Sound Morphing Strategies based on alterations of Time-Frequency representations by Gabor Multipliers

Anaïk Olivero[1,2], Philippe Depalle[3], Bruno Torrésani [1] and Richard Kronland-Martinet[2]

[1]*Laboratoire d'Analyse, Topologie et Probabilités UMR 6632, CNRS/Aix-Marseille Université. Technopôle Château-Gombert, 39, rue F. Joliot Curie, 13453 Marseille Cedex 13, France*

[2] *Laboratoire de Mécanique et d'Acoustique UPR 7051, CNRS/Aix-Marseille Université. 31, chemin Joseph Aiguier, 13402 Cedex 20, France*

[3]*Sound Processing and Control Laboratory - SPCL. Centre for Interdisciplinary Research in Media Music and Technology - CIRMMT. Schulich School of Music - McGill University 555, Sherbrooke Street West H3A 1E3 Montreal, Qc, Canada*

Correspondence should be addressed to Anaïk Olivero (olivero@lma.cnrs-mrs.fr)

ABSTRACT

Sounds morphing is an important topic in signal processing of musical sounds and covers a wide variety of techniques whose aim is to "interpolate" between two sound signals. We present here an approach based on the alteration of time-frequency representation. Time-frequency analysis is a classical tool in sounds analysis/synthesis. A time-frequency filter can be well-defined as a diagonal signal operator in a Gabor representation of sounds. Processing can be performed by multiplying a time-frequency representation with such a time-frequency filter, called a Gabor mask. After estimating such a Gabor mask between two sounds, we explore strategies to parametrize it for static morphing between two sounds. We then compare such an approach with standard and non standard approaches of morphing as different kind of sounds combination, notably classical means in the time-frequency domain.

1. INTRODUCTION

Sound morphing is an important topic of musical and audio processing. Morphing is usually based on underlying sound models. In this paper we present an alternate strategy which allows us to work without any prior assumption or model of the signal. Our approach is based on the alteration of time-frequency representation. Time-Frequency Representations (TFR) show the evolution of the spectral content of signals over time. We focus on invertible time-frequency representations, largely used in the context of analysis/transformation/synthesis of sounds, and exploit them to perform sounds morphing.

1.1. Time-varying filters

"Time-varying filter" design is an important issue in several areas of signal processing. In the most general setting, a linear "time-varying filter" can be defined by a linear operator, that can be represented as a matrix in finite dimensional situations. Depending in the desired properties, various approximations can be made [3, 4]. Some time-frequency filters can be modeled as Gabor multipliers, i.e. linear operators that act on a signal by pointwise multiplying its time-frequency representation with a time-frequency transfer function, called a Gabor mask. Gabor masks generalize the notion of convolution to the time-frequency domain, as the convolution is a linear diagonal operator in a frequency representation. Such models have been shown [8] to be fairly accurate for modeling underspread time-varying systems, i.e. "time-varying filters" which do not involve huge time-frequency shifts. Other designs can be adopted, a "time-varying" Wiener filter can be achieved locally by considering a block-thresholding operator as in [7], or globally by defining a time-frequency Wiener filter as described in [5, 6]. In the case of non-stationary signals, such approaches improve denoising results by matching the filter to the particular time-frequency structure of the signal.

1.2. Sound transformation and morphing

The definition of a sound morphing naturally comes from perceptual considerations, and it is not so clear how to derive signals processing methods to perform such transformations. The definition of a sound morphing must be properly set, and we refer to [11], which discusses this question. Here, we define a sound morph, as an hybrid sound which timbre is intermediate between the timbre of a source and a target sound, that share same fundamental frequency, duration and loudness, according to the timbre definition of [2]. Applications of sound morphing can be found in various domains, including speech processing, sound design for composers and industry, or definition of controlling timbres in psychoacoustic experiments relating to timbre studies [10].

