Adaptive Digital Audio Effects (A-DAFx): A New Class Of Sound Transformations

Vincent Verfaille, *Member, IEEE,* Udo Zölzer, *Member, IEEE,* Daniel Arfib

***Abstract*— After covering the basics of sound perception and giving an overview of commonly used audio effects (using a perceptual categorization), we propose a new concept called adaptive digital audio effects (A-DAFx). This consists of combining a sound transformation with an adaptive control. To create A-DAFx, low-level and perceptual features are extracted from the input signal, in order to derive the control values according to specific mapping functions. We detail the implementation of various new adaptive effects and give examples of their musical use.**

***Index Terms*— Signal processing, adaptive control, music, feature extraction, psychoacoustic models, information retrieval**

I. Introduction

An audio effect is a signal processing technique used to modulate or to modify an audio signal. The word 'effect' is also widely used to denote how something in the signal (cause) is being perceived (effect), thus sometimes creating confusion between the perceived effect and the signal processing technique that induces it (*e.g.* the Doppler effect).

Audio effects sometimes result from creative use of technology with an explorative approach (*e.g.* phase vocoder, distorsion, compressor); they are more often based on imitation of either a physical phenomenon (physical or signal models), or either a musical behaviour (signal models in the context of analysis–transformation–synthesis techniques), in which case they are also called 'transformations'. For historical and technical reasons, effects and transformations are considered as different, processing the sound at its surface for the former and more deeply for the latter. However, we use the word 'effect' in its general sense of musical sound transformations.

The use of digital audio effects has been developing and expanding for the last forty years for composition, recording, mixing, and mastering of audio signals, as well as real-time interaction and sound processing. Various implementation techniques are used such as filters, delay lines, time-segment and time-frequency representations, with sample-by-sample or block-by-block processing [1], [2].

The sound to be processed by an effect is synthesized by controlling an acoustico-mechanic or digital system, and may contains musical gestures [3] that reflects its control. These musical gestures are well described by sound features: the intelligence is in the sound. The adaptive control is a time-varying control computed from sound features modified by specific mapping functions. For that reason it is somehow related to the musical gesture already in the sound, and offers a meaningful and coherent type of control.

This adaptive control may add complexity to the implementation techniques the effects are based on; the implementation has to be designed carefully, depending on whether it is based on real-time or non real-time processing.

Using the perceptual categorization, we remind basic facts about sound perception and sound features, and briefly describe commonly used effects and the techniques they rely on in section II. Adaptive effects are defined and classified in section III; the set of features presented in section II-B is discussed in section III-C. The mapping strategies from sound features to control parameters are presented in section IV. New adaptive effects are presented in section V, as well as their implementation strategies for time-varying control.

Manuscript received May 21, 2004; revised September 30, 2004; February 12, 2005; May 26, 2005. This work was supported by the Centre National de la Recherche Scientifique (CNRS) and the Région Provence-Alpes-Côte-d'Azur.

Dr. Verfaille is with the Sound Processing and Control Laboratory, Faculty of Music – McGill University, 555, Sherbrooke Street West, Montréal, Québec - H3A 1E3, Canada. Email: vincent@music.mcgill.ca

Pr. Zölzer is with the Department of Electrical Engineering, Helmut-Schmidt-University, Holstenhofweg 85, 22043 Hamburg, Germany. Email: udo.zoelzer@hsu-hh.de

Dr. Arfib is with the Laboratoire de Mécanique et d'Acoustique, LMA–CNRS, 31, chemin Joseph Aiguier, F-13402 Marseille Cedex 20, France. Email: arfib@lma.cnrs-mrs.fr

II. Audio Effects and Perceptual Classification

A. Classifications of Digital Audio Effects

There exist various classifications for audio effects. Using the methodological taxonomy, effects are classified by signal processing techniques [1], [2]. Its limitation is redundancy as many effects appear several times (*e.g.* pitch-shifting can be performed by at least three different techniques). A sound object typology was proposed by Pierre Schaeffer [4], but does not correspond to an effect classification.

Using the perceptual categorization, audio effects are classified according to the most altered perceptual attribute: loudness, pitch, time, space and timbre [5]. This classification is the most natural to musicians and audio listeners, since the perceptual attributes are clearly identified in music scores.

B. Basics of Sound and Effect Perception

We now review some basics of psychoacoustics for each perceptual attribute. We also highlight the relationships between perceptual attributes (or high level features) and their physical counterparts (signal or low level features), which are usually simpler to compute. These features will be used for adaptive control of audio effects (*cf. sec.* III).

1) Loudness: Loudness is the perceived intensity of the sound through time. Its computational models perform time and frequency integration of the energy in critical bands [6], [7]. The sound intensity level computed by RMS (root mean square) is its physical counterpart. Using an additive analysis

and a transient detection, we extract the sound intensity levels of the harmonic content, the transient and the residual. We generally use a logarithmic scale named decibels: loudness is then $L_{dB} = 20 \log_{10} I$, with I the intensity. Adding 20 dB to the loudness is obtained by multiplying the sound intensity level by 10. The musical counterpart of loudness is called dynamics, and corresponds to a scale ranging from *pianissimo* (*pp*) to *fortissimo* (*ff*) with a 3 dB space between two successive dynamic levels. *Tremolo* describes a loudness modulation, which frequency and depth can be estimated.

2) Time and rhythm: Time is perceived through two intimately intricate attributes: the duration of sound and gaps, and the rhythm, which is based on repetition and inference of patterns [8]. Beat can be extracted with autocorrelation techniques, and patterns with quantification techniques [9].

3) Pitch: Harmonic sounds have their pitch given by the frequencies and amplitudes of the harmonics; the fundamental frequency is the physical counterpart. The attributes of pitch are height (high/low frequency) and chroma (or color) [10]. A musical sound can be either perfectly harmonic (*e.g.* wind instruments), nearly harmonic (*e.g.* string instruments) or inharmonic (*e.g.* percussions, bells). Harmonicity is also related to timbre.

Psychoacoustic models of the perceived pitch use both the spectral information (frequency) and the periodicity information (time) of the sound [11]. The pitch is perceived in the quasi-logarithmic *mel* scale which is approximated by the log-Hertz scale. Tempered scale notes are transposed up by one octave when multiplying the fundamental frequency by 2 (same chroma, doubling the height). The pitch organization through time is called melody for monophonic sounds and harmony for polyphonic sounds.

4) Timbre: This attribute is difficult to define from a scientific point of view. It has been viewed for a long time as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" [12]. However, this does not take into account some basic facts, such as the ability to recognize and to name any instrument when hearing just one note or listening to it through a telephone [13]. The frequency composition of the sound is concerned, with the attack shape, the steady part and the decay of a sound, the variations of its spectral envelope through time (*e.g.* variations of formants of the voice), and the phase relationships between harmonics. These phase relationships are responsible for the whispered aspect of a voice, the roughness of low-frequency modulated signals, and also for the phasiness[1] introduced when harmonics are not phase aligned. We consider that timbre has several other attributes, including:

- the brightness or spectrum height, correlated to spectral centroid[2] [15], and computed with various models [16];
- the quality and noisiness, correlated to the signal-to-noise ratio (*e.g.* computed as the ratio between the harmonics and the residual intensity levels [5]) and to the voiciness (computed from the autocorrelation function [17] as the second highest peak value of the normalized autocorrelation);
- the texture, related to jitter and shimmer of partials/harmonics [18] (resulting from a statistical analysis of the partials' frequencies and amplitudes), to the balance of odd/even harmonics (given as the peak of the normalized autocorrelation sequence situated half way between the first and second highest peak values [19]) and to harmonicity;
- the formants (especially vowels for the voice [20]) extracted from the spectral envelope; the spectral envelope of the residual; and the mel-frequency critical bands (MFCC), perceptual correlate of the spectral envelope.

Timbre can be verbalized in terms of roughness, harmonicity, as well as openness, acuteness and laxness for the voice [21]. At a higher level of perception, it can also be defined by musical aspects such as *vibrato* [22], *trill and flatterzung*, and by note articulation such as *appoyando, tirando and pizzicato*.

5) Spatial Hearing: In the last, spatial hearing has three attributes: the location, the directivity, and the room effect. The sound is localized by human beings in regards to distance, elevation and azimuth, through inter-aural intensity (IID) and inter-aural time (ITD) differences [23], as well as through filtering via the head, the shoulders and the rest of the body (head-related transfer function, HRTF). When moving, sound is modified according to pitch, loudness and timbre, indicating the speed and direction of its motion (Doppler effect) [24]. The directivity of a source is responsible for the differences of transfer function according to the listener position related to the source. The sound is transmitted through a medium as well as reflected, attenuated and filtered by obstacles (reverberation and echoes) thus providing cues for deducing the geometrical and material properties of the room.

6) Relationship between Low Level Features and Perceptual Attributes: We depict in Fig. 2 a feature set we used in this study. The figure highlights the relationships between the signal features and their perceptual correlates, as well as the possible redundancy of signal features.

C. Commonly Used Effects

We now present an overview of commonly used digital audio effects, with a specific emphasis on timbre, since that perceptive attribute is the more complex and offers a lot more possibilities than the other ones.

1) Loudness Effects : Commonly used loudness effects modify the sound intensity level: the volume change, the tremolo, the compressor, the expander, the noise gate and the limiter. The tremolo is a sinusoidal amplitude modulation of the sound intensity level with a modulation frequency between 4 and 7 Hz (around the 5.5 Hz frequency modulation of the vibrato). The compressor and the expander modify the intensity level using a non-linear function; they are among the first adaptive effects that were created. The former compresses the intensity level, thus giving more percussive sounds, whereas

[1]Phasiness is usually involved in speakers reproduction, where phases inproperties make the sound poorly spatialized. In the phase vocoder technique, the phasiness refers to a reverberation artifact, that appears when neighbor frequency bins representing a same sinusoid have different phase unwrapping.

[2]The spectral centroid is also correlated to other low level features: the spectral slope, the zero-crossing rate, the high frequency content [14].

