ISSN: 2321-8169
2023 - 2028

A Linear Hybrid Sound Generation of Musical Instruments using Temporal and Spectral Shape Features

Noufiya Nazarudin, PG Scholar, Arun Jose, Assistant Professor
Department of Electronics and Communication Engineering
TKM Institute of Technology
Kollam, India
noufishajil610@gmail.com

Abstract**—**The generation of a hybrid musical instrument sound using morphing has always been an area of great interest to the music world. The proposed method exploits the temporal and spectral shape features of the sound for this purpose. For an effective morphing the temporal and spectral features are found as they can capture the most perceptually salient dimensions of timbre perception, namely, the attack time and the distribution of spectral energy. A wide variety of sound synthesis algorithms is currently available. Sound synthesis methods have become more computationally efficient. Wave table synthesis is widely adopted by digital sampling instruments or samplers. The Over Lap Add method (OLA) refers to a family of algorithms that produce a signal by properly assembling a number of signal segments. In granular synthesis sound is considered as a sequence with overlaps of elementary acoustic elements called grains. The simplest morph is a cross-fade of amplitudes in the time domain which can be obtained through cross synthesis. A hybrid sound is generated with all these methods to find out which method gives the most linear morph. The result will be evaluated as an error measure which is the difference between the calculated and interpolated features. The extraction of morph in a perceptually pleasant manner is the ultimate requirement of the work.

Index Terms**—***Hybrid Instrument Sound, Temporal and Spectral Shape Features, Acoustic correlates of timbre.*

*****

I. INTRODUCTION

Every musical instrument sound is unique in nature. The modern instruments have been evolved from the older ones. The hybrid sounds should have the characteristics of both the sounds. Morphing is the best option to achieve such a sound. Morphing is seen commonly applied to images, where one picture smoothly changes to the other. Sound morphing also intends to produce such a smooth transition. It involves the hybridization of two sounds where the auditory features are fused together. One important requirement is that the result should blend into a single percept, but should not simply mix or cross fade the sounds. The musical instrument sounds can be analyzed in different contexts. Musical scale gives the perceived distance between two pitches when one is twice the frequency of the other and represented as logarithm of pitch. The frequency or pitch content of the sounds produced by musical instruments can be analyzed with different approaches of a musical scale[2]. A.Zlatintsi and P.Maragos has suggested that multiscale fractal dimension (MFD) [3] profile can be used as a short time descriptor to quantify the multiscale complexity and fragmentation of the different states of the music waveform.

In music, timbre is the quality of a musical note or sound that distinguishes different types of sounds. Based on the similarity and dissimilarity ratings, timbres can be arranged in an n-multidimensional space[4]. Timbre representation can be built upon spectral parameters extracted from samples of sounds performed along the entire pitch range of the single instrument [5]. It establishes a correspondence between the closest peak values in adjacent frames and associates these values to instantaneous frequency and amplitude values of harmonic components. It is seen that there is a correlation between the amplitude values at different time positions in the envelope [6]. Spectral modification on timbre can produce some perceptual effects [7]. Due to the desirable properties of Mel-frequency cepstral coefficients [8], like linearity, orthogonality and multi-dimensionality, it can be chosen as a hypothetical metric for spectral envelope perception.

Audio morphing generates a smooth transition from the source sound to the target sound by preserving their shared characteristics. Windowing and framing of the sound is considered as the preprocessing step [9]. The harmonic temporal variation of sound can be represented in terms of Wigner time frequency distribution as it gives a good localization in both time and frequency [10]. If two natural sounds with different timbres are given, there may arise a situation where a sound that interpolates the timbre has to be synthesized [11]. The cortical model is a computational model to observe how the brain is able to obtain and integrate the multitude of cues like loudness, location, timbre, and pitch arriving at the ears [12]. 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 [13]. For timbre morphing the features with high-level descriptors are measured so that a sound with intermediate descriptors should be perceived as intermediate [14]. Musical segmentation has found variety of applications in automatic musical accompaniment, sound modeling and manipulation techniques.

