSP*, Apr. 2015, pp. 5127–5131.

[17] J. Herre and J. D. Johnston, "Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS)," in *Proc AES Convention 101*, Los Angeles, CA, USA, Nov. 8–11 1996.

[18] J. Herre, H. Purnhagen, J. Koppens, O. Hellmuth, J. Engdegård, J. Hilper, L. Villemoes, L. Terentiv, C. Falch, A. Hölzer *et al.*, "MPEG spatial audio object coding—the ISO/MPEG standard for efficient coding of interactive audio scenes," *Journal of the Audio Engineering Society*, vol. 60, no. 9, pp. 655–673, 2012.

[19] T. Jähnel, T. Bäckström, and B. Schubert, "Envelope modeling for speech and audio processing using distribution quantization," in *Proc. EUSIPCO*, 2015.

[20] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Springer, 1992.

[21] C. Walck, *Handbook on statistical distributions for experimentalists*. University of Stockholm Internal Report SUF-PFY/96-01, 2007.

[22] J. Rissanen and G. G. Langdon, "Arithmetic coding," *IBM Journal of research and development*, vol. 23, no. 2, pp. 149–162, 1979.

[23] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H Audio – The new standard for universal spatial/3D audio coding," *Journal of the Audio Engineering Society*, vol. 62, no. 12, pp. 821–830, 2015.

[24] NTT-AT, "Super wideband stereo speech database," http://www.ntt-at.com/product/widebandspeech, accessed: 09.09.2014. [Online]. Available: http://www.ntt-at.com/product/widebandspeech

[25] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 1996.

[26] Recommendation BS.1534, *Method for the subjective assessment of intermediate quality levels of coding systems*, ITU-R, 2003.

[27] P. Kabal and R. P. Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 6, pp. 1419–1426, 1986.