Sound morphing is often achieved in two steps: estimation of low level features from input and output signals (followed by several processing steps including smoothing, rescaling,...), and application of some interpolation method to the selected features. Most of these methods consider the sinusoid+noise model and interpolate the time-varying parameters (frequencies and amplitudes) of the model, [12, 13, 14]. These methods differ from their strategy to interpolate the parameters. Haken [12] improves the parameter estimation of the model, Williams [14] proposes an iterative process in order to preserve a target value of a sound descriptor and Osaka [13] designs an algorithm which is able to match two spectra with a different number of partials. Another important sound model is the source-filter model, also used by [15, 11, 16], with different strategies for interpolation (interpolation of the coefficients of filter model, or computation of a dynamic ”audio flow”). Other authors propose to use a physical model in [17], or a dynamical model in [18]. Finally, in order to achieve a more intuitive control of the morphing strategy, perceptual constraints are added in the morphing process as in [11, 14]. The temporal structure of the sound have to be morphed too, but we will limit our discussion to the spectral aspects of sound morphing. We refer to [11] for the temporal aspect of the morphing.

1.3. Our approach

To perform such a sound morphing, we assume no signal model and deal with the time-frequency representations of sounds. We consider static morphing between complete sounds, that provides hybrid sounds with their own timbre. As a general statement, we aim to formalize a morphing of sounds as any combination of signals or TFR of signals, parametrized by a morphing parameter ranging from 0 to 1, where 0 will correspond to the source signal and 1 to the target signal. First of all, let us recall that a morphing cannot be an additive mixture of sounds, i.e. a simple mix, as the timbre of a sound is related to the fine structure of its different time-variant spectrum, that allow our hearing to distinguish each sources playing together. Here, we consider a time-frequency multiplicative morphing, that performs on both modulus and phase of the TFR of source sounds. Then, the temporal and spectral features of the sounds are globally taken into account. A multiplicative mixture can be achieved with convex paths between the TFR as follow

$$[0,1] \ni \alpha \longmapsto X_\alpha = X_0^{1-\alpha} X_1^{\alpha} \in \mathscr{S} \qquad (1)$$

where X_0 is the TFR of a source signal x_0 and X_1 is the TFR of a target signal x_1 The phases of the time-frequency representations play an important role and neglecting their influence generates artifacts and audible distortions. With such morphing, the phases have to be estimated (via an unwrap of the deterministic part of the TFR, a phase reconstruction algorithm, or the phase of the source or target signals). Our approach is to design a time-frequency transfer function $\mathbf{m}_\alpha$ such that the morphed TFR are given by a mapping like

$$[0,1] \ni \alpha \longmapsto X_\alpha = \mathbf{m}_\alpha X_0$$

where the Gabor mask $\mathbf{m}_\alpha$ will depend on both source and target TFR, and is a time-frequency matrix of complex numbers, we want to estimate. This approach acts globally in the time-frequency plane, needs no signal model, and takes implicitly into account both the temporal and spectral features of the sounds. The Gabor mask will act on the time-frequency plane by combining the coefficients of the source and the target. A Gabor multiplier between two signals can indeed be estimated by a simple pointwise regularized quotient of the Gabor representations of the output and input signals. Such a Gabor mask can be obtained as

$$\mathbf{m}_\mu^* = \mathrm{argmin}_{\mathbf{m}} \|X_1 - \mathbf{m}X_0\|_2^2 + \mu \||\mathbf{m}| - 1\|_2^2 \qquad (2)$$

A morphed sound is then obtained by synthesis from the resulting TFR $X_\mu = X_0 \mathbf{m}_\mu^*$. In this formulation, the regularization parameter μ may serve as an interpolation parameter between input and output signals. More precisely, setting it to very small values yields X_μ that is very close to X_1. Doing the same with a large value of the regularization parameter yields X_μ that is very close to X_0.

A Gabor multiplier can be estimated by iterative methods as proposed in [26], or by treating time-frequency coefficients independently of each other, leading to a diagonal approximation. In the diagonal case, we provide examples showing that intermediate values of the regularization parameter yield meaningful signals that interpolate between input and output signals.

1.4. Goal of the paper

This paper is a follow-up of previous work reported in [26], where multiplicative morphing constructed in the time-frequency domain was proposed (see equation (2))[1]. In the present paper, we aim at a better understanding of this morphing strategy. We also formalize how this morphing acts in the time-frequency plane, notably how it differs from a mixing strategy via arithmetic, geometric and harmonic means. We also address the problem of time-frequency phase reconstruction, posterior to modulus morphing. This paper is organized as follows: in section 2, the mathematical background of Gabor theory is briefly described and Gabor multipliers are defined. The applications to sound morphing are discussed in section 3.

