

L'intelligence artificielle au service du diagnostic acoustique : Apprendre à entendre les parois d'une salle

Offre de thèse

Durée : 3 ans

Encadrants : Cédric Foy et Antoine Deleforge (Chargés de recherche).

Contacts : cedric.foy@cerema.fr, antoine.deleforge@inria.fr

Emplacement principal: Cerema Est - 11 Rue Jean Mentelin, 67200 Strasbourg

Equipes: site de Strasbourg de l'UMRAE¹ et équipe MULTISPEECH de l'INRIA de Nancy²

Prérequis: Master en science des données/machine learning **ou** Master en audio/acoustique. Des connaissances ou de l'intérêt pour les deux domaines à la fois sont un grand plus.

Contexte et objectif :

Les nuisances sonores sont citées comme première source de gêne par les populations et constituent un **enjeu sanitaire et social important**, contribuant notamment au stress, aux troubles du sommeil, aux déficits d'attention en classe, à l'hypertension ou aux acouphènes³. La gêne sonore peut être liée au mauvais isolement acoustique d'un local par rapport à l'extérieur (bruits des transports) ou par rapport aux locaux avoisinants (bruit de voisinage). Elle peut aussi être directement liée à la mauvaise qualité acoustique de cette salle due à une réverbération trop importante (cantine, piscine, crèche...). Dans le cadre de la **réhabilitation acoustique** des bâtiments existants, les solutions acoustiques sont difficiles à définir et à optimiser car les contraintes sont nombreuses. La méthode utilisée par l'acousticien de bureau d'études est de réaliser des mesures audio (réponses impulsionnelles) de l'existant sur site puis d'utiliser ces mesures pour caler les **paramètres acoustiques d'entrée** (puissance acoustique des sources, isolement, absorption, superficie et diffusion des parois...) des modèles analytiques (théorie de Sabine) ou numériques (approche de particules). La connaissance de ces paramètres est nécessaire pour ensuite pouvoir quantifier le gain acoustique de la solution de réhabilitation proposée. Or, ce calage est **long et coûteux** car il se fait principalement par **approche itérative** et repose *in fine* sur l'expérience de l'acousticien.

Face à ce constat, le développement de méthodes permettant de **remonter automatiquement** aux paramètres d'entrée à partir de mesures audio constituerait une percée majeure pour l'acoustique du bâtiment. Il s'agit d'un **problème de physique inverse non-linéaire difficile** et ouvert à ce jour. Cette thèse s'attaquera à ce défi en tirant profit de **l'immense potentiel de l'intelligence artificielle**, et en particulier de l'apprentissage automatisé, qui connaît des avancées fulgurantes ces dernières années grâce aux progrès du deep learning, mais n'a encore que très peu percé dans le domaine de l'acoustique. La problématique centrale de la thèse est la suivante :

*Est-il possible, en utilisant des **mesures ponctuelles de champ sonore** obtenues à partir de microphones dans une salle, d'estimer précisément les **paramètres acoustiques d'entrée** de cette salle par **apprentissage automatisé préalable** sur larges jeux de données ?*

¹<http://www.umrae.fr/>

²<https://team.inria.fr/multispeech/>

³<https://www.eea.europa.eu/themes/human/noise>

La thèse sera co-encadrée par **l'équipe MULTISPEECH de l'Inria de Nancy**, spécialisée dans l'apprentissage automatisé pour le traitement du son et de la parole, et **l'UMRAE de Strasbourg**, qui bénéficie à la fois d'une forte expérience pratique et théorique en acoustique du bâtiment et d'un nombre important de données mesurées et simulées, accumulées sur des décennies.

Description du sujet de thèse :

