Master Thesis

Gaspard Ulysse Fragnière

August 2022

1 How to LaTeX

How to make a reference to a paper: [4]

2 Literature review

2.1 Introduction

[4] is a survey of several methods for sound source localization (SSL). Tradiontionally, this problem has been tackled using Signal Processing based methods. But in the recent years, methods based on deep learning have been developped and showed better results than traditional approaches. Those methods have been compiled in this paper. The survey is organized in the different following sections:

- **Section I**: Introduction
- **Section II**: Acoustic Environment and Sound Source Configuration
- **Section III**: Conventional SSL methods
- **Section IV**: Neural Network Architectures for SSL
- **Section V**: Input Features
- **Section VI**: Outputs strategies
- **Section VII**: Data
    1. Synthetic Data
    2. Real data
    3. Data augmentation techniques
- **Section VIII**: Learning Strategies
- **Section IX**: Conclusions and Perspectives

We are interested in the section about Synthetic Data and Data augmentation. Indeed those sections can be used as a statring point for building the state of the art.Its goal is to answer the following questions:

- Are there **existing methods** to generate virtually:
  - measured time data (single channel/multi-channel)?
  - measured source spectra (single channel/multi-channel)?
  - measured cross-spectral matrices in stationary environments (multi-channel only)?
- What **measurement scenarios** are used in the literature (time-stationary/non-stationary sources, number of microphones, temporal dimensions...)?
- What are the **existing setups** in multi-channel data generation with neural networks (conditioning variables, network architectures (convolutional, recurrent, Transformer,...), generative algorithms (GAN/VAE), ...)

In [4], a few methods about data generation are introduced. The first one is the following: Simulate the Room Impulse Response (RIR) in order to simulate realistically room acoustics (e.g. reverberation). This can provide suited training data, since RIR for rooms of different size, different source prosition as well as different dry signals can be used for the training.

In [1] and [2], an approach based on a GMR is used. TODO: read more.

In [6], a GAN is used to simulate data.

[5] proposed a low-complexity model-based training data generation method that includes a deterministic model for the direct path and a statistical model for late reverberation. It has been demonstrated that the SSL neural network, trained using the data generated by this method, achieves comparable localization performance as the same architecture trained on a dataset generated by the usual ISM.

An investigation of several simulation methods was done by [3], with extensions of ISM, namely, ISM with directional sources, and ISM with a diffuse field due to scattering. [3] compared the simulation algorithms via the training of an MLP (in both regression and classification modes) and showed that ISM with scattering effects and directional sources leads to the best SSL performance.

References

[1] Antoine Deleforge, Florence Forbes, and Radu Horaud. Variational em for binaural sound-source separation and localization. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 76–80. IEEE, 2013.

[2] Antoine Deleforge, Radu Horaud, Yoav Y Schechner, and Laurent Girin. Co-localization of audio sources in images using binaural features and locally-linear regression. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(4):718–731, 2015.

[3] Femke B Gelderblom, Yi Liu, Johannes Kvam, and Tor Andre Myrvoll. Synthetic data for dnn-based doa estimation of indoor speech. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4390–4394. IEEE, 2021.

[4] Pierre-Amaury Grumiaux, Srđan Kitić, Laurent Girin, and Alexandre Guérin. A survey of sound source localization with deep learning methods. *The Journal of the Acoustical Society of America*, 152(1):107–151, 2022.

[5] Fabian Hübner, Wolfgang Mack, and Emanuël AP Habets. Efficient training data generation for phase-based doa estimation. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 456–460. IEEE, 2021.

[6] Elizabeth Vargas, James R Hopgood, Keith Brown, and Kartic Subr. On improved training of cnn for acoustic source localisation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:720–732, 2021.