Channel Decorrelation for Stereo Acoustic Echo Cancellation in High-Quality Audio Communication

Presented by: Jean-Marc Valin

12th December 2006

CSIRO

How to Corrupt an Audio Signal and Get Away With It

Presented by: Jean-Marc Valin

12th December 2006

Introduction

Context: acoustic echo cancellation with stereo signals

Problem: channels in a stereo signal are highly-correlated

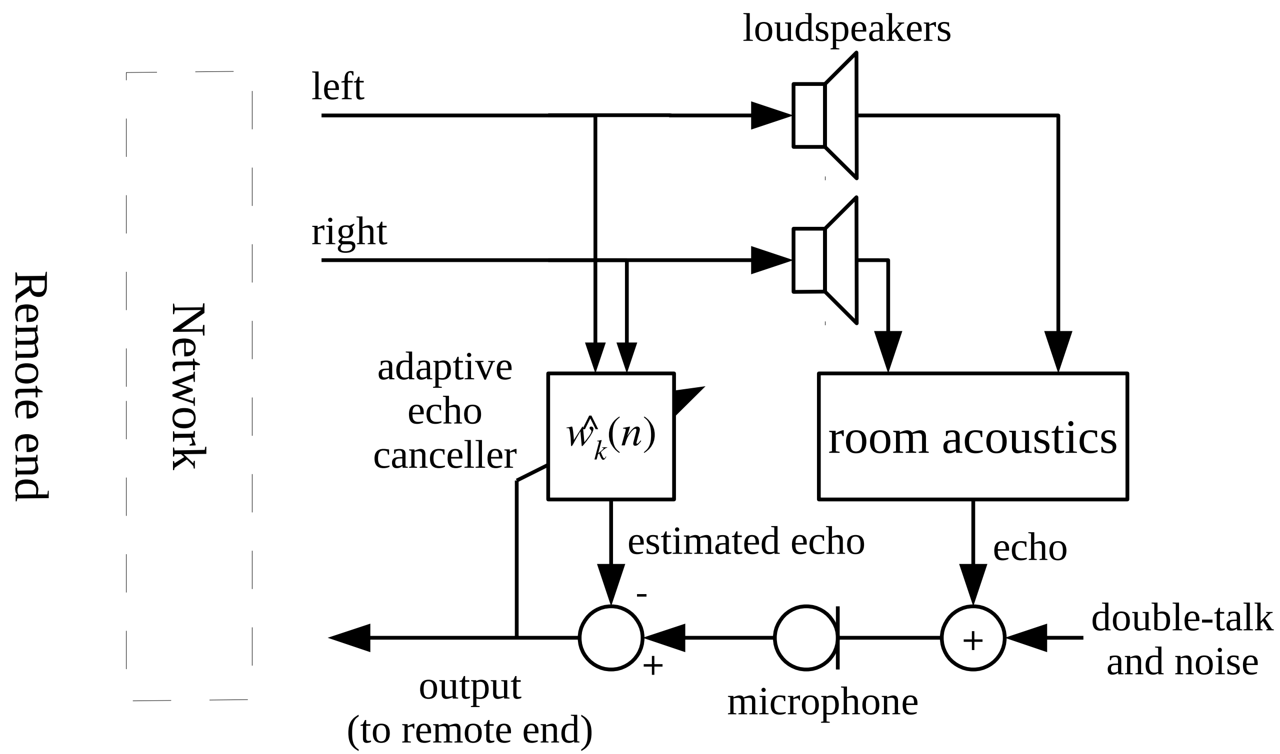
Approach: decorrelate channels by altering the received audio