2. TIME-FREQUENCY OPERATORS

Gabor multipliers are defined in the context of Gabor representations (see [19] and references therein), which may be thought of as a subsampled version of the Short Time Fourier Transform. For the sake of simplicity, we shall limit the present discussion to the finite-dimensional setting, i.e. to signals that are supposed to be finite length vectors $x \in \mathbb{C}^L$ (with periodic boundary conditions, i.e. restrictions to $\{0,\ldots L-1\}$ of L-periodic infinite sequences). Hereafter, $\|\cdot\|$ will denote the Euclidean norm. A similar theory can be developed in $\ell^2(\mathbb{Z})$ and $L^2(\mathbb{R})$. We use the formalism of Gabor frames to perform the time-frequency representations used here, as they provide a neat mathematical framework. In addition, this framework allows us to generalize our approach to other representations involving frames such as the non stationary Gabor frames [21, 22].

2.1. Gabor frames

A Gabor frame is an overcomplete family of time-frequency atoms generated by translation and modulation on a discrete lattice of a mother window, denoted by $g \in \mathbb{C}^L$. These atoms can be written

$$g_{mn}[l] = e^{2i\pi n\nu_0(l-mb_0)} g[l-mb_0],$$

where b_0 and ν_0 are two numbers (such that L is multiple of both b_0 and ν_0), which characterize the time-frequency lattice under consideration. Here, all operations have to be understood modulo L. We set $M = L/b_0$ and $N = L/\nu_0$.

The Gabor Transformation associates to each signal $x \in \mathbb{C}^L$ its Gabor transform $\mathscr{V}_g x \in \mathbb{C}^{M\times N}$, defined by $\mathscr{V}_g x[m,n] = \langle x, g_{mn}\rangle$, wich more precisely reads

$$\mathscr{V}_g x[m,n] = \sum_{l=0}^{L-1} x[l] e^{-2i\pi n\nu_0(l-mb_0)} \overline{g}[l-mb_0].$$

Under suitable assumptions on the mother window g and with a small enough $b_0\nu_0$ product, this transform is invertible. As well known, it is possible to find mother windows g so that the perfect reconstruction of the signal is achieved by

$$\forall x \in \mathbb{C}^L,\ \ x = \sum_{m,n} \mathscr{V}_g x[m,n] g_{mn}\ .$$

Such Gabor frames are called normalized tight frames. For now on, we limit the present discussion to this case. The extension to more general situations can be done easily.

2.2. Gabor Multipliers

Let $\mathbf{m} = \{\mathbf{m}[m,n], m = 1,..,M \text{ and } n = 1,..,N\}$ denote a finite sequence of complex numbers, the Gabor multiplier $\mathbb{M}_\mathbf{m}$ associated with $\mathbf{m}$ is then defined by :

$$\mathbb{M}_\mathbf{m} x = \sum_{m,n} \mathbf{m}[m,n] \mathscr{V}_g x[m,n] g_{mn}, \tag{3}$$

where $\mathbf{m}$ is called *Gabor mask* (or the upper symbol in the mathematics literature) and can be viewed as a *time-frequency transfer function* (so that $\mathbb{M}_\mathbf{m}$ is seen as a "time-varying filter"). $\mathbb{M}_\mathbf{m}$ is then a linear operator on the space of signals $\mathbb{C}^L$ and is diagonal in the Gabor representation g_{mn}. Approximation properties of linear operator by Gabor multipliers have been studied in [8, 9].

A Gabor multiplier acts on a signal x by pointwise multiplication of the Gabor mask with the Gabor transform

[1] In this context, the signals are supposed to be similar enough in the time-frequency domain so that these transformations can be modeled as Gabor multipliers.

$\mathscr{V}_g x$ of x. Pointwise multiplication by $\mathbf{m}$ is a linear operator denoted by $\Upsilon_{\mathbf{m}}$. Formally, the action of a Gabor multiplier is then written as follows:

$$\mathbb{M}_{\mathbf{m}} x = \mathscr{V}_g^* \Upsilon_{\mathbf{m}} \mathscr{V}_g x.$$

