Perceptual Audio Coding

Henrique Malvar
Director
Microsoft Research

UW Lecture – February '06

Contents

- Motivation
- "Sink coding": Auditory Masking
- Block & Lapped Transforms
- Audio compression
- Examples

2

Contents

- Motivation
- “Sink coding”: Auditory Masking
- Block & Lapped Transforms
- Audio compression
- Examples

3

Many applications need digital audio

- Business
  - Internet call centers
  - Multimedia presentations
- Communication
  - Digital TV, Telephony (VoIP) & teleconferencing
  - Voice mail, voice annotations on e-mail, voice recording
- Entertainment
  - solid-state music players
  - 150 songs on standard CD
  - thousands of songs on portable jukebox
  - Internet radio
  - Games

4

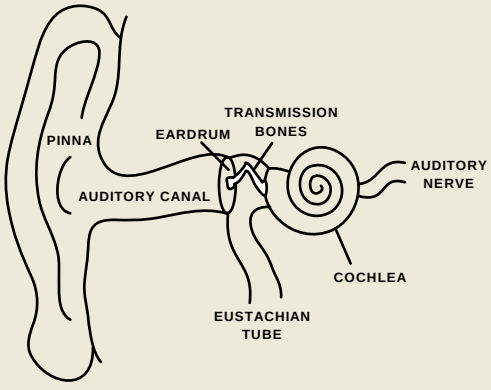
Contents

- Motivation
- “Sink coding”: Auditory Masking
- Block & Lapped Transforms
- Audio compression
- Examples

5

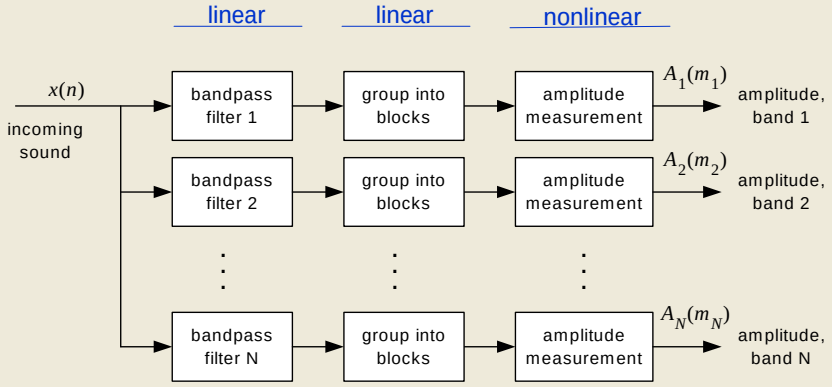
Physiology of the ear

- Automatic gain control
  - muscles around transmission bones
- Directivity
  - pinna
- Boost of middle frequencies
  - auditory canal
- Nonlinear processing
  - auditory nerve
- Filter bank separation
  - cochlea
- Thousands of “microphones”
  - hair cells in cochlea



6

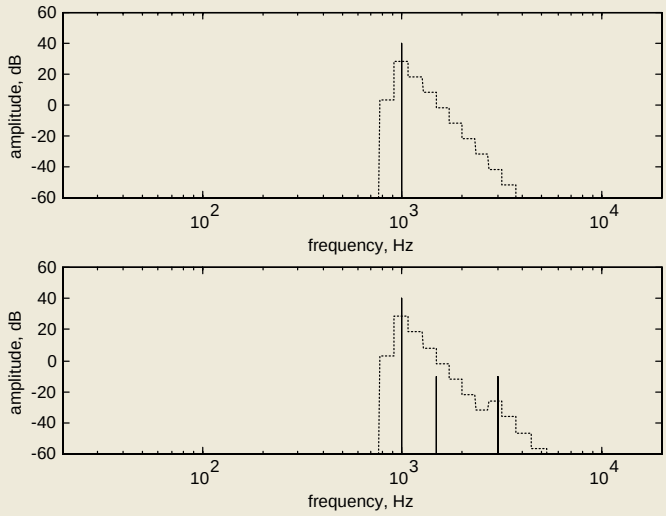
Filter bank model



- Explains frequency-domain masking

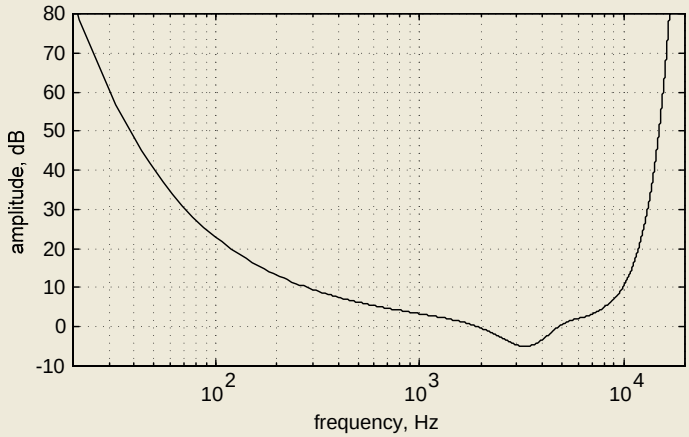
7

Frequency-domain masking



8

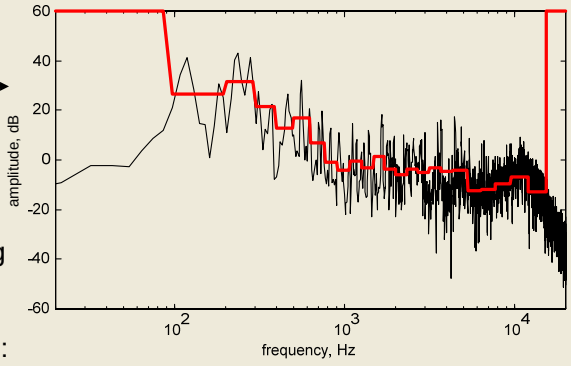
Absolute threshold of hearing



9

Example of masking

- Typical spectrum & masking threshold
- Original sound:
- Sound after removing components below the threshold (1/3 to 1/2 of the data):



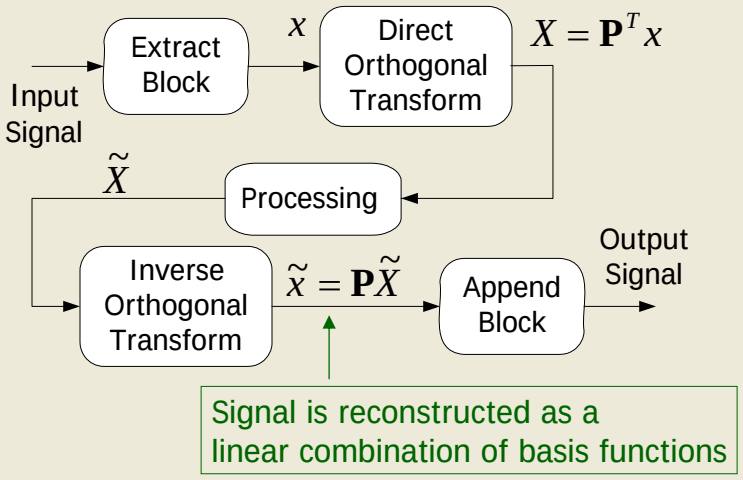
10

Contents

- Motivation
- “Sink coding”: Auditory Masking
- Block & Lapped Transforms
- Audio compression
- Examples

11

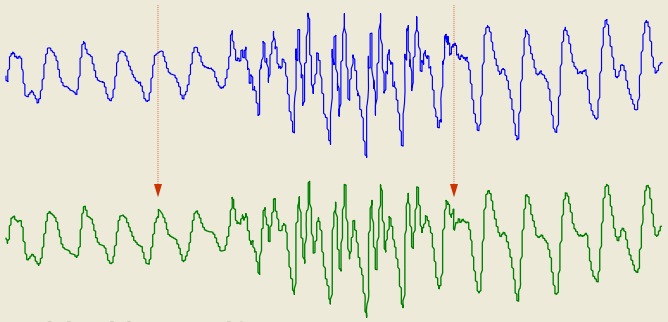
Block signal processing



12

Block processing: good and bad

- Pro: allows adaptability



- Con: blocking artifacts

13

Why transforms?

- More efficient signal representation
  - Frequency domain
  - Basis functions ~ “typical” signal components
- Faster processing
  - Filtering, compression
- Orthogonality
  - Energy preservation
  - Robustness to quantization

14

Compactness of representation

- Maximum energy concentration in as few coefficients as possible
- For stationary random signals, the optimal basis is the Karhunen-Loève transform:

$$\lambda_i p_i = R_{xx} p_i, \quad \mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{I}$$

- Basis functions are the columns of $\mathbf{P}$
- Minimum geometric mean of transform coefficient variances

15

Sub-optimal transforms

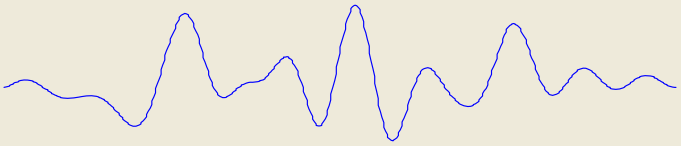
- KLT problems:
  - Signal dependency
  - $\mathbf{P}$ not factorable into sparse components
- Sinusoidal transforms:
  - Asymptotically optimal for large blocks
  - Frequency component interpretation
  - Sparse factors - e.g. FFT

16

Lapped transforms

- Basis functions have tails beyond block boundaries
  - Linear combinations of overlapping functions such as

  - generate smooth signals, without blocking artifacts



17

Modulated lapped transforms

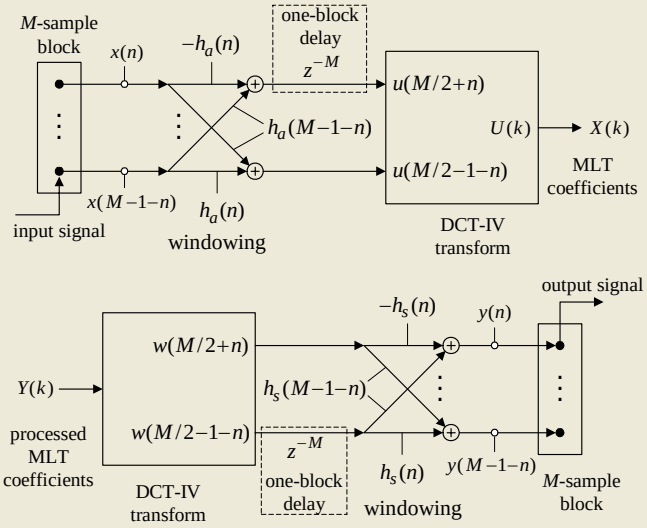
- Basis functions = cosines modulating the same low-pass (window) prototype $h(n)$:

$$p_k(n) = h(n)\sqrt{\frac{2}{M}}\cos\left[\left(n+\frac{M+1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right]$$

- Can be computed from the DCT or FFT
- Projection $X = \mathbf{P}^T x$ can be computed in $O(\log_2 M)$ operations per input point

