

ADAPTIVE EFFECTS BASED ON STFT, USING A SOURCE-FILTER MODEL

Vincent Verfaillie, Philippe Depalle

Sound Processing and Control Laboratory
Faculty of Music – McGill University
555, Sherbrooke Street West
Montréal, Québec H3A 1E3, Canada
{vincent | depalle}@music.mcgill.ca

ABSTRACT

This paper takes the opportunity of presenting a set of new adaptive effects to propose a generic scheme for adaptive effects built upon a spectral source-filter decomposition and a Short-Time Fourier analysis-resynthesis. This allows for a better formalization of the involved signal processing algorithms and leads to a simple classification of adaptive effects already presented in the literature, that falls into this category.

We discuss the motivation and the advantages of combining source-filter modeling and phase vocoder representation for the design of adaptive digital audio effects. Then we detail the general structure that includes STFT analysis and re-synthesis scheme, the source filter decomposition, and an adaptive control unit composed of a feature extraction system and a sound mapping unit that might be driven by a gestural control section.

1. INTRODUCTION

Adaptive effects are audio processing systems which controls are derived from sound features, in order to take into account the evolution of their structural properties (often called “musical gestures” [1]). As a canonical example, cross-synthesis (as defined in [2]) uses the information from one signal to modify the spectral envelope of a second one. In this paper, we present a generic scheme for adaptive effects built upon a Short-Time Fourier analysis-resynthesis and a spectral source-filter decomposition.

During the 1930ies, H. Dudley invented the vocoder (Voice Operated reCORDER) at the Bell labs [3]. The vocoder consisted of analyzing a voice signal with a filter-bank and was designed for voice coding in telecommunication applications. Its musical use has been popularized by The Beatles’ song “Tomorrow Never Knows” in 1966 and by W. Carlos’ song “TimeSteps” in 1971 (soundtrack of “A Clockwork Orange”).

An improved version of the vocoder was introduced in the mid 1960ies (namely the phase vocoder) [4], then explicitly linked to the Short-Time Fourier Transform (STFT) [5] and has benefited for many optimization, such as in [6]. At the same time, analysis and synthesis techniques were developed in the same context of voice coding and synthesis, namely source-filter model and linear prediction coding (LPC) [7]. Both techniques allows for many transformations of sounds, among which interpolation, pitch-shifting with formant preservation and robotization.

Recently, more and more effects that takes into account sound features of the sound under process have been developed, such as morphing [2, 8, 9], voice morphing [10, 11], etc. The generic

structure we propose allows for a better formalization of the involved signal processing algorithms and leads to a simple classification of adaptive effects already presented in the literature, that falls into this category. After introducing the techniques and the general structure (Section 2), we present a classified review of usual and new adaptive source-filter effects (Section 3) that processes the spectral envelope, the source or both of them.

2. TECHNIQUES

2.1. Motivations

Based on STFT, the phase vocoder represents exhaustively any sound without any error, allows for filtering with very accurate frequency response, for time-scaling without pitch-shifting and reciprocally, as well as many other exotic transformations [12, 13]. This time-frequency representation can be viewed as a block-by-block processing system [4, 5].

The source-filter model can be viewed in two ways: a signal production model that directly acts in the time domain, and a spectral classification of sound properties. It sorts the slow frequency-varying part of the spectrum which is interpreted as the frequency response of a filter, while the fast frequency-varying features are considered to be the spectrum of the source. Contrarily to the STFT, the source-filter model is not a representation, *i.e.* most of the signals cannot be exhaustively represented without an error.

However, a combination of the strengths of both methods is possible by separating the source and the filter in the spectral domain (*e.g.* using the STFT with the cepstrum technique [14]). Then we can benefit from the abstract and efficient coding of the source-filter model while preserving the processed signal from any modeling error.

2.2. Adaptive Digital Audio Effects (A-DAFx)

We call “adaptive digital audio effects” the generalization of effects and their control [15, 16], in the context of “intelligent effects” [17] and “content-based transformations” [18]. There are two types of control (see Fig. 1)

- adaptive control which is a time-varying control derived from sound features (*i.e.* an analysis step) [17] and modified by appropriate mapping functions;
- gestural control for realtime access through input devices.