2.3. Estimation of a Gabor Multiplier

The estimation problem is expressed as follows. Let x_0 and x_1 denote respectively source and target signals, assumed to be linked by the relation

$$x_1 = \mathbb{M}_{\mathbf{m}} x_0 + \varepsilon \ ,$$

where ε is an additive noise, and $\mathbf{m}$ is an unknown Gabor mask, which we aim at estimating. As the solution $\mathbf{m} = \mathscr{V}_g x_1 / \mathscr{V}_g x_0$, is not bounded in general, we turn to a regularized least squares solution. More precisely, we seek $\mathbf{m} \in \mathbb{C}^{M \times N}$ which minimizes the expression

$$\Phi[\mathbf{m}] = \|x_1 - \mathbb{M}_{\mathbf{m}} x_0\|^2 + \mu \ d(\mathbf{m}), \tag{4}$$

where $d(\mathbf{m})$ is a regularization term, whose influence on solution is controlled by the parameter μ. The equation (4) can be viewed as the following inverse problem: minimize

$$\Phi[\mathbf{m}] = \|x_1 - G\mathbf{m}\|^2 + \mu \ d(\mathbf{m}), \tag{5}$$

where the operator G and its adjoint read

$$G = \mathscr{V}_g^* \circ \Upsilon_{\mathscr{V}_g x_0} \ \text{ and } \ G^* = \Upsilon_{\overline{\mathscr{V}_g x_0}} \circ \mathscr{V}_g \tag{6}$$

$\Upsilon_{\mathscr{V}_g x_0}$ denoting the operator of pointwise multiplication with $\mathscr{V}_g x_0$. Notice that this operator depends on the source signal. Even in situations where a closed form expression for the solution of (5) exists (for example when the regularization term is the squared norm of the Gabor mask) the latter can hardly be exploited practically, as it involves huge matrix computation. In such cases, as well as cases where no closed form solution exist, we rather rely on dedicated numerical algorithms. We refer to [26] for the details of such iterative methods. Within this paper, we will work with a formulation of the problem defined directly in the Gabor domain, which expresses as

$$\tilde{\Phi}[\mathbf{m}] = \|X_1 - \mathbf{m} \cdot X_0\|^2 + \mu \ d(\mathbf{m}), \tag{7}$$

where X_i is the Gabor representation of the signal x_i. It is worth noticing that this amounts to a reduction of the operator G to its diagonal. Such an approximation yields a simple explicit solution for $\mathbf{m}$, which differs from the solution of problem (5). More precisely, this approach does not account for intrinsic correlations of redundant Gabor transforms, represented here by the non-diagonal terms of the matrix G. The experiments showed that the iterative method improve sounds quality of the morphed signals obtained for arbitrary choices of the the regularization parameter. Sound results of such algorithms can be found on the web page `www.lma.cnrs-mrs.fr/~kronland/olivero/aes.html` These results motivated us to try to better understand such sound morphings, in the simplified formulation (7).

We develop here some interpretations on our morphing strategy as it allows a first idea of the process. Hereafter, we will denote by S_i the modulus of X_i and φ_i its phase.

2.4. Choices of the regularization function d

More than a regularization term, the function d will provide us with a simple way to incorporate prior information on the time-frequency shape of the morphed signals. The choice $d(\mathbf{m}) = \||\mathbf{m}| - 1\|^2$ is preferred, as it avoids audible phase artifacts caused by the difficulty to handle a well-posed problem to estimate the phase of the Gabor mask of our morphing problem, and is motivated by the desire of retaining $|\mathbf{m}| = 1$ as a reference, corresponding to “no transformation”. As the phase of the Gabor transform depends only on the data-fidelity term, the phase of the Gabor mask is fixed for now on to the value

$$arg(\mathbf{m}) = \varphi_1 - \varphi_0$$

which implies that the phase of the morphed TFR equals to φ_1. Motivated by specific applications, weighted norm version can also be used; for example, introducing time-frequency-dependent weights $w_{kl} \in \mathbb{R}^+$, normalized so that $\sum_{k,l} w_{kl} = 1$, regularization terms of the following form can be used:

$$\||\mathbf{m}| - 1\|_{w,2}^2 = \sum_{k,l} w_{kl} (|\mathbf{m}(k,l)| - 1)^2 \ . \tag{8}$$

Such a choice for the regularization function d leads to an explicit solution for $\mathbf{m}$, obtained by differentiating the equation (7) with regards to the phase and the modulus of $\mathbf{m}$. This solution is the Gabor mask $\mathbf{m}_\mu$:

$$\mathbf{m}_\mu = \frac{S_0 S_1 + \mu w}{S_0^2 + \mu w} e^{i(\varphi_1 - \varphi_0)} \tag{9}$$

The weight matrix w will help us to understand some interesting comparison of the considered morphing. These weights and also μ can be used to control the morphing strategy.

