MORPHING OF TRANSIENT SOUNDS BASED ON SHIFT-INVARIANT DISCRETE WAVELET TRANSFORM AND SINGULAR VALUE DECOMPOSITION

Wasim Ahmad

I-Lab/CCSR,
University of Surrey,
Guildford, GU2 7XH,
United Kingdom
w.ahmad@surrey.ac.uk

Hüseyin Hacıhabiboğlu

CDSPR,
King's College London,
Strand, London, WC2R 2LS,
United Kingdom
huseyin.hacihabiboglu@kcl.ac.uk

Ahmet M. Kondoz

I-Lab/CCSR,
University of Surrey,
Guildford, GU2 7XH,
United Kingdom
a.kondoz@surrey.ac.uk

ABSTRACT

In this paper, a new morphing algorithm for transient sounds is introduced. Input sounds are first projected onto orthogonal bases from which intermediate or novel sounds can be generated. The proposed algorithm uses a shift invariant version of discrete wavelet transform and the singular value decomposition (SVD) to represent the input sound signals over a set of orthogonal bases. Interpolation is carried out between the weight vectors from the SVD to produce a new weight vector used for synthesising a new set of wavelet coefficients. The morphed sound is generated by taking the inverse discrete wavelet transform of the combined weighted bases. The proposed algorithm not only generates a range of new sounds, but also represents the input sounds in a more compact fashion.

Index Terms— Audio morphing, sound synthesis, interpolation, shift-invariant wavelet analysis, singular value decomposition

1. INTRODUCTION

Morphing is a general term referring to a set of widely used techniques in audio, speech, image and video processing domains. In the audio synthesis context, morphing is the process of: i) generating a smooth transition between two sounds; ii) hybridization of two sounds to generate an intermediate sound which has the characteristics of both sounds; and iii) hybridization of two or more sounds to obtain interesting and novel sounds which have no resemblance to the originals. These methods are commonly used in digital audio effects processing and to design innovative sounds for gaming, animation, and virtual reality applications.

Several algorithms have previously been proposed to morph or interpolate input sounds using signal-base models. The sound morphing algorithms presented in [1, 2, 3] used the sinusoidal model [4]. Sinusoidal model represents the input sounds as a summation of partials which are interpolated or modified to synthesise the morphed sound. Existing software tools [5, 6] have adopted these morphing algorithms. Sinusoids plus noise model [7], has also been used [8] to analyse the input sounds for the purpose of morphing. The magnitude of the partials were represented using Gaussian mixture models (GMM) and the target morphed sound was generated by interpolating between these mixtures. Time, spectral shape, and pitch were also used [9] as features and the new sound were obtained by performing a smooth interpolation between the matching features of the input sounds. These methods are based on the assumption that the input sounds are stationary or quasi-stationary making representation of the input signals as a summation of harmonics reasonable. However, these methods cannot in general synthesise or morph transient or non-stationary signals.

A wavelet-based analysis-synthesis method was previously proposed by the authors [10]. A new sound morphing algorithm based on discrete wavelet transforms and SVD is presented in this paper which allows a better representation and morphing of transient sounds. Section 2 presents a brief overview of wavelet transform and introduces the shift invariant version of discrete wavelet transform (DWT). The sound morphing algorithm is presented in Section 3. Several practical examples are presented in Section 4. Section 5 concludes the paper.

2. SHIFT-INVARIANT DISCRETE WAVELET TRANSFORM

The continuous wavelet transform of an input signal is computed by convolving the input signal with scaled and translated set of wavelets. These wavelets are generated from a single prototype wavelet $\psi(t)$ as:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{1}$$

where a and b represent scaling and translating parameters respectively. The prototype wavelet $\psi(t)$, also called *mother wavelet*, must satisfy $\int \psi(t)\mathrm{d}t = 0$ condition.

The CWT of an input signal $s(t) \in L^2(\Re)$ can be computed as:

$$\begin{aligned} \mathcal{W}(a,b) &= \langle s(t), \psi_{a,b}(t)\rangle \\ &= \int_{-\infty}^{+\infty} s(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right)\mathrm{d}t \qquad (2) \\ &= s(t) * \overline{\psi}_a(t-b) \qquad (3) \end{aligned}$$

where ψ^* denotes the complex conjugate of ψ, $(*)$ in Eq. (3) represents the convolution, and $\overline{\psi}_a = \frac{1}{\sqrt{a}}\psi^*\left(\frac{-t}{a}\right)$. Equation (3) computes the wavelet transform of the input signal that produces two sets of coefficients i.e., detail coefficients, $\mathbf{cd}$, and approximation coeffieients, $\mathbf{ca}$. Continuous wavelet transform is by itself shift-invariant, i.e. the CWT of the signal and its time-shifted versions are the same. However, computational complexity of CWT prohibits its use in real-time and limits its use mainly to high-resolution signal analysis applications.

