Watermarking for audio integrity protection

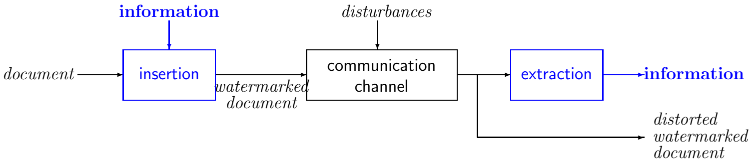
Gaël Mahé

Université Paris Descartes

March 2016

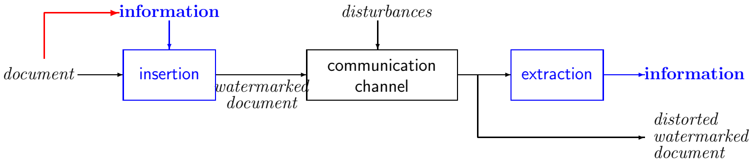
Reflexive WM: Embedding the signal in itself

From watermarking to reflexive watermarking:



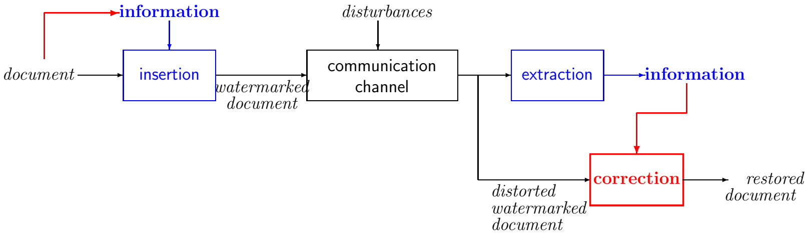
Reflexive WM: Embedding the signal in itself

From watermarking to reflexive watermarking:



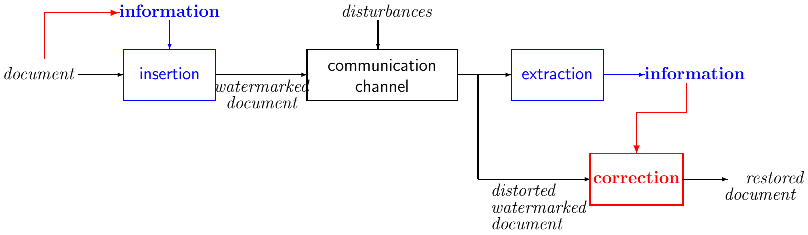
Reflexive WM: Embedding the signal in itself

From watermarking to reflexive watermarking:



Reflexive WM: Embedding the signal in itself

From watermarking to reflexive watermarking:



inaudibility

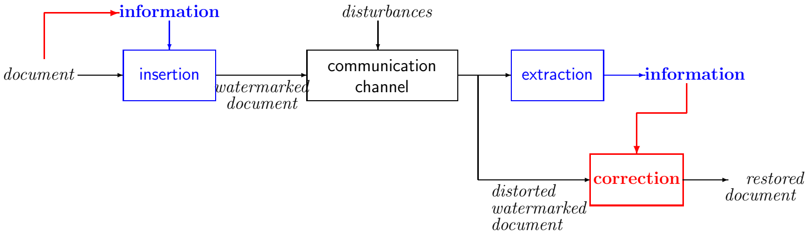
WATERMARKING

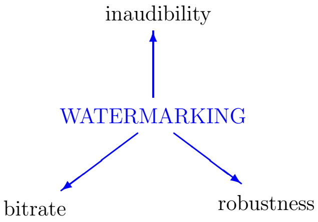
bitrate

robustness

Reflexive WM: Embedding the signal in itself

From watermarking to reflexive watermarking:





Why and how to "auto-watermark" audio signals?

