Etienne Thuillier

Real-Time Polyphonic Octave Doubling for the Guitar

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 14.03.2016

Thesis supervisor:

Prof. Vesa Välimäki

Thesis advisor:

M.Sc. (Tech.) Jens P. Kargo

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Etienne Thuillier

Title: Real-Time Polyphonic Octave Doubling for the Guitar

Date: 14.03.2016 Language: English Number of pages: 7+66

Department of Signal Processing and Acoustics

Professorship: Audio Signal Processing Code: S-89

Supervisor: Prof. Vesa Välimäki

Advisor: M.Sc. (Tech.) Jens P. Kargo

1 Summary

This thesis studies digital signal processing solutions for enriching live guitar sound by way of mixing-in octave-doubled versions of the chords and melodies performed on the instrument in real-time. Following a review of techniques applicable for real-time polyphonic octave doubling, four candidate solutions are proposed, amongst which two novel methods: ERB-PS2 and ERB-SSM2. Performance of said candidates is compared to that of three state of the art effect pedal offerings of the market. In particular, an evaluation of the added roughness and transient alterations introduced by each solution in the output sound is conducted. The ERB-PS2 method, which consists in doubling the instantaneous phases of the sub-bands signals extracted with a constant-ERB-bandwidth non-decimated IIR filter bank, is found to provide the best overall performance amongst the candidates. This novel solution provides greatly reduced latency compared to the baseline pedals, with comparable, and in some case improved, sound quality.

Keywords: digital signal processing, audio effect, pitch-shifting, multi-resolution filter bank, short-time Fourier transform, auditory roughness, transient smearing, audio systems, music, spectral analysis

Preface

The current Master's thesis work was essentially carried out in cooperation with the guitar effect team at TC Electronic's headquarters during my stay in Aarhus (Denmark), April and June 2014. Complementary research work and thesis drafting was conducted at Aalto University in May 2014 and throughout the first half of 2015.

Firstly, I would like to give a warm salute to Tore Lynggaard Mogensen for being accessible to discuss audio effects for the guitar with random Musikmesse fair visitors such as me, as well as for first introducing me to TC's team. Moreover, I thank René Østergaard who provided me the chance to work within TC Electronic's effect pedal department. I found the work environment and team organisation truly inspiring. I salute all the team members. I had a great time working amongst them. I would like to salute my thesis advisors Jens Peter Kargo and Lars Arknaes for initiating and orienting this work. It was truly inspiring and a privilege to work with them. More generally, I would like to thank them for making sure that I felt welcomed at TC, and at home in Aarhus. In that regard, special thanks are also due to "weltmeisters" Mike Nickel and Matthias Buhler. I warmly salute them. Thanks also to Chris, although Lars, Jens Peter and me only met him once: he was kind enough to offer a round of beer to three random engineers with a pair of nightshades on a dark and deserted Tuesday evening in Aarhus, which is surely worth a mention!

I thank Vesa Välimäki, my thesis supervisor at Aalto. Always clear, concise and constructive, his advices directed me towards a clearer and more streamlined thesis structure. He also helped me to provide a discussion of my subject that is more complete technically. I thank him again for taking the time to go through my draft in such an extensive fashion, despite it not being an easy read.

I would like to thank my familly for supporting me during my graduate studies. Finally, I warmly thank Heta for encouraging me throughout the stages of my thesis writing. I expect she might not believe that it is now trully complete!

Cupertino, 19.01.2016

Etienne Thuillier

Contents

Symbols and Abbreviations

Symbols

$\tilde{}$	Identifies non-heterodyned passband version of sub-band signal
$\hat{}$	Identifies analytic version of a time domain signal
$_\mathrm{a}$	Identifies analysis-related variable
A	Amplitude
BW_ERB	Effective output bandwidth (in ERB units)
d	Latency
D	Distortion (ratio)
Δf	Bandwidth specification
f	Subband modulation function
F	Synthesis filter, in particular synthesis window (frequency domain)
f_c	Center frequency of a filter or sub-band
f_s	Sampling frequency
G_a	Analysis crossover attenuation
GD	Group delay
G_stop	Stopband attenuation
G_t	Target crossover attenuation
h	Analysis filter, in particular analysis window (time domain)
H	Analysis filter, in particular analysis window (frequency domain)
$\mathcal{H}\{.\}$	Hilbert Transform
k	Subband index
K	Number of filter bank sub-bands
K_BQ	Parameter for bi-quad filter coefficient value determination
L	Window bandwidth from 0 Hz to first zero crossing (in bins)
L_BH	Blackman-Harris windows family parameter
m	Modulator function
n	Discrete time
N	Window length (in samples)
O	Overlap factor
q	Quality factor
q_C	Sub-band distance filter bank design parameter
q_ERB	Constant-ERB-bandwidth design parameter
q_PS	Sub-band frequency scaling factor
R	Downsampling factor (a.k.a. hop size)
$_\mathrm{s}$	Identifies synthesis-related variable
t	Continuous time
U	Audio-rate input sub-band signal (frequency domain)
V	Audio-rate output sub-band signal before interp. filter (freq. domain)
W_N	$\mathrm{e}^{-i2\pi/N}$
x	Input signal (time domain)
X	Input signal (frequency domain)
y	Output signal (time domain)
Y	Output signal (frequency domain)
z_left	Leftmost filter center frequency in ERB units

γ	Pitch shifting factor
τ	Downsampled discrete time (a.k.a. frame index)
θ	Phase
π	Pythagoras's constant
ω	Angular frequency (rad/s)

Abbreviations

ADC	Analog-to-Digital Converter
BLFWR	Band-Limited Full-Wave Rectifying
COLA	Constant Overlap Add
CPO2	Power of 2 (analytic signal)
DAC	Digital-to-Analog Converter
DFT	Discrete Fourier Transform
DSP	Digital Signal Processor
FWR	Full-Wave Rectifying
ERB	Equivalent Rectangular Bandwidth
FIR	Finite Impulse Response
FFT	Fast Fourier Transform
GPGPU	General Purpose Graphics Processing Unit
HWR	Half-Wave Rectifying
IIR	Infinite Impulse Response
ISTFT	Inverse Short-Time Fourier Transform
PSOLA	Pitch Synchronous Overlap and Add
PO2	Power of 2 (real-valued signal)
PS2	Phase Scaling for octave doubling
PV	Phase Vocoder
RM2	Ring-Modulation for octave doubling
SOLA	Synchronous Overlap and Add
SSM2	Single Sideband Modulation for octave doubling
STFT	Short-Time Fourier Transform
THD	Total Harmonic Distortion
WOLA	Weighted Overlap and Add

2 Introduction

A pitch shifting effect can easily be achieved by way of resampling an audio input and playing it back according to the original sample rate, much in the same manner as with a tape recording played at accelerated or reduced playback speed [18]. However, a contraction or extension of the input signal's duration also result from such operation. For real-time applications such contraction is workable only if it can be reversed within the real-time delay constraint by a corresponding expansion, or vice-versa, as in the case of the pitch vibrato effect[1]. By contrast, "pitch shifting" commonly designates methods which preserve the time duration of the signal such that an upwards or downwards pitch transposition effect can be sustained for an extended amount of time. Providing a pitch transposition effect at low latency is renowned to be a challenging task, especially for polyphonic audio inputs. This work studies the special case of real-time polyphonic octave doubling for the guitar.

Traditionally, pitch shifting methods are classified in two categories: time domain and frequency domain techniques. Most time domain techniques achieve pitch shifting through modification of the audio playback speed[2] and compensate the resulting time scale contractions or expansions by way of discarding or repeating waveform segments of the audio input during playback such as in [48], [41], [25], [18]. Frequency domain techniques rely on the short-time Fourier tranform (STFT) and achieve a pitch shifting effect at least in part by way of modifications brought to the Fourier transform bins. More specifically, these modifications aim to scale the instantaneous phase of each Fourier sub-band while preserving frame-to-frame phase coherence in the un-stretched output signal [23], [42], [12], [43], [36], [35].

The pitch shifting techniques mentioned above lead to artefacts that are specific to each categories of methods. An abundance of literature addresses this topic, e.g. [43], [34], [35]. However, the amount of latency incurred from these techniques seem often considered a secondary issue and is rarely tackled, e.g. see [30], [37]. Moreover, few publications aim specifically to provide short-delay implementations for real-time applications such as live audio effects, e.g. see [24]. This is especially the case of implementations applicable to polyphonic content with a relatively wide register such as that of the guitar. However, two such solutions were recently proposed by Juillerat in 2008 and 2010 [29], [28]. A similar observation can be made of the industry's offerings: effect pedals for the guitar carrying out single octave downwards pitch-shifting of monophonic melodic lines have been introduced a relatively long time ago, as early as 1982 in the case of Boss's OC-2 pedal [9]. But octave doubling of guitar chords with low latency and in a robust fashion has only been introduced recently in the form of the POG effect pedals, first commercialised by Electro-Harmonix in 2005 [22].

The current Master's thesis work primarily aims to identify polyphonic pitch shifting techniques of the art with a potential to provide an octave doubling effect in

[1] A vibrato effect of a given amplitude can be achieved with a fractional delay line of appropriate length as determined from a predetermined pitch modulation amplitude [20], [31].

[2] PSOLA does not rely on changing the playback speed, although it also implies repositioning of waveform segments one with regards to the other.

real-time for the guitar instrument. Like Electro-Harmonix's line of POG products, the intended application of the effect is to provide glitch-less[3] tonal enrichment of guitar chords and melodies performed on the instrument by providing a pedal output signal formed by a mix of the unprocessed input guitar signal with the octave-shifted version of it. Detailed requirements follow.

2.1 Low Latency Requirement

Low latency is required from the pitch shifting process to avoid the generation of an echo effect which is not intended to be part of the modified output sound in a perceptible fashion. The literature suggests a latency below 10 ms is required for percussive sound [32]. The guitar is a versatile instrument which allows various style of playing some of which include rhythmic ornamentations and sharply plucked notes. It can thus be assumed that this maximum latency requirement, formulated for percussive instruments, also apply to the guitar. However, this ideal low latency requirement can be relaxed to meet a lower performance standard found the case being in tested POG pedals.

2.2 Steady-State Sound Quality Requirement

Such low latency must be achieved within a level of artefacts introduced in the output sound by the pitch shifting stage. Specifically, the effect should not introduce perceivable mistuned harmonic components. Moreover, it should ideally not add more dissonance than that produced by an equivalent doubling of octaves as performed on the instrument by the guitar player.

2.3 Transient Preservation Requirement

As a result from mixing the processed signal with the dry input guitar signal, detrimental effects, introduced by the pitch shifting stage on signal transients are expected to be perceptively masked at the pedal's output. However, potential buyers of such pedals typically test the quality of the pitch shifting effect taken individually by way of setting an effect level knob to a 100% wet position, whereby only the pitch-shifted signal is transmitted at the output. In light of this commercial consideration, the effect should introduce as little alterations as possible to the transient components of the guitar sound.

2.4 Thesis Structure

A scientific background review for achieving low latency octave doubling is provided in the next three sections of the thesis. Section 3 presents essential scientific background on the short-time Fourier transform and decimated uniform filter-banks. A

[3] Techniques relying on explicit pitch estimation and tracking are disregarded by this study, in part for latency considerations but also because pitch estimation, and even more so polyphonic pitch estimation, is inherently glitch-prone (at least in the case of the guitar).

review of selected pitch shifting techniques of the literature is presented in Section 4, including time domain techniques, frequency domain techniques and Juillerat's Rollers method, which is based on a non-decimated constant-Q IIR filter bank. Section 3 serves as a foundation to the presentation of the frequency domain techniques as well as to the discussion of their incurred latencies. Two novel pitch shifting solutions are proposed in Section 5. It combines the sub-band phase scaling approach found in phase vocoder methods with a non-decimated constant-ERB-bandwidth complex IIR filter bank.

Section 6 presents four selected candidate solutions (amongst which said novel solutions) and proposes evaluation methods for comparing their performance in terms of the above mentioned requirements with that of three octave doubling effect pedal models from Electro-Harmonix. The results of this evaluation are presented and discussed in Section 7.

3 Analysis-Synthesis with the Short-Time Fourier Transform

3.1 Maximally Decimated Filter Banks

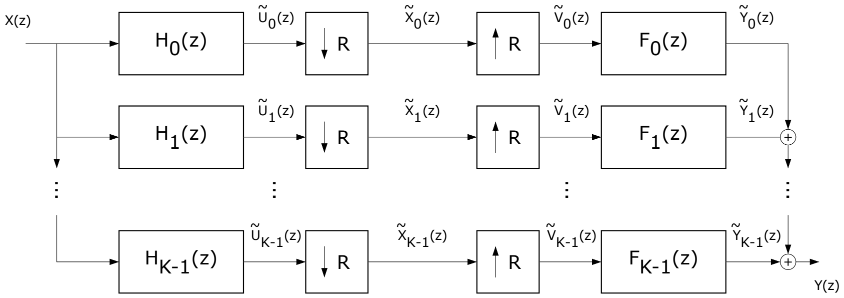


Figure 1: Block diagram of a typical decimated filter bank.

Decimated Filter Bank Definition A typical example of a decimated filter bank is represented in Figure 1 [54]. As pictured, a bank of narrowband filters splits an input signal $x(n)$ in K sub-band channels. Formally, these filters are called analysis filters and are said to form an analysis bank. As pictured, downsampling is applied at the output of each filter to minimise the bit rate of the system. In particular, the principles of undersampling (also known as bandpass sampling) apply for non-lowpass narrowbands of the bank, thus allowing for reconstruction of the input signal downstream despite fold-over of the band around the Nyquist frequency. Reconstruction, is achieved by upsampling the sub-band signal back to the input rate and recombination through adder operators. In particular, the upsampling operation is accomplished in each channel by an upsampler paired to a filter. These filters are called synthesis filters and are said to form a synthesis bank.

