Evolutionary Spectral Envelope Morphing by Spectral Shape Descriptors

Marcelo Caetano, Xavier Rodet

EVOLUTIONARY SPECTRAL ENVELOPE MORPHING BY SPECTRAL SHAPE DESCRIPTORS

Marcelo Caetano, Xavier Rodet
IRCAM, Analysis-Synthesis Team
{caetano,rodet}@ircam.fr

ABSTRACT

There has been a great collective effort in the search for perceptually meaningful sound transformation techniques. The transformation of sounds matching target sound descriptors is a promising candidate because the descriptors are thought to capture timbral dimensions corresponding to relevant perceptual features. However, matching the descriptors alone is not enough because there are a large number of perceptually different sounds with the same values of descriptors. In this work, we use evolutionary computation to search for the spectral envelope variation that best matches the target spectral shape descriptors. We were able to achieve a more independent control of the descriptors while preserving the overall perceptual features.

1. INTRODUCTION

Sound transformations are ubiquitous in a wide range of applications so there is a growing interest in models and techniques that enable us to independently control perceptual features as accurately as possible. Timbral transformations figure prominently among the most challenging because timbre and its relations to the parameters of most models are not yet very well understood, but timbre is known to be related to the spectral envelope shape. The music information retrieval community has been working to find good sound descriptors for retrieval and classification of audio files in large databases and spectral shape descriptors figure prominently among those [10]. However, the inverse task of synthesizing a sound that matches certain values of sound descriptors defined a priori is much more difficult.

There are many proposed approaches to attaining timbral transformations, depending on the choice of the sound model and the objectives. In general, the primary goal is to obtain sounds that contain specific perceptual features. One popular solution is to use a technique to tune the parameters of the model to produce a sound similar to a given target. Yee-King [11] uses a genetic algorithm to automatically find the settings of an FM synthesizer that produce an output sound similar to a given target. Similarity is an MFCC based measure, such that they do not compare spectral shape descriptors so it is not possible to just change one dimension/feature while trying to preserve the others. Another popular approach is to use a technique that performs a mapping from the sound model parameters to the desired perceptual features. Le Groux [4] presents a support-vector machine based system that maps the parameters of additive synthesis (after dimensionality reduction with PCA) to sound descriptors. However, the only features they control are pitch and loudness. Hoffman [3] presents a general theoretical framework for synthesizing sounds that match arbitrary sets of perceptually motivated sound descriptors. Although they pose the problem clearly, they only present very preliminary results. Park et al. introduced the concept of feature modulation synthesis (FMS) [7], which proposes modulating signal features that are related to sound descriptors so they are capable of varying the values of the descriptors, but the transformations are not independent, such that modulating one feature also changes the others in unexpected ways. Mintz [6] proposes the use of constrained linear optimization to find the sound whose descriptors are the closest to target values. Nevertheless, all the conditions have to be linear, so the spectral shape descriptors become a distorted measure. Instead of using a transformation domain that allows direct modification according to a target, Coleman [1] derives analytic relations for modelling parametric transformations with respect to target descriptors. The transformations are limited to resampling and bandpass equalization using the spectral centroid and spread, the inclusion of skewness and kurtosis are referred to as future work.

In this work we focus on morphing spectral envelopes governed by the spectral centroid, spread, skewness, kurtosis and slope [8], which allow an approximation of underlying perceptual dimensions [9]. In the present effort we aim at independently controlling the descriptors. The next section presents two simple techniques for interpolating the spectral parameters of sound models and shows the corresponding behaviour of the spectral shape descriptors. The main motivation for our approach is to improve these results in the feature space by using a genetic algorithm to search for a spectral envelope that matches more closely the target spectral shape descriptors. Next we describe the experiment followed by an evaluation of the results and the conclusions.

2. INTERPOLATION AND MORPHING BY DESCRIPTORS

In this work we avoid simply interpolating the parameters of a model regardless of the impact on the perceptual features of the result. We define sound morphing to be a transformation between source timbres governed by a factor α ranging from 0 to 1, in which the original sources are obtained at the extremes, and we operate in the feature space, which are the spectral shape descriptors. Using traditional additive models, we can interpolate the time-varying frequencies and amplitudes of corresponding partials in the source sounds. Throughout this article, we shall refer to this method as the naïve interpolation method, as opposed to morphing perceptually related parameters of spectral envelope models. It is not straightforward to obtain a certain envelope shape that matches target values of spectral shape descriptors, so we measure the success of the results as how closely they match the target values calculated as interpolations of the descriptor values of two extremes. In all examples presented hereafter, the bassoon sound is always the first extreme and the clarinet sound is always the second.

