

**Titre en français :** Vers un nouveau modèle auditif : décomposition empirique des signaux acoustiques par interpolations d'enveloppe adaptative

**Titre en anglais :** Towards a new auditory model: empirical decomposition of acoustic signals by adaptive envelope interpolations

**Nom du directeur de thèse :** Richard Kronland-Martinet (PRISM / CNRS)

**Tel :** 04 91 16 42 50

**E-Mail :** kronland@prism.cnrs.fr

**Co-direction de thèse :** Etienne Thoret (INT / CNRS)

**Laboratoire :** PRISM UMR 7061

**Financement :** demandé

**Type de financement :** contrat doctoral "Établissement Aix-Marseille Université"

### **Résumé en français :**

Les modèles auditifs actuels font l'hypothèse que l'oreille décompose les sons selon des bancs de filtres cochléaires caractérisés à partir d'études psychoacoustiques (Saremi et al., 2016). Si ces modèles ont montré leur efficacité pour modéliser un grand nombre de phénomènes perceptifs, ils ne rendent pas compte des phénomènes non linéaires et adaptatifs de l'oreille. Ces phénomènes se manifestent par exemple par la perception de composantes fréquentielles qui ne sont pas présentes dans le spectre du son (fondamentale manquante, sons de Tartini, ...). Actuellement ces phénomènes sont souvent modélisés en ajoutant a posteriori des corrections non linéaires en sortie des modèles linéaires.

Afin de proposer un modèle auditif qui unifie ces phénomènes tout en étant plausible du point de vue anatomique, nous avons récemment proposé au laboratoire PRISM une approche alternative et novatrice (Thoret et al., 2023) basée sur une décomposition en Modes Empiriques (méthode EMD, Huang et al., 1998). Cette approche consiste à décomposer le signal temporel en modes obtenus à partir d'interpolations sur ses extrema, en accord avec les observations anatomiques de la cochlée. Les premiers résultats montrent que cette nouvelle approche adaptative et attachée aux données permet de proposer un modèle auditif qui englobe aussi bien les phénomènes linéaires que non-linéaires, ouvrant ainsi la voie à une vision innovante et en rupture pour les applications liées à l'audition et en particulier à sa remédiation.

La thèse examinera en détail les principes théoriques sous-jacents de la décomposition empirique des signaux en se focalisant dans un premier temps sur la conception de nouvelles méthodes d'interpolation permettant de contrôler les caractéristiques des modes empiriques. En effet ces interpolations affectent grandement les propriétés de la décomposition en modes empiriques et in fine les caractéristiques du filtrage adaptatif résultant. Dans un second temps, la calibration des paramètres d'interpolation sera effectuée afin que la décomposition se comporte de façon équivalente à celle réalisée par le système auditif humain. Pour cela, il s'agira, par exemple, de caler les réponses impulsionnelles des bancs de filtres équivalents de la décomposition avec celles des filtres cochléaires largement décrits dans la littérature (par exemple gammatone), puis de contraindre l'optimisation du processus pour retrouver l'ensemble des caractéristiques perceptives (séparation de lignes spectrales, rugosité perceptive, effets de masquage, perception de l'inharmonicité, ...). Enfin, sur la base de ce nouveau paradigme de représentation des sons, la thèse explorera les différentes applications

potentielles liées à cette approche, notamment pour le traitement des signaux audio et cela afin d'améliorer, par exemple, les méthodes de séparation de sources perceptive dans le cas de l'effet « cocktail party » ou pour optimiser les processus d'amplification et d'intelligibilité des sons. Les résultats obtenus seront évalués à travers des expériences numériques mais également par des expérimentations sur des sujets humains, démontrant ainsi l'efficacité et la pertinence de cette nouvelle approche pour l'ingénierie de la perception auditive.

### **Résumé en anglais :**

Current auditory models assume that the ear decomposes sounds according to banks of cochlear filters characterized from psychoacoustic studies (Saremi et al., 2016). While these models have proven effective in modeling a wide range of perceptual phenomena, they do not account for the nonlinear and adaptive phenomena of the ear. These phenomena are manifested, for example, by the perception of frequency components that are not present in the sound spectrum (missing fundamental, Tartini tones, etc.). Currently, these phenomena are often modeled by adding nonlinear corrections post hoc to linear models.

In order to propose an auditory model that unifies these phenomena while being anatomically plausible, we recently proposed an alternative and innovative approach at the PRISM laboratory (Thoret et al., 2023) based on Empirical Mode Decomposition (EMD method, Huang et al., 1998). This approach involves decomposing the temporal signal into modes obtained from interpolations on its extrema, in accordance with anatomical observations of the cochlea. Initial results show that this new adaptive and data-driven approach allows for proposing an auditory model that encompasses both linear and nonlinear phenomena, thus paving the way for an innovative and disruptive vision for applications related to hearing and particularly its remediation.

The thesis will examine in detail the underlying theoretical principles of empirical signal decomposition, focusing initially on the design of new interpolation methods to control the characteristics of empirical modes. Indeed, these interpolations greatly affect the properties of empirical mode decomposition and ultimately the characteristics of the resulting adaptive filtering. Secondly, the calibration of interpolation parameters will be carried out so that the decomposition behaves equivalently to that performed by the human auditory system. For example, this will involve aligning the impulse responses of equivalent filter banks of the decomposition with those of cochlear filters widely described in the literature (e.g., gammatone), and then constraining the optimization process to retrieve all perceptual characteristics (spectral line separation, perceptual roughness, masking effects, perception of inharmonicity, etc.). Finally, based on this new paradigm of sound representation, the thesis will explore various potential applications related to this approach, notably for audio signal processing, to improve, for example, perceptual source separation methods in the case of the "cocktail party" effect or to optimize sound amplification and intelligibility processes. The results obtained will be evaluated through numerical experiments but also through experiments on human subjects, thus demonstrating the effectiveness and relevance of this new approach for auditory perception engineering.

### **Profil du candidat recherché :**

Titulaire d'un Master, d'un diplôme d'Ingénieur ou équivalent Bac+5 dans le domaine de du traitement du signal audio, informatique, l'acoustique, ou physique. Des expériences en optimisation et/ou en psychoacoustiques seront un plus.

### **Publications sur le sujet :**

Saremi, A., Beutelmann, R., Dietz, M., Ashida, G., Kretzberg, J., & Verhulst, S. (2016). A comparative study of seven human cochlear filter models. *The Journal of the Acoustical Society of America*, 140(3), 1618-1634.

Thoret, E., Ystad, S., & Kronland-Martinet, R. (2023). Hearing as adaptive cascaded envelope interpolation. *Communications Biology*, 6(1), 671.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., ... & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971), 903-995.

**Insertion professionnelle après thèse :**

Laboratoires de recherches académique ou grande institution du domaine du traitement du signal, de l'informatique, ainsi que le CNRS et les Universités des sciences en général mais aussi de nombreuses entreprises privées dans le domaine du signal.