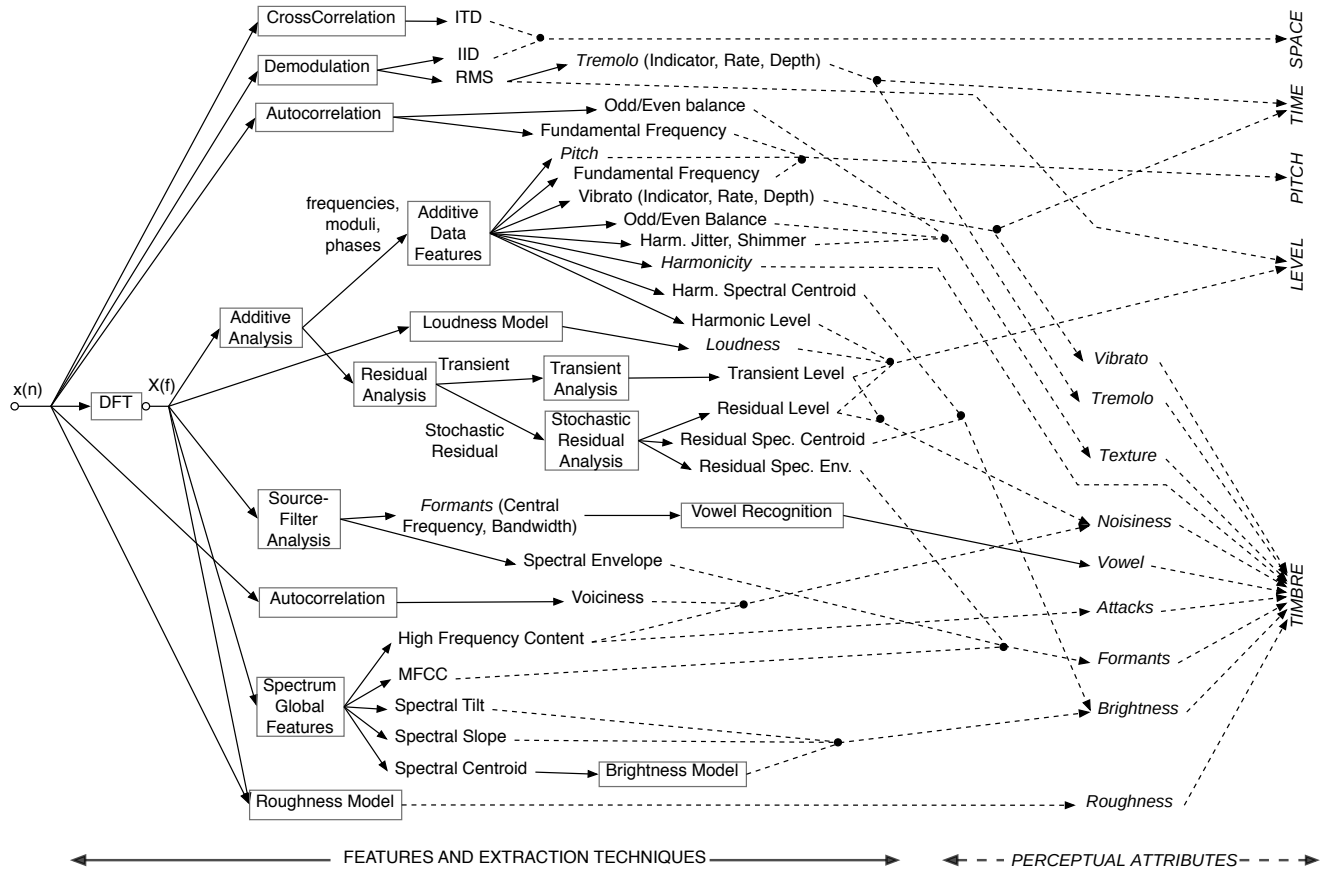


Fig. 1. *Set of features used as control parameters, with indications about the techniques used for extraction (left and plain lines) and the related perceptual attribute (right and dashed lines). Italic words refer to perceptual attributes.*

the latter has the opposite effect and is used to extend the dynamic range of the sound. With specific non-linear functions, we obtain noise gate and limiter effects. The noise gate bypasses sounds with very low loudness, which is especially useful to avoid the background noise that circulate throughout an effect system involving delays. Limiting the intensity level protects the hardware. Other forms of loudness effects include automatic mixers, automatic volume/gain control, which are sometimes noise-sensor equipped.

2) Time Effects: Time-scaling is used to fit the signal duration to a given duration, thus affecting rhythm. Resampling can perform time-scaling, resulting in an unwanted pitch-shifting. The time-scaling ratio is usually constant, and greater than 1 for time-expanding (or time-stretching, time-dilatation: sound is slowed down) and lower than 1 for time-compressing (or time-contraction: sound is sped up). Three block-by-block techniques permit to avoid this: the phase vocoder [25], [26], [27], SOLA [28], [29] and the additive model [30], [31], [32].

Time-scaling with the phase vocoder technique consists of using different analysis and synthesis step increments. The phase vocoder is performed using the Short-Time Fourier Transform (STFT) [33]. In the analysis step, the STFT of windowed input blocks is performed with a R_A samples step increment. In the synthesis step, the inverse Fourier transform delivers output blocks which are windowed, overlapped and then added with a R_S samples step increment. The phase vocoder step increments have to be suitably chosen to provide a perfect reconstruction of the signal [34], [33]. Phase computation is needed for each frequency bin of the synthesis STFT. The phase vocoder technique can time-scale any type of sound, but adds phasiness if no care is taken: a peak phase-locking technique solves this problem [35], [36].

Time-scaling with the SOLA technique[3] is performed by duplication or suppression of temporal grains or blocks, with pitch-synchronization of the overlapped grains in order to avoid low frequency modulation due to phase cancellation. Pitch-synchronization implies that the SOLA technique only correctly processes the monophonic sounds.

Time-scaling with the additive model results in scaling the time axis of the partial frequencies and their amplitudes. The additive model can process harmonic as well as inharmonic sounds while having a good quality spectral line analysis.

3) Pitch Effects: The pitch of harmonic sounds can be shifted, thus transposing the note. Pitch-shifting is the dual transformation of time-scaling, and consists of scaling the frequency axis of a time-frequency representation of the sound. A pitch-shifting ratio greater than 1 transposes up; lower than 1 it transposes down. It can be performed by a combination of time-scaling and resampling. In order to preserve the timbre and so forth the spectral envelope [19], the phase vocoder

[3]When talking about SOLA techniques, we refer to all the synchronized and overlap-add techniques: SOLA, TD-PSOLA, TF-PSOLA, WSOLA, etc.

decomposes the signal into source and filter for each analysis block: the formants are pre-corrected (in the frequency domain [37]), the source signal is resampled (in the time domain) and phases are wrapped between two successive blocks (in the frequency domain). The PSOLA technique preserves the spectral envelope [38], [39], and performs pitch-shifting by using a synthesis step increment that differs from the analysis step increment. The additive model scales the spectrum by multiplying the frequency of each partial by the pitch-shifting ratio. Amplitudes are then linearly interpolated from the spectral envelope. Pitch-shifting of inharmonic sounds such as bells can also be performed by ring-modulation.

Using a pitch-shifting effect, one can derive harmonizer and auto-tuning effects. Harmonizing consists of mixing a sound with several pitch-shifted versions of it, to obtain chords. When controlled by the input pitch and the melodic context, it is called smart harmony [40] or intelligent harmonization [41]. Auto-tuning consists of pitch-shifting a monophonic signal so that the pitch fits to the tempered scale [5], [42].

4) Timbre Effects : Timbre effects is the widest category of audio effects and includes vibrato, chorus, flanging, phasing, equalization, spectral envelope modifications, spectral warping, whisperization, adaptive filtering and transient enhancement or attenuation.

Vibrato is used for emphasis and timbral variety [43], and is defined as a complex timbre pulsation or modulation [44] implying frequency modulation, amplitude modulation, and sometimes spectral shape modulation [43], [45], with a nearly sinusoidal control. Its modulation frequency is around 5.5 Hz for the singing voice [46]. Depending on the instruments, the vibrato is considered as a frequency modulation with a constant spectral shape (*e.g.* voice, [20], string instruments [47], [13]), an amplitude modulation (*e.g.* wind instruments), or a combination of both, on top of which may be added a complex spectral shape modulation, with high-frequency harmonics enrichment due to non-linear properties of the resonant tube (voice [43], wind and brass instruments [13]).

A chorus effect appears when several performers play together the same piece of music (same in melody, rhythm, dynamics) with the same kind of instrument. Slight pitch, dynamic, rhythm and timbre differences arise because the instruments are not physically identical, nor are perfectly tuned and synchronized. It is simulated by adding to the signal the output of a randomly modulated delay-line [1], [48]. A sinusoidal modulation of the delay-line creates a flanging or sweeping comb filter effect [49], [50], [51], [48]. Chorus and flanging are specific cases of phase modifications known as phase shifting or phasing.

Equalization is a well-known effect that exists in most of the sound systems. It consists in modifying the spectral envelope by filtering with the gains of a constant-Q bank filter. Shifting, scaling or warping of the spectral envelope is often used for voice sounds since it changes the formant places, yielding to the so-called Donald Duck effect [19].

Spectral warping consists of modifying the spectrum in a non linear way [52], and can be achieved using the additive model or the phase vocoder technique with peak phase-locking [35], [36]. Spectral warping allows for pitch-shifting (or spectrum scaling), spectrum shifting, and in-harmonizing.

Whisperization transforms a spoken or sung voice into a whispered voice by randomizing either the magnitude spectrum or the phase spectrum STFT [27]. Hoarseness is a quite similar effect that takes advantage of the additive model to modify the harmonic-to-residual ratio [5].

Adaptive filtering is used in telecommunications [53] in order to avoid the feedback loop effect created when the output signal of the telephone loudspeaker goes into the microphone. Filters can be applied in the time-domain (comb filters, vocal-like filters, equalizer) or in the frequency-domain (spectral envelope modification, equalizer).

Transient enhancement or attenuation is obtained by changing the prominence of the transient compared to the steady part of a sound, for example using an enhanced compressor combined with a transient detector.

5) Spatial Effects: Spatial effects describe the spatialization of a sound with headphones or loudspeakers. The position in the space is simulated using intensity panning (*e.g.* constant power panoramization with two loudspeakers or headphones [23], vector-based amplitude panning (VBAP) [54] or Ambisonics [55] with more loudspeakers), delay lines to simulate the precedence effect due to ITD, as well as filters in a transaural or binaural context [23]. The Doppler effect is due to the behaviour of sound waves approaching or going away; the sound motion throughout the space is simulated using amplitude modulation, pitch-shifting and filtering [24], [56]. Echoes are created using delay-lines that can eventually be fractional [57]. The room effect is simulated with artificial reverberation units that use either delay-line networks or all-pass filters [58], [59] or convolution with an impulse response. The simulation of instruments' directivity is performed with linear combination of simple directivity patterns of loudspeakers [60]. The rotating speaker used in the Leslie/Rotary is a directivity effect simulated as a Doppler [56].

6) Multi-Dimensional Effects: Many other effects modify several perceptual attributes of sounds: we review a few of them. Robotization consists of replacing a human voice with a metallic machine-like voice by adding roughness, changing the pitch and locally preserving the formants. This is done using the phase vocoder and zeroing the phase of the grain STFT with a step increment given as the inverse of the fundamental frequency. All the samples between two successive non overlapping grains are zeroed[4] [27]. Resampling consists of interpolating the wave form, thus modifying duration, pitch and timbre (formants). Ring modulation is an amplitude modulation without the original signal; as a consequence, it duplicates and shifts the spectrum and modifies pitch and timbre, depending on the relationship between the modulation frequency and the signal fundamental frequency [61]. Pitch-shifting without preserving the spectral envelope modifies both pitch and timbre. The use of multi-tap monophonic or stereophonic echoes allow for rhythmic, melodic and harmonic constructions through superposition of delayed sounds.

[4]The robotization processing preserves the spectral shape of a processed grain at the local level. However, the formants are slightly modified at the global level when overlap-adding of grains with non phase-aligned grain (phase cancellation) or with zeros (flattening of the spectral envelope).

