GMM BASED ITERATIVE ENTROPY CODING FOR SPECTRAL ENVELOPES OF SPEECH AND AUDIO

Srikanth Korse[1], Guillaume Fuchs[1,2], Tom Bäckström[3]

[1]Fraunhofer IIS, Erlangen, Germany [2]International Audio Laboratories, Friedrich-Alexander University (FAU), Erlangen, Germany

[3]Aalto University, Helsinki, Finland

Introduction

- Spectral envelope models are integral part of speech and audio codecs
- Common parameterization techniques are linear predictive coding (LPC) parameters and scale factor bands (SFB) coded with vector quantizer (VQ) [1].
- We have recently proposed an alternative: distribution quantization (DQ)

	VQ	Existing GMM	GMM Proposed
Codebook Size	Function of dimension and bitrate	Independent of dimension and bitrate	Independent of dimension and bitrate
Domain	Original	Transform (KLT)	Original
Component classifier	No	Yes	No

Contribution of the current work:

- Iterative GMM approach for entropy coding of spectral envelopes.
- We derive a univariate probability distribution for each parameter using the previously quantized parameters as prior information. The conditional pdf is a scalar GMM.

Multivariate Gaussian Distribution

- Assume x can be modelled using Gaussian mixture model (GMM) with M components such that the probability distribution function is:

$$f(x) = \sum_{k=1}^{M} \lambda_k f_k(x) \quad \text{where } \sum_{k=1}^{M} \lambda_k = 1$$

$$f_k(x) = |2\pi\Sigma_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu_k)^T {\Sigma_k}^{-1}(x-\mu_k)\right)$$, where Σ: covariance matrix, μ: mean

- The vector x can be divided into 2 parts: x_0: coded coefficients and x_1: coefficients yet to be coded.
- It can be shown that, when x_0 is known, x_1 follows normal distribution with covariance ${A_1}^{-1}$ and mean $\hat{\mu}$ but scaled with $\alpha = \frac{e^c |A_1|^{\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}}$.
- In other words, the mean and the weights of x_1 are updated depending on the previously encoded samples. The covariances ${A_{k,1}}^{-1}$ are depended only Σ_k, hence can be computed offline.

The algorithm can be stated as follows:

- Encode the first component ξ_0 using the univariate distribution without priors.
- For h = 1 to N-1
  - Derive pdf for ξ_h using ξ_0 to ξ_{h-1} as priors.
  - Encode component ξ_h with the help of arithmetic coding using the univariate pdf obtained.