Other choices of regularization can be used, such as ℓ_1 regularization, which yields Gabor masks that are 1-sparse, i.e. whose coefficients tend to be shrunk to 1 rather than 0 in the usual approaches. Notice that the choice of regularization has to be guided by applications; for the morphing application we shall describe at the end of this paper, the ℓ_2 regularization appeared to be quite adequate and leads to an easy interpretation of the morphing process. Such a regularization models the transformation to be both smooth and as simple as possible.

2.5. Gabor multipliers for audio applications

The Gabor multiplier as proposed in [20] can be used to emphasize and measure quantitatively the differences between two sounds. It is actually a much finer analysis tool as it provides a time-frequency characterization of the differences between sounds signals, that has been exploited in various ways in [28, 27, 25]. Here, we will focus on the synthesis properties of such time-frequency filters.

3. MORPHING SOUNDS WITH A TIME-FREQUENCY TRANSFER FUNCTION

Our point is not here to propose a new sound morphing method directly comparable with the state of the art, but rather to further investigate the new paradigm proposed in [26], which exploits Gabor multipliers as described above. Gabor representation therefore serves as low level representation, and Gabor masks are used for interpolation. This method uses no explicit signal model, but assumes implicitly that source and target sounds have a similar time-frequency support so that a pertinent energy is captured in the Gabor masks. Such a transformation is also guided by the signals, according to their own features.

3.1. A penalization-based morphing

More precisely, as in [26] we approach the sound morphing problem as follows : given input and output sounds (or families of sounds), estimate the Gabor mask of a Gabor multiplier that maps input to output, and associate with it a one-parameter family of Gabor masks $\mathbf{m}_\mu$, $\mu \in [0,1]$ that interpolates between unity and the so-obtained Gabor mask. We assume that $\mathbf{m}$ allows perfect reconstruction of the target, so that $X_1 = \mathbf{m}X_0$.
On the one hand, a natural choice for the one-parameter family of Gabor masks could be :

$$\mathbf{m}_\mu[m,n] = \mathbf{m}[m,n]^\mu \ . \tag{10}$$

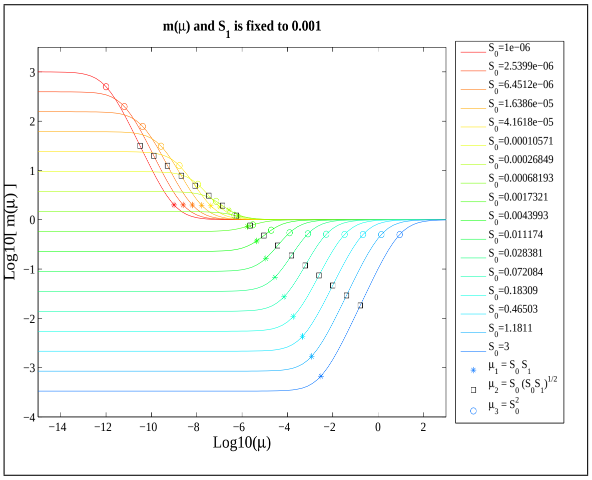

Fig. 1: Values of the Gabor mask as a function of μ, for different configurations of the couple (S_0, S_1)

This formulation is equivalent to the multiplicative combination (1) and we have to face the same phase estimation problem.
On the other hand, we will see that a similar formulation can be achieved in a different way by resolving the equation (7) with well-chosen values for the weights in the regularization function $d(\mathbf{m}) = \||\mathbf{m}| - 1\|^2_{w,2}$. The parameter μ is then used both to regularize the problem (4), and to control the morphing.

We propose in this section to study the latter approach, that uses the solutions of the above penalized approaches to estimate a mask able to perform a transformation between the two source signals. Such a morphing is closely related to a combination of the spectrograms as the Gabor mask is given by the equation (9), but the connection is far from obvious.