Overview

www.ict.csiro.au

Stereo acoustic echo cancellation is generally ill-conditioned

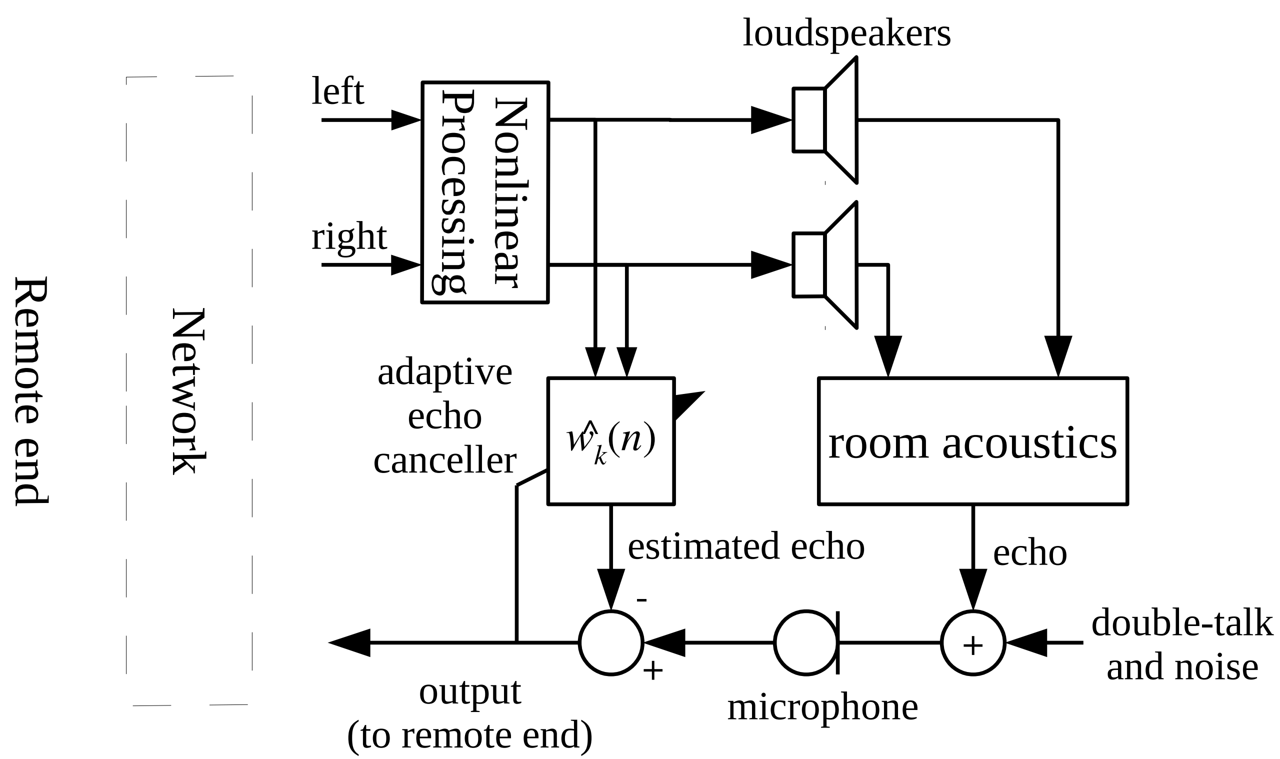
- If inter-channel coherence is high, the acoustic impulse response cannot be unambiguously identified
- Solution: reduce coherence before playback without affecting quality
- Popular implementation: memoryless non-linearity



Overview

Stereo acoustic echo cancellation is generally ill-conditioned

- If inter-channel coherence is high, the acoustic impulse response cannot be unambiguously identified
- Solution: reduce coherence before playback without affecting quality
- Popular implementation: memoryless non-linearity



Overview (cont.)

Goals

- Reduce coherence at all frequencies
- Keep noise (nearly) imperceptible
- **Preserve stereo image**
- Low delay