III. Adaptive Digital Audio Effects

A. Definition

We define adaptive digital audio effects (A-DAFx) as effects with a time-varying control derived from sound features transformed into valid control values using specific mapping functions [62], [63] as depicted in Fig. 2. They are also called 'intelligent effects' [64] or 'content-based transformations' [5]. They generalize observations of existing adaptive effects (compressor, auto-tune, cross-synthesis), and are inspired by the combination of amplitude/pitch follower combined with a voltage controlled oscillator [65]. We review the forms of A-DAFx depending on the input signal that is used for feature extraction, and then justify the sound feature set we chose in order to build this new class of audio effects.

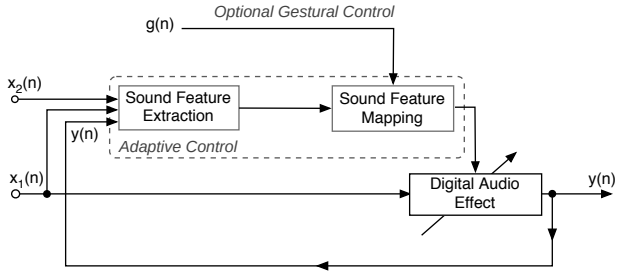


Fig. 2. *Diagram of the adaptive effect. Sound features are extracted from an input signal $x_1(n)$ or $x_2(n)$, or from the output signal $y(n)$. The mapping between sound features and the control parameters of the effect is modified by an optional gestural control.*

B. A-DAFx Forms

We define several forms of A-DAFx, depending on the signal from which sound features are extracted. Auto-adaptive effects have their features extracted from the input signal[5] $x_1(n)$. Adaptive or external-adaptive effects have their features extracted from at least one other input signal $x_2(n)$. Feedback adaptive effects have their features extracted from the output signal $y(n)$; it follows that auto-adaptive and external-adaptive effects are feed-forward. Cross-adaptive effects are a combination of at least two external-adaptive effects (not depicted in Fig. 2); they use at least two input signals $x_1(n)$ and $x_2(n)$. Each signal is processed using the features of another signal as controls. These forms do not provide a good classification for A-DAFx since they are not exclusive; they however provide a way to better describe the control in the effect name.

C. Sound Features

Sound features are used in a wide variety of applications such as coding, automatic transcription, automatic score following, and analysis-synthesis; they may require accurate computation depending on the application. For example, an automatic score following system must have accurate pitch and rhythm detection. To evaluate brightness, one might use the spectral centroid, with an eventual correction factor [66], whereas another may use the zero-crossing rate, the spectral slope, or psychoacoustic models of brightness [67], [68].

In the context of adaptive control, any feature can provide a good control: depending on its mapping to the effect's controls, it may provide a transformation that sounds. This is not systematically related to the accuracy of the feature computation, since the feature is extracted and then mapped to a control. For example, a pitch model using the autocorrelation function does not always provide a good pitch estimation; this may be a problem for automatic transcription or auto-tune, but not if it is low pass filtered and drives the frequency of a tremolo. There is a complex and subjective equation involving the sound to process, the audio effect, the mapping, the feature, and the will of the musician. For that reason, no restriction is given *a priori* to existing and eventually redundant features; however, perceptual features seem to be a better starting point when investigating the adaptive control of an effect.

We used the non-exhaustive set of features depicted in sec. II-B and in Fig. 1, that contains features commonly used for timbre space description (based on MPEG-7 proposals [69]) and other perceptual features extracted by the **PsySound** software [16] for non real-time adaptive effects. Note that also for real-time implementation, features are not really instantaneous: they are computed with a block-by-block approach so the sampling rate F_B is lower than the audio sampling rate F_A.

D. Are Adaptive Effects A New Class?

Adaptive control of digital audio effects is not new: it already exists in some commonly used effects. The compressor, expander, limiter and noise gate are feed-forward auto-adaptive effects on loudness, controlled by the sound intensity level with a non-linear warping curve and hysteresis effect. The auto-tuning (feedback) and the intelligent harmonizer (feed-forward) are auto-adaptive effects controlled by the fundamental frequency. The cross-synthesis is a feed-forward external adaptive effect using the spectral envelope of one sound to modify the spectral envelope of another sound.

The new concept that has been previously formulated is based on, promotes and provides a synthetic view of effects and their control (adaptive as described in this paper, but also gestural [63]). The class of adaptive effects that is built benefits from this generalization and provides new effects, creative musical ideas and clues for new investigations.

IV. Mapping Features to Control Parameters

A. The Mapping Structure

Recent studies defined specific strategies of mapping for gestural control of sound synthesizers [70] or audio effects [71], [72]. We propose a mapping strategy derived from the three-layer mapping that uses a perceptive layer [73] (more detailed issues are discussed in [63]).

To convert sound features $f_i(n), i = 1, ..., M$ into effect control parameters $c_j(n), j = 1, ..., N$, we use an M-to-N explicit mapping scheme[6] divided into two stages: sound

[5]The notation convention is small letters for time domain, *e.g.* $x_1(n)$ for sound signals, $g(n)$ for gestural control signal and $c(n)$ for feature control signal; and capital letters for frequency domain, *e.g.* $X(m, k)$ for short-time Fourier transform and $E(m, k)$ for the spectral envelope.

[6]M is the number of feature we use, usually between 1 and 5; N is the number of effect control parameters, usually between 1 and 20.

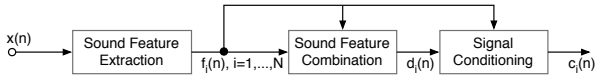


Fig. 3. *Diagram of the mapping between sound features and one effect control $c_i(n)$: sound features are first combined, and then conditioned in order to provide a valid control to the effect.*

feature combination and control signal conditioning (see Fig. 3 and [74], [63]).

The sound features may often vary rapidly and with a constant sampling rate (synchronous data) whereas the gestural controls used in sound synthesis vary less frequently and sometimes in an asynchronous mode. For that reason, we chose sound features for direct control of the effect and optional gestural control for modifications of the mapping between sound features and effect control parameters [63], [75], thus providing navigation by interpolation between presets.

B. Sound Feature Combination

The first stage combines several features, as depicted in Fig. 4. First, all the features are normalized in $[0, 1]$ for unsigned values features and in $[-1, 1]$ for signed value features. Second, a warping function – a transfer function that is not necessarily linear – can then be applied: a truncation of the feature in order to select an interesting part, a low pass filtering, a scale change (from linear to exponential or logarithmic), or any non-linear transfer function. Parameters of the warping function can also be derived from sound features (for example the truncation boundaries). Third, the feature combination is done by linear combination, except when weightings are derived from other sound features. Fourth and finally, a warping function can also be applied to the feature combination output in order to symetrically provide modifications of features before and after combination.

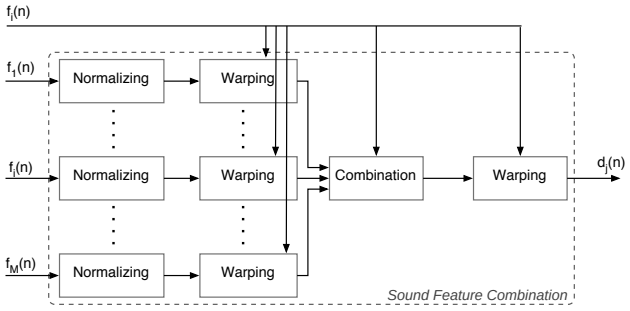


Fig. 4. *Diagram of the feature combination, first stage of the sound feature mapping. $f_i(n), i = 1, ..., M$ are the sound features, and $d_j(n), j = 1, ..., N$ are the combined features.*

C. Control Signal Conditioning

Conditioning a signal consists of modifying the signal so that its behaviour fits to pre-requisites in terms of boundaries and variation type; it is usually used to protect hardware from an input signal. The second mapping stage conditions the effect control signal $d_i(n)$ coming out from the feature combination box, as shown in Fig. 5, so that it fits the required behaviour of the effect controls. It uses three steps: an effect-specific warping, a low pass filter and a scaling.

First, the specific warping is effect-dependent. It may consist of quantizing the pitch curve to the tempered scale (auto-tune effect), quantizing the control curve of the delay time (adaptive granular delay, *cf.* sec. V-F.2), or modifying a time-warping ratio varying with time in order to preserve the signal length (*cf.* sec. V-B.2). Second, the low-pass filter ensures the suitability of the control signal for the selected application. Third and lastly, the control signal is scaled to the effect control boundaries given by the user, that are eventually adaptively controlled. When necessary the control signal, sampled at the block rate F_B, is resampled at the audio sampling rate F_A.

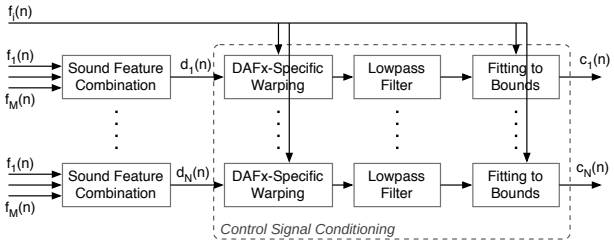


Fig. 5. *Diagram of the signal conditioning, second stage of the sound feature mapping. $c_i(n), n = 1, ..., N$ are the effect controls derived from sound features $f_i(n), i = 1, ..., M$. The DAFx-specific warping and the fitting to boundaries can be controlled by other sound features.*

D. Improvements Provided by the Mapping Structure

Our mapping structure offers a higher level of control and generalizes any effect: with adaptive control (remove the gestural control level), with gestural control (remove the adaptive control) or with both controls. Sound features are either short-term or long-term features; therefore they may have different and well identified roles in the proposed mapping structure. Short-term features (*e.g.* energy, instantaneous pitch or loudness, voiciness, spectral centroid) provide a continuous adaptive control with a high rate that we consider equivalent to a modification gesture [76] and useful as inputs (left horizontal arrows in Fig. 4 and 5). Long-term features computed after signal segmentation (*e.g.* vibrato, roughness, duration, note pitch or loudness) are often used for content-based transformations [5]. They provide a sequential adaptive control with low rate that we consider equivalent to a selection gesture, and that is useful as controls of the mapping (upper vertical arrow in Fig. 4 and 5).

V. Adaptive Effects and Implementations

Based on time-varying controls that are derived from sound features, commonly used A-DAFx were developed for technical or musical purposes, as answers to specific needs (*e.g.* auto-tune, compressor, and automatic mixer). In this section we illustrate the potential of this technique and investigate the effect control by sound features; we then provide new sound transformations by creative use of technology. For each effect presented in sec. V, examples are given with specific features and mapping functions in order to show the potential of the framework. Real-time implementations were performed

in the **Max/MSP** programming environment, and non real-time implementations in the **Matlab** environment.

A. Adaptive Loudness Effects

1) Adaptive Loudness Change: Real-time amplitude modulation with an adaptive modulation control $c(n)$ provides the following output signal:

$$y(n) = x(n) \cdot (1 + c(n)) \tag{1}$$

By deriving $c(n)$ from the sound intensity level, one obtains the compressor/expander (*c.f.* sec. II-C.1). By using the voiciness $v(n) \in [0, 1]$ and the mapping law $c(n) = (1 - \cos[\pi v(n)])/2$, one obtains a timbre effect: a 'voiciness gate' that removes voicy sounds and leaves only noisy sounds (which differs from the de-esser [77] that mainly removes the "s"). Adaptive loudness change is also useful for attack modification of instrumental and electroacoustic sounds (differently from compressor/expander), thus modifying loudness and timbre.