Maximally Decimated Filter Bank A bank is maximally decimated when the downsampling factor R is equal to the number K of bands of the filter bank [54], [40]. In such case, the overall bit rate at the output of the analysis bank is equal to that of the input signal. Figure 2 illustrates the magnitude spectrums of the signals at each stage of the processing chain of a first design example of such a filter bank. In this example, the analysis and synthesis filters are designed according to the same passband-stopband attenuation and transition bandwidth (see 2-b). In particular, the transition band is made sufficiently narrow so as to prevent any overlapping with the aliased images of the sub-band resulting from the downsampling-upsampling operation (see 2-d and 2-e). Such a design succeeds in preventing aliasing in the reconstructed signal (see 2-g). As pictured, said reconstructed signal unfortunately presents significant dips at the junction of the bank's

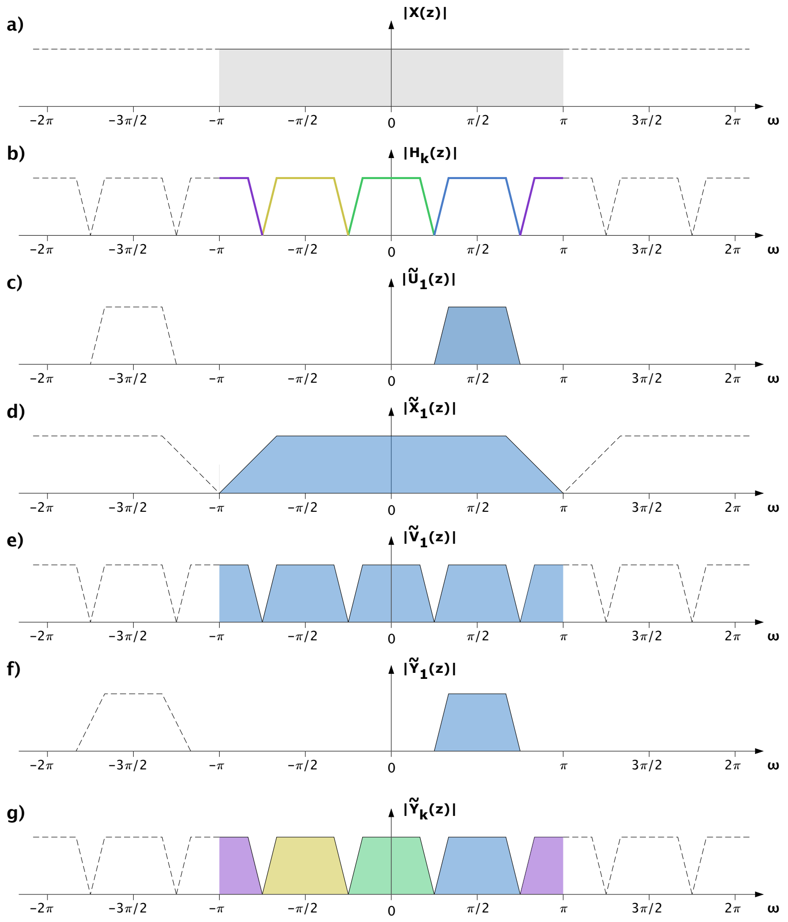


Figure 2: Magnitude spectrum at each stage of the processing chain of a maximally decimated filter bank with non-overlapping transition bands. Figure inspired from [54].

filters. As noted by Vaidyanathan [54], boosting the transition bands back up to the passband level could be achieved in theory by appropriate design of the synthesis filters. As he notes however, this would lead to drastic boosting of noise level as well. Specifying near-ideal transition bands is not practical either as it would require overly long filters and result in high latency.

Aliasing Cancellation Figure 3 illustrates a second design example of maximally decimated filter bank. As pictured, the analysis filters are designed with a broader transition band such that overlapping is allowed to occur (see 3-b). In this fashion no energy of the input signal is lost at the output of the analysis bank, by contrast to the example of Figure 2. However, aliasing is allowed to occur following downsampling as pictured in the Figure (see 3-d). Despite this, perfect reconstruction is nevertheless achieved at the output of the example as aliased components generated

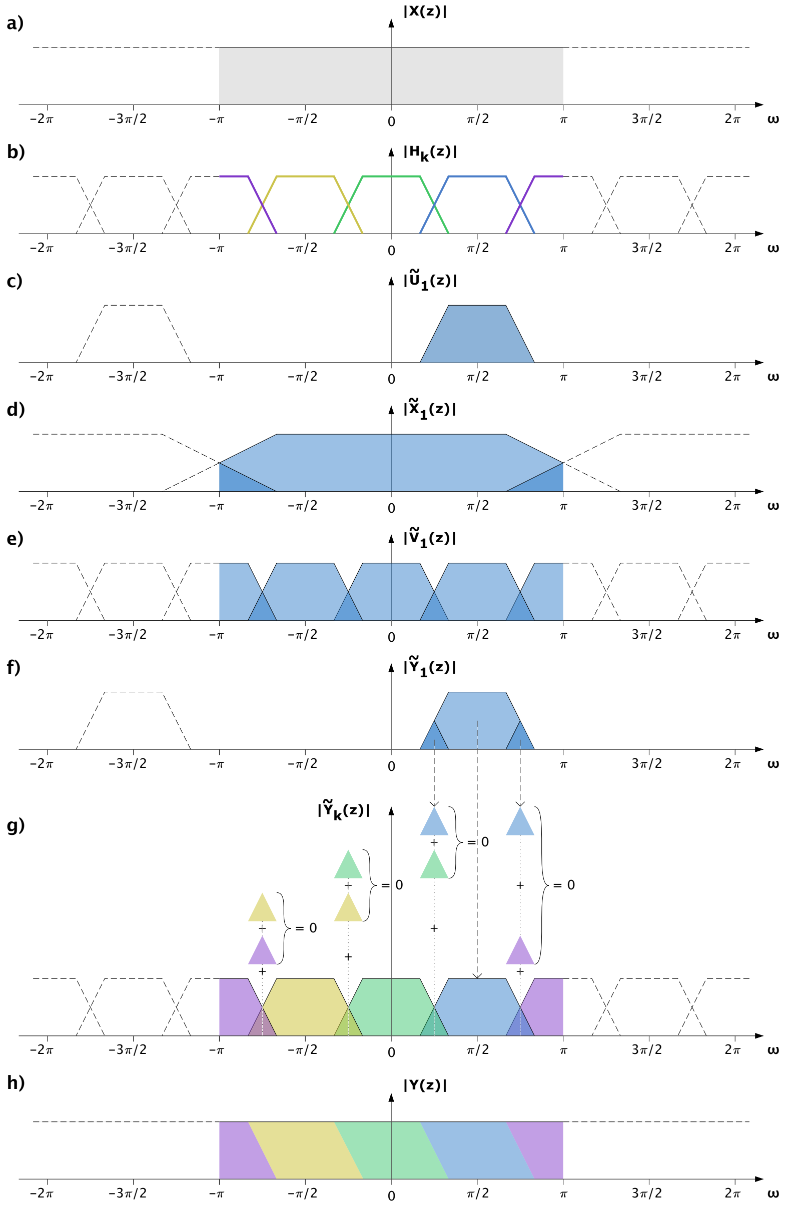


Figure 3: Magnitude spectrum at each stage of the processing chain of a perfect reconstruction maximally decimated filter bank. Figure inspired from [54].

in neighbouring channels and located in shared overlapping transition bands cancel out two-by-two as pictured in the Figure (see 3-g). This phenomenon is called aliasing cancellation. Its realisation requires appropriate design of the synthesis filters, the non-trivial conditions of which are thoroughly described in the literature [54].

Oversampled Filter Banks Achieving robust implementations of audio effects generally requires using oversampled filter banks, that is: decimated filter banks in which the downsampling factor is smaller than the number of sub-band channels, that is $R < K$ [51][4], [40]. Indeed, modifications brought to sub-band signals between the analysis and synthesis banks typically break the channel-to-channel relationship upon which aliasing cancelation depends. This is particularly the case of non-linear modifications such as for example those carried-out by the modulation techniques listed in table 3 (see Section 5). Carrying out sub-band processing at a higher rate allows to prevent aliasing from occurring at the output of the downsampler in the first place, a non-aliased reconstructed being thus achievable without relying on aliasing cancelation. The strong COLA requirement presented in Section 3.3.3 ensures that the downsampled sub-band does not overlapp with its aliases (unlike in Figure 3-d, see also Figure 7). In practice, an even higher processing rate is often required to provide a guard band that prevents aliasing to result from the non-linearities of the sub-band processing the case being.

3.2 Short-Time Fourier Transform

Classical Definition The classical definition of the Short-Time Fourier Transform (STFT) is given by [2], [54], [1], [51][5]:

$$X(m, \omega_k) = \sum_{n=-\infty}^{\infty} x(n)h(n - mR)\mathrm{e}^{-i\omega_k n}, \tag{1}$$

where R denotes the hop size of the STFT. In this definition, the input signal and complex exponential are fixed in time and the analysis window is shifted independently. This leads to the complex baseband interpretation of the STFT [2][6]. This interpretation is represented in Figure 4 (left), in which:

$$W_N = \mathrm{e}^{-i2\pi/N}. \tag{2}$$

As pictured for a single band k of the STFT, the product of input signal with the complex exponential corresponds to a frequency shift of the (real-valued) input signal by an amount corresponding to the band's center frequency ω_k. This intermediate signal is then lowpass filtered through window $h(n)$. As pictured, the result is a complex baseband signal.

Alternative Definition Another definition is often used, notably by Laroche [35]. It is that of the non-heterodyned STFT:

$$\tilde{X}(m, \omega_k) = \sum_{n=-\infty}^{\infty} x(n + mR)h(n)\mathrm{e}^{-i\omega_k n}. \tag{3}$$

This definition leads to the complex passband interpretation of the STFT [54][7]. This interpretation is represented in Figure 4 (right). In contrast to the complex

[4] http://ccrma.stanford.edu/~jos/sasp/Review_STFT_Filterbanks.html

[5] http://ccrma.stanford.edu/~jos/sasp/Mathematical_Definition_STFT.html

[6] p. 223

[7] pp. 466 & 480

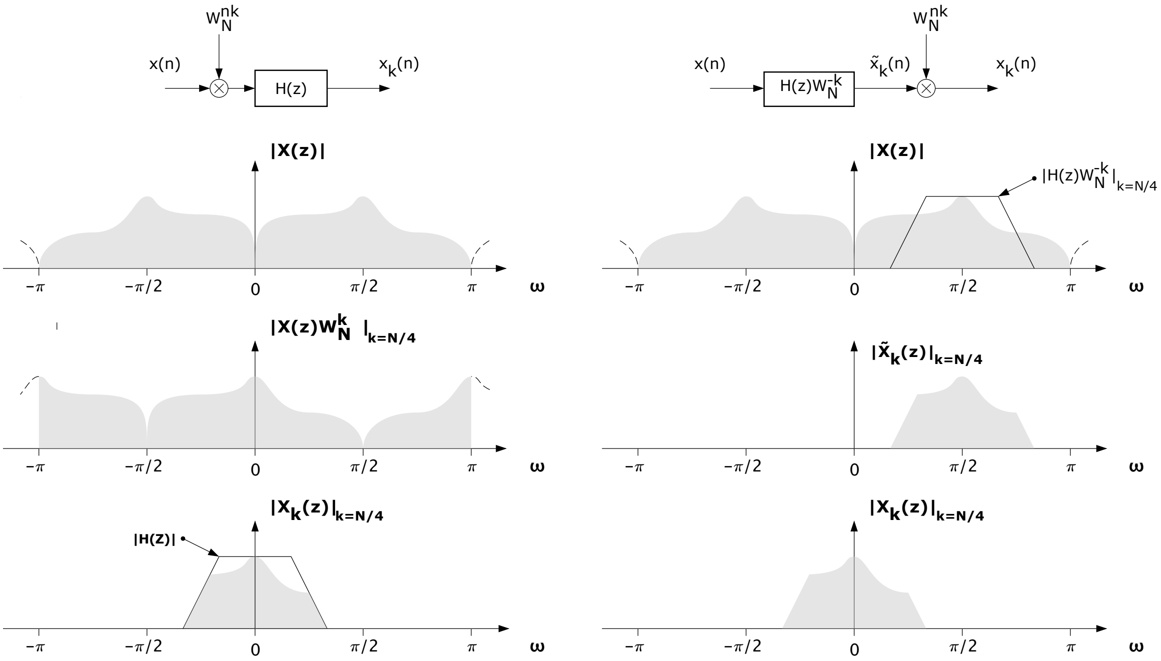


Figure 4: Complex baseband (left) and complex passband (right) block diagram representations of the STFT ($R = 1$). Corresponding frequency representation for the $k = N/4$ Fourier band. Inspired by [54] (pp. 465-467) and [2] (p. 223).

baseband interpretation, the input signal is time-shifted while the analysis window and the complex exponential are fixed in time. The window and exponential thus form a complex narrowband filter into which the input signal is filtered, resulting in a passband signal and identified as such by the tilde marking in the non-heterodyned STFT definition of Equation (3). In the complex passband representation of Figure 4 (right), said passband signal is frequency modulated to baseband so as to ensure equivalence with the complex baseband representation and classical definition of the STFT given by Equation (1).

3.3 The STFT Analysis-Synthesis Scheme

3.3.1 Decimated Filter Bank Representations

The Inverse Short-time Fourier Transform (ISTFT) is carried out by reversing, for each frequency band k, the operations accomplished to produce the complex baseband or complex passband representations described above. Cascading of STFT and a ISTFT yields a system represented in Figure 5. Such a system performs a frequency component analysis of the incoming signal followed by a re-synthesis of said input signal by way of recombination of its sub-band components extracted at the analysis stage.

An intermediate frequency domain processing of the sub-bands can be accom-

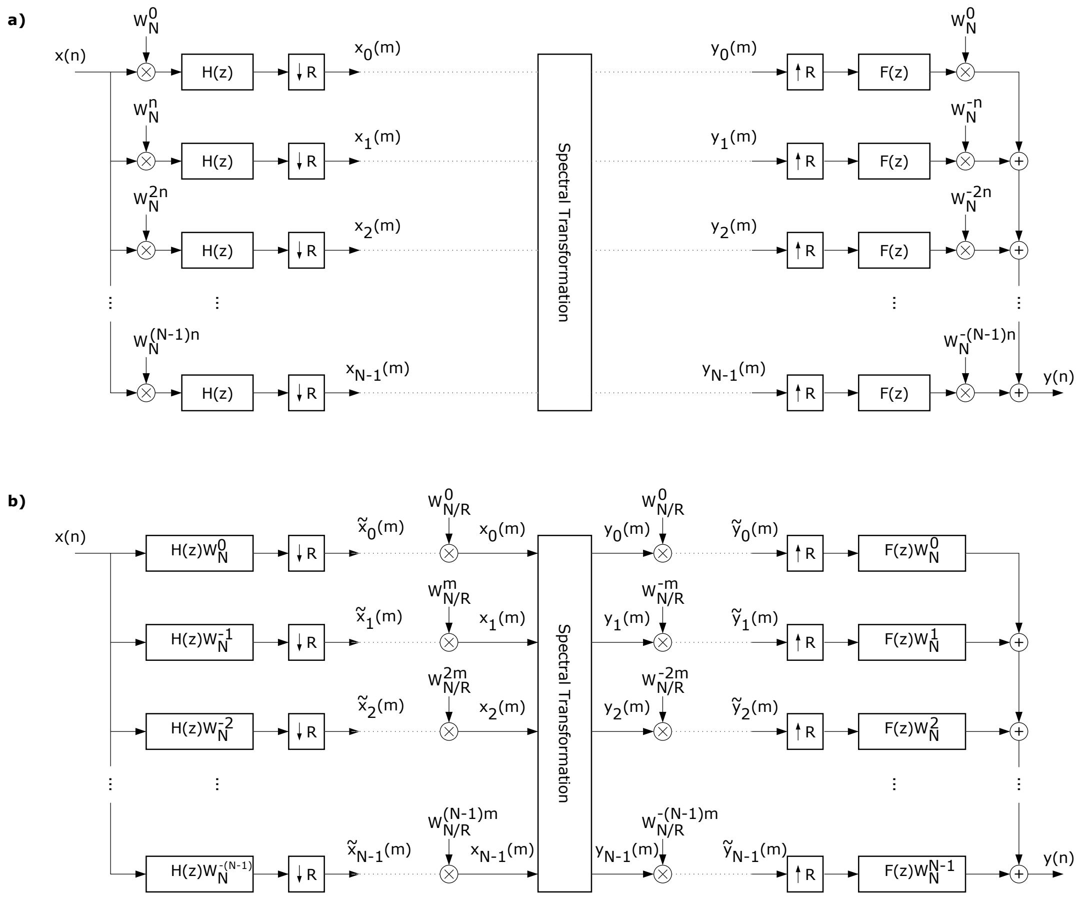


Figure 5: Analysis-synthesis with the STFT. a) Following complex baseband representation of STFT b) Following complex passband representation of STFT. Inspired by [2] (p. 223).

plished between the analysis and synthesis stage. The resulting analysis-transform-synthesis architecture serves as a foundation for numerous audio signal processing techniques, of which phase vocoder time stretching and phase vocoder pitch shifting.

