SOUND MORPHING BY AUDIO DESCRIPTORS AND PARAMETER INTERPOLATION

Savvas Kazazis
CIRMMT,
McGill University
Montreal, Canada
savvas.kazazis@mail.mcgill.ca

Philippe Depalle
CIRMMT,
McGill University
Montreal, Canada
depalle@music.mcgill.ca

Stephen McAdams
CIRMMT,
McGill University
Montreal, Canada
smc@music.mcgill.ca

ABSTRACT

We present a strategy for static morphing that relies on the sophisticated interpolation of the parameters of the signal model and the independent control of high-level audio features. The source and target signals are decomposed into deterministic, quasi-deterministic and stochastic parts, and are processed separately according to sinusoidal modeling and spectral envelope estimation. We gain further intuitive control over the morphing process by altering the interpolated spectrum according to target values of audio descriptors through an optimization process. The proposed approach leads to convincing morphing results in the case of sustained or percussive, harmonic and inharmonic sounds of possibly different durations.

1. INTRODUCTION AND RELATED WORK

Sound morphing plays an important role in many areas including sound design for compositional applications and video games, speech manipulation, and in generating stimuli with specific and controllable acoustic parameters that are used in psychoacoustic experiments [1, 2]. Despite the extensive literature on this topic, there is no consensus on a single definition of audio morphing, and an extensive discussion on different viewpoints can be found in [3]. In this paper we present a strategy for stationary morphing, as opposed to *dynamic morphing*, in which a source sound gets continuously transformed over time into a target sound. We consider *static morphing* as a process that hybridizes a source sound with target sounds, or target audio features, through the independent manipulation of acoustic parameters.

Additive synthesis is one of the most flexible techniques, and as such many morphing strategies rely on interpolating the parameters of a sinusoidal model [4, 5, 6, 7, 8]. Tellman et al. [4] first pair the partials of the two sounds by comparing their frequency ratios to the fundamental frequency, and afterwards they interpolate their frequency and amplitude values. They also time-scale the two sounds to morph between their tremolo and vibrato rates based on assumptions that usually do not hold in the case of most natural sounds. Osaka [5] first performs dynamic time warping (DTW), and then he finds partials' correspondences by dynamic programming. The residual is modeled with short partials and is morphed according to stochastic parameter interpolation with hypothesized distributions. Fitz et al. [6] estimate the parameters of the "Bandwidth Enhanced Model" [9] by reassigned spectrograms, and use morphing envelopes to control the evolution of the frequency, amplitude, bandwidth and noisiness of the morph. Haken et al. [7] use a similar technique to morph in real time between pre-analyzed sounds that are placed in a three-dimensional timbre control space. Boccardi and Drioli [8] use Gaussian Mixture Models (GMM) to morph only the partials' magnitudes, which are derived from Spectral Modeling Synthesis (SMS) [10]. According to Boccardi and Drioli, since the morphing is based only on magnitude transformations, the source and target signals should belong to the same instrument family.

Other morphing strategies rely on interpolating the parameters of a source-filter model. Slaney et al. [11] construct a multidimensional space that encodes spectral shape and fundamental frequency on orthogonal axes. Spectral shape is derived through Mel-Frequency Cepstral Coefficients (MFCC) and fundamental frequency by the residual spectrogram. The optimum temporal match between the source and target sounds is found using DTW based on MFCC distances. The smooth and pitch spectrograms are interpolated separately. Ezzat et al. [12] argue that interpolating the spectral envelopes by simple cross-fading, as in [11], does not account for proper formant shifting. They describe a method for finding correspondences between spectral envelopes so as to encode the formant shifting that occurs from a source to a target sound. The morphing is based on interpolating the warped versions of the two spectral envelopes, and morphing between the residuals is left for future work.

Other authors claim to control synthesis parameters or to morph according to perceptual dimensions by using high-level audio features. Hoffman and Cook [13] propose a general framework for feature-based synthesis according to an optimization scheme that maps synthesis parameters to target feature values. The results are very preliminary: the source sound consists of stationary sinusoids and noise that is spectrally shaped through MFCCs; the target features are limited to spectral centroid, spectral roll-off and fundamental frequency histograms. Park et al. [14] treat single features as modulation signals that are applied to a source sound. According to their proposed scheme, different features cannot be controlled independently and thus the combination of multiple target features leads to unpredictable results. Mintz [15] uses linear constrained optimization on audio descriptors to control the parameters of an additive-plus-noise synthesizer. Williams and Brookes [16] morph using SMS according to verbal attributes that correlate with audio descriptors and in [17] employ a similar technique to morph between prerecorded sounds and sounds captured in real time. Hikichi and Osaka [18] adjust the parameters of a physical model using the spectral centroid as a reference to morph between piano and guitar sounds, and Primavera et al. [19] focus on the importance of decay time when morphing between percussive sounds of the same family. Coleman and Bonada [20] derive analytic relations for the spectral centroid and standard deviation to control adaptive effects for resampling and band-pass equalization. Caetano and Rodet in [21] investigate spectral envelope representations, which lead to linearly varying values of audio descriptors when linearly interpolated according to a morphing factor, and in

[22] use optimization techniques based on genetic algorithms to obtain morphed spectral envelopes that approximate target audio descriptor values.

Other approaches rely strictly on the time domain [23] or on time-frequency representations [24, 25]. Röbel [23] models the signals as dynamical systems using neural networks and morphs by interpolating their corresponding attractors. According to the author, the attractors of the two sounds should be topologically equivalent for achieving a convincing morphing. Ahmad et al. [24] propose a scheme for morphing between transient and non-stationary signals using the discrete wavelet transform (DWT) along with singular value decomposition (SVD) for interpolating the wavelet coefficients. Olivero et al. [25] propose a sound morphing technique without making any presumptions about the nature of the signal or its underlying model. The technique relies on the interpolation of Gabor masks and its penalty-based version is shown to encompass typical cross-synthesis strategies used in computer music applications. Furthermore, the interpretation of one of the strategies in terms of Bregman divergences allows them to include constraints that force morphing intermediates to exhibit a predesigned temporal sequence of centroids. This approach works well only as long as there is overlapping energy between the sounds and in our opinion, certain presumptions about the nature of the signal are necessary for choosing an appropriate morphing strategy.

Table 1 shows a brief comparison between the above-presented methods that are applicable to static morphing and the current approach. In Section 2 we present an overview of our proposed approach. Section 3 describes in detail the morphing process based on parameter interpolation, and Section 4 presents the optimization scheme used for morphing based on higher-level audio features. In Section 5 we present our concluding remarks and future improvements of our method.

2. A HYBRID APPROACH TO SOUND MORPHING

The morphing scheme presented here requires a source sound, to which we apply timbral transformations according to a morphing factor "α" ($0 \leq \alpha \leq 1$), and a target. A value of $\alpha = 0$ corresponds to the source sound and a value of $\alpha = 1$ corresponds to the target sound. The target could consist only of specific audio descriptor values that are obtained according to a morphing factor α_{d} and applied to the source sound, or it could be a different sound from which we extract the audio descriptors that we want to morph accordingly, but we also interpolate between the spectrotemporal fine structures of the two according to a morphing factor α_{p}. Depending on their spectral content, the source and target sounds can be decomposed into three parts as in [5]: a deterministic part, which is related to harmonic and inharmonic qualities; a quasi-deterministic part, which is more related to transients and spectrotemporal irregularities; and a stochastic part, which is related to noise color. The deterministic and quasi-deterministic parts are estimated through sinusoidal modeling from which we obtain the time-varying frequencies, amplitudes and phases of the partials. The stochastic parts are derived by subtracting the deterministic and quasi-deterministic parts from the original signals [10] and are modeled by estimating their spectral envelopes.

In the next step, we compute the time-varying audio descriptors on each of the three parts and for each analysis frame. Audio descriptors that are applicable to the current approach are presented in detail in [26]. For the purposes of this study we have experimented with: spectral centroid and higher order statistical moments of the spectrum including the standard deviation (referred to as spectral spread), spectral skewness, and spectral kurtosis; spectral decrease; and spectral deviation, which is only computed on the deterministic part of the signal. Descriptors that are applicable exclusively to harmonic (or slightly inharmonic) signals, such as tristimulus values and the odd-to-even harmonic ratio, are also applicable. Natural sounds, however, rarely exhibit such well-defined properties, and thus such descriptors would be more suitable in the case of synthetic or simplified natural sounds. Once we calculate the descriptors of the source and target sounds we can compute intermediate values according to the morphing factor α_d, and we interpolate the model parameters of the deterministic, quasi-deterministic and stochastic parts separately. The intermediate values of audio descriptors are applied to the parameter-interpolated signals using the optimization scheme described in Section 4.

We chose to model differently the stochastic part, on the one hand, and the deterministic and quasi-deterministic parts, on the other hand, because not all sounds exhibit a strong formant structure. As such, spectral envelopes would be a poor estimation of the signal, unless they are estimated by the tracked partials, as in [10, 27, 28]. On the other hand, it is well known that if the signal is stochastic-only, sinusoidal modeling usually leads to artifacts and so a morphing scheme based exclusively on this model would degrade the sound quality. The separation into deterministic and quasi-deterministic parts is necessary for improving the estimation of partial-to-partial correspondences, as we discuss in Section 3.1.1. In the following we assume that the source and target sounds are equalized in loudness, have the same fundamental frequencies, and can be of different durations.

3. PARAMETER INTERPOLATION

In this section we describe the interpolation schemes based on the parameters of the sinusoidal model and the parameters that model the spectral envelopes of the residuals.

3.1. Deterministic and quasi-deterministic parts

The following scheme is used for both the harmonic and quasi-harmonic parts. Before interpolating the parameters of the sinusoidal model, it is necessary to find partial-to-partial correspondences between the source and target sounds.

3.1.1. Estimating partial-to-partial correspondences

The deterministic part consists of partials that are long in duration, with respect to the total duration of the analyzed sound, whereas the quasi-deterministic part consists of shorter partials that are generally unstable in frequency (short chirps), have lower amplitude values, and surround the harmonic or inharmonic partials of the deterministic part. Such partials may also occur as artifacts of the sinusoidal analysis algorithm, especially in cases where the sinusoids are of low amplitude and the tracking algorithm fails to perform a reliable peak-to-peak matching.

A one-to-one correspondence between the partials of the source and target sounds is very unlikely to occur unless we limit the number of tracked partials to the most prominent ones with respect to their durations and amplitude thresholds. However, there are

Table 1: *A brief comparison of methods for static morphing.*

Author(s) and papers	**Sound model & morphing strategy**	**Partial matching**	**High-level audio features**
Osaka [5]	Sinusoidal modeling. The residual is modeled with short partials according to hypothesized distributions.	Yes	No
Tellman et al. [4]	Sinusoidal modeling. No treatment of the residual.	Yes	No
Haken et al. [7]	Noise-enhanced sinusoidal modeling.	No	Amplitude and fundamental frequency
Boccardi and Drioli [8]	GMM applied to SMS. No treatment of the residual.	No	No
Caetano and Rodet [22]	Spectral envelopes for the deterministic and stochastic parts.	No	Spectral audio descriptors
Röbel [23]	Dynamical systems.	Not applicable	No
Ahmad et al. [24]	DWT with SVD.	Not applicable	No
Olivero et al. [25]	Gabor transform with constrained Gabor masks.	Not applicable	Arithmetic, harmonic and geometric centroids
Kazazis et al. [present document]	Sinusoidal modeling and spectral envelopes.	Yes	Spectral and harmonic audio descriptors

cases in which even if there is a limit to the number of tracked partials, the assumption of a one-to-one correspondence as described in [21] could be problematic. For example, when morphing from a sound that has odd and even harmonics to a sound that has only odd ones, we would ideally interpolate only the frequency and amplitude values of the odd harmonics of the two sounds to avoid the artifacts that would result from interpolating the odd with both the odd and even harmonics of the two sounds.

For finding correspondences between the partials of the source and target sounds, we use a k-nearest neighbors classifier (k-NN) based on Euclidean frequency proximity, and under the condition that the vector that is to be classified must have the same or a smaller number of partials. Obviously, the k-NN classifier does not return a one-to-one, but rather a many-to-one, mapping, so we choose the closest neighbor in frequency, and we treat the rest of the neighbors as unmatched partials. The unmatched partials retain their original frequencies but are initialized with zero amplitude levels, which gradually increase according to the morphing factor. After experimenting with different sounds, we concluded that such treatment does not lead to perceptual stream segregation, but rather to a seamless partial fade-in effect that facilitates the morphing between inharmonic sounds or between sounds that consist of unequal numbers of partials (see Fig. 1).

3.1.2. Interpolation of partials' breakpoint values

We represent each partial according to its start and end times, and with time breakpoints that are set according to its frequency and amplitude variations. If the source and target sounds have a different number of breakpoints, we simply interpolate the breakpoint values of the shorter one in order to match the number of breakpoints of the longer one. This representation enables us to interpolate the parameters at the level of events, which offers greater control over the morphing process as opposed to parameter interpolation between time frames. If the partials of the source and target sounds differ in duration, we are able to achieve intermediate durations by interpolating the breakpoint values of each partial according to the morphing factor. Interpolating between the start and end times of the partials also allows us to morph their onset asynchrony. We use the following expressions for calculating the interpolated values of partials' frequencies and amplitudes, respectively:

$$f(\alpha_\mathrm{p}) = \alpha_\mathrm{p} f_\mathrm{s} + (1-\alpha_\mathrm{p}) f_\mathrm{t} \tag{1}$$

$$\log_{10}(g(\alpha_\mathrm{p})) = \alpha_\mathrm{p} \log_{10}(g_\mathrm{s}) + (1-\alpha_\mathrm{p}) \log_{10}(g_\mathrm{t}) \tag{2}$$

where the subscripts "s", "t" denote the source and target, respectively, and α_p is the morphing factor related to parameter interpolation. Though Fig. 1 does not show a typical harmonic spectrum of the analyzed sounds because of the very low amplitude detection threshold ($-90\,\mathrm{dB}$) that was used in the partial-tracking algorithm, and which subsequently gave rise to auxiliary harmonic components, it clearly illustrates the estimation of partial-to-partial correspondences and the interpolation of the partials' breakpoint values.

3.2. Stochastic part

For morphing the stochastic part, we first estimate for every analysis frame its spectral envelope using Linear Predictive Coding (LPC), because we assume that the modeled signal is random, which fits exactly the basic assumption of LPC. We then get a temporal sequence of spectral envelopes (one for each frame), which allows us to render a time-varying Power Spectral Density (PSD) of the stochastic part. In order to morph, we interpolate for each time between the spectral envelope of the source and the target at this corresponding time. For a high-quality interpolation of the spectral envelopes, it is necessary to convert the LPC transverse coefficients to an alternative representation, because they do not interpolate well and might lead to unstable filters. Line Spectral Frequencies (LSF), Reflection Coefficients (RC) and Log Area Ratio (LAR) have been shown to interpolate smoothly, lead to stable intermediate filters, and lead to linear variations of audio descriptors

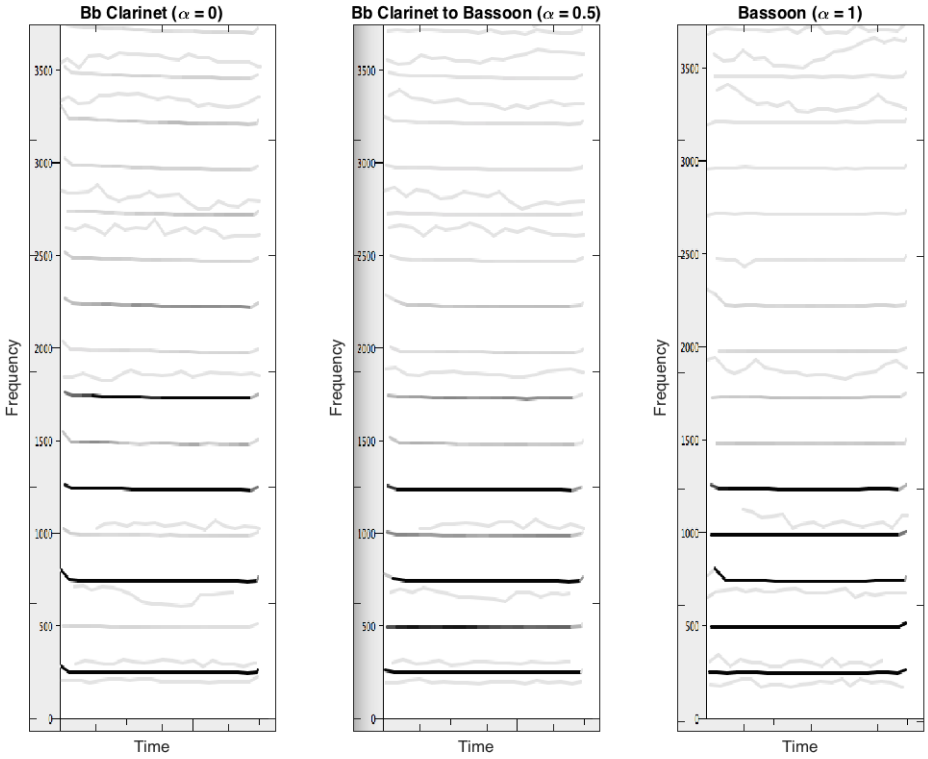

Figure 1: *Partial-to-partial correspondences and parameter interpolation of the deterministic part. Morphing from a clarinet sound to a bassoon with $\alpha_p = 0.5$. Gray-level values correspond to the partials' amplitude values.*

when linearly interpolated [21, 29]. We choose to interpolate the LAR coefficients (Eq. 3) as they both guarantee the filter's stability and have a physical interpretation, which could be specifically useful when trying to morph between sounds that were created by physical modeling synthesis as in [5, 2]. The filters' coefficients are interpolated according to Eq. (2).

$$\text{lar}(r_\alpha) = \alpha_\text{p}\text{lar}(r_\text{s}) + (1 - \alpha_\text{p})\text{lar}(r_\text{t}) \tag{3}$$

where lar is a vector the coefficients of which read:

$$\text{lar}(r)[i] = ln\left(\frac{1 - r(i)}{1 + r(i)}\right), \quad 1 \leq i \leq n \tag{4}$$

and n is the number of reflection coefficients r. The morphed residual is synthesized by filtered white noise after the inversion of the LAR coefficients to LPC coefficients.

3.3. Temporal Energy Envelope

In the present approach, the temporal energy envelope is a consequence of morphing. The parts of the signal that were morphed independently are added together to form the parameter-interpolated signal and thus, the energy envelope is constructed from the time-varying amplitudes of the partials and the gains of the filter.

4. FEATURE INTERPOLATION

The desired values of descriptors along with the interpolated spectrum form an underdetermined system because in theory there are an infinite number of sounds that have the same audio descriptor values. As previously described in Section 2, the target may consist only of target descriptor values D_a, in which case the morphing is based exclusively on high-level features. Fig. 2 shows an example of two sounds exchanging time-varying spectral centroids, where $\alpha_\text{p} = 0$, since the source is the Timpani without any parameter-based morphing, and $\alpha_\text{d} = 1$, because we apply to the source spectrum the spectral centroid values of the Tuba, which is the target. For each time frame, we match the audio descriptor values obtained according to a specific α_d to the interpolated spectrum by optimizing the amplitudes of the sinusoids or FFT bins of the interpolated spectrum x_j under the constraints of the target values of descriptors D_a. More formally this can be expressed as:

$$\min_x \sum_{j=1}^{N} |x_j - g_j| \quad \text{subject to} \quad D(x) = D_a \tag{5}$$

where g_j are the parameter-interpolated amplitude values according to α_p, N is the total number of partials or FFT bins, and D_a is the target value of $D(x)$, which can be one of the following descriptors (Eq. (6) – (11)).

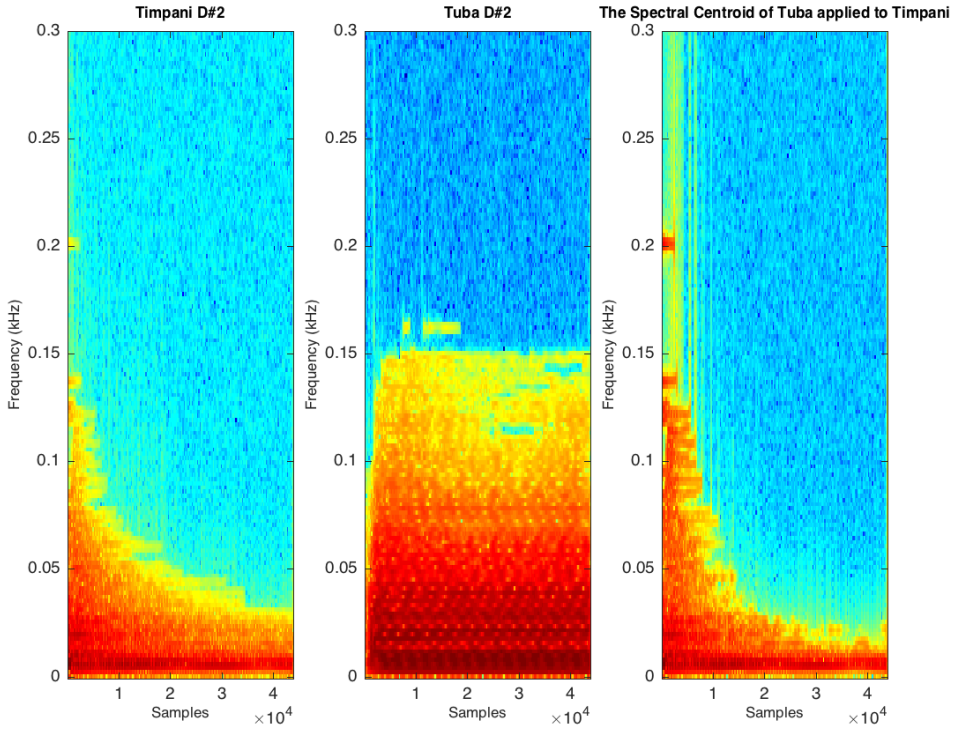

Figure 2: *The spectral centroid time series of a Tuba sound applied to a Timpani (the actual values of the time series are shown in Fig. 3.)*

$$m_1 = \sum_{j=1}^{N} f_j \cdot p_j \tag{6}$$

$$m_2 = (\sum_{j=1}^{N} (f_j - m_1)^2 \cdot p_j)^{1/2} \tag{7}$$

$$m_3 = (\sum_{j=1}^{N} (f_j - m_1)^3 \cdot p_j)/m_2^3 \tag{8}$$

$$m_4 = (\sum_{j=1}^{N} (f_j - m_1)^4 \cdot p_j)/m_2^4 \tag{9}$$

$$decr = \frac{1}{\sum_{j=2}^{N} x_j} \sum_{j=2}^{N} \frac{x_j - x_1}{j-1} \tag{10}$$

$$dev = \frac{1}{N} \sum_{j=1}^{N} (x_j - \mathrm{SE}(f_j)) \tag{11}$$

where p_j are the normalized values of x_j [27]:

$$p_j = \frac{x_j}{\sum_{j=1}^{N} x_j} \tag{12}$$

dev denotes the harmonic spectral deviation and $\mathrm{SE}(f_j)$ is the value of the spectral envelope at frequency f_j, which is estimated by averaging the values of three adjacent partials; $decr$ denotes the spectral decrease; m_1, m_2, m_3 and m_4 denote the spectral centroid, spectral spread, spectral skewness and spectral kurtosis respectively. The optimization is run in Matlab using the "fmincon" function along with the "sqp" method, which are suitable for solving constrained and non-linear problems [30]. Since the audio descriptors have different ranges, it is necessary to normalize them for assessing the convergence of the algorithm. Using this optimization scheme, we are able to set different morphing factors for each descriptor independently, as long as a feasible solution among these values exists. Furthermore, the choice of the objective function (Eq. 5) forces the optimized spectrum to be as close as possible to the interpolated one by keeping its frequency content unchanged and by altering its amplitude values as little as possible. Fig. 3 shows an example of morphing the parameter-interpolated signal according to varying morphing values of spectral centroid and spectral spread while preserving a constant value for the rest. Using a sinusoidal model for the deterministic and quasi-deterministic parts, the optimized values correspond directly to the parameters of additive synthesis, and the residual reaches its target values by altering the energy of the FFT bins. As in Section 3.1.2, if the source and target sounds are of different durations, we simply interpolate the descriptor values of the shorter one in order to match them to the analysis frames of the longer one.

5. CONCLUSIONS AND FUTURE WORK

We presented a hybrid approach to sound morphing based on sinusoid-plus-noise modeling and higher-level audio features. Dividing the signal into deterministic, quasi-deterministic, and stochastic parts and processing them separately allows for finer control of

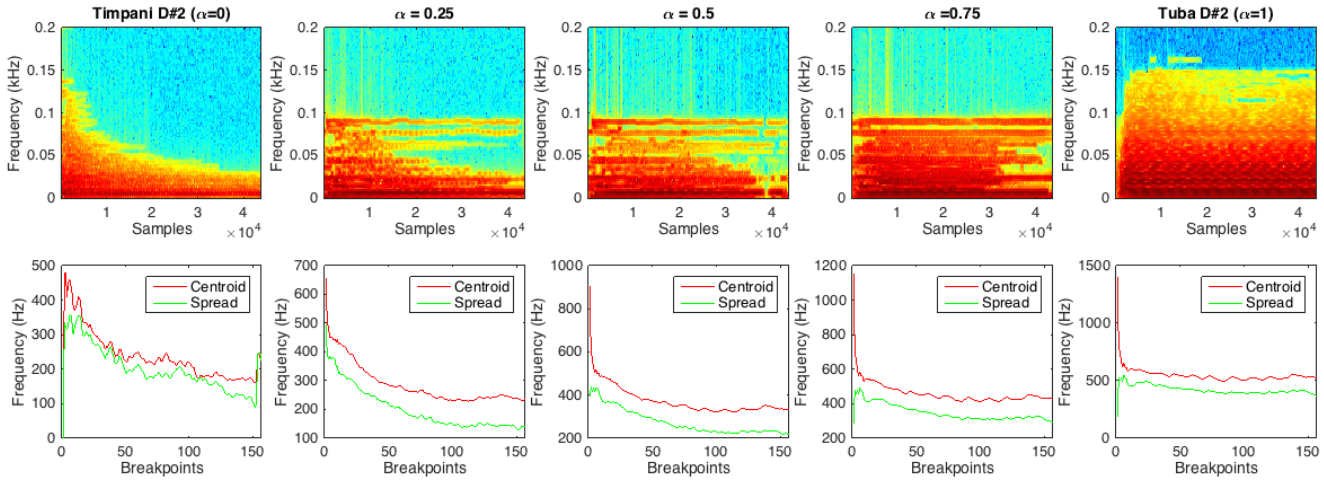

Figure 3: *Morphing the parameter interpolated signal by audio descriptors. Spectral centroid and spectral spread vary according to the morphing factor α. The rest of descriptors preserve constant target values according to their median when interpolated with $\alpha_d = 0.5$.*

the synthesis parameters and also enables us to morph between deterministic and quasi-deterministic signals of different durations. The morphed sound is synthesized using additive synthesis for the deterministic and quasi-deterministic parts and filtered white noise for the stochastic part. The spectrum of the morphed signal is further refined according to target audio descriptor values through an optimization process. We have shown that this process allows us to control accurately and independently several audio features, provided that a feasible solution among them exists. Audio examples are available at: https://www.mcgill.ca/mpcl/resources-0/supplementary-materials. The proposed scheme is more suitable for sustained and percussive sounds, which can either be harmonic or inharmonic, rather than textural sounds. Their residuals however, should be stationary (or pseudo-stationary) as opposed to sound texture, the residual of which is usually non-stationary and may consist of sharp transients. A refinement of our approach would be to find sophisticated ways to interpolate between different tremolo and vibrato rates while preserving the overall spectrotemporal complexity of the partials. Finally, we by no means claim that the use of high-level audio features enables a perceptually based sound morphing. Rather, it offers a more intuitive control over the morphing process, as in the case of adaptive effects [31]. Up to now only spectral centroid and log-attack time have been shown to be significantly correlated with perceptual dimensions, cf. [1, 32]. If and how such audio features collapse to single perceptual dimensions remains to be empirically determined.

6. ACKNOWLEDGMENTS

We would like to thank Philippe Esling for his advice on improving the computational efficiency of our optimization procedure, and the four anonymous reviewers for their helpful comments about improving the clarity of this document. This work was funded by Canadian Natural Sciences and Engineering Research Council grants awarded to Stephen McAdams and Philippe Depalle and a Canada Research Chair awarded to Stephen McAdams.

7. REFERENCES

[1] A. Caclin, S. McAdams, B.K. Smith, and S. Winsberg, "Acoustic Correlates of Timbre Space Dimensions: A Confirmatory Study Using Synthetic Tones," *J. Acoust. Soc. Am.,* 118 (1), pp. 471-482, 2005.

[2] S. Lakatos, P. Cook, and G. Scavone, "Selective attention to the parameters of a physically informed sonic model," *Acoustics Research Letters Online, J. Acoust. Soc. Am.,* March 2000.

[3] M. 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.

[4] E. Tellman, L. Haken, and B. Holloway, "Timbre Morphing of Sounds with Unequal Numbers of Features," *J. Audio Eng. Soc.,* vol. 43, no. 9, pp 678-689, September, 1995.

[5] N. Osaka, "Timbre Interpolation of Sounds Using a Sinusoidal Model," in *Proc. ICMC,* Banff Centre for the Arts, Canada, 1995.

[6] K. Fitz, L. Haken, S. Lefvert, C. Champion, and M. O'Donnell, "Cell-Utes and Flutter-Tongued Cats: Sound Morphing Using Loris and the Reassigned Bandwidth-Enhanced Model," *Computer Music Journal,* 27 (3), 2003.

[7] L. Haken, K. Fitz, and P. Christensen, *Sound of Music: Analysis, Synthesis, and Perception,* "Beyond Traditional Sampling Synthesis: Real-Time Timbre Morphing Using Additive Synthesis," J. W. Beauchamp Ed. Springer-Verlag, Berlin, 2006.

[8] F. Boccardi, and C. Drioli, " Sound Morphing with Gaussian Mixture Models," in *Proc. DAFx,* Limerick, Ireland, 2001.

[9] K. Fitz, L. Haken, and P. Christensen, "A New Algorithm for Bandwidth Association in Bandwidth Enhanced Additive Sound Modeling," in *Proc. ICMC*, Havana, Cuba, 2001.

[10] X. Serra and J. I. Smith, "Spectral Modelling Synthesis," *Computer Music Journal,* 14 (4),1990.

[11] M. Slaney, M. Covell, and B. Lassiter, "Automatic Audio Morphing," in *Proc. ICASSP,* Atlanta, Georgia, 1996.

[12] T. Ezzat, E. Meyers, J. Glass, and T. Poggio, "Morphing Spectral Envelopes using Audio Flow," in *Proc. ICASSP,* Philadelphia, Pennsylvania, 2005.

[13] M. Hoffman,and P. Cook, "Feature-based Synthesis: Mapping from Acoustic and Perceptual Features to Synthesis Parameters" in *Proc. ICMC,* New Orleans, USA, 2006.

[14] T. Park, J. Biguenet, Z. Li, R. Conner, and S. Travis, "Feature Modulation Synthesis," in *Proc. ICMC,* Copenhagen, Denmark, 2007.

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

[16] D. Williams, and T. Brookes, "Perceptually-Motivated Audio Morphing: Warmth," *AES 128th Convention,* London, UK, 2010.

[17] D.Williams, P. Randall-Page, E. R. Miranda, "Timbre morphing: near real-time hybrid synthesis in a musical installation," in *Proc. NIME,* London, UK, 2014.

[18] T. Hikichi and N. Osaka, "Sound Timbre Interpolation Based on Physical Modeling" *Acoustical Science and Technology,* 22 (2), 2001.

[19] A. Primavera, F. Piazza and J. D. Reiss, "Audio Morphing for Percussive Hybrid Sound Generation," *AES 45th Conference on Applications of Time Frequency Processing in Audio*, Helsinki, March 2012.

[20] G. Coleman, and J. Bonada, "Sound Transformation by Descriptor Using an Analytic Domain," in *Proc. DAFx,* Espoo, Finland, 2008.

[21] M. Caetano and X. Rodet, "Automatic Timbral Morphing of Musical Instrument Sounds by High-Level Descriptors," in *Proc. ICMC,* New York, USA, 2010.

[22] M. Caetano and X. Rodet, "Independent Manipulation of High-Level Spectral Envelope Shape Features for Sound Morphing by Means of Evolutionary Computation," in *Proc. DAFx,* Graz, Austria, 2010.

[23] A. Röbel, "Morphing Dynamical Sound Models," in *Proc. IEEE Workshop Neural Net Sig. Proc,* 1998.

[24] M. Ahmad, H. Hacihabiboglu, and A. M. Kondoz, "Morphing of Transient Sounds Based on Shift-Invariant Discrete Wavelet Transform and Singular Value Decomposition," in *Proc. ICASSP,* Taipei, Taiwan, 2009.

[25] A. Olivero, P. Depalle, B. Torresáni and R. Kronland-Martinet, "Sound Morphing Strategies Based on Alterations of Time-Frequency Representations by Gabor Multipliers", *AES 45th Conference, Applications of Time-Frequency Processing in Audio,* Helsinki, Finland, March 2012.

[26] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The Timbre Toolbox: Extracting Audio Descriptors from Musical Signals," *J. Acoust. Soc. Am.,* 130, pp. 2902-2916, 2011.

[27] A. Röbel, and X.Rodet, "Efficient Spectral Envelope Estimation and its Application to Pitch Shifting and Envelope Preservation," in *Proc. DAFx,* Madrid, Spain, 2005.

[28] T. Galas, and X. Rodet, "An Improved Cepstral Method for Deconvolution of Source-Filter Systems with Discrete Spectra: Application to Musical Sounds," in *Proc. ICMC,* Glasgow, Scotland, 1990.

[29] T. Islam, *Interpolation of Linear Prediction Coefficients for Speech Coding,* Master's thesis, McGill University, 2010.

[30] http://www.mathworks.com/help/optim/ug/constrained-nonlinear-optimization-algorithms.html#bsgppl4/

[31] V. Verfaille and P. Depalle, "Adaptive Effects based on STFT, using a Source-Filter model," in *Proc. DAFx*, Naples, Italy, 2004.

[32] C. W. Wun, A. Horner, and Bin. Wu, "Effect of Spectral Centroid Manipulation on Discrimination and Identification of Instrument Timbres," *J. Audio Eng. Soc.,* 62(9), 575-583, 2014.