

Projet CENSE


Codeur - Résultats



Félix GONTIER – Étudiant Master ESEO-LS2N
felix.gontier@reseau.eseo.fr

1. Contexte
2. Problématique
3. Méthode
4. Validation
5. Résultats
6. Conclusion

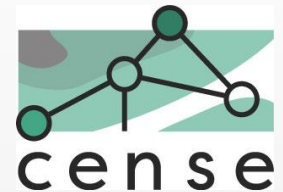


- Fusion IRCCYN – LINA en Janvier 2017
- Unité Mixte de Recherche CNRS (UMR6004)
-  Supervision: CNRS, Université de Nantes, IMT Atlantique, École Centrale de Nantes.
-  Installé à Nantes sur cinq sites différents (direction à l'ECN)
- 5 pôles de compétence, 5 thèmes transverses, 22 équipes rassemblant plus de 450 personnes

Projet CENSE

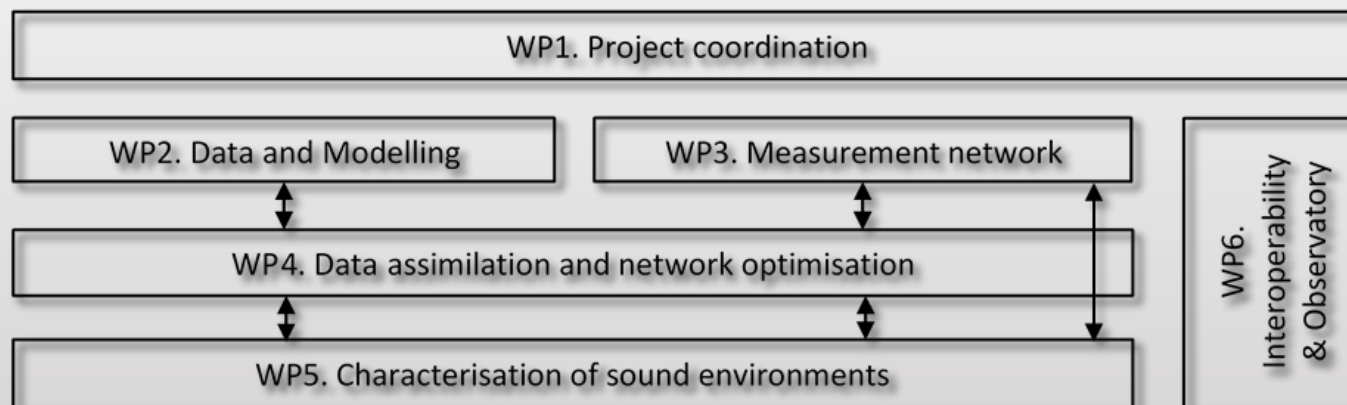
- **Objectif : Comprendre les scènes sonores urbaines**

- Réseau de capteurs acoustiques
- Association à des modèles de prédiction
- Création de cartes de bruit précises



- **Nombreux domaines d'expertises et entités engagés**

Schéma organisationnel du projet



Réseau de capteurs

- **Sujet**

Coder des enregistrements sonores urbains pour en permettre la transmission dans un réseau à large échelle en conservant l'information.

- **Contraintes multiples**



Enregistrements simultanés à large échelle

→ Faible débit de données (1.4 kbps)