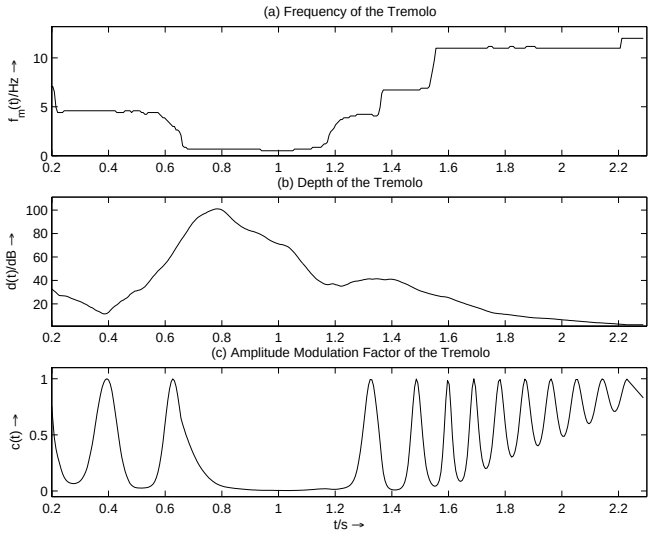


Fig. 6. *Control curves for the adaptive tremolo. (a) Tremolo frequency $f_m(n)$ is derived from the fundamental frequency as in Eq. (4). (b) Tremolo depth $d(m)$ is derived from the signal intensity level as in Eq. (5). (c) Amplitude modulation curve using the logarithmic scale given in Eq. (3).*

2) Adaptive Tremolo: This consists of a time-varying amplitude modulation with $f_m(n)$ the rate or modulation frequency in Hz, and $d(n)$ the depth, both being adaptively given by sound features. The amplitude modulation is expressed using the linear scale:

$$c(n) = d(n) \sin\left(2\pi \frac{f_m(n)}{F_A} n\right) \tag{2}$$

where F_A the audio sampling rate. It may also be expressed using the logarithmic scale:

$$c_{dB}(m) = 10^{d(m)\left(\sin\left(2\pi \frac{f_m(n)}{F_A} n\right) - 1\right)/40} \tag{3}$$

The modulation function is sinusoidal but may be replaced by any other periodic function (*e.g.* triangular, exponential, logarithmic or drawn by the user in a GUI). The real-time implementation only requires an oscillator, a warping function and an audio rate control. Adaptive tremolo allows for a more natural tremolo that accelerates/slows down (rhythm modification) and emphasizes/de-emphasizes (loudness modification) depending on the sound content. An example is given Fig. 6, where the fundamental frequency $F_0(m) \in [780, 1420]$ Hz and the sound intensity level $a(m) \in [0, 1]$ are mapped to the control rate and the depth according to the following mapping rules:

$$f_m(m) = 1 + 13 \cdot \frac{1420 - F_0(m)}{1420 - 780} \tag{4}$$

$$d(m) = 100 \cdot a(m) \tag{5}$$

B. Adaptive Time Effects

1) Adaptive Time-Warping: Time-warping is a non linear time-scaling. This non real-time processing uses a time-scaling ratio $\gamma(m)$ that varies with m the block index. The sound is then alternatively locally time-expanded when $\gamma(m) > 1$, and locally time-compressed when $\gamma(m) < 1$. The adaptive control is provided with the input signal (feedforward adaption). The implementation can be achieved either using constant analysis step increment R_A and time-varying synthesis step increment $R_S(m)$ or using time-varying $R_A(m)$ and constant R_S, thus providing more implementation efficiency. In the latter case, the recursive formulae of the analysis time index t_A and the synthesis time index t_S are:

$$t_A(m) = t_A(m-1) + R_A(m-1) \tag{6}$$

$$t_S(m) = t_S(m-1) + R_S \tag{7}$$

with the analysis step increment:

$$R_A(m-1) = \gamma(t_A(m-1)) \cdot R_S \tag{8}$$

Adaptive time-warping provides improvement to usual time-scaling, for example by minimizing the timbre modification. It allows for time-scaling with attack preservation when using an attack/transient detector to vary the time-scaling ratio [78], [79]. It also allows for time-scaling sounds with vibrato, when combined with adaptive pitch-shifting controlled by a vibrato estimator: vibrato is removed, the sound is time-scaled, and vibrato with same frequency and depth is applied [37].

Using auto-adaptive time-warping, we can apply fine changes in duration. A first example consists of time-compressing the gaps and time-expanding the sounding parts: the time-warping ratio γ is computed from the intensity level $a(n) \in [0, 1]$ using a mapping law such as $\gamma(n) = 2^{a(n)-a_0}$, with a_0 a threshold. A second example consists of time-compressing the voicy parts and time-expanding the noisy parts of a sound, using the mapping law $\gamma(n) = 2^{v(n)-v_0}$, with $v(n) \in [0, 1]$ the voiciness and v_0 the voiciness threshold. When used for local changes of duration, it provides modifications of timbre and expressiveness by modifying the attack, sustain and decay durations. Using cross-adaptive time-warping, time folding of sound A is slowed down or sped up depending on the sound B content. Generally speaking, adaptive time-warping allows for a re-interpretation of recorded sounds, for modifications of expressiveness (music) and perceived emotion (speech). Further research may investigate the

link between sound features and their mapping to the effect control on one side, and the modifications of expressiveness on the other side.

2) Adaptive Time-Warping That Preserves Signal Length: When applying a time-warping with an adaptive control, the signal length is changed. To preserve the original signal length, we must first evaluate the adaptive time-warped signal length according to the adaptive control curve given by the user, thus leading to a synchronization constraint. Second, we propose three specific mapping functions that modifies the time-warping ratio γ so that it verifies the synchronization constraints. Third, we modify the three functions so that they also preserve the initial boundaries of γ.

a) Synchronization constraint: Time indexes in Eq. (6) and (7) are functions of γ and m:

$$t_A(m) \quad = \quad \sum_{l=1}^{m} \gamma(t_A(l-1)) R_S \tag{9}$$

$$t_S(m) = m\, R_S, \quad 1 \leq m \leq M \tag{10}$$

The analysis signal length $L_A = \sum_{l=1}^{M} \gamma(t_A(l-1))\ R_S$ differs from the synthesis signal length $L_S = MR_S$. This is no more the case for $\tilde{\gamma}$ verifying the synchronization constraint:

$$M = \sum_{l=1}^{M} \tilde{\gamma}(t_A(l-1)) \tag{11}$$

b) Three Synchronization Schemes: The constrainted ratio $\tilde{\gamma}(t_A(m))$ can be derived from γ by:

1) addition:
   $\tilde{\gamma}_1(t_A(m)) = \gamma(t_A(m)) + 1 - \frac{\sum_{l=1}^{M} \gamma(t_A(l-1))}{M}$
2) multiplication:
   $\tilde{\gamma}_2(t_A(m)) = \gamma(t_A(m)) \cdot M / \sum_{l=1}^{M} \gamma(t_A(l-1))$
3) exponential weighting: $\tilde{\gamma}_3(t_A(m)) = [\gamma(t_A(m))]^{\gamma_3}$, with γ_3 the iterative solution[7] of:

$$\gamma_3 = \arg\min_{p} \left| \frac{\sum_{l=1}^{M} [\gamma(t_A(l-1))]^p}{M} - 1 \right| \tag{12}$$

An example is provided in Fig. 7.

Each of the three modification types of γ imposes a specific behavior to the time-warping control. For example, the exponential weighting is the only synchronization technique that preserves the locations where the signal has to be time-compressed or expanded: $\tilde{\gamma} > 1$ when $\gamma > 1$ and $\tilde{\gamma} < 1$ when $\gamma < 1$. However, none of these three methods take into account the boundaries of γ given by the user. A solution to this is provided below.

c) Synchronization That Preserves γ Boundaries: We define the clipping function:

$$\overline{\gamma}(p) = \begin{cases} \tilde{\gamma}_i(p) & \text{if } \tilde{\gamma}_i(p) \in [\gamma_{min}; \gamma_{max}] \\ \gamma_{min} & \text{if } \tilde{\gamma}_i(p) < \gamma_{min} \\ \gamma_{max} & \text{if } \tilde{\gamma}_i(p) > \gamma_{max} \end{cases} \tag{13}$$

[7]There is no analytical solution, so an iterative scheme is necessary.

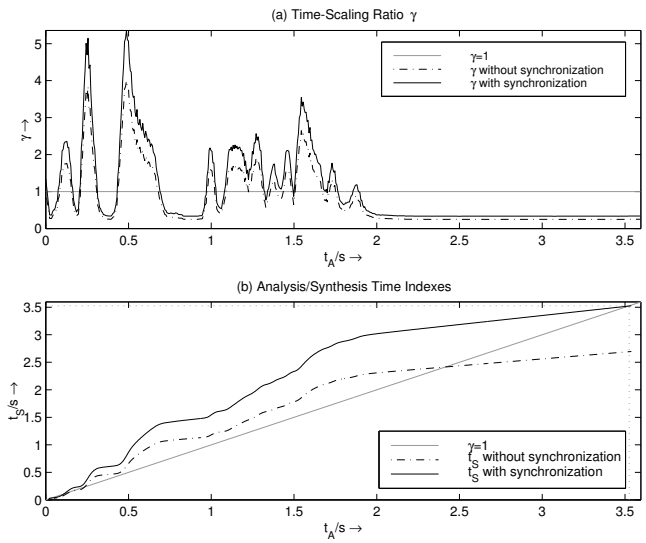


Fig. 7. *(a) The time-warping ratio γ is derived from the amplitude (RMS) as $\gamma(m) = 2^{4(a(m)-0.5)} \in [0.25, 4]$ (dashed line), and modified by the multiplication ratio $\gamma_m = 1.339$ (full line). (b) The analysis time index $t_A(m)$ is computed according to Eq. (6), verifying the synchronization constraint of Eq. (11).*

γ_{min} and γ_{max} denote the boundaries given by the user. The iterative solution $\overline{\overline{\gamma}}_i$ that both preserves the synchronization constraint of Eq. (11) and the initial boundaries is derived as:

$$\overline{\overline{\gamma}}_i = \arg\min_{\overline{\gamma}_i} \left| \frac{\sum_{l=1}^{M} \overline{\gamma}_i(t_A(l-1))}{M} - 1 \right| \tag{14}$$

where $i = 1, 2, 3$ respectively denotes addition, multiplication and exponential weighting.

The adaptive time-warping that preserves the signal length provides groove change when giving several synchronization points [63], that are beat-dependent for swing change [80] (time and rhythm effect). It also provides a more natural chorus when combined with adaptive pitch-shifting (timbre effect).

