Using Salient Envelope Features for Audio Coding

Joachim Thiemann and Peter Kabal

TSP Lab, McGill University

August 28, 2008

Outline

Perceptual Audio Coding

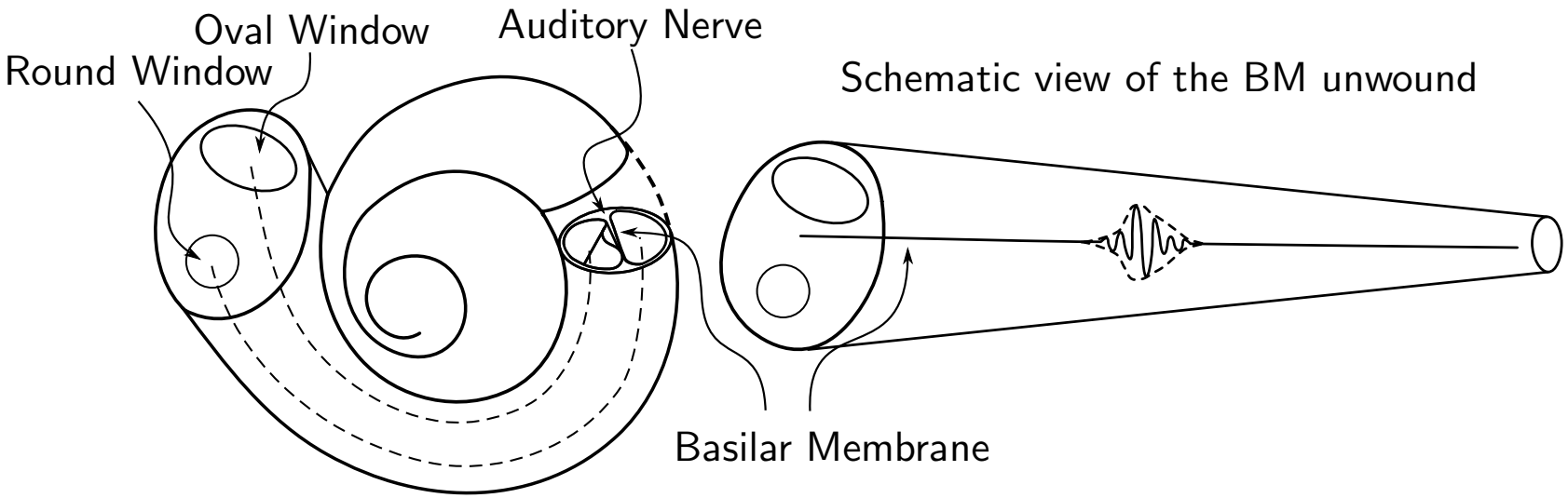
- Most commercial audio codecs use fast block-transforms (MDCT/FFT).
- While computationally efficient, these transforms do not match the physiological processes of the human ear very well.
- Additional processing is applied to the results to extract psychoacoustically relevant information and code it efficiently.
- This works reasonably well — but can we do better?

Motivation of Present Work

- Anthropomorphic coding (Feldbauer, Kubin, Kleijn) simulates parts of the auditory system, then encodes the model parameters.
- If the simulation is accurate, the model parameters represent *only* the audible information and thus are an efficient perceptual code (Smith and Lewicki).
- Reconstruction of audio from those parameters can be difficult.
- In the work presented here, we ensure explicitly that the *model parameters* of the reconstructed audio match the encoded data.

The Auditory System

Presently, the auditory model is limited to a passive model of the Basilar Membrane movement.