Réseau de capteurs bas coût

→ Faible complexité en mémoire/calcul

Traitement des données acquises

→ Calcul d'indicateurs acoustiques

→ Reconnaissance d'évènements

Respect de la vie privée

→ Intelligibilité réduite des extraits de voix

Information nécessaire

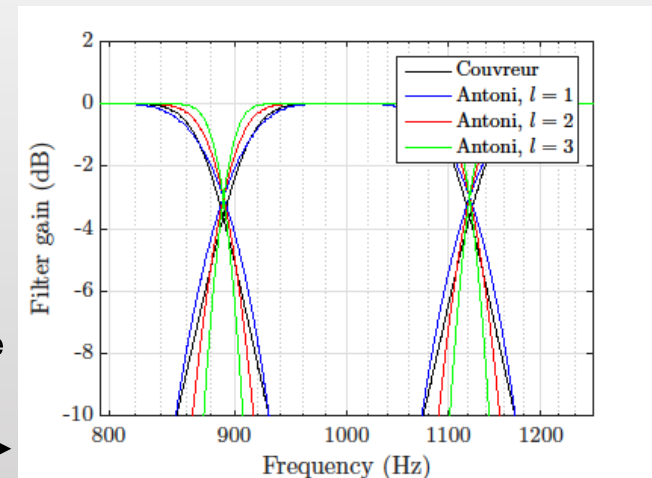
- **Monitoring acoustique des scènes sonores**
 - Calcul d'indicateurs spécifiques (énergétiques, psychoacoustiques)
 - Mesure rapide du niveau sonore équivalent L_{eq} suffisante pour obtenir la plupart des indicateurs
 - **Reconnaissance d'évènements**
 - Extraction de descripteurs de dimensions réduites différenciant au mieux les classes de données
 - Utilisation courante de spectrogrammes et bandes Mel
- Choix des L_{eq} par tiers d'octave pour permettre les deux applications

Représentation



- **Méthode courante: filtrage temporel et décimations successives**
 - Limitée par la conception de filtres dans le domaine temporel
 - S'apparente à une transformée en ondelettes
- **Alternative: transformation du spectre de magnitude**
 - Transformation linéaire dans le domaine fréquentiel
 - Permet des gains en fréquence plus précis
 - Apporte des erreurs causées par la transformée de Fourier
- **Algorithme**
 1. Transformée de Fourier à court-terme
 2. Analyse en tiers d'octave (multiplication matricielle)

Comparaison des gains en fréquence des filtres pour une implémentation temporelle et l'alternative proposée

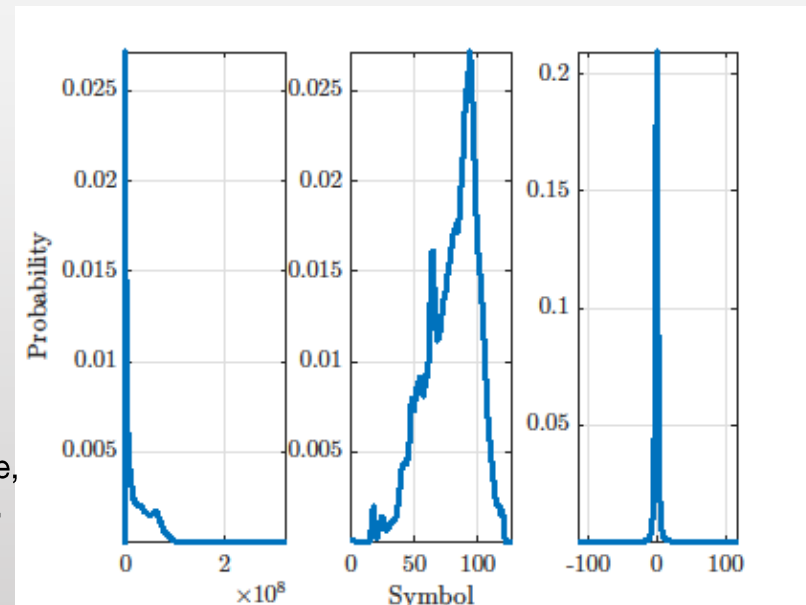


Encodage



- **Codage entropique: Huffman**
 - Étude de la distribution des données
 - Nombre et probabilité des symboles comme double facteur d'efficacité
- **Algorithme**
 1. Échelle logarithmique (dB)
 2. Quantification
 3. Compression Δ
 4. Codage de Huffman

Densités de probabilité au cours de l'algorithme: (gauche) à l'issue de l'analyse, (milieu) après application du logarithme, (droite) après compression Δ .