Several forms of adaptive effects exist, depending on the sound signal from which features are extracted [15, 16]. We name

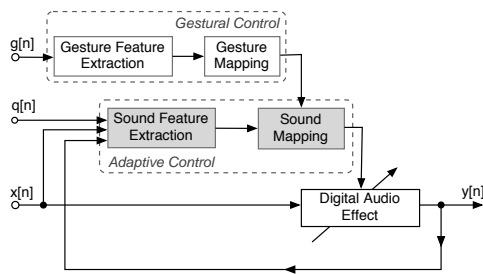


Figure 1: Diagram of the gesturally controlled adaptive effect.

- “adaptive” effects when features are extracted from $q[n]$;
- “auto-adaptive” effects when extracted from $x[n]$;
- “feedback¹ adaptive” effects when extracted from $y[n]$;
- “cross-adaptive” effects when two input signals are used.

The mapping between sound features and effect control values includes non-linearities as well as feature combinations [20]. The effect controls and their mapping can be modified by gestural control [21].

2.3. STFT

In order to describe the complete processing system, we briefly present the block processing structure (see Fig. 2) based on the STFT [4, 5]. The STFT $F_x[m, k]$ of a signal $x[n]$ is defined as

$$F_x[m, k] = \sum_{n=-\infty}^{\infty} x[n] \cdot w[mR_A - n] \cdot e^{-j2\pi nk/N}, \quad (1)$$

$$= \rho_x[m, k] \cdot e^{j\varphi_x[m, k]}, \quad (2)$$

where $w[n]$ is a window which length defines the block size, R_A is the step increment, m is the current time frame index, N is the number of spectral bins, indexed by $k = 0, \dots, N - 1$. The synthesis stage uses the classical overlap-add technique [22] with the same time increment, chosen to guaranty a perfect reconstruction of the signal [5].

2.4. Source-Filter Processing

In order to get a description of a sound signal in terms of a source-filter model, one need to deconvolve the signal by first estimating its spectral envelope $H_x[m, k]$. Depending on the chosen technique or the appropriate coding, H_x is going to be represented by a set of parameters² $P_{H_x}[m] = L_P^{-1}(H_x)[m, \bullet]$, that denotes either reflection coefficients, autoregressive coefficients, cepstral coefficients, correlation coefficients or formant coefficients. The filter estimation is achieved by various and well-known techniques such as LPC [7], cepstrum [14] or spectral breakpoint functions [23]. We then deduce the STFT of the source $S_x[m, k]$ (see Fig. 2) from

$$F_x[m, k] = H_x[m, k] \cdot S_x[m, k]. \quad (3)$$

¹The word “adaptive” is commonly used in this context, where the output signal is analyzed and used to minimize an error function, such as in adaptive filtering (telecommunications) [19].

²The notation $G[m, \bullet]$ stands for the frequency vector at time index m .

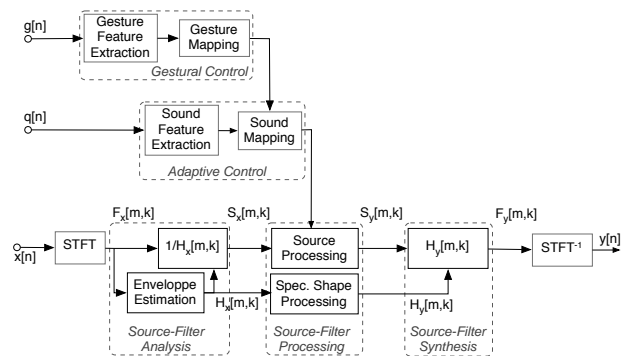


Figure 2: Diagram of adaptive source-filter block-by-block processing.

In case the adaptive control and the source-filter processing part of the A-DAFx use an identical analysis step and the same input signal, they can be factorized.

2.5. Useful Notations

The STFT of the output signal F_y is derived from input sources’ and filters’ STFTs by a given transformation noted T_F

$$F_y = T_F\{F_x, F_q\}. \quad (4)$$

When F_y is separable we consider the transform to be a source-filter one

$$F_y = T_S\{S_x, S_q\} \cdot T_H\{H_x, H_q\}. \quad (5)$$

Notice that in general, T_F , T_S or T_H might depend on features extracted from the signals x or q .