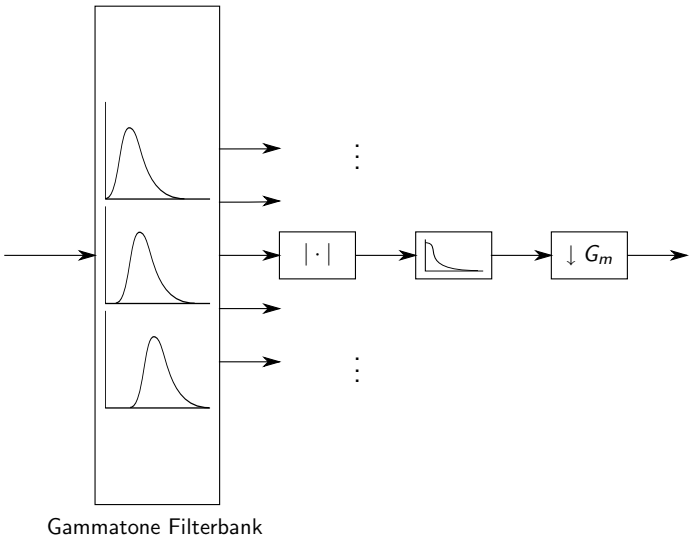


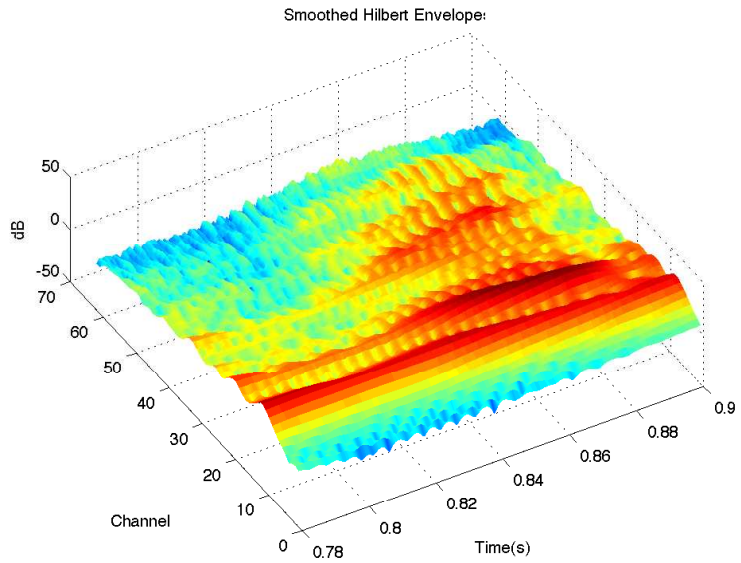
Analysis

- The analysis stage is based on the Dau perceptual model.
- The input is analysed by a set of complex gammatone filters.
- Using a magnitude operator, the phase components of the filter outputs are stripped off.
- The resulting signals are real, positive and strongly low-pass.
- Taking the temporal changes into account, we get an "Excitation Surface".
- This "Surface" is similar to the "Excitation Pattern" of the perceptual model within PEAQ.

Envelope Sampling

Envelopes are low-pass filtered and sampled at varying rates based on experiments by Ghitza.





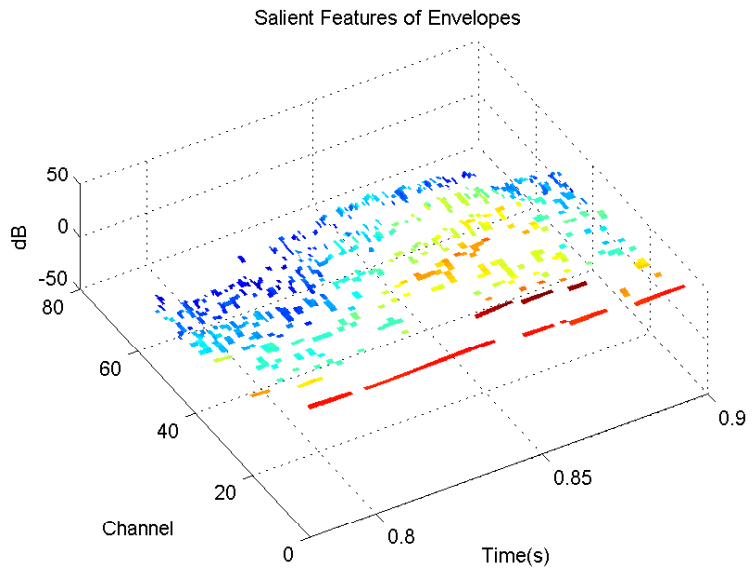
In the current implementation, 60 input samples ($f_s = 16$kHz) result in 170 envelope samples. The filterbank uses 62 filters at half-bark spacing.