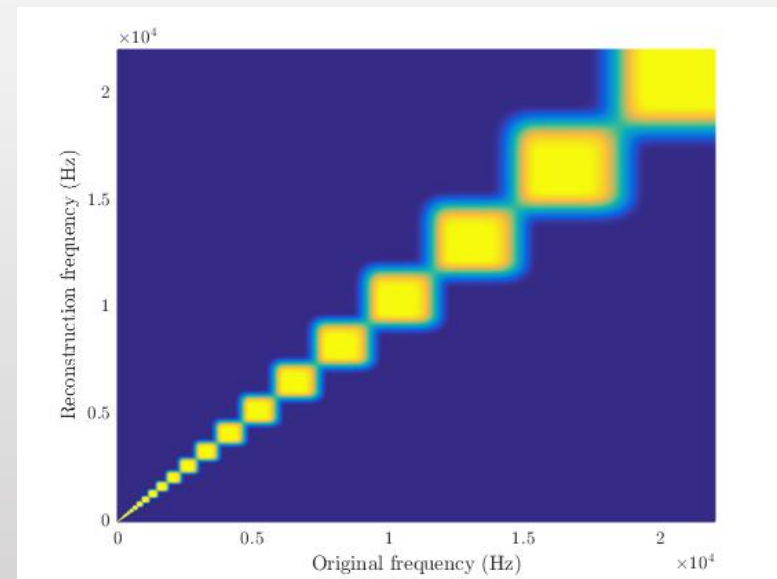


Reconstruction

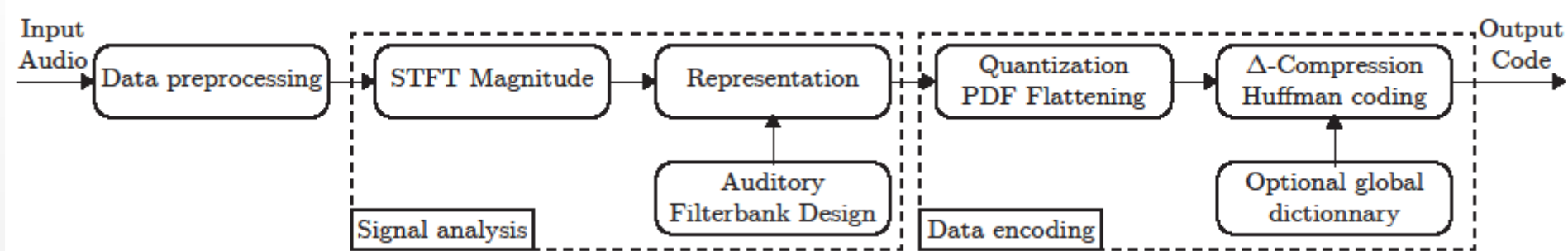


- **Récupération du spectrogramme linéaire**
 - Décodages Δ et Huffman triviaux
 - Opération directe d'analyse en tiers d'octave non réversible
 - Approximation: transposée de la matrice de transformation
 - Effet de perte de résolution dans les hautes fréquences
- **Reconstruction de phase**
 - Algorithme de Griffin&Lim
 - Spectrogramme de bruit

Transformation liant le spectrogramme d'origine au spectrogramme reconstruit.



Paramétrage



- **Analyse**
 - Résolution temporelle (durée, recouvrement)
 - Fonction de fenêtrage
- **Encodage**
 - Quantification



➤ Faible nombre de facteurs permettant un compromis entre les contraintes

- **Objectif: Valider la capacité de l'algorithme proposé à répondre à la problématique et ses contraintes**

- **Métriques**



- Erreur de mesure des niveaux sonores par tiers d'octave
- Performance de classification
- Débit de données en sortie du codeur
- Évaluation de l'intelligibilité

- **Données de test**

- Dataset d'enregistrements sonores urbains: UrbanSound8k
8732 fichiers (9h), 10 classes, 10 folds, résultats baseline
- Dataset de voix
9 phrases, 6 locuteurs, conditions d'enregistrements de studio

Erreur de mesure

- **Erreur d'analyse**
 - Implémentation de référence: *ita_toolbox*
 - Calcul des indicateurs sur plusieurs secondes de signal
 - Évaluation double: par rapport à une analyse complète ou suivant les mêmes paramètres
 - Estimation sur bruit blanc et enregistrements urbains
- **Erreur d'encodage**
 - Seule opération à pertes: processus de quantification



Reconnaissance d'évènements

- **Descripteurs**

- Comparaison bandes tiers-d'octave – bandes Mel
- Cepstrogrammes (25 coefficients DCT)
- Statistiques sur le temps (11 opérateurs)

- **Classifieurs**

- Support Vector Machine (C-SVM, noyau RBF)
- Decision Tree
- Random Forests (500 arbres)
- K-Nearest Neighbors ($k = 5$)