In cases where a given musical score consists of two parts: a solo part and an accompaniment it is possible to create a computer program that listens to a live performer playing the solo part and generates the accompaniment in real time [15]. Musical instrument sounds segmentation naturally depends on the correct detection of the boundaries of the regions [16]. The amplitude modulations of musical instrument sounds are important perceptual cues and should outline the waveform connecting the main peaks and should avoid over fitting. The classical amplitude envelope estimation techniques include low-pass filtering (LPF), root-mean square (RMS), analytic signal and frequency-domain linear prediction (FDLP)[17]. Partial Tracking has always been a challenging task in music audio processing systems. Linear prediction is a low complex algorithm that can be used to track and interpolate partials in the context of sinusoidal modeling [18].

II. OTHER SOUND SYNTHESIS ALGORITHMS

A wide variety of sound synthesis algorithms is currently available[19]. Each one of them exhibits their own individuality. Musicians can nowadays access a wide collection of synthesis techniques. Different types of sound synthesis methods are available to generate a new or hybrid sound.

1) WAVETABLE SYNTHESIS: A sound can be reproduced through recording if there exist a sample reference sound. Wavetable synthesis is such a method where a device called sampler is used to store and play back a large quantity of recorded sounds. Unlimited variety of sounds can be synthesized through wavetable synthesis. From the implementation viewpoint, computational simplicity is certainly an advantage of the technique, which contrasts with the need of huge memory capacities.

2) SYNCHRONOUS OVERLAP-ADD (SOLA) METHODS: The definition Overlap-Add (OLA) refers to a family of algorithms that produce a signal by assembling numerous signal parts. The segments $x_m[n]$ produced through windowing can be constructed from a given sound signal x[n],as

$$x_m[n] = x[n]w_a[n-mS_a], \quad (1)$$

where $w_a[n]$ is an analysis window and S_a indicates the time difference between two consecutive frames[19]. If the window w_a is N samples long, then the block size, i.e. the length of each frame $x_m[n]$, will be N. In order for the signal segments to actually overlap, the inequality $S_a <= N$ must be verified. When $S_a = N$ the segments are exactly juxtaposed with no overlap. In order to avoid phase discontinuities at the boundaries between frames, a proper time alignment of the blocks has to be chosen. The synchronous overlap-add (SOLA) algorithm realizes such a proper alignment, and provides a good sound quality while remaining computationally simple, which makes it suitable even for real-time applications.

3) GRANULAR SYNTHESIS: The term "granular synthesis" can be used to define a family of synthesis methods that share the basic idea of building complex sounds from simple sounds. According to Granular synthesis, grains are the elementary acoustic components which are sequenced together for sound formation. The sound timbre is determined by the features of the grains and their temporal location. This method allows the real sounds to be organized in succession, whether they are complex waveforms or spectra. In this way, it is possible to reproduce real sounds accurately and modify their dynamic characteristics as they are partially overlapped in time. In this respect granular synthesis can be viewed as an OLA technique in which segments $x_m[n]$ of a sound signal x[n] represent the grains, and are processed both in time and frequency before being reassembled.

4) CROSS SYNTHESIS: The simplest morph is a cross-fade of amplitudes in the time domain. This is not spectral audio morphing but involves fading out and fading in of the amplitudes of the first and second sound respectively. The transition from one sound to another will occur when these fade outs are added or overlapped. In cross synthesis the spectral envelope of two sounds are overlapped one over the other. The smooth spectral envelope is obtained through the cepstrum of the sound.

III. METHODOLOGY

There are some features which are to be derived from the two sounds before performing the morphing. These features are useful in allowing the timbre transition from one sound to another. In order to make the intermediate feature values to correspond to intermediate positions in the timbre space the features selected should correlate acoustically to timbre dimensions [20].

