Maltraitement du signal
Partie 3: tatouage

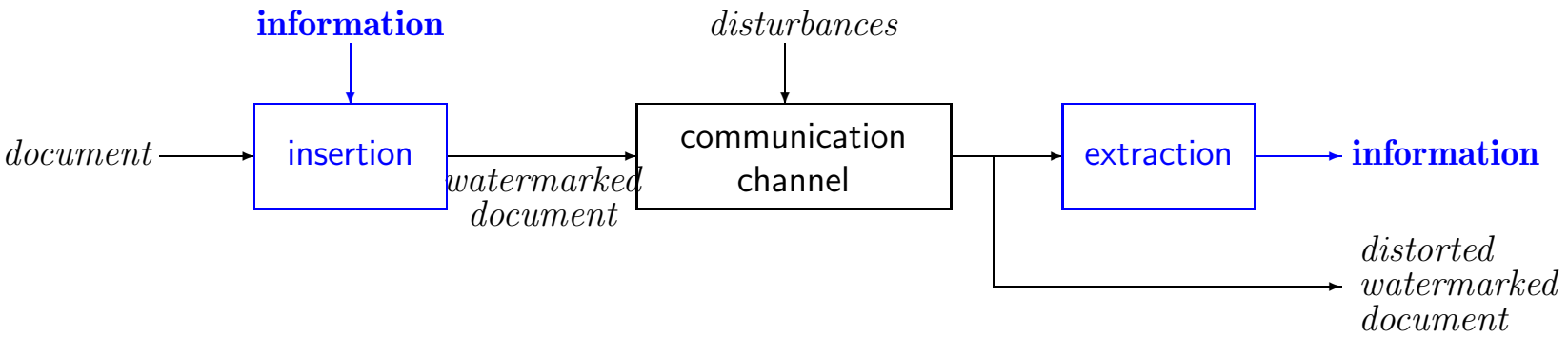
Gaël Mahé

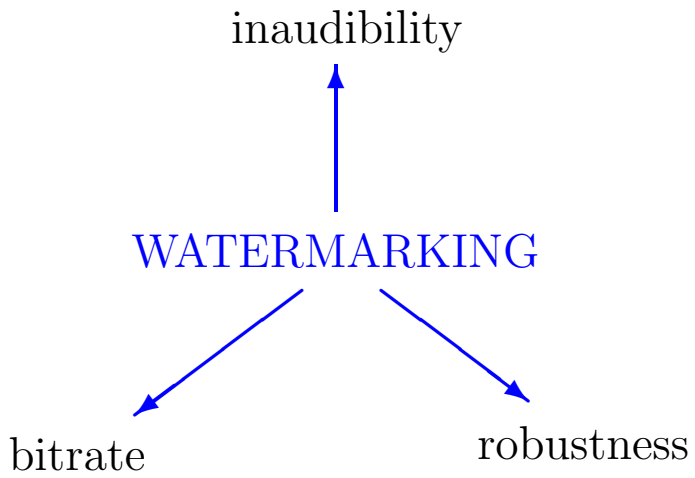
Université Paris Cité

octobre 2022

Maltraitement du signal
Partie 3: tatouage

Principes du tatouage audio





Techniques de tatouage

À suivre : 7 diapositives de Sonia Larbi et Cléo Baras

Classes applicatives (1/2)

Tatouage sécuritaire

- **Protection** :

  droits d'utilisation, droits d'auteur, preuve de propriété

- **Authenticité** du contenu :

  intégrité, falsification de documents

- **Identification** d'un document

  traçabilité dans un réseau de diffusion, estampillage

Craver *et al.* : What can we reasonably expect from watermarks ?, *WASPAA*, 2001

Kirovski et Malvar : Spread-spectrum audio watermarking : Requirements, applications, limitations, *IWMSP*, 2001

Classes applicatives (2/2)

Canaux cachés

- **Ajout d'informations**

  Annotation, auto-indexation

- **Amélioration** des systèmes de transmission existants :

  Manque de bande passante

- **Contrôle** d'applications cibles

  Projet RNRT[a] Artus

[a] Réseau National de Recherche en Télécommunications

Craver *et al.* : What can we reasonably expect from watermarks ?, *WASPAA*, 2001

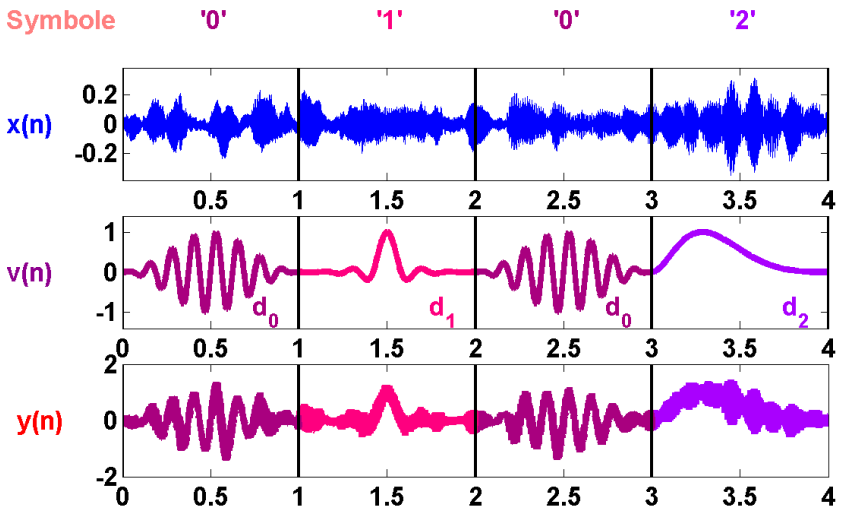
Kirovski et Malvar : Spread-spectrum audio watermarking : Requirements, applications, limitations, *IWMSP*, 2001

Une chaîne de communication BBAG

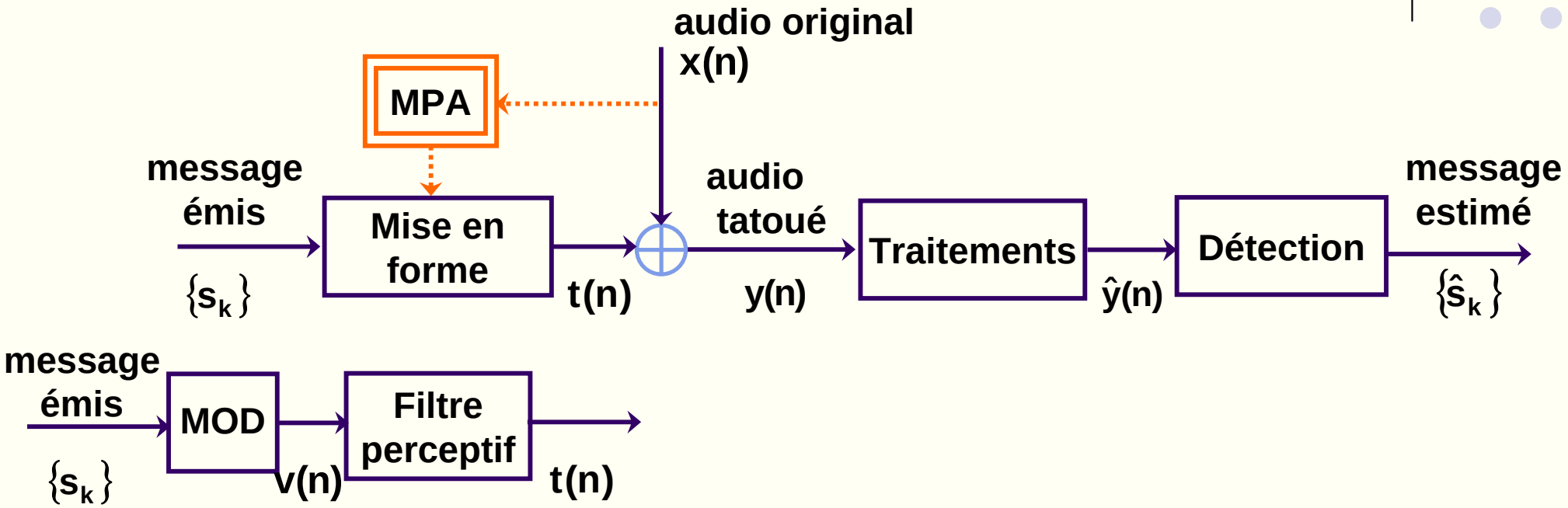
- **Source** : $s_l \in \{0, M-1\}$
- **Modulation** :