To make the wavelet transform available for discrete-time signals, mother wavelet is sampled and a scaling factor of two is used

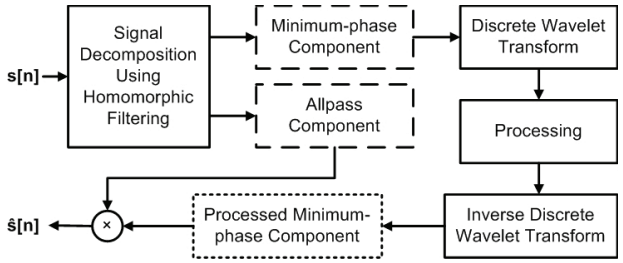

Fig. 1. Shift-invariant analysis of input signals using wavelet transform.

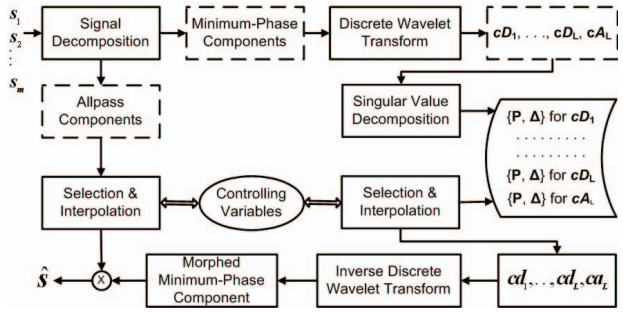

Fig. 2. Block diagram of the sound morphing algorithm.

such that the wavelet coefficients are computed only at dyadic scales. Discrete wavelet transform (DWT) also involves a decimation process which makes it a shift-variant transform. The lack of shift-invariance is a well-known drawback of the DWT and a number of shift-invariant wavelet transform schemes has been proposed in recent years [11, 12] to overcome this problem. However, these transforms are redundant and have a high computational complexity.

The authors have proposed [13] a simple shift-invariant analysis scheme for finite-length transient signals based on minimum-phase reconstruction and DWT. This shift-invariant version of DWT is non-redundant and has a low computational complexity. The input signal is decomposed into minimum-phase and allpass sequences by the cepstrum analysis. The minimum-phase sequences of the original signal and its time-shifted versions are identical for suitably band-limited signals. The time-shift present in the signal is extracted as an allpass sequence having the same phase as the original. The DWT is applied to the minimum-phase version of the input signal and the output signal can be reconstructed by reconstituting the phase and the output minimum-phase sequence. The block diagram of the shift-invariant DWT scheme used in this work is depicted in Fig. 1.

3. SOUND MORPHING ALGORITHM

The proposed morphing algorithm morphs the input sounds in the synthesis parameter domain. To generate the morphed sound, the synthesis parameters are interpolated to generate a new set of synthesis parameters used in the synthesis process. A block diagram which shows the main components of the algorithm is shown in Fig. 2. The proposed algorithm can be divided into four distinct phases: analysis, feature extraction, interpolation, and synthesis.

3.1. Analysis of input sounds

The proposed analysis block uses the shift-invariant analysis scheme using discrete wavelet transform, summarised in Sec. 2, that separates the non-minimum phase and extracts the salient features of the transient sounds. In Fig. 2, a set of sound signals, $\{\mathbf{s}_i : i = 1, \ldots, m\}$, to be morphed are input to the analysis block. These sound signals are first represented as a matrix,