Figure 1 shows the naïve interpolation method, and the interpolation of linear predictive coding (LPC) coefficients [5] to exemplify the effect of varying the interpolation factor linearly from 1 to 0 on both the envelope shape and the descriptor values. Comparing them, we readily see that the peaks of the spectral envelope of the hybrid versions on top of Figure 1 do not shift from one frequency region to the other, instead, they just increase or decrease following the interpolation factor. At the bottom, however, the behavior of the peaks of the hybrid versions follows a more natural and smooth path. The peaks in the spectrum represent frequency regions with more energy, and the

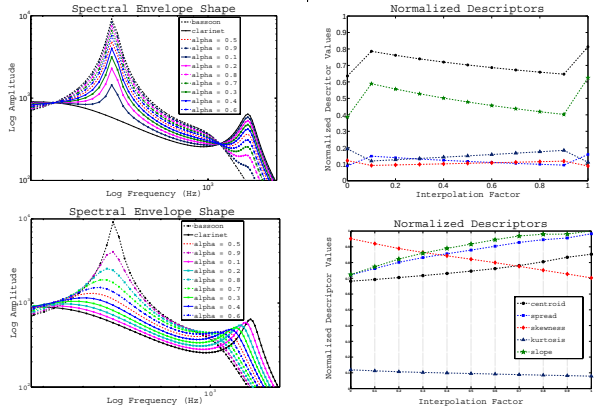

Figure 1. Top: Spectral envelopes (left) and spectral shape descriptor values (right) for the naïve interpolation method with linearly varying interpolation factor. Bottom: Spectral envelope (left). and spectral shape descriptor values (right) for the interpolation of the LPC coefficients with linearly varying interpolation factor.

Descriptor	*Target*	*Naïve*	*LPC*
Centroid	724	22	-4
Spread(10^{+6})	1.248	0.077	-0.138
Skewness	10	0	1
Kurtosis	153	2	24
Slope(10^{-12})	-5.06	-0.28	0.79

Table 1. Target ($\alpha = 0.5$) and differential spectral shape descriptor values for the naïve and LPC interpolation methods.

spectral centroid, for example, is usually attracted by these high energy regions. We should also notice that at the top of Figure 1 the variation is fairly counterintuitive, while at the bottom the descriptor values either increase or decrease as expected as the interpolation factor varies linearly. We want to verify if a specific factor α for either method generates descriptor values that also correspond to that factor, while preserving the desired hybrid envelope shape. Table 1 shows the values of the target and the difference between target and resultant descriptors for both methods for $\alpha = 0.5$. The aim of this work is not only to improve these values by matching them more closely, but also to allow independent control of the morphing factor for each descriptor investigated.

3. EVOLUTIONARY SPECTRAL ENVELOPE MORPHING BY SPECTRAL SHAPE DESCRIPTORS

We control the parameters of the LCP spectral envelope model independently with a genetic algorithm (GA). Since manipulating the LPC coefficients directly can lead to poles outside the unit circle (unstable filters), we manipulate the poles instead. On the right of Figure 3 we see that the poles of the first envelope drift to corresponding poles of the second as the interpolation factor varies from 0 to 1 when we interpolate the LPC coefficients. We refer to these poles in different positions of the interpolation path as hybrid poles and we are looking for a combination of hybrid poles that corresponds to an envelope model that matches the target descriptor values more closely while preserving the desired overall envelope shape.

3.1. Genetic Algorithms

A GA explores complex search spaces by codifying the parameters of a model into a chromosome-like structure where each individual corresponds to a point in the parameter space. The resulting search space contains the candidate solutions, and the evolutionary operators implement exploration and exploitation of the search space aiming at finding quasi-global optima. The GA iteratively manipulates populations of individuals at a given

generation by means of the simple genetic operations of crossover, mutation and selection [2].

