Flexible Quantization of Audio and Speech Based on the Autoregressive Model

Alexey Ozerov and W. Bastiaan Kleijn

ACCESS Linnaeus Center, Electrical Engineering, KTH - Royal Institute of Technology, Stockholm, Sweden

KTH VETENSKAP OCH KONST

FlexCode project

Abstract

- We present a coding scheme based on audio and speech quantization with an adaptive quantizer derived from the autoregressive (AR) model under high-rate assumptions.
- "The new coder can run under both constraint resolution (CR) (constant rate) and constraint entropy (CE) (variable rate) quantization scenarios.
- Compared to state-of-the-art training-based coders (e.g., CELP) the proposed scheme has the following advantages:
  - ability to run for any rate without re-training,
  - rate-independent computational complexity,
  - possibility of variable-rate quantization,
  - quantization cell shapes are optimal for local signal statistics.
- Experiments indicate that, compared to a CELP scheme, the proposed flexible coder performs at least as well in the CE case, and has nearly identical performance for the CR case.

Goals

Heterogeneous transmission networks => Need coders that are able to adapt in real time to continuously changing network conditions.

In *FlexCode* **project** we are aiming to develop a flexible audio coder that:

- performs as well as the state-of-the-art speech and audio coders,
- is able to run for any rate from a continuum of rates.

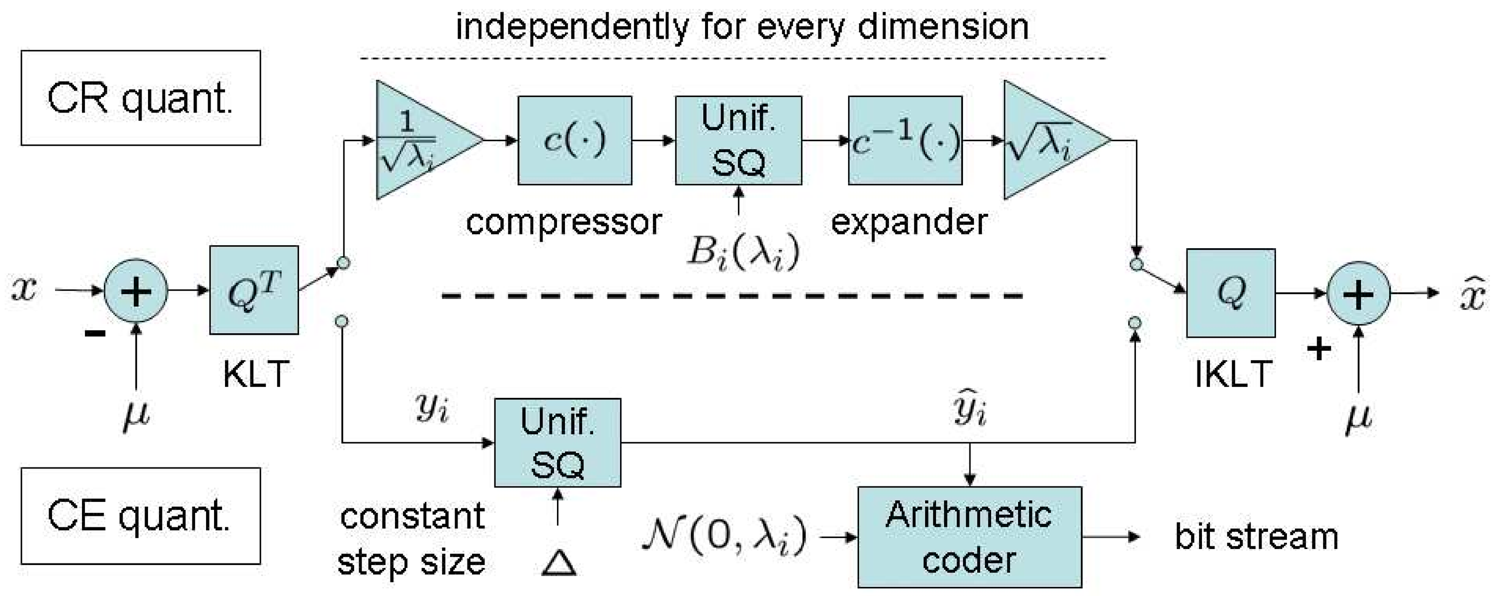
Motivation

- The modern heterogeneous network environment is not well served by the ubiquitous CELP algorithm:
  - non-adaptability for any particular rate,
  - computational complexity grows exponentially with rate,
  - variable-rate quantization is not facilitated,
  - quantization cell shapes are not locally optimal in signal domain (they can be only optimal in average).
- To address requirements of modern networks:
  - we cannot train and store codebooks,
  - we need adaptive codebooks, which can be computed in real time.
- Probabilistic source modeling combined with high-rate theory allows such adaptive quantization.

