Complex-valued and hybrid models for audio processing

MULTISPEECH team seminar - INRIA Nancy Grand-Est

Paul Magron

IRIT, Université de Toulouse, CNRS - Signal & Communications team

Academic background

- PhD – *Télécom ParisTech* (R. Badeau, B. David) 2013 - 2016
  Phase recovery based on signal modeling: application to audio source separation.
- Postdoc – *Tampere University* (T. Virtanen) 2017 - 2019
  Probabilistic phase modelling and real-time sound source separation.
- Postdoc – *IRIT* (C. Févotte) 2019 -
  Content-aware music recommendation.

Publications since 2015:

- 4 journal articles.
  - 2 *IEEE Transactions on Audio, Speech, and Language Processing*
  - 1 *IEEE Signal Processing Letters*
  - 1 *IEEE Journal of Selected Topics in Signal Processing*
- 19 conference articles.
  - 1 best paper award (iWAENC 2018)
- 13 co-authors from 7 institutions.

Research field

Audio signal processing

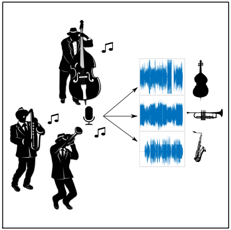
- Telecommunications: speech coding and synthesis.
- Medical devices: hearing aids.
- Educational softwares: language/music learning.
- Music information retrieval, music streaming.

Research field

Audio signal processing

- Telecommunications: speech coding and synthesis.
- Medical devices: hearing aids.
- Educational softwares: language/music learning.
- Music information retrieval, music streaming.

Main applications (of my work)



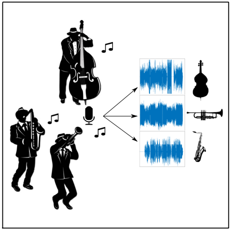
Source separation

Research field

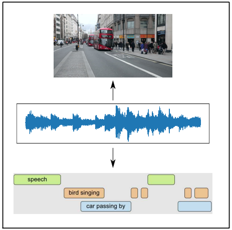
Audio signal processing

- Telecommunications: speech coding and synthesis.
- Medical devices: hearing aids.
- Educational softwares: language/music learning.
- Music information retrieval, music streaming.

Main applications (of my work)



Source separation



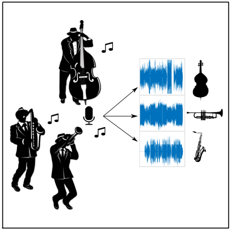
Acoustic scene analysis

Research field

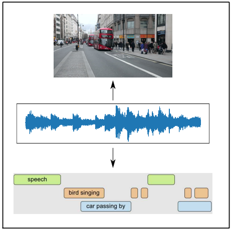
Audio signal processing

- Telecommunications: speech coding and synthesis.
- Medical devices: hearing aids.
- Educational softwares: language/music learning.
- Music information retrieval, music streaming.

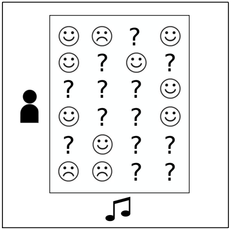
Main applications (of my work)



Source separation



Acoustic scene analysis



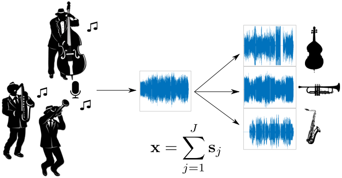
Music recommendation

Audio source separation

Audio source separation

Audio content is usually composed of several constitutive sounds:

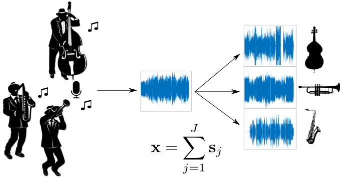
- ▷ One or several speakers + noise.
- ▷ Environmental / domestic sounds.
- ▷ Musical instruments.



Audio source separation

Audio content is usually composed of several constitutive sounds:

- ▷ One or several speakers + noise.
- ▷ Environmental / domestic sounds.
- ▷ Musical instruments.



$$\mathbf{x} = \sum_{j=1}^{J} \mathbf{s}_j$$

Source separation = recovering the sources from the mixture.

- ▷ Automatic speech recognition → clean speech vs. noise.
- ▷ Rhythm analysis → drums vs. harmonic instruments.
- ▷ Stationary / transient decomposition → time-stretching.

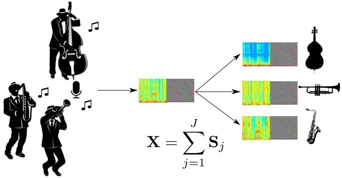
Audio source separation

Audio content is usually composed of several constitutive sounds:

▷ One or several speakers + noise.

▷ Environmental / domestic sounds.

▷ Musical instruments.



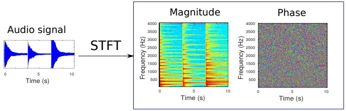
Source separation = recovering the sources from the mixture.

▷ Automatic speech recognition → clean speech vs. noise.

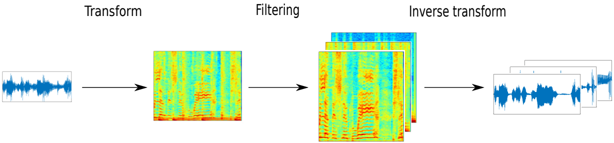
▷ Rhythm analysis → drums vs. harmonic instruments.

▷ Stationary / transient decomposition → time-stretching.

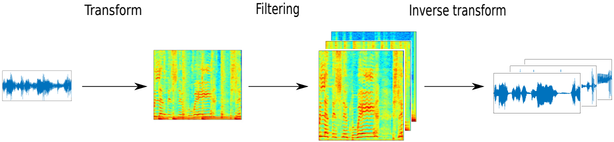
Time-frequency separation = acts on the short-time Fourier transform (STFT).



Source separation general framework

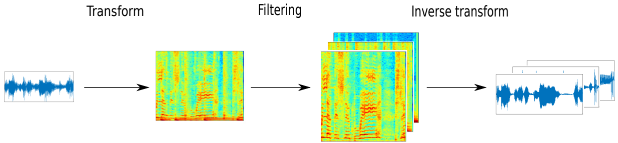


Source separation general framework

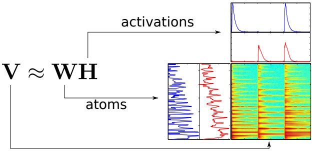


1. Nonnegative representation: $V = |STFT(x)|^2$.

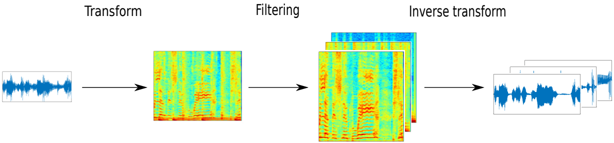
Source separation general framework



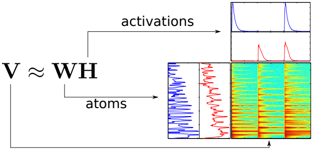
1. Nonnegative representation: $V = |STFT(x)|^2$.
2. Structured model:
   - ▷ Nonnegative matrix factorization (NMF).
   - ▷ Deep neural network (DNN).