As discussed above the purpose of the frequency shift terms $W_{N/R}^{km}$ in Figure 5 b) is to modulate the complex passband signal to baseband. These optional terms results from the difference between definitions (1) and (3). It is employed in the figure to ensure equivalence of the frequency representations at the entrance of the spectrum transformation block (Figure 5-a and 5-b). This step is typically discarded in practical applications and a transformation is carried out directly on the passband signals $\tilde{X}(m, \omega_k)$.

3.3.2 Weighted Overlap-Add Representation

In practice, the STFT can be implemented by a uniform discrete Fourier transform (DFT) filter bank in which, in the absence of zero-padding, the number of sub-band

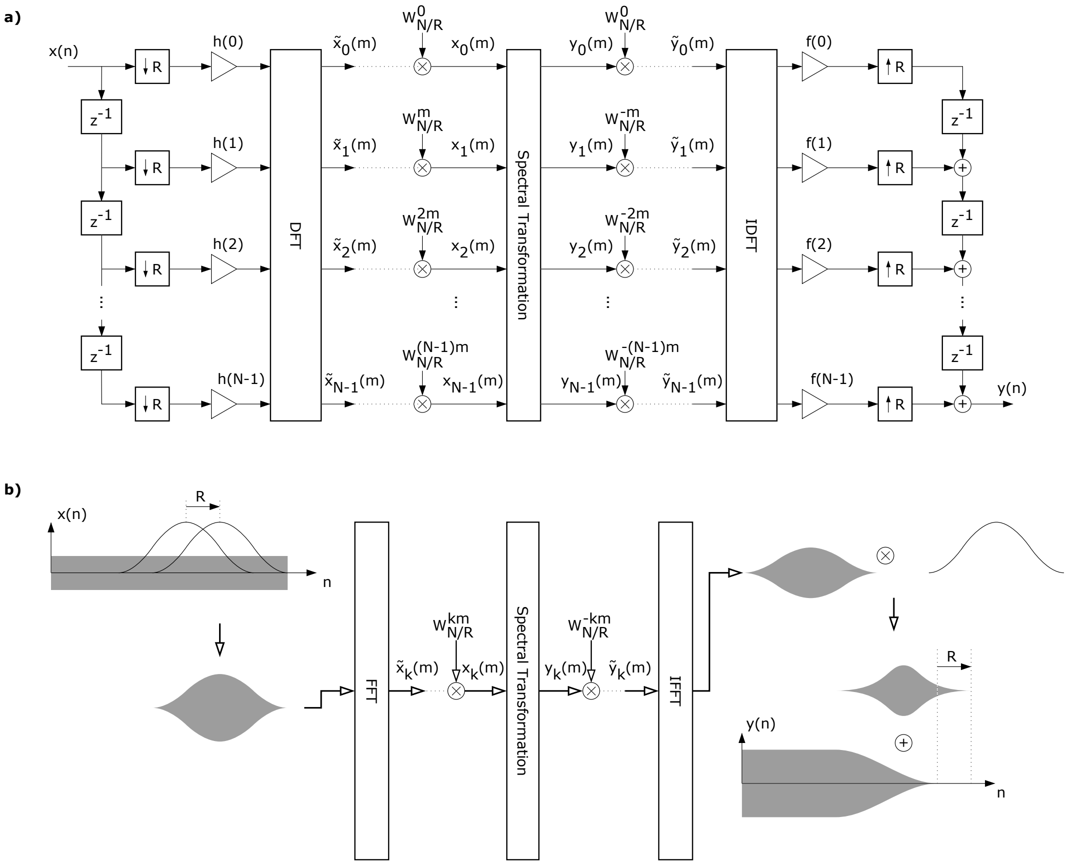


Figure 6: a) Uniform DFT filter-bank derived from the complex passband representation of the STFT found in Figure 5-b. The DFT and IDFT blocks follow non-causal definitions in this schematic representation. b) Equivalent Weighted Overlap-Add (WOLA) representation with a causal FFT algorithm. Inspired by [2] (p. 225).

channels K of the bank is equal to the number of window coefficients N. Correspondingly, the uniform DFT filter bank representation of Figure 6-a can be derived from the simple filter bank topology of Figure 5-b by application of polyphase decomposition and noble identities [54]. It follows that the coefficients of the window are distributed vertically in Figure 6-a. Inspection reveals this representation to be equivalent to the Weighted Overlap-Add representation (WOLA) [12], [1], represented in Figure 6-b and further discussed in what follows.

The WOLA representation of the STFT is pictured in the schematic diagram of Figure 6-b. It consists in a practical implementation of the STFT using the discreet Fourier transform (DFT) on a block-by-block basis for the processing of the input signal. Where N denotes the analysis window length. Thus the STFT is carried out by DFT computation over successive segments, called blocks or frames, of the input segment. Each input frame is resynthesized by means of the inverse discrete

Fourier transform (IDFT) and recombined in an overlap-add fashion at the output of the system.

As represented in Figure 6, the analysis-transformation-synthesis scheme is performed according to a step size, hereafter called hop size, whose value actually corresponds to the downsampling factor R. Increasing the hop size is thus equivalent to applying greater downsampling to the sub-bands. Conditions on the window and hop size regarding aliasing are discussed in the following Section. Advantageously, the DFT can be implemented using a Fast Fourier Transform (FFT) algorithm for reduced computation complexity [8].

3.3.3 Choice of Window and Overlap

Constant Overlap-Add Assuming no spectral transformation is performed in the sub-band, the analysis-synthesis scheme provides aliasing cancellation and perfect reconstruction under the Constant Overlap-Add (COLA) requirement formulated by Smith [51][8]:

$$\sum_{m=-\infty}^{\infty} h(n-mR) = 1 \qquad \forall n \in \mathbb{Z}. \tag{4}$$

The time-shifted replications of the window in WOLA can be interpreted as a time-domain aliasing phenomenon resulting from a sampling of said window in frequency [51][9]. This is expressed by the Poisson Summation Formula:

$$\sum_{m=-\infty}^{\infty} h(n-mR) = \frac{1}{R}\sum_{k=0}^{R-1} H(\omega_k)\mathrm{e}^{i\omega_k n} \qquad \omega_k = \frac{2\pi k}{R}, \tag{5}$$

where the right handside expression can be interpreted as the IDFT of the window's frequency response sampled with step $2\pi/R$.

Weak COLA The above mentioned time-domain aliasing phenomenon simplifies upon recombination to unit value, and thus meets COLA requirement (4), when said sampled frequency response is nill for all nonzero multiples of $2\pi/R$, i.e. if and only if the frequency sample values are nil except at $k = 0$, or formally [51][10,11]:

$$\sum_{m=-\infty}^{\infty} h(n-mR) = 1,\ \forall n \in \mathbb{Z} \ \Leftrightarrow\ H(\omega_k) = 0,\ |k| = 1, 2, 3 ... R-1. \tag{6}$$

$$\sum_{m=-\infty}^{\infty} h(n-mR) = \frac{1}{R}H(0). \tag{7}$$

This condition is met by the window of examples a) and b) of Figure 7.

[8] http://ccrma.stanford.edu/~jos/sasp/Mathematical_Definition_STFT.html

[9] http://ccrma.stanford.edu/~jos/sasp/Poisson_Summation_Formula.html

[10] http://ccrma.stanford.edu/~jos/sasp/Frequency_Domain_COLA_Constraints.html

[11] http://ccrma.stanford.edu/~jos/sasp/Strong_COLA.html

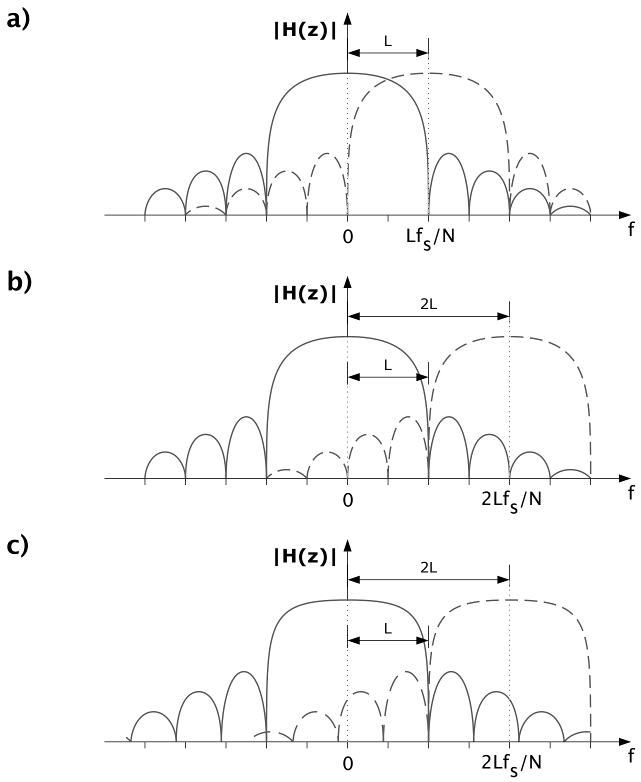


Figure 7: Magnitude response of window examples (full line) and their respective closest rightwards aliased image (dashed line) for maximum value of downsampling factor R as determined with: a) Weak COLA criterion $R = \frac{N}{L}$. b), c) Relaxed strong COLA criterion $R = \frac{N}{2L}$. L denotes the window's lowpass bandwidth in frequency bins as measured at the first zero crossing.

Strong COLA Of course such a condition is all the more ensured when the frequency response of the window is ideally bound to the downsampled Nyquist band $[-\pi/R, \pi/R]$, or formally [51][12]:

$$H(\omega) = 0, \qquad |\omega| \geq \frac{\pi}{R}. \tag{8}$$

Relaxed Strong COLA In practice, the strong COLA requirement can be partially met by means of window designs specified according to predefined stopband gains, such that the main lobe of the downsampled sub-band signal do not overlap with that of its aliased image, but nevertheless receives interference from the stopband components of said aliased image. Thus the author proposes a relaxed formulation of Smith's strong COLA criterion (8) as follows:

$$H(\omega) \leq G_{\text{stop}}, \qquad |\omega| \geq \frac{\pi}{R}. \tag{9}$$

[12] http://ccrma.stanford.edu/~jos/sasp/Strong_COLA.html

where G_{stop} denotes the stopband attenuation. This condition is met by examples b) and c) in Figure 7. The relaxed strong COLA requirement ensures that perfect reconstruction is at least approximately met.

Choice of window Perfect reconstruction depends both on the window(s) and downsampling factor used in the STFT analysis-synthesis system [51]. Figure 7 represents the magnitude response of two windows of identical main lobe bandwidth. The first window example has zero crossings uniformly distributed on the frequency axis, unlike the second window example. Two downsampling factors are considered for this first window. The weak COLA criterion provides a first downsampling value $R = \frac{N}{L}$ which results in the frequency representation in Figure 7-a. This value corresponds to a higher bound under which perfect reconstruction can be ensured for this specific window. The relaxed strong COLA criterion provide a more conservative bound of value $R = \frac{N}{2L}$ which prevents overlap between the main lobe of the window and that of its aliased image. The corresponding frequency representation is given in 7-b. It can be noted that the weak COLA requirement is ensured in this case as well, although with higher sampling rate in the sub-band compared to 7-a. By contrast, the second window example also meets the relaxed strong COLA requirement with said conservative value $R = \frac{N}{2L}$ as pictured in Figure 7-c. It thus suppresses aliased frequency components down to the level of its stop-band. However it does not meet the weak COLA criterion and therefore only approximates perfect reconstruction when $R > 1$.

Blackman-Harris Window Family Like the first window example of Figure 7-a and 7-b, the windows of the Blackmann-Harris family present uniform distribution of their zeros on the unit circle as required by the Weak COLA condition, and thus form admissible candidates when perfect reconstruction is sought [26], [51][13]. The coefficients of each window of the family is determined by the following general expression [51][14]:

$$h_{\text{BH}}(n) = h_{\text{R}}(n) \sum_{l=0}^{L_{\text{BH}}-1} \alpha_l \cos\left(l \frac{2\pi n}{N}\right), \tag{10}$$

where $h_{\text{R}}(n)$ is a rectangular window:

$$h_{\text{R}}(n) = \begin{cases} 1 & n \in [-\frac{N-1}{2}, \frac{N-1}{2}] \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Parameter L_{BH} controls the number of summation terms in the definition. In particular:

- $L_{\text{BH}} = 1$ yields the rectangular window,
- $L_{\text{BH}} = 2$ yields the generalized Hamming family of windows,

[13] http://ccrma.stanford.edu/~jos/sasp/Example_COLA_Windows_WOLA.html

[14] http://ccrma.stanford.edu/~jos/sasp/Blackman_Harris_Window_Family.html

- $L_{\mathrm{BH}} = 3$ yields the Blackman window family.

Table 1 gives the coefficient value for well-known windows of the Blackman-Harris family, amongst which the Hann, Hamming and Blackman windows. A summary of applicable higher bound values for R following the COLA criterions is given in Table 2. Part of these values were determined by the author while a significant number are reproduced from [51].

Window	L_{BH}	α
Hann	2	$\alpha_0 = 0.54,\ \alpha_1 = -0.46$
Hamming	2	$\alpha_0 = 0.5,\ \alpha_1 = -0.5$
Blackman (classic)	3	$\alpha_0 = 0.42,\ \alpha_1 = 0.5,\ \alpha_2 = 0.08$

Table 1: Examples of windows of the Blackman-Harris family [51].

Single Window	L_{BH}	Weak COLA	Overlap	Strong COLA	Overlap
Rectangular	1	$R \leq N - 1$	$\geq 0\%$	$R \leq (N-1)/2$	$\geq 50\%$
Hann	2	$R \leq (N-1)/2$	$\geq 50\%$	$R \leq (N-1)/4$	$\geq 75\%$
Hamming	2	$R \leq (N-1)/2$	$\geq 50\%$	$R \leq (N-1)/4$	$\geq 75\%$
Blackman	3	$R \leq (N-1)/3$	$\geq 66.6\%$	$R \leq (N-1)/6$	$\geq 83.3\%$
Blackman-Harris	4	$R \leq (N-1)/4$	$\geq 75\%$	$R \leq (N-1)/8$	$\geq 87.5\%$

Table 2: Weak and Strong COLA criterions as determined from parameter L_{BH} of window examples of the Blackman-Harris family.

Combined Analysis and Synthesis Windows In the case of a filter bank comprising both an analysis window and a synthesis window, the COLA conditions must be met for the frequency response of the combined (i.e. multiplied) window functions [51][15]. When both windows are not rectangular, this typically leads to lower admissible values for the downsampling factor. For example, perfect reconstruction is ensured for downsampling values under $R = N/3$ in the case where Hann windows of equal duration are used upon analysis and synthesis [51][16]. This justifies usage of the overlap ratio of 75% often mentioned in the literature in such case [44][17]. Meeting the relaxed Strong COLA criterion requires further halving the downsampling factor to $R = N/6$. Another design solution consist in taking the root values of the coefficient of a COLA-compliant window such as one listed in Table 2 to form the analysis and the synthesis window [51][18].

