

The Bivariate Mixture Space: A Compact Spectral Representation of Bivariate Signals

GIORGIO PRESTI

giorgio.presti@unimi.it

Lab. of Music Informatics, Department of Computer Science, Università degli Studi di Milano, Milan, Italy

The Fourier Transform (FT) is a widely used analysis tool. However, FT alone is not suited for the analysis of bivariate signals (*e.g.*, stereophonic recordings), as it is not sensitive to the relationship between channels. Different works addressing this problem can be found in the literature; the Bivariate Mixture Space (BMS) is introduced here as an alternative representation to the existing techniques. BMS is still based on the FT and can be thought of as an extension of it, such that the relationship between two signals is considered as additional information in the frequency domain. Despite being simpler than other techniques aimed at representing bivariate signals, this representation is shown to have some desirable characteristics that are absent in traditional representations, which lead to novel ways to perform linear and non-linear decomposition, feature extraction, and data visualization. As a demonstrative application, an Independent Component Analysis algorithm is derived from the BMS, who shows promising results with respect to existing implementations in terms of performance and robustness.

0 INTRODUCTION

The Fourier Transform (FT) is probably the most used method for spectral analysis. It allows to investigate magnitude and phase of frequency components in a straightforward and natural way. However, when dealing with bivariate signals such as stereophonic audio recordings, several issues arise. For example, the FT of each channel does not explicitly highlight any relationship between them, such as common components. On the contrary, it may be redundant (*e.g.*, in a graphical representation of the spectrum the same axis is represented twice) and discards much of the relational information, even if computed on a linear combination of the channels, like their sum or difference.

This problem is not new, and how such analysis could be done was already discussed in literature. In particular, many alternative techniques have been developed, such as examining bivariate cross-spectral densities [1], or by exploiting complex-valued [2, 3] and quaternion-valued [4] representations. Nevertheless – despite the aforesaid methods provide a complete view of bivariate signal properties – the correct use of these techniques is hard for non-experts in telecommunication engineering, and not intuitive for many experts in other fields, such as for many audio developers. Moreover, it would be desirable to have a representation of a bivariate signal combining the immediacy of FT together with an overview of the relationship between the channels, providing a comprehensive frame-

work for both linear and non-linear decomposition, visualization, and feature extraction. For these reasons, the Bivariate Mixture Space (BMS) is here introduced as a novel contribution to the field, providing powerful yet easy-to-use analysis tools. The BMS maps two signals of one argument into a continuous function of the original argument and a new angle argument, which, when evaluated for angles 0 and $\pi/2$, returns the original signals. From the study of this function, *relational information* (*i.e.* information concerning frequency component magnitude and phase differences, and correlation) can be retrieved and represented explicitly in the frequency domain, enabling new techniques useful for signal processing, information visualization, source separation, and feature extraction.

A good example of a context dealing with bivariate signals is the processing of stereophonic audio, where BMS displays a connection with a number of works that relies on the aforementioned *relational information*; like [5], where features such as the balance between the channels, average magnitude, and phase difference are used to provide an encoding strategy for the stereophonic pair. In [6], a source separation algorithm called ADress is presented, based on brute-force minimization in a frequency-domain phase-cancellation subspace. Unfortunately, it is slow to compute and lacks a proper phase reconstruction method.

Already in 2003, in [7], a better definition of a subspace equipotent to that of ADress, and features similar to those in [5] (but based on a more general model), were available.

In [8, 9, 10], the audio signal is seen in its vectorial nature, and features of these vectors are used to approximate separation of reverberated signal from direct signal.

Unfortunately, none of the aforesaid techniques provides a framework that allows different kinds of manipulations. On the contrary, they have been developed for specific tasks. Moreover, many of these formalizations are based on a signal model that (when used for source separation) underestimates phase differences between channels when computing the distribution of components in the mixture. Finally, the most interesting technique that can relate to BMS is probably Independent Component Analysis (ICA) [11, 12] since the signal model used in BMS is linked to the *mixing matrix* inferred by ICA.

Please note that the main goal of this work is to lay down a mathematical definition of BMS. Conversely, a detailed comparison between BMS and the aforesaid techniques will be carried out in future works since it deserves appropriate and exhaustive testing. Nevertheless, a BMS-based ICA algorithm has been tested against its most adequate counterparts in the context of ICA, so to provide empirical proof of the effectiveness of the proposed approach. ICA has been chosen as a benchmark since it is the most general and less domain-specific task.

This paper is organized as follows. Section 1 provides background definitions. Section 2 presents BMS. Section 3 describes some properties of the BMS together with some operators. Section 4 discusses the relationships with ICA. Section 5 concludes the paper. A.1 contain additional proof. A.2 describes a BMS-based ICA algorithm.

1 BACKGROUND

Even if the proposed technique works on general bivariate signals, to simplify the presentation it may be useful to see such signals as a *mixture* of latent sources, intended as proper underlying sources or, more in general, the single frequency components of the bivariate signal itself.

1.1 Brief Definition of a Bivariate Mixture

A bivariate signal $\mathbf{x}(n)$ can be seen as a set of 2 signals $x_i(n)$ for $i \in \{1, 2\}$:

$$\mathbf{x}(n) = \{x_1(n), x_2(n)\}, \quad (1)$$

where each $x_i(n)$ is called a *mixture* of J latent sources $u_j(n)$ weighted by a scalar value a_{ij} such as

$$x_i(n) = a_{i1}u_1(n) + a_{i2}u_2(n) + \dots + a_{iJ}u_J(n). \quad (2)$$

Introducing vector $\mathbf{u}(n)$ to collect all sources $u_j(n)$ and the $2 \times J$ mixing matrix \mathbf{A} to collect all weights a_{ij} , a more compact definition can be given

$$\mathbf{x}(n) = \mathbf{A}\mathbf{u}(n). \quad (3)$$

1.2 A Simpler Bivariate Mixture Model

In order to tackle the problem of finding a suitable signal model, let us prune some sources of complexity. In particular, let us suppose that $x_1(n)$ and $x_2(n)$ are two orthogonal

observations of a phenomenon $\mathbf{u}(n)$ such that:

$$\begin{aligned} x_1(n) &= \sum_{j=1}^J \cos(\sigma_j) u_j(n), \\ x_2(n) &= \sum_{j=1}^J \sin(\sigma_j) u_j(n). \end{aligned} \quad (4)$$