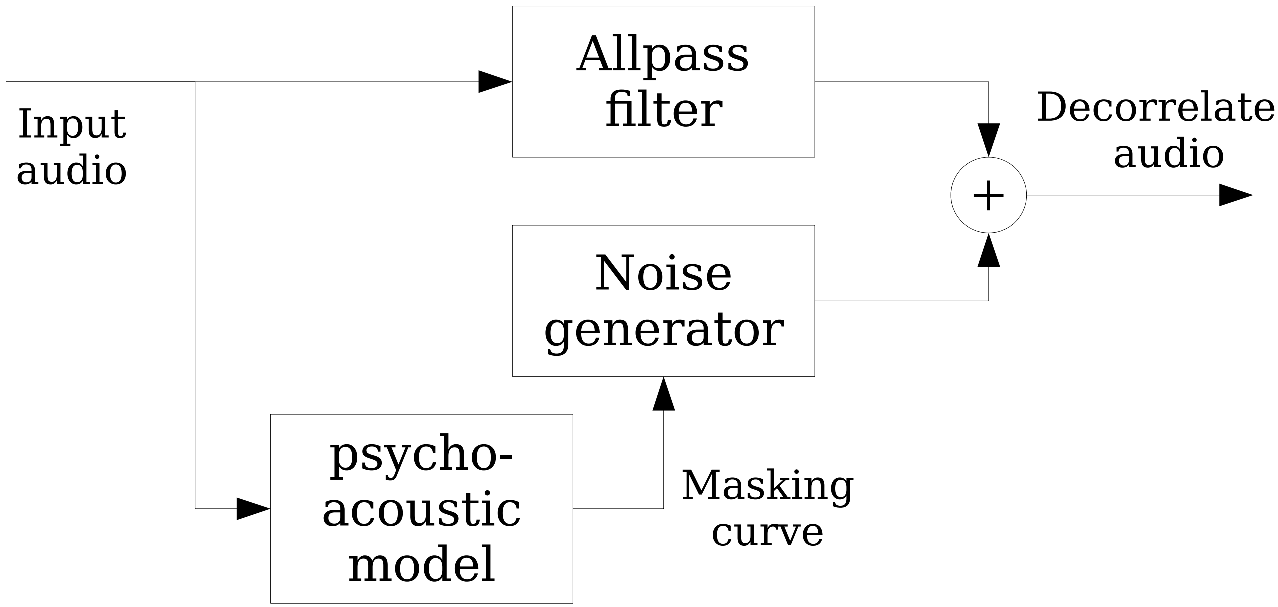
Considering

- Psychoacoustic masking
- Deafness to phase (single ear only)
- Interaural Phase Difference (IPD) as an important low-frequency localisation cue

Overview (cont.)

Proposed

- Time-varying all-pass filter with phase distortion only at higher frequencies
- Psychoacoustically-masked noise in the whole band



Allpass Filtering

Flat amplitude response, non-linear phase response

$$A(z) = \frac{\sum_{k=1}^{N} a_k z^{k-N} + z^{-N}}{1 + \sum_{k=1}^{N} a_k z^{-k}}$$