  M vecteurs/ 1 id de durée N_s
- **Psychoacoustique** :

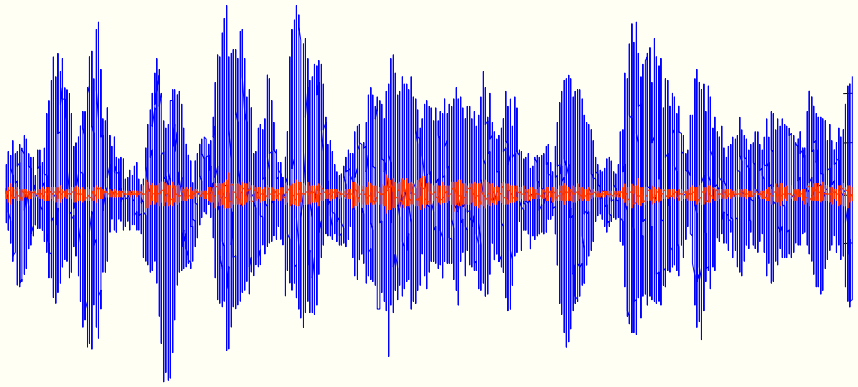
  $t(n) = \alpha v(n)$
- **Récepteur** : corrélateur



Tatouage audio temporel et perceptif



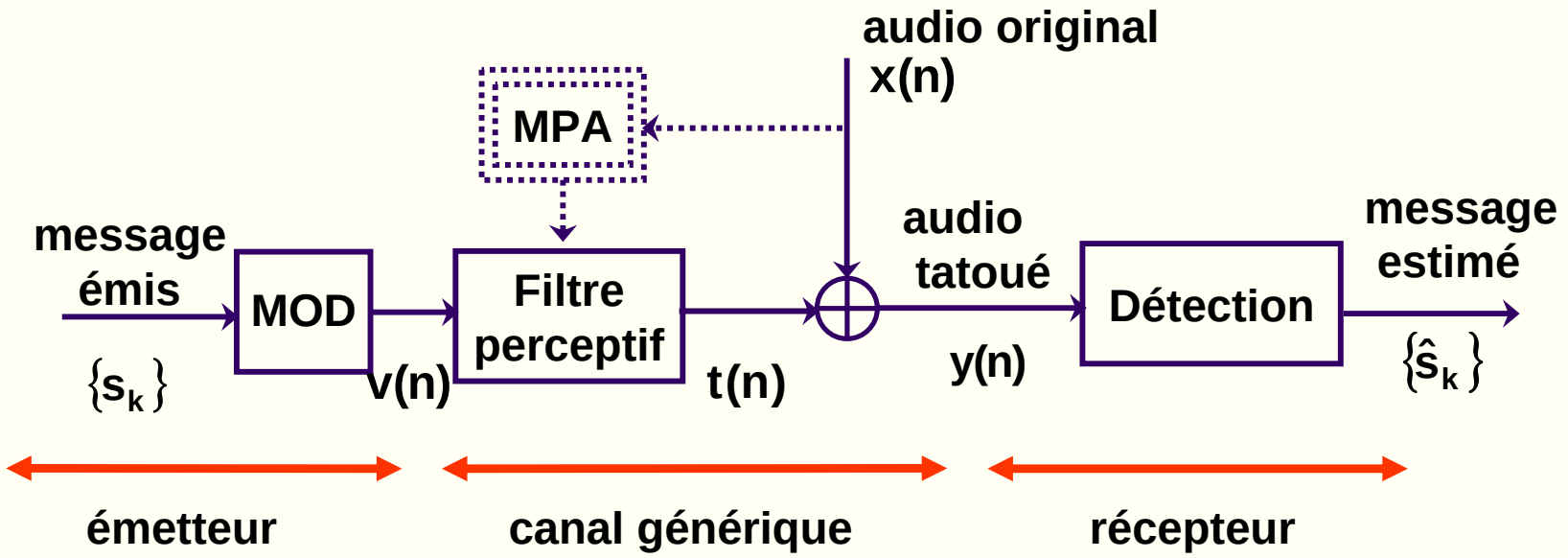
- Insertion temporelle



- Détection aveugle
  - Les traitements subis par le signal tatoué sont inconnus
  - Le message émis est inconnu
- Tatouage aveugle
  - L'audio original est inconnu

Tatouage = chaîne de Com. Num.

- **Equivalence** du système de tatouage avec une chaîne de **communications numériques**

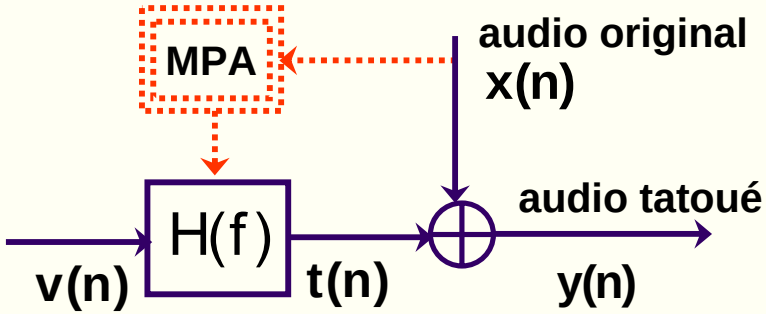


- **Contexte** : **Transmission de données cachées**
- **Objectif** : améliorer les performances du **récepteur** en termes de **débit** et de **TEB**

Transmission du tatouage : le canal

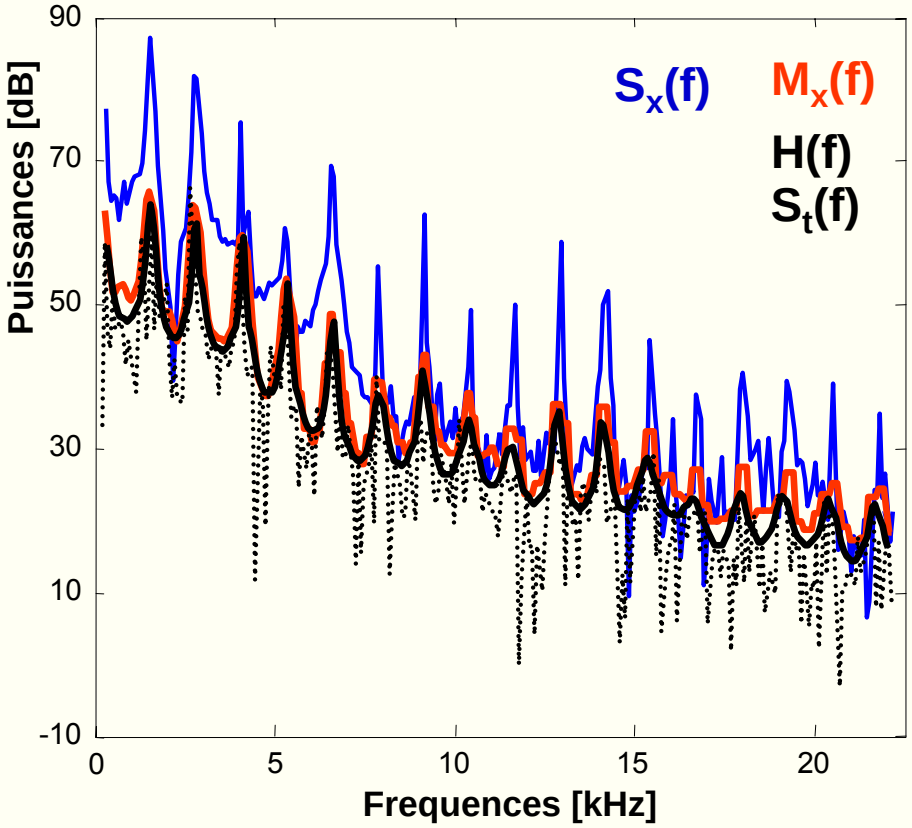
Mise en forme spectrale et amplification de v(n) :

- par filtrage linéaire : $S_t(f) = \sigma_v^2 \left|H(f)\right|^2 = M_x(f)$
- Avec contrainte perceptive :



