

Maltraitement du signal

Partie 4 : du bruit pour révéler le signal

Gaël Mahé ¹

¹Université Paris Cité

octobre 2022

Outline

- 1 Netteté du son
- 2 La Non Intrusive Audio Clarity (NIAC)
- 3 Evaluation : the NIAC as an intelligibility criterion

Plan

- 1 Netteté du son
- 2 La Non Intrusive Audio Clarity (NIAC)
- 3 Evaluation : the NIAC as an intelligibility criterion

Définitions

Intelligibilité et netteté

”La netteté concerne les mesures avec des symboles phonétiques dénués de sens ; l’intelligibilité se détermine, au contraire, avec des mots ou des phrases ayant un sens, support ou véhicules d’une idée qu’il faut comprendre... Le premier objet de la netteté est de ne pas faire intervenir, dans la mesure, la capacité de divination que nous acquérons par la pratique de notre langue et qui nous permet de reconnaître un mot, quoique nous ne l’ayons qu’imparfaitement perçu...”

Chavasse, 1962, cité par Calliope dans *La parole et son traitement automatique*, Masson, 1989.

Bonjour

Tests de netteté (1)

Reconnaissance de logatomes (CCITT, 1960)

- Mots de la forme consonne-voyelle-consonne (CVC)
- phonèmes et orthographe issus de l'esperanto (5 voyelles et 21 consonnes)
- 300 listes de 50 logatomes. Ex : stral ; psuc ; fliv ; sup ; cek ; car ; gleg
- opérateurs entraînés
- test précis mais lourd et onéreux

Tests de netteté (2)

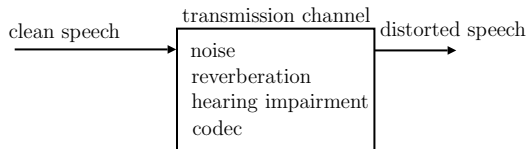
Tests de rime

Objectif : réduire la complexité de la tâche en limitant les choix du sujet

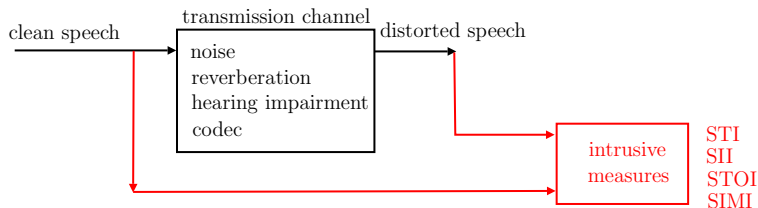
Exemple : Diagnostic Rhyme Test (DRT, 1973)

- Le sujet entend un mot mono-syllabique
- Il choisit entre 2 mots qui riment, dont les consonnes initiales ne diffèrent que d'un trait (voisé ou non, nasal ou non, ...)

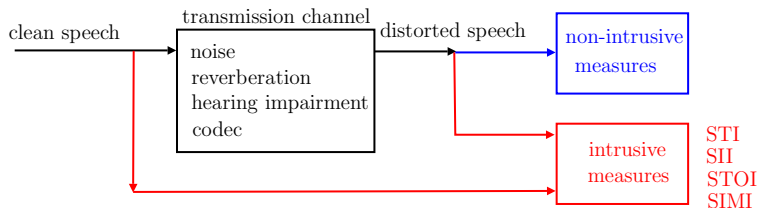
Mesures instrumentales de netteté et d'intelligibilité



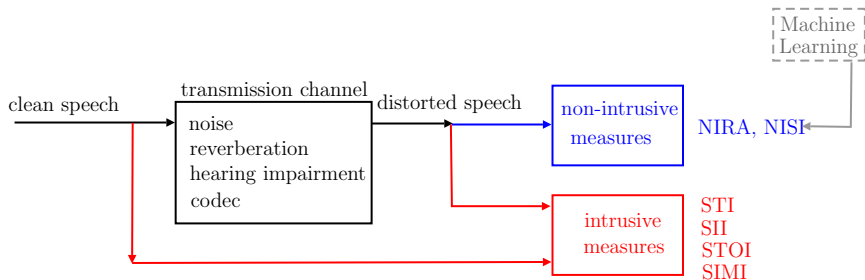
Mesures instrumentales de netteté et d'intelligibilité



