Comb-Filter Free Audio Mixing Using STFT Magnitude Spectra And Phase Estimation

Volker Gnann, Martin Spiertz

Institut für Nachrichtentechnik
RWTH Aachen University

Overview

- Audio Mixing and the Comb Filter Effect
- STFT Magnitude Mixing
- Phase Estimation
- Evaluation

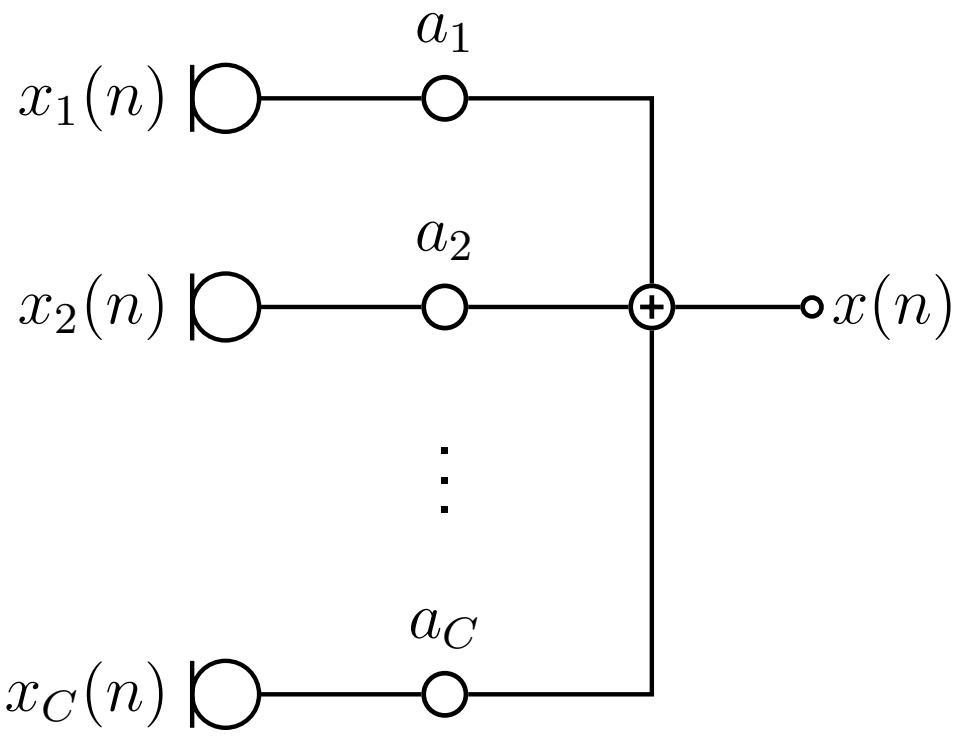
V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

Overview

- Audio Mixing and the Comb Filter Effect
- STFT Magnitude Mixing
- Phase Estimation
- Evaluation