Disturbances and impairments on the channel:

- Lossy compression at low bitrates → quality impairment
- Block erasure
  - ← Packet loss on IP channels (telefony or streaming)
  - ← Tampering due to malicious attacks
- Telefony: narrow-band filtering (300-3400 Hz)
- Telefony on PSTN: low-pass filtering due to analog lines
- Mobile phone: uncorrected binary errors → noises

New issues for watermarking

- High bitrate often required (>500 bit/s)
- Robustness required
  - ⊖ against adverse channel
  - ⊕ but generally not against malicious attacks
- Tradeoff on quality: impairments of the channel *vs* WM audibility + residual impairment after correction

Watermarking for audio integrity protection

1. Block erasure correction

2. Bandwidth extension

3. Correction of audio codec errors

Watermarking for audio integrity protection

1. Block erasure correction
2. Bandwidth extension
3. Correction of audio codec errors

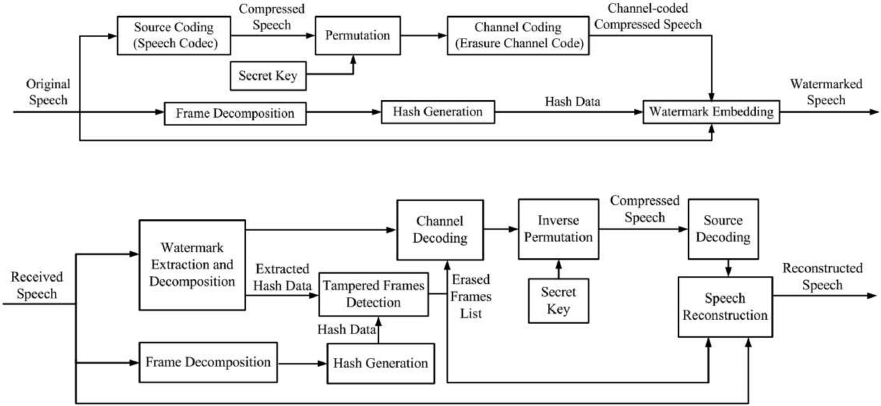
General principles

2 approaches:

- Embed a compressed version of the signal
  - → Needs high WM rate
  - → And if block A containing the compressed version of lost block B is also lost?
- Embed information to enhance interpolation from healthy blocks
  - → Lower WM rate
  - → More robust to multiple erasures

Ex: tampering correction (1)

Sarreshtedari *et al.*, “A Watermarking Method for Digital Speech Self-Recovery”, IEEE Trans. on Audio, Speech and Lang. Proc., nov. 2015:



Ex: tampering correction (2)

Implementation:

- Speech sampled at 8 kHz
- Watermark inserted in the 2 LSB of each sample
  $\rightarrow$ WM rate $=$ 16 kbit/s!
- Compression: G.723 speech codec at 6.6 kbit/s
- In each 10ms-frame:
  - 64 bits for compressed speech
  - 64 bits of redundancy for channel coding
  - 32 bits for hash code

Simulation results:

- MOS of watermarked speech > 4.2
- tampering of $1/3$ of speech $\rightarrow$ recovery $\rightarrow$ MOS around 3.6

But... robustness of WM not tested! (and surely catastrophic)

Ex: Packet loss concealment (1)

Geiser *et al.*, "Steganographic Packet Loss Concealment for Wireless VoIP", ITG-Fachtagung Sprachkommunikation, 2008.

Side information adapted to a specific speech codec (AMR wideband) and only **complements** classical blind concealment methods:

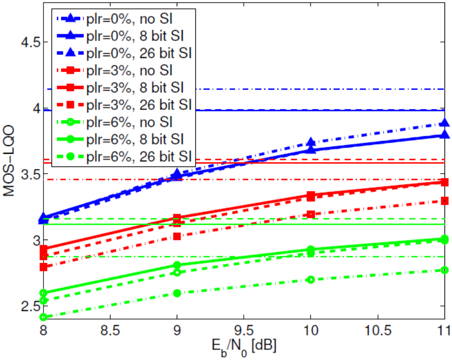
- **Spectral enveloppe (LSFs)** interpolated from previous and next frames
  → information = interpolation factor, 2 bit/frame, *i.e;* 100 bit/s
- **pitch** : information = method of estimation + correction of the estimation
  → 15 bit/frame, *i.e.* 750 bit/s
- **adaptive codebook gain**: information = interpolation method
  → 3 to 9 bit/frame, *i.e.* 150 to 450 bit/s