[15] http://ccrma.stanford.edu/~jos/sasp/PSF_Weighted_Overlap_Add.html

[16] http://ccrma.stanford.edu/~jos/sasp/Example_COLA_Windows_WOLA.html

[17] p. 276

[18] http://ccrma.stanford.edu/~jos/sasp/Choice_WOLA_Window.html

3.3.4 Incurred Delay

Group Delay Group delay represents the amount of delay in seconds suffered by a frequency component's amplitude envelope. In this definition, said amplitude envelope is band-limited such that its width covers the a frequency band around ω where the phase response of the Linear Time Invariant (LTI) system is approximately linear [49][19]. It is formally defined as:

$$GD(\omega) \triangleq -\frac{d}{d\omega}\theta(\omega), \tag{12}$$

where $\theta(\omega)$ denotes the phase response of the LTI system. In particular, group delay is constant across the frequency scale for linear phase Finite Impulse Response filters (FIR) and corresponds to half the order of the filter :

$$GD = \frac{N-1}{2}, \tag{13}$$

where N denotes the number of coefficients of the filter.

Audio Rate Filter Bank Inspection of Figure 5 reveals that, in the absence of spectral modifications, the delay incurred in each sub-band corresponds, in the absence of downsampling (i.e. when $R = 1$), to the cumulated group delays of the corresponding analysis and synthesis filters:

$$GD_{\text{fb}}(k,\omega) = GD_{\text{a}}(k,\omega) + GD_{\text{s}}(k,\omega). \tag{14}$$

In the case where the analysis lowpass prototype filter $H(z)$ and the synthesis lowpass prototype filter $F(z)$ are formed by symmetric windowing function, the above equation simplifies to :

$$GD_{\text{fb}} = GD_{\text{a}} + GD_{\text{s}}. \tag{15}$$

In the case where the windows are of equal lengths, the overall delay thus corresponds approximately to window duration:

$$GD_{\text{fb}} = N - 1. \tag{16}$$

Downsampled Filter Bank In the presence of downsampling (i.e. when $R \neq 1$), the filter bank is a Linear Periodically Time-Variant (LPTV) system with period R [40][20]. However, when the weak COLA requirement is met for the analysis-synthesis window pair, perfect reconstruction is ensured and the filter bank simplifies to a LTI system. Its group delay then corresponds approximately to the window duration as above.

[19] https://ccrma.stanford.edu/~jos/fp/Group_Delay.html

[20] p. 17

4 Pitch Shifting Techniques

4.1 Time Domain Techniques

Most time domain techniques, with the notable exception of PSOLA, achieve pitch shifting through accelerated or reduced audio playback speed and compensate the resulting time scale contractions or expansions by way of discarding or repeating waveform segments of the audio input during playback.

Ring Buffer Technique A simple solution consists in saving the digital audio input in a ring buffer following a revolving memory address while simultaneously providing a playback output read from another revolving memory address pointer progressing at a different speed, as proposed by Lee [38]. In this solution, the splicing of the waveform segments is carried out independently from the audio input content according to the difference of record and playback speeds: A waveform segment is discarded or repeated each time one memory address pointer overtakes the other. Upon the moment where splicing happens a “very objectionable discontinuity” occurs according to Lee.

SOLA Methods have been proposed to minimise the effect of such splicing artefacts. In particular, overlapping segments are employed in Synchronous Overlap Add (SOLA) methods first introduced by Roucos, Makhoul et al. [48], [41]. An extension or contraction of the input signal duration is achieved in these methods by repositioning each overlapping segment with regard to its neighbour in a telescopic fashion. Splicing is then conducted within each overlapping portion of the repositioned segments along an interval corresponding to the highest similarity between the spliced segments. In practice, this is achieved by way of a fine-tuned position reajustment to synchronize the pair of segments’ similar portions to said interval. Various measures of similarity have been proposed, a comparison of which is given by Dorran [16].

PSOLA Another notable solution is the Pitch-Synchronous Overlap and Add (PSOLA) method proposed by Moulines and al. [25], which was proposed specifically for processing of the human voice. It consists in determining glottal pulse timings with the assistance of a pitch detector and modifying the instantaneous period of the audio input by displacing (and eventually discarding or duplicating) waveform segments, each containing one glottal pulse, according to pitch-scaled glottal pulse timings. Pitch shifting is thus achieved without relying on resampling, unlike SOLA (see above). This yields a significant advantage as the spectrum enveloppe is preserved, thus avoiding the “chipmunk” effect which is characteristic of methods relying on resampling.

Real-Time Implementation An appealing real-time implementation of a time domain pitch shifting method related to SOLA has been proposed by Haghparast et al. [24]. The proposed method is based on Lee’s ring buffer technique. The signal read from the ring buffer is resampled on-the-fly according to a real-time user definable pitch factor parameter. Splicing artefacts are minimised by way of forcing the read pointer to jump backwards so as to prevent it from overtaking the write pointer, or forward so as to prevent the write pointer from overtaking it. When repositioned in such way, the read pointer is made to land in an area of the ring buffer memory bearing a high level of similarity to the area it has left and a cross fade is performed, as in traditional SOLA method.

Applicability to Polyphonic Octave Doubling Time domain methods, especially SOLA techniques, are particularly efficient but tend to produce transient skipping or repetition, especially for large pitch shifting factors [39]. Moreover, these methods explicitly (PSOLA) or implicitly (SOLA) rely on the assumption that the audio input is periodic at least to some degree which make them a priori ill-fitted for polyphonic pitch transposition. High quality polyphonic audio pitch transposition is actually achievable but limited to pitch shifting factors under 15% [33]. For higher pitch shifting factors, especially above 50%, frequency domain techniques generate less artefacts [33], [30].

4.2 Frequency Domain Techniques

Frequency domain techniques rely on the STFT (see previous Section 3.1) and achieve a pitch shifting effect at least in part by way of modifications brought to the Fourier transform bins [23]. In particular, these modifications aim to scale the instantaneous phase of each Fourier sub-band while preserving frame-to-frame phase coherence in the un-stretched output signal. Phase vocoder techniques achieves this through implicit or explicit estimation of the instantaneous frequency in the sub-band. By contrast, the OCEAN method, see Section 4.2.1, assumes the instantaneous frequency in each sub-band to be equal to its center frequency [28].

These techniques are known to provide better results for polyphonic material. They are further described hereafter. For notational convenience and to more directly reflect practical implementations, spectral transformations are denoted as carried out on the downsampled passband signals $\tilde{X}(\omega, k)$ as provided by common FFT algorithms, i.e. no modulation to baseband is conducted in the sub-bands contrary to the representations of Figures 5 and 6.

4.2.1 OCEAN Method

Juillerat teaches to scale the frequency location of the STFT sub-bands by shifting the bins of the transform to modified index positions so as to achieve a pitch shifting effect [28]. In particular, achieving a doubling of octaves requires repositioning each bins towards a doubled index position.

Maintaining horizontal phase coherence for each displaced bin is accomplished by rotating its phasor according to the bin displacement. In the specific case of octave doubling, the in-frequency modification can be formally expressed as:

$$\tilde{Y}(m,k) = \begin{cases} \tilde{X}(m, \frac{k}{2})\mathrm{e}^{-i2\pi\frac{k}{2}\frac{m \bmod O}{O}} & k \text{ even} \\ 0 & k \text{ odd}, \end{cases} \tag{17}$$

where $O = \frac{N}{R}$ denotes the overlaping factor, mod denotes the modulo operator and m denotes the frame index (in other words the downsampled sample time index).

Given the above the vertical phase coherence is exactly preserved every Oth frame, that is for each frame of index m that verifies:

$$m \bmod O = 0. \tag{18}$$

For other frames, it is preserved within each group of bins sharing the same phase rotating term according to Equation (17). Juillerat notes this results in a transient duplication artefact, as the frequency components of an impulse-like signal are distributed amongst as much as O groups of bins at each frame, and shifted according to the phase value of said group.

The bin repositioning operation and the phase propagation expression arbitrarily assumes that the frequency of the sub-band's content matches the center frequency of the sub-band. This results in mistuning as for the Rollers method, as well as in amplitude modulation.

In practice, a window must be applied upon the re-synthesis stage so as to taper sharp edges arising at the border of the pitch shifted frame as a result of the circular convolution effect of the phase adjustment according to (17) conducted in frequency domain.

Given the above, the OCEAN method can be interpreted as an efficient way of approximating a single-sideband modulation of each Fourier sub-band towards the next octave. Specifically, the modulation is accomplished from the combined operation of bin repositioning, phase rotation and inverse Fourier transform. By contrast to audio-rate sub-band single-sideband modulation which is conducted on a sample-by-sample basis, the single-sideband modulation is performed in the case of the OCEAN algorithm on a frame-by-frame basis. The two techniques are equivalent when the hop size is set to 1.

Latency Figure 8 represents the output of the OCEAN pitch shifting technique in response to an impulse test signal. In this example the analysis and synthesis windows are about 20 ms in length and the overlap is 75% ($O = 4$). As pictured, some impulse image duplicates fold back in time by effect of circular convolution such that the overall response spreads to an earlier start than the 20 ms latency that is expected from the STFT filter bank in the absence of in-frequency transformation. The response is given for different impulse time offsets with regards to the window timings of the STFT. As pictured the worst case offline latency produced as measured from the input pulse to the earliest transient duplicate corresponds to the window duration minus one hop in this case.

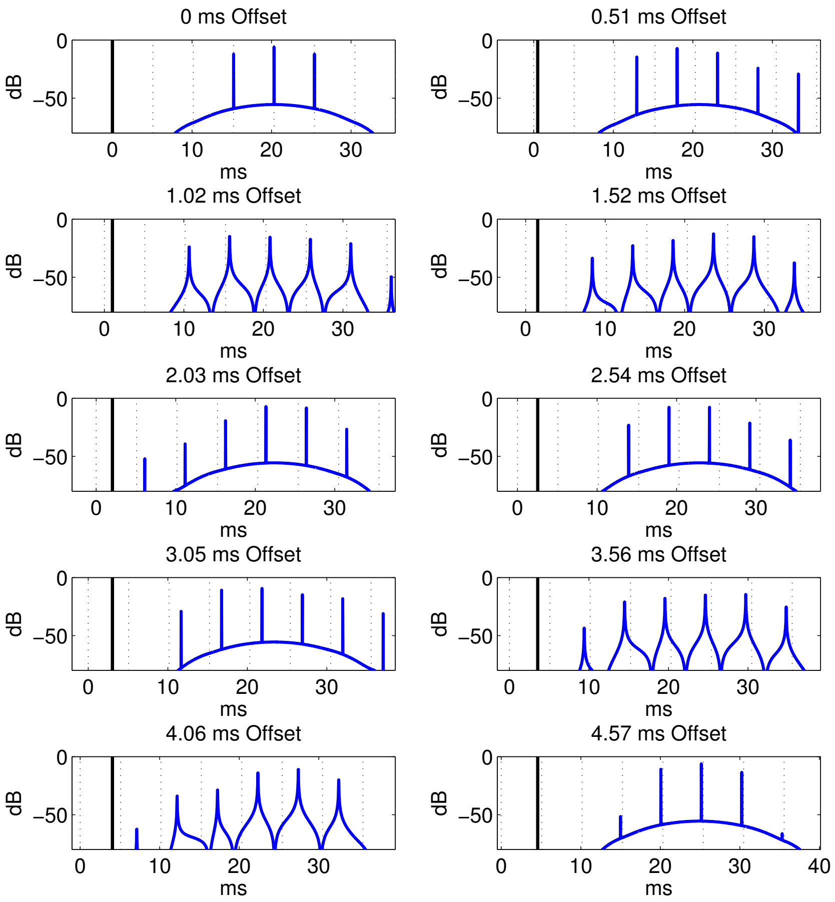


Figure 8: Output waveform of a causal implementation of the OCEAN pitch-shifting method (blue) in response to an impulse signal input (black). Various timing offsets of said impulse with regards to a the start of the analysis window (i.e. at 0 ms) are pictured. In this example, the overlap factor is four. The window and hop duration are approximately 20 ms and 5 ms respectively. Black dotted lines mark the timings for the start of each shifted window. As pictured the worst case offline latency produced by the method is 15 ms which corresponds to the window duration minus one hop.

4.2.2 Phase Vocoder: Time Stretching and Resampling Method

The phase vocoder can be defined as a “STFT representation by means of amplitude and frequency” [47], although some phase vocoder techniques only make implicit use of instantaneous frequency without estimating it explicitly.

According to a first technique, pitch shifting is carried out by the combination of time-stretching and resampling, [2], [17]. For example, in the case where an upwards pitch shift to the next octave is sought, a doubling of the input signal's length is first accomplished. The resulting stretched signal is then downsampled by the same factor, effectively shortening the signal back to its initial un-stretched duration while, at the same time, stretching the frequency scale. This perceptively translates to a doubling of octaves. Other integer or rational pitch-shifting factor values can be employed. In particular, for values lower than one, the input signal is first contracted before being upsampled back to its initial length, thereby achieving a downwards pitch shift.

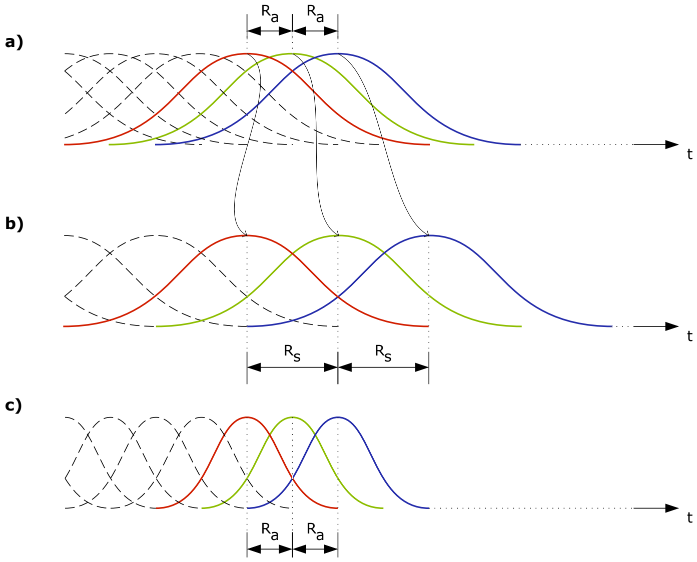


Figure 9: Schematical representation of the phase vocoder time stretching and resampling method for octave doubling. a) Input signal's analysis frames as distributed according to analysis hop size R_a. b) Synthesis frames as distributed according to synthesis hop size $R_\mathrm{s} = 2R_\mathrm{a}$. The signal is resampled after recombination of frames. c) Implementation variant in which each frame is resampled prior to recombination [2]. This variant is suitable for real time implementations of the method.

Horizontal Phase Propagation In practice, the STFT is typically implemented using an FFT algorithm and following the WOLA analysis-synthesis scheme (see Section 3.3.2) [42], [10]. However, unlike in the diagram of Figure 6, the synthesis and analysis hop sizes are set to differ by a factor corresponding to the pitch factor:

$$R_\mathrm{s} = \gamma R_\mathrm{a}, \tag{19}$$

where R_s denotes the (corrected) synthesis hop size, R_a the analysis hop size and γ the pitch factor. Depending on implementations, either R_a or R_s can be set to a fixed value while the other is ajusted according to (19) [4].

