Parametric Coding of High-Quality Audio

Prof. Dr. Gerald Schuller
Fraunhofer IDMT & Ilmenau Technical University
Ilmenau, Germany
Waveform vs Parametric

• Waveform
  – Filter-bank approach
  – Mainly exploits limitations of human auditory system
  – Mature technology

• Parametric
  – Source model approach
  – Exploits both source model as well as limitations of human auditory system

→ most audio coders use a combination of both
1. Perceptual Noise Substitution (PNS)
Perceptual Noise Substitution (1)

Background:
- Parametric coding of signals gives a very compact signal representation
- Parametric coding of noise-like signal components has been used widely e.g. in speech coding
- Can similar techniques be used in perceptual audio coding?

MPEG-4:
- Perceptual Noise Substitution (PNS) permits a frequency selective parametric coding of noise-like signal components
Perceptual Noise Substitution (2)

"Perceptual Noise Substitution" (PNS):
Perceptual coder + parametric representation of noise-like signals

Diagram:
- Encoder
  - Audio Input
  - Analysis Filterbank
  - Quantization & Coding
  - Bitstream Multiplexer
  - Noise Detection
  - Substituted signal energies

- Decoder
  - Synthesis Filterbank
  - Inverse Quantization
  - Bitstream Demultiplexer
  - Noise Generator
  - Noise subst. signaling
  - Substituted signal energies

- Bitstream Out
- Bitstream In
Perceptual Noise Substitution (3)

Principle:
• Noise-like signal components are detected on a scalefactor band basis
• Corresponding groups of spectral coefficients are excluded from quantization/coding
• Instead, only a "noise substitution flag" plus total power of the substituted band is transmitted in the bitstream
• Decoder inserts pseudo random vectors with desired target power as spectral coefficients

-> Highly compact representation for noise-like spectral components
2. Spectral Band Replication

Bandwidth Extension (1)

Background

• Audio coding at very low bitrates ⇒ artifacts
• To avoid excessive artifacts, bandwidth is usually sacrificed at low bitrates (<40kbit/s/ch) ⇒ Signal sounds unattractive (muffled)

Concept

• “re-generate” HF signal content at decoder end from LF part (and some helper information)
## Bandwidth Extension (2)

### Idea of Spectral Band Replication (SBR)
- Re-generate HF-part of signal spectrum by means of *transposition* of transmitted spectrum
  ⇒ ensures preservation of harmonic structure
- Subsequent shaping of signal towards original time/spectral envelope by an adaptive filter (*envelope adjuster*)
- Some more provisions for handling special situations
- SBR bitstream elements (ca. 2 kbit/s/ch) can be stored in AAC bitstream in a compatible way
  - Standard AAC decoders decode AAC part only

### Compatibility
- “MPEG-4 Audio Extension #1”
- SBR is used e.g. in High-Efficiency AAC (aacPlus) and MP3Pro
Bandwidth Extension (3)

Spectral Band Replication Scheme ( Principle)
Bandwidth Extension (4)

Figure 1: Spectrum and Masking Threshold

Dietz e.a., „Spectral Band Replication, a novel approach in audio coding“
Bandwidth Extension (5)

Figure 2: Ideal Perceptual Coding

Dietz e.a, „Spectral Band Replication, a novel approach in audio coding“
Bandwidth Extension (6)

Figure 4: Waveform coding beyond its limits

Dietz e.a., „Spectral Band Replication, a novel approach in audio coding“
Bandwidth Extension (7)

![Energy vs Frequency Graph]

**Figure 5: Limiting the audio bandwidth**

Dietz e.a., „Spectral Band Replication, a novel approach in audio coding“
Bandwidth Extension (8)

Figure 6: High frequency generation based on the waveform coded low frequency part

Dietz e.a., „Spectral Band Replication, a novel approach in audio coding“
Bandwidth Extension (9)

Figure 7: Spectrum after high frequency adjustment

Dietz e.a, „Spectral Band Replication, a novel approach in audio coding“
Audio Coding for Communication Applications

Prof. Dr. Gerald Schuller

Fraunhofer IDMT & Ilmenau University of Technology
Introduction (1)

- Audio coding/compression is used in:
  - Multimedia applications
  - Storage
  - Digital broadcast: Digital radio
    - Sirius, XM-Radio, iBiquity, DAB
- Examples:
  - MP3 (MPEG-1 Layer 3)
  - MPEG-2/4
  - AAC
  - AC-3
  - PAC
Introduction (2)