In this way, mixing weights a_{1j} and a_{2j} are collapsed to a single parameter σ_j (this simplification implicitly assumes that observing $u_j(n)$ is an instantaneous operation preserving sources magnitude), and mixing $u_j(n)$ in $\mathbf{x}(n)$ can be seen as a rotation encoded in the j -th column vector of \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} \cos(\sigma_1) & \dots & \cos(\sigma_J) \\ \sin(\sigma_1) & \dots & \sin(\sigma_J) \end{bmatrix}. \quad (5)$$

For what concerns audio signals, note that Eq. (4) is one of the many ways to *pan* sound – *i.e.* feed different amounts of a monophonic signal to a pair of speakers – [13, 14], giving the listener the illusion of a sound located somewhere between the speakers (the *ghost source* effect, for further readings see [15]).

1.3 Rotation of bivariate signals

To understand the rationale that leads to the definition of BMS, it helps to understand what it means to *rotate* a signal. In particular, rotating a mixture can be useful to highlight information that is not visible in the original mixtures. Let us discuss an example from the audio domain.

Given a discrete-time signal $x_{lr}(n)$, usually stereophonic information is encoded as a pair of channels, representing left- and right-speaker signals (corresponding to $x_l(n)$ and $x_r(n)$ respectively). Alternatively, it is common to represent the same information as $\mathbf{x}_{ms}(n)$, consisting in the sum $x_m(n)$ and difference $x_s(n)$ of the original channels

$$\begin{cases} x_m(n) &= 1/\sqrt{2} \cdot (x_r(n) + x_l(n)) \\ x_s(n) &= 1/\sqrt{2} \cdot (x_r(n) - x_l(n)) \end{cases}, \quad (6)$$

that is a $\pi/4$ rotation, as shown in Fig. ??:

$$\mathbf{x}_{ms}(n) = \begin{bmatrix} \cos(-\pi/4) & -\sin(-\pi/4) \\ \sin(-\pi/4) & \cos(-\pi/4) \end{bmatrix} \cdot \mathbf{x}_{lr}(n). \quad (7)$$

This method is called *Mid-Side* and offers the ability to process what is perceived *in front* of the listener separately from what can be perceived *laterally*, in opposition to *Left-Right* technique where the distinction is made for signals that come from the *left* or from the *right* direction.

Fig. 1: Scatterplot of two bivariate signals, obtained distributing a cosine signal into the output observations x_l and x_r with different σ_j parameters. In signal 2 some phase difference between the channels is introduced. Vectors \mathbf{x}_m and \mathbf{x}_s are the basis of the $-\pi/4$ rotation (*i.e.* mid-side encoding).

1.4 The Frequency Domain Interpretation

Now let us consider the DFT¹ of $x(n)$, notated with uppercase letters $\mathbf{X}(f) = \{X_1(f), X_2(f)\}$, so that – for a fixed frequency f – Eq. (4) can be thought as the weighted sum of J cosine components with the same frequency f but different amplitudes and phases, depending on the input sources and the mixing matrix. If a source $U_j(f)$ does not overlap with other sources at any f (a condition which is sometime referred as *ω -disjoint orthogonality* [16]), then $X_1(f)$ and $X_2(f)$ only differs by a gain factor depending on σ_j (Signal 1 in Fig. ??). Conversely, if U_j overlaps for some f with other sources, output components may also present phase differences (Signal 2 in Fig. ??), since each component is fed with different mixtures of the sources in $\mathbf{U}(f)$ (*i.e.* the FT of $\mathbf{u}(n)$), which are very likely different in terms of phase. In this case, also magnitude differences will no longer depend only on σ_j .

Note that many real-world discrete-time signals may fall in the first scenario if a very large analysis window is used in the Fourier Transform (leading to a small Δf). The only drawback is that time-variant changes in the mixing matrix can be tracked accurately only by using a short-time frequency analysis, thus increasing Δf .

Finally, phase and magnitude differences in $\mathbf{X}(f)$ may also occur in the case of convolutive phenomena, not considered by Eq. (3). At least for what concerns magnitude differences, these aspects can be accounted by making the mixing parameter of the signal model also dependent on f (*i.e.* each frequency may have a different gain factor),

$$\begin{aligned} X_1(f) &= \sum_{j=1}^J \cos(\sigma_j(f)) U_j(f), \\ X_2(f) &= \sum_{j=1}^J \sin(\sigma_j(f)) U_j(f), \end{aligned} \quad (8)$$

such that if $\sigma_j(f)$ is constant across all frequencies, it is equivalent to that of Eq. (4), but at the same time it is also able to explain more complex mixtures. Examples of signals that can be described by this model include (but are not limited to): complex stereo mixes, pairs of tracks bleeding into each other (such as in analog tape recordings), or *input signal* against *output signal* of a delay-free system.

2 DEFINITION

2.1 Rationale

The idea behind BMS is to interpret the original mixtures X_1 and X_2 as special cases of a continuous rotation of the mixture along an angle α . This should not sound new, since the idea behind PCA is to find – under precise statistical assumptions – the angle α that maximizes the variance of the output representation. The novelty introduced by BMS is that the function of α is not computed in the time domain, where signals can be considered as random variables with disparate distributions, but in the frequency domain, where cosine signals – with well-known proper-

ties – are handled. This domain shift simplifies the study of the function and manifests some interesting properties.