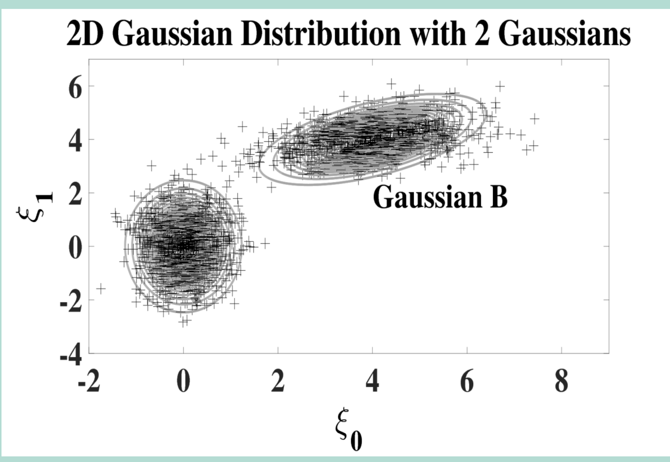

Fig. 1. Illustration of 2D Gaussian mixture model with 2 Gaussians

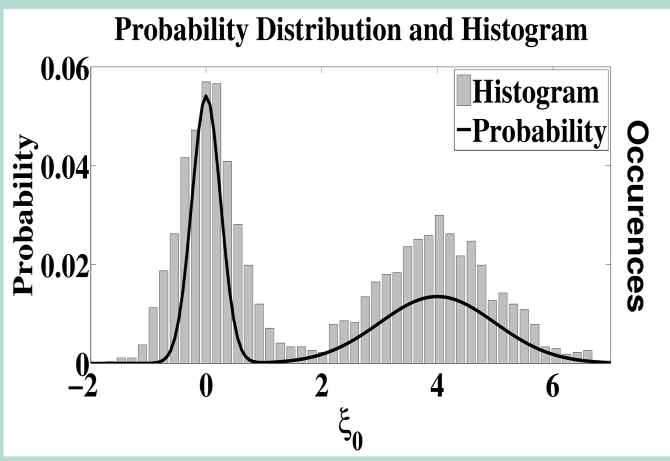

Fig. 2. Histogram and probability distribution model of first parameter ξ_0

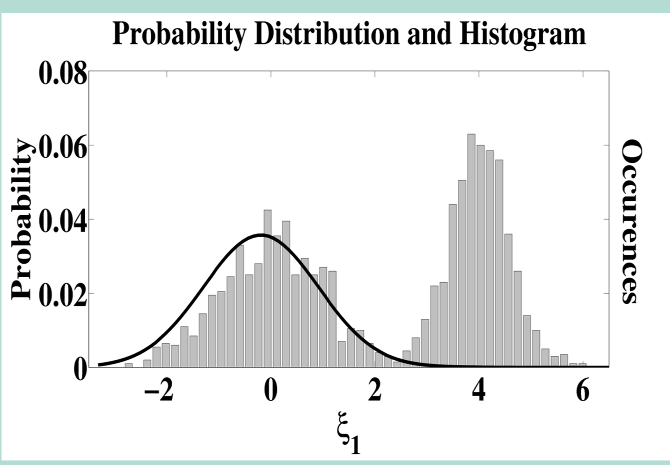

Fig. 3. Histogram and probability distribution model of ξ_1 w.r.t Gaussian A

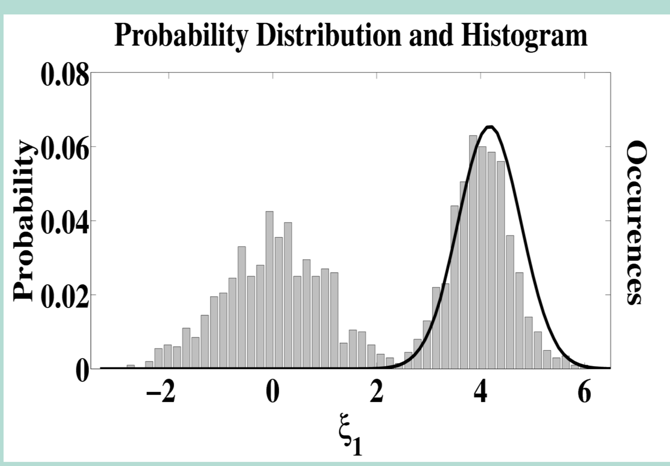

Fig. 4. Histogram and probability distribution model of ξ_1 w.r.t Gaussian B

Experiments and Results

Database used for Training and Testing	TIMIT[2]
Tested versions	VQ, GMM
Tested bandwidth	Narrowband (NB) and Wideband (WB)
Number of Gaussians tested	3, 5 and 10 (NB), 5, 10 and 15 (WB)
Objective measure	Log Spectral Distance (LSD)
Bitrate tested	24 and 33 bits (NB), 36 and 43 bits (WB)

- 8 different parameterizations of the envelope: 1. LSF, 2. Delta-LSF, 3. LSF with inverse sigmoid mapping, 4. distribution quantization (DQ), 5. DQ with log-mapping, 6. scale factor bands (SFB), 7. SFB with log-mapping, 8. SFB with inverse sigmoid mapping.
- Baseline system: Tree-search multi-stage VQ.
- Logistic distribution approximation is used to obtain the cumulative distribution function.

Results:

- LSF is better than D-LSF and LSF-IS
- DQ-LD is a bit better than DQ-ER
- VQ has slightly better LSD than GMM for almost all points
- Number of outliers are similar
- Complexity of VQ is $O(MN2^B)$ and for GMM it is $O(N^2)$, where M is the number of stages, N is the vector length and B is the number of bits.
  - Complexity of GMM is lower than VQ.

Condition	2-4 dB	> 4 dB	Bits	Mean SD
VQ-NB-24	3.87%	0.59%	24.00	0.68 dB
GMM10-NB-24	3.91%	0.66%	24.32	0.67 dB
VQ-WB-43	5.85%	0.94%	43.00	0.81 dB
GMM15-WB-43	6.87%	1.31%	42.40	0.87 dB

Table 1. Outlier comparison for LSF parameter at 24 bits (NB) and 43 bits (WB).