C. Adaptive Pitch Effects

1) Adaptive Pitch-Shifting: As for the usual pitch-shifting, three techniques can perform adaptive pitch-shifting with formant preservation in real-time: PSOLA, the phase vocoder technique combined with a source-filter separation [81], and the additive model. The adaptive pitch-shift ratio ρ is defined in the middle of the block as:

$$\rho(m) = \frac{F_{0,\text{out}}(m)}{F_{0,\text{in}}(m)} \tag{15}$$

where $F_{0,\text{in}}(m)$ (resp. $F_{0,\text{out}}(m)$) denotes the fundamental frequency of the input (resp. the output) signal. The additive model allows for varying pitch-shift ratios, since the synthesis can be made sample by sample in the time domain [30]. The pitch-shifting ratio is then interpolated sample by sample between two blocks. PSOLA allows for varying pitch-shifting ratios as long as one performs at the block level and performs energy normalization during the ovelap-add technique.

The phase vocoder technique has to be modified in order to permit that two overlap-added blocks have the same pitch-shifting ratio for all the samples they share, thus avoiding

phase cancellation of overlap-added blocks. First, the control curve must be low pass filtered to limit the pitch-shifting ratio variations. Doing so, we can consider that the spectral envelope does dot vary inside a block, and then use the source-filter decomposition to resample only the source. Second, the variable sampling rate $F_A(m)$ implies a variable length of the synthesis block $N_S(m) = N_A/\rho(m)$ and so a variable energy of the overlap-added synthesis signal. The solution we chose consists in imposing a constant synthesis block size $N_c = N_U/\max\rho(m)$, either by using a variable analysis block size $N_A(m) = \rho(m)N_U$ and then $N_S = N_U$, or by using a constant analysis block size $N_A = N_U$ and post-correcting the synthesis block x_{ps} according to:

$$y(n) = x_{ps}(n) \cdot \frac{h_{N_c}(n)}{w_{N(m)}(n)} \tag{16}$$

h is the Hanning window; $N(m) = 1 + \sum_{l=2}^{N_A} \rho(l)$ is the number of samples of the synthesis block m; $x_{ps}(n)$ is the resampled and formant-corrected block $n = 1, ..., N(m)$; w is the warped analysis window defined for $n = 1, ..., N(m)$ as $w_{N(m)}(n) = h(\sum_{l=1}^{n} \rho_s(l))$; and $\rho_s(n)$ is the pitch-shifting ratio $\rho(m)$ resampled at the signal sampling rate F_A.

A musical application of adaptive pitch-shifting is the adaptive detuning, obtained by adding to a signal its pitch-shifted version with a lower than a quarter-tone ratio (this also modifies timbre): an example is the adaptive detuning controlled by the amplitude as $\rho(n) = 2^{0.25 \cdot a(n)}$, where louder sounds are the most detuned. Adaptive pitch-shifting allows for melody change when controlled by long-term features such as the pitch of each notes of a musical sentence [82]. The auto-tune is a feedback adaptive pitch-shifting effect, where the pitch is shifted so that the processed sound reaches a target pitch. Adaptive pitch-shifting is also useful for intonation change, as explained below.

2) Adaptive Intonation Change : Intonation is the pitch information contained in prosody of human speech. It is composed of the macro-intonation and the micro-intonation [83]. To compute these two components, the fundamental frequency $F_{0,\text{in}}(m)$ is segmented over time. Its local mean $\overline{F_{0,\text{in}}^{loc}}$ is the macro-intonation structure for a given segment, and the reminder $\Delta F_{0,\text{in}}(m) = F_{0,\text{in}}(m) - \overline{F_{0,\text{in}}^{loc}}$ is the micro-intonation structure[8], as seen in Fig. 8. This yields the following decomposition of the input fundamental frequency:

$$F_{0,\text{in}}(m) = \overline{F_{0,\text{in}}^{loc}} + \Delta F_{0,\text{in}}(m) \tag{17}$$

The adaptive intonation change is a non real-time effect that modifies the fundamental frequency trends by deriving (α, β, γ) from sound features, using the decomposition:

$$F_{0,\text{out}}(m) = \gamma\overline{F_{0,\text{in}}} + \alpha(\overline{F_{0,\text{in}}^{loc}} - \overline{F_{0,\text{in}}}) + \beta\Delta F_{0,\text{in}}(m) \tag{18}$$

where $\overline{F_{0,\text{in}}}$ is the mean of $F_{0,\text{in}}$ over the whole signal [72]. One can independently control the mean fundamental frequency (γ *e.g.* controlled by the first formant frequency), the macro-intonation structure (α *e.g.* controlled by the second formant frequency) and the micro-intonation structure (β *e.g.* controlled by the intensity level $a(m)$); as well as strengthen ($\alpha \geq 1$ and $\beta \geq 1$), flatten ($0 \leq \alpha < 1$ and $0 \leq \beta < 1$) or inverse ($\alpha < 0$ and $\beta < 0$) an intonation, thus modifying the voice ambitus. Another adaptive control is obtained by replacing $\Delta F_{0,\text{in}}(m)$ by a sound feature.

[8]In order to avoid the rapid pitch-shifting modifications at the boundaries of voiced segments, the local mean of unvoiced segments is modified as the linear interpolation between its bound values (see Fig. 8.(b)). The same modification is applied to the reminder (micro-intonation).

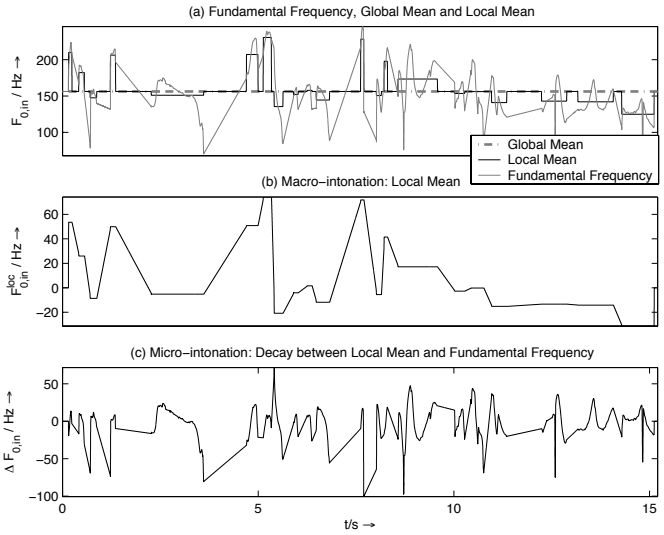


Fig. 8. *Intonation decomposition using an improved voiced/unvoiced mask. (a) Fundamental frequency $F_{0,in}(m)$, global mean $\overline{F_{0,in}}$ and local mean $\overline{F_{0,in}^{loc}}$. (b) Macro-intonation $\overline{F_{0,in}^{loc}}$ with linear interpolation between voiced segments. (c) Micro-intonation $\Delta F_{0,in}(m)$ with the same linear interpolation.*

D. Adaptive Timbre Effects

Since timbre is the widest category of audio effects, many adaptive timbre effects were developed such as voice morphing [84], [85], spectral compressor (also known as Contrast [52]), automatic vibrato [86], martianization [74] and adaptive spectral tremolo [63]. We present two other effects, namely adaptive equalizer and spectral warping.

1) Adaptive Equalizer: This effect is obtained by applying a time-varying equalizing curve $H_i(m, k)$ which is constituted of N_q filter gains of a constant-Q filter bank. In the frequency domain, we extract a vector feature of length N denoted[9] $f_i(m, \bullet)$ from the STFT $X_i(m, k)$ of each input channel i (the sound being mono or multichannel). This vector feature $f_i(m, \bullet)$ is then mapped to $H_i(m, \bullet)$, for example by averaging its values in each of the constant-Q segments, or by taking only the N_q first values of $f_i(m, \bullet)$ as the gains of the filters. The equalizer output STFT is then:

$$Y_i(m, k) \quad = \quad H_i(m, k) \cdot X_i(m, k) \tag{19}$$

If $H_i(m, k)$ varies too rapidly, the perceived effect is not varying equalizer/filtering but ring modulation of partials, and potentially phasing. To avoid this, we low pass filter $H_i(\bullet, k)$ in time [81], with L the under sampling ratio, $F_I = F_B/L$ the equalizer control sampling rate, and F_B the block sampling

[9]The notation $f_i(m, \bullet)$ corresponds to the frequency vector made of $f_i(m, k), k = 1, ..., N$.

rate. This is obtained by linear interpolation between two key vectors denoted C_{P-1} and C_P (see Fig. 9). For each block position $m, PL \leq m \leq (P+1)L$, the vector feature $f_i(m, \bullet)$ is given by:

$$f_i(m,k) = \alpha(m) \cdot C_{P-1}(m,k) + (1-\alpha(m)) \cdot C_P(m,k) \quad (20)$$

with $\alpha(m) = (m - PL)/L$ the interpolation ratio. The real-time implementation requires to extract a fast computing key vector $C_P(m,k)$, such as the samples buffer $C_P(m,k) = x(PLR_A + k)$, or the spectral envelope $C_P(m,k) = E(PL,k)$. However, non real-time implementations allow for using more computationally expensive features, such as a harmonic comb filter, thus providing an odd/even harmonics balance modification.

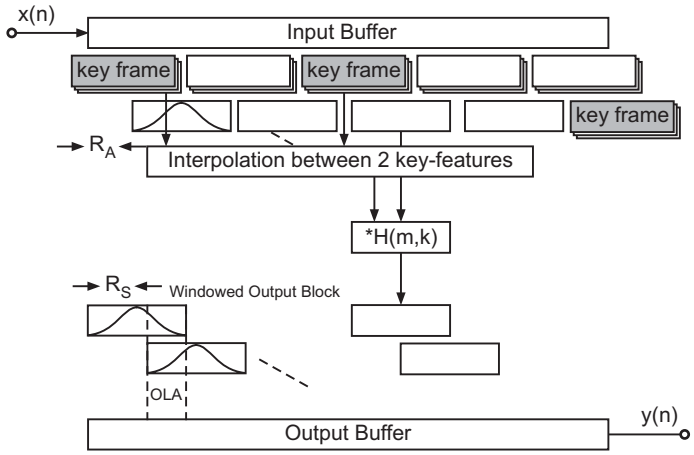


Fig. 9. *Block-by-block processing of adaptive equalizer. The equalizer curve is derived from a vector feature that is low pass filtered in time, using interpolation between key frames.*

2) Adaptive Spectral Warping: Harmonicity is adaptively modified when using spectral warping with an adaptive warping function $W(m,k)$. The STFT magnitude is:

$$|Y(m,k)| = |X(m, W(m,k))| \quad (21)$$

The warping function is:

$$W(m,k) = C_1(m) \cdot C_2(m,k) + (1 - C_1(m)) \cdot k \quad (22)$$

and varies in time according to two control parameters: a vector $C_2(m,k) \in [1,N]$, $k = 1,...,N$ (*e.g.* the spectral envelope E or its cumulative sum) which is the maximum warping function, and an interpolation ratio $C_1(m) \in [0,1]$ (*e.g.* the energy, the voiciness), which determines the warping depth. An example is given in Fig. 10, with $C_2(m,k)$ derived from the spectral envelope $E(m,k)$ as:

$$C_2(m,k) = k - (N-1) \cdot \frac{\sum_{l=2}^{k} E(m,l)}{\sum_{l=2}^{N} E(m,l)} \quad (23)$$

This mapping provides a monotonous curve, and prevents from folding over the spectrum.

Adaptive spectral warping allows for dynamically changing the harmonicity of a sound. When applied only to the source, it allows for better in-harmonizing a voice or a musical instrument since formants are preserved.

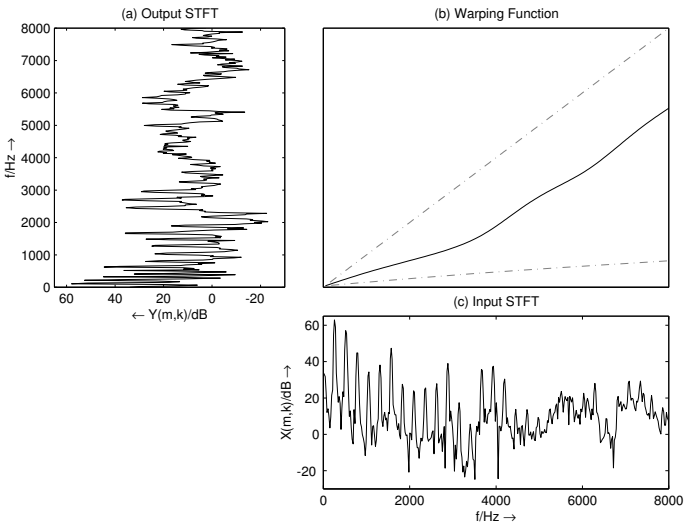


Fig. 10. *A-Spectral Warping: (a) Output STFT, (b) Warping function derived from the cumulative sum of the spectral envelope, (c) Input STFT. The warping function gives to any frequency bin the corresponding output magnitude. The spectrum is then non-linearly scaled according to the warping function slope p: compressed for $p < 1$ and expanded for $p > 1$. The dashed lines represent $W(m,k) = C_2(m,k)$ and $W(m,k) = k$.*

E. Adaptive Spatial Effects

We developed three adaptive spatial effects dealing with sound position in space, namely adaptive panoramization, adaptive spectral panoramization, and adaptive spatialization.

1) Adaptive Panoramization: It requires intensity panning (modification of left and right intensity levels) as well as delay, that are not taken into account in order to avoid the Doppler effect. The azimuth angle $\theta(m) \in [-\frac{\pi}{4}, \frac{\pi}{4}]$ varies in time according to sound features; constant power panoramization with the Blumlein law [23] gives the following gains:

$$L_l(n) = \frac{\sqrt{2}}{2}(\cos\theta(n) + \sin\theta(n)) \quad (24)$$

$$L_r(n) = \frac{\sqrt{2}}{2}(\cos\theta(n) - \sin\theta(n)) \quad (25)$$

A sinusoidal control $\theta(n) = \sin(2\pi f_{pan} n / F_A)$ with $f_{pan} > 20$ Hz is not heard anymore as motion but as ring modulations (with a phase decay of $\pi/2$ between the two channels). With more complex motions obtained from sound feature control, this effect does not appear because the motion is not sinusoidal and varies most of the time under 20 Hz. The fast motions cause a stream segregation effect [87], and the coherence in time between the sound motion and the sound content gives the illusion of splitting a monophonic sound into several sources.

An example consists of panoramizing synthesis trumpet sounds (obtained by frequency modulation techniques [88]) with an adaptive control derived from brightness, that is a strong perceptual indicator of brass timbre [89], as:

$$\theta(m) = \pi \cdot \frac{cgs(m)}{F_A} - \frac{\pi}{4} \quad (26)$$

Low-brightness sounds are left panoramized whereas high brightness sounds are right panoramized. Brightness of trumpet sounds evolves differently during notes attack and decay,

implying that the sound attack moves fastly from left to right whereas the sound decay moves slowly from right to left. This adaptive control then provides a spatial spreading effect.

2) Adaptive Spectral Panoramization: Panoramization in the spectral domain allows for intensity panning by modifying the left and right spectrum magnitudes as well as for time delays by modifying the left and right spectrum phases. Using the phase vocoder, we once again only used intensity panning in order to avoid the Doppler effect. To each frequency bin of the input STFT $X(m,k)$ we attribute a position given by the panoramization angle $\theta(m,k)$ derived from sound features. The resulting gains for left and right channels are then:

$$L_l(m,k) = \frac{\sqrt{2}}{2}(\cos\theta(m,k) + \sin\theta(m,k)) \quad (27)$$

$$L_r(m,k) = \frac{\sqrt{2}}{2}(\cos\theta(m,k) - \sin\theta(m,k)) \quad (28)$$

In this way, each frequency bin of the input STFT is panoramized separately from its neighbors (see Fig. 11): the original spectrum is then split across the space between two loudspeakers. To avoid the phasiness effect due to the lack of continuity of the control curve between neighbor frequency bins, a smooth control curve is needed, such as the spectral envelope. In order to control the variation speed of the spectral panoramization, $\theta(m,k)$ is computed from a time-interpolated value of a control vector (see the adaptive equalizer, sec. V-D.1). Adaptive spectral panoramization adds envelopment

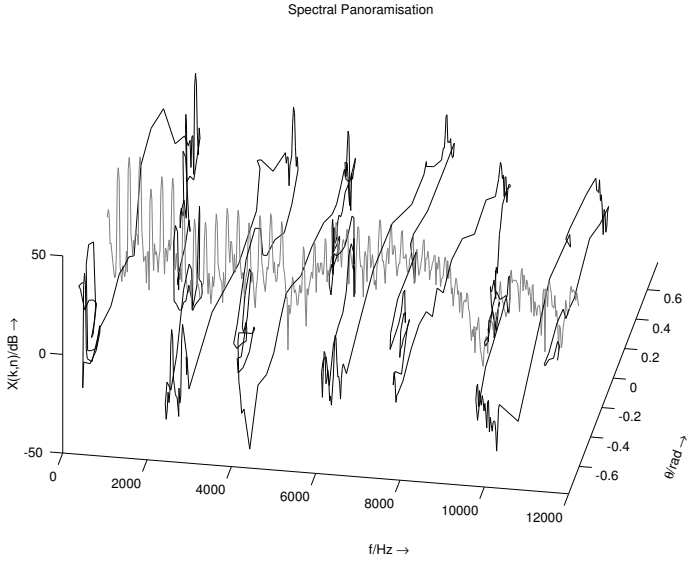


Fig. 11. *Frequency-space domain for the adaptive spectral panoramization (in black). Each frequency bin of the original STFT $X(m,k)$ (centered with $\theta = 0$, in gray) is panoramized with constant power. The azimuth angles are derived from sound features as $\theta(m,k) = x(mR_A - N/2 + k) \cdot \pi/4$.*

to the sound when the panoramization curve is smoothed. Otherwise, the signal is split into virtual sources having more or less independent motions and speeds. In the case the panoramization vector $\theta(m,\bullet)$ is derived from the magnitude spectrum with a multi-pitch tracking technique, it allows for source separation. When derived from the voiciness $v(m)$ as $\theta(m,k) = \frac{\pi v(m)\cdot(2k-N-1)}{4(N-1)}$, the sound localization varies between a point during attacks and a wide spatial spread during steady state, simulating width variations of the sound source.

3) Spatialization: Using VBAP techniques on an octophonic system, we tested the adaptive spatialization [90], [63]. A trajectory is given by the user (for example an ellipse), and the sound moves onto that trajectory, with adaptive control onto the position, the speed or the acceleration. Concerning the position control, the azimuth can depend on the chroma $a(n) = \frac{\pi \log_2 F_0(n)}{6}$, then splitting the sounds onto a spatial chromatic scale. The speed control adaptively depending on voiciness as $\dot{x}(n) = (1 - v(n))$ allows for the sound to move only during attacks and silences; on the contrary an adaptive control of speed given as $\dot{x}(n) = v(n)$ allows for the sound to move only during steady states, and not during attacks and silences.

F. Multi-Dimensional Adaptive Effects

Various adaptive effects affect several perceptual attributes simultaneously: adaptive resampling modifies time, pitch and timbre; adaptive ring modulation modifies only harmonicity when combined to formant preservation, and harmonicity and timbre when combined with formants modifications [81]; gender change combines pitch-shifting and adaptive formant-shifting [91], [86] to transform a female voice into a male voice, and *vice-versa*. We now present two other multi-dimensional adaptive effects: adaptive robotization that modifies pitch and timbre, and adaptive granular delay that modifies spatial perception and timbre.

1) Adaptive Robotization: Adaptive robotization changes expressiveness on two perceptual attributes, namely intonation (pitch) and roughness (timbre), and allows for transforming a human voice into an expressive robot voice [62].

This consists of zeroing the phase $\varphi(m,k)$ of the grain STFT $X(m,k)$ at a time index given by sound features: $Y(m,k) = |X(m,k)|$, and zeroing the signal between two blocks [62], [27]. The synthesis time index $t_S(m) = t_A(m)$ is recursively given as:

$$t_S(m) = t_S(m-1) + R_S(m) \quad (29)$$

The step increment $R_S(m) = \frac{F_A}{F_0(m)}$ is also the period of the robot voice, *ie.* the inverse of the robot fundamental frequency to which sound features are mapped (*e.g.* the spectral centroid as $F_0(m) = 0.01 \cdot cgs(m)$, in Fig. 12). Its real-time implementation implies the careful use of a circular buffer, in order to allow for varying window and step increments [92]. Both the harmonic and the noisy part of the sound are processed, and formants are locally preserved for each block. However, the energy of the signal is not preserved, due to the zero-phasing, the varying step increment and the zeroing process between two blocks, thus resulting in giving a pitch and modifying the loudness of noisy contents. An annoying buzz sound is then perceived, and can be easily removed by reducing the loudness modification: after zeroing the phases, the synthesis grain is multiplied by the ratio of analysis to synthesis intensity level computed on the current block m:

$$y_{norm}(n) = y(n) \cdot \frac{a_x(m)}{a_y(m)}, \quad n \in [m - N/2, m + N/2] \quad (30)$$

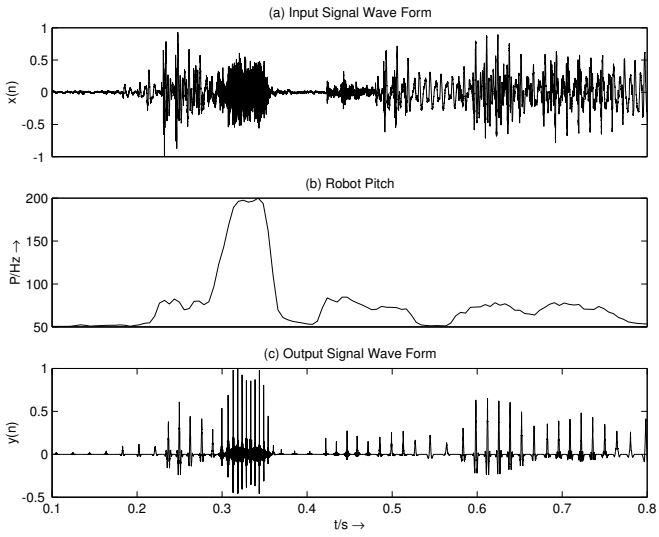


Fig. 12. *Robotization with a 512 samples block. (a) Input signal wave form. (b)* $F_0 \in [50; 200]Hz$ *derived from the spectral centroid as* $F_0(m) = 0.01 \cdot cgs(m)$*. (c) A-robotized signal wave form before amplitude correction.*

A second adaptive control is given on the block size $N(m)$ and allows for changing the robot roughness: the lower the block length, the higher the roughness. At the same time, it allows for preserving the original pitch (*e.g.* $N \geq 1024$) or removing it (*e.g.* $N \leq 256$), with an ambiguity in between. This is due to the fact that zero phasing a small block creates a main peak in the middle of the block and implies amplitude modulation (and then roughness). Inversely, zero phasing a large block creates several additional peaks in the window, the periodicity of the equally-spaced secondary peaks being responsible for the original pitch.