Adaptive Model-Based Quantization

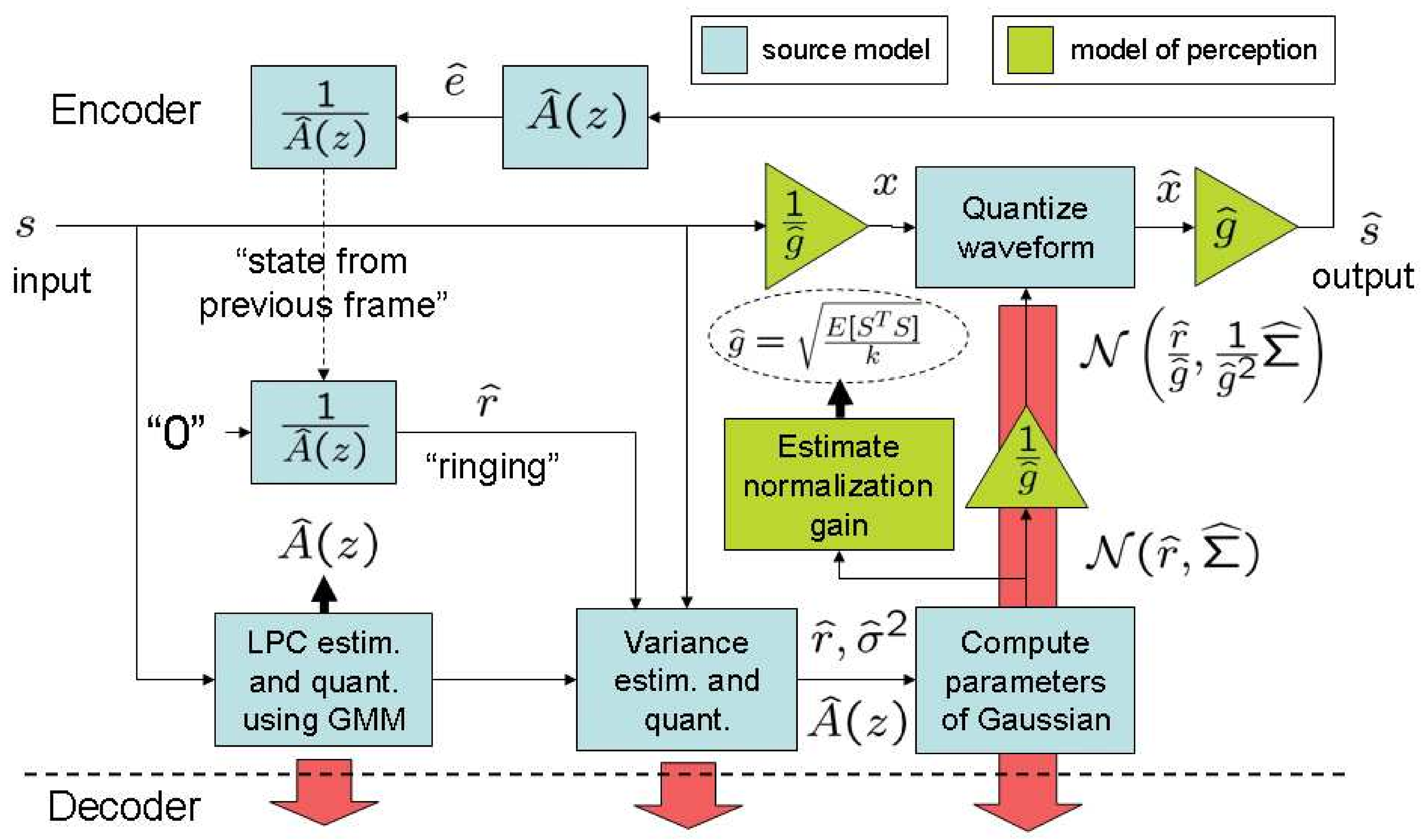
$X \in \mathbb{R}^k, \quad X \sim \mathcal{N}(\mu, \Sigma)$