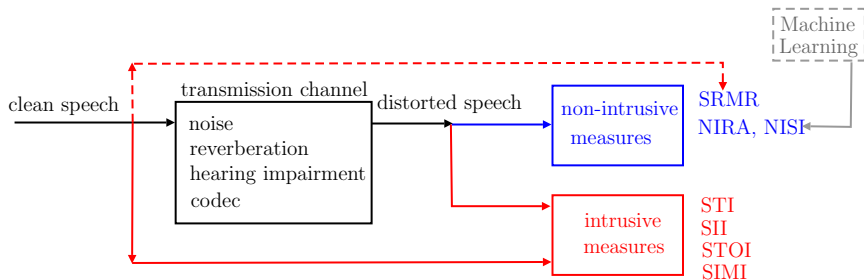
Mesures instrumentales de netteté et d'intelligibilité



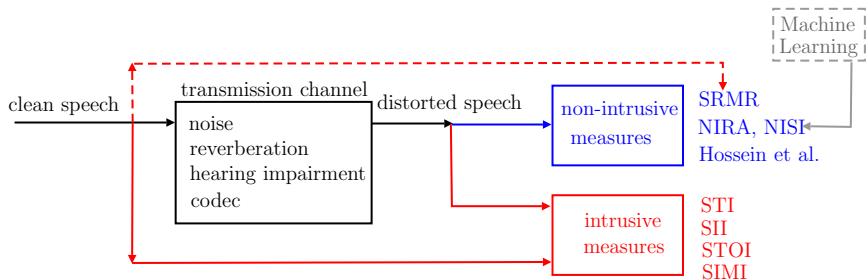
Mesures instrumentales de netteté et d'intelligibilité



Mesures instrumentales de netteté et d'intelligibilité



Mesures instrumentales de netteté et d'intelligibilité



Plan

- 1 Netteté du son
- 2 La Non Intrusive Audio Clarity (NIAC)
- 3 Evaluation : the NIAC as an intelligibility criterion

Principe de la NIAC

- Ici, le bruit n'est ajouté au signal que dans un calcul
- \sim poudre révélatrice d'empreintes digitales
- Idée estimer la netteté d'un son
en mesurant la sensibilité de sa parcimonie à un bruitage

Pourquoi ?

Un détour par le traitement d'images

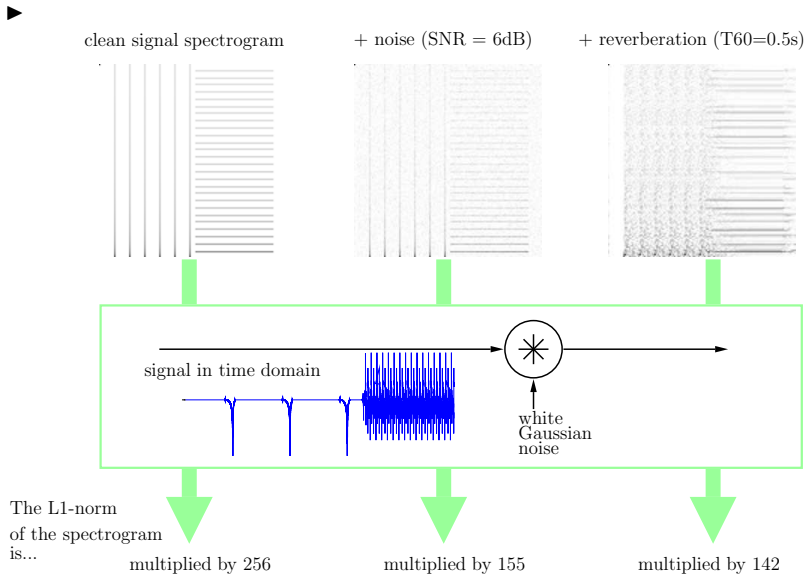
Blanchet, G., Moisan, L., "An Explicit Sharpness Index Related to Global Phase Coherence", Proceedings of Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp. 1065-1068, 2012.

- Gradient d'une image $u(x, y)$:

$$\nabla u(x, y) = \begin{pmatrix} \partial_x u(x, y) \\ \partial_y u(x, y) \end{pmatrix} = \begin{pmatrix} u(x+1, y) - u(x, y) \\ u(x, y+1) - u(x, y) \end{pmatrix}$$

- Régularité d'une image mesurée par la parcimonie de son gradient
- → Mesure de la variation totale : $TV(u) = \sum_{x,y} |\partial_x u(x, y)| + |\partial_y u(x, y)|$
- Plus une image est floue ou bruitée,
plus convoluer une image par un bruit blanc augmente sa TV
- → Mesure de netteté :
Sharpness Index = sensibilité de la TV à convolution par un bruit blanc

