Enhanced Waveform Interpolative Coding at 4 kbps

Oded Gottesman, and Allen Gersho

Signal Compression Lab.
University of California, Santa Barbara
E-mail: [oded, gersho]@scl.ece.ucsb.edu
http://scl.ece.ucsb.edu/current/oded
Outline

• Description of Waveform-Interpolative coder
• Our WI Coder’s Novel Techniques
  – AbS SEW VQ
  – AbS Dispersion-Phase VQ
  – Pitch search for transitions
  – AbS Switched-Predictive Gain VQ
• Subjective Tests
• Summary
• Demo
Outline

• Description of Waveform-Interpolative coder
  • Our WI Coder’s Novel Techniques
    – AbS SEW VQ
    – AbS Dispersion-Phase VQ
    – Pitch search for transitions
    – AbS Switched-Predictive Gain VQ
  • Subjective Tests
  • Summary
  • Demo
Waveform Interpolation (WI)

• Voiced speech is nearly periodic. Succeeding pitch cycles, have a **slowly evolving shape**.

• A continuously evolving sequence of pitch cycle waveforms can be generated.

• Extract a subsequence of these waveforms (“subsampling”) for quantization, then synthesis of speech can be performed by interpolating missing waveforms (“upsampling”).
Waveform Decomposition

- *Slowly Evolving Waveform* (SEW) - voiced component, is coded:
  - (1) at low temporal resolution,
  - (2) high spectral resolution,
  - (3) with spectral masking distortion.

- *Rapidly Evolving Waveform* (REW) - unvoiced component, is coded:
  - (1) at high temporal,
  - (2) low spectral resolution,
  - (3) with spectral and temporal masking.

- The system is universal for all speech sounds (no V/UV classification needed)
Waveform Decomposition (cont’d)

• A speech segment typically contains both voiced and unvoiced attributes.
• Different perceived character of voiced and unvoiced components suggests a separation of the components and distinct perceptually based coding.
Waveform Extraction

*From Kleijn and Haagen, ICASSP95*
Waveform Decomposition

* From Kleijn and Haagen, ICASSP95
Waveform Synthesis

\[ \phi(t) = \phi(t_m) + \int_{t_m}^{t} \frac{2\pi}{p(\tau)} d\tau \]

*From Kleijn and Haagen, ICASSP95*
WI Decoder

Signal Compression Lab., University of California, Santa Barbara
Continuous Waveform Interpolation

Over the interpolation interval \( t_m \leq t \leq t_{m+1} \), the continuous reconstructed excitation signal is given by the time dependent Fourier series:

\[
 r(t) = \sum_{k=0}^{K(t)} \left\{ \left[ (1 - \alpha(t)) A_k(t_m) + \alpha(t) A_k(t_{m+1}) \right] \cos(k\phi(t)) + \right. \\
\left. \left[ (1 - \alpha(t)) B_k(t_m) + \alpha(t) B_k(t_{m+1}) \right] \sin(k\phi(t)) \right\}
\]

\[
 \phi(t) = \phi(t_m) + \int_{t_m}^{t} \frac{2\pi}{p(\tau)} d\tau
\]

where \( \phi(t) \), is the instantaneous phase contour, and \( \alpha(t) \) is some (increasing) interpolation function in the range \( 0 \leq \alpha(t) \leq 1 \), which can be a simple linear function of the time, \( t \), or of the pitch-normalized phase \( \phi \).

Signal Compression Lab., University of California, Santa Barbara
Outline

• Description of Waveform-Interpolative coder
• Our WI Coder’s Novel Techniques
  – AbS SEW VQ
  – AbS Dispersion-Phase VQ
  – Pitch search for transitions
  – AbS Switched-Predictive Gain VQ
• Subjective Tests
• Summary
• Demo
SEW Quantization

- SEW is downsampled - quantized - upsampled
- Non-ideal filters used: aliasing + mirroring
- Distortion most notable in the transitions