Usual effects often involve frequency shifting, scaling or warping of a time-frequency function G_α (i.e. F_α , S_α , or H_α), as well as multiplication by another time-frequency function. Frequency shifting of $G_\alpha[m, k]$ by $\beta[m]$ frequency bins is denoted

$$\text{Shift}\{G_\alpha, \beta\}[m, k] = G_\alpha[m, \beta[m] + k]. \quad (6)$$

As soon as we shift G_α , we have to prevent aliasing due to spectral content moved out of the bounds, and fill in emptied regions at the other bound [24]. *Shift* includes this process.

Scaling (or dilation) can be applied to the spectral envelope and to the source (it is then called pitch-shifting), with the same boundary management than the one used for shifting. Scaling by a ratio $\lambda[m]$ is denoted

$$\text{Scale}\{G_\alpha, \lambda\}[m, k] = G_\alpha[m, \lambda[m]k]. \quad (7)$$

Warping (or arbitrary modifying) consists of applying an arbitrary function $W[m, k]$ to $G_\alpha[m, k]$, and is denoted

$$\text{Warp}\{G_\alpha, W\}[m, k] = G_\alpha[m, W[m, k]]. \quad (8)$$

Shifting and scaling are simple cases of warping function. We note *Mul* the multiplication transform ($\text{Mul}\{G_\alpha, E\} = G_\alpha \cdot E$) and *Id* the identity transform ($\text{Id}\{G_\alpha\} = G_\alpha$).

We define the time-varying interpolation $G_y[m, \bullet]$ of spectral functions $G_x[m, \bullet]$ and $G_q[m, \bullet]$ by using a time-varying interpolation coefficient $\gamma[m]$ that evolves in the range $[0; 1]$

$$G_y = \text{Interp}\{G_x, G_q, \gamma\}. \quad (9)$$

Notice that $\gamma[m]$ is not necessarily a monotonic function of time. Warping does not necessarily guaranty a valid STFT. Therefore, special frequency warping functions need to be used [25].

3. EFFECTS BASED ON A SOURCE-FILTER PARADIGM

We will present effects that are based on a source-filter modeling of spectral information³. We will successively consider effects that modify the spectral envelope only, the source only or both.

3.1. Effects on the Spectral Envelope

When applying the identity function to the source, only the spectral envelope is transformed

$$F_y = S_x \cdot T_H\{H_x, H_q\}. \quad (10)$$

In this class of effect, we can regroup not only spectral envelope interpolation such as cross-synthesis and hybridization, but also adaptive spectral envelope modifications. When the processing is linear, T_H simply becomes a multiplication of the spectral envelope by a time-frequency function, and then the explicit source-filter separation is no more needed.

3.1.1. Adaptive Spectral Envelope Modifications

Modifying a spectral envelope is useful not only for transforming voiced sounds as it distorts and moves the formants, but also for creating new sounds for electro-acoustic music as it strongly changes the spectral envelope. To prevent the generation of annoying sounds, it is well known that one has to change the modification of the spectral envelope over time. Based on the functions defined in sec. 2.5, we have built several adaptive spectral envelope modifiers to generate timbre modulation according to the content of the signal [16]. Adaptive shifting of the spectral envelope is then given by

$$H_y = Shift\{H_x, \beta\}. \quad (11)$$

Adaptive spectral envelope scaling depends on a scaling ratio $\lambda[m]$ greater or lower than 1 (expanding/compressing) that is given by

$$H_y = Scale\{H_x, \lambda\}. \quad (12)$$

Adaptive spectral envelope warping uses a non linear curve W which varies in time. We have found useful to express W as a linear interpolation between the identity warping ($W[m, k] = k$) and a vector $c_2[m, k] \in [0; 1]$, $k = 1, \dots, N$. This allows for an easy balance between a bypass effect and a warping effect. In practice, the interpolation ratio $c_1[m] \in [0; 1]$ is for example derived from the RMS, or the voiciness, whereas c_2 is derived from the spectral envelope or its integral

$$W[m, k] = c_1[m] \cdot c_2[m, k] \cdot N + (1 - c_1[m]) k. \quad (13)$$

The spectral envelope is given by

$$H_y = Warp\{H_x, W\}. \quad (14)$$

³Existing effects using source-filter processing are referenced in the title of the Subsection, whereas adaptive effects that we developed are not.