Les modèles utilisés aujourd'hui en acoustique du bâtiment se divisent en approches ondulatoires, intraitables numériquement aux hautes fréquences, ou énergétiques, plus accessibles mais approximatives. Dans les deux cas, il n'existe à ce jour aucune méthode générique pour **inverser ces modèles**, c'est-à-dire, remonter aux paramètres acoustiques d'entrées à partir de mesures audio. L'apprentissage automatisé, encore peu exploité dans le domaine de l'acoustique, possède un fort potentiel pour la résolution de problèmes inverses non-linéaires comme celui-ci. En effet, à partir d'un large jeu de données d'entrées et de sorties dit *d'entraînement*, ces méthodes permettent d'estimer une fonction non-linéaire reliant les deux, par exemple avec l'aide d'un réseau de neurones profond. Des premiers travaux ont exploité cette approche pour l'estimation de l'absorption moyenne [1], du volume [2], ou des échos précoces [3] dans une salle à partir d'enregistrements de sources sonores inconnues. Dans le cadre de cette thèse, des jeux de données simulées [4] et réelles seront exploités pour apprendre à estimer automatiquement des paramètres acoustiques tels que la géométrie, l'absorption ou la diffusivité [5] des parois d'une salle à partir d'une ou plusieurs mesures de *réponses impulsionnelles*. Les verrous scientifiques inclus :

- Comment combiner des simulateurs acoustiques de pointe et des mesures réelles pour obtenir des jeux de données d'apprentissage suffisamment larges, riches, réalistes, et précisément annotés? L'**apprentissage par transfert** [6] du simulé au réel est-il possible?
- Quels paramètres d'entrées des modèles physiques peuvent être déduits à partir de réponses impulsionnelles seules? Peut-on dépasser les limites des approches énergétiques classiques (ex: **milieux inhomogènes** [7]) en exploitant des indices fins tels que les échos précoces?
- Quelles indices d'entrées et quelles architectures pour les réseaux de neurones utilisés? Comment y **injecter des connaissances physiques** [8]? Comment prendre en compte dans les fonctions de coûts optimisées les invariants fondamentaux du problème tels que la **permutation des parois** d'une salle [9]?
- Comment gérer le **bruit**, l'**incertitude** et les **facteurs latents non-annotés** (ex: occlusions, parois mixtes) dans les modèles d'apprentissage? Leurs effets peuvent-ils être dissipés par l'**exploitation de plusieurs mesures** et de **modèles génératifs** [10]?

Tâches et déroulement de la thèse:

Le doctorant effectuera les tâches suivantes :

- Etudes bibliographiques
- Synthèse et collecte de données
- Elaboration de modèles physiques et appris
- Prétraitement des données et entraînement
- Validation des méthodes
- Publication des travaux
- Rédaction de la thèse

La thèse sera co-encadrée par Cédric Foy pour la partie acoustique, et Antoine Deleforge pour les aspects traitement audio et machine learning. Elle se déroulera principalement au Cerema à Strasbourg en compagnie de Cédric Foy. Antoine Deleforge, qui travaille à l'Inria de Nancy, sera présent deux fois par semaine au Cerema. Des déplacements ponctuels du doctorant vers l'Inria de Nancy, à la journée, seront également à envisager.

Bibliographie:

- [1] Saurabh Kataria, Clément Gaultier, and Antoine Deleforge. "Hearing in a shoe-box: binaural source position and wall absorption estimation using virtually supervised learning." *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.
- [2] Andrea F. Genovese, Hannes Gamper, Ville Pulkki, Nikunj Raghuvanshi, and Ivan J Tashev, "Blind room volume estimation from single-channel noisy speech," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.
- [3] Diego Di Carlo, Antoine Deleforge, and Nancy Bertin, "Mirage: 2d source localization using microphone pair augmentation with echoes," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.
- [4] S. M. Schimmel, M. F. Muller, and N. Dillier. "A fast and accurate "shoebox" room acoustics simulator." *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009.
- [5] Cedric Foy, Judicael Picaut, and Vincent Valeau, "Including the wall diffusivity within the room acoustics diffusion model: an analytical approach," *Journal of the Acoustical Society of America*, 2016.
- [6] Lisa Torrey and Jude Shavlik, "Transfer learning," *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, 2010.
- [7] Huw G Davies, "Noise propagation in corridors," *Journal of the Acoustical Society of America*, 1973.
- [8] Anuj Karpatne, William Watkins, Jordan Read, and Vipin Kumar, "Physics-guided neural networks (PGNN): An application in lake temperature modeling," *arXiv preprint 1710.11431*, 2017.
- [9] Dong Yu, Morten Kolbæk, Zheng-Hua Tan, and Jesper Jensen, "Permutation invariant training of deep models for speaker-independent multi-talker speech separation," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.
- [10] Diederik P. Kingma and Max Welling, "Auto-encoding variational Bayes," *arXiv preprint 1312.6114*, 2013.