Internship subject: Convolutional neural networks for sound source localization

Direction/entity/department : OLS/HOME/VIBES/CVA
Employement domain : Master/end-of-studies internship

Department : Ille et Vilaine (35)
Location : Cesson-Sévigné, France
Site : **Orange Labs**

Address: 4 rue du Clos Courtel, 35510 Cesson-Sévigné, France

Preferred starting date : February 2020 onwards

Job description

Internship context

Nowadays it has become clear that voice would become the interface of choice for our interactions with machines. Boosted by recent advances in deep neural networks, voice interfaces in our smartphones are gradually replacing the textual inputs for various tasks: Internet browsing, making phone calls, address book search, SMS writing… Arrival of smart speakers on the home market, such as Amazon Echo or Google Home, to name only a few, shows that we are only at the beginning of the expansion of voice interfaces to other sectors, such as home automation. However, this rises significant challenges in terms of quality of far field sound capture: the more distant is the speaker from the device, the more pronounced are the effects of reverberation and household noise (TV, children, …), which degrades the performance of speech recognition engines.

Hence the interest for microphone arrays embedded in these devices -- 2 for Google Home, and up to 7 for Amazon Echo -- to obtain an optimal sound capture, and progress towards the human-machine interface Holy Grail: achieve a hands-free communication in difficult acoustic environments. Orange is also engaged in the smart speaker market by launching the first in the long series of its virtual assistants Djingo[1], in collaboration with Deutsche Telekom, at the end of the year.

Subject description

Speech enhancement by microphone arrays is done by beamforming, which emphasizes the signal coming from a target direction, and suppresses the others [1]. This raises the question of accurate localization of sound sources, as a mandatory prerequisite. Unfortunately, traditional methods suffer from considerable drop in performance in the presence of noise and/or reverberation [2]. Lately, deep neural networks, trained on simulated data, have shown surprising robustness to these adverse acoustic conditions [3] [4]. These networks usually incorporate recurrent layers (e.g. LSTM or GRU), in order to capture the temporal pattern in the audio data. However, such networks are somewhat more difficult to train, have larger number of parameters, and their parallelized implementation is hindered.

[1] https://djingo.orange.fr/

Recently, the architectural improvements in convolution neural networks (CNNs) made them an appealing alternative to recurrent versions [5]. Indeed, our research on deep learning-based source separation suggests that the performance can be preserved, or even surpassed, using the CNNs with dilated convolutional layers. The goal of the internship is to investigate if such networks could be successfully used for sound source localization task as well.

The intern will thus devise a source localization CNN, and compare its complexity and performance against the recurrent model [3]. To facilitate the task, he/she would adapt an internal deep learning framework and the training/validation datasets. The localization performance is to be evaluated on public [5] and our private datasets, using a benchmark library written in Python. Depending on the progress, the intern would also embed the model into a processing chain comprising the existing source counting and source separation modules.

Bibliography

[1] S. Gannot, E. Vincent, S. Markovich-Golan and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* , vol. 4, no. 25, pp. 692-730, 2017.

[2] C. Evers, H. Loellmann, H. Mellmann, A. Schmidt, H. Barfuss, P. Naylor and W. Kellermann, "The LOCATA Challenge: Acoustic Source Localization and Tracking," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* 2019.

[3] L. Perotin, R. Serizel, E. Vincent and A. Guérin, "CRNN-based joint azimuth and elevation localization with the Ambisonics intensity vector," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2018.

[4] S. Adavanne, P. Archontis, N. Joonas and T. Virtanen, "Sound event localization and detection of overlapping sources using convolutional recurrent neural networks," *IEEE Journal of Selected Topics in Signal Processing,* vol. 1, no. 13, pp. 34-48, 2019.

[5] H. Löllmann, C. Evers, A. Schmidt, H. Mellmann, H. Barfuss, P. Naylor and W. Kellermann, "The LOCATA challenge data corpus for acoustic source localization and tracking," in *IEEE 10th Sensor Array and Multichannel Signal Processing Workshop*, 2018.

Preferred profile

- **Education:**
  - You are the final year master-level student in signal processing and/or machine learning
- **Technical requirements**:
  - Understanding of audio and/or array signal processing.
  - Understanding of statistical learning methods, preferably neural networks
  - Exposure to Matlab and/or Python software. Ideally: hands-on experience in neural network frameworks such us Keras/TensorFlow/PyTorch.
- **Personal traits:**
  - Strong interest in signal processing and/or machine learning
  - Working knowledge of English language.

Additional information

- **Team**
  You will work with the team of researchers in the domain of speech and acoustics signal processing. You will have the opportunity to discuss with different team members, experts in the domains of multichannel signal processing, coding, and machine learning applied to audio.You will work directly with a PhD student whose thesis subject concerns sound source localization by deep learning methods.

- **Added values of this offer**
  You would contribute, along with the team, to the research in two domains which strongly influence the development of vocal assistents: machine learning, and microphone array signal processing. You would work on the technologies which will be used more and more in the coming future. Finally, you would have the opportunity to submit patents, and participate in the publication of scientific articles as well.