A. Temporal and Spectral Shape Features

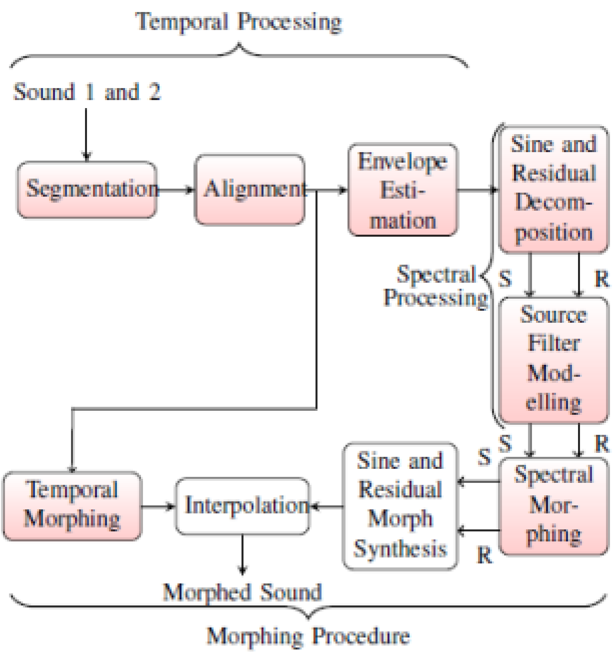

Fig. 1. Basic blocks of musical instrument sound morphing

Temporal features are global descriptors as they are computed for the whole signal. They are directly acquired from the waveform or the signal energy. The spectral shape features are instantaneous features as they are computed for each time frame. They are calculated from the Short Time Fourier Transform (STFT) of the signal.

1) Temporal Features: In all sounds the beginning of the acoustic stimulus is defined by the term attack. It is one of the most perceptually salient features of musical instrument sounds. The log attack time Λ is the logarithm of the time duration between the time the signal starts λ_1 to the time it reaches its stable part λ_2. The measure of the balance of energy distribution along the course of a sound is calculated in terms of temporal centroid τ. It is defined as the time averaged over the energy envelope a(t). The temporal centroid helps us to compare and distinguish between two sound classes because it varies more significantly for both. The log attack time is defined as follows:

$$\Lambda = \log(\lambda_2 - \lambda_1) \quad (2)$$

2) *Spectral Shape Features:* In psychoacoustic studies, there is a salient feature correlated with the verbal attribute "brightness" which is the spectral centroid. Spectral spread is a measure of the bandwidth. The spectral shape features δ_i are the first four standardized moments of the normalized magnitude spectrum p(k) viewed as a probability distribution defined as

$$p(k) = |X(k)| / \Sigma_k |X(k)| \quad (3)$$

where |X(k)| is the magnitude spectrum, the frequencies k are the possible outcomes, and the probabilities to observe them are p(k). The spectral centroid δ_1 is the mean of p(k) and the spectral spread δ_2 is the variance around the mean. The spectral skewness δ_3, measures the asymmetry of p(k) around the spectral centroid, while spectral kurtosis δ_4, is a measure of the peak relative to the normal distribution.

B. Basic Block Diagram

The hybrid musical instrument sound generation technique includes mainly three steps namely temporal processing, spectral processing and morphing procedure. The most advantageous and challenging factor is that the interpolation is based on a single parameter. Selective frequency tuning is possible as the morph for different morphing factors ranging from 0 to 1 can be obtained. The most hybrid sound will be given by the morph for an interpolation factor of 0.5.

In Fig.1 the coloured blocks indicate their duplication, i.e, these processes have to be performed separately on both the source and target sounds. The white blocks indicate the processing of a single input which is the resultant after morphing of the sounds. These three steps are described in the following sections.

I. Temporal Processing

The source and target sounds are subjected to temporal processing. As the name implies it is a time domain activity. From Fig.1 it is evident that temporal processing includes three steps. They are segmentation, alignment and envelope estimation.

1) *Segmentation Based on ACT model:* Temporal segmentation is the procedure to estimate the boundaries of four perceptually important regions like attack, transition, sustain, and release [17]. It is commonly known as the ATSR regions. The Amplitude/Centroid Trajectory(ACT) model is used to automatically segment the temporal evolution of musical instrument sounds and estimate the boundaries (1-5) of the regions (A,T,S,R). The ACT model exploits both the temporal envelope and the spectral centroid for this purpose which is represented as solid line and dashed line respectively in the Fig. 2.