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_m \end{bmatrix} = \begin{bmatrix} s_1(1) & s_1(2) & \cdots & s_1(n) \\ s_2(1) & s_2(2) & \cdots & s_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ s_m(1) & s_m(2) & \cdots & s_m(n) \end{bmatrix} \quad (4)$$

where each row contains samples of a transient sound signal, m represents the number of sounds, and n represents the length of each sound. The shift-invariant analysis using discrete wavelet transform is applied to extract the features of the input sounds. The input matrix $\mathbf{S}$ is decomposed into minimum-phase sequences matrix, $\mathbf{S}_{\text{min}}$, and allpass sequences matrix, $\mathbf{S}_{\text{ap}}$. The DWT is applied to the minimum-phase sequences matrix, $\mathbf{S}_{\text{min}}$, which decomposes it into two sets of wavelet coefficient matrices: the approximation coefficients, $\mathbf{cA_1}$, and the detail coefficients, $\mathbf{cD_1}$, where the subscript represents the level of decomposition. The approximation coefficient matrix $\mathbf{cA_1}$ is further split into two parts, $\mathbf{cA_2}$ and $\mathbf{cD_2}$, using the same scheme. This decomposition process continues up to L^{th} level which produces the following set of coefficient matrices:

$$\begin{aligned} \mathbf{cD}_i &= [\mathbf{cd}_i^{s_1} \quad \mathbf{cd}_i^{s_2} \ \ldots \ \mathbf{cd}_i^{s_m}]^{\mathbf{T}}, \quad \text{for } i = 1 \cdots L \\ \mathbf{cA}_L &= [\mathbf{ca}_L^{s_1} \quad \mathbf{ca}_L^{s_2} \ \ldots \ \mathbf{ca}_L^{s_m}]^{\mathbf{T}} \end{aligned} \quad (5)$$

where each row represents the sound feature vector from a input sound signal. The approximation coefficients represent the low-frequency components. The detail coefficients represent the high-frequency components.

3.2. Representation of sound features

The proper representation of the sound features is an essential element of the morphing process. The extracted sound features should be presented in such a way that their similarities, differences, and relationships with the input sounds are preserved and reflected in the sound parameters. In the proposed model, singular value decomposition (SVD) is used which is a powerful data analysis technique. SVD helps to identify any existing patterns in the input data, and highlights the similarities and differences [14]. The SVD of a real valued $m \times n$ rectangular matrix $\mathbf{X}$ (where $m < n$) can be written as:

$$\mathbf{X} = \mathbf{P}\ \mathbf{\Omega}\ \mathbf{Q}^T \quad (6)$$

where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m]$ is an $m \times m$ orthonormal matrix (i.e. $\mathbf{PP}^T = \mathbf{I}$), $\mathbf{\Omega}$ is an $m \times n$ rectangular diagonal matrix equal to $[diag\{w_1,\ w_2,\ \ldots, w_m\} : 0]$, and $\mathbf{Q} = [\mathbf{q}_1,\ \mathbf{q}_2,\ \ldots,\ \mathbf{q}_n]$ is an $n \times n$ orthonormal matrix (i.e. $\mathbf{QQ}^T = \mathbf{I}$). The diagonal elements w_i are the singular values, the vectors $\mathbf{p}_i$ are the left singular vectors, and the vectors $\mathbf{q}_i$ are the right singular vector of $\mathbf{X}$. Furthermore, the ω_i, $\mathbf{p}_i$, and $\mathbf{q}_i$ are sorted according to the amount of variation. Rows of matrix $\mathbf{Q}^T$ are orthonormal and form a linearly independent basis which spans the input matrix $\mathbf{X}$, and the matrix $\mathbf{\Omega}$ is a rectangular diagonal matrix with diagonal elements $\omega_1, \omega_2, \ldots, \omega_m \in \Re$. Therefore, the matrix product $\mathbf{\Omega Q}^T$ produces a matrix $\mathbf{\Delta}$ whose rows are orthogonal and also form the linearly independent basis which spans the input matrix $\mathbf{X}$. Thus, (6) becomes:

$$\mathbf{X} = \mathbf{P\Delta} \quad (7)$$

where $\mathbf{\Delta}$ is an $m \times n$ matrix. The matrices in Eq. (7) can be expanded as:

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_m \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{m1} \\ p_{12} & p_{22} & \cdots & p_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m} & p_{2m} & \cdots & p_{mm} \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_m \end{bmatrix} \tag{8}$$

where $\{\mathbf{x}_i : i = 1, \ldots, m\}$ and $\{\delta_i = w_i \mathbf{q}_i : i = 1, \ldots, m\}$ are the row vectors. An important relationship between the input matrix $\mathbf{X}$ and the orthogonal basis matrix $\mathbf{\Delta}$ is revealed in (7) and (8) i.e., each input row vector $\mathbf{x}_i \in \mathbf{X}$ can be written as a weighted linear combination of orthogonal basis vectors $[\delta_1,\ \delta_2,\ \ldots,\ \delta_m]$. Therefore, Eq. (8) can be generalised as:

$$\mathbf{x}_i = \mathbf{p}_i\ \mathbf{\Delta} = \sum_{j=1}^{m} p_{ji}\ \delta_j \quad \text{for } i = 1 \cdots m \ \text{ and } \ r \leq m \tag{9}$$

where $\mathbf{p}_i = \{p_{ij} : 1 \leq j \leq m\}$ is the weight vector for the input vector $\mathbf{x}_i$. This means that any row vector $\mathbf{x}_i \in \mathbf{X}$ can be perfectly reconstructed using the corresponding weight vector $\mathbf{p} \in \mathbf{P}$ and the orthogonal bases $\mathbf{\Delta}$ obtained using SVD.

For obtaining a common feature bases for the input sounds, the wavelet coefficient matrices $\{\mathbf{cD}_1, \ldots, \mathbf{cD}_L, \mathbf{cA}_L\}$ are decomposed using SVD and represented as in (7) where a weight matrix $\mathbf{P}$ and a set of orthogonal basis matrix $\mathbf{\Delta}$ are obtained to represent each coefficient matrix. Therefore, (5) can be expressed as:

$$\begin{aligned} \mathbf{cD}_i &= \mathbf{P}_{cD_i} \mathbf{\Delta}_{cD_i}, \quad \text{for } i = 1 \cdots L \\ \mathbf{cA}_L &= \mathbf{P}_{cA_L} \mathbf{\Delta}_{cA_L} \end{aligned} \tag{10}$$

This presents a parametrisation of several sounds on a set of common bases and can be used to synthesise any one of the input sounds perfectly using the full set of orthogonal bases ($r = m$) or approximately using first few orthogonal bases ($r < m$).

3.3. Interpolation in the synthesis parameter domain

In the proposed parameterisation, the weight vectors control the characteristics of the generated sound. Different types of sounds can be generated by interpolating between these weight vectors. A simple linear interpolation scheme is easier for obtaining intermediate parameters and simplifies their mapping. However, other interpolation strategies with a better perceptual correlation can also be applied. Linear interpolation between two weight vectors $\mathbf{p}_i$ and $\mathbf{p}_j$ can be expressed as:

$$\overline{\mathbf{p}} = \alpha\ \mathbf{p}_i + (1 - \alpha)\ \mathbf{p}_j \quad \text{where } i \neq j \text{ and } 0 \leq \alpha \leq 1 \tag{11}$$

where α is the interpolation coefficient and $\overline{\mathbf{p}}$ is the morphed weight vector that is used to synthesise the morphed sound over the common bases. Linear interpolation can be carried out between more than two weight vectors. For example, to generate a new weight vector $\overline{\mathbf{p}}$ by interpolating between three weight vectors can be written as:

$$\overline{\mathbf{p}} = \alpha\ \mathbf{p}_i + \beta\ \mathbf{p}_j + (1 - \alpha - \beta)\ \mathbf{p}_k. \tag{12}$$

In Eq. (12), if $0 \leq \alpha + \beta \leq 1$ and $0 \leq \alpha, \beta \leq 1$ then the interpolated weight vector $\overline{\mathbf{p}}$ resides on the hypertriangle formed by the original weight vectors in the multidimensional vector space. Once the SVD of the approximation and detail wavelet coefficients of the input sounds are obtained, the interpolation is carried out between the weight vectors.

3.4. Synthesis of the morphed sound

In the synthesis process, the target sound is generated by taking the inverse discrete wavelet transform (IDWT) of the set of approximation and detail coefficients which are obtained by weighting the orthogonal bases with the interpolated weight vectors. The generation of target sound and the interpolation of weight vectors in synthesis parameters domain can thus be expressed as:

$$\hat{\mathbf{s}}(n) = \text{IDWT} \begin{cases} \mathbf{cd}_i &= \overline{\mathbf{p}}_{cD_i} \mathbf{\Delta}_{cD_i}, \quad \text{for } i = 1 \cdots L \\ \mathbf{cd}_L &= \overline{\mathbf{p}}_{cA_L} \mathbf{\Delta}_{cA_L}. \end{cases} \tag{13}$$

where $\hat{\mathbf{s}}(n)$ is the target morphed sound, $\overline{\mathbf{p}}_{cD_i}$ and $\overline{\mathbf{p}}_{cA_L}$ are the interpolated weight vectors for detail and approximation components respectively. These weight vectors can be calculated independently for each set of bases using (11).