To put the rationale in a more formal perspective, the BMS is the interpretation of $\mathbf{X}(f)$ as two observations of an underlying continuous space $\tilde{X}(f, \alpha)$ such that

$$\begin{aligned} X_1(f) &= \tilde{X}(f, 0), \\ X_2(f) &= \tilde{X}(f, \pi/2). \end{aligned} \quad (9)$$

Moreover, their rotation $X_m(f)$ and $X_s(f)$ (see Eq. (7)) should sample the space in $\alpha = \pi/4$ and $\alpha = 3/4\pi$ respectively. In other words, $\tilde{X}(f, \alpha)$ can be seen as a revolving surface that interpolates $\mathbf{X}(f)$. A graphical representation focusing only on the magnitude of a single frequency component is presented in Fig. ?? . It is important to stress that this is done on the Fourier decomposition of the input, thus each cosine component of the mixture is considered independently from the others.

No assumptions are made regarding sources $\mathbf{U}(f)$ mixed in $\mathbf{X}(f)$, nevertheless a different behavior is expected for time-correlated and time uncorrelated frequency components of the mixtures (*i.e.* cosine components with different phases), as discussed in section 1.4. In particular, when observations in $\mathbf{X}(f)$ have an absolute correlation $|\rho_x| \neq 0$ in the time domain, peaks in $|\tilde{X}(f, \alpha)|$ should appear for some α , while time uncorrelated frequency components should present no peaks² (see Fig. ??, ??, and ??). Finally, by studying $\tilde{X}(f, \alpha)$ it should be possible to represent $\mathbf{X}(f)$ in a more compact form, which encapsulates all the properties of the starting set of variables but in a more meaningful way. This will be discussed in Section 2.3.

2.2 Bivariate Mixture Space

Suppose a latent signal $U_1(f)$ is distributed in $\mathbf{X}(f)$ with some angle $\sigma_1(f)$, and let define $\tilde{X}(f, \alpha)$ the transformation of $\mathbf{X}(f)$ into the BMS, satisfying the desiderata described in Section 2.1:

$$\tilde{X}(f, \alpha) := X_1(f) \cdot \cos(\alpha) + X_2(f) \cdot \sin(\alpha). \quad (10)$$

As shown in Fig. ??, the magnitude of this complex function has a periodicity of π , and it peaks in correspondence of angle $\sigma_1(f)$, taking a value equal to $|U_1(f)|$ (demonstration in A.1). The behavior of $\tilde{X}(f, \alpha)$ for variations of $\sigma_1(f)$ for a fixed f can be seen in Fig. ??.

If $U_1(f)$ is represented in $\mathbf{X}(f)$ by some convolutive phenomena, or other sources are present on the same f (as

²in complex-valued statistical processing this kind of signals are generally said to be *circular*.

Fig. 2: Magnitude of the BMS $|\tilde{X}(f, \alpha)|$ for a fixed f . Original $X_1(f)$ and $X_2(f)$ have different magnitude and phase.

Fig. 3: $|\tilde{X}(f, \alpha)|$ (first two columns) and scatterplot in the time domain (third column) for a fixed f . In (A) $X_1(f) = X_2(f)$, in (B) $X_2(f) = 0$ and in (C) $X_1(f) = -0.2X_2(f)$.

Symbols and axes legend in Fig. ??.

¹When not specified, the spectrum of the whole signal is considered, nevertheless, the same concepts can be applied to each frame of a short-time Fourier Transform.

discussed in Section 1.4), a difference in phase between $X_1(f)$ and $X_2(f)$ occurs, which in turn affects $\tilde{X}(f, \alpha)$ as depicted in Fig. ???. In the same figure, it can be seen that, despite those phase differences, $|\tilde{X}(f, \alpha)|$ may peak in correspondence of some principal component angle $\sigma(f)$ (the source index subscript j is dropped, since frequency components of the mixtures do no longer belong to a single j -th source; the mixtures can be considered as a generic bivariate signal, where the J latent sources are its own frequency components, as explained in Section 1).

2.3 Bivariate Spectrum

To better represent information present in the two spectra $\mathbf{X}(f)$, a more compact transformation can be used. First, the angle $\sigma(f)$ that maximizes $|\tilde{X}(f, \alpha)|$ can be inferred by finding the zeros of its first derivative, occurring at:³

$$\sigma(f) = \frac{1}{2} \arctan \frac{2\langle X_1, X_2 \rangle}{|X_1|^2 - |X_2|^2}, \quad (11)$$

or, in a slightly more computationally efficient form (\Re and \Im denotes real and imaginary part respectively),

$$\sigma(f) = \frac{1}{2} \arctan \frac{2\Im X_1 \Im X_2 + 2\Re X_1 \Re X_2}{|X_1|^2 - |X_2|^2}. \quad (12)$$

Then, note that the Pearson linear time correlation $C(f)$ between frequency components in $X_1(f)$ and $X_2(f)$ can be computed as a function of their phase difference:

$$C(f) = \frac{\langle X_1(f), X_2(f) \rangle}{|X_1(f)| |X_2(f)|} = \cos(\angle X_1(f) - \angle X_2(f)). \quad (13)$$

$C(f)$ can be useful since those $\sigma(f)$ values resulting from overlapping partials or convolutive phenomena can be highlighted by values of $C(f) \neq \pm 1$ (see Section 1.4). Moreover, instead of computing the whole $\tilde{X}(f, \alpha)$ surface, just the principal components for each f can be saved as a *principal spectral content* (PSC) $\bar{X}(f)$ which discards the *relational information* mentioned in Section 0:

$$\bar{X}(f) = \tilde{X}(f, \sigma(f)). \quad (14)$$

Finally, let the following also be introduced, for the sake of completeness. The relational information contained in

³For the sake of readability, f argument has been omitted for X_1 and X_2 . The angle notation $\angle C$ denotes the argument of C . $\langle X_1, X_2 \rangle$ is the internal product of X_1 and X_2 , i.e. $|X_1| |X_2| \cos(\angle X_1 - \angle X_2)$.