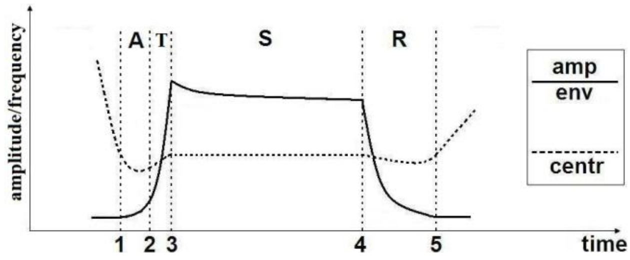

Fig. 2. Amplitude/Centroid Trajectory (ACT) model

2) *Alignment using DTW:* After segmentation both the musical instrument sounds has to be aligned with one another. The beginning of the note will be characterized by fast transients and the sustain region is much more stable due to absence of fast changes. When a sound with long attack time is combined with another sound of short attack time without prior temporal alignment, the result would not sound natural. Only those sounds which are synchronized at their A,T,S and R regions are capable of producing a perceptually seamless morph.

In order to find out which features of the first sound correspond to any particular feature of the second alignment is a necessary step. Dynamic Time Warping (DTW) is used to find the best temporal match between two sounds [21]. Dynamic time warping (DTW) is a well-known technique used to find an optimal alignment between two given time-dependent sequences under certain restrictions. To match each other, the sequences are warped in a nonlinear fashion. DTW algorithm helps in measuring similarity between two time or speed varying time series. Dynamic time warping (DTW) aims at aligning two sequences of feature vectors by iteratively warping the time axis until an optimal match between them is found.

3) *Envelope Estimation using TAE method:* The musical instrument sounds have amplitude modulations as their important perceptual cues. The amplitude envelope is an

outline of the waveform which connects the main peaks and avoids over fitting. The temporal envelope estimation can be performed with the True Amplitude Envelope (TAE) technique which is based on cepstral smoothing [18]. A curve which follows the general shape of the waveform without representing the harmonic structure is ideally known as the amplitude envelope. This curve should be smooth during stable regions of the waveform and should react to sudden changes. So amplitude envelope is expected to match the amplitude peaks corresponding to the period of the waveform. The main idea of TAE is that the structure of the spectrum is mimicked with the time-domain signal. The basic steps to estimate the TAE are as follows.

- Inorder to avoid negative amplitudes, the rectified version of the waveform is obtained
- This rectified waveform is zero-padded to the nearest power of two. It is similar to mimicking the DFT
- A time-reversed version of the zero-padded rectified waveform is added to represent the negative frequencies
- Obtain the true amplitude envelope (TAE)which represents a solid line outlining the rectified waveform

II. Spectral Processing

After temporal processing both the source and target sounds are subjected separately for spectral processing. In spectral processing the sound has to be decomposed into its sinusoidal and residual components, which are modeled independently as source and filter.

1) Sine and Residual Decomposition: A sound model assumes certain characteristics of the sound waveform or the sound generation mechanism. The sounds produced by musical instruments, or by any physical system, can be modeled as the sum of a set of sinusoids plus a noise residual [22]. The sinusoidal component of musical instrument sounds contain most of the acoustic energy present in the signal as they are designed to have very steady and clear modes of vibration. Each sinusoid models a narrow band component of the original sound and is described by an amplitude and a frequency function. The residual component which contains mostly noisy modulations is obtained by subtraction of the sinusoidal component from the original signal. A stochastic, or noise, signal is fully described by its power spectral density which gives the expected signal power versus frequency.

2) Source-Filter Model: The source-filter (SF) model represents both the sinusoidal and residual components as source and filter independently. The sinusoidal component constitutes the time-varying frequency values for the partials as source driving a time-varying spectral envelope filter. The residual component comprises a white noise (source time-varying spectral envelope (filter). The time-varying transfer function of the filter [23] can be written as

$$H(f,t) = |H(f,t)| \exp[j\Psi(f,t)] \tag{8}$$

where $|H(f,t)|$ and $\Psi(f,t)$ are respectively the amplitude and phase of the system. The musical instrument sound processing is done on a frame-by-frame basis. Inside each frame, the filter H(f,t) is considered linear shift-invariant (LSI). The output of the system is the convolution of the impulse response of the LSI filter and the excitation signal as

$$y(t) = x(t) * h(t) = [x_s(t) + x_r(t)] * h(t) = y_s(t) + y_r(t) \tag{9}$$

So it is evident that the filter response $h_s(t)$ is estimated as the spectral envelope of the sinusoidal spectrum $Y_s(f)$. The True Envelope (TE) method which minimizes spectrum peak estimation error is chosen to estimate the spectral envelope curve of $Y_s(f)$ as it is interpreted as the best band limited interpolation of the spectral peaks. The partials are the frequency values at which the spectral envelope curve is sampled. The residual signal $y_r(t)$ is modeled as a white noise source $x_r(t)$ driving the response of the system. The response of the resonant cavity to the excitation is modeled as the spectral envelope of using linear prediction. The SF residual is mixed into the SF sinusoidal after re-synthesis.