$\Sigma = Q \Lambda Q^T, \quad Q^T Q = I, \quad \Lambda = \text{diag}\{\lambda_i\}_i$



Proposed Coding Scheme

- Coded signal is segmented into frames s.
- Redundancy removal:
  - Intra-frame redundancy: AR model-based KLT.
  - Inter-frame redundancy: AR model-based ringing subtraction.
- Model for frame s: $S \sim \mathcal{N}(r, \Sigma)$, where
  - r is "ringing", and
  - Σ is computed from AR model defined by LPC $A(z)$ and σ^2.
- AR model parameters are quantized as well (forward adaptation).
- Rate distribution between model and signal. General result [1]:
  - *The optimal rate for the model is independent on the overall rate.*



Results

- Comparison with a CELP scheme using a codebook trained to minimize the MSE in signal domain.
  - 10 sentences of 8kHz speech, frame length = 5 samples,
  - rate = 19.2 kbps (12 bits per frame), LPC are not quantized.

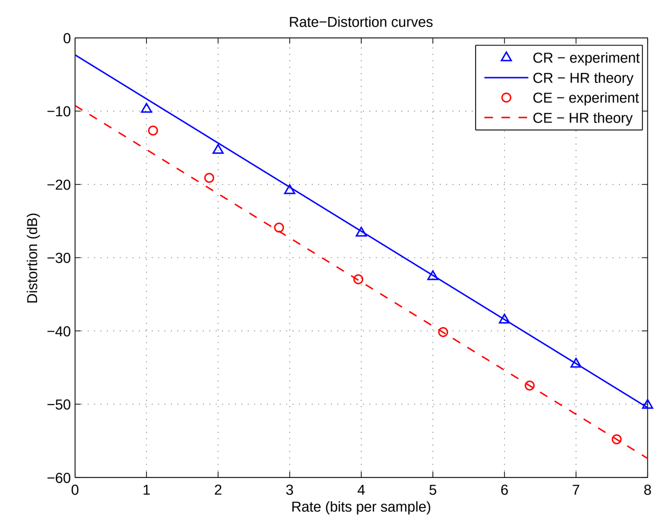
	Proposed (CR)	Proposed (CE)	CELP
Variance rate (bpf)	3	2.7	5
Signal rate (bpf)	9	9.2	7
SSNR (dB)	16.07	17.96	17.82
fr. len. = 10 samp.	*18.84*	*20.43*	-

  - CELP uses VQ while proposed scheme uses scalar quantizers.

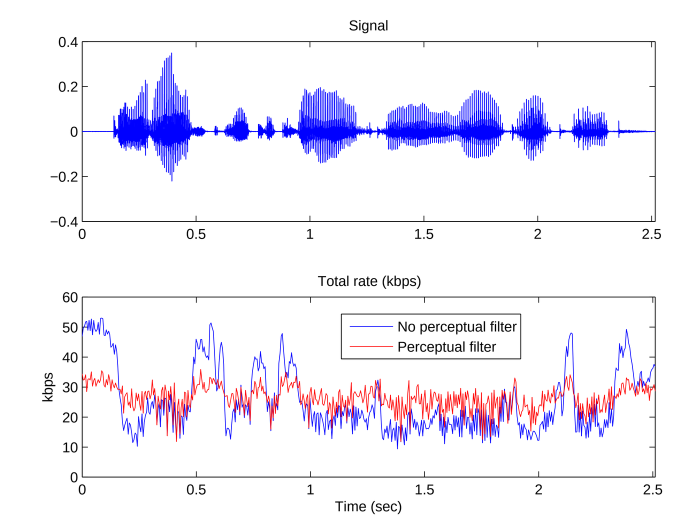
- Rate vs. distortion.

$R_{\mathrm{CR}} = -(k/2)\log_2 \bar{D}_{\mathrm{CR}} + G_{\mathrm{CR}},$

$\bar{R}_{\mathrm{CE}} = -(k/2)\log_2 D_{\mathrm{CE}} + G_{\mathrm{CE}}.$



- Rate variation (for CE quantization).



Conclusion

- We developed a coder for heterogeneous networks that:
  - can be reconfigured in real-time,
  - uses only computable quantizers,
  - does not require storage of quantization tables,
  - has complexity independent of rate,
  - has performance equivalent to CELP.

References

[1] W. B. Kleijn and A. Ozerov, "Rate distribution between model and signal," in IEEE WASPAA, Mohonk, NY, Oct. 2007.