- **Validation croisée**

Intelligibilité

- **Indicateurs objectifs: CSII et fwSNRseg**
 - Comparaison signal original – signal décodé
 - Utilisation des spectrogrammes (Magnitude ou complet)
 - Corrélation démontrée avec l'intelligibilité subjective
 - Uniquement testé pour de faibles distorsions
- **Test perceptif**
 - 12 participants, 18 phrases avec différents paramètres d'encodage
 - Intelligibility Ratio (IR): Ratio de bonne transcription des mots
 - Average Intelligibility Score (AIS): Note globale subjective

Erreur de mesure

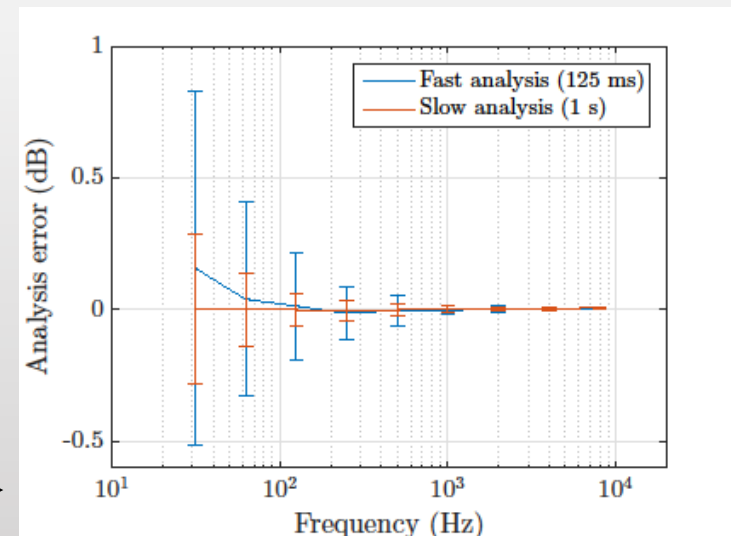
- **Bruit blanc**

- Erreur centrée sur zéro: biais d'estimation faible
- Variance décroissant en fonction de la fréquence



➤ Erreurs liées à la résolution d'analyse lors de la transformée de Fourier (périodicité, effets de fuite)

Erreur d'analyse en fonction de la fréquence pour des signaux de bruit blanc. Les bandes tiers d'octave sont calculées pour un fenêtrage rectangulaire et aucun recouvrement.

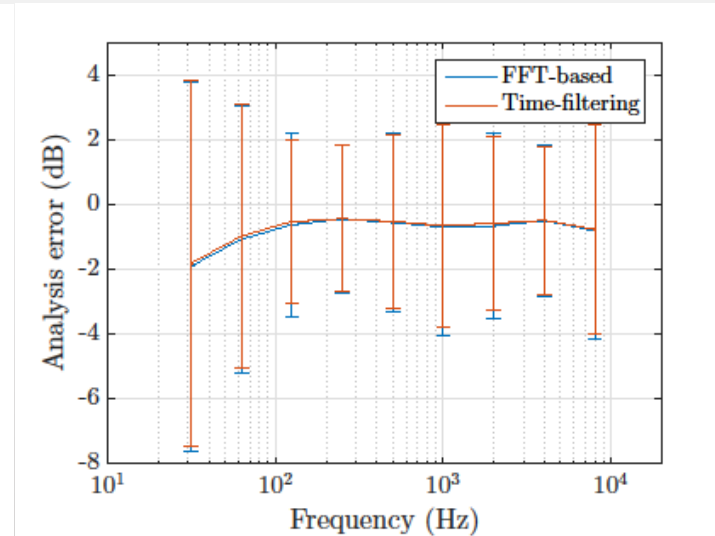
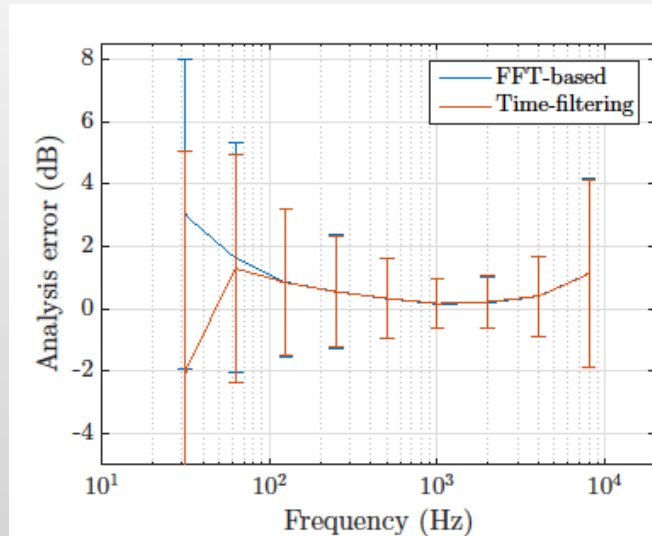


Erreur de mesure

- **Enregistrements sonores urbains**
 - Erreur faible dans les fréquences moyennes, importante dans les premières bandes
 - Corrélation entre les deux implémentations

Erreur d'analyse pour des enregistrements sonores urbains.