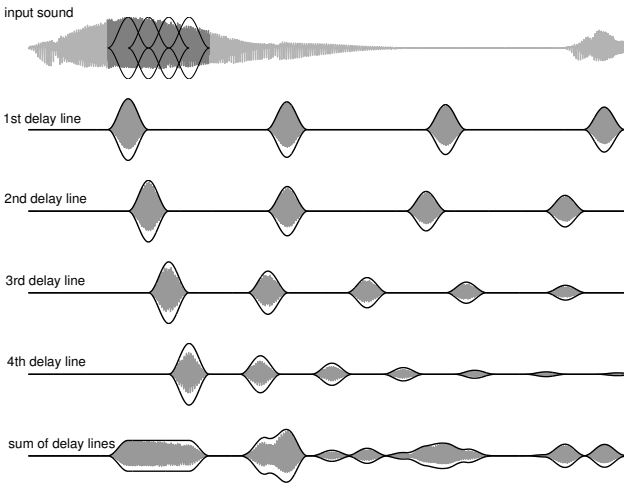


Fig. 13. *Illustration of the adaptive granular delay: each grain is delayed, with feedback gain* $g(m) = a(m)$ *and delay time* $\tau(m) = 0.1 \cdot a(m)$ *both derived from intensity level. Since intensity level of the four first grains is going down, the gains* $(g(m))^n$ *and delay time* $n\tau(m)$ *of the repetitions are also going down with* n*, resulting in a granular time-collapsing effect.*

2) Adaptive Granular Delay : This consists of applying delays to sound grains, with constant grain size N and step increment R_A [62], and varying delay gain $g(m)$ and/or delay time $\tau(m)$ derived from sound features (see Fig. 13). In non real-time applications, any delay time is possible, even fractional delay times [57], since each grain repetition is overlaped and added into a buffer. However, real-time implementations require to limit the number of delay lines, and so forth to quantize delay time and delay gain control curves to a limited number of values. In our experience, 10 values for the delay gain and 30 for the delay time is a good minimum configuration, yielding 300 delay lines.

In the case where only $g(m)$ varies, the effect is a combination between delay and timbre morphing (spatial perception and timbre). For example, when applying this effect to a plucked string sound and controlling the gain with a voiciness feature as $g(m) = 0.5 \cdot (1 + \cos(-\pi v(m)))$, the attacks are repeated a much longer time than the sustain part. With the complementary mapping $g(m) = 0.5 \cdot (1 + \cos(\pi(1 - v(m))))$, the attacks rapidly disappear from the delayed version whereas the sustain part is still repeated.

In the case where only $\tau(m)$ varies, the effect is a kind of granular synthesis with adaptive control, where grains collapse in time, thus implying modifications of time, timbre, and loudness. With a delay time derived from voiciness $\tau(m) = v(m)$ (in seconds), attacks and sustain parts of a plucked string sound have different delay times, so sustain parts may be repeated before the attack with repetitions going on, as depicted Fig. 13: not only time and timbre are modified, but also loudness, since the grains superposition is uneven.

Adaptive granular delay is a perfect example of how the creative modification of an effect with adaptive control offers new sound transformation possibilities; it also shows how the frontiers between the perceptual attributes modified by the effect may be blurred.

VI. Conclusions

We introduced a new class of sound transformations that we call adaptive digital audio effects (A-DAFx), and that generalizes audio effects and their control from observations of existing adaptive effects. Adaptive control is obtained by deriving effect controls from signal and perceptual features, thus changing the perception of the effect from linear to evolutive and/or from simple to complex. This concept also allows for the definition of new effects, such as adaptive time-warping, adaptive spectral warping, adaptive spectral panoramization, prosody change, and adaptive granular delay.

A higher level of control can be provided by combining the adaptive control with a gestural control of the sound feature mapping, thus offering interesting interactions including interpolation between adaptive effects and between presets.

A classification of effects was derived relying on the basis of perceptual attributes. Adaptive control provides creative tools to electro-acoustic music composers, musicians and engineers. This control allows for expressiveness changes and for sound re-interpretation, as especially noticeable in speech (prosody change, robotization, ring modulation with formant preservation, gender change, or martianization).

Further applications concern the study of emotion and prosody, for example to modify the prosody or to generate it appropriately. Formal listening tests are needed to evaluate

the mapping between sound features and prosody, thus giving new insights on how to modify the perceived emotion.

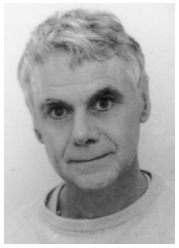
ACKNOWLEDGEMENTS

We would like to thank E. Favreau for discussions about creative phase vocoder effects, J.-C. Risset for discussions about creative use of effects in composition, and A. Sédès for spatialization experiments at MSH–Paris VIII. We also thank the reviewers for their comments and the significative improvements of the first drafts they proposed.

REFERENCES

[1] S. Orfanidis, *Introduction to Signal Processing*. Prentice Hall Int. Editions, 1996.
[2] U. Zölzer, Ed., *DAFX - Digital Audio Effects*. U. Zoelzer ed., J. Wiley & Sons, 2002.
[3] E. Métois, "Musical gestures and audio effects processing," in *Proc. COST-G6 Workshop on Digital Audio Effects (DAFx-98), Barcelona*, 1998, pp. 249–53.
[4] P. Schaeffer, *Le Traité des Objets Musicaux*. Seuil, Paris, 1966.
[5] X. Amatriain, J. Bonada, A. Loscos, J. L. Arcos, and V. Verfaille, "Content-based transformations," *J. New Music Research*, vol. 32, no. 1, pp. 95–114, 2003.
[6] E. Zwicker and B. Scharf, "A model of loudness summation," *Psychological Review*, vol. 72, pp. 3–26, 1965.
[7] E. Zwicker, "Procedure for calculating loudness of temporally variable sounds," *J. Ac. Soc. of America*, vol. 62, no. 3, pp. 675–82, 1977.
[8] P. Desain and H. Honing, *Music, Mind and Machine: Studies in Computer Music, Music Cognition, and Artificial Intelligence*. Thesis Publishers, Amsterdam, 1992.
[9] J. Laroche, "Estimating tempo, swing and beat locations in audio recordings," in *Proc. IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics*, 2001, pp. 135–8.
[10] R. Shepard, "Geometrical approximations to the structure of musical pitch," *Psychological Review*, vol. 89, no. 4, pp. 305–33, 1982.
[11] A. de Cheveigné, *Pitch*. C. Plack and A. Oxenham eds, Springer-Verlag, Berlin, 2004, ch. Pitch Perception Models.
[12] ANSI, *USA Standard Acoustic Terminology*. American National Standards Institute, 1960.
[13] J.-C. Risset and D. L. Wessel, *Exploration of timbre by analysis and synthesis*. D. Deutsch, Academic Press, New York, 1999, pp. 113–69.
[14] P. Masri and A. Bateman, "Improved modelling of attack transients in music analysis-resynthesis," in *Proc. Int. Computer Music Conf. (ICMC'96), Hong Kong*, 1996, pp. 100–3.
[15] S. McAdams, S. Winsberg, G. de Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes," *Psychological Research*, vol. 58, pp. 177–92, 1995.
[16] D. Cabrera, ""PsySound": a computer program for the psychoacoustical analysis of music," in *Proc. Australasian Computer Music Conf., MikroPolyphonie*, vol. 5, Wellington, New Zealand, 1999.
[17] J. C. Brown and M. S. Puckette, "Calculation of a narrowed autocorrelation function," *J. Ac. Soc. of America*, vol. 85, pp. 1595–601, 1989.
[18] S. Dubnov and N. Tishby, "Testing for gaussianity and non linearity in the sustained portion of musical sounds," in *Proc. Journées Informatique Musicale (JIM'96)*, 1996.
[19] D. Arfib, F. Keiler, and U. Zölzer, *DAFX - Digital Audio Effects*. U. Zoelzer ed., J. Wiley & Sons, 2002, ch. Source-Filter Processing, pp. 299–372.
[20] J. Sundberg, *The Science of the Singing Voice*. Dekalb, IL: Northern Illinois University Press, 1987.
[21] W. Slawson, *Sound Color*. Berkeley: Univ. of California Press, 1985.
[22] S. Rossignol, P. Depalle, J. Soumagne, X. Rodet, and J.-L. Collette, "Vibrato: Detection, estimation, extraction, modification," in *Proc. COST-G6 Workshop on Digital Audio Effects (DAFx-99), Trondheim*, 1999.
[23] J. Blauert, *Spatial Hearing: the Psychophysics of Human Sound Localization*. MIT Press, Cambridge, Massachussetts, 1983.
[24] J. Chowning, "The simulation of moving sound sources," *J. Audio Eng. Soc.*, vol. 19, no. 1, pp. 1–6, 1971.
[25] M. Portnoff, "Implementation of the digital phase vocoder using the fast fourier transform," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 24, no. 3, pp. 243–8, 1976.
[26] M. Dolson, "The phase vocoder: a tutorial," *Computer Music J.*, vol. 10, no. 4, pp. 14–27, 1986.
[27] D. Arfib, F. Keiler, and U. Zölzer, *DAFX - Digital Audio Effects*. U. Zoelzer ed., J. Wiley & Sons, 2002, ch. Time-Frequency Processing, pp. 237–97.
[28] E. Moulines and F. Charpentier, "Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Com.*, vol. 9, no. 5/6, pp. 453–67, 1990.
[29] J. Laroche, *Applications of Digital Signal Processing to Audio & Acoustics*. M. Kahrs and K. Brandenburg, eds., Kluwer Academic Publishers, 1998, ch. Time and Pitch Scale Modification of Audio Signals, pp. 279–309.
[30] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–54, 1986.
[31] X. Serra and J. O. Smith, "A sound decomposition system based on a deterministic plus residual model," *J. Ac. Soc. of America, sup. 1*, vol. 89, no. 1, pp. 425–34, 1990.
[32] T. Verma, S. Levine, and T. Meng, "Transient modeling synthesis: a flexible analysis/synthesis tool for transient signals," in *Proc. Int. Computer Music Conf. (ICMC'97), Thessaloniki*, 1997.
[33] J. B. Allen and L. R. Rabiner, "A unified approach to short-time fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–64, 1977.
[34] J. B. Allen, "Short term spectral analysis, synthesis and modification by discrete fourier transform," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 25, no. 3, pp. 235–8, 1977.
[35] M. S. Puckette, "Phase-lockedvocoder," in *Proc. IEEE ASSP Conf. on Appl. of Sig. Processing to Audio and Acoustics (Mohonk, N.Y)*, 1995.
[36] J. Laroche and M. Dolson, "About this phasiness business," in *Proc. Int. Computer Music Conf. (ICMC'97), Thessaloniki*, 1997, pp. 55–8.
[37] D. Arfib and N. Delprat, "Selective transformations of sound using time-frequency representations: An application to the vibrato modification," in *104th Conv. Audio Eng. Soc., Amsterdam*, 1998.
[38] R. Bristow-Johnson, "A detailed analysis of time-domain formant-corrected pitch-shifting algorithm," *J. Audio Eng. Soc.*, vol. 43, no. 5, pp. 340–52, 1995.
[39] E. Moulines and J. Laroche, "Non-parametric technique for pitch-scale and time-scale modification," *Speech Com.*, vol. 16, pp. 175–205, 1995.
[40] S. Abrams, D. V. Oppenheim, D. Pazel, and J. Wright, "Higher-level composition control in lusic sketcher: Modifiers and smart harmony," in *Proc. Int. Computer Music Conf. (ICMC'99), Beijing*, 1999, pp. 13–6.
[41] TC-Helicon, "Voice One, Voice Prism, http://www.tc-helicon.tc/," 2002.
[42] Antares, "Autotune, http://www.antarestech.com/," 2003.
[43] R. C. Maher and J. Beauchamp, "An investigation of vocal vibrato for synthesis," *Applied Acoustics*, vol. 30, pp. 219–45, 1990.
[44] C. E. Seashore, "Psychology of the vibrato in voice and speech," *Studies in the Psychology of Music*, vol. 3, 1936.
[45] V. Verfaille, C. Guastavino, and P. Depalle, "Perceptual evaluation of vibrato models," in *Colloquium on Interdisciplinary Musicology, Montréal (CIM'05)*, 2005.
[46] H. Honing, "The vibrato problem, comparing two solutions," *Computer Music J.*, vol. 19, no. 3, pp. 32–49, 1995.
[47] M. Mathews and J. Kohut, "Electronic simulation of violin resonances," *J. Ac. Soc. of America*, vol. 53, no. 6, pp. 1620–6, 1973.
[48] J. Dattoro, "Effect design, part 2: Delay-line modulation and chorus," *J. Audio Eng. Soc.*, pp. 764–88, 1997.
[49] B. Bartlett, "A scientific explanation of phasing (flanging)," *J. Audio Eng. Soc.*, vol. 18, no. 6, pp. 674–5, 1970.
[50] W. M. Hartmann, "Flanging and Phasers," *J. Audio Eng. Soc.*, vol. 26, pp. 439–43, 1978.
[51] J. O. Smith, "An allpass approach to digital phasing and flanging," in *Proc. Int. Computer Music Conf. (ICMC'84), Paris*, 1984, pp. 103–8.
[52] E. Favreau, "Phase vocoder applications in GRM tools environment," in *Proc. of the COST-G6 Workshop on Digital Audio Effects (DAFx-01), Limerick*, 2001, pp. 134–7.
[53] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, Third Edition, 1996.
[54] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–66, 1997.
[55] M. A. Gerzon, "Ambisonics in Mutichannel Broadcasting and Video," *J. Audio Eng. Soc.*, vol. 33, no. 11, 1985.
[56] J. O. Smith, S. Serafin, J. Abel, and D. Berners, "Doppler simulation and the Leslie," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-02), Hamburg*, 2002, pp. 13–20.
[57] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay," in *IEEE Signal Processing Magazine*, 1996, pp. 30–60.
[58] M. R. Schrœder and B. Logan, ""colorless" artificial reverberation," *J. Audio Eng. Soc.*, vol. 9, pp. 192–7, 1961.