Fig. 4: $|\tilde{X}(f, \alpha)|$ and scatterplot in the time domain for a fixed f with $X_1(f)$ and $X_2(f)$ with same magnitude but different phase. Phase difference in each row is (A): 0; (B): $\frac{1}{4}\pi$; (C): $\frac{1}{2}\pi$; (D): $\frac{3}{4}\pi$; (E): π . Note the negative angle for anti-phase cases (D) and (E), also characterized by destructive interference in the space between $X_1(f)$ and $X_2(f)$. Also, note the ambiguity in case (C) of uncorrelated signals. Symbols and axes legend in Fig. ??.

$\sigma(f)$ and $C(f)$ can be stored as a *Relational vector* $R(f)$:⁴

$$R(f) = |\sigma(f)| \cdot e^{i(\angle X_1(f) - \angle X_2(f))}. \quad (15)$$

All of this information can be packed into a vector $\mathfrak{X}(f)$ called *bivariate spectrum* (BS):

$$\mathfrak{X}(f) = \{\bar{X}(f), R(f)\}, \quad (16)$$

which – in the perspective of a bivariate analysis – is more meaningful than $\mathbf{X}(f)$ since it discards no information, but organizes it in a more straightforward form, separating the *spectral content* from the *relational content*:

- $\bar{X}(f)$ accounts for the overall magnitude and phase of the spectral content of the bivariate mixture (note that by simply summing $X_1(f)$ and $X_2(f)$ destructive phase interference may occur, which is not true for $\bar{X}(f)$);
- $C(f) = \cos(\angle R(f))$ provides information about correlation and phase differences of the input mixtures at a single frequency level. This provides an insight into the presence of convolutive phenomena, the presence of overlapping sources, and (in case of mixing parameters estimation, Section 3.2) reliability of found $\sigma(f)$ values;
- $\sigma(f) = |R(f)| \cdot \text{sgn}(C(f))$ accounts for the *balance* of the frequency components, i.e. it provides information about how each input mixture contributes to the PSC and can be used to retrieve the mixing matrix of the signal (see Section 4 and A.1).

3 PROPERTIES AND METHODS

3.1 Signal Manipulation

$\tilde{X}(f, \alpha)$, $\sigma(f)$ and $C(f)$ can be exploited to perform two different kinds of manipulation: *spectral masking* and *mixture resampling*. The former method is intrinsically non-linear, nevertheless, both methods can be used so that the outputs can exactly sum up to the input mixture.

Spectral masking is a simple way to control the magnitude of specific frequency components of the mixture by using $\sigma(f)$ or other variables as keys to mask parts of the spectrum, thus isolating components at particular positions of the BMS. For example, given an angle θ and upper and lower bounds h and l , a simple notation of masking can be introduced as:

$$\text{mask}(\theta, l, h) = \begin{cases} 1, & \text{if } \theta - l < \sigma(f) < \theta + h \\ 0, & \text{otherwise} \end{cases}, \quad (17a)$$

$$\mathbf{X}(f) \langle \theta_l^h \rangle := \mathbf{X}(f) \circ \text{mask}(\theta, l, h), \quad (17b)$$

where \circ is the Hadamard product between frequency bins and respective binary mask values.⁵ Special masks can be

⁴ $R(f)$ can be thought as a complex angle that further generalizes the model of Eq. (8).

⁵In principle also non-binary masks are possible. For example, interesting results were obtained with masks as Gaussian functions of $\sigma(f)$ with $\mu = \theta$, saturated with arctangent function,

realized by using the correlation $C(f)$ as key, for example, to split any $\mathbf{X}(f)$ in $\mathbf{X}(f)\langle + \rangle + \mathbf{X}(f)\langle - \rangle$ containing respectively positively and negatively correlated components. The first may also be referred to as the *in-phase* signal, while the latter is the *anti-phase* signal

$$\text{mask}^+ = \begin{cases} 1, & \text{if } C(f) \geq 0 \\ 0, & \text{if } C(f) < 0 \end{cases}, \quad (18a)$$

$$\mathbf{X}(f)\langle + \rangle := \mathbf{X}(f) \circ \text{mask}^+, \quad (18b)$$

$$\mathbf{X}(f)\langle - \rangle := \mathbf{X}(f) \circ (1 - \text{mask}^+). \quad (18c)$$

$C(f)$ can also split $\mathbf{X}(f)$ in $\mathbf{X}(f)\langle 1 \rangle + \mathbf{X}(f)\langle 0 \rangle$ containing respectively highly correlated and poorly correlated components. The first may also be referred to as the *clear* signal, while the latter is the *cluttered* signal (γ is a *clarity parameter* used to emphasize the relevance of $|C(f)|$). Finally, note that for values of γ close to 0, only components with correlation significantly different from 0 are selected, thus providing an approximation of the separation between circular and non-circular components

$$\mathbf{X}(f)\langle 1 \rangle := \mathbf{X}(f) \circ |C(f)|^\gamma, \quad (19a)$$

$$\mathbf{X}(f)\langle 0 \rangle := \mathbf{X}(f) \circ (1 - |C(f)|^\gamma). \quad (19b)$$

Mixture resampling is a resynthesis process that relays on $\tilde{X}(f, \alpha)$ to generate new mixtures or to rotate the mixture space. A new mixture $X_\theta(f)$ can be synthesized by choosing any $\theta(f)$ as argument for $\tilde{X}(f, \theta(f))$:

$$X_\theta(f) = \tilde{X}(f, \theta(f)). \quad (20)$$

If $\theta(f)$ is the same for all f , this operation is not different from a linear combination of the original observations, otherwise, it results in a more customized resampling.

