Digital Audio Effects 2010

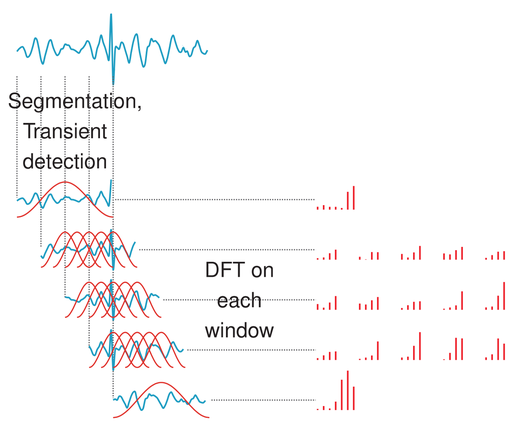
Improving RTISI Phase Estimation with Energy Order and Phase Unwrapping

Volker Gnann, Martin Spiertz
Institute of Communications Engineering / RWTH Aachen

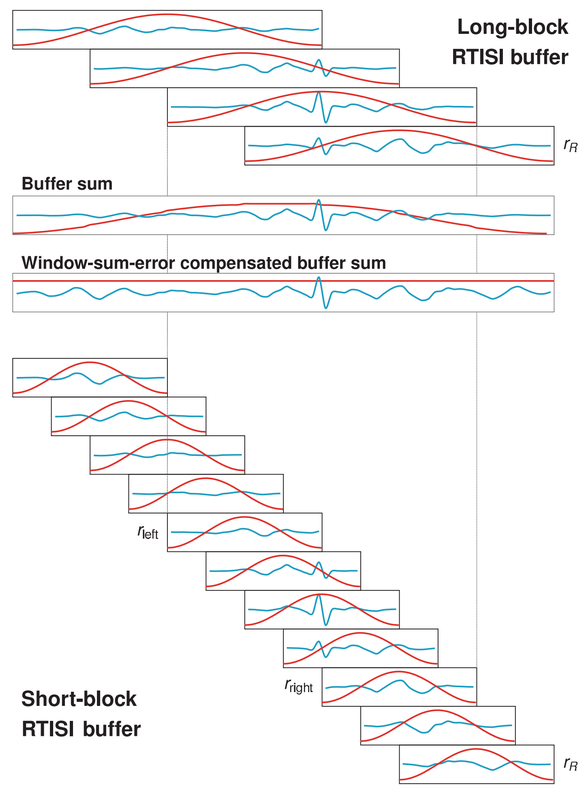
Motivation

- Phase estimation has a wide range of applications, like
  - time/pitch scale modification
  - source separation
  - comb-filter free audio mixing
- Reference algorithm: Real-Time Iterative Spectrogram Inversion (RTISI)
  - Online-capable version of Griffin/Lim algorithm
  - Iterative combination of buffer sum phase with target magnitude
  - Extension to dual window length (time/frequency resolution) available
- We present two RTISI improvements:
  - determination of the processing order by energy
  - initialization of the phase estimator by phase unwrapping

Dual-Resolution Spectrogram Generation



Dual-Resolution RTISI



Processing Order

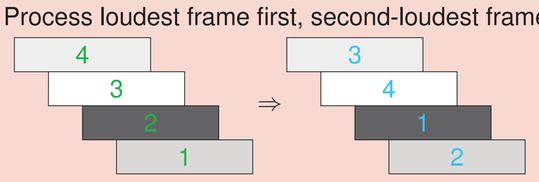
- Standard RTISI: Process last row, second-last row, ... to the row to commit.
- Disadvantage: Estimation of loud segments depends on result of previous and future quieter segments.

Improvement: Energy Order

- Determine processing order by energy:

$$\text{order} = \underset{i}{\text{argsort}} \left(-\sum_{n} (\boldsymbol{r}_i[n])^2 \right),$$

- Process loudest frame first, second-loudest frame next, ..., quietest frame last.



Standard Initialization

- Before phase estimation, the rows are filled with zeros.
- Phase is estimated by the window-compensated sum of previous rows

Improvement: Phase Unwrapping

- In steady-state signals, the phase for a frame can be derived from the difference of the previous frames (like in the phase vocoder).

$$\begin{aligned}\boldsymbol{r}_R &= A \cdot \text{IDFT}\left\{ |S_R[k]| \cdot e^{j(\angle S_{R-1}[k] + (\angle S_{R-1}[k] - \angle S_{R-2}[k]))} \right\} \\ &= A \cdot \text{IDFT}\left\{ |S_R[k]| \cdot e^{j(2\angle S_{R-1}[k] - \angle S_{R-2}[k])} \right\} \\ &= A \cdot \text{IDFT}\left\{ \frac{|S_R[k]| \cdot S^2_{R-1}[k] \cdot |S_{R-2}[k]|}{|S_{R-1}[k]|^2 \cdot S_{R-2}[k]} \right\}.\end{aligned}$$