(gauche) Fenêtre rectangulaire de 125 ms, pas de recouvrement. (droite) Fenêtre de Hann de 1 s, recouvrement de 66%.

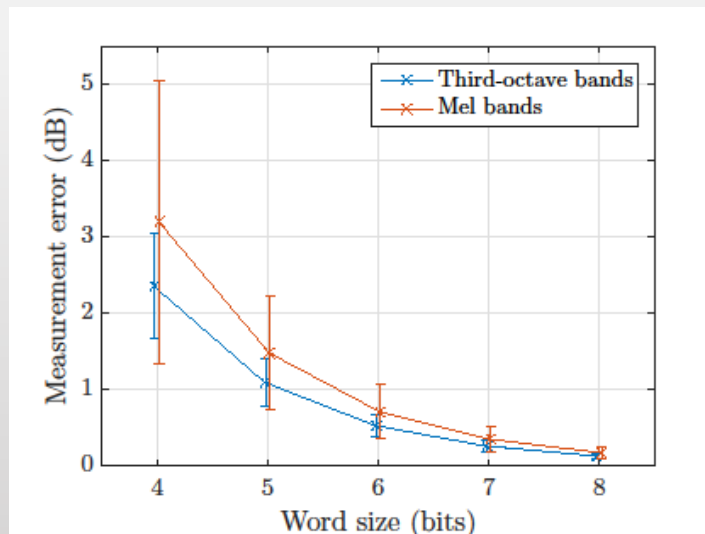


Erreur de mesure

- Erreur d'encodage

- Fonction de la distribution des données
- Proche de l'approximation théorique (décroissance exponentielle)
- Négligeable pour des pas de quantification faibles

Erreur d'encodage en fonction de la taille de mot utilisée.



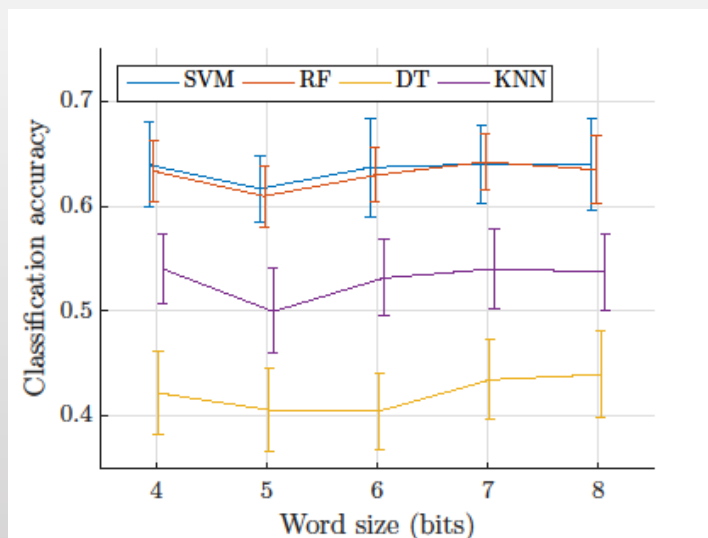
Reconnaissance d'évènements

- Impact de la quantification négligeable
- Efficacité des Mel et tiers d'octave similaires
- Possibilité de réduire la résolution temporelle d'analyse sans effet majeur sur les performances

Performances de classification pour des coefficients cepstraux à base tiers d'octave, en fonction du pas de quantification. →

Performances de classification pour des MFCC en fonction de la résolution d'analyse.