Transposition du Sharpness Index à l'audio



Définition de la NIAC (1)

Parcimonie du spectrogramme

mesurée par :

$$\|S\|_1 = \sum_{f,t} |S(f,t)|$$

NIAC = sensibilité de la parcimonie de $S(f,t)$ à la convolution de s par un bruit blanc gaussien

- Soit $s' = s * w$ avec w bruit blanc $\sim \mathcal{N}(0, \sigma_w^2)$
- Soit S' le spectrogramme de s'
- On définit la NIAC comme :

$$-\log \text{Prob}[\|S'\|_1 \leq \|S\|_1]$$

Définition de la NIAC (2)

Assuming that $\|S'\|_1$ is nearly Gaussian,

Audio Sharpness Index

$$NIAC(s) \triangleq -\log \left(\Phi \left(\frac{\mathbb{E}[\|S'\|_1] - \|S\|_1}{\sqrt{\text{Var}[\|S'\|_1]}} \right) \right)$$

where $\Phi =$ tail of the normalized Gaussian distribution.

$\mathbb{E}[\|S'\|_1]$ and $\text{Var}[\|S'\|_1]$ can be computed from

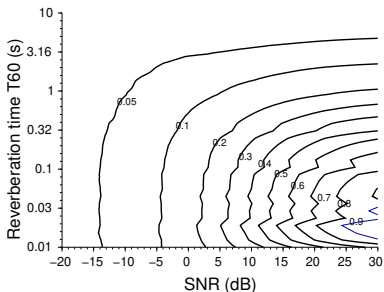
$$\Gamma_{S'}(f, f', \tau) \triangleq \sigma_w^2 \mathfrak{T}[\mathcal{R}_{s,\tau}(n, n')]$$

- for $f \in \{0, 1, \dots, N_f - 1\}$ and $\tau \in \lambda N \mathbb{Z}$
- where $\mathcal{R}_{s,\tau}(n, n') \triangleq R_s(\tau + n - n')h(n)h(n')$
with $R_s =$ auto-correlation of s (finite and deterministic)
- and \mathfrak{T} is the transform used for spectrogram computation

Comparaison de la NIAC à un indicateur de référence

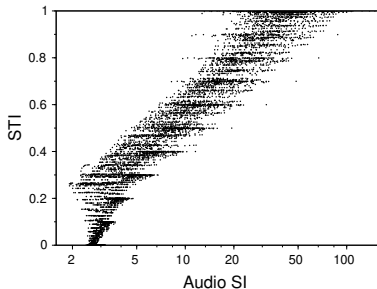
Iso-NIAC lines in the SNR-T60 plane

For each (SNR, T60) condition, the NIAC is averaged over the 16 speakers.



Correlation btw log NIAC and STI

- Each point = one condition (speaker, SNR, T60)
- global correlation coef : 0.96
- $0.94 < \text{individual coef} < 0.98$



Plan

- 1 Netteté du son
- 2 La Non Intrusive Audio Clarity (NIAC)
- 3 Evaluation : the NIAC as an intelligibility criterion

Using the NIAC to separate an audio mixture

Idea

A separated source is more intelligible than a mixture

► drive a separation algorithm by NIAC maximization.

- Instantaneous stereo mix $x = As$,
with unknown $A(., i) = [\cos \theta_i \sin \theta_i]^T$
- To extract one source : $y_\theta = x_1 \sin \theta - x_2 \cos \theta$, with $\theta \in \{\theta_1, \theta_2\}$
- Estimate θ_1 and θ_2 by NIAC maximization :

$$\{\hat{\theta}_1, \hat{\theta}_2\} = \arg \text{local max}_\theta NIAC(y_\theta)$$

Using the NIAC to separate an audio mixture

Idea

A separated source is more intelligible than a mixture

► drive a separation algorithm by NIAC maximization.

- Instantaneous stereo mix $x = As$,
with unknown $A(., i) = [\cos \theta_i \ \sin \theta_i]^T$
- To extract one source : $y_\theta = x_1 \sin \theta - x_2 \cos \theta$, with $\theta \in \{\theta_1, \theta_2\}$
- Estimate θ_1 and θ_2 by NIAC maximization :

$$\{\hat{\theta}_1, \hat{\theta}_2\} = \arg \text{local max}_\theta NIAC(y_\theta)$$

Using the NIAC to separate an audio mixture

Idea

A separated source is more intelligible than a mixture

► drive a separation algorithm by NIAC maximization.

- Instantaneous stereo mix $x = As$,
with unknown $A(., i) = [\cos \theta_i \sin \theta_i]^T$
- To extract one source : $y_\theta = x_1 \sin \theta - x_2 \cos \theta$, with $\theta \in \{\theta_1, \theta_2\}$
- Estimate θ_1 and θ_2 by NIAC maximization :

$$\{\hat{\theta}_1, \hat{\theta}_2\} = \arg \text{local max}_\theta NIAC(y_\theta)$$

Using the NIAC to separate an audio mixture

Idea

A separated source is more intelligible than a mixture

► drive a separation algorithm by NIAC maximization.

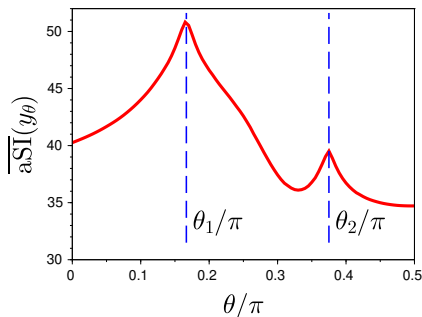
- Instantaneous stereo mix $x = As$,
with unknown $A(., i) = [\cos \theta_i \sin \theta_i]^T$
- To extract one source : $y_\theta = x_1 \sin \theta - x_2 \cos \theta$, with $\theta \in \{\theta_1, \theta_2\}$
- Estimate θ_1 and θ_2 by NIAC maximization :

$$\{\hat{\theta}_1, \hat{\theta}_2\} = \arg \text{local max}_\theta NIAC(y_\theta)$$

Results 1 : speech separation

Average NIAC of the demixed signal y_θ
as a function of the demixing parameter θ :

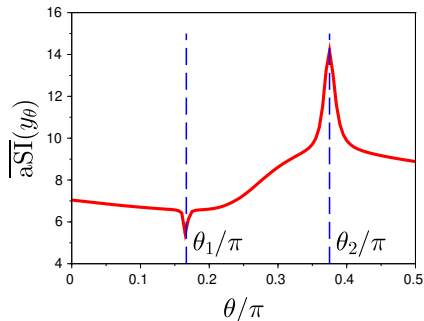
- male + female mix
- sampling rate 16kHz, duration 2.9 s
- $\theta_1 = \pi/6$, $\theta_2 = 3\pi/8$.
- NIAC computed by blocks of 512ms
- averaged on the whole signal $\rightarrow \overline{NIAC}(y_\theta)$



Results 2 : music separation (1)

- singing voice
+ piano accompaniment
- sampling rate 32kHz,
duration 13 s
- $\theta_1 = \pi/6$, $\theta_2 = 3\pi/8$.
- NIAC computed by
blocks of 256ms
- averaged on the whole
signal $\rightarrow \overline{NIAC}(y_\theta)$

Average NIAC of the demixed signal y_θ
as a function of the demixing parameter θ :



Results 2 : music separation (2)

- If the 2 sources have very different NIACs, **the base assumption does not hold anymore**

► Use iterative deflation :

- ① extract voice signal by $\max \overline{NIAC}(y_\theta)$
- ② estimate its contribution to the mixture
- ③ estimate the piano signal by subtraction

• Results :

- estimation error on θ_1 and $\theta_2 < \pi/1000$
- ► song-piano mix
- ► song ► piano from FastICA separation
- ► song ► piano from separation by NIAC maximization

		SDR	SIR	SAR
voice	FastICA	37	37	73
	Max NIAC	70	109	70
piano	FastICA	34	34	67
	Max NIAC	43	43	67

SDR = source to distortion ratio, SIR = source to interference ratio, SAR = source to artifact ratio

Results 2 : music separation (2)

- If the 2 sources have very different NIACs, **the base assumption does not hold anymore**
- ▶ **Use iterative deflation :**
 - ① extract voice signal by $\max \overline{NIAC}(y_\theta)$
 - ② estimate its contribution to the mixture
 - ③ estimate the piano signal by subtraction
- **Results :**
 - estimation error on θ_1 and $\theta_2 < \pi/1000$
 - ▶ song-piano mix
 - ▶ song ▶ piano from FastICA separation
 - ▶ song ▶ piano from separation by NIAC maximization

		SDR	SIR	SAR
voice	FastICA	37	37	73
	Max NIAC	70	109	70
piano	FastICA	34	34	67
	Max NIAC	43	43	67

SDR = source to distortion ratio, SIR = source to interference ratio, SAR = source to artifact ratio