V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

Audio Mixing

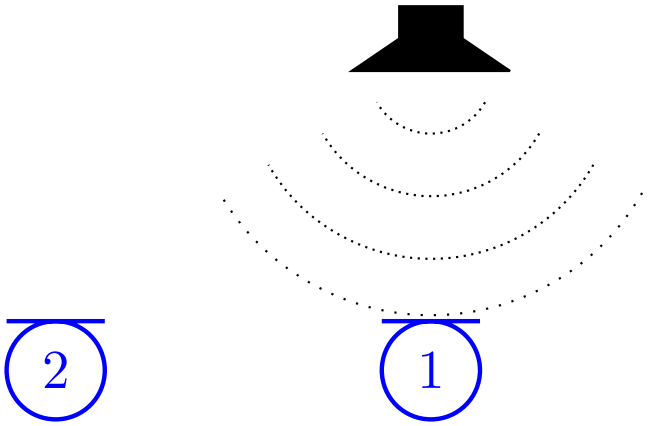


Audio mixing

output = linear combination of input

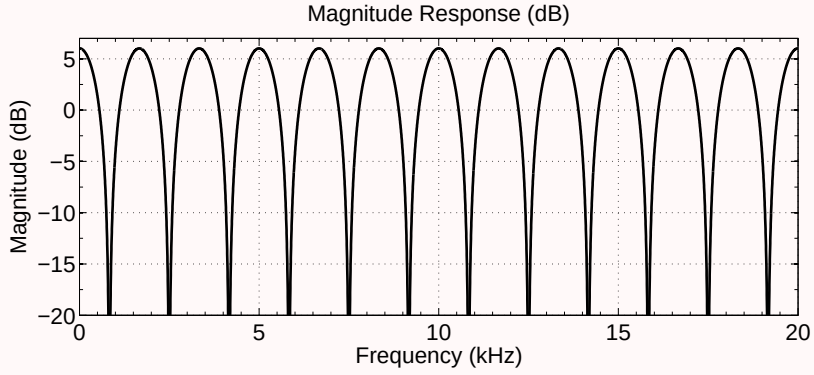
$$x(n) = \sum_{c=1}^{C} a_c x_c(n)$$

Comb Filter Distortions



- scenario: one source recorded with multiple mics
- delays between mics (wave propagation)
- mix = original + delayed version

Magnitude response vs frequency

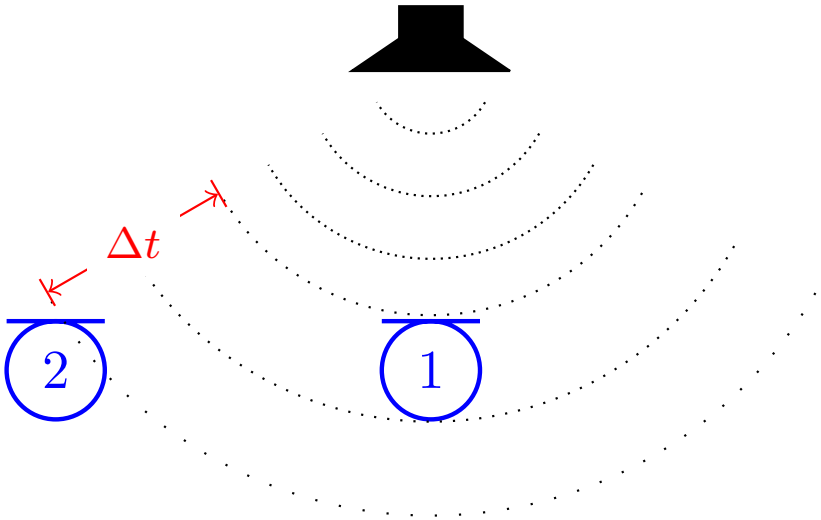


Listening example

- original ♫
- comb filter (1 ms delay) ♫

V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

Comb Filter Distortions

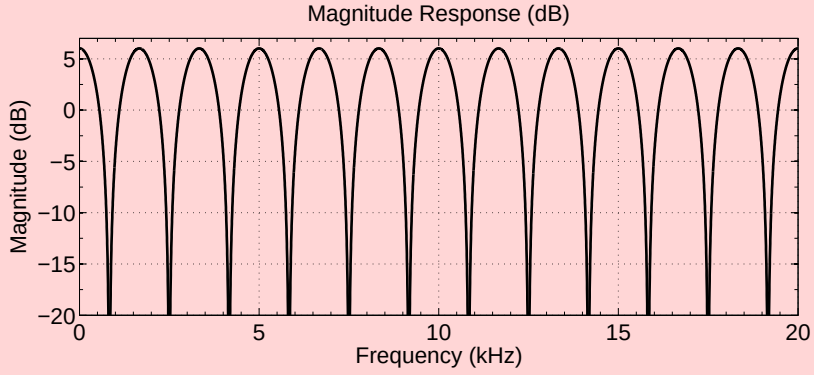


- scenario: one source recorded with multiple mics
- delays between mics (wave propagation)
- mix = original + delayed version