18

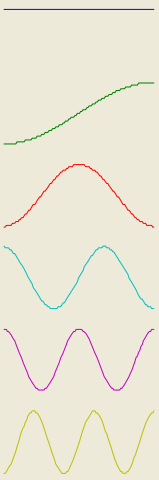
Fast MLT computation



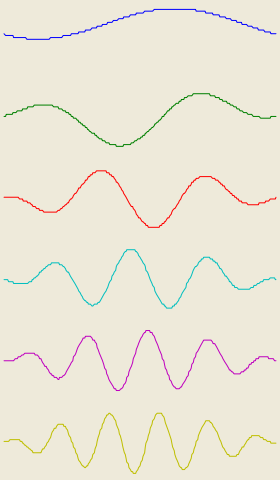
19

Basis functions

DCT:



MLT:



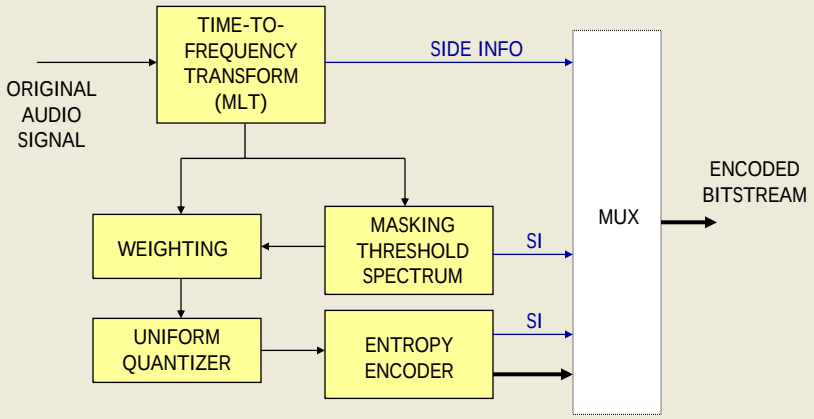
20

Contents

- Motivation
- “Sink coding”: Auditory Masking
- Block & Lapped Transforms
- Audio compression
- Examples

21

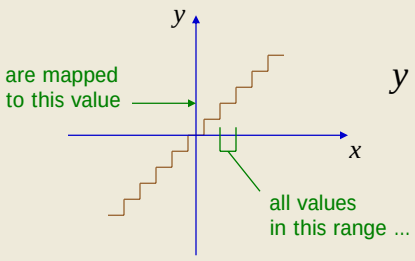
Basic architecture



22

Quantization of transform coefficients

- Quantization = rounding to nearest integer.
- Small range of integer values = fewer bits needed to represent data
- Step size T controls range of integer values

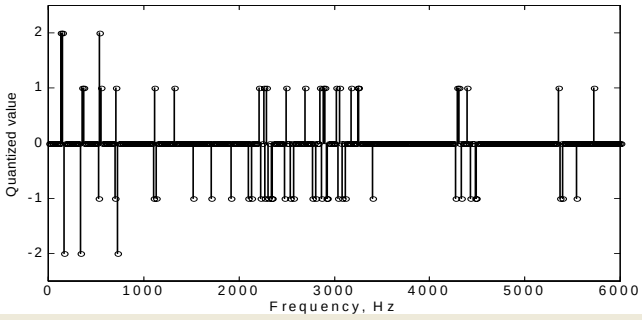


$$y = T\,\mathrm{int}(x / T)$$

23

Encoding of quantized coefficients

- Typical plot of quantized transform coefficients



- Run-length + entropy coding

24

Basic entropy coding

- Huffman coding: less frequent values have longer codewords
- More efficient if groups of values are assembled in a vector before coding

Value	Codeword
-7	'1010101010001'
-6	'10101010101'
-5	'101010100'
-4	'10101011'
-3	'101011'
-2	'1011'
-1	'01'
0	'11'
+1	'00'
+2	'100'
+3	'10100'
+4	'1010100'
+5	'1010101011'
+6	'101010101001'
+7	'1010101010000'