3.2. Interpretation of the morphing

The penalized-based morphing is closely connected to the combination approach. However this connection is not easy to establish explicitly, and we will study this connection in some details, based on a simplified example. We consider the above mask formulation (9), that we recall here

$$\mathbf{m}_\mu = \frac{S_0 S_1 + \mu w}{S_0^2 + \mu w} e^{i(\varphi_1 - \varphi_0)} \tag{11}$$

To analyze this process, we consider different values of a couple of time-frequency point (S_0, S_1) (where the indices $[m,n]$ have been omitted for the sake of clarity), as depicted on figure No. 1. The value of S_1 is fixed and the value of S_0 evolves linearly over a logarithmic scale. The figure also shows for a logarithmic (log10) scale of its axes the behavior of the mask values $|\mathbf{m}_\mu|$ as a function of μ. All curves exhibit a similar structure. As expected, when μ increases, the Gabor mask coefficients go from large values (where the target is reconstructed) to 1 (no transformation). Some interesting features on these curves can be emphasized. Two regions appear clearly, separating the points such that $S_1 > S_0$ on the top (red colors) and those such that $S_1 < S_0$ on the bottom (blue colors). And, three particular points (represented as stars, circles and squares) correspond to particular values of μ, which depend on the values of the current couple. We will denote these values by

$$\mu_1 = S_0 S_1 \ , \quad \mu_2 = S_0\sqrt{S_0 S_1} \ , \quad \text{and} \quad \mu_3 = S_0^2 \ .$$

The role of μ_3 and μ_1 are swapped depending on whether $S_1 > S_0$ or $S_1 < S_0$.

Comparison with classical means : These adapted masks allow us to recover classical morphing effects known as cross-synthesis in computer music [1]. The three particular values of μ defined correspond to morphed signals obtained as one of the three classical means evaluated between S_0 and S_1. More precisely, it is easy to show that

$$\begin{aligned} |\mathbf{m}_{\mu_1}|S_0 &= A = \frac{S_0+S_1}{2} \\ |\mathbf{m}_{\mu_2}|S_0 &= G = \sqrt{S_1 S_0} \\ |\mathbf{m}_{\mu_3}|S_0 &= H = \frac{2S_0S_1}{S_0+S_1} \end{aligned}$$

where the mask $\mathbf{m}_\mu$ is given by equation (11) and μ_i (more precisely $w\mu$) are adapted to correspond to the points observed on figure No. 1. We recognize the arithmetic, geometric and harmonic means, respectively denoted by A, G and H. A and G correspond to morphing obtained by an additive or multiplicative combination of the two spectrograms S_0 and S_1. A is simply a mix of the spectra, that leads to a mix of the two signals, while G refers to the filtering of one sound by the other one. We then conclude that classical interpolations [1] can be obtained by our approach, for suitable choices of the parameter μ.

For example, the harmonic mean H appears in the morphing process with a similar behavior comparing to the arithmetic mean. These two means follow a symmetrical behavior in figure No. 1, with respect to the "no transformation" case $|\mathbf{m}| = 1$. Then, varying μ leads to evolve from the source to the target through a path that includes classical cross-synthesis effects for specific values of μ.

We now derive another interpretation of the harmonic mean, using the notion of centroid defined thanks to Bregman divergences, studied in [24].

Interpretations of such means in terms of Bregman divergences : The harmonic mean of a data set $\{S_i : i = 1,..,n\}$ can be achieved by

$$H = \text{argmin}_S \ \frac{1}{n}\sum_{i=1}^{n} d_F(S||S_i) \tag{12}$$

where d_F is the Itakura-Saito divergence, defined as the Bregman divergence associated with $F(x) = -Log(x)$. This divergence is used in audio applications to compare two spectra, for example [23]. Similarly, the geometric mean is achieved by

$$G = \text{argmin}_S \ \frac{1}{n}\sum_{i=1}^{n} d_F(S||S_i) \tag{13}$$

where d_F is the Kullback-Leibler divergence, associated with $F(x) = x\,Log(x) - x$. On the contrary, the arithmetic mean is defined for any F, but by inverting the role of S and S_i, such that

$$A = \text{argmin}_S \ \frac{1}{n}\sum_{i=1}^{n} d_F(S_i||S) \tag{14}$$

Finally, the symmetrized version of the Itakura-Saito divergence leads to an interesting mean

$$\sqrt{AH} = \text{argmin}_S \ \frac{1}{n}\sum_{i=1}^{n} \frac{d_F(S||S_i) + d_F(S||S_i)}{2} \tag{15}$$

This mean is obtained by developing the divergences in (15) for $F(x) = -Log(x)$ and differentiating with respect to S (see `www.lma.cnrs-mrs.fr/~kronland/olivero/aes.html` for details). As the Bregman divergences are not symmetric, the right (14), left (12)-(13) and symmetrized

(15) versions lead to different centroids. A detailed study of the interpretation and the calculation of such centroids of data can be found in [24].