Listening example

- original ♫
- comb filter (1 ms delay) ♫

Magnitude response vs frequency



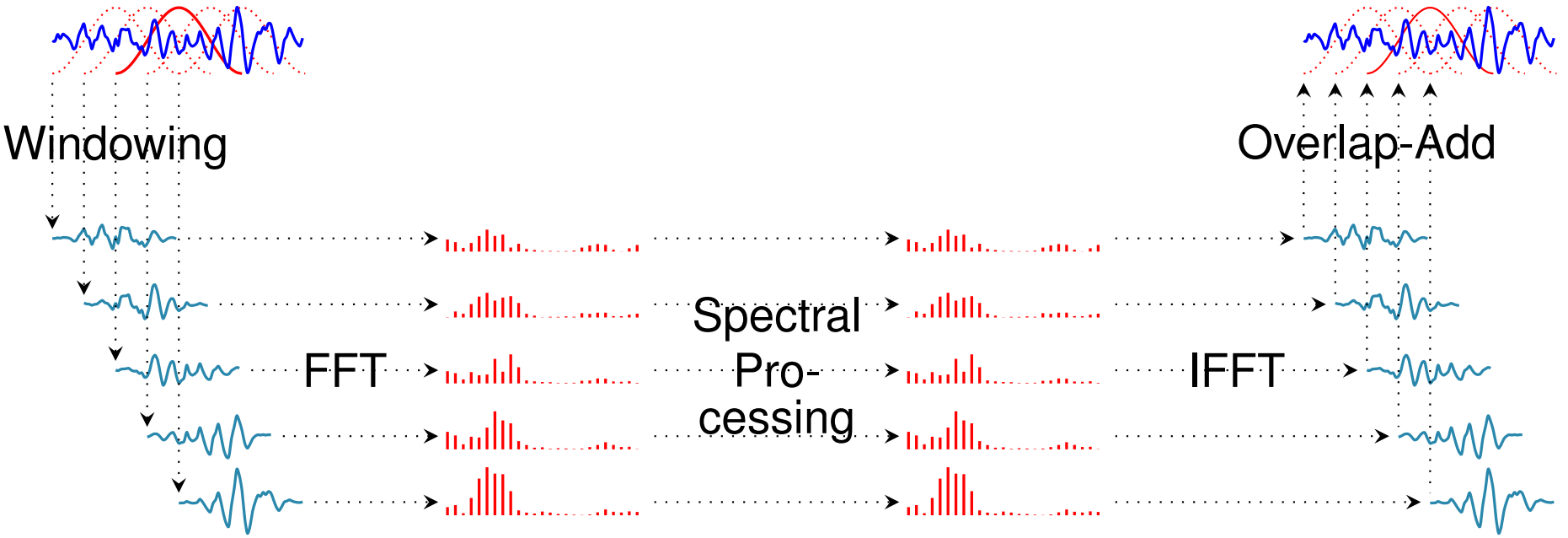
V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

Overview

- Audio Mixing and the Comb Filter Effect
- STFT Magnitude Mixing
- Phase Estimation
- Evaluation