The residual phase response obtained as a result of minimum-phase reconstruction can first be interpolated and then reconstituted either by convolving the obtained morphed output with the allpass sequence containing the phase information, or by designing and using an allpass IIR filter modeling the excess phase response.

4. MORPHING EXAMPLES

The presented sound morphing algorithm is implemented on everyday impact sounds to generate intermediate and novel sounds. A group of impact sounds were recoded in an acoustical booth ($T_{60} < 100$ ms). These contain bumping sounds of football, basketball, and taped tennis ball (tennis ball covered with plastic insulation tape) on laminate floor. A sequence of sound events was recorded for each sound source at a sampling rate of 44.1 kHz.

4.1. Intermediate sounds

Two transient sounds from the same type of acoustical interaction under different conditions can be used in the morphing operation to generate physically plausible intermediate sounds. The presented morphing model is used to morph between two sounds of a football dropped from a height of 40 cm and from 120 cm, respectively, to generate the sound of a football dropped from an intermediate height. The input sounds were decomposed up to 5^{th} level using the 'db4' wavelet and the features were represented as orthogonal bases and weight vectors using the SVD. The interpolation between the weight vectors are performed as in (11) with $\alpha_{cD_1} = 0.4$, $\alpha_{cD_2} = \alpha_{cD_3} = 0.2$, $\alpha_{cD_4} = \alpha_{cD_5} = 0.6$, $\alpha_{cA_5} = 0.5$, and the intermediate sound is generated as described. The original input sounds and the generated intermediate sound $\hat{\mathbf{s}}$ are plotted in Fig. 3. Their magnitude spectra are shown in Fig. 4. It may be observed from Figs. 3 and 4 that the generated sound has a magnitude spectrum in between the magnitude spectra of the two input sounds. The interpolation and manipulation of allpass sequence also play an important role in the perception of the generated sound.

4.2. Novel Sounds

Two or more sounds from different sound sources can be interpolated to generate novel sounds which do not correspond to any physical interaction but are merely hybrid sounds. The presented morphing model is used to interpolate taped tennis ball with the basketball to generate a novel sound. The synthesis weight vectors were interpolated to generate new weight vector using Eq. (11) with $\alpha_{cD_2} = \alpha_{cD_4} = 0.2$, $\alpha_{cD_1} = \alpha_{cD_3} = \alpha_{cD_5} = \alpha_{cA_5} = 0.5$. The input sounds and the synthesised sound are plotted in Fig. 5.

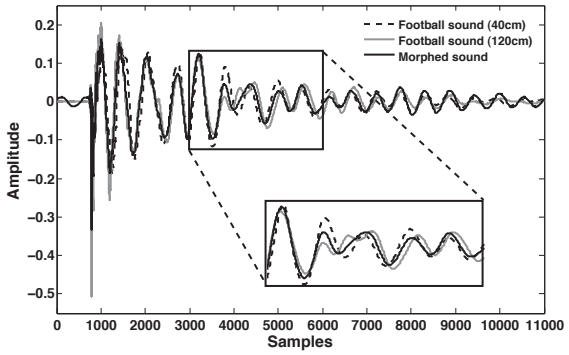

Fig. 3. Input football sounds and the generated intermediate sound.

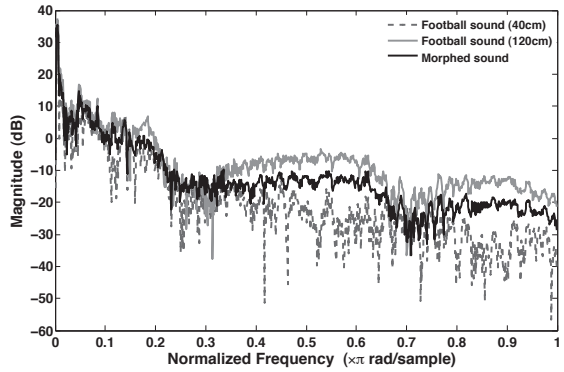

Fig. 4. Magnitude spectrum of input football sounds and the generated intermediate sound.