[59] J. A. Moorer, "About this reverberation business," *Computer Music J.*, vol. 3, no. 2, pp. 13–8, 1979.
[60] O. Warusfel and N. Misdariis, "Directivity synthesis with a 3D array of loudspeakers - Application for stage performance," in *Proc. of the COST-G6 Workshop on Digital Audio Effects (DAFx-01), Limerick*, 2001, pp. 232–6.
[61] P. Dutilleux, "Vers la machine à sculpter le son, modification en temps-réel des caractéristiques fréquentielles et temporelles des sons," Ph.D. dissertation, University of Aix-Marseille II, 1991.
[62] V. Verfaille and D. Arfib, "ADAFx: Adaptive digital audio effects," in *Proc. of the COST-G6 Workshop on Digital Audio Effects (DAFx-01), Limerick*, 2001, pp. 10–4.
[63] V. Verfaille, "Effets audionumériques adaptatifs : Théorie, mise en œuvre et usage en création musicale numérique," Ph.D. dissertation, Université de la Méditerranée (Aix-Marseille II), 2003.
[64] D. Arfib, *Recherches et applications en informatique musicale*. Hermès, 1998, ch. Des Courbes et des Sons, pp. 277–86.
[65] R. Moog, "A voltage-controlled low-pass, high-pass filter for audio signal processing," in *17th Annual AES Meeting*, no. Preprint 413, 1965.
[66] J. W. Beauchamp, "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," *J. Audio Eng. Soc.*, vol. 30, no. 6, pp. 396–406, 1982.
[67] W. von Aures, "Der sensorische wohlklang als funktion psychoakustischer empfindungsgröfsen," *Acustica*, vol. 58, pp. 282–90, 1985.
[68] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Springer-Verlag, Berlin, 1999.
[69] G. Peeters, S. McAdams, and P. Herrera, "Instrument sound description in the context of MPEG-7," in *Proc. Int. Computer Music Conf. (ICMC'00), Berlin*, 2000.
[70] M. M. Wanderley, "Mapping strategies in real-time computer music," *Org. Sound*, vol. 7, no. 2, 2002.
[71] M. M. Wanderley and P. Depalle, "Gesturally controlled digital audio effects," in *Proc. COST-G6 Workshop on Digital Audio Effects (DAFx-00), Verona*, 2000, pp. 165–9.
[72] D. Arfib and V. Verfaille, "Driving pitch-shifting and time-scaling algorithms with adaptive and gestural techniques," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-03), London*, 2003, pp. 106–11.
[73] D. Arfib, J.-M. Couturier, L. Kessous, and V. Verfaille, "Strategies of mapping between gesture parameters and synthesis model parameters using perceptual spaces," *Org. Sound*, vol. 7, no. 2, pp. 135–52, 2002.
[74] V. Verfaille and D. Arfib, "Implementation strategies for adaptive digital audio effects," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-02), Hamburg*, 2002, pp. 21–6.
[75] V. Verfaille and M. M. Wanderley, "Mapping strategies for gestural control of adaptive digital audio effects," *in preparation*, 2005.
[76] C. Cadoz, *Les Nouveaux gestes de la musique*. H. Genevois and R. de Vivo, ed. Parenthèses, 1999, ch. Musique, geste, technologie, pp. 47–92.
[77] P. Dutilleux and U. Zölzer, *DAFX - Digital Audio Effects*. U. Zoelzer ed., J. Wiley & Sons, 2002, ch. Nonlinear Processing, pp. 93–135.
[78] J. Bonada, "Automatic technique in frequency domain for near-lossless time-scale modification of audio," in *Proc. Int. Computer Music Conf. (ICMC'00), Berlin*, 2000.
[79] G. Pallone, "Dilatation et transposition sous contraintes perceptives des signaux audio: application au transfert cinéma-vidéo," Ph.D. dissertation, University of Aix-Marseille III, 2003.
[80] F. Gouyon, L. Fabig, and J. Bonada, "Rhythmic expressiveness transformations of audio recordings: swing modifications," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-03), London*, 2003, pp. 94–99.
[81] V. Verfaille and P. Depalle, "Adaptive effects based on STFT, using a source-filter model," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-04), Naples*, 2004.
[82] E. Gómez, G. Peterschmitt, X. Amatriain, and P. Herrera, "Content-based melodic transformations of audio material for a music processing application," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-03), London*, 2003.
[83] A. D. Cristo, *Prolégomènes à l'étude de l'intonation*. Editions du CNRS, 1982.
[84] P. Depalle, G. Garcia, and X. Rodet, "Reconstruction of a castrato voice: Farinelli's voice," in *Proc. IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics*, 1995.
[85] P. Cano, A. Loscos, J. Bonada, M. de Boer, and X. Serra, "Voice morphing system for impersonating in karaoke applications," in *Proc. Int. Computer Music Conf. (ICMC'00), Berlin*, 2000, pp. 109–12.
[86] X. Amatriain, J. Bonada, A. Loscos, and X. Serra, *DAFX - Digital Audio Effects*. U. Zoelzer ed., J. Wiley & Sons, 2002, ch. Spectral Processing, pp. 373–438.
[87] A. Bregman, *Auditory Scene Analysis*. MIT Press, Cambridge, Massachussetts, 1990.
[88] J. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *J. Audio Eng. Soc.*, vol. 21, pp. 526–34, 1971.
[89] J.-C. Risset, "Computer study of trumpet tones," *J. Ac. Soc. of America*, vol. 33, p. 912, 1965.
[90] A. Sédès, B. Courribet, J.-B. Thiébaut, and V. Verfaille, *Espaces Sonores – Actes de Recherches*. CICM - Editions Musicales Transatlantiques, 2003, ch. Visualisation de l'Espace Sonore, vers la Notion de Transduction : une Approche Intéractive Temps-Réel, pp. 125–43.
[91] X. Amatriain, J. Bonada, A. Loscos, and X. Serra, "Spectral modeling for higher-level sound transformations," in *MOSART Workshop on Current Research Dir. in Computer Music, IUA-UPF, Barcelona*, 2001.
[92] D. Lebel, "Adaptive Digital Audio Effects using Audio Units," SPCL, McGill University, Faculty of Music, Tech. Rep., 2004.



Daniel Arfib (born 1949) is Research Director at the Laboratoire de Mécanique et d'Acoustique (LMA-CNRS), in Marseille (France). After studies in computer science (Engineer Degree from the École Centrale of Paris and Ph.D. at the University of Aix-Marseille II), he has joined the LMA computer music team and in parallel has followed composer activities. Former coordinator of the "DAFx" European COST action, he is now collaborating to 'ConGAS' (gestural control of audio systems).

Vincent Verfaille (born 1974) holds an Engineer Degree (Ing.) in Applied Mathematics with honors (Institut National des Sciences Appliquées, Toulouse, France, 1997), and a Ph.D. in Music Technology (ATIAM, University of Aix-Marseille II, France, 2003). He is pursuing postdoctoral research at the Faculty of Music in McGill University (Montréal, Canada). His research interests are analysis/synthesis techniques, sound processing, gestural and automated control and psychoacoustics.

Udo Zölzer received the Diplom-Ingenieur degree in electrical engineering from the University of Paderborn in 1985, the Dr.-Ingenieur degree from the Technical University Hamburg-Harburg (TUHH) in 1989 and completed a *habilitation* in Communications Engineering at the TUHH in 1997. Since 1999 he has been a Professor and head of the Department of Signal Processing and Communications at the Helmut Schmidt University, University of the Federal Armed Forces in Hamburg, Germany. His research interests are audio and video signal processing and communications. He is a member of the AES and the IEEE.