- New networks:
  - Higher rate wireless services (for instance with space-time block coding)
  - Quality of service (->low delay)
  - In-home or local networks
New Communication Applications

- High quality teleconferencing

- Reporting for radio or TV stations, using wireless networks

- → Delay wise most critical:
  - Virtual presence (for instance musicians playing together over long distance)
  - Concerts with wireless microphones and speakers (data compression for transmitted power and bandwidth)
  - Desired delay < 10 ms (Ultra Low Delay)
Goals

- Audio coding for communications:
  - Ultra low encoding/decoding delay (<10ms)
  - High quality for music and speech
  - Bit-rates about 50-100 kb/s
Problems

- Conventional audio coders:
  - Good audio quality
  - But very high encoding/decoding delay (>100 ms)

- Speech coders:
  - Low encoding/decoding delay (order of 10…50ms), suitable for communications applications
  - But not high quality audio, don’t perform well on non-speech signals like music or room noise
Conventional Audio Coders

Audio Input → Analysis Filter Bank → Q → Coder → Synthesis Filter Bank → Decoded Signal

Psycho-acoustic model
Compression

• Irrelevance (Psycho Acoustics)
  – What the receiver (ear) cannot detect
  – Sound below the threshold of hearing
  – In general sound below the psycho-acoustic “masking threshold”

• Redundancy
  – The predictability or statistical dependencies in a signal
Basics of Psychoacoustics (Irrelevance)

- Temporal masking threshold

- Spectral masking threshold
Major Sources of Delay

- System delay of analysis and synthesis filter bank
- Buffering for bit-rate smoothing
- Previous Approaches:
  - Low delay filter banks
  - MPEG-4 low delay coder: reduced number of subbands
Limitations of previous Approaches

• High coding gain requires high numbers of subbands
  
  − Low delay filter banks: delay lower bounded by downsampling factor (= number of bands)
  
  − MPEG 4 low delay coder: reduced coding efficiency (higher bit-rate), not very low delay (ca. 30 ms at 32 kHz sampling).
  
  − Reason: subband coding leads to trade-off between coding efficiency and delay.
New Approach

• Subband coding has same asymptotic gain as predictive coding (Jayant, Noll, 1984; Nitadori, 1970).

• But predictive coding has lower delay → Replace filter bank by predictor
How to apply predictive Coding

- Problem: output of psycho-acoustic model is a time/frequency description.

- Approach:
  - Separate stages for application of irrelevance (psycho-acoustics) and redundancy reduction
  - Apply psycho-acoustic quantization noise shaping with linear filters (irrelevance red.)
  - Use lossless predictive coding after quantization (redundancy reduction)
Pre- and Post-Filter Approach

Encoder

full band connection

Decoder

Audio Input

Pre-Filter

Q

Lossless Coding

Lossless Decoding

Post-filter

Audio Output

Psycho-acoustic model

Irrelevance Reduction

Redundancy Reduction
Function

- Pre- and post-filters form the quantization noise over frequency and time
- Post-filter is inverse of pre-filter
- Pre-filter normalizes the signal to its psychoacoustic masking threshold
- Simple uniform constant step size quantizer is used (rounding operation)
- Added benefit: more precise control over quantization noise shape than conventional approach
Pre-Filter Structure

Short delay for synchronization with psycho-acoustic model
(our implementation: 128 samples + 128 samples blocking delay)
Properties

• Disadvantage:
  – Computationally complex structure

• But advantage:
  – No inherent delay, suitable for communications applications
Example Frequency Response

Spectrum of the signal (i) compared to masking threshold (ii) and freq. resp. of post-filter (iii)
Redundancy Reduction

- Irrelevance removed after pre-filter and quantizer
  - Lossless compression needed for perceptual lossless coding

Audio

Irrelevance Reduction

Pre-filter and quantization

Redundancy Reduction

Lossless coding or compression

Coded Audio
Redundancy Reduction

- Irrelevance removed after pre-filter and quantizer
  - Lossless compression needed for perceptual lossless coding

Audio → Irrelevance Reduction → Redundancy Reduction → Coded Audio

Pre-filter and quantization

Lossless coding or compression
Lossless Coding Unit

- **Goals:**
  - high compression ratios
  - low delay

- **Previous approaches:**
  - General purpose or text: Lempel-Ziv, PPMZ
  - Audio: Shorten (Softsound, GB), LPAC, LTAC
    (TU-Berlin, Germany), WaveZip
    (Soundspace, CA), MLP (Meridian, GB, for DVD)
Lossless Coding

• Previous approaches:
  – Made for file compression
  – Based on forward (block based) prediction, or transforms