Finally, rotation is realized simply by resampling the original mixture at an angle $\theta(f)$ and $\theta(f) + \pi/2$, in order to create new orthogonal output mixtures. Again, the difference with the same operation done in the time domain, is that a different $\theta(f)$ can be chosen for each f :

$$\mathbf{X}_\theta(f) = \{X_\theta(f), X_{\theta+\frac{\pi}{2}}(f)\}. \quad (21)$$

3.2 Distributions of Components in the BMS

Among the information that can be collected from the BS, it is important to cite the density distribution of $\sigma(f)$. However collecting the mere density distribution of $\sigma(f)$ can produce misleading plots, since frequency components with different magnitudes have the same influence on the distribution. So, to plot the dispersion of the components correctly, instead of *counting* the components inside a σ bin, it may be useful to *weight* the count by $|\bar{X}(f)|$ to account for the actual signal content. Nevertheless, in some cases, also the $|\bar{X}(f)|$ weighting can be misleading, since components with low $|C(f)|$ are placed at an angle $\sigma(f)$

where the binary behavior is a special case provided with extreme saturation.

which is very likely incorrect if compared with that in **A** (see Fig. ??, row c). To overcome this issue, it may be useful to consider the more sophisticated weighting $|\bar{X}(f)\langle 1 \rangle|$ to see the distribution of highly correlated components, or $|\bar{X}(f)\langle 0 \rangle|$ to see the contribution of poorly correlated components. Differences are shown in Fig. ???. In general, let us refer to any kind of σ *weighted-distribution* as σ_{WD} .

Note that in the case of short-time Fourier transform, the same distributions may be computed also on a *per-frame* or *per-frequency* base, providing more punctual information about the mixture, as shown in Fig. and .

3.3 BS-Enhanced Spectrogram

When analyzing signals, visualizing data is part of the process. Usually, spectrograms (*i.e.* plotting *STFT* across time and frequency, using color for magnitude) are the first tool that comes to mind for inspecting signal frequency content over time. Unfortunately, the relationship between the components of bivariate signals are hard to see with this method. A naive approach is to display the spectrogram of the sum of $X_1(f)$ and $X_2(f)$. In this way, those components which are not in phase are canceled from the plot. So, a more common way to mix the spectra is to sum only the magnitude. Anyhow, in the $|X_1(f)| + |X_2(f)|$ spectrogram, no relational information is preserved. At the same time, the display of two separate sonograms for a bivariate mixture is not easy to interpret and introduces redundancy due to the duplication of time and frequency axes.

A new kind of sonogram, visible in Fig. ??, is introduced as a way to represent BS, which encodes $|\bar{X}(f)|$ with brightness, $\sigma(f)$ with hue, and $|C(f)|^\gamma$ with saturation. The use of a *Hue Saturation Value* color mode is particularly effective in this context since it maps angular data $\sigma(f)$ to angular color information (hue), and correlation to saturation, such that uncorrelated signals (which $\sigma(f)$ is unreliable, as in Fig. ??, row c) are shown with no color information. Since variation in saturation may be hard to notice, correlation visualization may be tuned by choosing proper γ . Empirical tests show that values between $\gamma = 0$ (ignoring $C(f)$, thus relative phase) and $\gamma = 4$ (strongly emphasizing correlation) may fit most of the situations.

This method reduces the redundancy of having two separate spectrograms and highlights mixtures differences, without discarding any information but absolute phase (rel-

Fig. 5: Distribution of signal energy over the BMS for a mixture of 5 musical instruments in a stereo recording. Solid line displays overall σ_{WD} , dotted lines show σ_{WD} of highly and poorly correlated components. Horizontal axis shifted to display common components in the middle.



Fig. 6: σ_{WD} over frequency () and time () for a mixture of 5 signals (*i.e.* musical instruments in a stereo recording).

lofsubfigure“numberline()lofsubfigure“numberline()

ative phase is encoded in $|C(f)|$ and $\sigma(f)$ sign). Moreover, low-level visual cues such as brightness, hue, and saturation are processed by our brain faster than the pattern recognition task needed to compare two spectrograms [17]. For example, if a colorful image appears, it means that the channels are strongly correlated, while if a grayscale image appears, they have a very low absolute correlation.

This kind of spectrogram helps interpret the bivariate pair more as a continuum than a discrete set of mixtures, and packs a wide range of information in a compact area, letting the user recognize signal properties at a glance. Further aspects regarding the use of a prototypical version of BMS for visualization purposes have been explored in [18].

4 COMPARISON WITH INDEPENDENT COMPONENT ANALYSIS

4.1 Contextualizing ICA

As a first and solely demonstrative use case, an ICA algorithm based on BMS (BMS-ICA) has been realized and tested against other ICA implementations. Independent Component Analysis [11, 12] is a technique for linear decomposition of multivariate signals based on a 2 steps approach: a whitening phase and an iterative independence-maximization phase. It may come in different flavors depending on the definition of *independence*. The main implementations of this technique may fall in the following categories: *non-gaussianity maximization* or *mutual information minimization*. Moreover, the most general approaches only account for the marginal distributions of the samples, while other techniques (e.g., suitable for time series) also exploit spatial, temporal, or spectral diversity. In very simple terms, whitening is used to balance the contribution of latent sources by uniforming the variance and to make those sources orthogonal to each other, then the independence-maximization is used to find a suitable rotation of the new mixture such that sources are maximally separated in the new space.

One of the common drawbacks of ICA is the inconsistent behavior in terms of polarity and overall magnitude of the extracted sources. Here a simple technique to solve such an issue in the bivariate context has been adopted: given a mixing matrix \mathbf{A} , the output polarity is fixed by flipping the angle of each column vector of \mathbf{A} in the range $\pm\pi/2$ (if outside that range), while the output gain issue is fixed by dividing each of the column vectors of \mathbf{A} by its norm. This operation applied to \mathbf{A} is indicated as $\text{fix}(\mathbf{A})$. Solving

Fig. 7: BS-enhanced spectrogram. Hue is linked to σ , saturation to $|C(f)|^\gamma$ (in this case $\gamma = 1$) and brightness to $\log(|\bar{X}(f)|)$. In this example, a mix of instruments panned in the stereo image is shown (same of Fig. ?? and ??). Visible sources are: a violin in green; a kick drum and a snare in light blue; a bass guitar in teal; an electric piano in dark blue; a hi-hat in yellow, and a taiko drum in purple. $\sigma = 0$ is left, $\sigma = \pi/2$ is right and $\sigma = \pi/4$ is the middle position. $\sigma = -\pi/4$ denotes anti-phase signals.

these issues was necessary because BMS-ICA, presented in Section 4.3, is potentially free from these problems, so by avoiding these ICA drawbacks a more fair comparison is possible.