- How can we improve quantization?
SEW Filter Bank

CW

$C_0(t)$ → LPF $h(t)$
Anti-Aliasing (Decomposition)
$\downarrow M$

SEW

$C_1(t)$ → LPF $h(t)$
Anti-Aliasing (Decomposition)
$\downarrow M$

$C_K(t)$ → LPF $h(t)$
Anti-Aliasing (Decomposition)
$\downarrow M$

Quantized SEW

$e^{j\omega_0 t}$

SEW excitation

$e^{j\omega_0 Kt}$

Signal Compression Lab., University of California, Santa Barbara

15
Waveform Based SEW AbS

$M$ - Number of waveforms per frame
$L$ - Number of lookahead waveforms
SEW Optimization Examples

Original

Optimized

Non-optimized

Time (sec)

Amplitude

Original

Optimized

Non-optimized

Time (sec)

Amplitude
Outline

• Description of Waveform-Interpolative coder
• Our WI Coder’s Novel Techniques
  – AbS SEW VQ
  – AbS Dispersion-Phase VQ
  – Pitch search for transitions
  – AbS Switched-Predictive Gain VQ
• Subjective Tests
• Summary
• Demo
Phase Quantization

• Observations:
  – Phase is of secondary perceptual significance
  – No efficient phase quantization scheme is known.

• Two extremes:
  – Waveform coders (CELP) implicitly allocate a perceptually-excessive number of bits to the phase.
  – In parametric coders the phase information is commonly not transmitted:

• How can the phase be quantized efficiently?
Dispersion Phase $\phi$

• Dispersion Phase:
  – Pitch cycle extracted from residual signal
  – cyclically shifted such that its pulse is located at position zero.
  – resulting DFT phase: $\phi$

• $\phi$ determines (along with the magnitude), the waveform’s pulse shape.
AbS Phase Dispersion Quantization

Pitch-Cycle Waveform’s DFT → Crude Linear-Phase Alignment

Magnitude Codebook → | r | → X

Phase Codebook → \( e^{j\hat{\Phi}} \)

Refined Linear-Phase Alignment

\( \frac{W(z)}{A(z)} \)

\( \min \| r \|^2 \)
Vector Quantization Design

- Variable-Dimension VQ
  - Eight pitch-range dependent codebooks
- Pitch changes over time cause the quantizer to switch among the pitch-range codebooks.
  - Solution: overlapped training clusters used, to achieve smooth phase variations whenever such switch occurs.
- 0-6 bit codebooks were designed and tested.
Vector Quantization Design (cont’d)

Range for codebook #1

Range for codebook #7

Range for codebook #8

Cluster 1

Cluster 2

Cluster 3

Cluster 4

Cluster 5

Cluster 6

Cluster 7

Cluster 8

Number of harmonics
(vector dimension)

K

10 20 30 40 50 60 70

Range for codebook #8
Range for codebook #7
Range for codebook #1

Signal Compression Lab., University of California, Santa Barbara
Phase Code-Vector Examples
Segmental Weighted SNR of Phase VQ

Seg. Weighted SNR: dB

Phase Bits

Non-MIRS (Flat)

MIRS

Signal Compression Lab., University of California, Santa Barbara
Subjective Results

Results of subjective A/B test for comparison between the 4-bit phase VQ, and male extracted fixed phase.
Outline

• Description of Waveform-Interpolative coder
• Our WI Coder’s Novel Techniques
  – AbS SEW VQ
  – AbS Dispersion-Phase VQ
  – Pitch search for transitions
  – AbS Switched-Predictive Gain VQ
• Subjective Tests
• Summary
• Demo
**Pitch Search**

- Pitch search robustness needed for:
  - high and low pitch
  - transitions
  - segments with rapidly varying pitch

- Our suggested solutions:
  - Combined temporal/spectral domain search
  - Higher rate temporal domain search
  - Varying boundaries based search
  - Average pitch - using normalized correlation as weighting
Pitch Search Algorithm

Speech

Spectral domain pitch search + tracker

100 Hz

Good Pitch?

Yes

Temporal domain pitch refinement

500 Hz

No

Weighted speech

Temporal domain pitch search

Good Pitches?