• Problems:
  – High encoding/decoding delay
  – Compression ratio can be improved

• New approach:
  – Backward adaptation (based on past) for low delay
  – Cascading predictors for improved compression
Lossless Predictive Coding – Encoder

For backward adaptation: Predictor coefficient vector $\mathbf{h}$ updated with LMS algorithm.
Lossless Predictive Coding – Decoder

Observe: Quantization / rounding of predicted value does not affect lossless property
Low Coding Delay

- Backward adaptation with LMS algorithm
  - Define a vector of input samples:
    \[ \mathbf{x}^T(n) = [x(n - L + 1),...,x(n)] \]
  - The predicted value
    \[ [P(n)] = \text{round}(\mathbf{x}^T(n-1)\mathbf{h}(n)) \]
  - Update of the predictor coefficient vector \( \mathbf{h} \) with normalized LMS (Widrow, Hoff, 1960).

Prediction error:
\[ e(n) = x(n) - [P(n)] \]

\[ \mathbf{h}(n+1) = \mathbf{h}(n) + \frac{e(n)}{1 + \lambda \| \mathbf{x}(n) \|^2} \mathbf{x}(n) \]
Increased Compression Ratio

- Cascading LMS predictors, using the final output, has advantages:
  - Increased adaptation speed
  - Improved prediction accuracy
  - Better numerical stability (see Prandoni, Vetterli, 1998, for a special case)
Cascading and Combining Predictors

• For us important: Availability of predictors of different orders as additional outputs.
Reasons:
  – Very non-stationary signals (attacks) require fast adaptation/ short filters
  – Stationary signals require long filters

• Approach:
  – Combine predictors of different orders adaptively (analog to block switching)
Combination of Predictors (1)

• Assume predictors $P_1$, $P_2$, $P_3$ with different orders for different signal statistics. How can they be combined?

• Use predictive minimum description length principle for the “optimal” combination of predictors

\[ P = \sum_{i=1}^{3} w_i P_i \]

$w_i$ : probability of $P_i$ being “correct” on past signal.
Combination of Predictors (2)

- Assume that the prediction error has a Laplacian distribution. Then the weights $w_i$ are:

$$w_i \propto e^{-c \sum_n |x(n) - P_i(n)|}$$

- Weights “reward” predictors with good past performance
Weighted Cascaded LMS (WCLMS) Prediction

The \( w \)'s are adapted based on previous prediction errors, \( 0 < w < 1 \), to adapt to signal statistics (orders: 120, 80, 40)
Entropy Coding of Residuals

• Take known algorithms, e.g.:
  – Adaptive Golomb-Rice Codes
  – Adaptive Arithmetic Coding
  – Block based Huffman using pre-calculated code books

• No additional delay introduced, because ULD implementation already is block-based (128 samples)

• Inherently variable bit rate
Comparison of different lossless compression schemes

- Signals are at 32 khz sampling rate, bit-rate in bit/sample. Application after pre-filter and quantization.

<table>
<thead>
<tr>
<th>Signal</th>
<th>Cascaded LMS</th>
<th>Shorten</th>
<th>Wavezip</th>
<th>LPAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop</td>
<td>1.94</td>
<td>2.52</td>
<td>3.22</td>
<td>2.23</td>
</tr>
<tr>
<td>Jazz</td>
<td>1.99</td>
<td>2.67</td>
<td>3.35</td>
<td>2.48</td>
</tr>
<tr>
<td>mixed</td>
<td>2.16</td>
<td>2.58</td>
<td>3.19</td>
<td>2.35</td>
</tr>
<tr>
<td>Speech</td>
<td>1.96</td>
<td>2.48</td>
<td>3.09</td>
<td>2.12</td>
</tr>
</tbody>
</table>
Controlling the Bit-Rate: constant bit rate mode

- A factor of less than 1 leads to quantization noise above threshold of audibility, but a reduced bit-rate.

- By adapting the attenuation factor and iterating the lossless coder a target bit rate can be approximated.
Alternative Predictor Structure

- Closed-Loop Predictor instead of Open-Loop Predictor
- Advantage: only one quantizer
- Disadvantage: Quantization and Prediction not separated anymore (Irrelevance and Redundancy Reduction)
Conclusions

• Predictive coding can be used to obtain Ultra Low Delay audio coders

• Obtained delay: 6 ms (<10 ms)

• Subjective audio quality comparable to conventional high delay coder at same bitrate

• Price: higher complexity

• Can also be used to obtain high audio quality (unlike speech coders)