- Synthèse d'un filtre récursif :

$$H(z) = \frac{\beta_0}{1 + \sum_{i=1}^{P} \alpha_i z^{-i}}$$

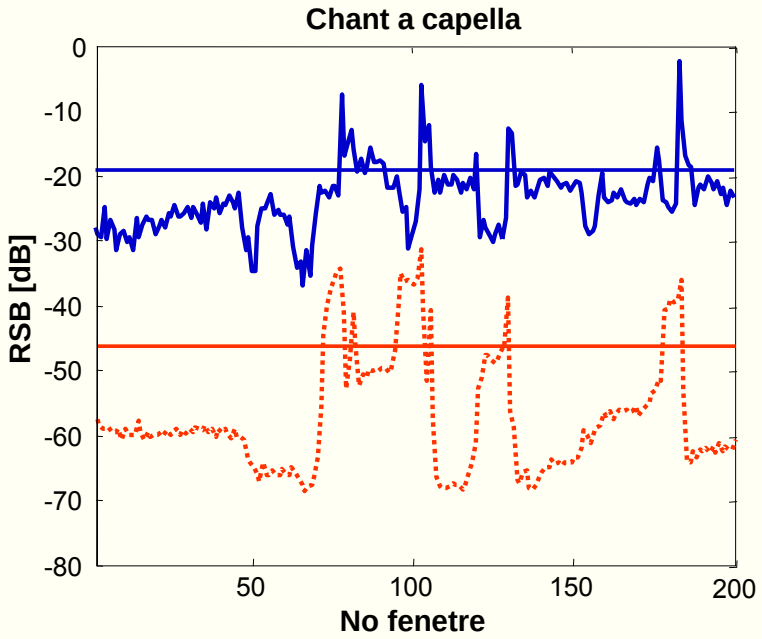


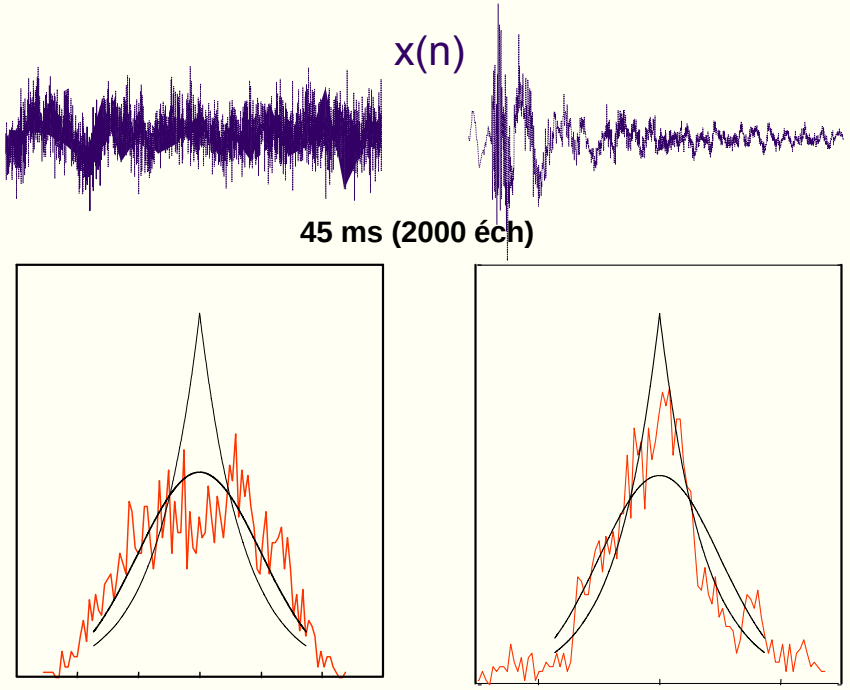
L'audio original comme bruit de canal

- Le signal audio représente un bruit additif :
  - non stationnaire
  - très puissant
  - très corrélé
  - non gaussien

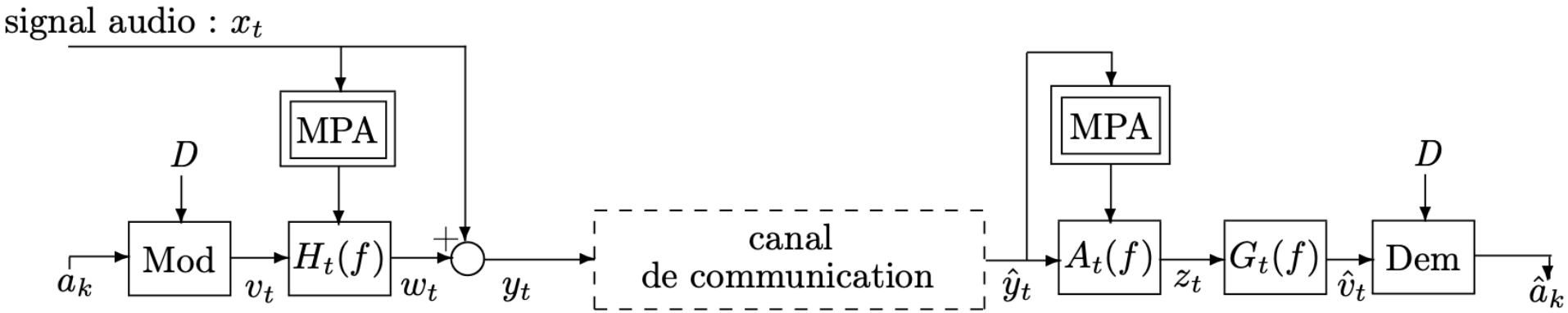
RSB moyen à l'entrée du canal : $\sigma_v^2 / \sigma_x^2 = -45$ dB

RSB moyen à la sortie du canal : $\sigma_t^2 / \sigma_x^2 = -20$ dB





Chaîne de tatouage par étalement de spectre



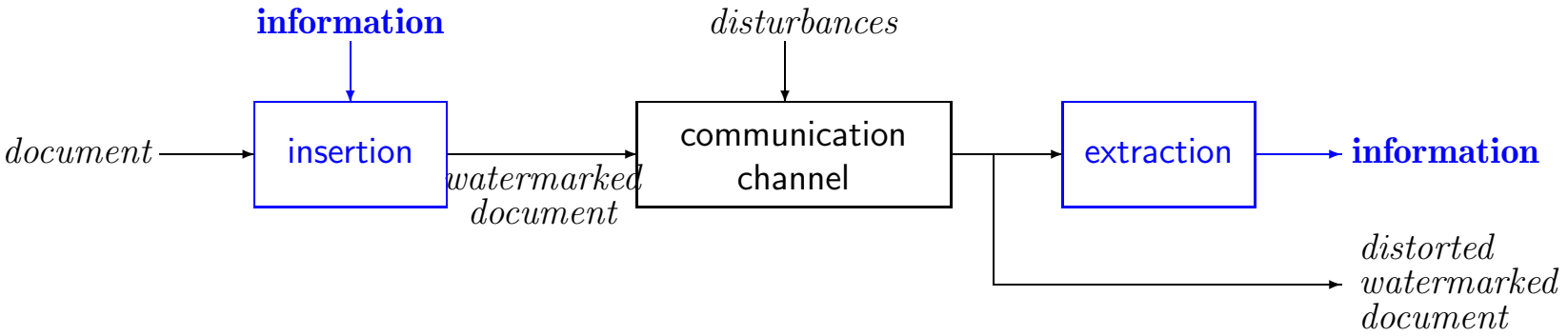
En réception :

- égaliseur par zero-forcing : A inverse filtre de mise en forme H,
  - H inconnu en réception
  - $\rightarrow$ hypothèse : seuil de masquage de $\hat{y}(t)$ très proche de celui de $x(t)$
- filtre de Wiener G
  - Principe : minimisation de $\mathrm{E}[\hat{v}(n) - v(n)^2]$
  - $\rightarrow$ expression à partir des fonctions d'autocorrélation de z et de v,
  - autocorrélation de v obtenue par connaissance du dictionnaire D

Maltraitement du signal
Partie 3: tatouage

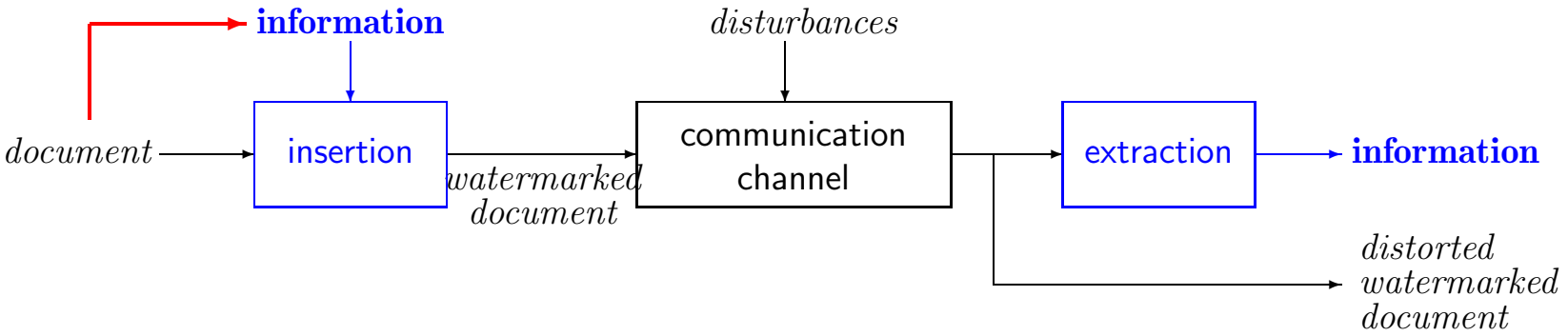
Reflexive WM: Embedding the signal in itself