No

Yes

500 Hz

Use 4 ms waveform length

100 Hz

Weighted-Average Pitch

Yes

No
Outline

• Description of Waveform-Interpolative coder
• Our WI Coder’s Novel Techniques
  – AbS SEW VQ
  – AbS Dispersion-Phase VQ
  – Pitch search for transitions
  – AbS Switched-Predictive Gain VQ
• Subjective Tests
• Summary
• Demo
Switched-Predictive Gain VQ using Temporal Weighting

• VQ of Log-Gain (no gain down/up-sampling).
• Temporal weighting used for quantization avoids smear of plosives and onsets.
• Switched-Predictive VQ allows different gain predictors, and reduces outliers.
• Switched DC levels improve performance.
Switched-Predictive AbS

Gain VQ

Log-Gain

\[ g(m) \]

\[ D_i \]

\[ P_i \]

\[ c_{ij}(m) \]

\[ t(m) \]

DC Codebook

Predictor Codebook

Vector Quantizer Codebook

Synthesis Filter

\[ \frac{1}{1 - P_i z^{-1}} \]

Temporal Weighting

\[ \min \| * \|^2 \]
### Bit Allocation of EWI Coder

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bits / Frame</th>
<th>Bits / second</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPC</td>
<td>18</td>
<td>900</td>
</tr>
<tr>
<td>Pitch</td>
<td>2 x 6 = 12</td>
<td>600</td>
</tr>
<tr>
<td>Gain</td>
<td>2 x 6 = 12</td>
<td>600</td>
</tr>
<tr>
<td>REW</td>
<td>20</td>
<td>1000</td>
</tr>
<tr>
<td>SEW magnitude</td>
<td>14</td>
<td>700</td>
</tr>
<tr>
<td>SEW phase</td>
<td>4</td>
<td>200</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>80</strong></td>
<td><strong>4000</strong></td>
</tr>
</tbody>
</table>

- Frame size: 20 ms
- Two subframes per frame
Outline

• Description of Waveform-Interpolative coder
• Our WI Coder’s Novel Techniques
  – AbS SEW VQ
  – AbS Dispersion-Phase VQ
  – Pitch search for transitions
  – AbS Switched-Predictive Gain VQ

• Subjective Tests
• Summary
• Demo
A/B subjective Test

- 14 Listeners
- 12 male + 12 female sentences.
- MIRS filtered speech.

<table>
<thead>
<tr>
<th>Test</th>
<th>4 kbps WI</th>
<th>4 kbps MPEG-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>65%</td>
<td>35%</td>
</tr>
<tr>
<td>Male</td>
<td>62%</td>
<td>38%</td>
</tr>
<tr>
<td>Total</td>
<td>64%</td>
<td>36%</td>
</tr>
</tbody>
</table>

With 95% certainty the WI preference lies in [59%, 69%]
WI is preferred in 99.999984% certainty.
## A/B subjective Test (cont’d)

<table>
<thead>
<tr>
<th>Test</th>
<th>4 kbps WI</th>
<th>5.3 kbps G.723.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>58%</td>
<td>42%</td>
</tr>
<tr>
<td>Male</td>
<td>61%</td>
<td>39%</td>
</tr>
<tr>
<td>Total</td>
<td>60%</td>
<td>40%</td>
</tr>
</tbody>
</table>

With 95% certainty the WI preference lies in [54%, 65%]
WI is preferred in 99.98% certainty.

<table>
<thead>
<tr>
<th>Test</th>
<th>4 kbps WI</th>
<th>6.3 kbps G.723.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>55%</td>
<td>45%</td>
</tr>
<tr>
<td>Male</td>
<td>53%</td>
<td>47%</td>
</tr>
<tr>
<td>Total</td>
<td>54%</td>
<td>46%</td>
</tr>
</tbody>
</table>

With 95% certainty the WI preference lies in [49%, 59%]
WI is preferred in 92.21% certainty.
Summary

• AbS VQ for the SEW enhances coding efficiency.
• AbS VQ for the dispersion phase improves speech naturalness and waveform matching.
• New pitch search improves accuracy and robustness.
• Switched-Predictive Gain VQ avoids gain smear, and reduces outliers.
• Enhanced WI coder slightly exceeds quality of G.723.1 at 6.3 kbps.
References


Ericsson-Nokia Best Paper Award

for the paper

Enhanced Waveform Interpolative Coding at 4 kbps

by

Oded Gottesman
Allen Gersho

Tor Björn Minde
Ericsson

Jari Hagqvist
Nokia