3.3. Discussion

This study prevents us from prompt interpretations of such morphing process. For example, finding a way to readjust the curves depicted on the figure No.1 might finally be equivalent to calculate a mean between S_0 and S_1. Controlling such a morphing process is now a question of interest that can be addressed with this new knowledge in a proper way. To control a such penalized-based morphing, the choice for the weights seems to be a good way. Another information brought that this study is that we enlighten the asymmetry of such a morphing process, when we identify the role of the source and target signals in the process.

Finally, it is worth noticing that the iterative process described in section 2 is preferably used in practice, as it leads to a better estimation of the phases of the Gabor mask. However, it prevents from such simple interpretations, as we don't have any explicit formulation for the solution of the problem (4). As a conclusion, once a regularization function d and a set of weights w have been chosen, which leads to a satisfying solution of the form (7), the use of the iterative process allows to reconstruct the phase of the Gabor mask in a proper way, so that the mask verifies the equation (4). Solutions of (7) and (4) are not the same, morphed sounds are different, but they satisfy a similar problem (a signal problem, and its version in the Gabor domain).

Sounds examples can be found on the web page `www.lma.cnrs-mrs.fr/~kronland/olivero/aes.html`

4. CONCLUSION AND PERSPECTIVES

This paper investigated the concept of Gabor Multiplier as a synthesis tool for audio signals. We enlightened how the problem of morphing sounds by estimating a Gabor Multiplier proposed in [26] between two sounds can be interpreted and controlled. This work is also a first step towards more sophisticated approaches such as the addition of stronger priors in the sounds morphing process to better control the perceived quality of the morphing. The control of such morphing processes is currently under study.

5. REFERENCES

[1] `http://support.ircam.fr/forum-ol-doc/audiosculpt/2.9.2/co/generalized-cross-synthesis_02.html`, last access on Oct, 5th 2011.

[2] ANSI (1973). Psychoacoustical terminology. American National Standards Institute, New York.

[3] Michael R. Portnoff ”Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-28, No. 1, pp 55-69, February 1980.

[4] F. Hlawatsch, G. Matz. *Linear time-frequency filters, in Time-Frequency Signal Analysis and Processing : A comprehensive Reference.* ed. B. Boashash, Oxford, UK Elsevier, 2003, ch. 11.1, pp. 466-475.

[5] J. Le Roux, E. Vincent, Y. Mizuno, H. Kameoka, N. Ono and S. Sagayama. ”Consistent Wiener Filtering : Generalized Time-Frequency Masking Respecting Spectrogram Consistency”. *IProc. 9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2010)*, pp 89-96, September 2010.

[6] F. Hlawatsch, G. Matz, H. Kirchauer and W. Kozek. ”Time-Frequency Formulations, Design and Implementation of Time-Varying Optimal Filters for Signal Estimation”. *IEEE Transactions on Signal Processing*, Vol. 48, No. 5, pp 1417-1432, May 2000.

[7] G. Yu, S. Mallat, and E. Bacry. ”Audio Denoising by Time-Frequency Block-Thresholding”. *IEEE Transactions on Signal Processing*, Vol. 56, No. 5, pp 1830-1839. May 2008.

[8] M. Dörfler and B. Torrésani, ”On the time-frequency representation of operators and generalized Gabor multiplier approximations”, *Journal of Fourier Analysis and Applications*, Vol. 16, No 2, pp. 261 –293, (2010).

[9] M. Dörfler and B. Torrésani, ”Approximation of operators by sampling in the time-frequency domain”, *Sampling Theory in Signal and Image Processing*, Vol. 10, No 2, pp. 171 –190, (2011).

[10] A. Caclin, S. McAdams, B.K. Smith and S. Winsberg. ”Acoustic correlates of timbre space dimensions : A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America,* Vol. 118, Issue 1, pp. 471-482 (2005).