From watermarking to reflexive watermarking:



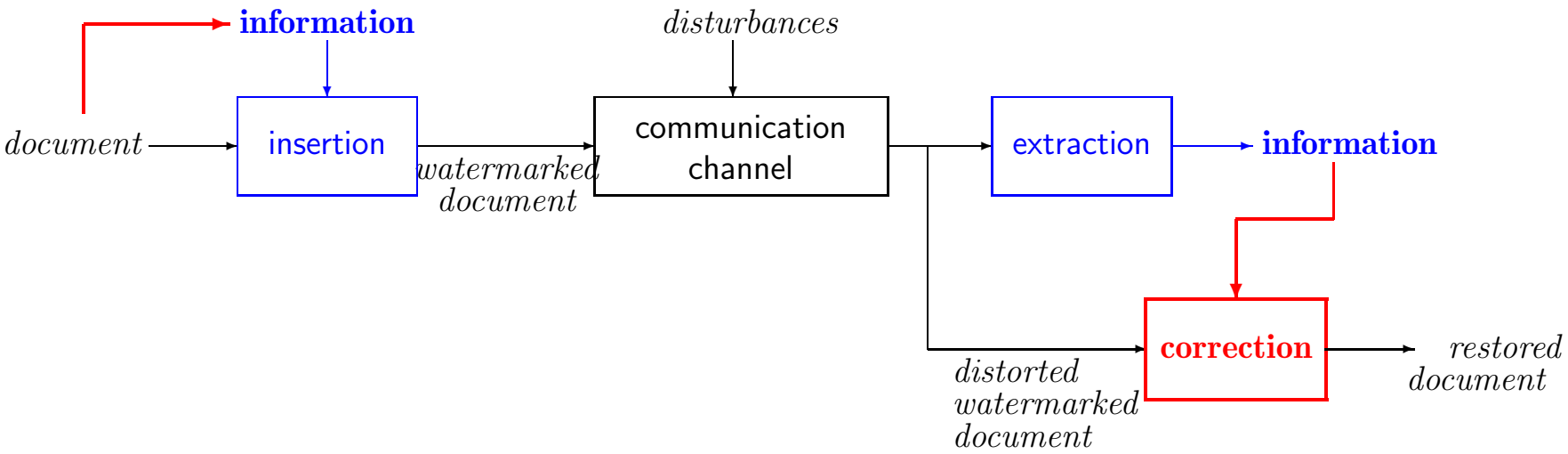
Reflexive WM: Embedding the signal in itself

From watermarking to reflexive watermarking:



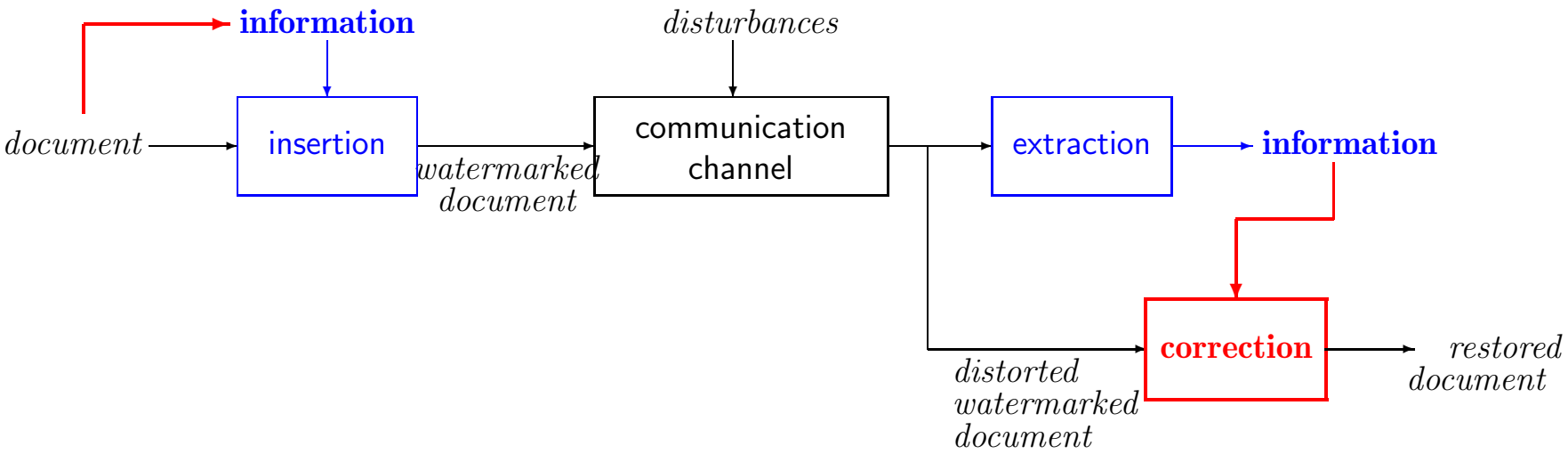
Reflexive WM: Embedding the signal in itself

From watermarking to reflexive watermarking:



Reflexive WM: Embedding the signal in itself

From watermarking to reflexive watermarking:



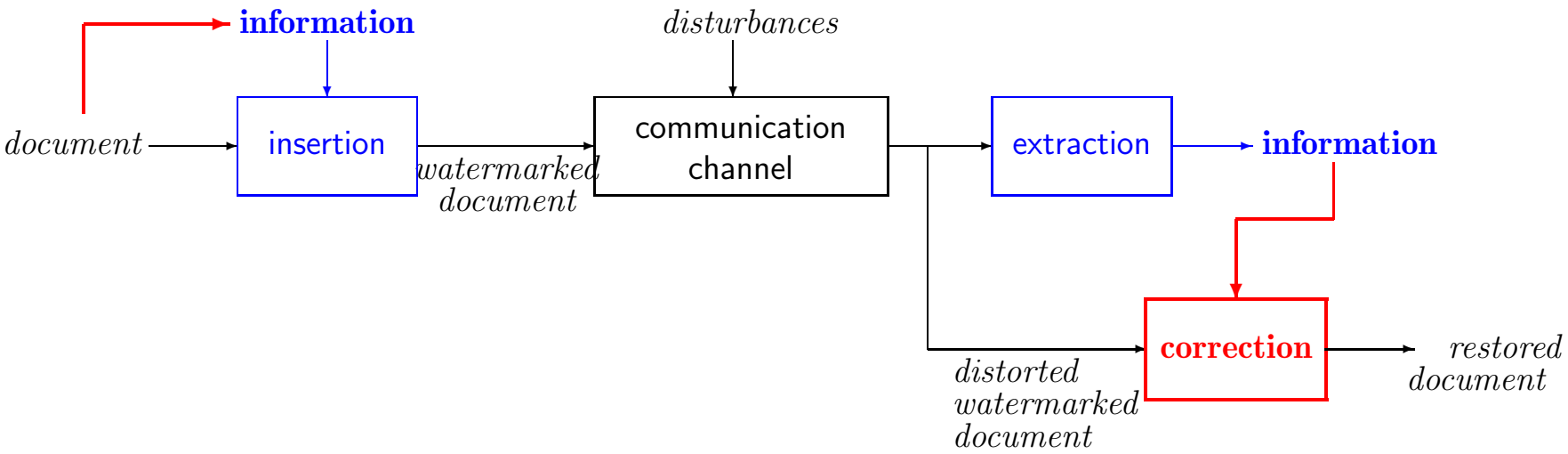
inaudibility

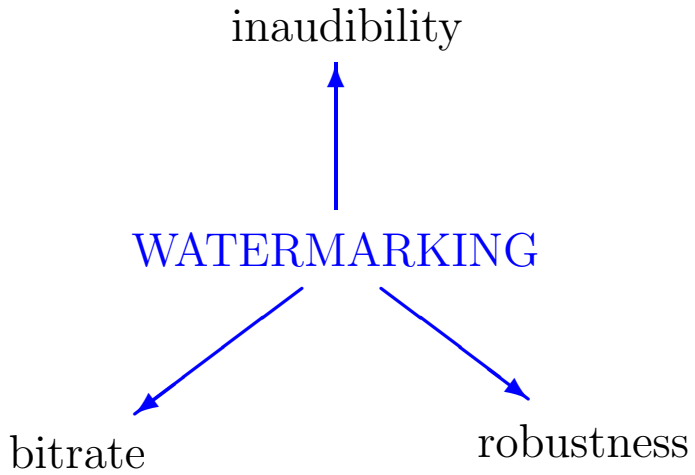
WATERMARKING

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

Block erasure correction

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: 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:

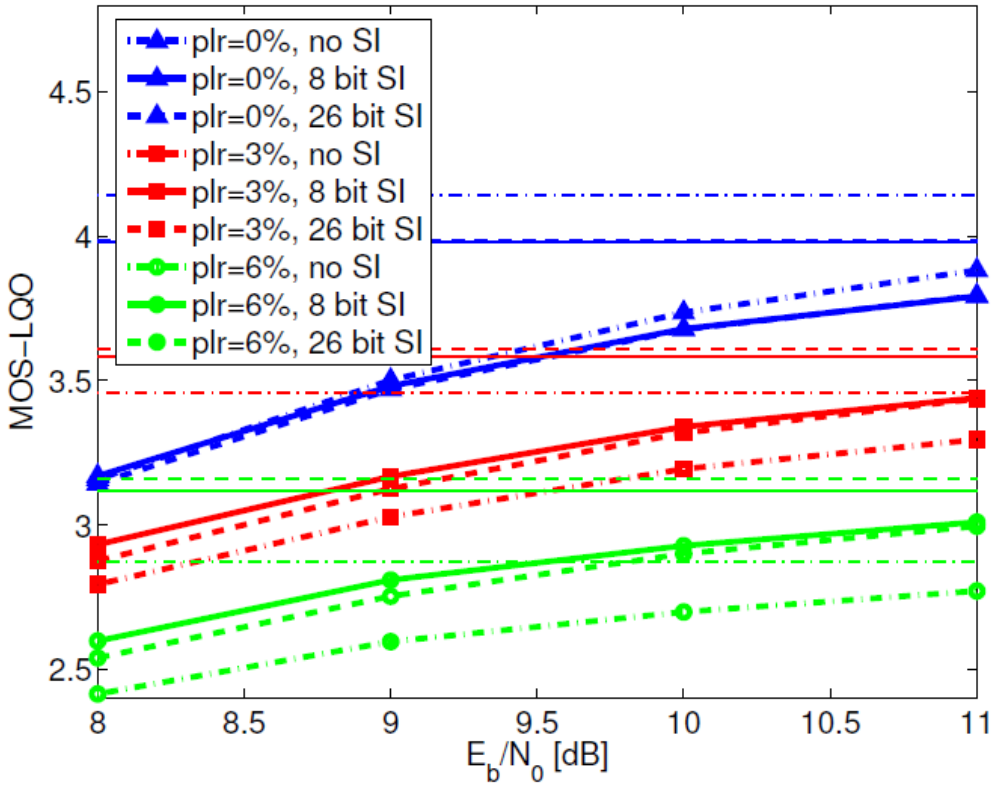
- **Spectral enveloppe (LSFs)** interpolated from previous and next frames
  $\rightarrow$ information = interpolation factor, 2 bit/frame, *i.e;* 100 bit/s
- **pitch** : information = method of estimation + correction of the estimation
  $\rightarrow$ 15 bit/frame, *i.e.* 750 bit/s
- **adaptive codebook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

Finally, WM rate of 400 to 1300 bit/s + channel coding
$\rightarrow$ 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



Bandwidth extension

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

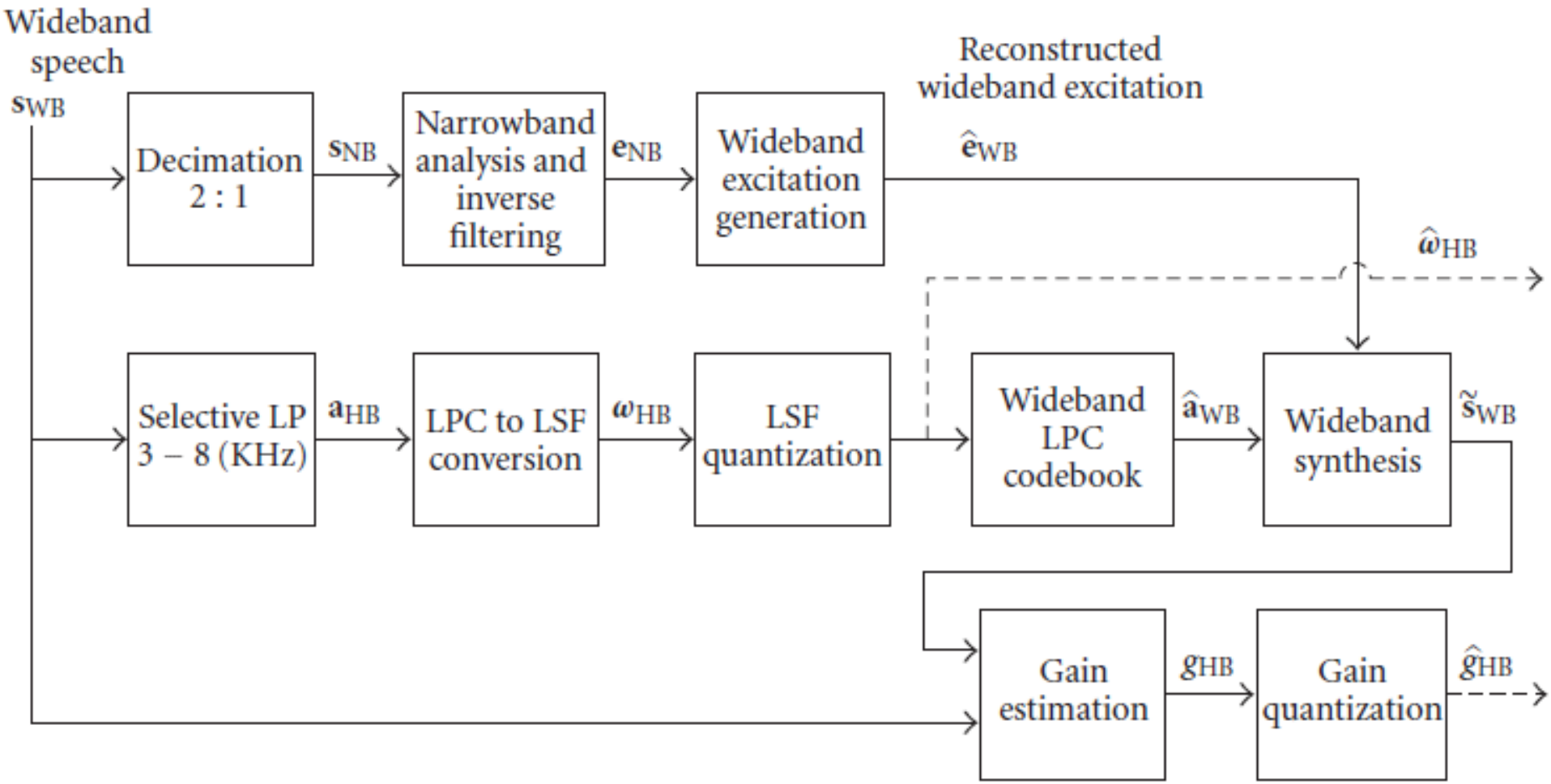
2 approaches:

- 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.

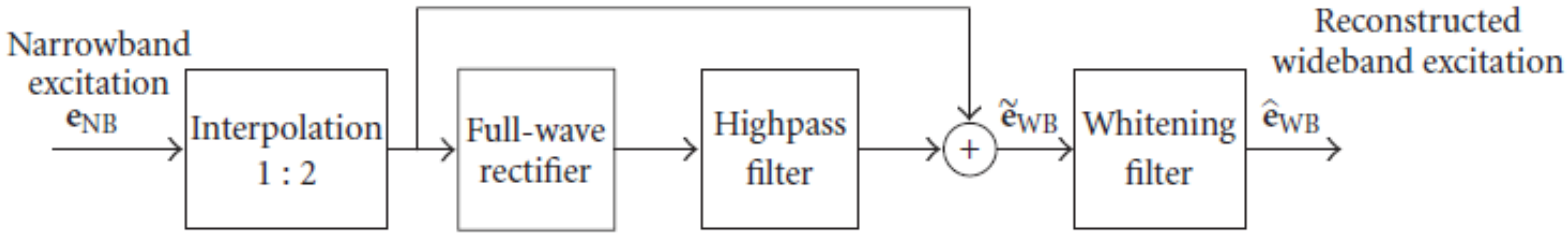
Transmitting part:



Bandwidth extension using side information (2)

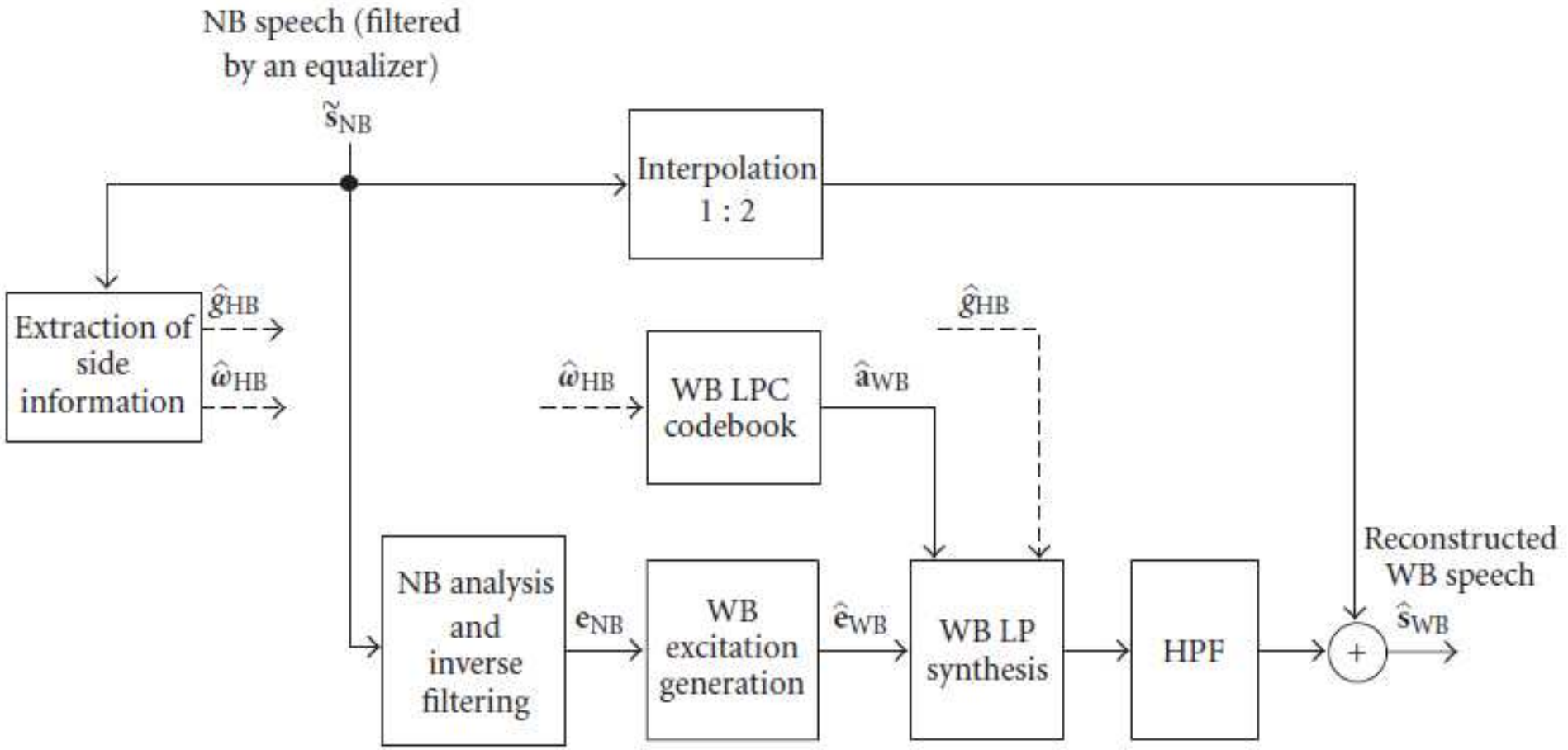
Transmitting part:

Artificial WB excitation generation:



Bandwidth extension using side information (3)

Receiving part



Bandwidth extension using side information (4)

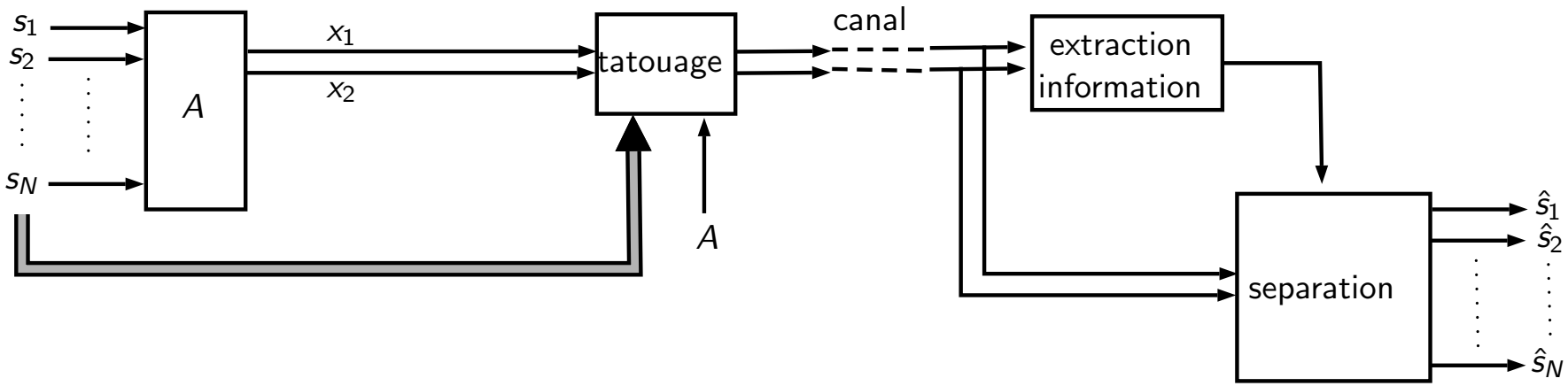
Simulation:

- 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:

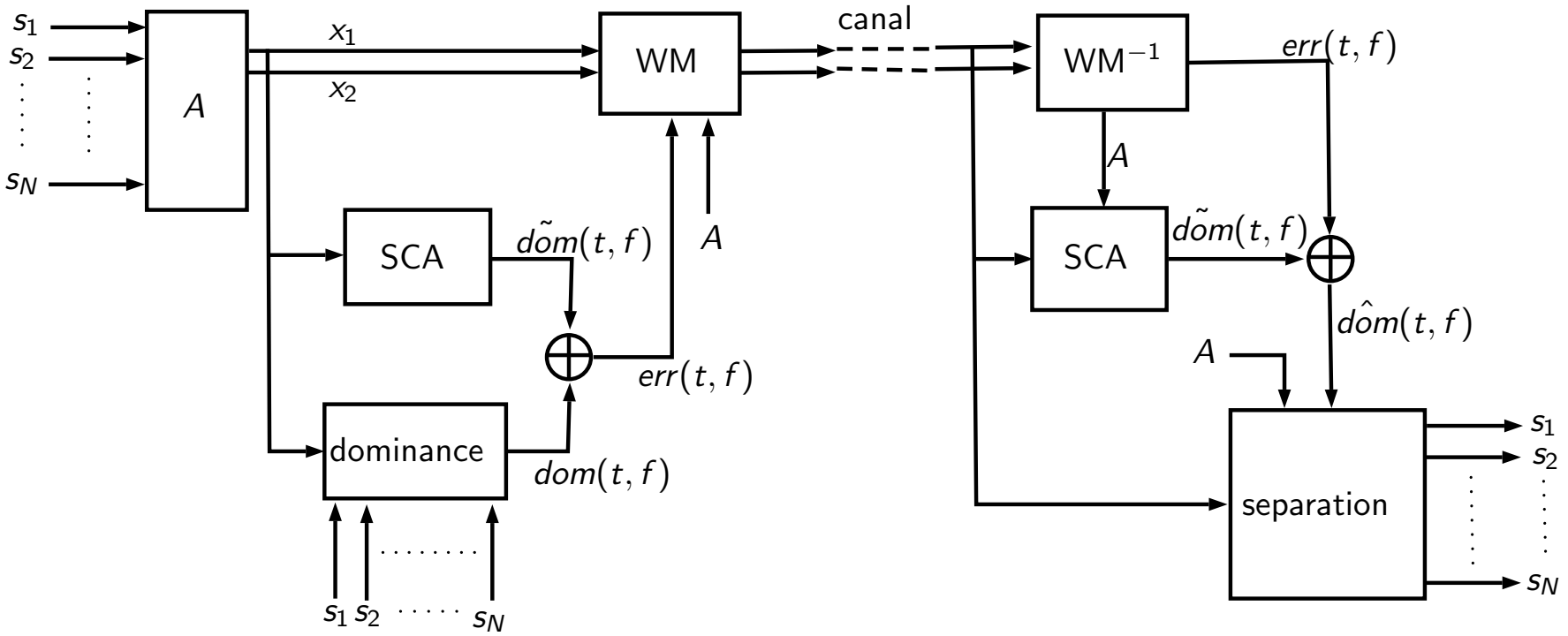
- MOS of watermarked NB speech $=$ 3.625 *vs* 3.7 without WM
- BER in WM detection: 3.10^{-4}
- Reconstructed WB speech preferred to NB speech in 92.5% of test utterances