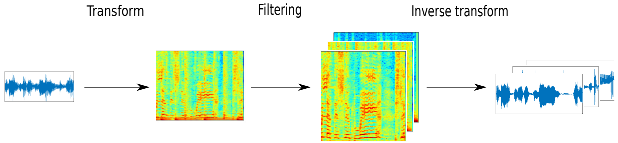
Source separation general framework



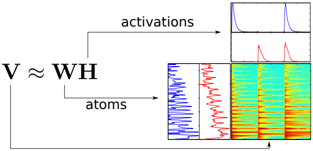
1. Nonnegative representation: $V = |STFT(x)|^2$.
2. Structured model:
   - ▷ Nonnegative matrix factorization (NMF).
   - ▷ Deep neural network (DNN).
3. Estimation:
   - ▷ $\min_{W,H} \mathcal{D}(V, WH)$
   - ▷ $\min_\theta \mathcal{L}(\{V_j\}, \phi_\theta(V))$



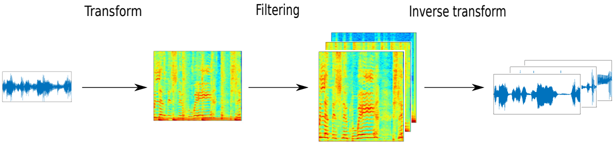
Source separation general framework



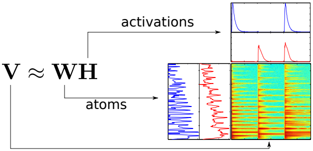
1. Nonnegative representation: $\mathsf{V} = |\mathsf{STFT}(\mathsf{x})|^2$.
2. Structured model:
   - ▷ Nonnegative matrix factorization (NMF).
   - ▷ Deep neural network (DNN).
3. Estimation:
   - ▷ $\min_{\mathsf{W},\mathsf{H}} \mathcal{D}(\mathsf{V}, \mathsf{WH})$
   - ▷ $\min_{\theta} \mathcal{L}(\{\mathsf{V}_j\}, \phi_\theta(\mathsf{V}))$
4. Set of nonnegative masks M_j.



Source separation general framework



1. Nonnegative representation: $V = |\mathrm{STFT}(x)|^2$.
2. Structured model:
   - ▷ Nonnegative matrix factorization (NMF).
   - ▷ Deep neural network (DNN).
3. Estimation:
   - ▷ $\min_{W,H} \mathcal{D}(V, WH)$
   - ▷ $\min_\theta \mathcal{L}(\{V_j\}, \phi_\theta(V))$
4. Set of nonnegative masks M_j.
5. Synthesis: $\tilde{s}_j = \mathrm{STFT}^{-1}(M_j \odot X)$.



The phase problem

Wiener filter

$$\underbrace{\boxed{\hat{\mathsf{S}}_j}}_{\text{Source STFT (Estimated)}} = \underbrace{\boxed{\frac{\hat{\mathsf{V}}_j}{\sum_{k=1}^{J} \hat{\mathsf{V}}_k}}}_{\text{Mask}} \odot \underbrace{\boxed{\mathsf{X}}}_{\text{Mixture STFT}}$$

$\angle \mathbf{S}_j = \angle \mathbf{X} \rightarrow$ Issues in sound quality when sources overlap in the TF domain.

The phase problem

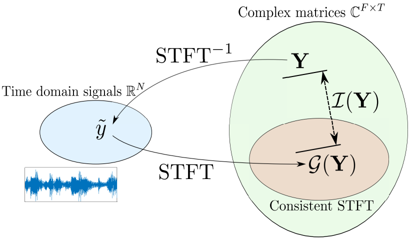
Wiener filter

$$\underbrace{\boxed{\hat{\mathsf{S}}_j}}_{\text{Source STFT (Estimated)}} = \underbrace{\boxed{\frac{\hat{\mathsf{V}}_j}{\sum_{k=1}^{J} \hat{\mathsf{V}}_k}}}_{\text{Mask}} \odot \underbrace{\boxed{\mathsf{X}}}_{\text{Mixture STFT}}$$

$\angle \mathbf{S}_j = \angle \mathbf{X} \rightarrow$ Issues in sound quality when sources overlap in the TF domain.

Inconsistency

$$\mathcal{I}(\mathsf{Y}) = ||\mathsf{Y} - \mathcal{G}(\mathsf{Y})||^2$$
$$\mathcal{G} = \mathsf{STFT} \circ \mathsf{STFT}^{-1}$$



The phase problem

Wiener filter

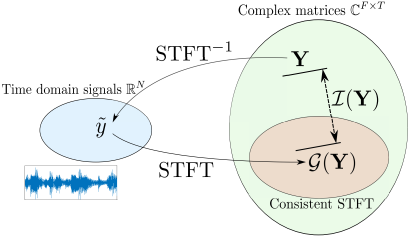
$$\underbrace{\boxed{\hat{\mathsf{S}}_j}}_{\text{Source STFT (Estimated)}} = \underbrace{\boxed{\frac{\hat{\mathsf{V}}_j}{\sum_{k=1}^{J} \hat{\mathsf{V}}_k}}}_{\text{Mask}} \odot \underbrace{\boxed{\mathsf{X}}}_{\text{Mixture STFT}}$$

$\angle \mathbf{S}_j = \angle \mathbf{X} \rightarrow$ Issues in sound quality when sources overlap in the TF domain.

Inconsistency

$$\mathcal{I}(\mathsf{Y}) = ||\mathsf{Y} - \mathcal{G}(\mathsf{Y})||^2$$
$$\mathcal{G} = \mathsf{STFT} \circ \mathsf{STFT}^{-1}$$

- ▷ Griffin-Lim (GL): minimization of $\mathcal{I}$.
- ▷ Extension to source separation.
- ▷ Combination with Wiener filtering.



The phase problem

Comparative study [ICASSP 15]

- Room for improvement for phase recovery.
- Test alternative separation methods accounting for the phase:
  - Consistency-based.
  - Signal model-based.
- The most promising uses a signal model for structuring the phase / STFT.

P. Magron, R. Badeau, B. David, "Phase recovery in NMF for audio source separation: an insightful benchmark", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, April 2015.

The phase problem

Comparative study [ICASSP 15]

- Room for improvement for phase recovery.
- Test alternative separation methods accounting for the phase:
  - Consistency-based.
  - Signal model-based.
- The most promising uses a signal model for structuring the phase / STFT.

My approach

- Leveraging model-based phase properties in source separation.
- A phase-aware probabilistic framework.
- Joint estimation of magnitude and phase.

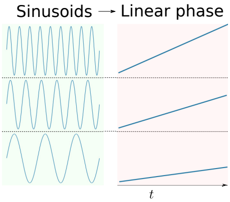
P. Magron, R. Badeau, B. David, "Phase recovery in NMF for audio source separation: an insightful benchmark", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, April 2015.

Model-based phase recovery

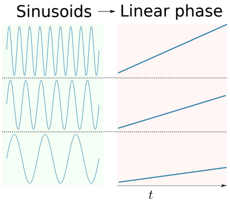
Sinusoidal phase model

For a mixture of sinusoids, the phase is:

$$\mu_{f,t} = \mu_{f,t-1} + 2\pi \underbrace{\nu_{f,t}}_{\text{normalized frequency}}$$



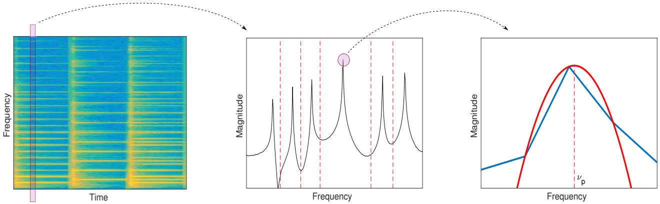
Sinusoidal phase model



For a mixture of sinusoids, the phase is:

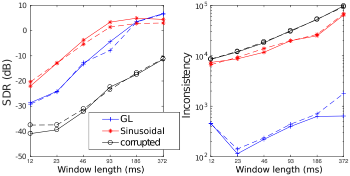
$$\mu_{f,t} = \mu_{f,t-1} + 2\pi \underbrace{\nu_{f,t}}_{\text{normalized frequency}}$$

Recursive estimation in each time frame:



Sinusoidal phase model

Restoration of piano pieces:

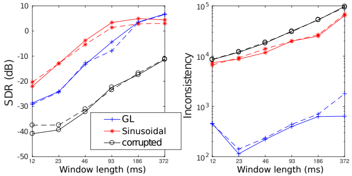


- ▷ Better performance than the consistency-based GL algorithm.
- ▷ But overall low SDR: error propagates over time frames.

P. Magron, R. Badeau, B. David, "Phase reconstruction of spectrograms with linear unwrapping: application to audio signal restoration", *Proc. European Signal Processing Conference (EUSIPCO)*, August 2015.

Sinusoidal phase model

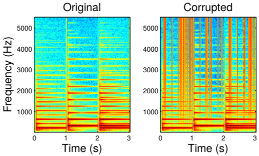
Restoration of piano pieces:



- ▷ Better performance than the consistency-based GL algorithm.
- ▷ But overall low SDR: error propagates over time frames.

Applications

- ▷ Click removal [EUSIPCO 15].
- ▷ Source separation.



P. Magron, R. Badeau, B. David, "Phase reconstruction of spectrograms with linear unwrapping: application to audio signal restoration", *Proc. European Signal Processing Conference (EUSIPCO)*, August 2015.

Repeating attacks

Why treating attacks?

- ▷ Perceptive quality of the sound.
- ▷ Initialize the sinusoidal model.

P. Magron, R. Badeau, B. David, "Phase reconstruction of spectrograms based on a model of repeated audio events", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2015.

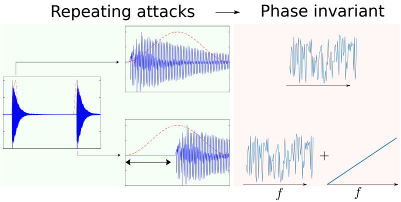
Repeating attacks

Why treating attacks?

- ▷ Perceptive quality of the sound.
- ▷ Initialize the sinusoidal model.

Model within onset frames:

$$\mu_{f,t} = \underbrace{\psi_f}_{\text{invariant}} + \underbrace{\eta_t}_{\text{offset}} f$$



P. Magron, R. Badeau, B. David, "Phase reconstruction of spectrograms based on a model of repeated audio events", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2015.

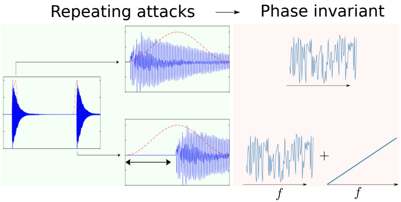
Repeating attacks

Why treating attacks?

- ▷ Perceptive quality of the sound.
- ▷ Initialize the sinusoidal model.

Model within onset frames:

$$\mu_{f,t} = \underbrace{\psi_f}_{\text{invariant}} + \underbrace{\eta_t}_{\text{offset}} f$$



Incorporation in a mixture model

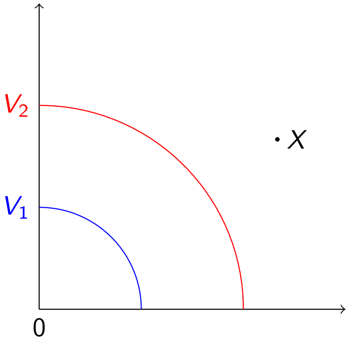
- ▷ Estimation with coordinate descente.
- ▷ Slight improvement over using the mixture's phase [WASPAA 15].

P. Magron, R. Badeau, B. David, "Phase reconstruction of spectrograms based on a model of repeated audio events", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2015.

Source separation iterative algorithm

Problem

$$\min_{\boldsymbol{\mu}} ||\mathsf{X} - \sum_{j=1}^{J} \mathsf{V}_j e^{\mathrm{i}\boldsymbol{\mu}_j}||^2$$



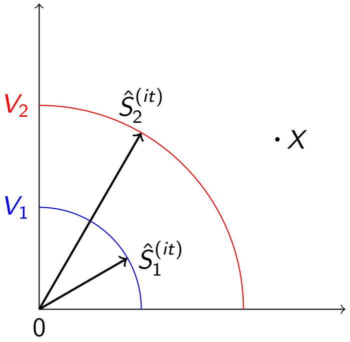
Source separation iterative algorithm

Problem

$$\min_{\boldsymbol{\mu}} ||\mathsf{X} - \sum_{j=1}^{J} \mathsf{V}_j e^{\mathrm{i}\boldsymbol{\mu}_j}||^2$$

Optimization

- ▷ Use a phase model for initialization.



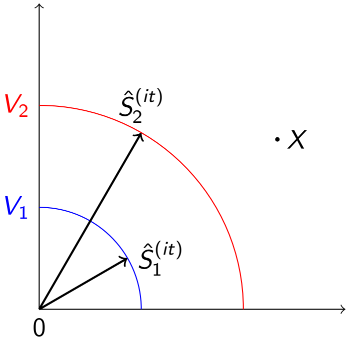
Source separation iterative algorithm

Problem

$$\min_{\boldsymbol{\mu}} ||\mathsf{X} - \sum_{j=1}^{J} \mathsf{V}_j e^{\mathrm{i}\boldsymbol{\mu}_j}||^2$$

Optimization

- Use a phase model for initialization.
- Auxiliary function method: iterative procedure.



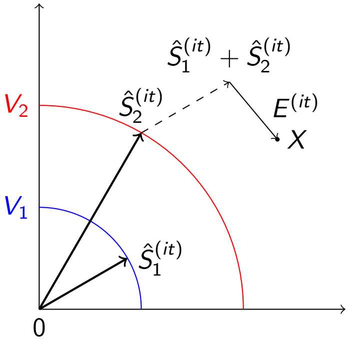
Source separation iterative algorithm

Problem

$$\min_{\boldsymbol{\mu}} ||\mathrm{X} - \sum_{j=1}^{J} \mathrm{V}_j e^{\mathrm{i}\boldsymbol{\mu}_j}||^2$$

Optimization

- Use a phase model for initialization.
- Auxiliary function method: iterative procedure.



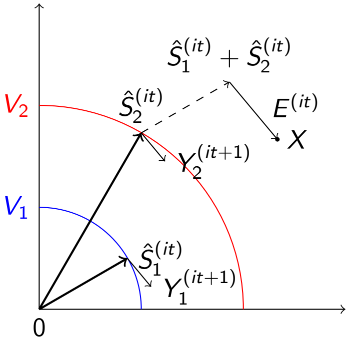
Source separation iterative algorithm

Problem

$$\min_{\boldsymbol{\mu}} ||\mathrm{X} - \sum_{j=1}^{J} \mathrm{V}_j e^{\mathrm{i}\boldsymbol{\mu}_j}||^2$$

Optimization

- ▷ Use a phase model for initialization.
- ▷ Auxiliary function method: iterative procedure.



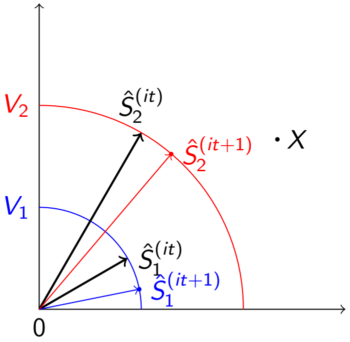
Source separation iterative algorithm

Problem

$$\min_{\boldsymbol{\mu}} ||\mathsf{X} - \sum_{j=1}^{J} \mathsf{V}_j e^{\mathrm{i}\boldsymbol{\mu}_j}||^2$$

Optimization

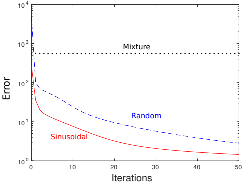
- Use a phase model for initialization.
- Auxiliary function method: iterative procedure.



Source separation iterative algorithm

Initialization impact (DSD100 dataset):

	SDR	SIR	SAR
Mixture	7.5	13.7	8.9
Random	9.5	22.8	9.7
Sinusoidal	13.6	31.0	13.7

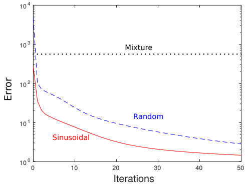


P. Magron, R. Badeau, B. David, "Model-based STFT phase recovery for audio source separation", *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26, no. 6, pp. 1095–1105, June 2018.

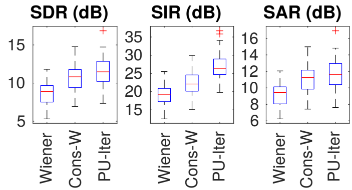
Source separation iterative algorithm

Initialization impact (DSD100 dataset):

	SDR	SIR	SAR
Mixture	7.5	13.7	8.9
Random	9.5	22.8	9.7
Sinusoidal	13.6	31.0	13.7



Comparison with Wiener filters:



P. Magron, R. Badeau, B. David, "Model-based STFT phase recovery for audio source separation", *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26, no. 6, pp. 1095–1105, June 2018.

How to re-introduce consistency?

Time-domain formulation

$$\min_{\{\tilde{s}_j\}} \sum_j |||\text{STFT}(\tilde{s}_j)| - V_j||^2 \text{ s. t. } \sum_j \tilde{s}_j = \tilde{x}$$

P. Magron, T. Virtanen, "Online spectrogram inversion for audio source separation", *IEEE Signal Processing Letters*, vol. 27, pp. 306–310, January 2020.

How to re-introduce consistency?

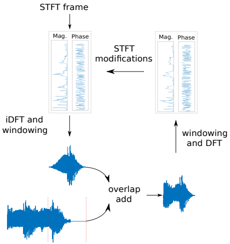
Time-domain formulation

$$\min_{\{\tilde{s}_j\}} \sum_j |||\mathrm{STFT}(\tilde{s}_j)| - V_j||^2 \text{ s. t. } \sum_j \tilde{s}_j = \tilde{x}$$

- Auxiliary function method: the (well-known) MISI algorithm, but convergence-guaranteed.

Online implementation

- Useful for real-time applications (hearing aids): for a latency of 16 ms, same results as the offline counterpart.
- Allows initialization with the sinusoidal phase: better results in some cases.



P. Magron, T. Virtanen, "Online spectrogram inversion for audio source separation", *IEEE Signal Processing Letters*, vol. 27, pp. 306–310, January 2020.

Alternative divergences

- Euclidean distance: not the most appropriate in audio.
- Popular alternative: the beta-divergences.

$\beta = 2$	$\beta = 1$	$\beta = 0$
Euclidean	Kullback-Leibler (KL)	Itakura-Saito (IS)
Emphasis on high-energy components	$\leftarrow$ In between $\rightarrow$	Scale invariance

P. Magron, P.-H. Vial, T. Oberlin, C. Févotte, "Phase recovery with Bregman divergences for audio source separation", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, June 2021.
P.-H. Vial, P. Magron, T. Oberlin, C. Févotte, "Phase retrieval with Bregman divergences and application to audio signal recovery", *IEEE Journal of Selected Topics in Signal Processing*, January 2021.

Alternative divergences

- Euclidean distance: not the most appropriate in audio.
- Popular alternative: the beta-divergences.

$\beta = 2$	$\beta = 1$	$\beta = 0$
Euclidean	Kullback-Leibler (KL)	Itakura-Saito (IS)
Emphasis on high-energy components	$\leftarrow$ In between $\rightarrow$	Scale invariance

Phase retrieval with beta-divergences

$$\min_{\{\tilde{s}_j\}} \sum_j D_\beta(|\mathrm{STFT}(\tilde{s}_j)|^d, V_j) \text{ s. t. } \sum_j \tilde{s}_j = \tilde{x}$$

- Optimization with accelerated gradient descent or ADMM.
- Experimentally: alternative divergences (e.g., KL or $\beta = 0.5$) $>$ Euclidean.

P. Magron, P.-H. Vial, T. Oberlin, C. Févotte, "Phase recovery with Bregman divergences for audio source separation", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, June 2021.

P.-H. Vial, P. Magron, T. Oberlin, C. Févotte, "Phase retrieval with Bregman divergences and application to audio signal recovery", *IEEE Journal of Selected Topics in Signal Processing*, January 2021.

Probabilistic phase modelling

Probabilistic framework

Why?

- Modeling uncertainty.
- Incorporating prior information.
- Obtaining estimators with nice statistical properties.
- Deriving inference schemes with convergence guarantees.

Probabilistic framework

Why?

- ▷ Modeling uncertainty.
- ▷ Incorporating prior information.
- ▷ Obtaining estimators with nice statistical properties.
- ▷ Deriving inference schemes with convergence guarantees.

Traditionally

Circularly-symmetric (or **isotropic**) sources $\iff$ Uniform phase

$\Rightarrow$ Phase-unaware estimators.

Probabilistic framework

Why?

- Modeling uncertainty.
- Incorporating prior information.
- Obtaining estimators with nice statistical properties.
- Deriving inference schemes with convergence guarantees.

Traditionally

Circularly-symmetric (or **isotropic**) sources $\iff$ Uniform phase

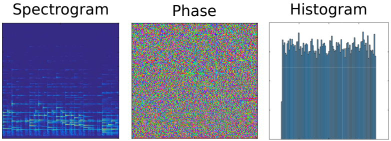
$\Rightarrow$ Phase-unaware estimators.

My approach

A phase-aware probabilistic framework for source separation.

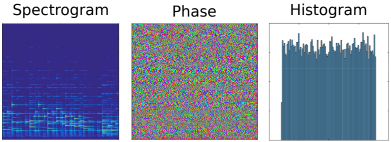
Is the phase really uniform?

A simple example (piano piece), where the phase appears uniformly-distributed.



Is the phase really uniform?

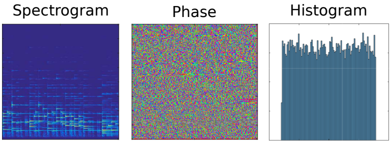
A simple example (piano piece), where the phase appears uniformly-distributed.



But: is that consistent with, e.g., the sinusoidal model?

Is the phase really uniform?

A simple example (piano piece), where the phase appears uniformly-distributed.



But: is that consistent with, e.g., the sinusoidal model?

Interpretation

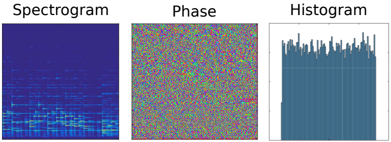
- ▷ The histogram validates an iid assumption on $\{\phi_{f,t}\}$:

$$\phi_{f,t} \sim \mathcal{D} \text{ and independent } \rightarrow \mathcal{D} = \mathcal{U}_{[0,2\pi[}$$

- ▷ This model only conveys a **global** information.

Is the phase really uniform?

A simple example (piano piece), where the phase appears uniformly-distributed.



But: is that consistent with, e.g., the sinusoidal model?

Interpretation

- The histogram validates an iid assumption on $\{\phi_{f,t}\}$:

$$\phi_{f,t} \sim \mathcal{D} \text{ and independent } \rightarrow \mathcal{D} = \mathcal{U}_{[0,2\pi[}$$

- This model only conveys a **global** information.

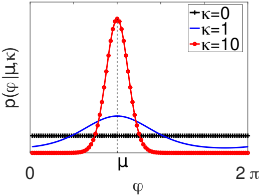
What about the **local structure** of the phase?

Von Mises phase model

Von Mises distribution

$$\phi_{ft} \sim \mathcal{VM}(\mu_{ft}, \kappa)$$

- $\mu_{f,t}$ = phase location.
- κ = concentration (quantifies non-uniformity).

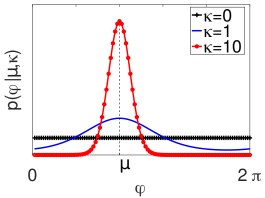


Von Mises phase model

Von Mises distribution

$$\phi_{ft} \sim \mathcal{VM}(\mu_{ft}, \kappa)$$

- ▷ $\mu_{f,t}$ = phase location.
- ▷ κ = concentration (quantifies non-uniformity).



Model

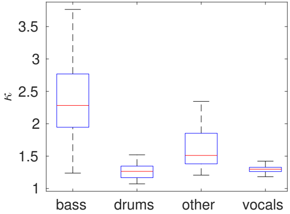
- ▷ $\mu_{f,t}$ = sinusoidal phase.
- ▷ Center the phases: $\psi_{f,t} = \phi_{f,t} - \mu_{f,t}$

Distribution	Uniform	VM	Centered VM
	$\phi_{f,t} \sim \mathcal{U}_{[0,2\pi[}$	$\phi_{ft} \sim \mathcal{VM}(\mu_{f,t}, \kappa)$	$\psi_{f,t} \sim \mathcal{VM}(0, \kappa)$
iid	✓	✗	✓
Local structure	✗	✓	✓

Von Mises phase model [iWAENC 18]

Estimation of κ (maximum likelihood):

$$\frac{I_1(\kappa)}{I_0(\kappa)} = \frac{1}{FT} \sum_{f,t} \cos(\psi_{ft})$$



- ▷ Solved with fast numerical schemes.
- ▷ κ quantifies the "sinusoidality" of the sources.

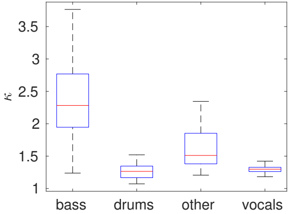
P. Magron, T. Virtanen, "On modeling the STFT phase of audio signals with the von Mises distribution", *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, September 2018.

Von Mises phase model [iWAENC 18]

Estimation of κ (maximum likelihood):

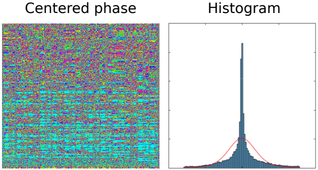
$$\frac{I_1(\kappa)}{I_0(\kappa)} = \frac{1}{FT} \sum_{f,t} \cos(\psi_{ft})$$

- ▷ Solved with fast numerical schemes.
- ▷ κ quantifies the "sinusoidality" of the sources.



Validation

- ▷ Both uniform and VM models are statistically relevant.
- ▷ They convey different information about the phase (global vs. local).



P. Magron, T. Virtanen, "On modeling the STFT phase of audio signals with the von Mises distribution", *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, September 2018.

Multiple sources model

In each TF bin, $x = \sum_{j=1}^{J} s_j$.

Isotropic gaussian model: $s_j \sim \mathcal{N}_{\mathbb{C}}(m_j, \Gamma_j)$ with $\Gamma_j = \begin{pmatrix} \gamma_j & c_j \\ \bar{c}_j & \gamma_j \end{pmatrix}$.

- m_j: mean (location) / γ_j: variance (energy).
- c_j: relation term, usually 0.

Multiple sources model

In each TF bin, $x = \sum_{j=1}^{J} s_j$.

Isotropic gaussian model: $s_j \sim \mathcal{N}_{\mathbb{C}}(m_j, \Gamma_j)$ with $\Gamma_j = \begin{pmatrix} \gamma_j & c_j \\ \bar{c}_j & \gamma_j \end{pmatrix}$.

- m_j: mean (location) / γ_j: variance (energy).
- c_j: relation term, usually 0.

Polar coordinate equivalent model: $s_j = r_j e^{\mathrm{i}\phi_j}$ where:

- $r_j \sim \mathcal{R}(v_j)$ (Rayleigh magnitude).
- $\phi_j \sim \mathcal{U}_{[0,2\pi[}$ (uniform phase).

Multiple sources model

In each TF bin, $x = \sum_{j=1}^{J} s_j$.

Isotropic gaussian model: $s_j \sim \mathcal{N}_{\mathbb{C}}(m_j, \Gamma_j)$ with $\Gamma_j = \begin{pmatrix} \gamma_j & c_j \\ \bar{c}_j & \gamma_j \end{pmatrix}$.

- m_j: mean (location) / γ_j: variance (energy).
- c_j: relation term, usually 0.

Polar coordinate equivalent model: $s_j = r_j e^{\mathrm{i}\phi_j}$ where:

- $r_j \sim \mathcal{R}(v_j)$ (Rayleigh magnitude).
- $\phi_j \sim \mathcal{U}_{[0,2\pi[}$ (uniform phase).

Naive idea

Consider a VM phase instead $\rightarrow$ Phase-aware

Multiple sources model

In each TF bin, $x = \sum_{j=1}^{J} s_j$.

Isotropic gaussian model: $s_j \sim \mathcal{N}_{\mathbb{C}}(m_j, \Gamma_j)$ with $\Gamma_j = \begin{pmatrix} \gamma_j & c_j \\ \bar{c}_j & \gamma_j \end{pmatrix}$.

- m_j: mean (location) / γ_j: variance (energy).
- c_j: relation term, usually 0.

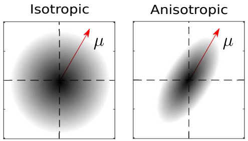
Polar coordinate equivalent model: $s_j = r_j e^{\mathrm{i}\phi_j}$ where:

- $r_j \sim \mathcal{R}(v_j)$ (Rayleigh magnitude).
- $\phi_j \sim \mathcal{U}_{[0,2\pi[}$ (uniform phase).

Naive idea

Consider a VM phase instead $\rightarrow$ Phase-aware... but not tractable.

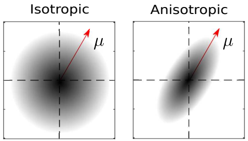
Anisotropic Gaussian model



	Phase-aware	Tractable
Isotropic Gaussian	✗	✓
Rayleigh + von Mises	✓	✗
Anisotropic Gaussian	✓	✓

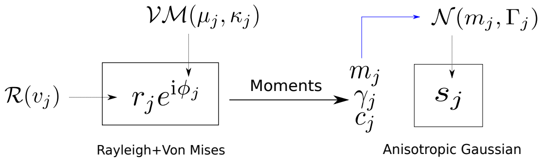
Source model: $s_j \sim \mathcal{N}_{\mathbb{C}}(m_j, \Gamma_j)$ with $\Gamma_j = \begin{pmatrix} \gamma_j & c_j \\ \bar{c}_j & \gamma_j \end{pmatrix}$.

Anisotropic Gaussian model

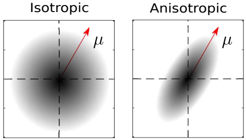


	Phase-aware	Tractable
Isotropic Gaussian	✗	✓
Rayleigh + von Mises	✓	✗
Anisotropic Gaussian	✓	✓

Source model: $s_j \sim \mathcal{N}_{\mathbb{C}}(m_j, \Gamma_j)$ with $\Gamma_j = \begin{pmatrix} \gamma_j & c_j \\ \bar{c}_j & \gamma_j \end{pmatrix}$.

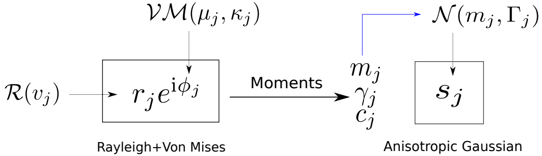


Anisotropic Gaussian model



	Phase-aware	Tractable
Isotropic Gaussian	✗	✓
Rayleigh + von Mises	✓	✗
Anisotropic Gaussian	✓	✓

Source model: $s_j \sim \mathcal{N}_{\mathbb{C}}(m_j, \Gamma_j)$ with $\Gamma_j = \begin{pmatrix} \gamma_j & c_j \\ \bar{c}_j & \gamma_j \end{pmatrix}$.



- ▷ Parameters: v_j (energy), μ_j (phase location), κ (non-uniformity).

Anisotropic Wiener filter

- ▷ Posterior mean of the sources: anisotropic Wiener filter [ICASSP 17].
- ▷ Performance (oracle separation results):

	SDR	SIR	SAR
Wiener	8.5	19.1	9.1
Anisotropic Wiener	**9.7**	**21.9**	**10.1**

Main message

- ▷ Including phase information in the filter improves the separation quality.
- ▷ Potential of a phase-aware statistical framework.

P. Magron, R. Badeau, B. David, "Phase-dependent anisotropic Gaussian model for audio source separation", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, March 2017.

Again... consistency?

Problem

- ▷ The (anisotropic) Wiener filter produces inconsistent matrices.

P. Magron, J. Le Roux, T. Virtanen, "Consistent anisotropic Wiener filtering for audio source separation", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2017.

Again... consistency?

Problem

- The (anisotropic) Wiener filter produces inconsistent matrices.

Consistent anisotropic Wiener [WASPAA 17]

- Consider the loss function:

  {posterior distribution of the (anistropic) sources}+{consistency constraint}

- Minimization with preconditioned conjugate gradient descent.

P. Magron, J. Le Roux, T. Virtanen, "Consistent anisotropic Wiener filtering for audio source separation", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2017.

Again... consistency?

Problem

▷ The (anisotropic) Wiener filter produces inconsistent matrices.

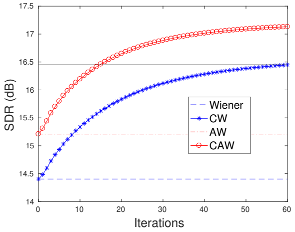
Consistent anisotropic Wiener [WASPAA 17]

▷ Consider the loss function:

{posterior distribution of the (anistropic) sources}+{consistency constraint}

▷ Minimization with preconditioned conjugate gradient descent.

	Sinusoidal model	Consistent
Wiener	✗	✗
CW	✗	✓
AW	✓	✗
CAW	✓	✓



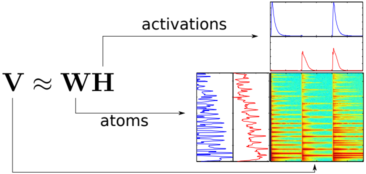
P. Magron, J. Le Roux, T. Virtanen, "Consistent anisotropic Wiener filtering for audio source separation", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2017.

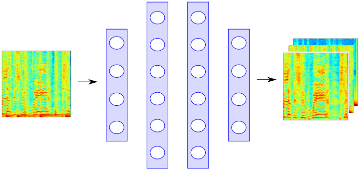
Joint estimation of magnitude and phase

Realistic source separation

Goal: estimate the magnitude **and** the phase of the sources.

- ▷ Needs an additional spectrogram-like model and estimation technique.





Approaches

- ▷ Two-stage: first estimate the magnitude, and then recover the phase.
- ▷ One-stage: jointly estimate the magnitude and the phase.

Two-stage approaches

NMF + phase recovery

- ▷ Phase recovery induces a slight improvement (interference reduction).

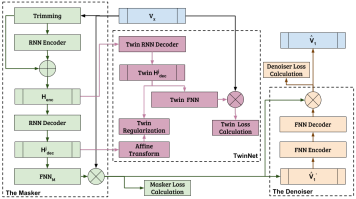
P. Magron, K. Drossos, S. I. Mimilakis, T. Virtanen, "Reducing interference with phase recovery in DNN-based monaural singing voice separation", *Proc. Interspeech*. September 2018.
K. Drossos, P. Magron, S. I. Mimilakis, T. Virtanen, "Harmonic-percussive source separation with deep neural networks and phase recovery", *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, September 2018.

Two-stage approaches

NMF + phase recovery

- ▷ Phase recovery induces a slight improvement (interference reduction).

DNN + phase recovery [Interspeech 18, iWAENC 18]



- ▷ More significant results (DNNs > NMF).
- ▷ Phase recovery makes sense on top of good magnitude estimates.

P. Magron, K. Drossos, S. I. Mimilakis, T. Virtanen, "Reducing interference with phase recovery in DNN-based monaural singing voice separation", *Proc. Interspeech*. September 2018.

K. Drossos, P. Magron, S. I. Mimilakis, T. Virtanen, "Harmonic-percussive source separation with deep neural networks and phase recovery", *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, September 2018.

Complex NMF

Baseline NMF $|X| \approx \hat{V} = WH$

- ▷ Assumes the additivity of the sources' magnitudes $\rightarrow$ phase?

P. Magron, R. Badeau, B. David, "Complex NMF under phase constraints based on signal modeling: application to audio source separation", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, March 2016.

Complex NMF

Baseline NMF $|\mathsf{X}| \approx \hat{\mathsf{V}} = \mathsf{WH}$

▷ Assumes the additivity of the sources' magnitudes $\rightarrow$ phase?

Complex NMF $\mathsf{X} \approx \hat{\mathsf{X}} = \sum_j \underbrace{[\mathsf{W}_j \mathsf{H}_j]}_{\text{NMF}} e^{i\boldsymbol{\mu}_j}$

▷ Regularize the phases with model-based properties [ICASSP 16]:

$$\min ||\mathsf{X} - \hat{\mathsf{X}}||^2 + \underbrace{\lambda_{\text{sp}} \mathcal{C}_{\text{sp}}(\mathsf{H})}_{\text{sparsity}} + \underbrace{\lambda_{\text{sin}} \mathcal{C}_{\text{sin}}(\boldsymbol{\mu})}_{\text{sinusoidal phase}} + \underbrace{\lambda_{\text{rep}} \mathcal{C}_{\text{rep}}(\boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\eta})}_{\text{repeating attacks}}$$

▷ Optimization: coordinate descent or auxiliary function method.

▷ Tuning λ_{sin} and λ_{rep}: trade-off between interference and artifact reduction.

P. Magron, R. Badeau, B. David, "Complex NMF under phase constraints based on signal modeling: application to audio source separation", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, March 2016.

Complex ISNMF

Hard to extend complex NMF to non-Euclidean metrics (e.g., beta-divergences).

P. Magron, T. Virtanen, "Complex ISNMF: a phase-aware model for monaural audio source separation", *IEEE/ACM Transactions on Audio, Speech and Language*, vol. 27, no. 1, pp. 20–31, January 2019.

Complex ISNMF

Hard to extend complex NMF to non-Euclidean metrics (e.g., beta-divergences).

Probabilistic view

(Real) Gaussian	$r \sim \mathcal{N}(m, \sigma^2)$	$m = wh$	Euclidean NMF
Poisson	$r \sim \mathcal{P}(v)$	$v = wh$	KLNMF
Isotropic Gaussian	$x \sim \mathcal{N}_{\mathbb{C}}(0, v^2 I)$	$v^2 = wh$	ISNMF
Isotropic Gaussian	$x \sim \mathcal{N}_{\mathbb{C}}(m, \sigma^2 I)$	$m = whe^{\mathrm{i}\mu}$	Complex NMF

P. Magron, T. Virtanen, "Complex ISNMF: a phase-aware model for monaural audio source separation", *IEEE/ACM Transactions on Audio, Speech and Language* , vol. 27, no. 1, pp. 20–31, January 2019.

Complex ISNMF

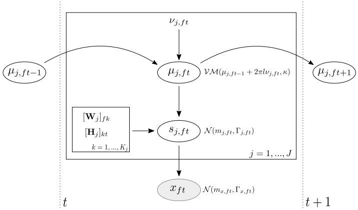
Hard to extend complex NMF to non-Euclidean metrics (e.g., beta-divergences).

Probabilistic view

(Real) Gaussian	$r \sim \mathcal{N}(m, \sigma^2)$	$m = wh$	Euclidean NMF
Poisson	$r \sim \mathcal{P}(v)$	$v = wh$	KLNMF
Isotropic Gaussian	$x \sim \mathcal{N}_{\mathbb{C}}(0, v^2 I)$	$v^2 = wh$	ISNMF
Isotropic Gaussian	$x \sim \mathcal{N}_{\mathbb{C}}(m, \sigma^2 I)$	$m = whe^{\mathrm{i}\mu}$	Complex NMF

Complex ISNMF [TASLP 19]

- Anisotropic Gaussian model (NMF variance).
- Markov chain prior on the phase parameter.
- Estimation: EM algorithm.
- Better results than Complex NMF and ISNMF.



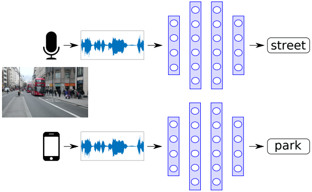
P. Magron, T. Virtanen, "Complex ISNMF: a phase-aware model for monaural audio source separation", *IEEE/ACM Transactions on Audio, Speech and Language*, vol. 27, no. 1, pp. 20–31, January 2019.

Acoustic scene classification

Domain adaptation for acoustic scene classification

Problem

- ▷ Classify an audio segment into classes.
- ▷ Mismatch between recording conditions.

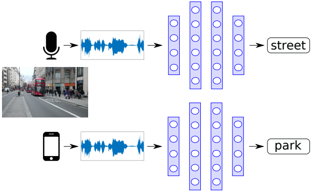


K. Drossos, P. Magron, T. Virtanen, "Unsupervised adversarial domain adaptation based on the Wasserstein distance for acoustic scene classification", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2019.

Domain adaptation for acoustic scene classification

Problem

- ▷ Classify an audio segment into classes.
- ▷ Mismatch between recording conditions.



Domain adaptation [WASPAA 19]

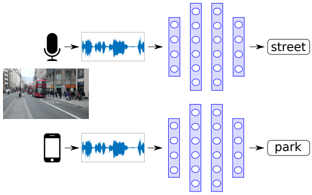
- ▷ *Source* and *target* domain data $\rightarrow$ same latent distribution.
- ▷ Wasserstein GAN: minimize scene classification error / maximize domain classification error.

K. Drossos, P. Magron, T. Virtanen, "Unsupervised adversarial domain adaptation based on the Wasserstein distance for acoustic scene classification", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2019.

Domain adaptation for acoustic scene classification

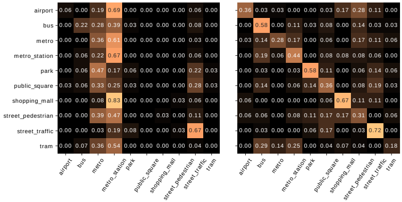
Problem

- ▷ Classify an audio segment into classes.
- ▷ Mismatch between recording conditions.



Domain adaptation [WASPAA 19]

- ▷ *Source* and *target* domain data $\rightarrow$ same latent distribution.
- ▷ Wasserstein GAN: minimize scene classification error / maximize domain classification error.



K. Drossos, P. Magron, T. Virtanen, "Unsupervised adversarial domain adaptation based on the Wasserstein distance for acoustic scene classification", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2019.

Music recommendation

Content-aware music recommendation

Collaborative filtering with matrix factorization: $Y \approx WH$.

- $\triangleright$ Y / W / H = interactions / users preferences / songs attributes.
- $\triangleright$ Cannot recommend novel items: cold-start problem.

P. Magron, C. Févotte, "Leveraging the structure of musical preference in content-aware music recommendation", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, June 2021.

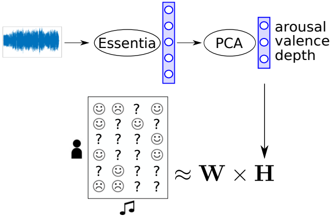
P. Magron, C. Févotte, "Neural content-aware collaborative filtering for cold-start music recommendation", to be submitted in the *ACM Transactions on Information Systems*.

Content-aware music recommendation

Collaborative filtering with matrix factorization: $Y \approx WH$.

- ▷ Y / W / H = interactions / users preferences / songs attributes.
- ▷ Cannot recommend novel items: cold-start problem.

Content-aware recommendation



P. Magron, C. Févotte, "Leveraging the structure of musical preference in content-aware music recommendation", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, June 2021.

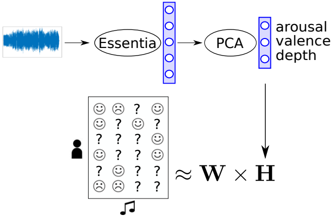
P. Magron, C. Févotte, "Neural content-aware collaborative filtering for cold-start music recommendation", to be submitted in the *ACM Transactions on Information Systems*.

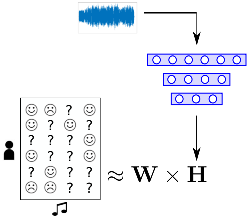
Content-aware music recommendation

Collaborative filtering with matrix factorization: $Y \approx WH$.

- ▷ Y / W / H = interactions / users preferences / songs attributes.
- ▷ Cannot recommend novel items: cold-start problem.

Content-aware recommendation





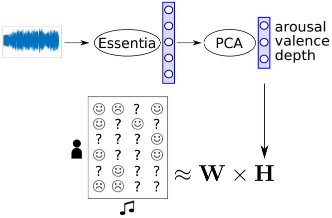
P. Magron, C. Févotte, "Leveraging the structure of musical preference in content-aware music recommendation", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, June 2021.
P. Magron, C. Févotte, "Neural content-aware collaborative filtering for cold-start music recommendation", to be submitted in the *ACM Transactions on Information Systems*.

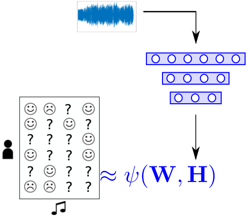
Content-aware music recommendation

Collaborative filtering with matrix factorization: Y $\approx$ WH.

- ▷ Y / W / H = interactions / users preferences / songs attributes.
- ▷ Cannot recommend novel items: cold-start problem.

Content-aware recommendation





P. Magron, C. Févotte, "Leveraging the structure of musical preference in content-aware music recommendation", *Proc. IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*, June 2021.

P. Magron, C. Févotte, "Neural content-aware collaborative filtering for cold-start music recommendation", to be submitted in the *ACM Transactions on Information Systems*.

Thanks!

- Webpage: `https://magronp.github.io/`
- Code: `https://github.com/magronp`