4.2 ICA and BMS

To bring ICA and BMS in the same context, it must be assumed that the mixtures are meaningful arguments for the Fourier transform (e.g., time series, images, or any uniformly sampled signal). With this assumption another definition of independence can arise: to be considered independent, latent sources must have some degree of *ω -disjoint orthogonality* [16]. In other words: source spectra must present enough non-overlapping components to let clear peaks emerge in the σ_{WD} (such as in Fig. ??). Given this assumption and the FT at the base of BMS, BMS-ICA falls into the category of ICA techniques that exploit spectral (thus spatial or temporal) diversity.

An analytical comparison of ICA and BMS is outside the scope of this work, what is being presented is an empirical demonstration of how BMS can be used to compute ICA of real-world signals as a representative use case of BMS. To ensure a fair comparison, Second-Order Blind Identification (SOBI) [19] will be used as a baseline, since – unlike methods such as FastICA [20] – it also considers temporal diversity. SOBI is an iterative source separation technique that also allows the separation of gaussian sources and relies on the diagonalization of multiple covariance matrices. Nevertheless, FastICA will also be tested, since it would be interesting to see how it performs in terms of execution time against BMS-ICA.

4.3 Comparing SOBI and FastICA with BMS-ICA

To use BMS as a tool to compute ICA, the BMS-ICA algorithm in A.2 has been implemented in MATLAB. It finds the peaks of the σ_{WD} and uses the resulting angles to reconstruct the mixing matrix as shown in Fig. ??.

Ideally, in BMS-ICA whitening should not be necessary, since the mixing matrix can be computed directly from σ_{WD} peaks. Consequently, there is no need to subtract the mean from the data. Furthermore, since the mean is represented in the first of many frequency bins of the FT, it does not influence the results as much as it does in ICA. Nevertheless, avoiding whitening can become a problem when the mixing angles σ_j are very close to each other, thus becoming indistinguishable due to the finite resolution of σ_{WD} , therefore whitening has been enabled for the tests. This and other implementation details and parameters are reported in A.2.

Fig. 8: A bivariate distribution and its actual independent components vectors, and those estimated by SOBI and BMS-ICA. Also, σ_{WD} is plotted in polar coordinates.

4.3.1 Experimental setup

To measure the fitness of BMS in the task of computing ICA, the following experiment has been realized. A dataset of audio recordings [21] is used as a pool of sources needed to generate 500 random test tasks. The dataset is preprocessed such that all recordings are brought to a sample rate of 44100 Hz, and multichannel signals are converted to single-channel by channels averaging.

Each task is created by mixing a non-silent portion of two randomly picked sources with a uniformly distributed random σ_j mixing parameter. The duration in seconds of the task mixture is randomly chosen as an integer between 1 and 30 seconds. It can be reasonably assumed that the mixed samples will be sufficiently ω -disjoint, since they are taken into the frequency domain with a very-high frequency resolution (*i.e.* the whole mixture duration).

The mixture is then fed into SOBI, FastICA, and BMS-ICA functions. These will estimate the mixing matrices. The performance of the three implementations is measured by comparing the quality of the estimated mixing matrices and the average time needed to process 1 s.

The quality is measured as Mixing Error Ratio (MER) score [22, 23]; a value expressed in dB that can be interpreted as a signal-to-noise ratio: a MER of $+\infty$ denotes an exact estimate, a MER of 0 dB means the estimate mix vector forms a 45° angle with the ground truth, and a MER of $-\infty$ when the estimated vector is orthogonal to the ground truth. The score of a single task is determined by the mean of the two associated MER scores (one for each source).

For what concerns complexity, SOBI is reported to be in the order of $\mathcal{O}(d^4m + d^2mn)$ [24], where d is the number of mixtures, n is the number of data points per mixture, and m is the number of correlation matrices used. FastICA is reported to be $\mathcal{O}(d(d+1)np)$ [25, 26], which is basically $\mathcal{O}(d^2np)$, where p is the number of iterations needed to converge to a solution. Finally, BMS-ICA is dominated by the computation of FFT repeated for d mixtures, thus having a complexity of $\mathcal{O}(dn \log n)$. Since in this context d is fixed to 2, the former complexity simplifies to $\mathcal{O}(nm)$, the second to $\mathcal{O}(np)$, and the latter to $\mathcal{O}(n \log n)$.

The advantage of BMS-ICA is to be dependent only on n , while SOBI is very sensitive to m which, according to [19], in case of noisy sources should be set at least at $m = 100$ (the experiment has been set up with this value since it deals with real-world noisy signals), and FastICA depends on p , which cannot be known *a priori*. Even if both SOBI and FastICA appear to be asymptotically faster than BMS-ICA, hidden factors and the large m or p values may play an important role, thus, to provide an idea of actual performances, the average time needed to process 1 s of audio is measured for each algorithm. The test is run over MATLAB R2021b software, using a commercial laptop, equipped with an Intel Core i7-11390H processor.

4.3.2 Results and discussion

A paired t -test revealed that SOBI and BMS-ICA MER scores (visible in Fig.) are significantly different by a small amount (Δ MER mean: -4.561 dB, SD: 13.289,

$p \ll 0.001$, DF: 499, d: 0.23). Nevertheless, the great variance of the Δ MER scores indicates that SOBI performs better in some cases, while BMS-ICA performs better in others. In detail, SOBI performed better than BMS-ICA 296 times (59.0% of tasks). Significant correlation with Δ MER scores has been found neither with signal length, mixing parameters nor signal type (*i.e.* musical instrument). A further investigation of this aspect is thus called for and will be the subject of future works.