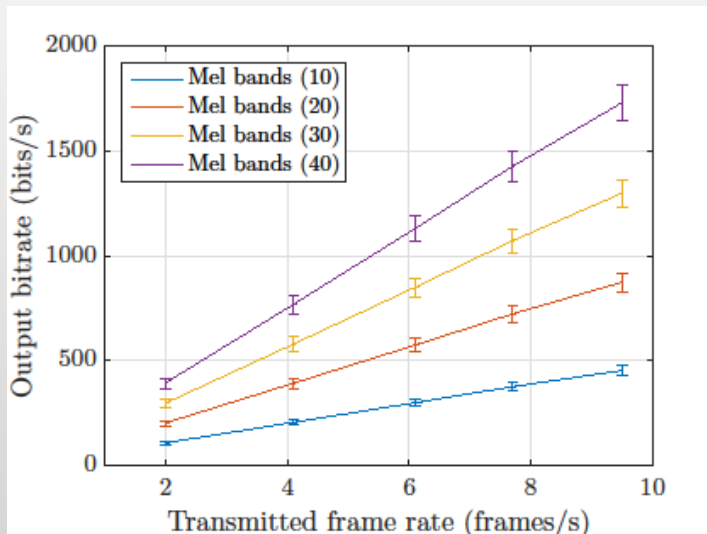
SVM		Frames per second						
		2	4 (4.1)	6 (6.1)	8 (7.7)	10 (9.5)	20 (21)	85
Mel bands	10	55±3	60±3	61±4	62±3	62±4	63±6	65±6
	20	58±4	62±4	63±4	64±4	63±4	65±5	67±6
	30	60±3	64±4	64±4	65±4	65±3	67±4	68±4
	40	60±3	63±4	64±4	64±4	64±4	66±4	68±5



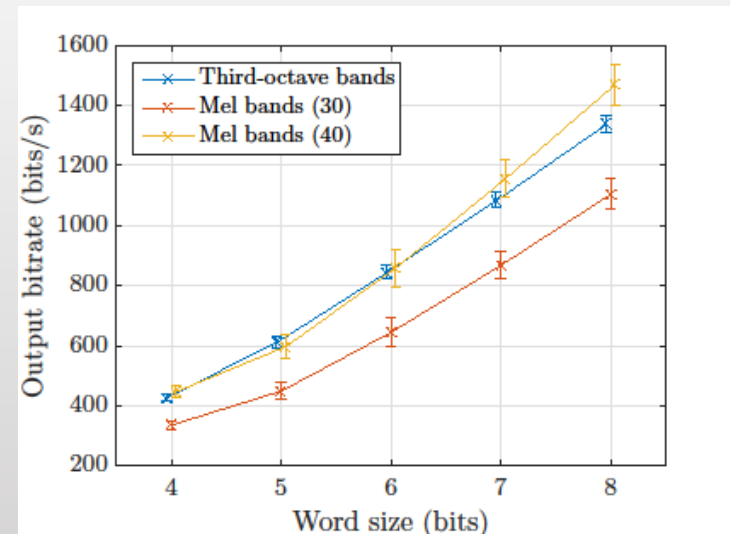
Débit de données

- Relation simple entre débit et dimensions de l'analyse
- Débit acceptable (< 1.5 kbps) jusqu'à 8 fenêtres par seconde

Débit de sortie en fonction des dimensions de la représentation.



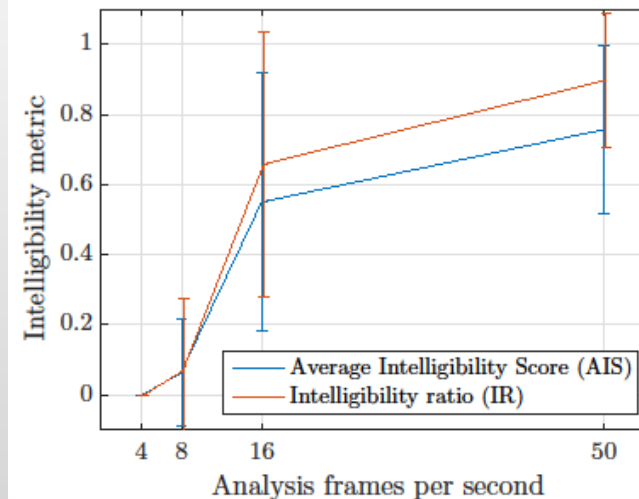
Débit de sortie en fonction du pas de quantification.