Finally, WM rate of 400 to 1300 bit/s + channel coding
→ WM at 2 kbit/s, embedded through joint speech coding / data hiding

Ex: Packet loss concealment (2)

Simulations:

- Channel = packet network + GSM network (circuit switch)
- Various packet loss rates: 0, 3 and 6%
- Noisy GSM channel ($E_b/N_0 = 8$ to 11dB) $\rightarrow$ residual bit errors
- side-information used only if not detected as corrupted



Watermarking for audio integrity protection

1. Block erasure correction
2. Bandwidth extension
3. Correction of audio codec errors

Principles

Telephony narrow-band (NB): 300-3400 Hz

High-frequency band (3-8 kHz) re-synthetized at receiver part from:

- wide-band (WB) excitation
- wide-band spectral envelope

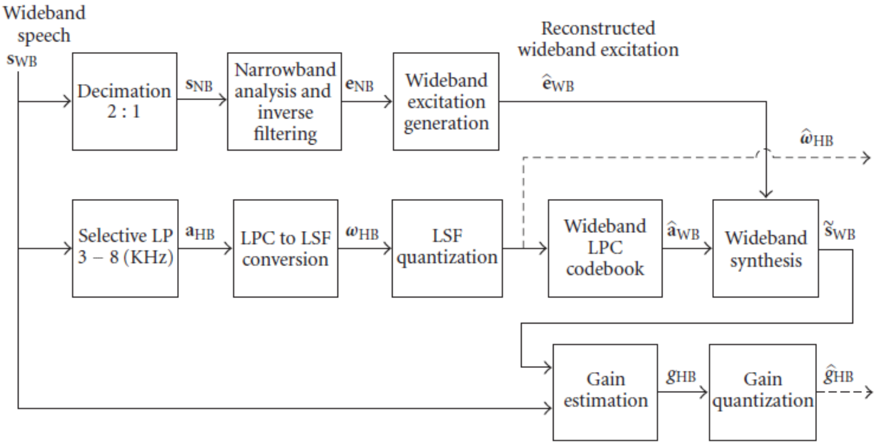
2 approaches:

- blind scheme: use correlation between low and high frequencies
- hybrid scheme : reconstruction of HF both from BF and side information

Bandwidth extension using side information (1)

A. Sagi and D. Malah, "Bandwidth Extension of Telephone Speech Aided by Data Embedding", EURASIP J. on Advances in Signal Processing, 2007.

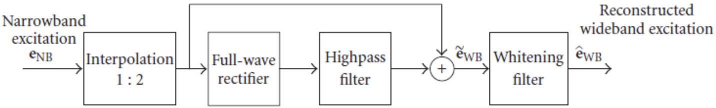
Transmitting part:



Bandwidth extension using side information (2)

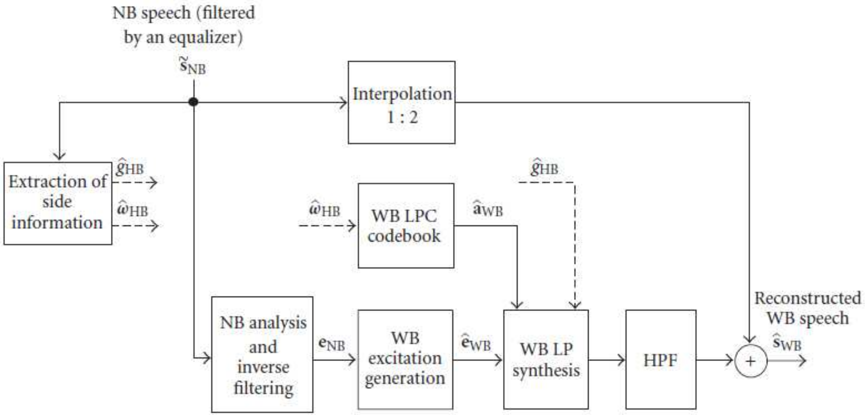
Transmitting part:

Artificial WB excitation generation:



Bandwidth extension using side information (3)

Receiving part



Bandwidth extension using side information (4)

Simulation:

- WM based on scalar Costa scheme ($\simeq$QIM) applied to Discrete Hartley Transform (DHT)
- In each 32ms frame with 50% overlap, insert: 16 bits for LSF, 8 bits for gain and 40 bits for error correction $\rightarrow$ WM rate = 4 kbit/s
- Psycho-acoustical model: MPEG-1
- Channel models:
  1. telephone channel model ITU-T V.56bis (amplitude and phase distortions) + PCM quantization + white Gaussian noise
  2. μ-law 8 bit quantization only
  3. white Gaussian noise with 35dB SNR

Results:

- MOS of watermarked NB speech = 3.625 *vs* 3.7 without WM
- BER in WM detection: 3.10^{-4}
- Reconstructed WB speech preferred to NB speech in 92.5% of test utterances

Watermarking for audio integrity protection

Pre-echo in MP3 and AAC codecs (1)

Quantization in the transform domain

- q. noise: frequency-shaped, uniform in time-domain
- ▶▶ pre-echo in attacks

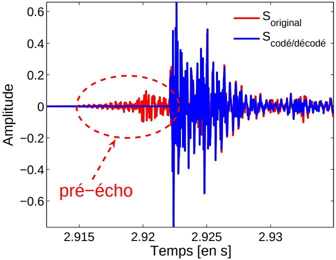


Figure: Castagnet signal, coded by MP3 at 48 kbit/s

Pre-echo in MP3 and AAC codecs (2)

How to avoid pre-echo? Options implemented in the standards:

- Unaudible if duration $<$ 5ms and level $<$ pre-masking threshold

$\rightarrow$ MP3 and AAC use variable window lengths.

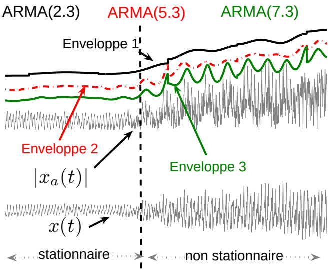
$\rightarrow$ Option Temporal Noise Shaping (TNS) in AAC

- But do not cancel all pre-echoes and increase bitrate

Pre-echo in MP3 and AAC codecs (3)

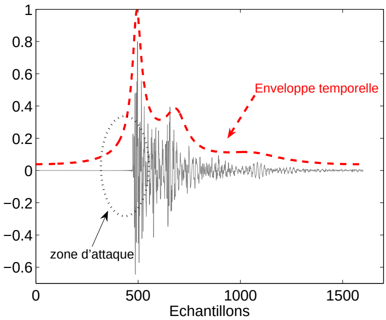
- Solution proposed in: I. Samaali *et al.*, "Watermark-aided pre-echo reduction in low bit-rate audio coding", J. of the Audio Engineering Society, 2012
- **Principle:** transmit the temporal envelop by watermarking and correct after decoding

→ How to model the envelop with few parameters?



Pre-echo in MP3 and AAC codecs (4)

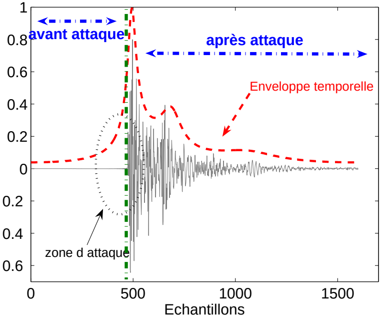
Example: attack in a castanet signal, envelop ARMA(7,3)-modeled



- ⊖ Under-modelling in case of strong energy variation

Pre-echo in MP3 and AAC codecs (4)

Example: attack in a castanet signal, envelop ARMA(7,3)-modeled



- ⊖ Under-modelling in case of strong energy variation
- → Transmit 2 successive models: before and after attack
- attack position does not need to be transmitted (robust detection though pre-echo)