Selecting Salient Features

- We refer to Salient Features (SF) as envelope samples whose *levels* are considered important to perception.
- The algorithm determining relevance is similar to Feldbauer's Auditory Pulse sparsification.
- We consider the excitation surface that a single isolated gammatone pulse would create.
- This envelope of an isolated pulse gives us an indication of which other envelope samples in the vicinity we need to consider.
- If the excitation surface surrounding a "loud" sample exceeds a surface that a single pulse would create, additional points of the "target" surface are relevant.

Selection Algorithm

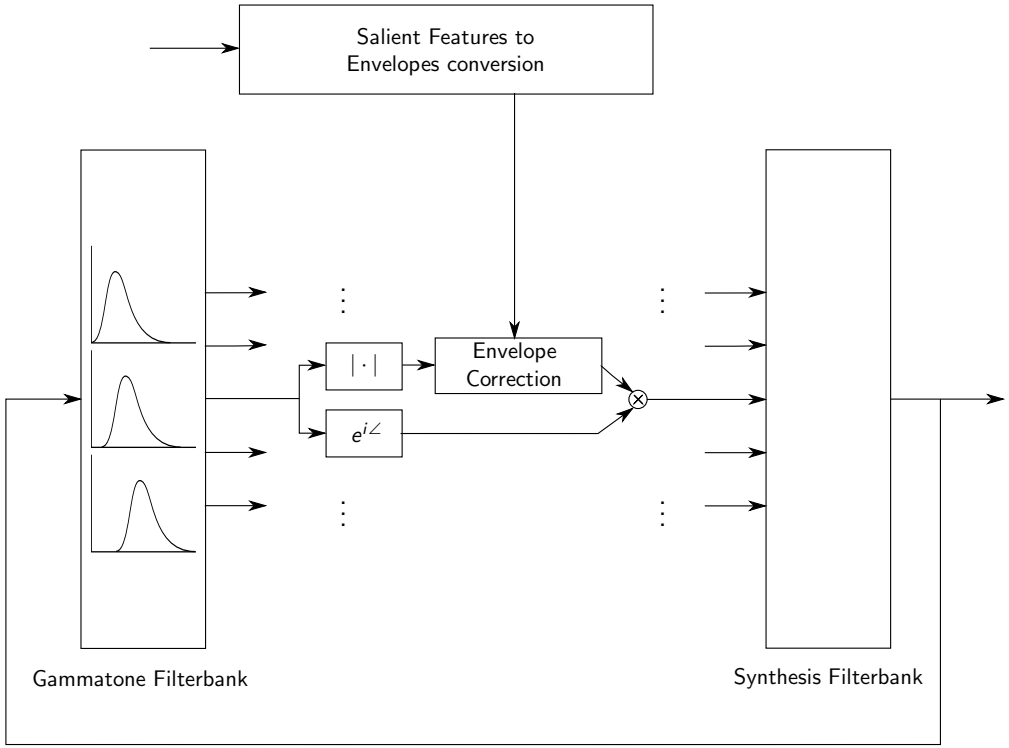
- We begin by initializing a working excitation surface to the absolute threshold of hearing.
- The isolated-pulse effect of the largest sample of the target surface that is *above the working surface* is added to the working surface.
- This is repeated until all points of the target surface are at or below the working surface points.



Reconstructing the Audio at the Decoder

- The decoder does not know the exact excitation surface, only the samples the encoder considered relevant.
- However, we have some implied knowledge about the surface at the other points.
- Reconstruction is iterative. An initial guess is made of the audio signal, which is then refined.
- Refinement is done by analyzing the reconstructed audio by the same process as the encoder.
- At each iteration, the envelopes are corrected while retaining the phase of the previous estimate.

Reconstruction Block Diagram



Data Rate Estimations

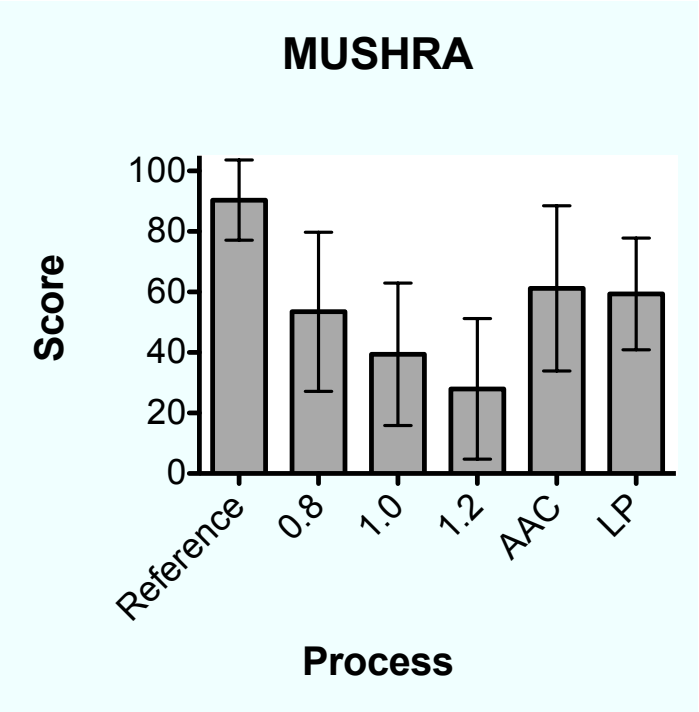
		SF per Frame		
Sample	Frames	$\gamma = 0.8$	$\gamma = 1.0$	$\gamma = 1.2$
SQAM35	2322	26.45	18.78	11.63
SQAM48	2870	70.53	40.07	24.68
SQAM49	1770	61.02	35.93	22.58
SQAM50	1983	61.52	35.74	21.92
SQAM51	1797	60.64	35.72	22.12
SQAM54	1853	58.76	34.23	21.28
SQAM60	3318	57.20	31.85	19.56

Samples were resampled to 16kHz, frames are 3.75ms (60 samples).

Subjective Tests

- Initial informal tests indicate that with most items the sound quality of Enhanced aacPlus at 16kbps can be achieved.
- Best performance was with male speech items.
- Tonal sounds (SQAM35, "Glockenspiel") are very problematic.

▸ Detailed breakdown



Work to be done

- The biggest problem with the presented method is the poor performance with pitched tonal sounds.
  - The phase could be encoded in some way during tonal sections.
  - Another possibility is adding an *adaptive* component to the filterbank.

  However, either would be difficult to integrate with the SF algorithm and the iterative reconstruction.
- To reduce the bit rate, a well designed vector quantizer is needed.
- The SF selection algorithm can be optimised.

Conclusion

- It is possible to reconstruct an audio signal with good quality from a sparse representation of the auditory envelopes.
- This works well for transient or spectrally rich sounds but less so for tonal sounds.
- Analysis and SF selection is of low complexity, but reconstruction is computationally expensive.
- The deliberate exclusion of phase information — and the problems caused by that — is interesting for the development of future codecs.

References

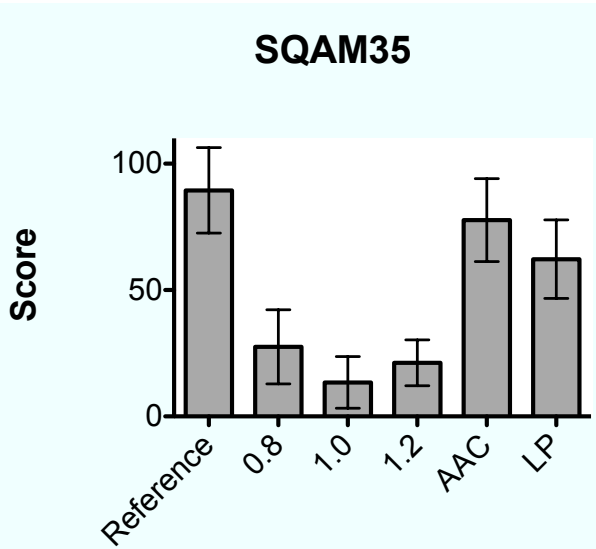

C. Feldbauer, G. Kubin, and W. B. Kleijn, "Anthropomorphic coding of speech and audio: A model inversion approach," *EURASIP J. Applied Signal Processing*, pp. 1334–1349, Sep 2005.

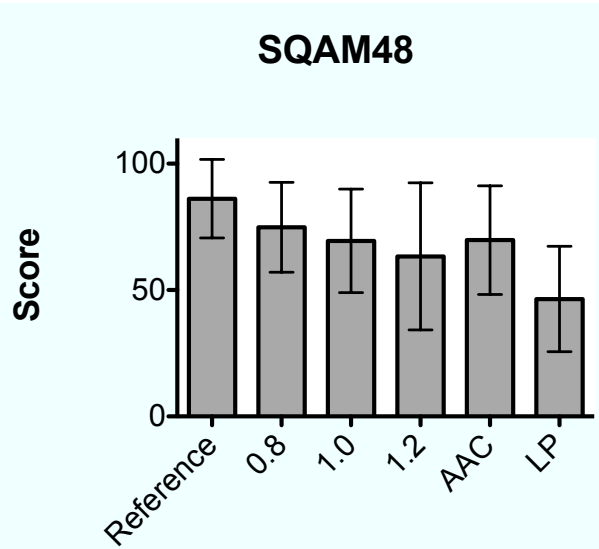

E. C. Smith and M. S. Lewicki, "Efficient auditory coding," *Nature*, vol. 439, pp. 978–982, Feb. 2006.

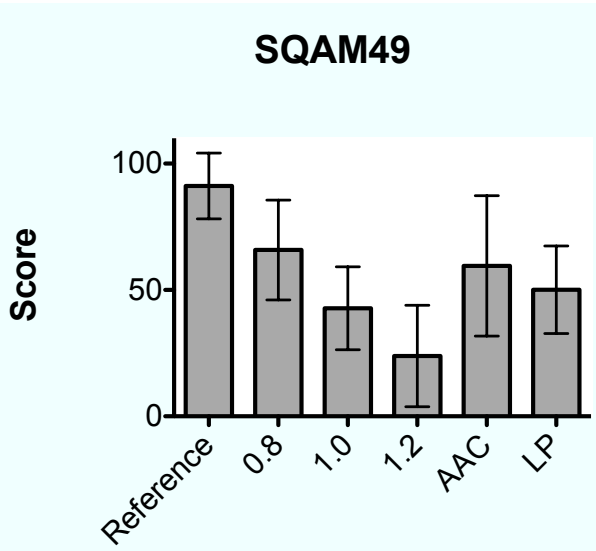

T. Dau, D. Puschel, and A. Kohlrausch, "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," *J. Acoust. Soc. Am*, Jan 1996.

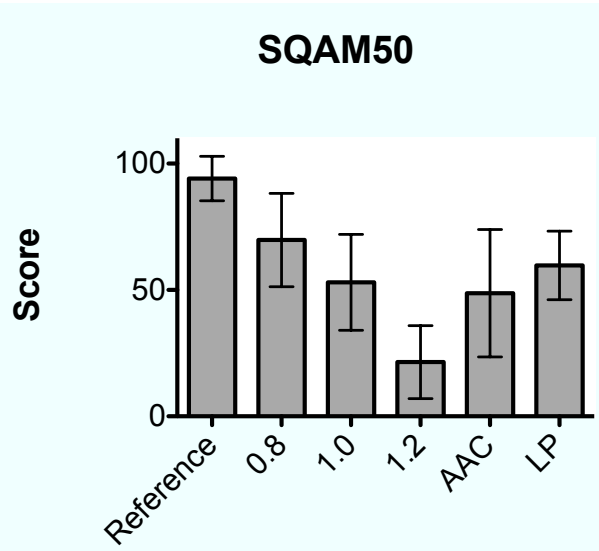

O. Ghitza, "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," *J. Acoust. Soc. Am.*, vol. 110, pp. 1628–1640, Sep. 2001.

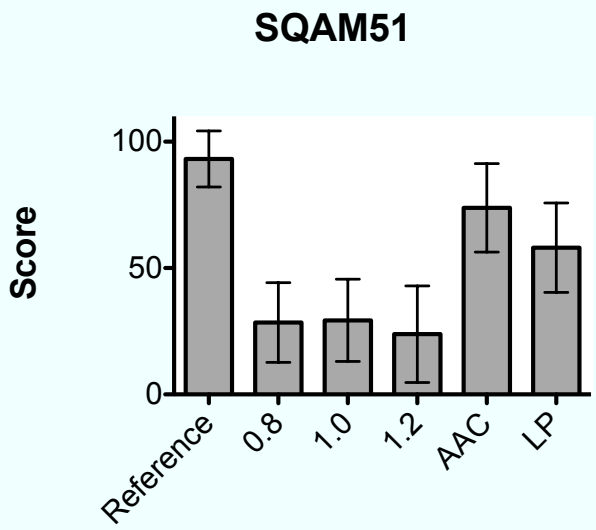

J. Thiemann and P. Kabal, "Reconstructing Audio Signals from Modified Non-Coherent Hilbert Envelopes," in *Proc. Interspeech 2007*, (Antwerpen, Belgium), pp. 534–537, Aug. 2007.

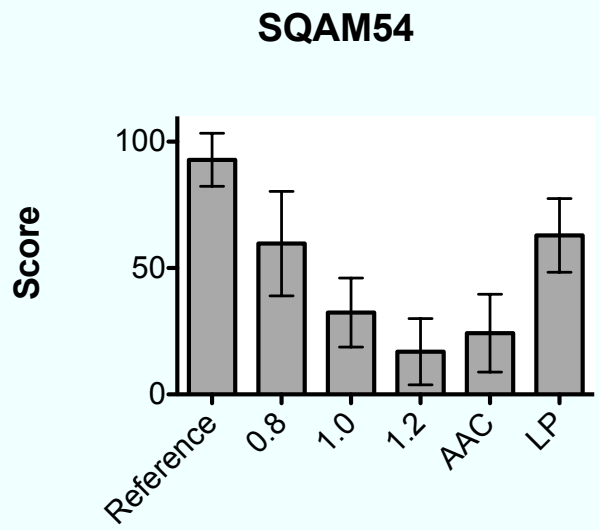

SQAM35
Score
100
50
0
Reference
0.8
1.0
1.2
AAC
LP
Process

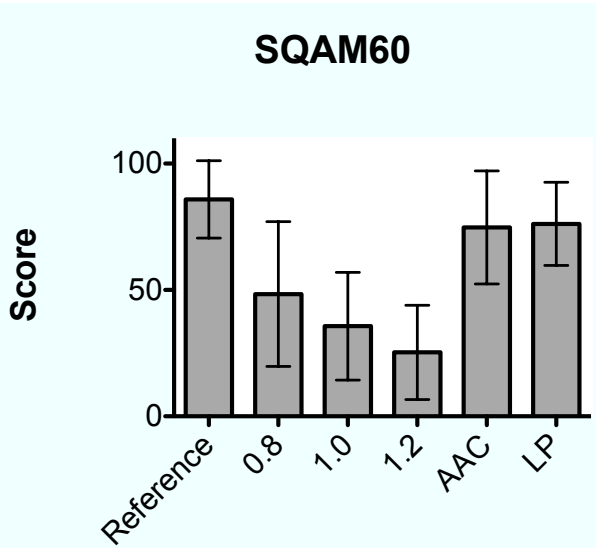

SQAM48
Score
100
50
0
Reference
0.8
1.0
1.2
AAC
LP
Process


SQAM49
Score
100
50
0
Reference
0.8
1.0
1.2
AAC
LP


SQAM50
Score
100
50
0
Reference
0.8
1.0
1.2
AAC
LP


SQAM51
Score
100
50
0
Reference
0.8
1.0
1.2
AAC
LP
Process


SQAM54
Score
100
50
0
Reference
0.8
1.0
1.2
AAC
LP
Process


SQAM60
Score
100
50
0
Reference
0.8
1.0
1.2
AAC
LP