Intelligibilité

- **Test perceptif**
 - Nombre moyen de phonèmes par seconde
 - Corrélation des deux métriques
 - Intelligibilité faible pour une analyse « fast » avec 8 fenêtres par seconde

Indicateurs AIS et IR en fonction du nombre de fenêtres d'analyse par seconde



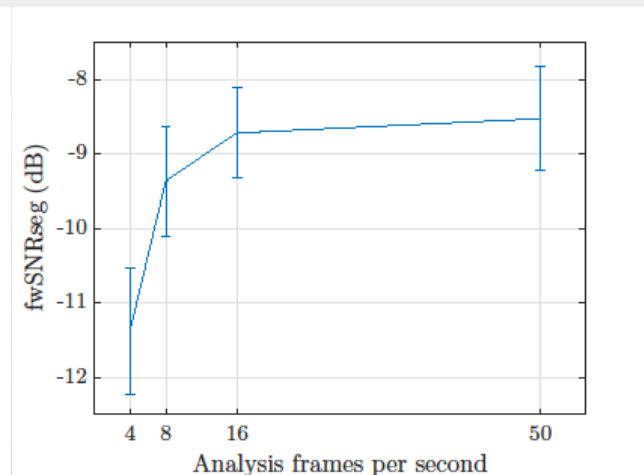
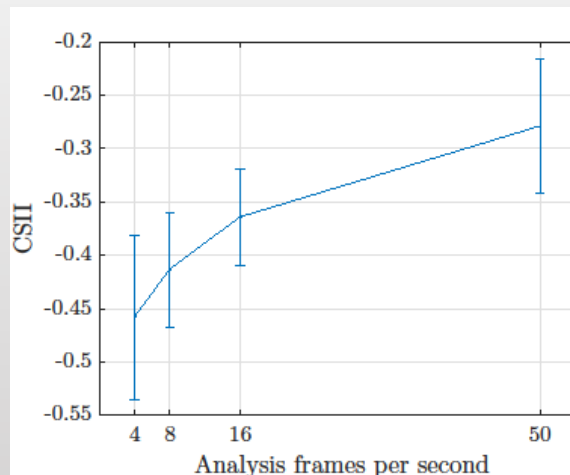
Intelligibilité

- **Indicateurs objectifs**

- Faible corrélation avec les résultats du test perceptif
- Absence de seuil d'intelligibilité
- Comparaison point par point des spectrogrammes

➤ Distorsions trop importantes pour que ces indicateurs soient efficaces

Indicateurs CSII (gauche) et fwSNRseg (droite) en fonction de la résolution temporelle d'analyse.



- Le codeur proposé permet de répondre aux problématiques exposées, avec un compromis entre complexité, débit de données et conservation de l'information.
- Les paramètres restent à définir lors de l'implémentation sur les capteurs. Le développement futur de cette partie du projet sera suivi avec intérêt.
- Un article scientifique a été écrit et sera proposé à la publication dans les prochains mois, ainsi que les ressources nécessaires à la reproduction.

- ANSI S1.1-1986, (ASA 65-1986), Specifications for Octave-Band and Fractional-Octave-Band Analog and Digital Filters, 1993.
- J. Antoni. Orthogonal-like fractional-octave-band filters. J. Ac. Soc. Am., 127(2):884-895, 2010.
- M. Berzborn, R. Bomhardt, J. Klein, J.-G. Richter, and M. Vorländer. The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing. 43th Annual German Congress on Acoustics, Kiel (Germany), 6 Mar 2017 - 9 Mar 2017, Mar 2017.
- S. Davis. Octave and fractional-octave band digital filtering based on the proposed ANSI standard. In 1986 IEEE International Conference on Acoustics, Speech and Signal Processing, 1986.
- D. Grin and J. Lim. Signal estimation from modified short-time Fourier transform. IEEE Transactions on Acoustics, Speech, and Signal Processing, 32(2):236-243, 1984.
- Y. Hu and P. Loizou. Evaluation of objective quality measures for speech enhancement. IEEE Transactions on Audio, Speech, and Language Processing, 16(1):229-238, 2008.
- J. Kates and K. Arehart. Coherence and the speech intelligibility index. J. Ac. Soc. Am., 115(5):2224-2237, 2005.
- J. Salamon, C. Jacoby, and J. Bello. A dataset and taxonomy for urban sound research. In 22nd ACM international conference on Multimedia, 2014.