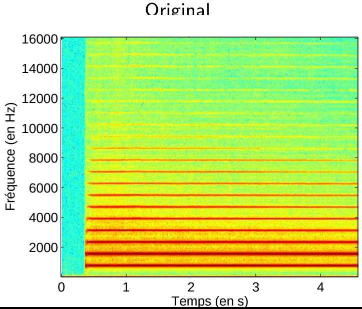
Harmonicity disruption in SBR codecs (1)

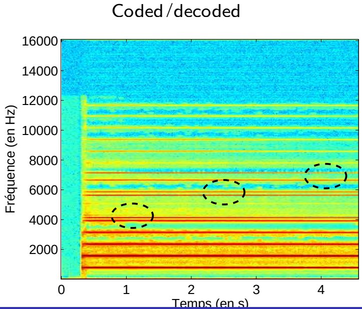
Principles of **Spectral Band Replication** (SBR: AAC+, MP3Pro...):

- transform coding of low-frequency band (AAC, MP3...)
- side info for HF synthesis: spectral envelop + tone-to-noise ratio
- HF reconstruction in decoder =
  - copy low-frequency bands to HF
  - correct spectrum using side information

▶ **Harmonicity disruption / tones alteration** (unharmonic tonals)

Trumpet, AAC+ coded at 16 kbits/s:

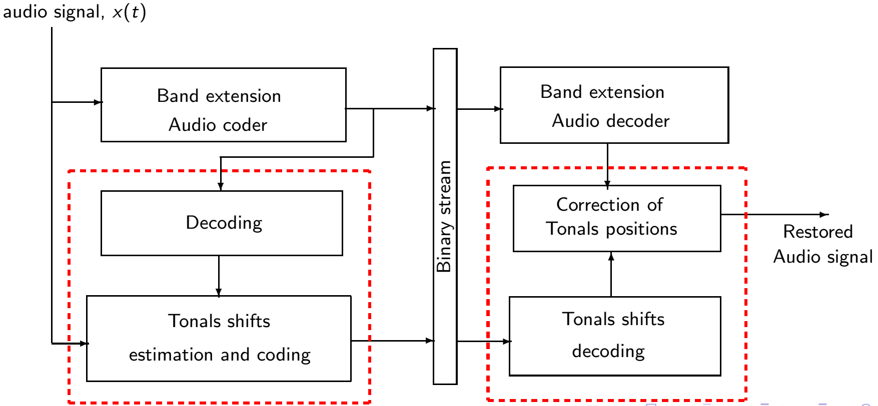




Harmonicity disruption in SBR codecs (2)

Proposition in: I. Samaali *et al.*, "High-Frequency Tonal Components Restoration In Low-Bitrate Audio Coding Using Multiple Spectral Translations", Eusipco 2015:

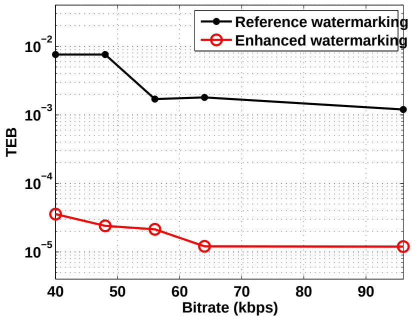
1. Transmit by WM the offsets between original and synthetized HF tonals
2. In receiving part, translate HF tonals



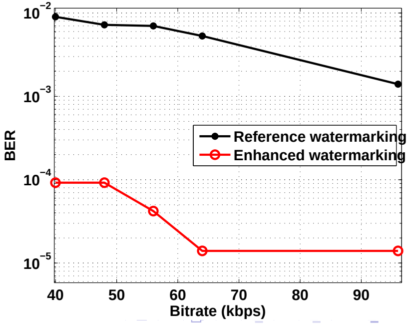
How to embed side information? (1)