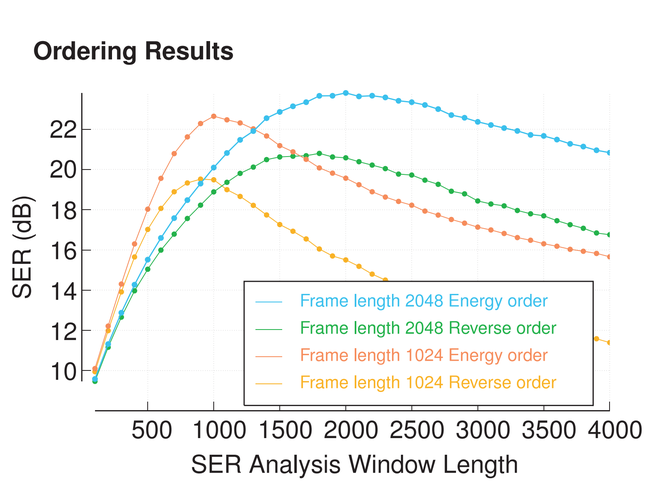
- A: weighting factor for initialization
- $A = 0$: standard initialization

Evaluation

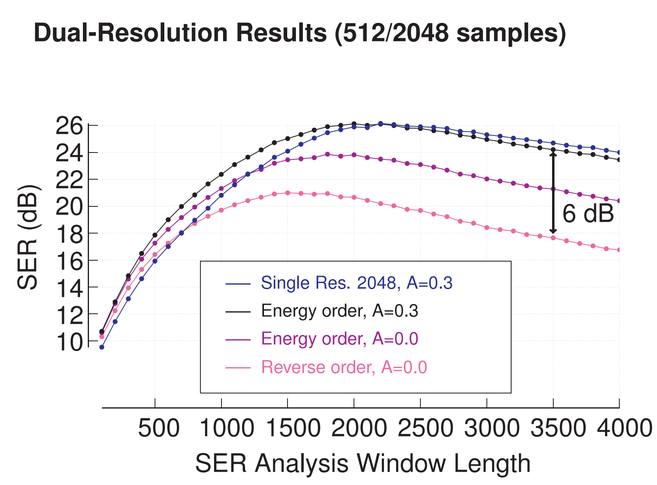
- EBU-SQAM test set
- 70 files with speech, singing vocals, and instruments
- Sampling freq. 48 kHz
- Hamming window, Overlap=75%
- Objective: Maximization of signal-to-error ratio (SER) for different SER analysis window lengths.

$$\text{SER} = 10 \log \frac{\sum_{m=-\infty}^{\infty} \sum_{k=0}^{L-1} |X[mS,k]|^2}{\sum_{m=-\infty}^{\infty} \sum_{k=0}^{L-1} (|X[mS,k]| - |X'[mS,k]|)^2}$$

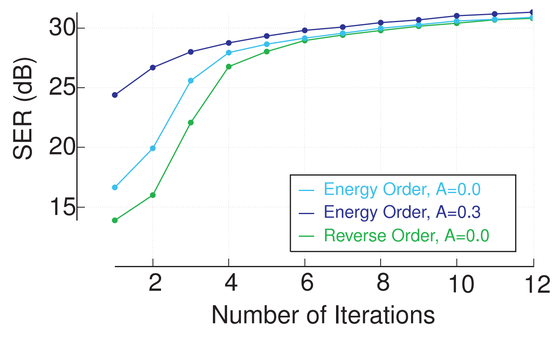
Ordering Results



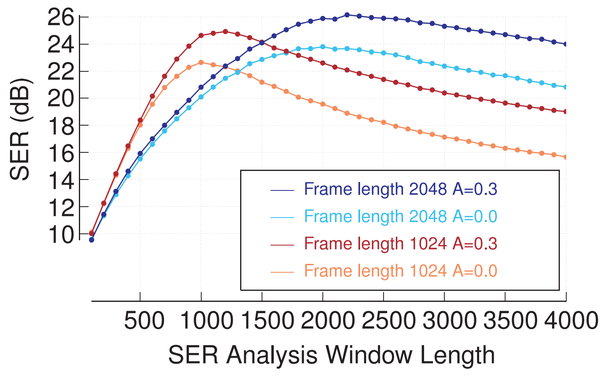
Dual-Resolution Results (512/2048 samples)



Number of Iterations



Phase Unwrapping Results



Conclusions

- Both methods lead to phase estimation improvements.
- With an increasing number of iterations, the improvements become smaller. The advantage of energy ordering diminishes completely.
- The combination of both improvements leads to an SER gain of up to 6 dB for dual-resolution RTISI.

{gnann, spiertz}@ient.rwth-aachen.de
www.ient.rwth-aachen.de
Institut für Nachrichtentechnik, Melatener Str. 23, 52072 Aachen, Germany
RWTH RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN