Phase-dependent anisotropic Gaussian model for audio source separation

Paul MAGRON, Roland BADEAU, Bertrand DAVID
LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

Introduction

Problem

- Many source separation techniques act on spectrograms (e.g. NMF).
- Phase recovery is necessary for time-domain signal synthesis.
- Traditional estimators (e.g. Wiener filtering) assume a uniform phase.

Our approach

A probabilistic model with **non-uniform** phase.

- Von Mises phase (non-tractable) $\rightarrow$ Anisotropic Gaussian model.
- Exploit a phase prior based on a sinusoidal model.
- Compute an estimator of the sources.

Von Mises mixture model

$X \in \mathbb{C}^{F\times T}$: Short-Term Fourier Transform of a mixture of K sources:

$$X = \sum_k Z_k = \sum_k V_k e^{i\phi_k}.$$

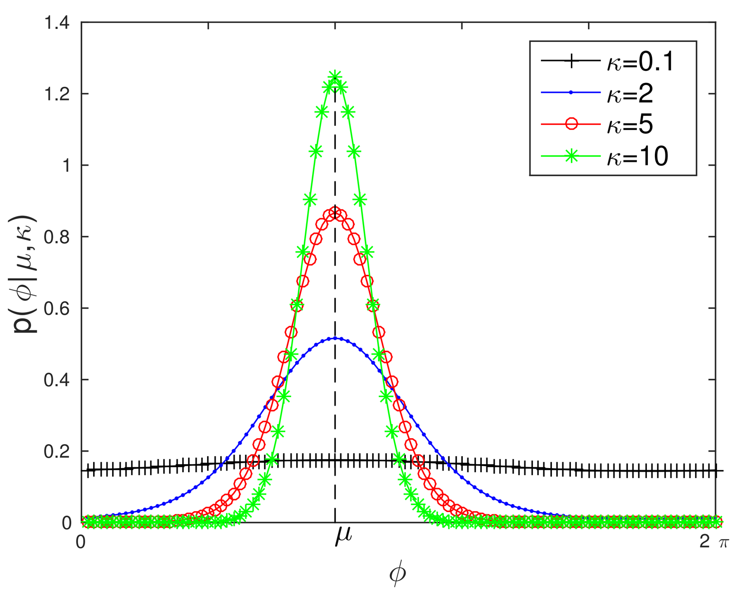
V_k is deterministic (assumed known or estimated beforehand).

Von Mises (VM) phase

$$\phi_k \sim \mathcal{VM}(\underbrace{\mu_k}_{\text{Location}}, \underbrace{\kappa_k}_{\text{Concentration}})$$

+ Max entropy distribution;
+ A tractable PDF:

$$p(\phi|\mu,\kappa) = \frac{e^{\kappa\cos(\phi-\mu)}}{2\pi I_0(\kappa)}.$$



Phase unwrapping prior

Each source is modeled as a $\sum$ of sinusoids [1]:

- Frequency peaks are estimated with QIFFT;
- Each channel f is assigned to one sine frequency $\nu_k(f,t)$;
- Phase unwrapping:

$$\mu_k(f,t) = \mu_k(f,t-1) + 2\pi S\nu_k(f,t).$$

Main drawback

A non-tractable likelihood $\rightarrow$ costly numerical methods (MCMC).

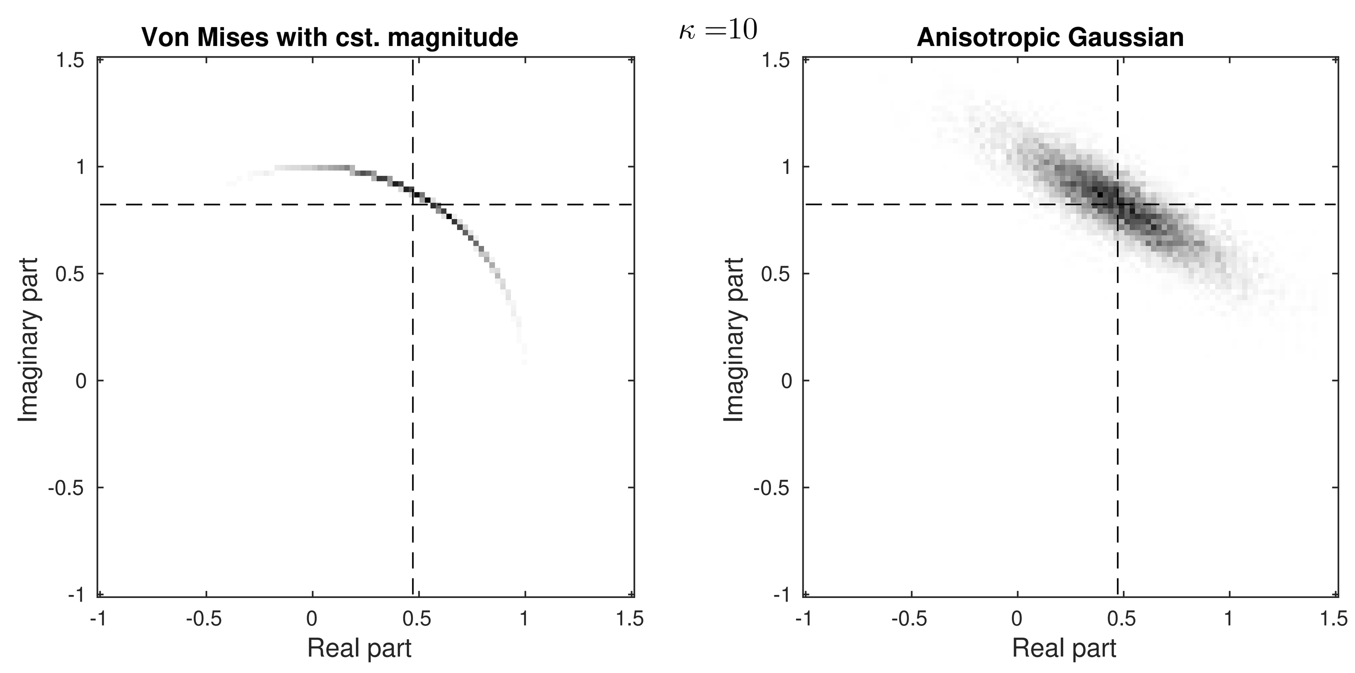
$\rightarrow$ **Approximate the VM model by a Gaussian model which keeps the phase dependencies.**

Anisotropic Gaussian model

Anisotropic Gaussian (AG) sources:

$$X_k \sim \mathcal{N}(\underbrace{m_k}_{\text{Mean}}, \underbrace{\gamma_k}_{\text{Variance}}, \underbrace{c_k}_{\text{Relation}}),\ \Gamma_k = \begin{pmatrix}\gamma_k & c_k \\ \bar{c}_k & \gamma_k\end{pmatrix}.$$

Key idea: the moments are the same ones in VM and AG models.



Mixture: $X = \sum_k X_k \sim \mathcal{N}(m_X, \gamma_X, c_X)$ with

$$(m_X, \gamma_X, c_X, \Gamma_X) = \sum_k (m_k, \gamma_k, c_k, \Gamma_k).$$

Source separation

Estimator of the sources

MMSE estimator $\hat{X}_k = \mathbb{E}(X_k|X)$.
For Gaussian mixtures:

$$\underline{\hat{X}}_k = \underline{m}_k + \Gamma_k \Gamma_X^{-1}(\underline{X} - \underline{m}_X) \text{ where } \underline{u} = \begin{pmatrix} u \\ \bar{u}\end{pmatrix}. \quad (1)$$

- Conservative: $\sum_k \hat{X}_k = X$;
- When $\kappa \rightarrow 0$: Wiener filtering $\frac{V_k^2}{\sum_l V_l^2} X$!

$\rightarrow$ Optimal combination of **prior** and **mixture phases**.

Source separation procedure

1 Phase from previous estimate: $\mu_k(f,t) = \angle \hat{X}_k(f,t-1) + 2\pi S\nu_k(f,t)$;
2 MMSE estimator given by (1);
3 Proceed to next frame.

Experimental results

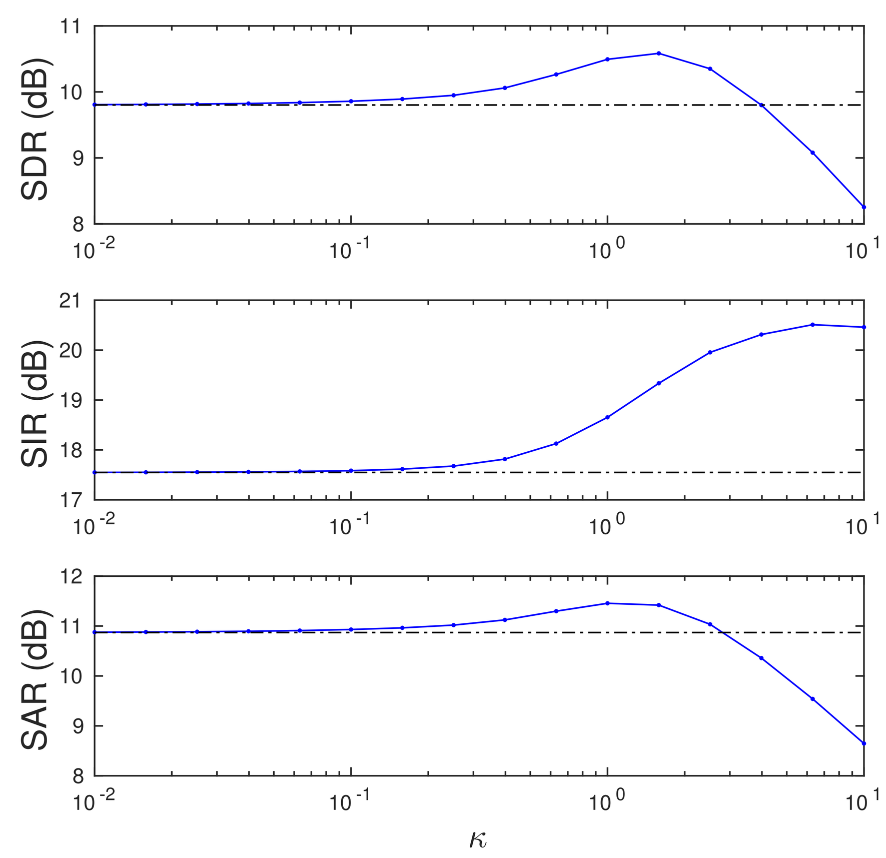
- 100 songs from the Demixing Secrets Database, $K = 4$ sources;
- Separation quality measured with the SDR/SIR/SAR (in dB).

Influence of the concentration parameter

- Constant concentration:

$$\kappa_k(f,t) = \kappa;$$

- Optimal κ tuned on the learning database (50 songs);
- For a range of κ: better results than with Wiener.



Source separation

- Test database (50 songs).
- Methods: Wiener, Consistent Wiener [2] or proposed (MMSE).

	Oracle magnitudes			Approx. magnitudes		
	SDR	SIR	SAR	SDR	SIR	SAR
Wiener	9.1	16.4	10.4	7.9	14.5	9.3
Consistent Wiener	**11.1**	**19.7**	**12.0**	**8.8**	**16.3**	**10.1**
MMSE	9.8	18.1	10.8	8.0	15.1	9.3

+ Better results than with Wiener;
− Slightly worse results than with Consistent Wiener in terms of SDR/SIR/SAR but not in simple listening tests;
+ Significantly faster than Consistent Wiener ($\approx \times 7$).

Conclusion

Model-based prior phase information
$\rightarrow$ **efficient source separation procedure.**

Future research

- Refinement of onset phase estimation;
- Modeling the uncertainty about the magnitude estimates;
- Joint estimation of magnitudes and phases: novel Complex NMF.

References

[1] P. Magron, R. Badeau and B. David, "Phase reconstruction of spectrograms with linear unwrapping: application to audio signal restoration", EUSIPCO 2015.
[2] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation", IEEE SPL 2013.