Figure 9 represents the case of a doubling of octaves. Correspondingly, the analysis hop size in a) is twice that in b) which translates into a wider positioning offset of each frame relative to the previous, i.e. an increase of $R_\mathrm{s} - R_\mathrm{a}$ (compared to the perfect reconstruction case). Phase continuity of the frequency components of the input signal is preserved across the repositioned frames by preventively rotating, prior to synthesis, said components short-time phase values provided by the STFT. Such conservation of phase continuity of the components' phases along the time axis is called horizontal phase coherence, in reference to the abscissae axis which usually represents time in time-frequency representations (while frequency is represented extending along the vertical axis). Under the ideal assumption that each sub-band contains nothing more than a single pure tone, the corrected phase value for the sub-band follows from frame to frame, a recursive phase propagation formula from frame to frame [35]:

$$\tilde{\theta}_\mathrm{s}(m,k) = \tilde{\theta}_\mathrm{s}(m-1,k) + \frac{R_\mathrm{s}}{R_\mathrm{a}} \int_{t_{m-1}}^{t_m} \tilde{\omega}(t,k)dt, \tag{20}$$

where $\tilde{\theta}_\mathrm{s}$ denotes the synthesis phase, $\tilde{\omega}(t,k)$ denotes the time-continuous instantaneous frequency of said pure tone in the kth band. The integral term of the above expression represents a phase increment of the pure tone that occured between current frame m and previous frame $m-1$.

Phase propagation depends on proper estimation of said phase increment. Denoting $\tilde{\omega}'(m,k)$ the estimated value at frame m of the instantaneous frequency $\tilde{\omega}$ for the pure tone located in the kth bin, an estimation $\Delta\tilde{\theta}'(m,k)$ of the phase increment can be formulated as follows:

$$\Delta\tilde{\theta}'(m,k) = R_\mathrm{a}\tilde{\omega}'(m,k). \tag{21}$$

Phase increment estimation and instantaneous frequency estimation thus form a single problem. A discussion on this question follows.

Instantaneous Frequency Estimation For values of the analysis hop size greater than one, instantaneous frequency estimation requires the implementation of an unwrapping procedure to take account for the (intentional, see Section 3.1) frequency fold-over effect that occur in the sub-band as a result of undersampling. This unwrapping procedure is based on the following decomposition of the instantaneous frequency $\tilde{\omega}(t,k)$ [2] [35]:

$$\tilde{\omega}(t,k) = \frac{2\pi k}{N} + \omega(t,k), \tag{22}$$

where $\omega(t,k)$ denotes the complex baseband frequency offset and $\frac{2\pi k}{N}$ represents the kth sub-band's center frequency as pictured in Figure 10 a). Accordingly, the estimated phase increment from frames $m-1$ to m is given by [35]:

$$\Delta\tilde{\theta}'(m,k) = R_\mathrm{a}\frac{2\pi k}{N} + \Delta\theta'(m,k), \tag{23}$$

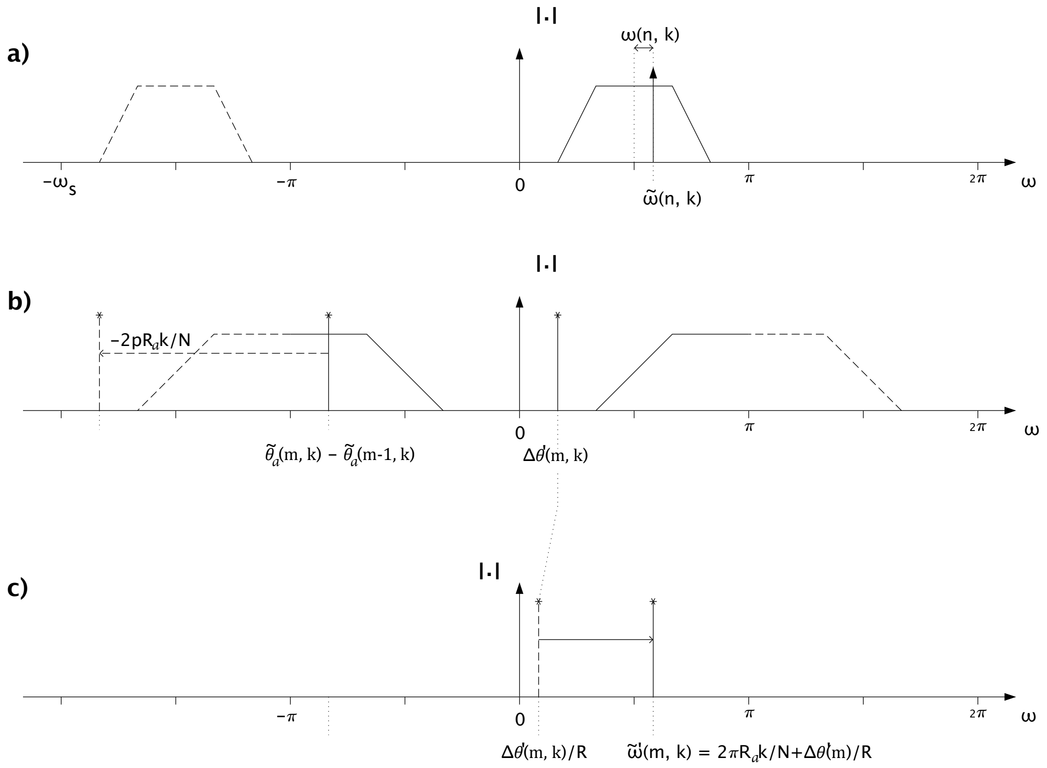


Figure 10: Illustrative representation of the phase unwrapping operation. a) Audio rate band: Frequency representation of a pure tone located in the passband of a sub-band channel before downsampling (in this example $R = 2$ and $k = 1$). b) Downsampled sub-band: Corresponding passband phase increment (left) and calculated unwrapped baseband increment (right). c) Audio rate band: Estimated instantaneous frequency as determined from said unwrapped increment.

where $\Delta\theta'(m, k)$ denotes the baseband phase increment estimate. This estimate is itslef given by [35]:

$$\Delta\theta'(m, k) = \text{princarg}[\tilde{\theta}_{\text{a}}(m, k) - \tilde{\theta}_{\text{a}}(m-1, k) - R_{\text{a}}\frac{2\pi k}{N}], \tag{24}$$

where $\tilde{\theta}_{\text{a}}(m, k)$ and $\tilde{\theta}_{\text{a}}(m-1, k)$ denotes the phase measurement acquired for the current frame and previous frame respectively, and princarg is the principal argument function.

An illustrative representation of formulas (23) and (24) the above formula is given in Figure 10. As pictured, the procedure essentially consists in finding a valid value for the phase increment in the sub-band's bandwith given the downsampling factor and a measured (wrapped) phase increment.

Vertical Phase Coherence The phase vocoder is known for creating a phasiness artefact [34], [35], [28]. Puckette teaches that the source of this artefact lies in the phase misalignment upon re-synthesis of contributions generated by neighbouring sub-bands channels [43]. Such misalignment produces amplitude modulations on an input pure tone simultaneously located in the overlapping sub-bands of said neighbouring channels. A solution to this problem is to conduct in-frequency processing of the signal in such way that the relative phase relation between neighbouring sub-bands corresponding to the same pure tone is maintained. Such alignement of the phase across the frequency bands of the short-time spectrum is called vertical phase coherence. This name results from the fact that the frequency is usually represented by the vertical axis in time-frequency representations.

Laroche teaches to reformulate Equation (20) to express the synthesis phase value at frame m in terms of an initial synthesis phase value and a cumulative phase increment [35]:

$$\tilde{\theta}_{\mathrm{s}}(m,k) = \tilde{\theta}_{\mathrm{s}}(0,k) + \frac{R_{\mathrm{s}}}{R_{\mathrm{a}}}\int_0^{t_m} \tilde{\omega}(t,k)dt, \tag{25}$$

where said cumulated phase increment is given by the integral term and can be estimated by frame-to-frame accumulation of the individual estimated increments according to (23):

$$\tilde{\theta}_{\mathrm{s}}(m,k) = \tilde{\theta}_{\mathrm{s}}(0,k) + \frac{R_{\mathrm{s}}}{R_{\mathrm{a}}}\sum_{l=1}^{m} R_{\mathrm{a}}\frac{2\pi k}{N} + \Delta\theta'(m,k) \tag{26}$$

Substituting $R_{\mathrm{s}}/R_{\mathrm{a}} = \gamma$ and (24) into (26) yields the following result [35]:

$$\tilde{\theta}_{\mathrm{s}}(m,k) = \tilde{\theta}_{\mathrm{s}}(0,k) - \gamma\tilde{\theta}_{\mathrm{a}}(0,k) + \gamma\tilde{\theta}_{\mathrm{a}}(m,k) + \gamma\sum_{l=1}^{m} 2\lambda_l(k)\pi, \tag{27}$$

where $\lambda_l(k)$ is an integer and $2\lambda_l(k)\pi$ denotes the wrapped phase component at frame m.

Inspection of (27) reveals two sources of phase incoherence if we assume that the instantaneous phase measurement $\tilde{\theta}_{\mathrm{a}}(m,k)$ is identical across the neighbouring (and overlaping) channels in which the pure tone is located[21]. The first source of error lies in the initial phase terms $\tilde{\theta}_{\mathrm{s}}(0,k)$ and $\tilde{\theta}_{\mathrm{a}}(0,k)$. Laroche teaches to set, in each band, the synthesis phase term $\tilde{\theta}_{\mathrm{s}}(0,k)$ to the measured phase term $\tilde{\theta}_{\mathrm{a}}(0,k)$ scaled by the stretch factor γ [35]. This ensures that these terms vanish in the above expression which thus becomes:

$$\tilde{\theta}_{\mathrm{s}}(m,k) = \gamma\tilde{\theta}_{\mathrm{a}}(m,k) + \gamma\sum_{l=1}^{m} 2\lambda_l(k)\pi. \tag{28}$$

Integer Pitch Factor Case The second source of phase incoherence is the cumulated phase unwrapping term. Fortunately this terms simplifies to a multiple of

[21] This is the case in the absence of interference provided by a circular shift of the input signal frame has been performed before application of the Fourier transformation as described in Section 3.3.2.

2π for integer values of γ, λ being also an integer. In that case equation (28) further simplifies to:

$$\tilde{\theta}_{\mathrm{s}}(m,k) = \gamma\tilde{\theta}_{\mathrm{a}}(m,k). \tag{29}$$

Of course, the above equation applicable to octave doubling, i.e. for $\gamma = 2$. A possible formulation of the corresponding spectral transformation is given by:

$$\tilde{Y}(m,k) = |\tilde{X}(m,k)|\mathrm{e}^{2i\angle\tilde{X}(m,k)}. \tag{30}$$

Explicit evaluation of the trigonometric term can be avoided following the equivalent formulation below:

$$\tilde{Y}(m,k) = \tilde{X}(m,k)\frac{\tilde{X}(m,k)}{|\tilde{X}(m,k)|}. \tag{31}$$

Non-Integer Pitch Factor Case Although not directly releavant in the case of octave doubling, the case of non-integer pitch factor is discussed hereafter for completeness. For non-integer values of γ, the cumulated phase unwrapping term is not a multiple of 2π but is nevertheless expected to be of equal value across neighbouring sub-bands under the influence of a common pure tone. Laroche notes that phase coherence across the bands is maintained as long as this equality remains, which is however doubtful outside the ideal case of an ever lasting sinusoid of constant frequency: any temporary interruption of the pure tone will lead to permanent divergence of the values of the summation term of equation (28) across the sub-bands as a consequences of determinations of said summation term values upon the noise background [35]. Since these values are determined frame by frame in a cumulative fashion, the loss of vertical phase coherence will affect permanently any future resynthesized pure tone, as noted by Laroche.

Some level of vertical phase coherence can nevertheless be ensured by means of phase-locking techniques which attempt to re-establish vertical phase coherence a posteriori following the horizontal phase propagation step and before the synthesis bank. In particular, loose phase locking was proposed by Puckette [43]. This technique consists in assigning to each sub-band a synthesis phase term corresponding to the weighted average phase of the sub-bands of the close neighbourhood. Laroche proposes an alternative phase-locking solution with improved results and which rely on a peak detection algorithm [35].

Choice of Window Duration and Hop Size The above presentation of the Phase Vocoder technique implicitly relied on the assumption that at most only a single pure tone is present at once in each band. For such assumption to approximately hold typically requires using relatively high order analysis filters to ensure sufficient frequency resolution. This is in particular required for polyphonic signals which typically contain numerous and therefore potentially closely located partials. Lack of frequency selectivity leads to increased instantaneous frequency estimation error [45].

Moreover, the downsampling factor should be set in such way that a guard band is provided between the passband of the downsampled sub-band and its aliased image, so as to prevent high-frequency components generated from the non-linear processing to leak into said passband as aliased terms. In practice the downsampling factor should not be set above half the value defined by the relaxed strong COLA requirement (see Section 3.3.3).

Latency The amount of latency introduced by the Phase Vocoder technique is not a well-documented phenomenon [37]. Empirical investigations conducted by the author have lead to the hypothesis that, in the specific case of an integer pitch factor value and following Equation (29), it corresponds to the perfect reconstruction filter bank delay as expressed by Equation (15). In general, the latency thus corresponds to the WOLA frame duration when the analysis and synthesis windows are of the same length [6].

Use of the time-stretching / resampling phase vocoder method for octave doubling provides a significant advantage in that regard as its synthesis window is half the length of the analysis window as a result from the downsampling operation, thus leading to reduced latency. An empirical expression for offline latency can thus be expressed as:

$$d = \frac{N}{f_s}(\frac{1}{2} + \frac{1}{2\gamma}). \tag{32}$$

In practical real-time implementations, added latency is introduced by a buffering of the output (see Section 6.5).

4.2.3 Other Phase Vocoder Methods

Oscillator Bank Synthesis An alternative to the time-stretching/resampling approach described above consists in synthesising a sinusoid for each sub-band according to the scaled value of the instantaneous frequency estimate. In practice, the analysis bank can still be implemented using a WOLA scheme as usual. The phase argument of each oscillator is determined on a per-sample basis (at input signal sample rate) by integration of the instantaneous frequency estimate calculated in the corresponding sub-band. Horizontal phase coherence is also ensured, as in the time-stretching/resampling scheme by means of phase propagation from one frame to the next. Dutilleux suggests ensuring smooth transition between frame is by means of a triangular-shaped synthesis window [2].

Because an oscillator running at high-quality audio sampling rate is implemented at the output of each sub-band channel, the oscillator bank synthesis method is often significantly more costly than the time-stretch/resample methods described above. However, unlike this method, it allows independent scaling of the instantaneous frequency in each synthesis band. Sophisticated effects based on non-unifom scaling of the partials are therefore achievable with this method which is not the case of the former.

Bin Repositioning Laroche describes yet another alternative method for pitch shifting under the phase vocoder framework [36]. In this solution, the WOLA scheme is implemented for both the analysis bank and synthesis bank. The scaling of the frequency estimate is carried out through repositioning of the frequency bins inside the frequency transform as for the OCEAN algorithm (see Section 4.2.1). By contrast however, a peak detection step is carried out before the bin repositioning operation. This ensures that frequency bins receiving a common pure tone are translated grouped together. Moreover, instantaneous frequency estimates are found for each peak in order to determine precisely the repositioning offset to be applied for the group of bins.

Sliding Phase Vocoder Jacobsen teaches that the Sliding Discrete Fourier Transform is more efficient than the radix-2 Fast Fourier Transform in the absence of downsampling [27], that is when $R = 1$. Such lack of downsampling ensures the sub-bands are processed at the input audio rate. Consequently, pitch shifting can be readily applied by offsetting the frequency of each sub-band of the Fourier Transform by any amount within the auditory band [6]. Such operation does not imply repositioning of the bins within the frequency representation unlike in the previous method.