[11] Marcelo Caetano. ”Morphing isolated quasi-harmonic acoustic musical instrument sounds guided by perceptually motivated features”. *PhD Thesis*, IRCAM, Université Pierre et Marie Curie, Paris, France. June 2011.

[12] Haken, L., Fitz, K., and Christensen, P. (2006). Beyond traditional sampling synthesis: Real-time timbre morphing using additive synthesis. In Beauchamp, J. W., editor, *Sound of Music: Analysis, Synthesis, and Perception*. Springer-Verlag, Berlin.

[13] N. Osaka. ”Concatenation and stretch/squeeze of musical instrumental sound using sound morphing”. *Proc. International Computer Music Conference*. (2005). Barcelona, Spain.

[14] D. Williams and T. Brookes. ”Perceptually motivated audio morphing: brightness”. *Proc. 122th AES Convention*. 2007 May 5-8. Vienna, Austria

[15] M. Slaney, M. Covell and B. Lassiter, ”Automatic Audio Morphing”, *Proc. IEEE ICASSP*, Atlanta Georgia, May 7-10. 1996.

[16] T. Ezzat, E. Meyers, J. Glass and T. Poggio. ”Morphing Spectral Envelopes Using Audio Flow”. *Proc. Interspeech/Eurospeech*. Lisbon, Portugal. September 2005.

[17] T. Hikichi and N. Osaka. ”Sound timbre interpolation based on physical modeling”. *Acoustical Science and Technology* Vol. 22, No. 2, pp. 101-111 (2001). (formerly J. Acoust. Soc. Jpn. (E))

[18] A. Röbel. ”Morphing Sound Attractors.” *Proc. of the 3rd. World Multiconference on Systemics, Cybernetics and Informatics (SCI’99) and the 5th. Int’l Conference on Information Systems Analysis and Synthesis (ISAS’99)*. Florida, 1999.

[19] H. G. Feichtinger and T. Strohmer, *Gabor Analysis and Algorithms: Theory and Applications*, ISBN: 0817639594, Birkhauser Boston, 1997.

[20] P. Depalle, R. Kronland-Martinet and B. Torrésani. *Time-Frequency mutlipliers for sound synthesis*. Proceedings of the Wavelet XII conference, SPIE annual Symposium. San Diego, 4-8 September 2007, pp. 221–224.

[21] F. Jaillet, P. Balazs, M. Dörfler. ”Non stationary Gabor frames”. *International Conference on Sampling Theory and Applications*. SAMPTA’09, International Conference on Sampling Theory and Applications, pp. 227–230, Marseille, France (2009).

[22] G. A. Velasco, N. Holighaus, M. Dörfler, T. Grill. ”Constructing a invertible Constant-Q Transform with nonstationary Gabor frames”. *Proc. International Conference on Digital Audio Effects*, pp 93-99, (Paris, France), September 19-22 2011.

[23] R. Hennequin, R. Badeau, B. David. ”Spectral similarity measure invariant to pitch shifting and amplitude scaling” in *Proc. Congrès Français d’Acoustique*, (Lyon, France), April 12-16. 2010.

[24] F. Nielsen and R. Nock. ”Sided and Symmetrized Bregman Centroids”. *IEEE Transactions on Information Theory*, Vol. 55, No. 6, June 2009.

[25] Ph. Guillemain, Ch. Vergez, D. Ferrand and A. Farcy, “An instrumented saxophone mouthpiece and its use to understand how an experienced musician play”, *Acta. Acustica united with Acustica*, Vol. 96, No. 4, pp. 622-634. (2010),

[26] A. Olivero, L. Daudet, R. Kronland-Martinet and B. Torrésani. ”A new method for Gabor Multiplier estimation : Application to sound morphing”. *Proc. European Signal Processing Conference, EUSIPCO 2010.* (Aalborg, Danemark).

[27] Anaïk Olivero. ”Identification of Time-Frequency Maps for timbre discrimination” *Proc. Internatiional Conference on Digital Audio Effects*. (Paris, France), September 19-22, 2011.

[28] A. Olivero, L. Daudet, R. Kronland-Martinet and B. Torrésani. ”Analyse et Catégorisation de sons par multiplicateurs temps-fréquence”. *Proc. XXIIe colloque GRETSI* (Dijon), 8-11 septembre 2009.