Multimodal Music Recording Remastering

Giorgia Cantisani, Slim Essid, Gael Richard

Multimodal Music Recording Remastering

Giorgia Cantisani[1], Slim Essid[1], Gaël Richard[1]

[1] LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France.

MIPFrontiers

TELECOM ParisTech

Objectives

My proposal for this Ph.D. is to develop methods for a **user-centered** remastering of **music performance** recordings for giving the user an **interactive multimedia experience**. The idea is to guide audio **source separation/enhancement** using the user's attention as a high-level control/feedback to select which, for him/her, is the desired source to enhance. In the case of music performances, the source to enhance is represented by a particular **instrument** in the ensemble, thus we have a **polyphonic music** source separation problem.

Audio Source Separation

Source separation refers to *extracting one or more target sources in a mixture while suppressing interfering sources and noise, excluding dereverberation and echo cancellation* [10].

When talking about music...

- *mixture* ⇒ **polyphonic music**
- *target sources* ⇒ **individual instruments**

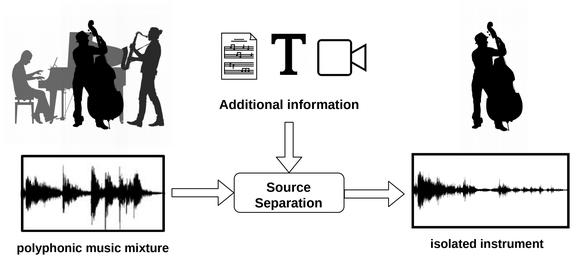


Figure 1: Informed audio source separation process

Informed source separation algorithms exploit all the available prior information about the sources (e.g. text, score, visual features) and the mixing process along with the audio signal to enhance the source separation process [4]. They perform better than the blind ones for music tasks.

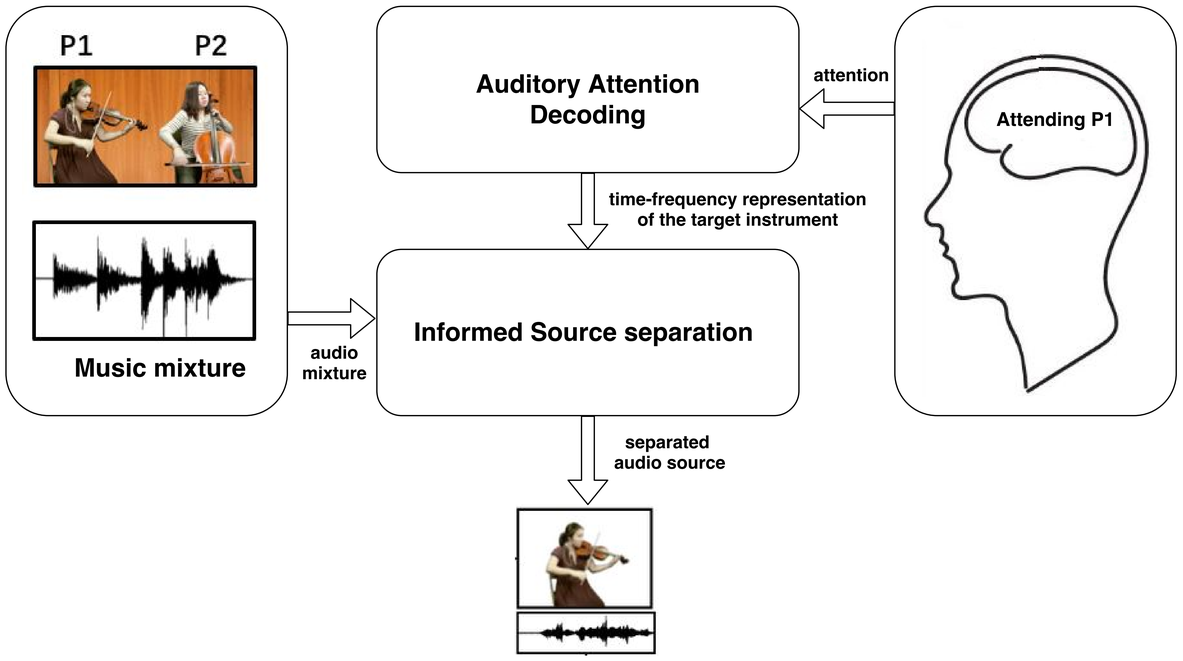
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowsa-Curie grant agreement No. 765068

DMRN+13: Digital Music Research Network One-day Workshop, December 18th 2018, London

Attention-Driven Source Separation

The user's attention can be used as a high-level **control/feedback** to select which, for him/her, is the target instrument to enhance. It would be useful for:

- General audience;
- Musicians;
- Sound Engineers;
- People with impaired hearing.



Attention to sounds

Attention *is a set of processes that allow the cognitive system to* ***select relevant information*** *in a given context* [8] and it can be characterized in many ways.

There are a few works on **auditory attention decoding** from the subject's **neural activity** [6], but they focus mostly on speech and they had access to the **isolated audio sources**.

Proposed Approach

Multimodal source separation *which exploits modalities that have never been considered before such as the* ***user's attention*** *to the source.*

The selective attention to the source will be explored first using only **audio stimuli**, and in a second phase using **audio-visual stimuli**.

The information extracted from the user while he/she is interacting with the music/video will be fully exploited to obtain:

- **source recognition**;
- **source estimation**;
- **source separation/enhancement**.

State of the Art

Speech stimuli

[2], [9], [5] tried to separate each sound source and use them to identify and enhance the attended speaker using the neural activity of the subject.

Audiovisual stimuli

In **noisy/multispeaker scenarios**, the visual modality, enhances speech reconstruction [1], [3].

Music stimuli

- **No previous work on music**;
- few works try to extract music information from the subject's neural activity;
- they all rely on extracting stimulus-related brain responses by averaging a high number of stimulus repetitions [7].

References

[1] M. J. Crosse, G. M. Di Liberto, and E. C. Lalor.
Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration.
Journal of Neuroscience, 36(38):9888–9895, 2016.

[2] N. Das, S. Van Eyndhoven, T. Francart, and A. Bertrand.
Eeg-based attention-driven speech enhancement for noisy speech mixtures using n-fold multi-channel wiener filters.
In *EUSIPCO*, pages 1660–1664. IEEE, 2017.

[3] E. Z. Golumbic, G. B. Cogan, C. E. Schroeder, and D. Poeppel.
Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party".
Journal of Neuroscience, 33(4):1417–1426, 2013.

[4] A. Liutkus, J.-L. Durrieu, L. Daudet, and G. Richard.
An overview of informed audio source separation.
In *WIAMIS*, pages 1–4. IEEE, 2013.

[5] J. O'Sullivan, Z. Chen, S. A. Sheth, G. McKhann, A. D. Mehta, and N. Mesgarani.
Neural decoding of attentional selection in multi-speaker environments without access to separated sources.
In *EMBC*, pages 1644–1647. IEEE, 2017.

[6] J. A. O'sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor.
Attentional selection in a cocktail party environment can be decoded from single-trial eeg.
Cerebral Cortex, 25(7):1697–1706, 2014.

[7] I. Sturm.
Analyzing the perception of natural music with eeg and ecog.
Ph.D. thesis, 2016.

[8] M. Turatto.
notes and lessons, 2006.

[9] S. Van Eyndhoven, T. Francart, and A. Bertrand.
Eeg-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses.
IEEE Trans. Biomed. Engineering, 64(5):1045–1056, 2017.

[10] E. Vincent, T. Virtanen, and S. Gannot.
Audio source separation and speech enhancement.
John Wiley & Sons, 2018.