3.1.2. Vocoding Effect [3, 4, 2]

The classical vocoding effect consists of applying the spectral envelope H_q to F_x . This adaptive effect is historically implemented using the channel vocoder [3] but can also be implemented through the use of the phase vocoder [4]

$$T_H\{H_x, H_q\} = H_x \cdot H_q, \quad (15)$$

$$F_y = F_x \cdot H_q. \quad (16)$$

In practice, the input signal $x[n]$ that provides the source must be a rich and complex one (e.g. a monophonic or polyphonic harmonic sound, or an inharmonic sound such as a bell). The “filtering sound” $q[n]$ must exhibit strong formants, such as the human voice. Indeed, most of vocoding effect examples consist of making instruments talk⁴.

Cross-synthesis such as defined in [2] is another occurrence of the vocoding effect. The only difference when compared to the classical one is that the spectral envelope H_q is explicitly extracted by using an LPC analysis and then applied by using an IIR filter [2] or a phase vocoder [26]. The output STFT is

$$F_y = S_x \cdot H_q. \quad (17)$$

Even though cross-synthesis is used to describe the process involved in [2], we prefer to define cross-synthesis as a more general processing, that blends parameters coming from two different sounds to generate a third one.

3.1.3. Interpolation Between Two Spectral Envelopes

We will consider the “L-interpolation” H_y between two spectral envelopes H_x and H_q as the interpolation between time-varying parameter sets $P_{H_x}[m, \bullet]$ and $P_{H_q}[m, \bullet]$ that represent the spectral envelopes (H_x itself, cepstral coefficients, auto-regressive coefficients, formant coefficients, autocorrelation coefficients, reflection coefficients). The interpolated value

$$P_{H_y}[m, \bullet] = L_P^{-1}(H_y)[m, \bullet] \quad (18)$$

is given by

$$L_P^{-1}(H_y) = Interp\{L_P^{-1}(H_x), L_P^{-1}(H_q), \gamma\}, \quad (19)$$

with the time-varying interpolation ratio $\gamma[m] \in [0; 1]$. We denote

$$H_y = L-Interp\{H_x, H_q, \gamma\} \quad (20)$$

this interpolation between parameters of H_x and H_q . The cross adaptive effect may imply complex strategies when the sets to interpolate do not have the same number of parameters (e.g. from 5 to 4 formants). We now focus on the particular case of hybridization.

3.1.4. Hybridization (Voice Morphing) [10, 11]

We call hybridization or voice morphing the cross-synthesis of one voice by another voice previously analyzed. Its purpose is to make a singing voice fit to specific properties: e.g. recreating the castrato’s voice of Farinelli [10] (see Fig. 3), or singing with someone else’s voice such as in the Voice Impersonator in a karaoke

⁴A famous example is the talk box, popularized by Pete Frampton in the 1970ies. The guitar signal goes through a tube into the mouth, and the spectral envelope of the mouth is superimposed to the guitar formants.

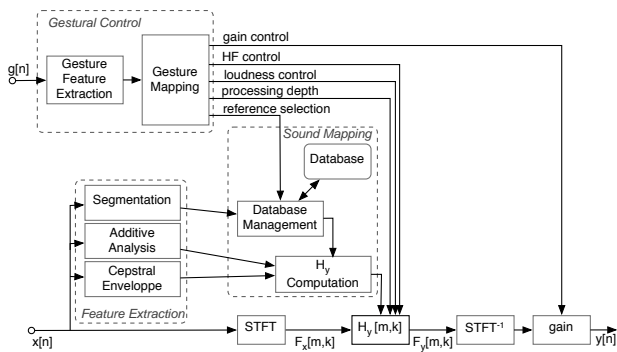


Figure 3: Diagram of the voice morphing developed to recreate a castrato's voice: Farinelli [10].

context [11]. The spectral envelope H_y is computed by interpolating between the envelope of the signal itself and the envelope of the signal stored in the database, according to specific mapping rules. The output STFT is given by

$$F_y = S_x \cdot H_y. \quad (21)$$

3.1.5. Adaptive Equalizer

Starting from the structure of an equalizer, which is essentially to filter a signal by a given spectral frequency response, we have extended it to an adaptive effect, where the frequency response H_q is provided through the adaptive control section of the processing system displayed on Fig. 2. In this adaptive effect, H_q can be generated from any vector feature of the input signal q after some specific mapping (see [16] for more details). The only constraint imposed to H_q is that for any time m , $H_q[m, \bullet]$ evolves sufficiently slowly as a function of frequency to be considered as a spectral envelope. The output STFT is given by

$$F_y = H_q \cdot F_x. \quad (22)$$

Moreover, we have to ensure that the time evolution of $H_q[\bullet, k]$ is sufficiently slow so that, given the time increment, its spectrum fits in the bandwidth mentioned in sec. 2.3 [5] in order to prevent aliasing of parameters.

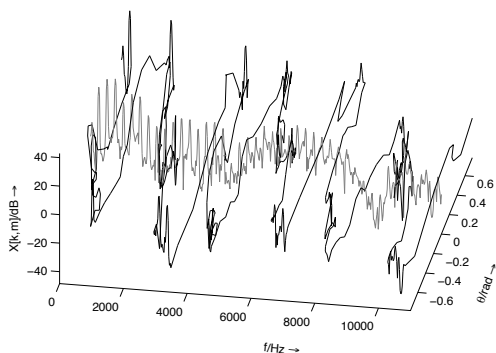


Figure 4: Frequency-azimuth domain of the spectral panoramization STFTs.

One of the typical example we have developed is the adaptive spectral panoramization [16]. It consists of generating a stereophonic signal by an adaptive splitting of the spectrum of the input signal x . We evaluate a panoramization angle vector $\theta[m, \bullet]$, which is derived from input sound features (e.g. the waveform). STFTs of the two output signals y_l and y_r are computed from the Blumlein law

$$\begin{cases} Y_l = \frac{\sqrt{2}}{2} (\cos \theta + \sin \theta) \cdot F_x, \\ Y_r = \frac{\sqrt{2}}{2} (\cos \theta - \sin \theta) \cdot F_x. \end{cases} \quad (23)$$

In practice, each frequency bin of the STFT F_x is moved to a specific location in space, see Fig. 4. This adaptive effect increase the perception of surrounding sound, and the signal is split with more or less independent motions and speeds.

3.2. Effect on the Source

Effects on the source preserve the spectral envelope so $H_y = H_x$. The possible effect that falls into this category are pitch-shifting, ring modulation (particular case of sec. 3.3.1) and frequency warping.

3.2.1. Pitch-Shifting with Formant Preservation [27, 28]

Pitch-shifting consists of resampling the source and preserving the sound duration, e.g. using the overlap-add of the STFT technique [24]. However, the spectral envelope is also scaled (leading to the ‘‘Donald Duck’’ effect). In order to preserve formants, a pitch-shifting algorithm must be applied to the source only [27, 28]

$$S_y = Scale\{S_x, \beta\}, \quad (24)$$

with β being the pitch-shifting ratio. When $\beta[m]$ varies in time, the pitch-shifting becomes adaptive (allowing intonation changes and adaptive resampling [21]).

3.2.2. Adaptive Spectral Warping

Time-varying spectral warping allows to change a musical sound from an harmonic one to an inharmonic one, while preserving the spectral envelope [16]. The source is computed as

$$S_y = Warp\{S_x, W\}, \quad (25)$$

where the warping function W has the same structure than the one specified in Eq. (13).

3.3. Effect on the Source and the Spectral Envelope

3.3.1. Adaptive Ring-Modulation

The famous ring modulation is a simple amplitude modulation with no DC component⁵. It consists of multiplying a signal $x[n]$ by a modulator signal. We only consider the case where the modulator is a simple sinusoidal wave $x_{mod}[n] = \sin(2\pi f_{mod}n)$, with $f_{mod} = \frac{k_{mod}}{N} f_s$, where f_s is the audio signal sampling rate. The whole spectrum is then duplicated and shifted

$$F_y = Shift\{F_x, k_{mod}\} + Shift\{F_x, -k_{mod}\}, \quad (26)$$

⁵The modulation frequency is over 20 Hz, so that it is timbre and not amplitude (and so rhythm) that is modified.

which implies that both source and spectral envelope are modified. The adaptive ring modulation uses a time-varying modulation frequency $f_{mod}[n]$. We designed an adaptive ring modulation by controlling the frequency of the modulator from mapped input signal features [16].

By a careful selection of a sound feature, a ring modulator that preserve the fundamental frequency can be designed [29]. If $f_{mod}[n] = M f_0[n]$ with M an integer, pitch is unchanged. When $f_{mod}[n] = M f_0[n]/P$ with P an integer, pitch is transposed down. These particular cases only affect the spectral envelope (sec. 3.1).

For any other values $f_{mod}[n] \neq M f_0[n]/P$, the sound remains inharmonic. In this case, the adaptive ring-modulation can be applied to the source only [16]

$$S_y = Shift\{S_x, k_{mod}\} + Shift\{S_x, -k_{mod}\}. \quad (27)$$

As it only modifies the source, it can be considered as an effect of sec. 3.2. This effect provides an inharmonization of a voice while preserving its intelligibility; this is also useful for transposing inharmonic sound, such as bell sounds.

3.3.2. Adaptive Spectral Envelope and Source Modification

Adaptive spectral envelope and source modifications (shifting, scaling or warping) can be combined [16] to provide interesting effects for processing electroacoustic sounds. The output STFT is

$$F_y = T_S\{S_x\} \cdot T_H\{H_x\}. \quad (28)$$

A particular case is the gender change.

3.3.3. Gender Change [30]

To transform a female voice into a male voice, and vice-versa [30, 18], pitch has to be shifted (up for male-to-female, down for female-to-male) and the formants have to be warped, more precisely this is a combination of scaling the spectral envelope in order to take into account changes in the length of the vocal tract and a shift of the spectral envelope in order to take into account synchronization between the fundamental frequency and the first formant's frequency for high pitched sounds. It is an auto-adaptive effect, controlled by $f_0[n]$, with the output STFT

$$F_y = Scale\{S_x, \alpha\} \cdot Warp\{H_x, \beta\}. \quad (29)$$

3.3.4. Robotization with Spectral Envelope Modifications

Robotization consists of nulling the phase for every block of a voice signal $x[n] \cdot w[mR_A - n]$ [31], such as defined in Eq. (1). The hop size between blocks imposes the robot's pitch, that can also adaptively evolve in time according to sound features [15]. This effect only modifies the pitch. Formants can be independently modified, for example to obtain a Donald Duck robot effect.

$$F_y = |S_x| \cdot T_H\{H_x\}. \quad (30)$$

3.3.5. Effects Based on Interpolation [31, 32, 33, 34, 35, 36]

Effects based on spectra L-interpolation such as defined in Eq. (9) appear in the literature under many different names, not only depending on the interpolated parameter sets, but also depending on the authors. Here is a list of existing effects that falls into this category, with the names proposed by the authors

- *timbral morphing* [32, 33] when $\gamma[m] = \gamma_0$ and P_x, P_q are additive parameters (partials' amplitude and frequency);
- *mutation* [34, 35, 31] when

$$F_y[m, \bullet] = Interp\{\rho_x, \rho_q, \gamma_\rho\} \cdot e^{jInterp\{\varphi_x, \varphi_q, \gamma_\varphi\}};$$

with the specific case found in [37] when $\gamma_\rho = 0$ or 1, and $\gamma_\varphi = 1 - \gamma_\rho$. This is called hybridization by the author. Notice that it differs from the definition of hybridization we use in sec. 3.1.4;

- *spectral interpolation* [36] when

$$F_y[m, \bullet] = Interp\{S_x, S_q, \gamma_s\} \cdot Interp\{H_x, H_q, \gamma_h\};$$

with the particular case of voice morphing (hybridization in our definition) [10, 11] when $\gamma_s = 0$, and vocoding effect/cross-synthesis [3, 4, 2, 34, 26] when $\gamma_s = 0$ and $\gamma_h = 1$.

The so called *timbral metamorphosis* [38] is the general case when $\gamma[m]$ monotonically evolves from 0 to 1 and takes continuous or discrete values. Timbral morphing and mutations cannot be considered as based on the source-filter model.

3.4. About Resampling Control Parameters

According to the structure of the adaptive audio effects considered in this study, the result of the mapping of extracted features $G_q[m, \bullet]$ has to be a vector of size N , whatever the features are. Another constraint is that G_q is sampled at frequency f_s/R_A . Therefore, it has to be low-pass filtered to prevent aliasing and under sampled when the feature is generated at a higher rate, and re-sampled when provided at a lower rate. Nevertheless, in a musical situation, one may want to use and control the degree of aliasing to add modulations in the resulting sound.

4. CONCLUSIONS

We presented a general structure for adaptive effects built upon a Short-Time Fourier analysis-re-synthesis and a spectral source-filter decomposition. This structure demonstrated to be very flexible and allows to reformulate in an unified framework many already existing adaptive audio effects. As a part of the review of these existing effects, we have also presented several new adaptive effects that we developed: adaptive shifting, scaling and warping of the spectral envelope, adaptive equalizer and spectral panormization, adaptive ring modulation of the source and robotization with spectral envelope modifications. We hope that this formulation which combines STFT, source-filter decomposition and adaptive control will give rise to the development of new adaptive audio effects.

5. REFERENCES

- [1] E. Métois, "Musical Gestures and Audio Effects Processing," in *Proc. COST-G6 Workshop on Digital Audio Effects (DAFx-98)*, Barcelona, Spain, 1998, pp. 249–253.
- [2] J. A. Moorer, "The Use of Linear Prediction of Speech in Computer Music Applications," *Journal of the Audio Engineering Society*, vol. 27, no. 3, pp. 134–140, 1979.
- [3] H. W. Dudley, "Remaking speech," *Journal of the Acoustical Society of America*, vol. 17, pp. 169–177, 1939.

- [4] J. L. Flanagan and R. M. Gloden, "Phase Vocoder," *Bell Syst. Tech.*, vol. 45, pp. 1493–1509, 1966.
- [5] J. B. Allen and L. R. Rabiner, "A unified approach to Short-Time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, 1977.
- [6] J. Laroche and M. Dolson, "About this Phasiness Business," in *Proc. Int. Comp. Music Conf. (ICMC'97)*, Thessaloniki, 1997, pp. 55–58.
- [7] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, Springer-Verlag, Berlin, 1976.
- [8] C. Dodge, *Current Directions in Computer Music Research*, M. V. Mathews and J. Pierce, Eds, chapter On Speech Songs, pp. 9–17, MIT Press, Cambridge, Massachusetts, 1989.
- [9] P. Lansky, *Current Directions in Computer Music Research*, M. V. Mathews and J. Pierce, Eds, chapter Compositional Applications of Linear Predictive Coding, pp. 5–8, MIT Press, Cambridge, Massachusetts, 1989.
- [10] P. Depalle, G. Garcia, and X. Rodet, "Reconstruction of a castrato voice: Farinelli's voice," in *Proceedings of the IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics*, 1995.
- [11] P. Cano, A. Loscos, J. Bonada, M. de Boer, and X. Serra, "Voice Morphing System for Impersonating in Karaoke Applications," in *Proc. Int. Comp. Music Conf. (ICMC'00)*, Berlin, 2000, pp. 109–112.
- [12] J. Laroche and M. Dolson, "New Phase Vocoder Technique for Real-Time Pitch-Shifting, Chorusing, Harmonizing and Other Exotic Audio Modifications," *Journal of the Audio Engineering Society*, vol. 47, no. 11, 1999.
- [13] U. Zölzer, Ed., *DAFX - Digital Audio Effects*, U. Zölzer ed., J. Wiley & Sons, 2002.
- [14] A. M. Noll, "Short-time Spectrum and "Cepstrum" Techniques for Vocal Pitch Detection," *Journal of the Acoustical Society of America*, vol. 36, no. 2, pp. 296–302, 1964.
- [15] V. Verfaillie and D. Arfib, "ADAFx: Adaptive Digital Audio Effects," in *Proc. COST-G6 Workshop on Digital Audio Effects (DAFx-01)*, Limerick, Ireland, 2001, pp. 10–14.
- [16] V. Verfaillie, *Effets Audionumériques Adaptatifs : Théorie, Mise en Œuvre et Usage en Création Musicale Numérique*, Ph.D. thesis, Université Aix-Marseille II, 2003.
- [17] D. Arfib, *Recherches et applications en informatique musicale*, chapter Des Courbes et des Sons, pp. 277–286, Hermès, 1998.
- [18] X. Amatriain, J. Bonada, A. Loscos, J. L. Arcos, and V. Verfaillie, "Content-based Transformations," *Journal of New Music Research*, vol. 32, no. 1, pp. 95–114, 2003.
- [19] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Third Edition, 1996.
- [20] V. Verfaillie and D. Arfib, "Implementation Strategies for Adaptive Digital Audio Effects," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-02)*, Hamburg, Germany, 2002, pp. 21–26.
- [21] D. Arfib and V. Verfaillie, "Driving Pitch-Shifting and Time-Scaling Algorithms with Adaptive and Gestural Techniques," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-03)*, London, England, 2003, pp. 106–111.
- [22] G. Fairbanks, W. L. Everitt, and R. P. Jaeger, "Method for Time or Frequency Compression-Expansion of Speech," *IEEE Transactions on Audio and Electroacoustics*, vol. AU, no. 2, pp. 7–12, 1954.
- [23] X. Serra and J. O. Smith, "A Sound Decomposition System Based on a Deterministic plus Residual Model," *Journal of the Acoustical Society of America, Supp. 1*, vol. 89, no. 1, pp. 425–434, 1990.
- [24] E. Moulines and F. Charpentier, "Pitch Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones," *Speech Communication*, vol. 9, no. 5/6, pp. 453–467, 1990.
- [25] G. Evangelista, *DAFX - Digital Audio Effects*, chapter Time and Frequency Warping Musical Signals, pp. 439–463, U. Zölzer ed., J. Wiley & Sons, 2002.
- [26] P. Depalle and G. Poirot, "Svp: A Modular System for Analysis, Processing and Synthesis of Sound Signals," in *Proc. Int. Comp. Music Conf. (ICMC'91)*, Montréal, 1991, pp. 161–164, ICMA, San Francisco.
- [27] R. Bristow-Johnson, "A detailed analysis of time-domain formant-corrected pitch-shifting algorithm.," *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 340–352, 1995.
- [28] E. Moulines and J. Laroche, "Non-parametric technique for pitch-scale and time-scale modification.," *Speech Communication*, vol. 16, pp. 175–205, 1995.
- [29] P. Dutilleul, *Vers la machine à sculpter le son, modification en temps-réel des caractéristiques fréquentielles et temporelles des sons*, Ph.D. thesis, Université Aix-Marseille II, 1991.
- [30] X. Rodet, G. Poirot, and P. Depalle, "Synthèse multivoix," in *Séminaire 'Variabilité du locuteur'*, CIRM, Luminy, France, 1989.
- [31] D. Arfib, F. Keiler, and U. Zölzer, *DAFX - Digital Audio Effects*, chapter Time-Frequency Processing, pp. 237–97, U. Zölzer ed., J. Wiley & Sons, 2002.
- [32] J. M. Grey, *An Exploration of Musical Timbre*, Ph.D. thesis, Stanford University, 1975.
- [33] T. L. Peterson, "Vocal Tract Modulation of Instrumental Sounds by Digital Filtering," in *Proc. Second Annual Music Computation Conference*, 1975, pp. 33–41.
- [34] L. Polansky and M. McKinney, "Morphological Mutation Functions: Applications to Motivic Transformation and a New Class of Cross-Synthesis Techniques," in *Proc. Int. Comp. Music Conf. (ICMC'91)*, Montréal, 1991, pp. 234–242.
- [35] L. Polansky and T. Erbe, "Spectral Mutation in Soundhack," *Computer Music Journal*, vol. 20, no. 1, pp. 92–101, 1995.
- [36] D. Arfib, F. Keiler, and U. Zölzer, *DAFX - Digital Audio Effects*, chapter Source-Filter Processing, pp. 299–372, U. Zölzer ed., J. Wiley & Sons, 2002.
- [37] F. Boyer and R. Kronland-Martinet, "Granular resynthesis and transformation of sounds through wavelet transform analysis," in *Proc. Int. Comp. Music Conf. (ICMC'91)*, Colombo, 1989, pp. 51–54, ICMA, San Francisco.
- [38] L. Landy, "Sound Transformations in Electroacoustic Music," 1991.