Phase unwrapping errors result from the undersampling of the sub-band as discussed above. This issue vanishes in the case of the sliding phase vocoder as noted by Brown [7]. Consequently, integer as well as non-integer pitch shifting values can be used with the same expected sound quality.

Moreover, the sliding phase vocoder provides reduced latency according to Bradford [6]:

> *While the contents of the frame for the first few samples understandably bear little relation to the source, we find that useful output is available when just one third of the frame has been filled. This appears to be an exact figure - tests with N a multiple of 3 confirm that viable output (directly matching the input signal) commences exactly at sample $N/3$. Figure 4 illustrates the case for $N = 1500$, where useful output starts exactly at sample 500, excluding some startup transient behaviour 2 . The SPV thus reduces the latency imposed by standard pvoc, for a given value of N, by some 75%, with clear and valuable advantages for real-time performance*

where N denotes the window duration in samples.

The added computation cost of the sliding phase vocoder technique can be addressed by exploiting the intrinsic parallel properties of the Sliding Discrete Fourier Transform by way of implementation on a General Purpose Graphics Processing Unit (GPGPU) [5].

4.3 Rollers

Multi-resolution filter bank designs provide a solution when the latency and frequency resolution requirements cannot be simultaneously met by uniform filter-bank designs, that is filter banks exhibiting equal bandwidths across the bands such as with the STFT. Multi-resolution filter banks, by contrast, exhibit non-uniform passband widths. In particular, achieving a reduction in perceived latency can be accomplished using passband filter bandwidth specifications that increase with frequency. The Rollers pitch-shifting solution proposed by Juillerat is an example of such use of a multi-resolution filter-bank design [29]. A schematic block diagram of this solution is pictured in Figure 11. In particular, Juillerat teaches to combine a constant-Q multi-resolution filter-bank with single-sideband modulation in each sub-band. The filter bank is composed of real-valued fourth order IIR passband filters. As pictured in the block diagram, the quadrature component of the input signal to each single-sideband modulation block, denoted SSM2 in the figure, is provided using a Hilbert transformer.

Juillerat notes that a downward chirp results from the increase of group-delay value across the sub-bands from high to low frequencies. In particular, he notes the level of this artefact becomes unacceptable under his selected constant-Q specification value. As a solution, he teaches to use reduced quality factor values for the low-frequency analysis filters so as to meet a predetermined maximum frequency resolution threshold. The low end resonance being thus limited, the level of perceived chirp effect is reduced. According to Juillerat, his design achieves acceptable results with as low as 10 ms of latency.

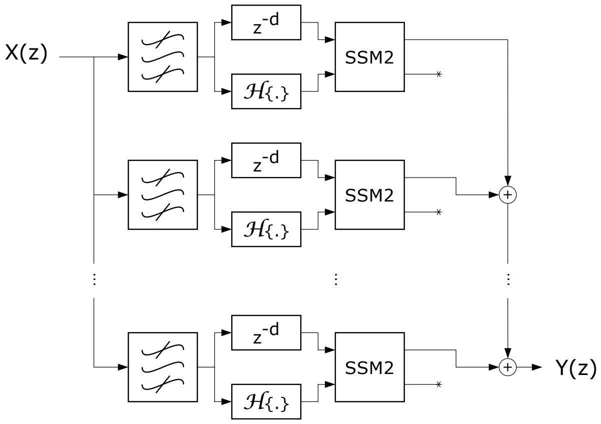


Figure 11: Multi-resolution filter bank implementation with sub-band single-sideband modulators (SSM2) as proposed by Juillerat [29]. The analysis filters are real valued.

5 Novel Octave Doubler Designs

Novel octave doubling solutions are developed in the current section. In a first part of this section, candidate general purpose sub-band modulation techniques are presented and compared in light of their potential applicability to the specific problem at hand as formulated in this thesis's Introduction. In a second part of this section, two selected sub-band modulation techniques are combined with a non-decimated multi-resolution IIR filter bank similar to that found in the Rollers method proposed by Juillerat's (see Section 4.3): notably, a constant-ERB-bandwidth specification is employed for the design instead of a constant-Q one. Tests of the two novel solutions are presented, alongside that of Phase Vocoder and OCEAN octave doubler implementations, in following Sections 6 and 7.

5.1 Sub-Band Modulator Candidates for Octave Doubling

We propose the following general notation for the modulation techniques discussed in the current section:

$$y_k(t) = x_k(t)m_k(t), \tag{33}$$

where $x_k(t)$, $m_k(t)$ and $y_k(t)$ denote the input signal, the modulation signal, and the output signal respectively for sub-band channel k. Following this formulation, the modulation process of a memoryless waveshaper modulator $f(.)$ can be expressed as follows:

$$\begin{aligned} y_k(t) &= f\big(x_k(t)\big) \\ &= m_{k,x}(t)x_k(t), \end{aligned} \tag{34}$$

where:

$$m_{k,x}(t) = \frac{f\big(x_k(t)\big)}{x_k(t)}. \tag{35}$$

It is brought to the reader's attention that the modulator function $m_{k,x}(t)$ is in that case dependent on the input as indicated by the subscript. A relatively extensive list of modulators with a potential for octave doubling application is given in Table 3. In this table:

- $\hat{x}_k(t) = x_k(t) + i\mathcal{H}\{x_k(t)\}$ denotes an analytic input signal (complex-valued), $\mathcal{H}\{x_k(t)\}$ denoting the Hilbert transform [21][22].
- γ represents the pitch shifting factor, in our case $\gamma = 2$.
- ω_k denotes the center frequency of sub-band channel of index k.
- Band-Limited Full-wave Rectifying (BLFWR) is based on a sixth order Taylor series expansion [50][23] of the absolute value operator and thus provides a band-limited approximation of the full-wave rectifying function, which provides a

[22] p. 96

[23] http://ccrma.stanford.edu/~jos/mdft/Taylor_Series_Expansions.html

finite bandwidth alternative to full-wave rectifying with an improved dynamic response compared to the power-of-2 (PO2) modulator. However, the graph of Figure 29 in Appendix A of the appendix demonstrates that the dynamic response is still highly non-linear. Values for coefficient c_1, c_2 and c_3 are derived in the Appendix.

	$y_k(t)$	$\gamma = 2$	Modulator
Input-independent meth.:			
Ring Modulation	$x_k(t)\cos((\gamma-1)\omega_k t)$	RM2	$m_k(t) = \cos(\omega_k t)$
Single Sideband Modulation	$\hat{x}_k(t)\mathrm{e}^{i(\gamma-1)\omega_k t}$	SSM2	$\hat{m}_k(t) = \mathrm{e}^{i\omega_k t}$
Static waveshaping meth.:			
Full-Wave Rectifying	$\lvert x_k(t)\rvert$	FWR	$m_{k,x}(t) = \mathrm{sign}(x_k(t))$
Half-Wave Rectifying	$x_k(t)(x(t) > 0)$	HWR	$m_{k,x}(t) = x_k(t) > 0$
Band-Limited FWR	$c_1 x_k(t)^2 - c_2 x_k(t)^4 + c_3 x_k(t)^6$	BLFWR	$m_{k,x}(t) = c_1 x_k(t) - c_2 x_k(t)^3 + c_3 x_k(t)^5$
Power (real-valued signal)	$x_k(t)^\gamma$	PO2	$m_{k,x}(t) = x_k(t)$
Phase scaling methods.:			
Power (analytic signal)	$\hat{x}_k(t)^\gamma$	CPO2	$\hat{m}_{k,x}(t) = \hat{x}_k(t)$
Phase Scaling	$\hat{x}_k(t)(\frac{\hat{x}_k(t)}{\lvert\hat{x}_k(t)\rvert})^{\gamma-1}$	PS2	$\hat{m}_{k,x}(t) = \frac{\hat{x}_k(t)}{\lvert\hat{x}_k(t)\rvert}$

Table 3: List of sub-band modulators with a potential for octave doubling applications. General expressions for sub-band output (left). Specific abbreviations and expressions for modulator function in the specific case of octave doubling operation, i.e. for $\gamma = 2$ (right)

Other possible sub-band techniques for polyphonic octave doubling include sub-band time-domain stretching, [14], [15], [13]. These methods have not been considered in the current study but could form the object of a future work.

5.1.1 Added Roughness

Critical Band A critical band is a frequency interval within which sounds with distinct frequencies cannot be discriminated by the human ear. Critical bands thus define the frequency resolution of the human auditory system. Auditory scales such as the Bark scale and the ERB scale are derived from measurements of the critical bandwiths. They provide a tonotopical representation of the human perception of sound frequencies [46].

Roughness Roughness is caused by the beating resulting from the combination of (at least) two harmonics into a critical band [52], [46]. The frequency of the amplitude modulation corresponds to the difference of the harmonics' frequencies. It has been shown that under a threshold of about 10 to 16 Hz the amplitude modulation is perceived as fluctuation and can be found to be pleasant. Above that threshold and within the width of the critical band however it is perceived as roughness. The overall perceived roughness can be though as corresponding to the sum of roughness specific to each critical band. Roughness is in part determined by the pitch ratios of the combined notes and sensory dissonance rises when neighbouring harmonics align two by two in such imperfect manner that it produces roughness. In the context of polyphonic octave doubling, the consonance or dissonance of the input sound should be preserved at the output. A discussion of the modulation methods of Table 3 is made below in terms of their respective impact on the perceived roughness of the output sound.

Out-of-Band Distortion Firstly, added roughness can occur despite a lack of interfering harmonics in the sub-bands. Figure 12 illustrates the output spectrum produced by each modulation technique in response to a pure tone. As pictured, harmonic distortion by-product components[24] or side-band by-product components[25] are generated for all methods but SSM2, CPO2 and PS2. In a polyphonic context where the harmonic components are related to multiple coexisting pitches, these out-of-band distortion components do not align exactly as would happen with a perfectly harmonic monophonic input, but combine in beating pairs with each other or with pitch-transposed harmonics, which is susceptible to generate roughness.

Given this, selecting sub-band modulators exhibiting low out-of-band distortion should help to prevent occurrence of added roughness at the output. The amount of distortion contributed by each sub-band in the single pure tone case can be quantified according to the principles of the traditional total harmonic distortion (THD) formulae [19][26]. In the following modified version, all frequency components but the pitch-transposed harmonic, in our case the second rank, is counted as distortion in accordance to the specific goal of octave doubling:

$$D = \sqrt{\frac{A_1^2 + A_3^2 + \cdots + A_K^2}{A_1^2 + A_2^2 + \cdots + A_K^2}}. \tag{36}$$

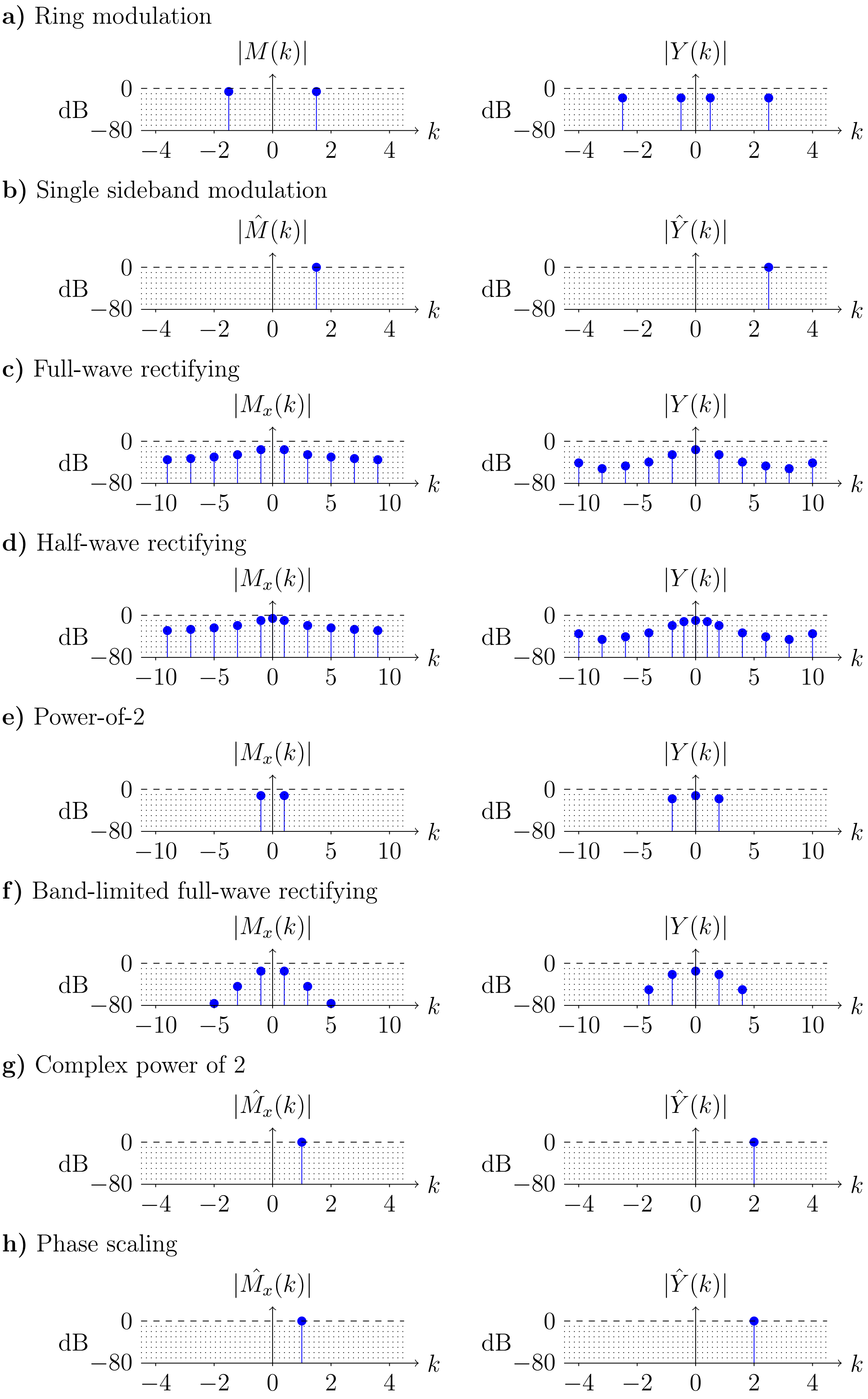
Intermodulation distortion In the non-ideal case, each sub-band is susceptible to receive multiple frequency components (of a monophonic or polyphonic input signal). In such case, intermodulation distortion should result from all modulation techniques of Table 4, with the notable exception of RM2 and SSM2. The impact of intermodulation distortion on perceived roughness depends on the composition of the sub-band input signal as well as the modulation method and would require further study in order to be characterised precisely for each case. Figure 13 illustrates the intermodulation distortion that occurs in the complex power-of-2 case (CPO2). As an immediate requirement for the modulation techniques subject to intermodulation distortion, the analysis filters should be specified with the narrowest possible passbands so as to minimise the count of interfering frequencies in the sub-bands.

Beating Pairs in Transition Bands Another source for roughness lies, according to Juillerat, in the potential creation of beating pairs within the overlapping transition bands of the filters upon combination of the sub-bands for synthesis of an overall output sound. This is of particular concern for the RM2 and SSM2 methods as it arises even in the ideal single pure tone case. If it is located in the crossover

[24] Static waveshaper cases: FWR, HFWR, POW2 and BLFWR.

[25] Ring-modulator case [21].

[26] p. 102



a) Ring modulation

b) Single sideband modulation

c) Full-wave rectifying

d) Half-wave rectifying

e) Power-of-2

f) Band-limited full-wave rectifying

g) Complex power of 2

h) Phase scaling

Figure 12: Frequency representations of modulator and output signals in response to a pure tone input. The abscissa represents the frequency relative to that of the pure tone (not represented), itself thus being located at $k = 1$.

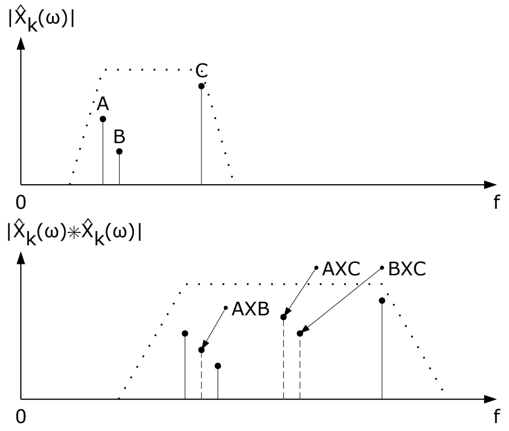


Figure 13: Schematical example of intermodulation occuring in the sub-band channel for the complex power-of-2 case (CPO2). The intermodulation components are represented in dashed lines.

band, a beating pair is generated from the single pure tone as a result of the difference in frequency translation applied for the corresponding partially overlapping bands (which depend on a fixed modulation frequency for each band). This effect cannot be totally avoided although lowering the crossover amplitude specification of neighbouring filters should limit the amount of added roughness it introduces. There is however a limit to which the attenuation crossover can be set without excessively denaturing the processed output sound. Juillerat teaches that “removing more than about a quarter of the frequencies is not desired” [29]. He suggests using analysis filters having an attenuation crossover of -12 dB.

5.1.2 Mistuning

Unlike the other modulation techniques listed in Table 3, Ring Modulation (RM2) and Single-Sideband Modulation (SSM2) do not introduce intermodulation distortion. As mentioned above, the oscillator is set with a predetermined fixed modulation frequency which prevents the occurrence of intermodulation terms. Consequently, the output sub-band spectrum is as sparse as that of the input, and thus do not lead in itself to significant increase in roughness, if any[27]. However, unlike the other listed methods, RM2 and SSM2 shifts the frequency components of a given sub-band by an equal amount as determined by the fixed predetermined modulation frequency value, which leads to individual mistuning errors with regards to target pitch values as specified by the desired pitch transposition. Such an error increases with the analysis passband width. Thus narrow band analysis passband filters are also required for RM2 and SSM2 as the other modulation techniques, although

[27] A slight increase or decrease can occur even in the absence of intermodulation as critical bands do not contribute equally to perceived roughness. Thus the content in the original critical band can yield a different amount of perceived roughness after translation to a different targeted critical band as a result of the pitch shifting operation.

for a different reason, that is prevention of shifted frequency mistunings instead of intermodulation distortion.

5.1.3 Comparative Summary

A comparative summary of the sub-band modulation techniques is provided by Table 4. As shown, SSM2, CPO2 and PS2 methods do not produce out-of-band distortion in the ideal pure tone case. Low expected sub-band output distortion allows to contemplate re-synthesis of the mixed output signal without sub-band post filtering. In turn this allows for sharper analysis filter designs at a given overall maximum latency specification. This is significant given the fact that intermodulation distortion nor mistuning can be cured by post-filtering: these artefacts can only be prevented as much as possible by ensuring narrower analysis passbands. Thus the above mentioned methods are expected to yield, when combined with an adequate filter-bank design, the lowest artefacts under stringent real-time constraints.

Unlike CPO2, both SSM2 and PS2 methods provide a linear dynamic response. SSM2 and PS2 thus seem to form the best candidates for real-time polyphonic octave doubling applied to the guitar. In addition to roughness introduced by beating pairs in overlapping transition bands, each of these methods introduces a specific processing artefact: intermodulation-generated roughness in the case of PS2, and mistuning in the case of SSM2[28].

Given the above, SSM2 and PS2 are both selected for prototyping and testing (see Section 7 for test results). As pictured in Figure 14 both methods are implemented within a novel multi-resolution filter bank design, which is described in the following section.

	Out-of-Band Dist.	Interm. Dist.	Mistuning	Lin. dyn. resp.
RM2	+++	No	Yes	Yes
SSM2	None	No	Yes	Yes
FWR	+++	Yes	No	Yes
HWR	+++	Yes	No	Yes
BLFWR	++	Yes	No	No
PO2	+	Yes	No	No
CPO2	None	Yes	No	No
PS2	None	Yes	No	Yes

Table 4: Comparison summary table of potentially applicable sub-band modulation techniques. Out-of-band distortion is given for the pure tone case in light of Figure 12. Plus sign indicates increased level of distortion. Red, orange and green colors respectively indicate poor, average and good preliminary performance assessment ranking relative to the other modulators.

[28] Concurrent partials in overlapping shifted transition bands can also generate added roughness in the case of SSM2.

5.2 Improved Multi-Resolution Filter Bank Design

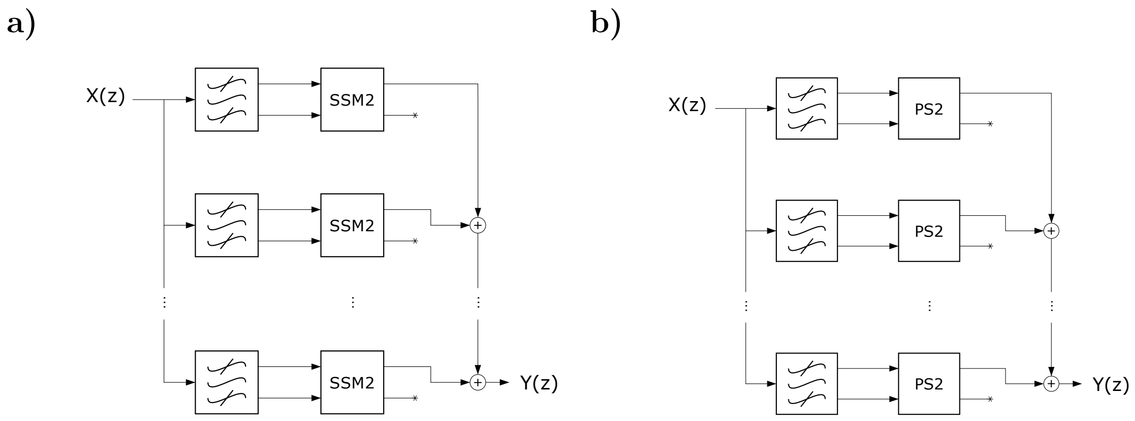


Figure 14: Novel octave doubling solutions based on a identical multiresolution filter bank design with complex IIR passband filters. a) SSM2-based solution. b) PS2-based solution. The operation caried out in SSM2 and PS2 modulator blocks are given in Table 4. In this schematical representation, each block produce an in-phase signal and a quadrature signal, the later being discarded at the output of the block.

5.2.1 Complex IIR Multi-Resolution Filter Bank Design

The block diagrams of Figure 14 represents two novel short-delay octave doubling solutions similar to that of Juillerat's (see Figure 11). By contrast, the analysis banks are formed of complex filters instead of real-valued ones. This provides the analytic sub-band directly at the output of the passband filter without need for the additional step of Hilbert transformation. This reduces the system's latency by the amount introduced from the Hilbert transformer in Juillerat's solution. As pictured, the filter bank systems of Figure 14 are used to carry out the selected PS2 and SSM2 candidate octave doubling modulators of Section 5.1.3.

In practice, complex minimum-phase passband designs can be derived from a real-valued prototype by a simple manipulation of the conjugate pair's pole values as represented by the Z-plane representations of Figure 15.

In particular, the analysis filter bank can be for example implemented following a bi-quad filter design. Design of such filters are practical for prototyping as their coefficients can be readily derived from a quality factor q and and a center frequency f_{c} specification [20][29] as defined by Table 5, where $K_{\mathrm{BQ}} = \tan\left(\frac{\pi f_{\mathrm{c}}}{f_{\mathrm{s}}}\right)$. The real-valued coefficient filter prototype should be specified with a quality factor value half of that desired for the derived complex filter.

As pictured in Figure 14, the quadrature component is discarded at the output of each sub-band modulator block. In practice, a modified version of the equations

[29] p. 49

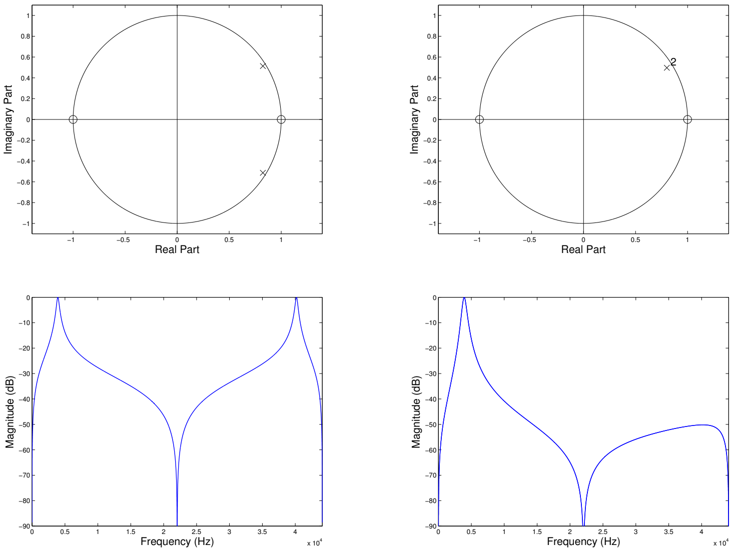


Figure 15: Complex minimum-phase IIR passband filter design. Left: Z-plane for real-valued passband prototype (top) and corresponding magnitude response (bottom). Right: Z-plane for the derived complex design (top) and corresponding magnitude response (bottom).

b_0	b_1	b_2	a_1	a_2
$\frac{K_{\mathrm{BQ}}}{K_{\mathrm{BQ}}^2 q + K_{\mathrm{BQ}} + q}$	0	$\frac{-K_{\mathrm{BQ}}}{K_{\mathrm{BQ}}^2 q + K_{\mathrm{BQ}} + q}$	$\frac{2q(K_{\mathrm{BQ}}^2 - 1)}{K_{\mathrm{BQ}}^2 q + K_{\mathrm{BQ}} + q}$	$\frac{K_{\mathrm{BQ}}^2 q - K_{\mathrm{BQ}} + q}{K_{\mathrm{BQ}}^2 q + K_{\mathrm{BQ}} + q}$

Table 5: Bi-quad coefficient values (adapted from [20])

of Table 3 can be used to avoid the unnecessary computation of the imaginary part (quadrature) and thus improve computing efficiency.

5.2.2 Constant-ERB-Bandwidth Specification

Moreover, the multi-resolution filter bank specification is based on a constant-ERB-bandwidth approach rather than the constant-Q approach of Juillerat's design. This specific multi-resolution approach yields the same advantages and drawbacks than those described in Section 4.3, that is an increased frequency resolution for low frequency components, lowered latency for high frequency components and the generation of a downward chirp artefact that is rendered audible with percussive transients in the input signal.

The comparative advantage of using a constant-ERB-bandwidth approach for setting the time-frequency resolution of the filter bank lies in the possibility of adjusting the amount of generated roughness in a more uniform matter across the sub-bands than what can be achieved through a non-perceptually approaches such as the constant-Q. In this approach the filter bandwidths of the analysis filters are set to a same fraction of ERB critical bandwidth as pictured in Figure 16. The portion of critical bandwidth occupied by each sub-band is thus equal across the auditory band. The roughness introduced by the intermodulation components located in the sub-bands can thus be expected to be the same across the bands (actually it can only be expected to be more uniform than the constant-Q method, see next paragraph). In turn, a more optimal reduction of the amount of roughness follows when the ERB bandwidth specification is sharpened at the expense of higher latency and increased downwards chirp effect.

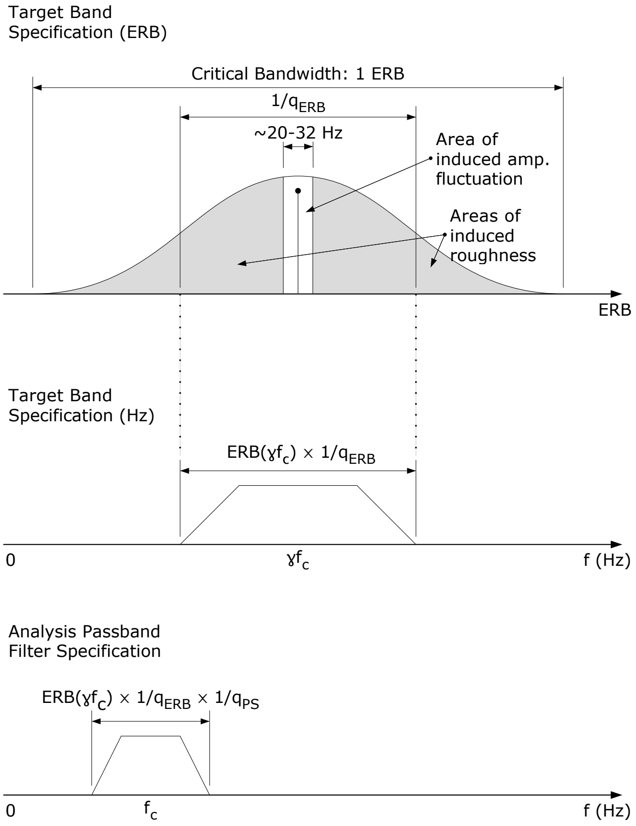


Figure 16: Constant-ERB-bandwith specification principle represented for a single analysis filter of the filter bank centered around f_c Hz. q_{ERB} is the constant-ERB-bandwidth design parameter that determines the time-frequency resolution tradeoff of the filter bank. $\mathrm{ERB}\big(\gamma f_c\big)$ denotes the corresponding target critical bandwidth in Hz for the analysis filter of the example. q_{PS} is the scaling factor that takes account for the broadening of the sub-band following phase scaling: $q_{\mathrm{PS}} = \gamma$ for PS and $q_{\mathrm{PS}} = 1$ for SSM.

This constant-ERB-bandwidth specification approach could be improved further to take account for the fact that critical bands do not contribute equally to the roughness percept. In particular, amplitude modulations produce proportionally more perceived roughness in the critical band around 1 kHz band than in any other

[46]. A future study could take account of that phenomenon to provide further sharpening of frequency resolution where sensitiveness to roughness is highest.