General form hard to design, instead comb-allpass filter

$$A(z) = \frac{\alpha + z^{-N}}{1 + \alpha z^{-N}}$$

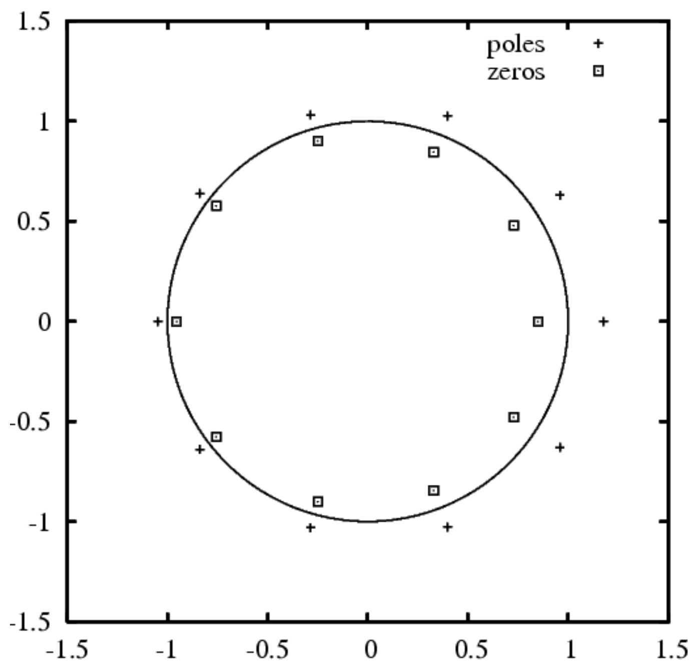
Needs to be time-varying

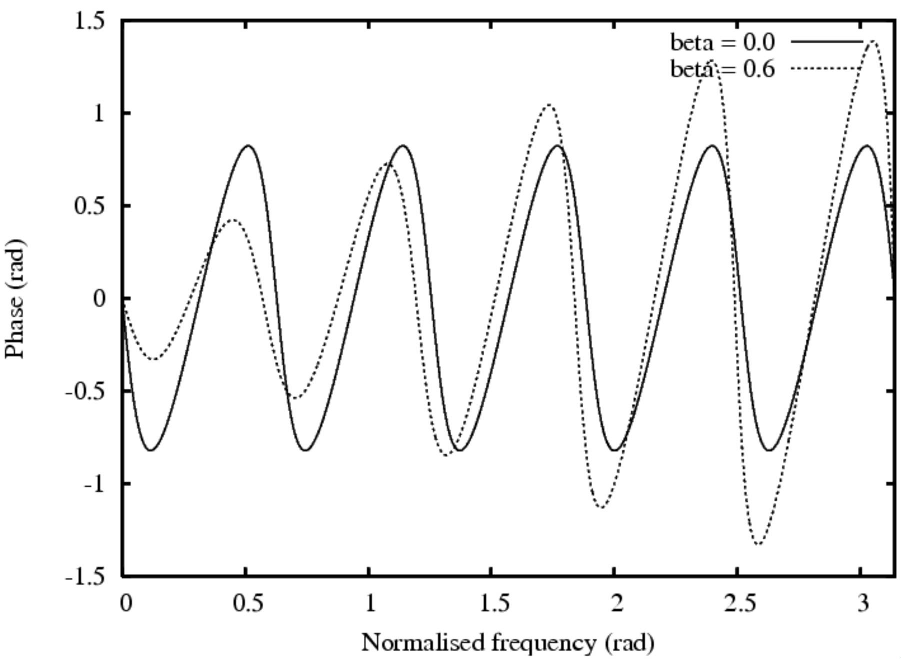
- Both *N* and α
- Held constant over a frame, apply weighted overlap-add (WOLA)

Shaped Comb-Allpass (SCAL) Filtering

Why not “shape” the phase response to minimise phase distortion at low frequency?

$$A(z) = \frac{\alpha + z^{-N}}{1 + \alpha z^{-N}} \longrightarrow A(z) = \frac{\alpha\left(1 - \beta z^{-1}\right) + z^{-N}}{1 + \alpha\left(-\beta z^{-N+1} + z^{-N}\right)}$$





Psychoacoustically-Masked Noise

- The human ear sometimes cannot perceive noise when there are other noise or tone signals (simultaneous masking)
- Can be exploited to inject imperceptible noise
- Using masking curve from the Vorbis audio codec (with minimal additional tuning)
  - Exploits simultaneous masking
  - Curve computed in the frequency domain and used to shape a white noise signal
- Weighted overlap add (WOLA)
  - Noise is delayed to eliminate delay (exploiting temporal masking)

Evaluation

Comparing with

- Smoothed absolute value (best memoryless non-linearity)
- First-order all-pass filter (filter coefficients change every sample)

Stereo processing from mono input

- 44.1 kHz
- Four music samples, four speech samples

Coherence measured as:

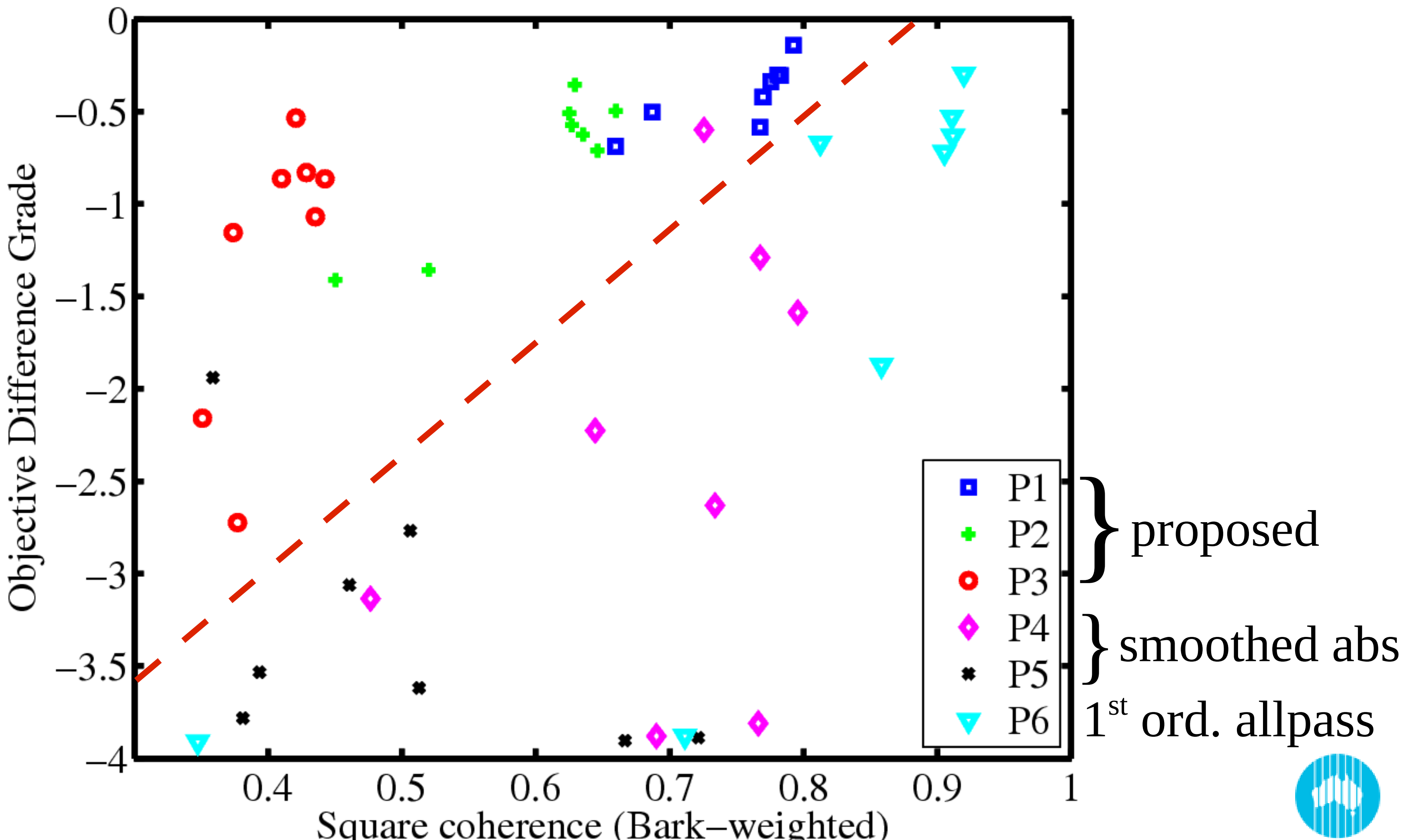
$$\gamma_{xy}^2(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)}$$

- Averaged over critical bands (Bark scale)

Quality evaluated with PEAQ (validated with informal MUSHRA-like test)

CSIRO

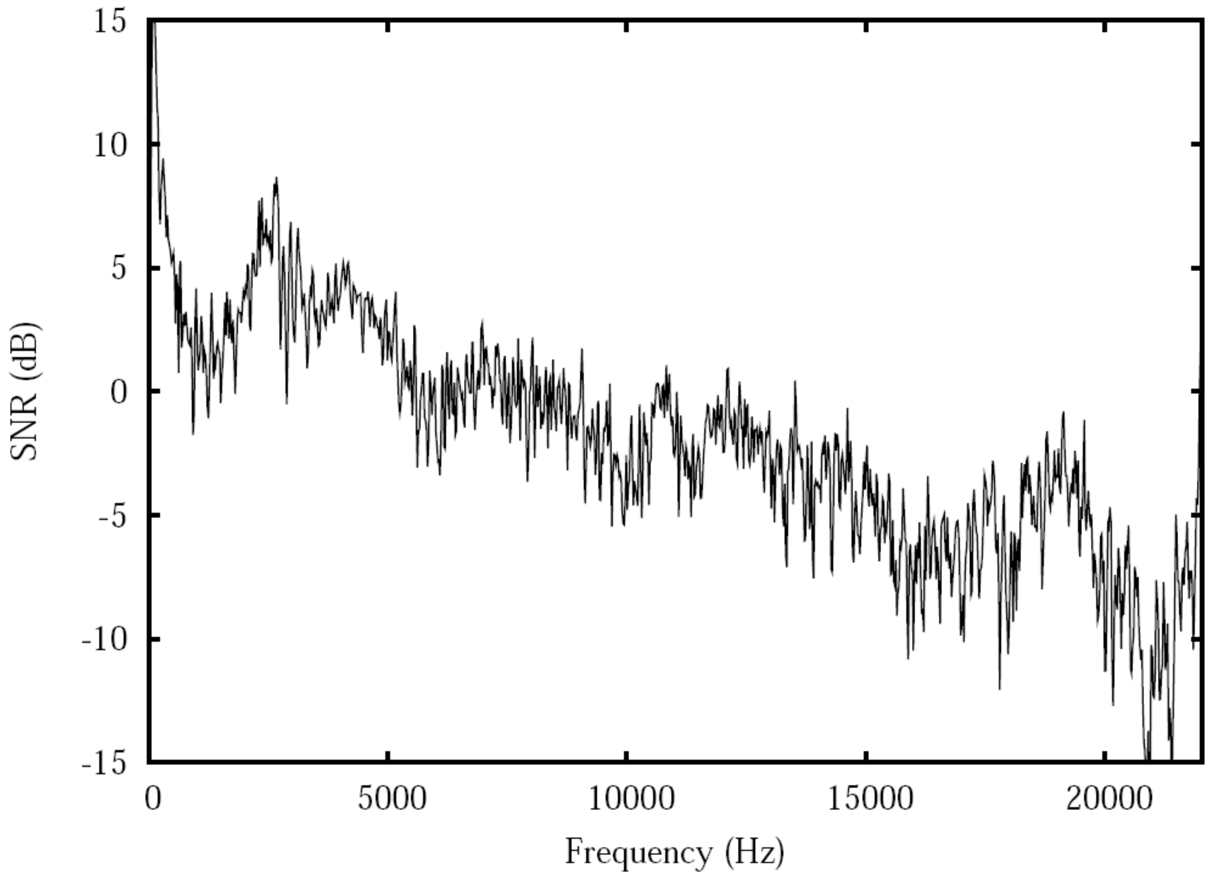
Results



Objective Difference Grade
Square coherence (Bark–weighted)
0
–0.5
–1
–1.5
–2
–2.5
–3
–3.5
–4
0.4
0.5
0.6
0.7
0.8
0.9
1
P1
P2
P3
P4
P5
P6
proposed
smoothed abs
1st ord. allpass

Results (cont.)

SNR measure on a guitar sample (nearly transparent quality)



Audio Artifacts (Worst Examples)

Smoothed absolute value

- Inter-modulation distortion on tonal audio
- “Ping-pong” stereo effect on impulsive audio

First order allpass filter

- High-frequency crackling noise (first-order allpass)

Proposed (allpass+noise)

- Mild stereo “flanging”

Conclusion

Observation: it is surprising how much abuse an audio signal can take

- No sensitivity to high frequency phase

De-correlation method:

- Shaped comb-allpass filter for high frequencies
- Wideband psychoacoustically masked noise

Both more effective and better quality than other methods

Next step: measure improvement in stereo acoustic echo cancellation context

Questions???