Séparation de sources assistée par tatouage (1)



- Parcimonisation des sources :
  suppression des composantes du spectre masquées
- $\rightarrow$ Pour chaque (f, t), peu probable que plus de 2 sources actives
- Idée : transmettre pour chaque source une matrice de dominance dans le plan temps-fréquence
  Pour chaque (f, t), elle dit si la source est une des 2 sources dominantes
- $\rightarrow$ Séparation de sources déterminée en chaque (f, t) : facile
- Problème : débit de tatouage $=$ 44,1 kbit/s par source

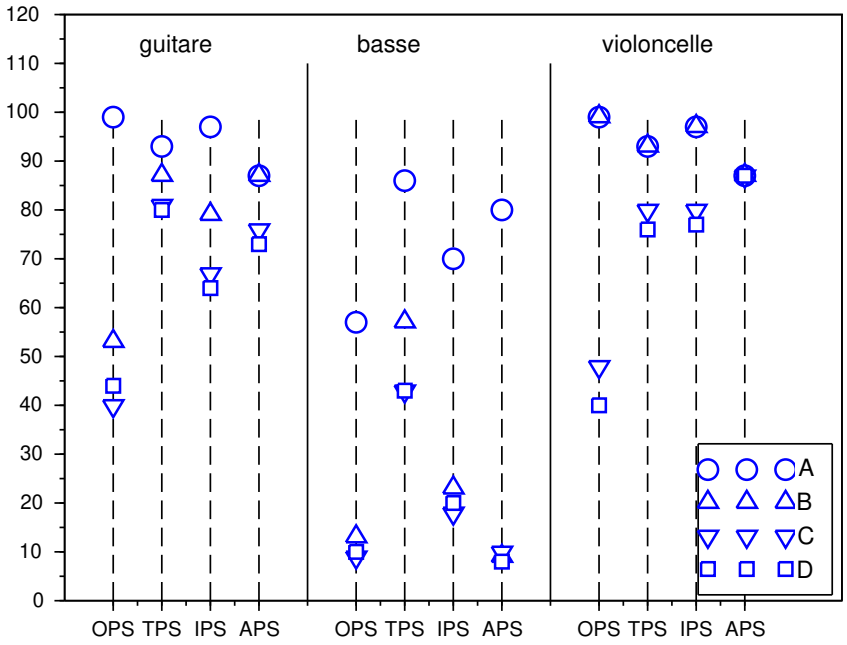
Séparation de sources assistée par tatouage (2)



- 2e idée : en réception, pour chaque (f, t),
  la SCA peut identifier les 2 sources dominantes, mais avec erreurs
  $\rightarrow$ ne transmettre que l'erreur de matrice de dominance estimée
- Intérêt : matrice creuse (pleine de 0), donc compressible
- $\rightarrow$ débit de tatouage $\simeq$ 6 kbit/s par source : c'est encore trop

Séparation de sources assistée par tatouage (3)

- 3e idée : en plus de parcimoniser les sources, on peut les bruiter (imperceptiblement) pour forcer la SCA à identifier sans erreur les sources dominantes
- $\rightarrow$ Il ne reste plus que 1 % de 1 dans la matrice erreur
- $\rightarrow$ débit de tatouage < 1 kbit/s par source : réaliste



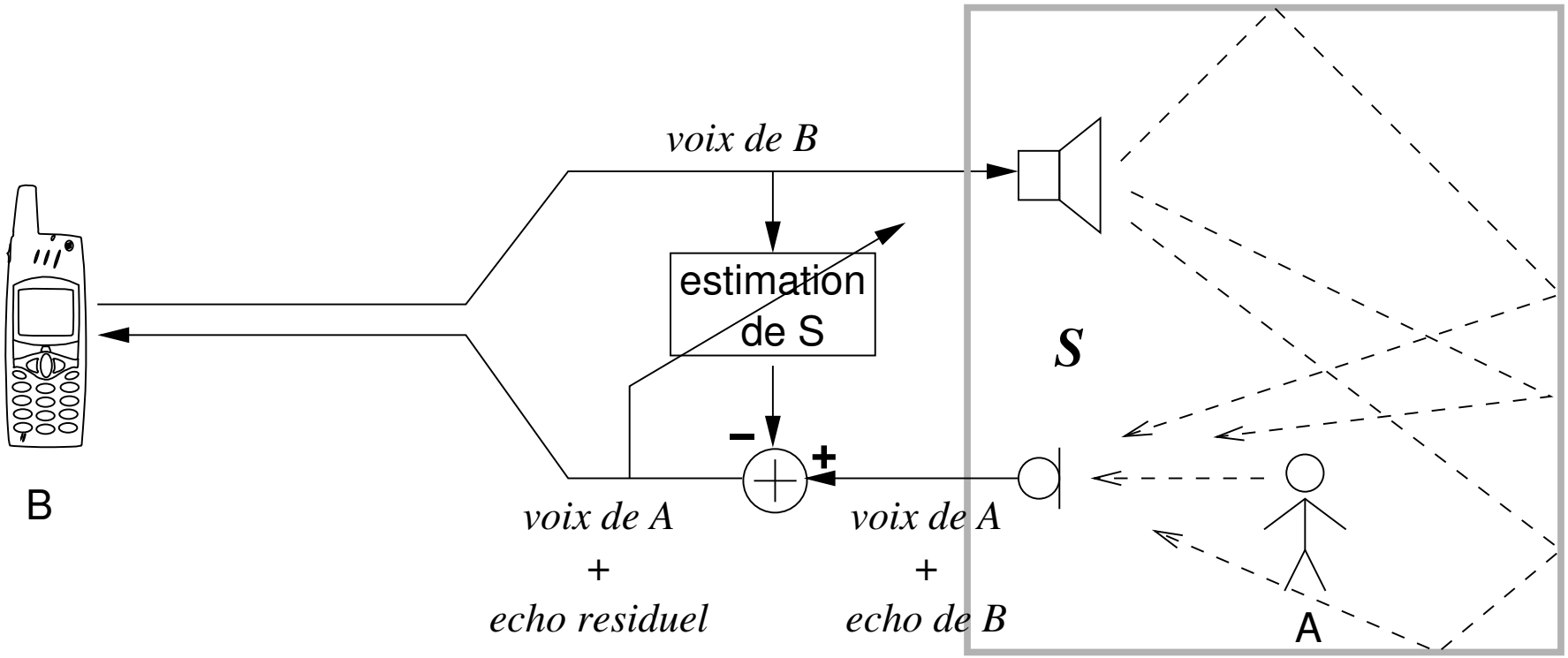
- (A) Sources parcimonisées forcées, information auxiliaire = matrices erreur des matrices de dominance
- (B) Sources parcimonisées forcées
- (C) Sources parcimonisées non-forcées
- (D) Sources originales