Concerning FastICA, as expected, it has the lowest mean score. In particular, it performed significantly worse than BMS-ICA (Δ MER mean: 4.416 dB, SD: 14.724, $p \ll 0.001$, DF: 499, d: 0.22). Again, this difference is not very big, even if this time it is in favor of BMS-ICA.

A more substantial difference between the three strategies emerges when analyzing the execution time (Fig.): a Mann-Whitney U test shows that BMS-ICA is significantly faster than SOBI in terms of average time needed to process 44100 samples. ($p \ll 0.001$). In particular, SOBI never outperformed BMS-ICA, and on average it performed 12.78 times slower. BMS-ICA also outperformed FastICA ($p \ll 0.001$). In particular, FastICA was faster than BMS-ICA only 35 times (7.0% of tasks) and, on average, it performed 5.08 times slower. In conclusion, the above results show that BMS is a valid representation space for the bivariate mixture analysis, capable of providing a basis for competitive algorithms. In particular, tests showed that BMS-ICA is preferable to SOBI and FastICA in terms of execution speed, while the accuracy of SOBI is just slightly better.

5 CONCLUSION

A new technique has been presented, aimed at better representing bivariate spectral information in a compact and effective way, *i.e.* by being able to represent relational information of bivariate mixtures in the frequency domain. This technique is based on the idea of Bivariate Mixture Space, which is an interpolation of two mixtures in the frequency domain. From this auxiliary space, different representations can be rendered, such as the decomposition into *Principal Spectral Content* and *Relational Content*.

The proposed representations enable new techniques for visualization (BS-enhanced spectrogram), manipulation (Spectral Masking and Mixture Resampling), and analysis (σ weighted-distribution). To prove the usefulness of these concepts, a simple implementation of BMS-based ICA has been provided and tested, demonstrating how BMS can outperform SOBI and FastICA in terms of speed and can be considered equally effective in practice.



Fig. 9: Comparison between mean MER scores of SOBI, FastICA, and BMS-ICA in the estimation of mixing matrices () and mean time to process 1 s of audio ().

lofsubfigure“numberline()lofsubfigure“numberline()

A MATLAB implementation of these functions is provided as a GitHub repository⁶, together with instructions to reproduce the discussed experiments and other examples. The provided implementation has been tested on Fourier transform and short-term Fourier transform of audio signals. Nevertheless, it should work on any uniformly sampled signal, such as sensor readings, etc. For what concerns the limitations of the proposed techniques, it must be said that: (i) BMS only works on bivariate signals and mixtures, and (ii) unlike ICA, BMS-ICA cannot handle the separation of uniformly (and non-uniformly) distributed noise-only sources, due to the complete overlapping in frequency of this kind of signals.

6 REFERENCES

- [1] B. V. Hamon and E. J. Hannan, "Spectral Estimation of Time Delay for Dispersive and Non-Dispersive Systems," *Journal of the Royal Statistical Society*, vol. 23, no. 2, pp. 134–142 (1974 Jun.), doi:10.2307/2346994.
- [2] A. M. Sykulski, S. C. Olhede, J. M. Lilly, and J. J. Early, "Frequency-Domain Stochastic Modeling of Stationary Bivariate or Complex-Valued Signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 12, pp. 3136–3151 (2017 Mar.), doi:10.1109/TSP.2017.2686334.
- [3] J. P. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals* (Cambridge university press, Cambridge, UK, 2010).
- [4] J. Flamant, N. L. Bihan, and P. Chainais, "Spectral Analysis of Stationary Random Bivariate Signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 23, pp. 6135–6145 (2017 Aug.), doi:10.1109/TSP.2017.2736494.
- [5] M. Briand, D. Virette, and N. Martin, "Parametric Representation of multichannel Audio Based on Principal Component Analysis," in *120th Convention of the Audio Engineering Society* (Paris, FR) (2006 May), paper 6813.
- [6] D. Barry and R. Lawlor, "Sound source separation: Azimuth discrimination and resynthesis," in *Proceedings of the 7th. Digital Audio Effects (DAFx)* (2004 Oct.).
- [7] C. Avendano, "Frequency-domain source identification and manipulation in stereo mixes for enhancement, suppression and re-panning applications," in *Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 55–58 (2003 Apr.).
- [8] M. M. Goodwin and J.-M. Jot, "Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. I–9 (2007 Jun.).
- [9] S. Kraft and U. Zölzer, "Stereo signal separation and upmixing by mid-side decomposition in the frequency-domain," in *Proceedings of the 18th Digital Audio Effects (DAFx)* (2015 Nov.).
- [10] J. He, W.-S. Gan, and E.-L. Tan, "Primary-Ambient Extraction Using Ambient Spectrum Estimation for Immersive Spatial Audio Reproduction," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1431–1444 (2015 May), doi:10.1109/TASLP.2015.2434272.
- [11] J. Herault and C. Jutten, "Space or time adaptive signal processing by neural network models," in *AIP conference proceedings*, vol. 151, pp. 206–211 (1986).
- [12] P. Comon, "Independent Component Analysis, a New Concept?" *Signal processing*, vol. 36, no. 3, pp. 287–314 (1994 Apr.), doi:10.1016/0165-1684(94)90029-9.
- [13] B. A. Dower, "Sound-Transmission, Sound-Recording, and Sound-Reproducing System," US Patent 2,093,540 (1937 Sep.).
- [14] J. M. Eargle, "Stereo/Mono Disc Compatibility: A Survey of the Problems," *Journal of the Audio Engineering Society*, vol. 17, no. 3, pp. 276–281 (1969 Oct.).
- [15] B. Bauer, "Phasor Analysis of Some Stereophonic Phenomena," *IRE Transactions on Audio*, vol. AU-10, no. 1, pp. 18–21 (1962 Jan.), doi:10.1109/TAU.1962.1161613.
- [16] S. Rickard and O. Yilmaz, "On the approximate W-disjoint orthogonality of speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. I–529 (2002).
- [17] C. Ware, *Information Visualization: Perception for Design*, 3rd ed. (Elsevier, Waltham, USA, 2013).
- [18] G. Presti, G. Haus, and D. A. Mauro, "Visualization and manipulation of stereophonic audio signals by means of IID and IPD," in *40th ICMC joint with 11th SMC* (2014).
- [19] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A Blind Source Separation Technique Using Second-Order Statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444 (1997 Feb.), doi:10.1109/78.554307.
- [20] A. Hyvarinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634 (1999 May), doi:10.1109/72.761722.
- [21] M. Vinyes, "MTG MASS database," <https://www.upf.edu/web/mtg/mass> (acc. Mar. 03, 2023).
- [22] E. Vincent, S. Araki, and P. Bofill, "The 2008 signal separation evaluation campaign: A community-based approach to large-scale evaluation," in *International Conference on ICA and Signal Separation*, pp. 734–741 (2009).
- [23] E. Vincent, R. Gribonval, and C. Févotte, "Performance Measurement in Blind Audio Source Separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469 (2006 Jun.), doi:10.1109/TSA.2005.858005.
- [24] P. Tichavský, E. Doron, A. Yeredor, and J. Nielsen, "A computationally affordable implementation of an asymptotically optimal BSS algorithm for AR sources," in *14th EUSIPCO*, pp. 1–5 (2006 Mar.).
- [25] V. Zarzoso, P. Comon, and M. Kallel, "How fast is FastICA?" in *14th EUSIPCO*, pp. 1–5 (2006 Mar.).
- [26] V. Laparra, G. Camps-Valls, and J. Malo, "Iterative Gaussianization: From ICA to Random Rotations," *IEEE Transactions on Neural Networks*, vol. 22, no. 4, pp. 537–549 (2011 Feb.), doi:10.1109/TNN.2011.2106511.