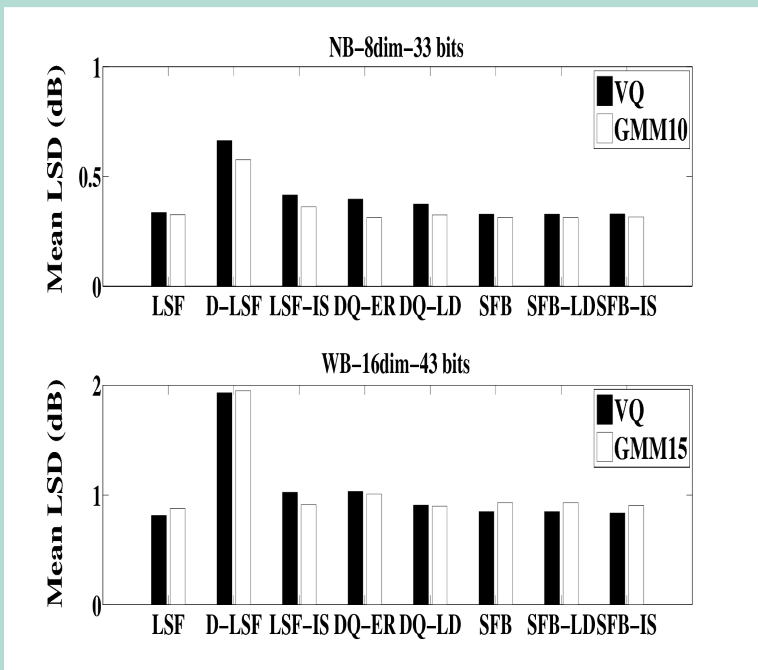

Fig. 5. Mean Log Spectral distance (LSD) (dB) vs all parameters at 33 bits (NB) and 43 bits (WB).

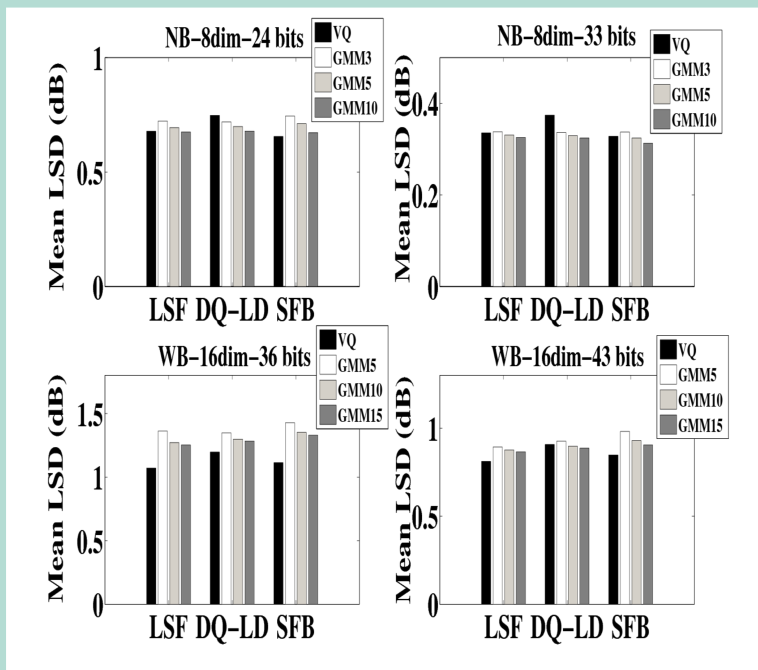

Fig. 6. Mean Log Spectral distance (LSD) (dB) vs all configurations for LSF, DQ-LD and SFB at NB (24 and 33 bits) and WB (36 and 43 bits).

Conclusion

- We propose an iterative GMM based entropy coder to encode the spectral envelope parameters.
- Coding efficiency (LSD and outliers) of VQ and GMM are similar.
- VQ is not flexible (training specific to bitrate).
- Proposed GMM has flexible bitrate (training independent of biteate) and low complexity.
- Compared to previous GMM methods, proposed scheme does not require decorrelation and component classification.

References

1) Tom Bäckström, *"Speech Coding with Code-Excited Linear Prediction"*, Springer, 2017

2) J S Garofolo, Linguistic Data Consortium, et al., TIMIT: acoustic-phonetic continuous speech corpus, Linguistic Data Consortium, 1993.

IN COOPERATION WITH

The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS