AES Milan 2018 - Audio Re-purposing using Sound Separation

Sound Source Separation: An Introduction

Dr. Estefanía Cano

Semantic Music Technologies

Fraunhofer Institute for Digital Media Technology IDMT

cano@idmt.fraunhofer.de

Fraunhofer IDMT

Outline

1. Definition of Sound Source Separation (SSS)
2. Methods for SSS
   - Models of the source position
   - Models of the source
   - Models for interference reduction
3. Applications
   - Examples
4. Evaluation of SSS algorithms

1. What is Sound Source Separation?

- A Definition -

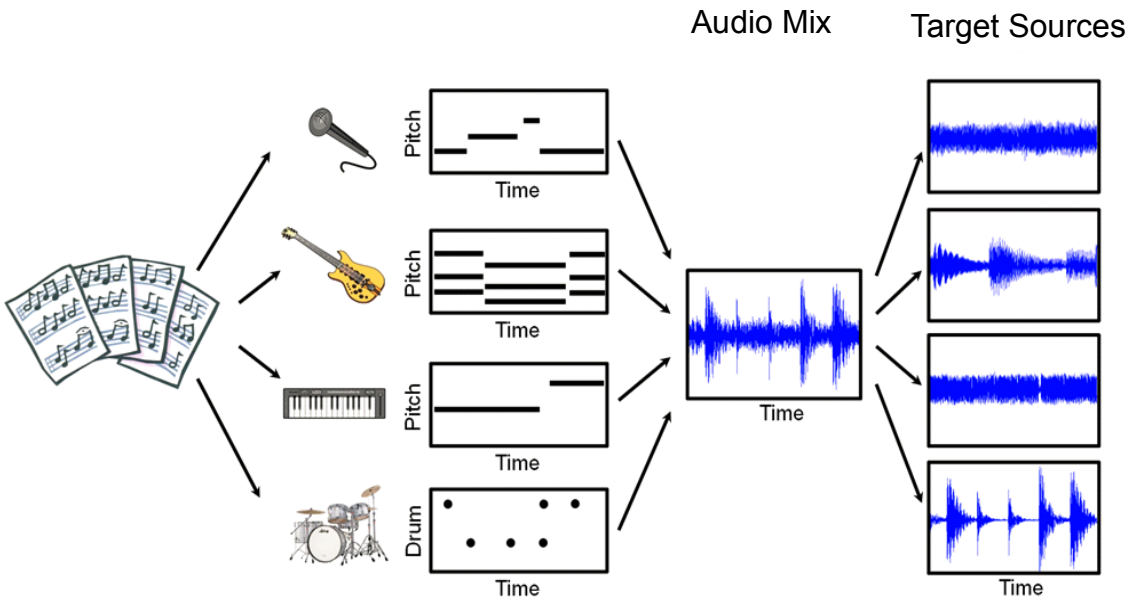


Figura: Separación de Señales de Audio [1]

1. What is Sound Source Separation?

- A Definition -

The definition of the target source is often lose:

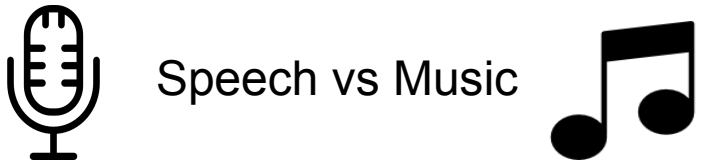
- Harmonic/Percussive Separation
- Solo/Accompaniment Separation
- Singing Voice Separation

→ Extremely relevant both for model design and evaluation.

1. What is Sound Source Separation?

- A Definition -

The creation process of the audio mixture varies greatly depending on the application:

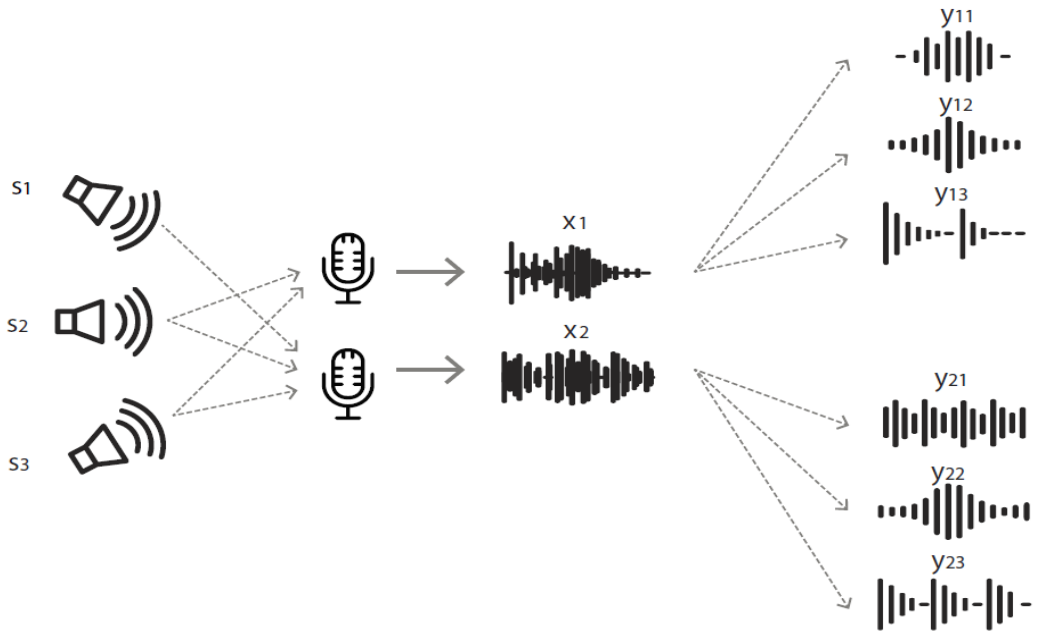


1. What is Sound Source Separation?

- A Definition -

The creation process of the audio mixture varies greatly depending on the application:

Speech

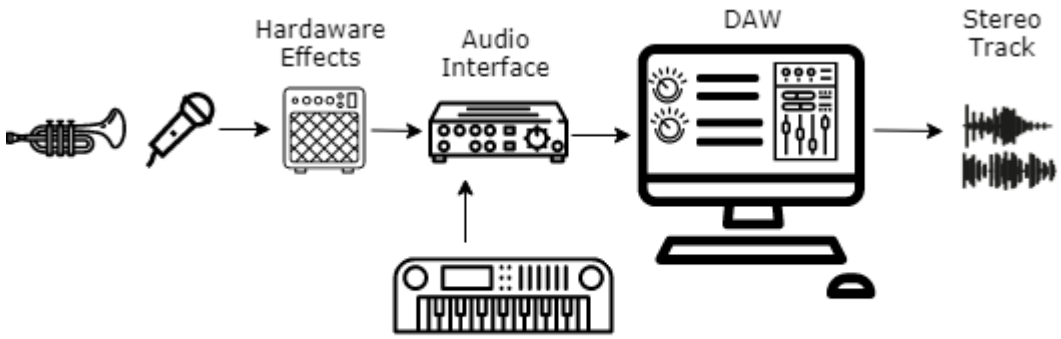


1. What is Sound Source Separation?

- A Definition -

The creation process of the audio mixture varies greatly depending on the application:

Music



→ Understanding the creation process of the mixture is critical for the development of separation algorithms.

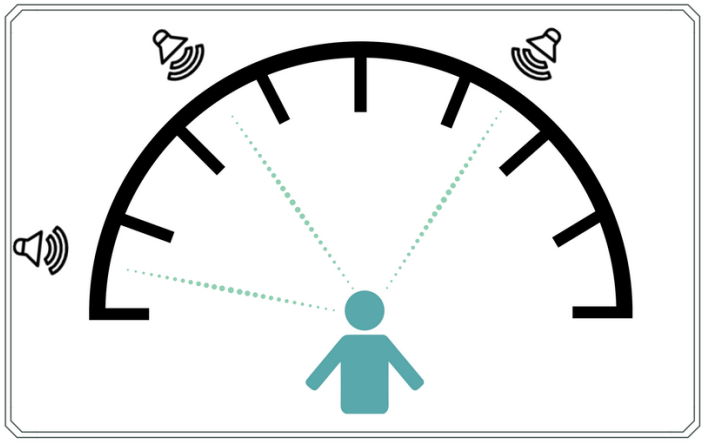
2. Methods for Sound Source Separation
- Categorization -

- Models of the source position
- Models of the source
- Models for interference reduction

2. Methods for Sound Source Separation - Models of the source position -

Assumption:
The target source is predominant in a given panning direction.

Implications:
- Prior knowledge of the source location is necessary.
- The model cannot handle overlapping sources in the stereo panning
- Minimum stereo recordings are needed.

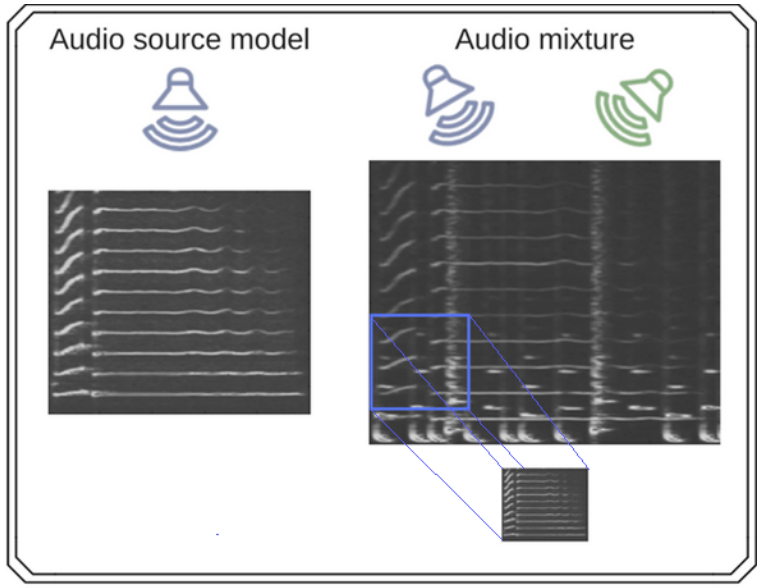


Models of the source position

2. Methods for Sound Source Separation - Models of the sound source -

Assumption:
The spectral characteristics of the target source - in time and frequency – are well known and can be modeled.

Implications:
- Source characteristics need to be well-understood
- An independent model for each source is needed
- Non-harmonic sources can be more difficult to model.

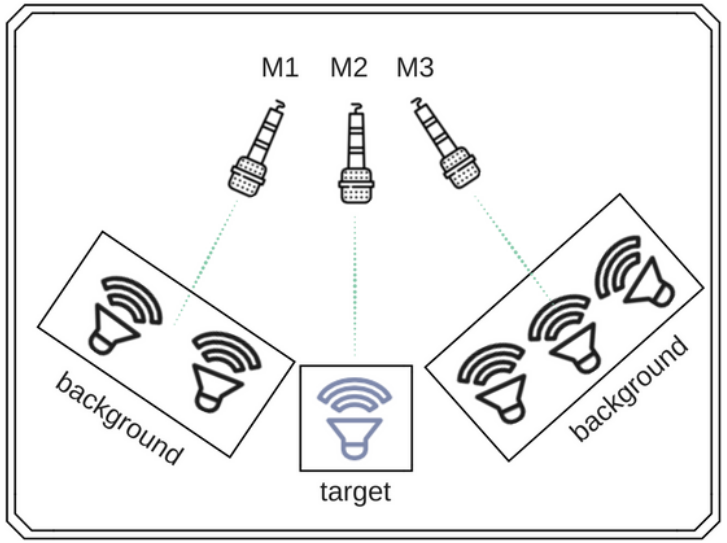


Models of the sound source

2. Methods for Sound Source Separation - Models for interference reduction -

Assumption:
There is at least one microphone that predominantly captures the target source and at least one microphone that predominantly captures the interference.

Implications:
- Multi-microphone recordings are needed
- Performance usually increases with increased number of microphones
- Depending on the number of microphones, processing times can be long.



Models for interference reduction

3. Applications of Sound Source Separation
- Examples -

Pre-Processing	Final Goal
Music Transcription [2]	Re-mixing [5]
Audio Classification	Upmixing [6]
• Musical instruments [3]	Gaming [7]
• Engine sounds [4]	Leakage removal

Applications

accusonus

Audionamix
SEPARATE2CREATE®

technicolor

4. Evaluation of Sound Source Separation

- An open challenge -

BSS: SDR, SAR, SIR, ISR [8]

↓

PEASS: OPS, IPS, APS, TPS [9]

↓

Reserachers reporting inconstencies with the metrics [10]

↓

Roll back to BSS → SiSEC Campaign

What now?

Other alternatives? [11-13]

What about listening tests?

Take home messages

- A clear definition of the target source is critical both for model design and evaluation.
- Understanding the creation process of the mixture will result in better separation algorithms.
- Do no neglect the importance of a clear evaluation procedure!

References

[1] Müller M., Fundamentals of Music Processing, Figure 8.1, Springer 2015

[2] Dittmar C., Gärtner D., "Real-time transcription and separation of drum recordings based on NMF decomposition", Proc Int. Conf on Digital Audio Effects (DAFx), 2014.

[3] Bosch, J., Janer, J., Fuhrmann, F., & Herrera, P. ,"A Comparison of Sound Segregation Techniques for Predominant Instrument Recognition in Musical Audio Signals", Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR), 2012.

[4] Cano, E. & Nowak, J. & Grollmisch, S.."Exploring sound source separation for acoustic condition monitoring in industrial scenarios". (EUSIPCO), 2017.

[5] Matz, Daniel et al. "New Sonorities for Early Jazz Recordings Using Sound Source Separation and Automatic Mixing Tools." (ISMIR), 2015.

[6] D. FitzGerald, "Upmixing from mono - A source separation approach," 2011 17th International Conference on Digital Signal Processing (DSP), 2011.

[7] Cano E., Dittmar C., Abeßer J., Kehling C., Grollmisch S. (2018) Music Technology and Education. In: Bader R. (eds) Springer Handbook of Systematic Musicology. Springer Handbooks. Springer, Berlin, Heidelberg.

[8] Vincent E., Gribonval R., Févotte C.. Performance measurement in blind audio source separation. IEEE Transactions on Audio, Speech and Language Processing, Institute of Electrical and Electronics Engineers, 2006, 14 (4), pp.1462–1469

References

[9] Emiya V., Vincent E., Harlander N., and Hohmann V., Subjective and objective quality assessment of audio source separation, IEEE Transactions on Audio, Speech and Language Processing, 2011, 19 (7), pp. 2046-2057.

[10] Cano E., FitzGerald D. and Brandenburg K., "Evaluation of quality of sound source separation algorithms: Human perception vs quantitative metrics," 2016 24th European Signal Processing Conference (EUSIPCO), Budapest, 2016, pp. 1758-1762.

[11] Cartwright M., Pardo B., Mysore G.J.and Hoffman M., "Fast and easy crowdsourced perceptual audio evaluation," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016, pp. 619-623.

[12] Cano E., Liebetrau J.,Fitzgerald D., Brandenburg K., “The dimensions of perceptual quality of sound separation”, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, 2018.

[13] Ward D., Wierstorf H., Mason R.D, Grais E.M. and Plumbley M.D, “BSS EVAL or PEASS? Predicting the Perception of Singing Voice Separation”, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, 2018.

Thank you!

cano@idmt.fraunhofer.de