⁶<https://github.com/Kuig/LIM-Toolbox>

[27] J. H. Friedman, “Exploratory Projection Pursuit,” *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 249–266 (1987), doi:10.2307/2289161.

A.1 Proof of $|\tilde{X}(f, \alpha)|$ peaking in $\sigma_j(f)$

Given Equations (8) and (10), and omitting f for better readability, $\tilde{X}(f, \alpha)$ can be written as:

$$\tilde{X}(\alpha) = \cos(\sigma_j) \cos(\alpha) U_j + \sin(\sigma_j) \sin(\alpha) U_j. \quad (1)$$

We prove that:

$$\arg \max_{\alpha} |\tilde{X}(\alpha)| = \pi \sigma_j. \quad (2)$$

Supposing $\alpha \in \mathbb{R}$ and $\sigma \in (-\pi/2, \pi/2]$,

$$\begin{aligned} |\tilde{X}(\alpha)| &= |U_j \cdot (\cos(\sigma_j) \cos(\alpha) + \sin(\sigma_j) \sin(\alpha))| \\ &= |U_j| \cdot |\cos(\alpha - \sigma_j)| \end{aligned}$$

Eq. (2) is thus equivalent to:

$$\arg \max_{\alpha} |\cos(\alpha - \sigma_j)| = \pi \sigma_j. \quad \blacksquare \quad (3)$$

The same procedure can prove that $\tilde{X}(\sigma_j) = U_j$.

A.2 BMS-ICA Algorithm

The basic idea is shown in Fig. ??:

1. Optional: whiten input (keep the de-whitening matrix \mathbf{D});
2. Compute the FFT of the two channels;
3. Compute the Bivariate Spectrum (Eq.s (12), (13), (14));
4. Compute the σ_{WD} (Section 3.2);
5. Smooth the σ_{WD} (e.g., with a moving average filter);
6. Find the top J bins of the σ_{WD} (i.e. 2 σ intervals):
 - textbf–** If signal is whitened: find the peaking σ_{WD} bin as first component σ_1 , and use the bin at $\sigma_1 + \pi/2$ as second component σ_2 ;

textbf– If the signal is not whitened, find the two highest local maxima of σ_{WD} σ_1 and σ_2 ;

7. Look inside the σ intervals previously found to refine actual signal peaks. Currently implemented heuristics include (but are not limited to):

textbf– Use the σ corresponding to the center of the distribution bin (i.e. no refinement);

textbf– Select the σ of the component in the interval with maximum magnitude;

textbf– Mean of σ of all the components in the interval, weighted by magnitude;

textbf– Mean of σ of a selection of components in the interval, weighted by magnitude; the selection includes only the top-5 percentile of the frequency components in terms of magnitude;

8. From found angles compute mixing matrix \mathbf{A} and unmixing matrix $\mathbf{W} = \mathbf{A}^{-1}$ using following equations:

textbf– in case of whitening:

$$\mathbf{A} = \text{fix} \left(\mathbf{D} \cdot \begin{bmatrix} \cos(\sigma_1) & \cos(\sigma_2) \\ \sin(\sigma_1) & \sin(\sigma_2) \end{bmatrix} \right); \quad (4)$$

textbf– in case of no whitening:

$$\mathbf{A} = \begin{bmatrix} \cos(\sigma_1) & \cos(\sigma_2) \\ \sin(\sigma_1) & \sin(\sigma_2) \end{bmatrix}. \quad (5)$$

9. Unmix the signal $\mathbf{u} = \mathbf{W}\mathbf{x}$.

In this work the following parameters were used:

- PCA Whitening [27] enabled;
- σ_{WD} calculated with weight $|X(f)| \cdot |C(f)|^{0.5}$;
- σ_{WD} calculated with 180 bins;
- σ_{WD} smoothing radius of 1 degree (1 bin)
- Final guess of σ_{WD} peak found with *top-5%* heuristic.

THE AUTHOR



Giorgio Presti

Giorgio Presti is a researcher and adjunct Professor at the Department of Computer Science at the University of Milan. He carries out research in the field of Sound and

Music Computing. Specialized in the creation of new digital tools for music signal analysis and production.