- Spread spectrum watermarking system from [Larbi2005]
- MPEG-1 psycho-acoustical model (1992)
- WM bandwidth adapted to that of core-codec at low bitrates: 5 kHz for MP3 and AAC, 3.5 kHz for AAC+
- insertion only in frames without attacks

MP3 codec, WM rate = 78bit/s

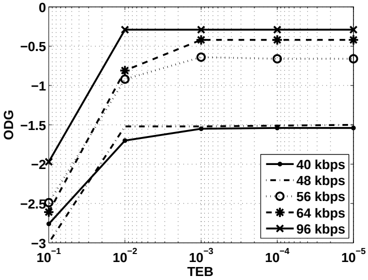


AAC codec, WM rate = 78bit/s



How to embed side information? (2)

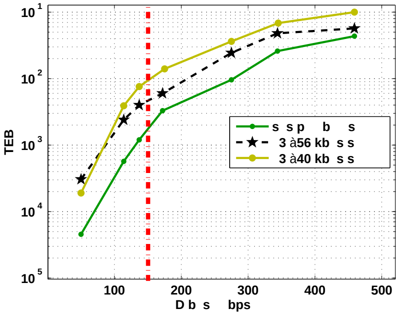
Quality of corrected audio *vs* BER on side information (MP3):



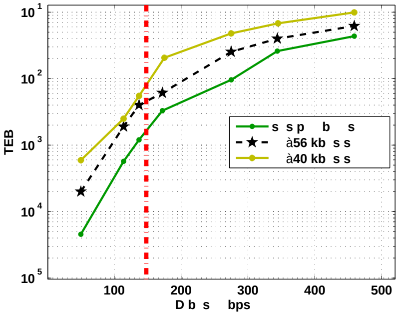
How to embed side information? (3)

BER on side information *vs* WM bitrate

MP3 codec



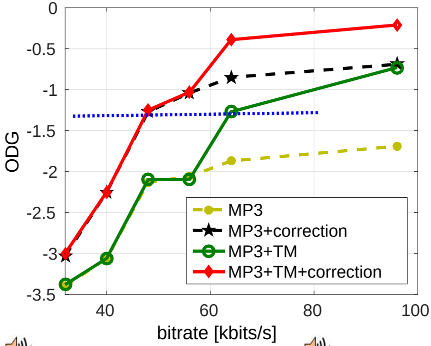
AAC codec



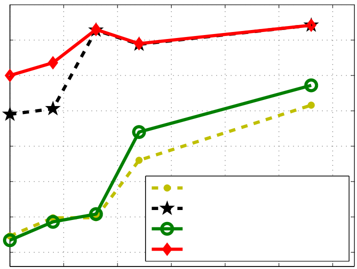
Pre-echo reduction: results

Side information at 50 bit/s, MP3 coding

Castanets



Darbouka



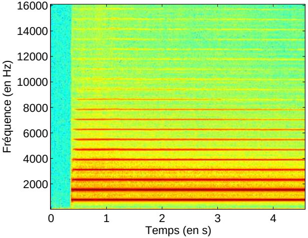
Original

Coded/decoded

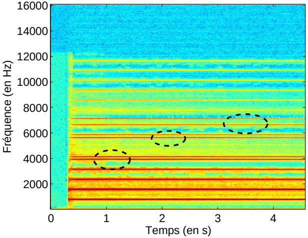
Restored

Harmonicity correction: results

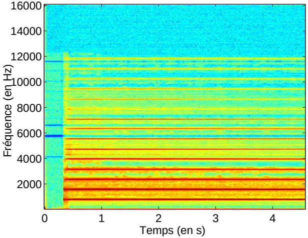
Trumpet at 32 kHz



Original



Coded/decoded



Restored

Conclusion

How to build a WM system for audio integrity protection?

- Contradiction high WM rate bitrate / high robustness
  - To **reduce the amount of data to insert**, hybrid approach = classical blind estimation complented by side information
  - Known channel "attacks" → **insert WM in the less sensitive part**
- **Inaudibility constraint can be relaxed**
  if WM + correction less annoying than channel impairment