Conclusion / tatouage réflexif

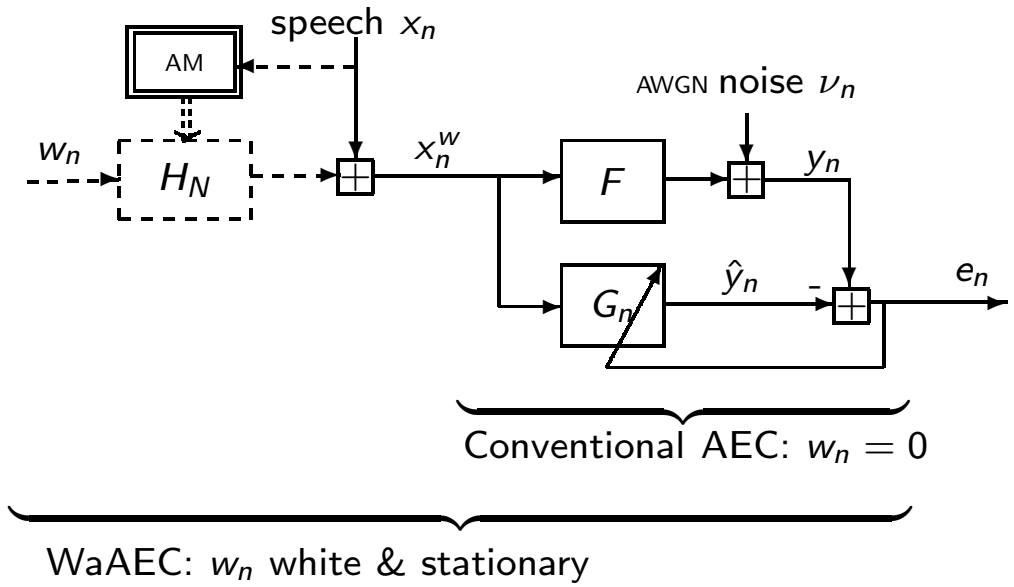
How to build a reflexive WM system for audio ?

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

Contexte applicatif : annulation d'écho acoustique (AEC)



Identification de système linéaire

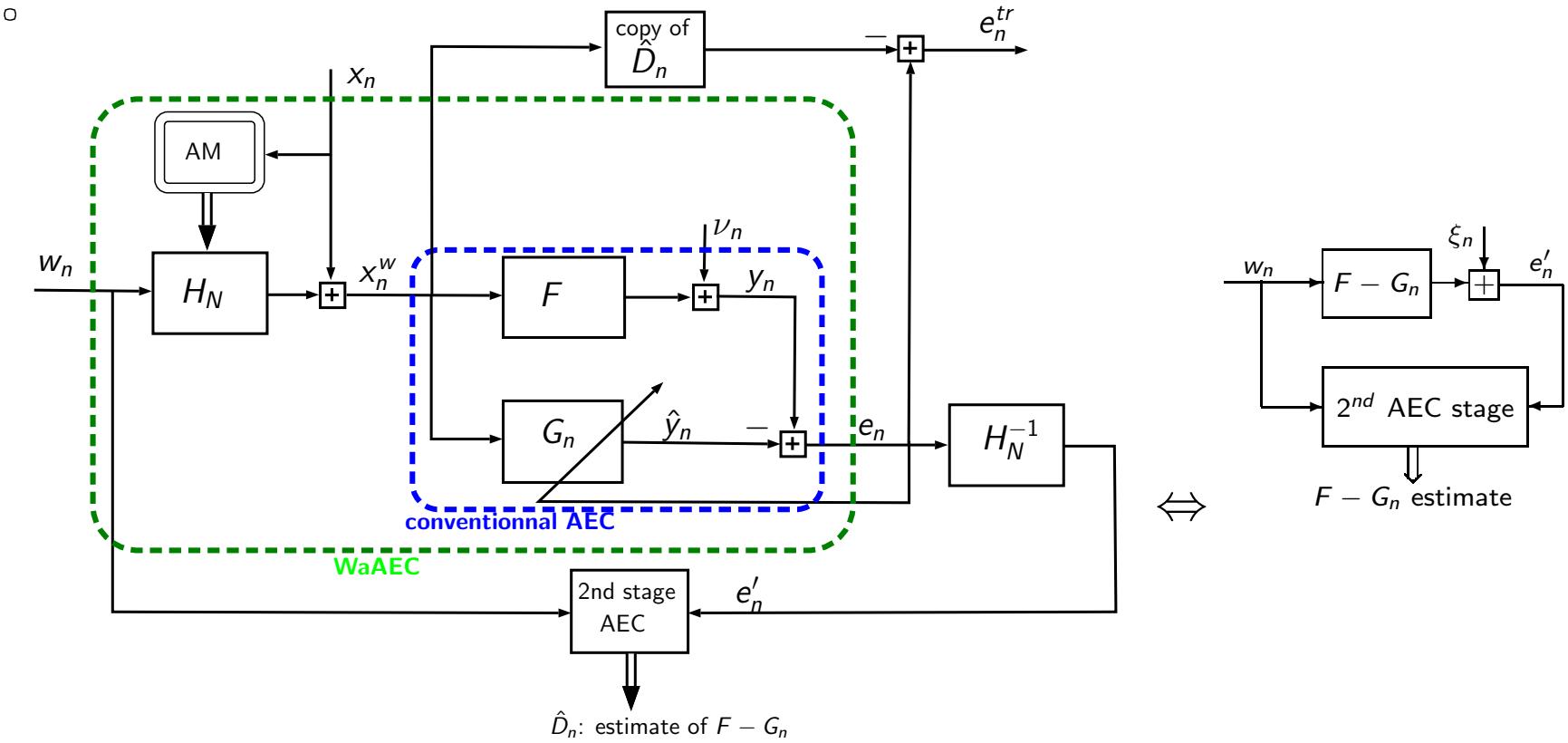


- Identification par filtrage adaptatif : ex. NLMS : $G_{n+1} = G_n + \frac{\mu}{\|X_n\|^2} e_n X_n$
- Vitesse de convergence dépend de la blancheur du signal
- Performances en régime permanent dépendent de la stationarité
- Or le signal de parole n'est ni blanc ni stationnaire...
- 1ère idée : dopage par un bruit inaudible (voir S. Larbi et M. Jaïdane, Audio watermarking : A way to stationarize audio signals. IEEE Trans. Signal Processing, 53(2) :816-823, 2005.) $\rightarrow$ watermark aided AEC (WaAEC) : accélère convergence et réduit l'écho résiduel de 2 à 5 dB en régime permanent

Le tatouage témoin

- Idée : le bruit inséré a toutes les bonnes propriétés...
  Pourquoi ne pas piloter l'identification par le bruit seul ?
- Principe du tatouage témoin :
  il s'imprègne des mêmes altérations que le signal hôte
  $\rightarrow$ comparer le tatouage altéré et sa version originelle permettrait d'identifier ces altérations
- Difficulté : il faut pouvoir extraire de l'écho du signal tatoué l'écho du tatouage seul

Mise en œuvre : le WdAEC



- $e'_n = \underbrace{(f - g_n)}_{d_n} * w_n + \xi_n$, où : $\xi_n = [(f - g_n) * x_n + \nu_n] * h_N^{-1}$
- 2nd étage identifie désalignement $d_n = f - g_n$ du 1er étage
- Écho résiduel transmis : $e_n^{tr} = e_n - \hat{d}_n * x_n^w = [d_n - \hat{d}_n] * x_n^w + \nu_n$

AEC piloté adaptativement par le tatouage (A-WdAEC)

- w_n : bruit blanc gaussien de variance unité.
- 2nd étage = filtre adaptatif selon l'algorithme NLMS
- vitesse de convergence dépend du conditionnement de la matrice de corrélation normalisée du signal de référence,
  - $\mathbf{R_{xw}}(n) = \mathrm{E}\left[\frac{X_n^w (X_n^w)^t}{\|X_n^w\|^2}\right]$ pour le premier étage
  - $\mathbf{R_w}(n) = \mathrm{E}\left[\frac{W_n (W_n)^t}{\|W_n\|^2}\right]$ pour le second étage
  - w_n est blanc alors que x_n^w est fortement coloré,
  - $\rightarrow$ vitesse de convergence du second étage $\gg$ celle du premier.
- Comportement en régime permanent dépend de la puissance instantanée de
  - $\mu\nu_n \frac{X_n^w}{\|X_n^w\|^2}$ pour le premier étage
  - $\mu^w \xi_n \frac{W_n}{\|W_n\|^2}$ pour le second étage,
  - Le bruit ξ_n est plus puissant et moins stationnaire que ν_n,
  - mais $\frac{W_n}{\|W_n\|^2}$ a des variations temporelles plus douces que $\frac{X_n^w}{\|X_n^w\|^2}$
  - Empiriquement : 2nd étage réduit l'écho d'environ 10 dB