V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

Review of STFT Processing

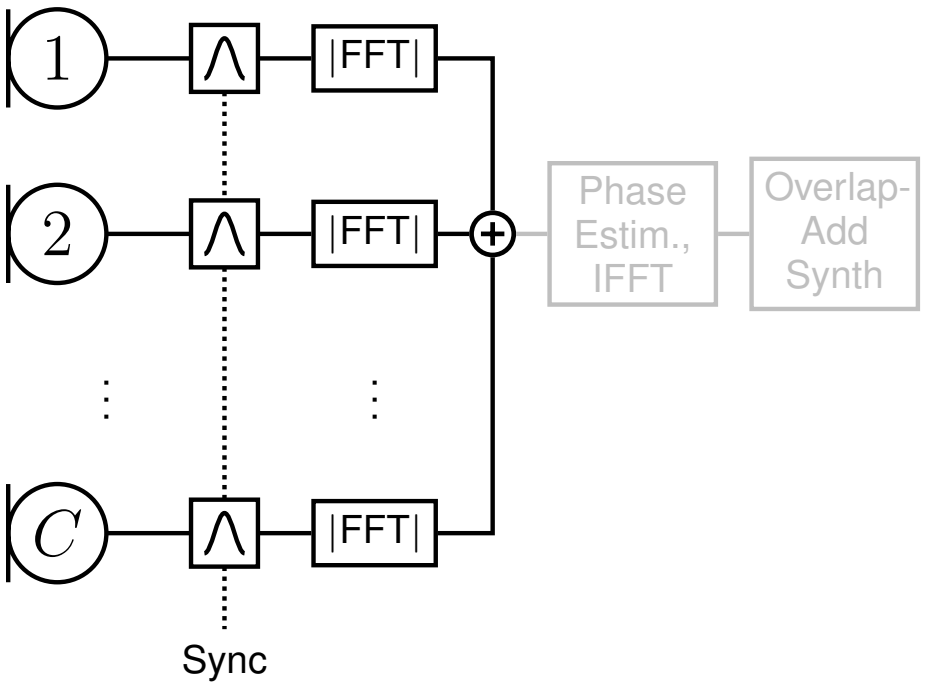


FFT coefficients are complex

- $C = |C| \cdot e^{j\varphi}$ with $|C|$ = magnitude, φ = phase
- reconstruction requires magnitude **and** phase information

V. Gnann, M. Spiertz | Institut für Nachrichtentechnik | RWTH Aachen University

STFT Magnitude Spectrum Mixing



Advantage

magnitude-only mixing $\Longrightarrow$
no phase cancellations $\Longrightarrow$
no comb filter effect

Disadvantage

magnitude-only mixing $\Longrightarrow$
phase information missing

Listening example

overlap-add on magnitude-only spectrum IFFT: ♫

V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

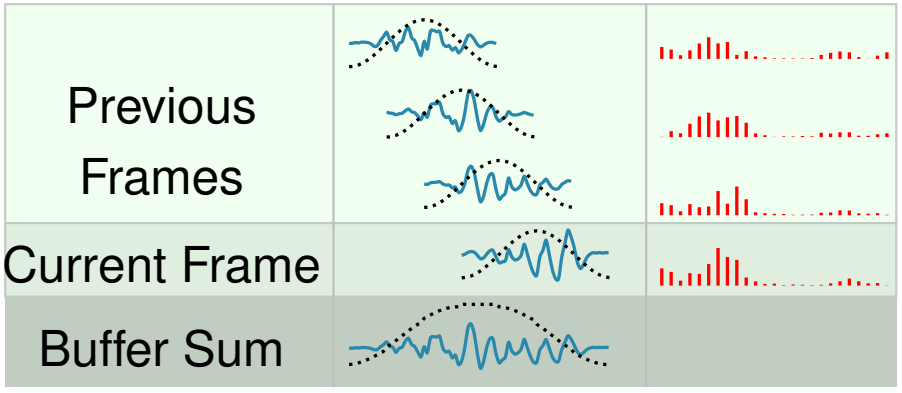
Overview

- Audio Mixing and the Comb Filter Effect
- STFT Magnitude Mixing
- Phase Estimation
- Evaluation

V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

Review: Phase Estimation with RTISI

RTISI = Real-Time Iterative Spectrogram Inversion (Zhu, Beauregard, Wyse, IEEE ASLP, 2007)



- consecutive frames stored in consecutive rows
- row content:
  - recent estimation of time domain signal
  - STFT magnitude spectrum (given by mixer)

Characterization of algorithm

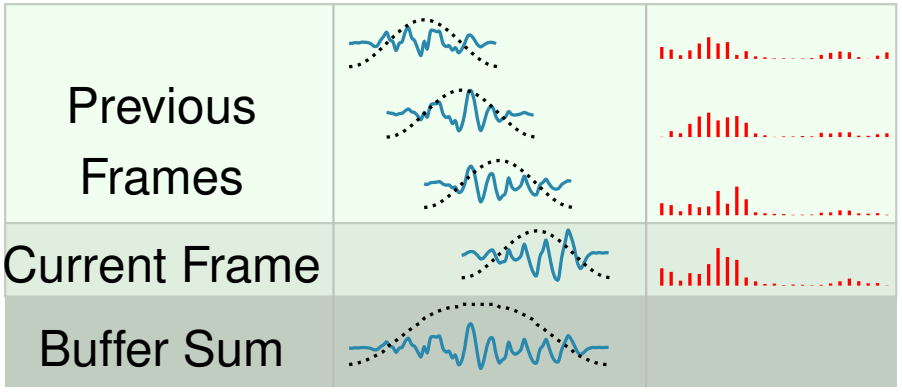
- estimation of the **last** buffer row (time domain).
- iterative combination of
  - the magnitude stored in the row
  - the phase from the buffer sum FFT
- result IFFT is windowed and written back