We initialize the population of candidate solutions by interpolating the LPC coefficients of the spectral envelopes of the extremes with a factor α varying linearly from 0 to 1, as shown at the bottom of Figure 1. Each individual consists of a chromosome that contains the hybrid poles at one point of the path, as depicted at the top of Figure 2. Crossover consists of selecting two parent individuals, the crossover points, and swapping the chromosome segments (represented by different shades in Figure 2) between them, thus generating two offspring. We mate each individual of the current population with one randomly chosen partner (uniform distribution) using a one-point crossover operator with a uniform distribution [2]. Both offspring are inserted in the population and the parents are also kept. The mutation operation, applied to all individuals in this increased population, is depicted at the bottom of Figure 2 and consists of randomly (uniform distribution) choosing a mutation point, represented in black in the figure, and replacing the hybrid pole in that position by a new interpolated value with a factor α randomly (uniform distribution) chosen between 0 and 1. We measure the fitness of all individuals in the current population using the fitness function in Equation 1, the absolute value of the difference between target descriptors (T) and the descriptor values calculated for each individual in the current generation (c), normalized by the target value. Selection is done by sorting the individuals of the population of the current generation by increasing values of fitness and selecting N for the next generation.

$$ff = \sum_i \left| \frac{T_i - c_i}{T_i} \right| \qquad (1)$$

4. EXPERIMENT AND RESULTS

The experiment was designed to show that we can obtain evolutionary hybrid envelopes that match closely the target descriptor values while retaining the desired envelope shape, even if we use independent morphing factors for the

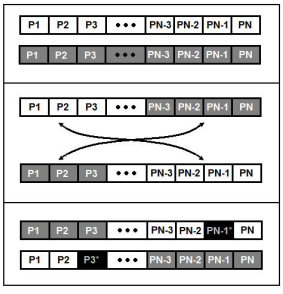

Figure 2. Chromosomes at the top, the crossover operation in the middle and mutation operation at the bottom.

descriptors. We generated evolutionary hybrid spectral envelopes for different values of the same morphing factor applied to all descriptors or an independent factor for each descriptor. Table 2 shows the target descriptor values on the left and the difference between target and result on the right for each morphing factor α indicated. Mixed factor means that an independent factor α was set for each descriptor. Figure 3 shows all the spectral envelopes resulting from the experiment on the left. If we compare the bottom of Figures 1 and 3 we see that all evolutionary hybrid envelope shapes keep the somewhat desired smooth transitions of peaks evidenced earlier. This means that, in general, manipulating the hybrid poles does not lead to unexpected envelope shapes. However, the two envelopes marked mixed alpha in Figure 3 present some interesting characteristics that deserve to be mentioned. For example, one of them presents a slightly higher peak than the corresponding one in the original envelope to accommodate the independence of the morphing factors. This is not possible to achieve with either method presented in section 2.

The fitness values measured for all the results ranged from ff = 0.01 to ff = 0.1, with varying degrees of accuracy. We do not control the individual accuracy of descriptors, so because they all have different ranges, the precision of matching the centroid is different from the spread, etc. The connection between envelope shape and descriptor values is an important part of the validation of the method because we are assuming that the closer this relation, the more perceptually meaningful the results. We can see in Table 2 that the descriptors whose ranges are smaller tend to be matched with greater precision. This problem arises because we use the same weight for all descriptors in the fitness function, so we could even out the importance of each descriptor in the final result with different weights. It is impossible to compare the results numerically with other approaches because there are no published results. We can, however, compare the column that corresponds to α = 0.5 in Table 1 with the corresponding column in Table 2 to see if the descriptor values obtained with the GA are closer than those obtained with simple LPC coefficient interpolation.

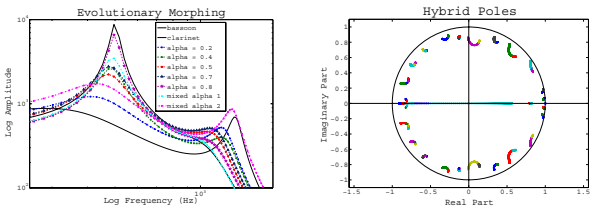

Figure 3. Evolutionary spectral envelopes for different target values of spectral shape descriptors (left) and hybrid poles resulting from the interpolation of the LPC coefficients with linearly varying interpolation factor (right).

Descriptor	*α = 0.2*		*α = 0.4*		*α = 0.5*		*α = 0.8*		*Mixed α*		*Mixed α*	
Centroid	777	9	742	-14	724	2	671	3	688	0.26	795	1
Spread(10^{+6})	1.456	0	1.317	-0.01	1.248	0	1.040	0	1.387	-0.002	1.248	-0.001
Skewness	9	0	10	0	10	0	11	0	11	0.72	11.56	0.15
Kurtosis	128	0	144	0	152	0	177	0	128	-1	161	-1
Slope(10^{-12})	-5.76	-0.02	-5.30	0	-5.06	0	-4.35	0	-5.06	0	-4.11	0.01

Table 2. Target and differential spectral shape descriptor values for the evolutionary spectral envelope morphing.

The values were significantly improved with the application of the GA for this specific example. However, we would have to perform listening tests to verify how this difference reflects perceptually. The most important aspect of the results lies in the independent control of all the descriptors given that the relevance of the descriptors values is only relative and there is no scale at present with which to perform a deep quantitative analysis. All the sounds synthesized with the spectral envelopes resulting from the naïve and the LPC interpolation methods as well as with the evolutionary spectral envelope morphing by spectral shape descriptors technique can be heard on http://recherche.ircam.fr/anasyn/caetano/icmc2009.html.

5. CONCLUSION AND FUTURE PERSPECTIVES

We proposed a method to obtain perceptually meaningful sound transformations guided by descriptor values because of their relation to perceptual features of sounds. We focused on timbral morphing between two extreme spectral envelopes, and we used a GA to search for the variation that best matches target values of spectral shape descriptors that were set between those of the extremes. Our approach enabled us to match the descriptors with independent morphing factors. However, the perceptive impact of these results is yet to be tested.

Future perspectives of this work include using different models, such as MFCC or LSF because LPC coefficients and poles are too sensitive and do not quantize well. We also plan to conduct perceptive experiments to try and validate the method and even to investigate if there is a scale or even a Just Noticeable Difference (JND) for the spectral shape descriptors.

6. ACKNOWLEDGEMENTS

This work is supported by the Brazilian Governmental Research Agency CAPES (process 4082-05-2). The main author would like to thank Juan José Burred, who followed this work from its early stages and made insightful suggestions during the course of the development and Diemo Schwarz, who took an interest and provided me with an overview of the literature.

7. REFERENCES

[1] Coleman, G., Bonada, J. Sound Transformation by Descriptor Using an Analytic Domain. Proc. DAFx, 2008.

[2] Davis, L. "Handbook of Genetic Algorithms". New York:Van Nostraud Reinhold,1991.

[3] Hoffman, M., Cook, P. "Feature-based Synthesis: Mapping from Acoustic and Perceptual Features to Synthesis Parameters" Proc. ICMC, 2006.

[4] Le Groux, S.; Verschure, P. Perceptsynth: Mapping Perceptual Musical Features to Sound Synthesis Parameters, Proc. ICASSP, pp. 125-128, 2008.

[5] Makhoul, J. "Linear prediction: A tutorial review" *Proc. IEEE*, vol. 63, pp. 561-580, Apr. 1975.

[6] Mintz, D. "Toward Timbral Synthesis: a New Method for Synthesizing Sound Based on Timbre Description Schemes". Master's thesis, Univ. Cal, 2007.

[7] Park, T., Biguenet, J., Li, Z., Conner, R., Travis, S. Feature Modulation Synthesis, Proc. ICMC, 2007.

[8] Peeters, G. "A large set of audio features for sound description (similarity and classification) in the CUIDADO project". Project Report, 2004.

[9] Peeters, G., McAdams, S., Herrera, P. Instrument Sound Description in the Context of MPEG-7. Proc. ICMC, 2000.

[10] Tzanetakis, G., Cook, P. "Musical Genre Classification of Audio Signals" IEEE Trans. Speech Audio Processing, 10(5), July 2002.

[11] Yee-King, M., Roth, Synthbot: An Unsupervised Software Synthesizer Programmer, proc ICMC, 2008.