III. Morphing Procedure

In the spectral domain each frame is morphed separately. The morphed spectral frames are modulated by the morphed temporal envelope upon re-synthesis. For each frame, the morphed spectral envelope gives the amplitude of each partial at the value of the interpolated frequencies.

4) Spectral Envelope Morphing: The spectral energy is concentrated at the frequency regions where peaks of the spectral envelope are present. The peaks of the spectral envelope must be shifted in frequency for this technique. The balance of spectral energy should gradually shift from source to target when the spectral envelope morph is perceived linearly. For morphing between smooth spectral magnitudes envelopes of sound a method based on the notion of audio flow [24] is used. Following the morphing by feature interpolation principle [25], the objective of the spectral envelope morphing step is to obtain a morphed spectral envelope that has intermediate formant peaks and intermediate values of spectral shape features. The idea is to interpolate the spectral feature values and invert this representation to obtain spectral envelope parameters corresponding to the interpolated feature values.

1) Sine and Residual Morph Synthesis: The sinusoidal component of the spectral morph, results from the magnitude and frequency trajectories, or their transformation through additive synthesis [22]. This can either be implemented in the time domain with the traditional oscillator bank method or in the frequency domain using the inverse-FFT approach. The synthesized stochastic signal is the result of generating a noise

 ISSN: 2321-8169

signal with the time-varying spectral shape obtained in the analysis. A time varying filtering of white noise can be implemented using the time domain convolution of white noise with the impulse response corresponding to the spectral envelope of the frame. In the frequency domain a complex spectrum is created for every spectral envelope of the residual and an inverse-FFT is performed.

2) *Interpolation of Frequencies of Partials:* The source in the SF model which are the frequencies of the partials, carry perceptually important information in the form of temporal frequency modulations. For the morphing of musical instrument sounds, there exists a direct one to one correspondence between the partials of both sounds. This can be achieved by interpolating the interval ζ in cents between frequency f_{n1} and frequency f_{n2} as

$$\zeta = 1200\log_2(f_{n1}/ f_{n2}) \qquad (10)$$

where f_{n1} represents the frequency value of the n^{th} partial of the first sound, and f_{n2} the frequency value of the n^{th} partial of the second sound. Matching of the partial number will be enough for near harmonic musical instrument sounds. If one sound has more partials than the other, then the unmatched partials are discarded. A harmonic estimate of the unmatched partial f_n based on the fundamental frequency f_1 and the harmonic number n as $f_n = n\, f_1$ can be used if both sounds are nearly harmonic.

3) *Temporal Envelope Morphing:* Morphing the amplitude envelope is similar to the spectral envelope because the techniques for estimating the amplitude envelope are inspired by spectral envelope estimation techniques. Also, the temporal centroid is the time-domain analogous of the spectral centroid and its values behave in the same fashion under the same transformations. The temporal envelope morphing techniques considered are interpolation of the envelope curve (ENV) directly and interpolation of the cepstral coefficients (CC) used to represent it. The morphing techniques that shift peaks of the envelope is discarded because this behavior is undesirable for the temporal envelope.

IV. EVALUATION FOR LINEARITY IN MORPHING

The musical instrument sound morphing aims at creating an auditory illusion which gradually blurs the distinction between the source and target sounds by transforming across timbre dimensions. The morph has to be controlled both on the algorithmic and perceptual levels with a coefficient α. Controlling the morph with this single coefficient called morphing or interpolation factor is a challenging task.

