2018-00724 - PhD Position/ Deep learning-based phase-aware audio signal modeling and estimation [S]

Type de contrat : CDD de la fonction publique
Niveau de diplôme exigé : Bac + 5 ou équivalent
Fonction : Doctorant

Mission confiée

Decades of research in audio signal processing have brought tremendous progress in sound source enhancement, recognition and localization, to the point that this technology is now available on most smartphones. Yet, quality is still far from perfect, especially in challenging scenarios, e.g., separating multiple musical instruments such as piano, drums, guitar and vocals from a single recording. To address such scenarios, the main line of research over the past 15 years has been to rely on models of the magnitude short-time Fourier spectra of the target class of audio signals [1]. These are known to carry rich structures, while phase spectra are less easily predictable, due to the nonlinear and nonlocal nature of phase relationships across time and frequency. In particular, recent breakthroughs in audio source separation have relied on magnitude models learned with deep neural networks (DNNs) [2]. Building on traditional statistical signal processing, the outputs of these models are interpreted as the variances of Gaussian source signals, and optimal linear filters (Wiener) are derived. Few methods have been proposed to estimate complex-valued spectra, based either on a non-informative uniform phase prior [3] or on heuristic phase models which are applicable to certain sounds (e.g., periodic) only [4].

The goal of this thesis is to step out of this historically motivated pipeline, towards phase-aware audio signal modeling and estimation, by leveraging the superior modeling power of DNNs. Two key scientific challenges are at hand:

1. How can the nonlinear, nonlocal and stochastic relationships between the phases and magnitudes of general audio signals across the time-frequency plane be efficiently captured by DNNs?

2. Given such joint models of phases and magnitudes, how can a complex target signal be efficiently recovered from linear (e.g., mixed) or nonlinear (e.g., clipped) noisy observations?

Challenge 1 will build on the methodological framework of complex-valued neural networks [5] and the recent development of DNN architectures enforcing nonlocal relationships such as WaveNet [6]. Rather than using off-the-shelf architectures borrowed from image processing or attempting to model the raw waveform in a brute-force fashion, as classically done today, we will seek to design new neuron
types and new layers that effectively incorporate prior knowledge about the structure of audio signals. The difficulty is to obtain a generative model that is both applicable to any type of sound (periodic, quasi-periodic, transient, noise-like, etc, unlike [5]) and computationally tractable (unlike WaveNet). Recent works on phase modelling for speech and audio [4] will guide the development of such a new architecture. Success will be measured by the tradeoff achieved between computational efficiency and the ability to accurately predict complex real-world audio signals.

Challenge 2 will be tackled using recent theoretical and methodological breakthroughs from the field of phase retrieval [7]. Recent work from our team showed that the problem of phase retrieval bears strong connections to the problem of signal separation from magnitude [3] or phase [8] models. To alleviate the difficulty of this generally NP-hard non-convex problem, spectral initialization [7], iterative methods [4] and state-of-the-art stochastic optimization techniques [9] will be jointly investigated. The success of this challenge will be measured using standard metrics for the considered audio tasks, e.g., source separation, speech enhancement, or audio inpainting.


**Principales activités**
The first year will be devoted to modeling the invariances of audio signals and
incorporating prior knowledge into DNN architectures for audio generation (Challenge 1).

The second year will be devoted to exploiting the methodological tools introduced during the first year for speech and music source separation, and possibly other tasks (Challenge 2). We aim to submit our results to at least one relevant evaluation campaign, namely CHiME for speech separation or MIREX for music separation.

**Compétences**

MSc in computer science, machine learning, or signal processing

Experience with Python programming language

**Avantages sociaux**

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- French courses

**Rémunération**

Gross Salary per month: 1982€ brut per month (year 1 & 2) and 2085€ brut/month (year 3)

**Informations générales**

- **Ville** : Villers-lès-Nancy
- **Centre Inria** : CRI Nancy - Grand Est
- **Date de prise de fonction souhaitée** : 01-10-2018
- **Durée de contrat** : 3 ans
- **Date limite pour postuler** : 01-05-2018

**Contacts**

- **Equipe Inria** : MULTISPEECH
- **Recruteur** :
  
  Vincent Emmanuel / emmanuel.vincent@inria.fr

L'essentiel pour réussir

**Application deadline**

May 1st, 2018 (Midnight Paris time)

**How to apply**

Upload your file on jobs.inria.fr in a single pdf or zip file, and send it as well by email to emmanuel.vincent@inria.fr. Your file should contain the following documents:

- Your CV.
- A cover/motivation letter describing your interest in this topic.
- A short (max one page) description of your Master thesis (or equivalent) or of the work in progress if not yet completed.
- Your degree certificates and transcripts for Bachelor and Master (or the last 5 years).
• Master thesis (or equivalent) if it is already completed and publications if any (it is not expected that you have any). Only the web links to these documents are preferable, if possible.

In addition, one recommendation letter from the person who supervises(d) your Master thesis (or research project or internship) should be sent directly by his/her author to emmanuel.vincent@inria.fr.

Applications are to be sent as soon as possible.

**Conditions pour postuler**

**Sécurité défense :**
Ce poste est susceptible d’être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L’autorisation d’accès à une zone est délivrée par le chef d’établissement, après avis ministériel favorable, tel que défini dans l’arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l’annulation du recrutement.

**Politique de recrutement :**
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

**Attention:** Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.