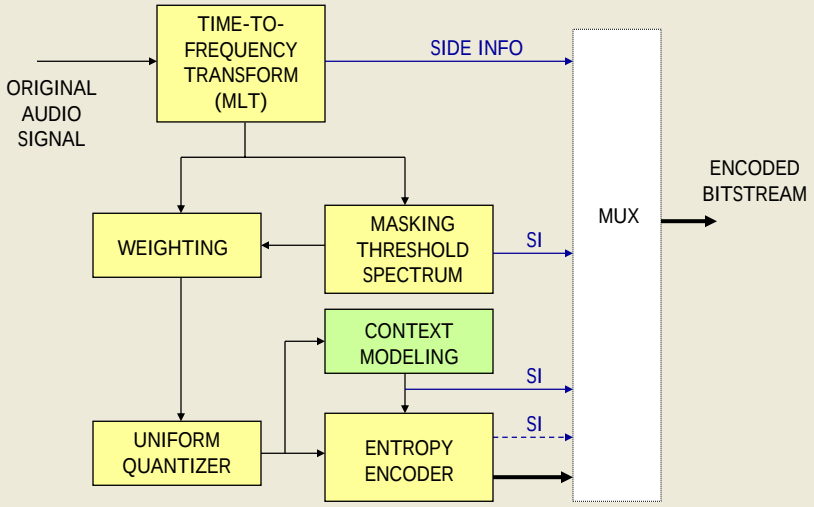
25

Side information & more about EC

- SI: model of frequency spectrum
  - e.g. averages over subbands
- Quantized spectral model determines weighting
  - masking level used to scale coefficients
- Backward adaptation reduces need for SI
- Run-length + Vector Huffman works
  - Context-based AC can be better
  - Room for better context models via machine learning?

26

Improved architecture



27

Examples of context modeling

- For strongly voiced segments, spectral energies may be well predicted by a “Linear Prediction” model, similar to those used in VoIP coders.
- For strongly periodic components, spectral energies may be predicted by a pitch model.
- For noisy segments, a noise-only model may allow for very coarse quantization → lower data rate.

28

Other aspects & directions

- Stereo coding
  - (L+R)/2 & L-R coding, expandable to multichannel
  - Intensity + balance coding
  - Mode switching – extra work for encoder only
- Lossless coding
  - Easily achievable via integer transforms
  - exactly reversible via integer arithmetic
  - example: lifting-based MLT (see Refs)
- Using complex subband decompositions (MCLT)
  - Potential for more sophisticated auditory models
  - Efficient encoding is an open problem

29

Contents

- Motivation
- “Sink coding”: Auditory Masking
- Block & Lapped Transforms
- Audio compression
- Examples

30

WMA examples:

- Original clip (~1,400 kbps) 
- 64 kbps (MP3) 
- 64 kbps (WMA) 
- Original clip 
- WMA @ 32 kbps (Internet radio) 
- More examples at windowsmedia.com

31

References

- T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, pp. 451–513, Apr. 2000.
  - Available at http://www.eas.asu.edu/~spanias/papers.html
- H. S. Malvar, "Auditory Masking in Audio Compression," chapter in *Audio Anecdotes*, K. Greenebaum, Ed., A. K. Peters Ltd., 2004.
- H. S. Malvar, "Fast Algorithms for Orthogonal and Biorthogonal Modulated Lapped Transforms," *IEEE Symposium Advances Digital Filtering and Signal Processing*, Victoria, Canada, pp. 159–163, June 1998.
- H. S. Malvar, "Enhancing the performance of subband audio coders for speech signals," *IEEE International Symposium on Circuits and Systems*, Monterey, CA, vol.5, pp. 98–101, June 1998.
- H. S. Malvar, "A modulated complex lapped transform and its applications to audio processing," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, AZ, pp. 1421–1424, March 1999.
- J. Li, "Reversible FFT and MDCT via matrix lifting," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Canada, pp. IV-173–176, May 2004.
- H. S. Malvar, "Adaptive run-length/Golomb-Rice encoding of quantized generalized Gaussian sources with unknown statistics," *IEEE Data Compression Conference*, Snowbird, UT, March 2006.

32