Linearity is required in both the temporal and spectral shape feature domains. Evaluation is done on the variation of temporal centroid and spectral shape features. The deviation between the calculated feature values $\delta(\alpha_m)$ represented as "o" and the interpolated feature values α_m represented as "x" for each normalized feature $\delta_i(\alpha)$ gives the feature interpolation error illustrated in Fig. 3. The interpolated feature values are obtained as a linear regression by connecting the calculated feature values for the source and target with a straight line. The condition $\delta_i(\alpha) = \alpha$ holds for the interpolated features as all the features are normalized between 0 and 1.

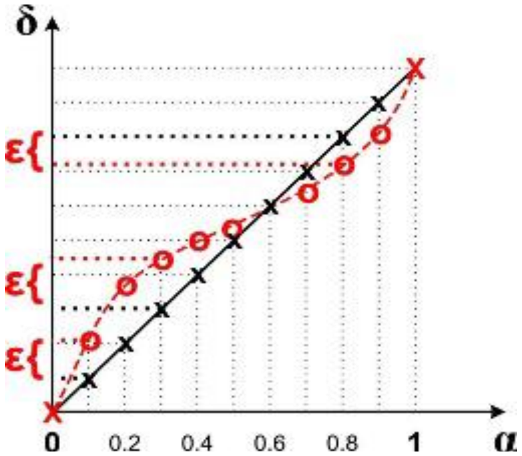

Fig. 3. Error calculation for interpolated and calculated feature values

The error function $\varepsilon(\delta_i)$ is defined as

$$\varepsilon(\delta_i) = \sqrt{\Sigma_m\, \varepsilon_m^{\,2}} = \sqrt{\Sigma_m(\delta(\alpha_m)\text{-}\alpha_m^{\,2})} \qquad (11)$$

where the square root of the sum of the quadratic deviations $\varepsilon_m^{\,2}$ between the calculated feature values ($\delta(\alpha_m)$ and the interpolated feature values α_m for each normalized feature $\delta_i(\alpha)$, where M is the number of linear steps between α_1= 0 and α_M=0 and the subscript i represents each temporal or spectral shape feature. For a given pair of sounds, for all considered temporal and spectral envelope representations the error $\varepsilon(\delta_i)$ is evaluated for each feature i. This error is then averaged across features to obtain an error estimation for each temporal and spectral envelope morphing method.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

As the first phase of my project the methodology has been completed and the hybrid sound through morphing is generated. The temporal and spectral shape features for the source and target sounds is found out. The intermediate sounds exhibit intermediate features.

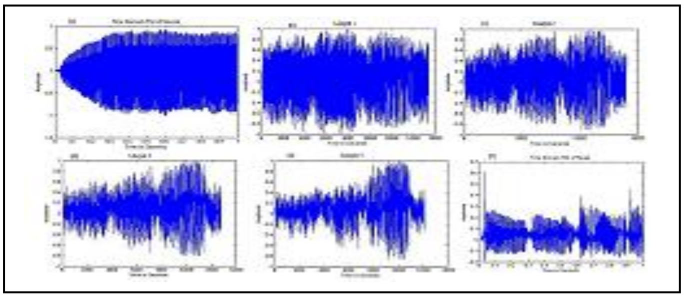

Fig. 4 The [a] source, [b.c.d.e] morphed and target sounds

VI. CONCLUSION

For an effective hybrid musical instrument sound generation three steps are required. The temporal and spectral shape features of the sound are extracted. In time domain the feature extraction is direct but spectral features are extracted for each segment of the signal. In temporal processing the input sounds are segmented using the ACT model. To align the boundaries of these sounds to the same stable or transient regions Dynamic Time warping is used. Alignment is required so that the morphed sound does not contain any howling or void space due to combination of different regions. The temporal envelope of the sound is estimated using the TAE estimation method. The envelope formed approximates the amplitude when compared to the other envelope estimation methods. For the spectral processing the both sounds are analyzed as a combination of sinusoidal and residual components. These two components are modeled into a source and filter separately and are spectrally processed. These form the input signal for spectral morphing procedure.

VII. REFERENCES

[1] Marcelo Caetano, Xavier Rodet, "Musical Instrument Sound Morphing Guided by Perceptually Motivated Features", *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 21, August 2013.

[2] Jeremy F. Alm, James S. Walker, "Time-Frequency Analysis of Musical Instruments", *SIAM Review, Society for Industrial and Applied Mathematics*, 2002, Vol. 44, pp. 457-476.

[3] A. Zlatintsi and P. Maragos, "Multiscale Fractal Analysis of Musical Instrument Signals With Application to Recognition", *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 21, April 2013.

[4] David L. Wessel, "*Timbre Space as a Musical Control Structure*", Computer Music Journal, Volume 3, 1980.

[5] Mauricio Loureiro, Hugo Paula, Hani Yehia, "Timbre Classification of a Single Musical Instrument", *Centre for Research on Speech, Acoustics Language and Music*, 2001.

[6] Thomas Lysaght, Diarmuid O'Donoghue, David Vernon, "Timbre morphing using The Wigner Time-Frequency Distribution", *National University of Ireland*, 1997.

[7] Wasim Ahmad, Huseyin Hacıhabiboglu and Ahmet M. Kondoz, "Perceptual effects of spectral modifications on musical timbres", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009.

[8] Hiroko Terasawa, Jonathan Berger, Shoji Makino, "In Search of a Perceptual Metric for Timbre: Dissimilarity Judgments among Synthetic Sounds with MFCC-Derived Spectral Envelopes", *J. Audio Eng. Soc.*, Vol. 60, 2012 September.

[9] Malcolm Slaney, Michele Covell and Bud Lassiter, "Automatic Audio Morphing", *International Conference on Acoustics ,Speech, and Signal Processing*, Atlanta, 1996.

[10] Thomas Lysaght, Diarmuid O'Donoghue, David Vernon, "Timbre Morphing Using The Wigner Time-Frequency Distribution", *National University of Ireland*, June 1997.

[11] Naotoshi Osaka, "Timbre Interpolation of Sounds Using a Sinusoidal Model", *ICMC Proceedings*, 1995.

[12] D.N.Zotkin, S.A.Shamma, R.Duraiswami, L.S.Davis, "Pitch and Timbre Manipulations Using Cortical Representation of Sound", *Perceptual Interfaces and Reality Laboratory, UMIACS*, 2004.

[13] John Grey, John Gordon, "Morphing of Transient Sounds Based on Shift Invariant Discrete Wavelet Transform and Singular Value Decomposition", *J. Acoust. Soc. America*, May 1978.

[14] Marcelo Caetano, Xavier Rodet, "Automatic Timbral Morphing of Musical Instrument Sounds by High Level Descriptors", *Analysis/synthesis Team, IRCAM*, 2010.

[15] Christopher Raphael, "Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 21, April 1999.

[16] Marcelo Caetano, Juan Burred, Xavier Rodet, "Automatic Segmentation of The Temporal Evolution of Isolated Acoustic Musical Instrument Sounds Using Spectro Temporal Cues", *Proc. of the 13th Int. Conference on DAFx, September*, 2010.

[17] Marcelo Caetano, Xavier Rodet, "Improved Estimation of Time Domain Signals Using True Envelope Cepstral Smoothing", Analysis/synthesis Team, IRCAM, 2011.

[18] Mathieu Lagrange, Sylvain Marchand and Jean-Bernard Rault "Using Linear Prediction To Enhance The Tracking of Partials", *France Telecom,University Bordeaux*, 2004.

[19] Giovanni De Poli and Federico Avanzini, "Sound Modeling: Signal Based Approaches", *Algorithms for Sound and Music Computing* , October 30, 2009.

[20] G.Peeters, "A Large Set of Audio Features for Sound Description", *CUIDADO Project*, 2004.

[21] R.B.Shinde, V.P.Pawar, "Dynamic time Warping using MATLAB and PRAAT", *International Journal of Scientific & Engineering Research*, Volume 5, May-2014.

[22] Xavier Serra, "Musical Sound Modeling with Sinusoids plus Noise", *Musical Signal Processing*, 1997.

[23] Marcelo Caetano, Xavier Rodet, "A Source Filter Model for Musical Instrument Sound Transformation", ICASSP, 2012.

[24] Tony Ezzat, Ethan Meyers, Jim Glass and Tomaso Poggio, "Morphing Spectral Envelopes Using Audio Flow", *Center for Biological and Computational Learning*, 2003.

[25] Marcelo Caetano, Xavier Rodet, "Sound Morphing by Feature Interpolation", *Analysis/synthesis Team, IRCAM*, 2011.