Improvement

look-ahead extension

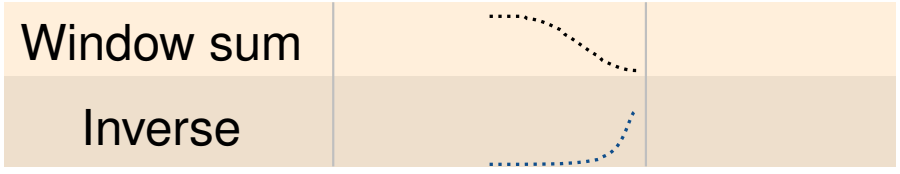
V. Gnann, M. Spiertz | Institut für Nachrichtentechnik | RWTH Aachen University

Window Sum Compensation



Window sum error

- buffer sum implicitly windowed with window sum (dotted line)
- mismatch analysis window
- FFT does not match given spectrograms even with correct signal

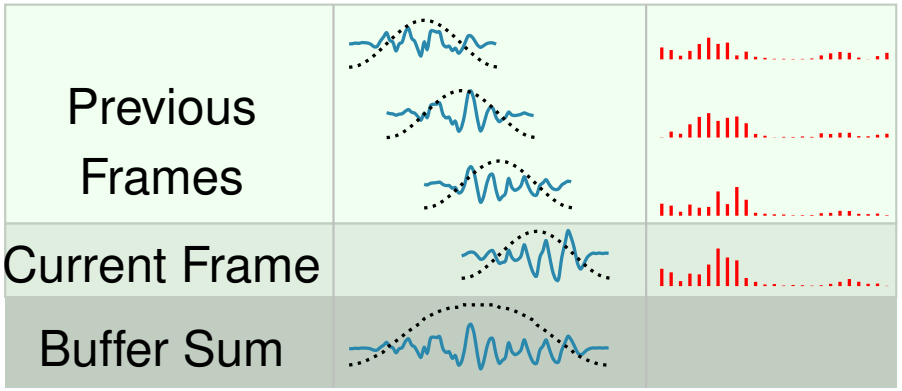


Compensation

- applying the window sum inverse to the buffer sum
- re-windowing the buffer sum with the Hamming window

V. Gnann, M. Spiertz | Institut für Nachrichtentechnik | RWTH Aachen University

RTISI Initialization for Mixing



Initialization

- iterative algorithm → approaching a local minimum
- result depending on the initial buffer values
- good initialization heuristic: **time domain mix**

Comb-Filter effect removal

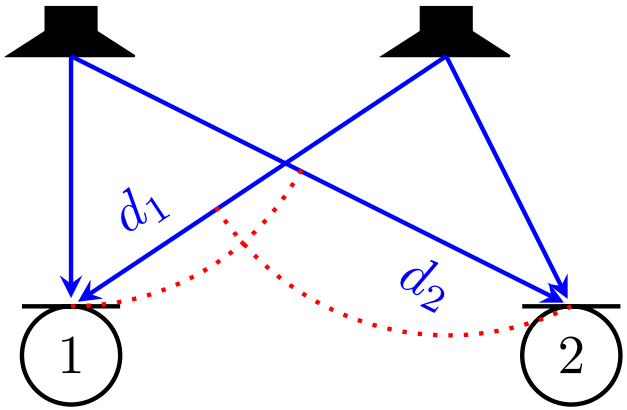
comb filter effects are removed **automatically** (combination with the comb-filter-free magnitude spectrum).

Overview

- Audio Mixing and the Comb Filter Effect
- STFT Magnitude Mixing
- Phase Estimation
- Evaluation

V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

Listening Examples



Recording simulation

- two instruments, two microphones
- delay differences d_1, d_2 between long and short distances

Listening examples

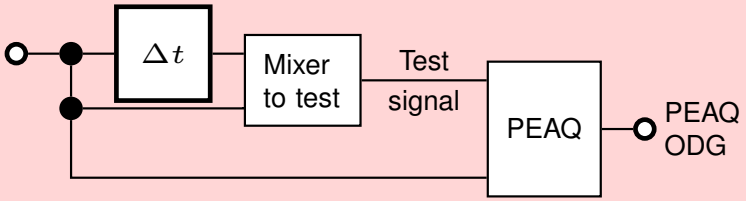
- Sources:
  - english horn
  - glockenspiel
- Delays:
  - d_1 (english horn) = 1 ms
  - d_2 (glockenspiel) = 0.3 ms
- reference = delay-free time-domain mix ♫
- time-domain mix with delays ♫
- rtisi mix with delays ♫

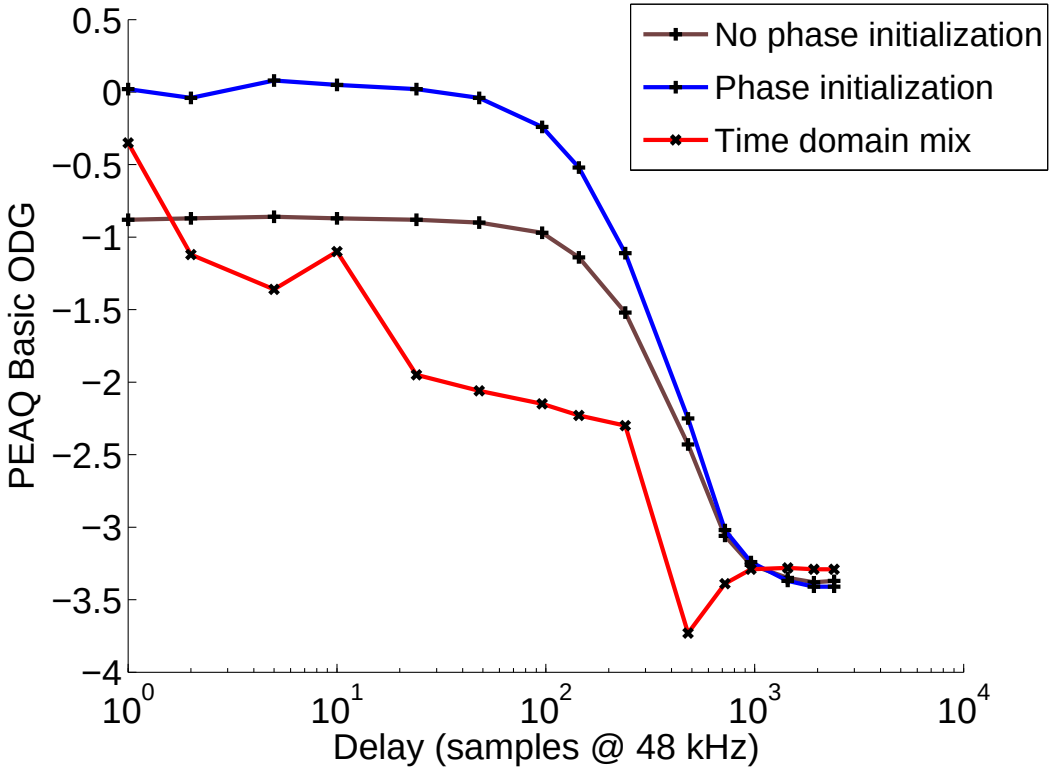
Evaluation with PEAQ

Source

- only single source
- german hip hop music, 15 s
- mono

Setup





V. Gnann, M. Spiertz Institut für Nachrichtentechnik RWTH Aachen University

Thank you for your attention!