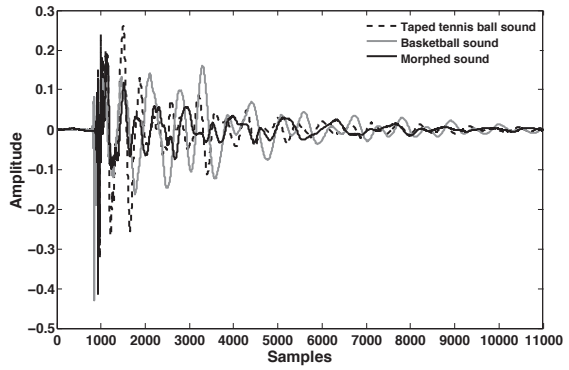

Fig. 5. Input taped tennis ball, basketball, and the generated novel sounds.

5. CONCLUSIONS

A sound morphing algorithm was presented in this paper which can be used to generate intermediate and novel sounds from a set of input sounds. It was suggested that the interpolation between the input sounds can be performed in the synthesis parameter domain. The presented method uses a shift-invariant version of discrete wavelet transform to analyse transient impact sounds and extracts the salient sound features. The wavelet coefficients of the input signals are projected onto a set of common bases by the singular value decomposition. The morphing is carried out between the weight vectors obtained from the singular value decomposition and the common bases for each set of detail and approximation wavelet coefficients. The morphed sound is obtained by inverting the wavelet transform and incorporating the interpolated phase. Two examples were presented. The first example shows how an intermediate sound can be obtained from two transient input sounds obtained from the same type of acoustic interaction. The second example shows how a novel sound can be obtained from two sounds from different types of acoustic interaction.

6. REFERENCES

[1] M.-H. Serra, D. Rubine, and R. Dannenberg, "Analysis and synthesis of tones by spectral interpolation," *J. of Audio Eng. Soc.*, vol. 38, no. 3, pp. 111–128, March 1990.

[2] E. Tellman, L. Haken, and B. Holloway, "Timbre morphing of sounds with unequal numbers of features," *J. of Audio Eng. Soc.*, vol. 43, no. 9, pp. 678–689, 1995.

[3] N. Osaka, "Timbre interpolation of sounds using a sinusoidal model," in *Proc. Int. Computer Music Conf.*, Banff, Canada, Sep. 1995, pp. 408–411.

[4] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Tran. on Acoustics, Speech, and Signal Proc.*, vol. 34, no. 4, pp. 744–754, August 1986.

[5] L. Haken, "Real-time timbre modifications using sinusoidal parameter streams," in *Proc. Int. Computer Music Conf.*, Banff, Canada, Sep. 1995, pp. 162–1163.

[6] K. Fitz and L. Haken, "Sinusoidal modeling and manipulation using lemur," *Computer Music Journal*, vol. 20, no. 4, pp. 44–59, 1996.

[7] X. Serra and J. Smith, "Spectral modeling synthesis: A sound analysis/synthesis based on a deterministic plus stochastic decomposition," *Computer Music J.*, vol. 14, no. 4, pp. 12–24, 1990.

[8] F. Boccardi and C. Drioli, "Sound morphing with Gaussian mixture models," in *Proc. 2001 Int. Conf. Digital Audio Effects*, Limerick, Ireland, Dec. 2001, pp. 44–48.

[9] M. Slaney, M. Covell, and B. Lassiter, "Automatic audio morphing," in *Proc. IEEE Int. Conf. Acoust. Speech and Signal Process.*, Atlanta, GA, May 1996, pp. 1001–1004,.

[10] W. Ahmad, H. Hacıhabiboğlu, and A. M. Kondoz, "Analysis-synthesis model for transient impact sounds by stationary wavelet transform and singular value decomposition," in *Proc. Int. Computer Music Conf.*, 2008.

[11] G. P. Nason and B. Silverman, "The stationary wavelet transform and some statistical applications," in *Lecture Notes in Statistics, 103*, 1995, pp. 281–299.

[12] R. R. Coifman and D. L. Donoho, "Translation invariant denoising," in *Lecture Notes in Statistics, 103*, 1995, pp. 125–150.

[13] W. Ahmad, H. Hacıhabiboğlu, and A. M. Kondoz, "Shift-invariant analysis of transient signals based on minimum-phase reconstruction and discrete wavelet transforms," manuscript in preparation, 2008.

[14] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, ser. Springer Series in Statistics. Springer-Verlag, 1997, pp. 109–132.