It is understood that the above principle holds when applied, as pictured in Figure 16 relative to the target critical bands in which the sub-band components will land after the pitch-shifting operation and not the critical bands from which the sub-band components originate: it is indeed the amount of roughness in the pitch shifted output that we wish to control. In practice, the bandwidth specification for each analysis filter can be set according to the following expression:

$$\Delta f(k) = \frac{1}{q_{\mathrm{PS}} q_{\mathrm{ERB}}} \mathrm{ERB}\big(\gamma f_{\mathrm{c}}(k)\big), \qquad k = 0..K-1, \tag{37}$$

where:

- $f_{\mathrm{c}}(k)$ denotes the center frequency of the analysis filter k.
- $\Delta f(k)$ denotes the bandwidth specification for the analysis filter k.
- K denotes the number of sub-band filters in the filter bank design.
- γ denotes the pitch shifting factor, that is $\gamma = 2$ in our case. Thus $\gamma f_{\mathrm{c}}(k)$ represents the target center frequency of the sub-band after the pitch shifting operation.
- $\mathrm{ERB}\big(\gamma f_{\mathrm{c}}(k)\big)$ denotes the target critical bandwidth (in Hz) as defined by Equation (38), that is the ERB bandwidth in which the content of sub-band k will land after the pitch shifting operation.
- q_{ERB} is the constant-ERB-bandwidth design parameter that determines the time-frequency resolution of the multi-resolution filter bank. When set to values higher than 1, it indicates by what factor the amount of intermodulation has been reduced in the critical bands.
- q_{PS} is a scaling factor that takes account for the broadening of the sub-band following phase scaling: $q_{\mathrm{PS}} = \gamma$ for PS and $q_{\mathrm{PS}} = 1$ for SSM (no broadening occurs from this modulation technique).

The ERB critical bandwidth (in Hz) is given as a function of its center frequency f_{ERB} (in Hz) by the following equation [46][30]:

$$\mathrm{ERB}(f_{\mathrm{ERB}}) = 24.7 + 0.108 f_{\mathrm{ERB}}. \tag{38}$$

Crossover Attenuation The sub-bands' target center frequencies $\gamma f_{\mathrm{c}}(k)$ are preferably defined so as to be uniformly distributed on the ERB scale within the target auditory band:

$$\gamma f_{\mathrm{c}}(k) = g_{\mathrm{ERBtoHz}}\Big(\mathrm{z}_{\mathrm{left}} + k\frac{q_{\mathrm{C}}}{q_{\mathrm{ERB}}}\Big) \qquad k = 0..K-1, \tag{39}$$

where:

[30] p. 167

- $\mathrm{z}_{\mathrm{left}}$ denotes the leftmost center frequency of the filter bank sub-bands in ERB units, and K denotes the number of bands required to cover the auditory band.
- $g_{\mathrm{ERBtoHz}}(.)$ denotes the ERB to Hz scale conversion function [46] as defined by equation (40).
- q_{C} is a design parameter that determines the distance on the ERB scale between the center frequency of two consecutive sub-bands.

The ERB to Hz scale conversion function yields the values for the target sub-band center frequency in Hz [46]:

$$g_{\mathrm{ERBtoHz}}(z) = 228.7(10^{\frac{z}{21.3}} - 1). \tag{40}$$

Given the above, the design parameter q_{C} determines the crossover attenuation between two neighbouring bands, or equivalently, the number of required bands K. High attenuation values at crossover requires spacing the sub-bands further appart leading to fewer number of analysis filters, and thus reduced complexity. Moreover, relatively low amplitude is ensured in the overlapping transition bands, thus limiting the emergence of additional roughness upon summation of the sub-bands (see Section 5.1.1).

6 Material and Methods

A set of candidate solutions for real-time polyphonic octave doubling are tested and compared in terms of their performance with regards to that of industry standards offerings of the market. These comprise the POG line of effect pedals for the guitar manufactured by Electro-Harmonix. In particular, we take the original POG, the Micro POG and the Slammi models as baselines for our study.

6.1 Evaluation Criteria

Baseline and candidate methods are evaluated according to their response to a set of test signals including synthetic and pre-recorded guitar performances.

6.1.1 Steady-State Sound Quality

Distortion An informal evaluation of the minimum amount of distortion generated at the output by each candidate and baseline is ensured by means of spectrogram plots of a sine sweep synthetic signal covering a band extending from 20 Hz to 20 kHz.

Roughness Estimation A measure of the roughness of the sound at the output of each candidate and baseline is used as an objective indicator of the perceived steady-state sound quality of the solution. Modelling of human hearing is an ongoing area of research and designing an accurate model for such a task requires considerable knowledge and know-how in the field. This would extend well beyond the scope of the current work. Instead we used an existing implementation provided by Vassilakis and made available online through a web form interface [55][31].

A fixed 1kHz pure sinusoid ($\sim$ 15.5 ERB) crossed by a slowly swept sinusoïd of equal amplitude are used as the input test signal for the estimation of roughness. Specifically, the swept sinusoïd's frequency is made to cover the range from 700 Hz to 1370 Hz in a logarithmic fashion within a time of 47 seconds.

Informal Listening Test Further confirmation of the sound quality level is ensured through informal listening test of recorded guitar performances processed by the baseline pedals and candidates methods.

6.1.2 Latency

The overall latency incurred for each candidate method (including DSP platform model latency) is evaluated to that of the baseline effect pedals directly by means of spectrogram plots of candidates' and baselines' responses to an impulse test signal.

[31] This implementation was used with the normalised amplitude option, absolute threshold set to 0, time interval of 250 ms and frequency resolution of 10 Hz.

6.1.3 Transient Preservation

According to Blauert, the human ear is most sensitive to group delay distortion when it is incurred on brief sound impulses [3]. Accordingly, smearing in the time domain is assessed informally by means of spectrogram plots of the candidate and baseline responses to an impulse signal as for the evaluation of latency. Moreover, an informal listening comparison test is provided by means of a guitar performance recording consisting in plucks, rhythmic ornamentations and scratchings of muted guitar strings.

6.2 Baseline Measurement Setup

Measurement Platform Setup A block diagram of the setup employed for recording the output of the POG pedals in response to the test signals is provided in Figure 17 (bottom). As represented, the test signal is sent simultaneously through two recording loops. The first recording loop links a first output channel of the sound card to the input of the POG pedal and the output of the POG pedal to a first recording input of the sound card. A second recording loop links a second output channel of the sound card directly to a second recording input of the sound card. This setup allows to record a loop-back dry signal and a wet signal in response to each test signal. The latency induced by the sound card and buffer setup of the digital audio workstation used to carry out the measurements is thus compensated. In particular, the loop-back dry signals were used as inputs to the candidate solutions so as to render their output directly comparable to that of the POG baseline pedals.

Baseline Pedals' Setup The response of each baseline POG pedal is measured with the pedal set at a 100%[32] wet setting with transposition set to the next octave: the dry signal is set to zero as well as any higher or lower octave transpositions. Pictures of the pedals during measurement are provided in Figure 17 as further reference of their setup.

In the case of the original POG pedal, the rightmost filter slider was erroneously set to the middle position. Consequently the low-pass post-processing filter of the pedal was not appropriately bypassed. Moreover, the Slammi's pitch shifting control was set coarsely around a pitch factor of two: a dissonant chord was noticed in the late development of this work when blending this pedal's wet and dry signals (see this thesis's online companion web page[33]). It should be noted that author believes no error was committed in setup of the Micro POG pedal.

6.3 Latency Specification

DSP Platform Model Figure 18-a) is a schematic representations of the integration of a pitch-shifting algorithm on a target real-time DSP platform. During each

[32] With the sole exception of the measurement of Figure 19.

[33] http://research.spa.aalto.fi/publications/theses/thuillier_mst/

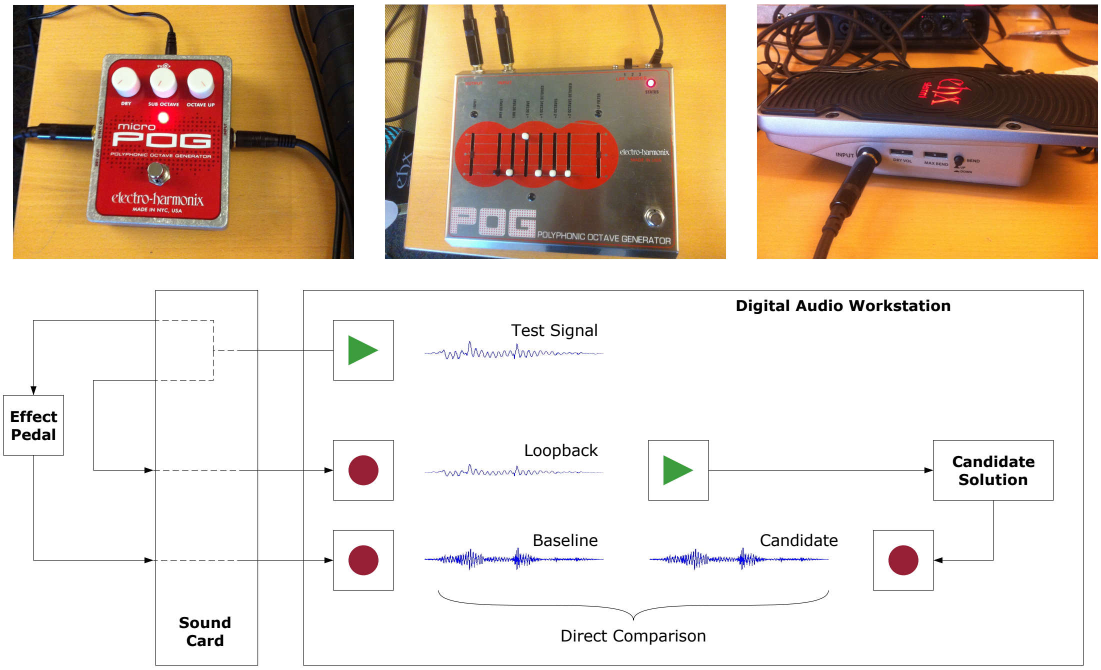


Figure 17: Measurement setup block diagram (bottom) and photograph pictures of effect pedal setups: Micro POG (top left), POG (top middle) and Slammi (top right).

execution of the pitch shifting algorithm, an input and output buffer of the platform continuously stores and output samples respectively. These input/output operations are carried out across the executions of the pitch-shifting algorithm, alternatively in one of the two memory slots of each buffer. The pitch shifting algorithm processes the block of samples stored during the previous iteration and stores the result in the idle output buffer slot for it to be output at the next iteration. This architecture ensures optimal dimensioning of the DSP processor, leading to minimal platform realisation cost. Indeed, the DSP can be chosen to fit a target signal block pitch shift computation load to be carried out within the block's duration so as to prevent as much as possible DSP idle time while still preventing output buffer under-run.

DSP Platform Latency The latency incurred from the hardware DSP platform is not negligible, typically a couple of milliseconds [29]. It corresponds to twice the slot size of the input and output buffers in the above model.

Baseline DSP Platform Latency Figure 19 represents the responses to an impulse signal of various models of pitch-shifter pedals from Electro-Harmonix when set to a 100% dry output. As pictured, the Micro POG pedal has a 0 ms platform-induced latency at that setting, which indicates a true bypass circuit. Such a circuit links the input signal wire to the output, thus circumventing the ADC-DSP-DAC circuit and its buffer-induced latency. However, the POG and Slammi models do

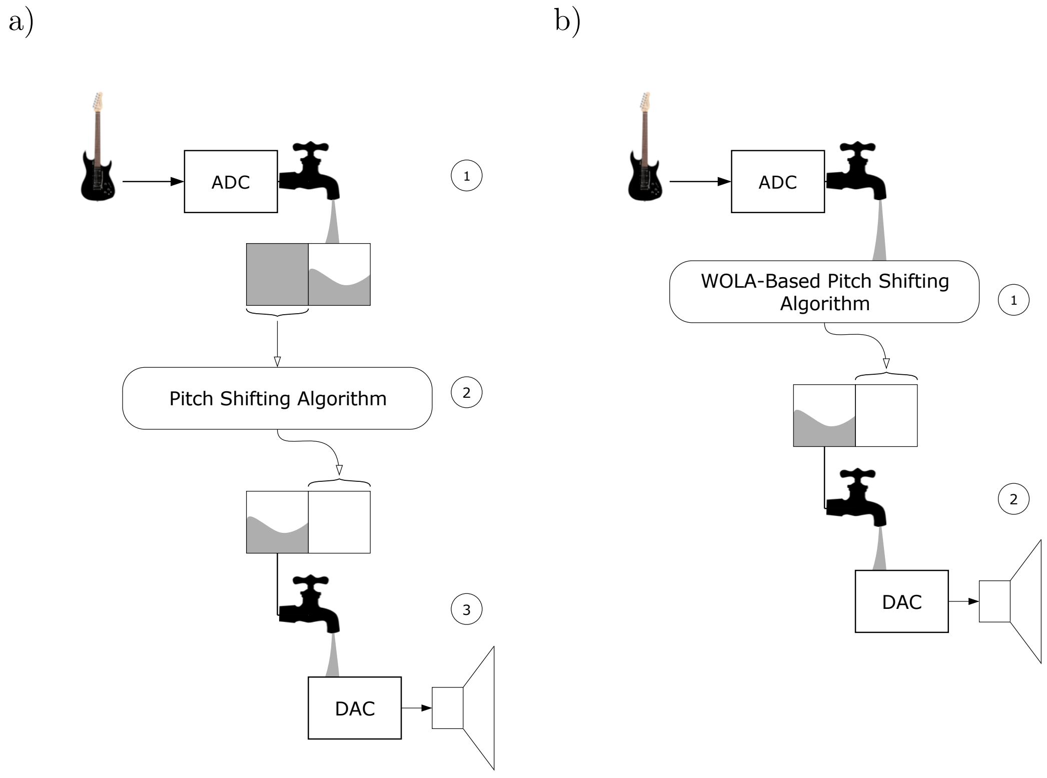


Figure 18: a) Real-time implementation with optimal DSP usage. Process 1, 2 and 3 run in parallel leading to an incurred latency of twice the block size. b) Real-time implementation for STFT-based candidates PV2 and OCEAN2 (the input buffer is implicit to WOLA scheme).

not seem to include such a functionality. They indicate a platform latency of about 1 ms, which corresponds to a block size of approximately 0.5 ms.

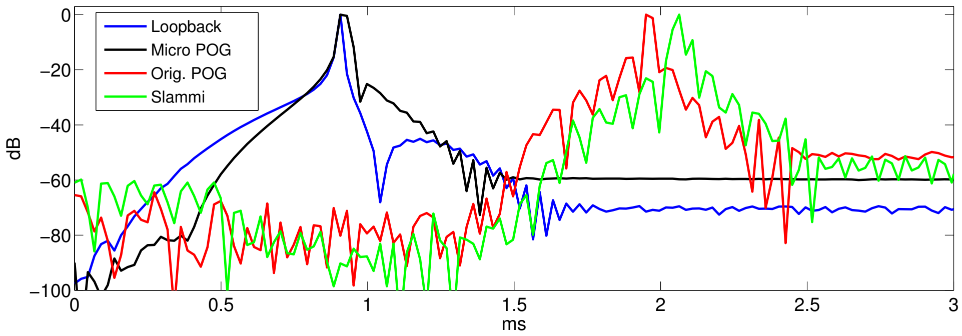


Figure 19: Responses to an impulse signal of various models of pitch-shifter pedals from Electro-Harmonix set to a 100% dry output

Real-Time DSP Platform Simulation Setup The candidate solutions are executed offline under the Matlab environment. The methods are nevertheless implemented and executed in a block-by-block fashion as would be the case on a real-time DSP platform. In the case of audio-rate methods[34], input and output buffers are moreover implemented according to the model of Figure 18-a) so as to achieve similar platform-induced latency than in the baseline pedals. Specifically, a block size of 16 samples and a sample rate of 44100 Hz are employed in the simulations of the audio-rate candidate methods. This corresponds to a combined input and output buffer latency of 0.73 ms, which is similar to the 1 ms platform latency found in the baseline pedals. In the case of STFT-based methods[35], the input buffer of Figure 18-a) is already integrated as part of WOLA operation, see Figure 6, which leads to the DSP platform model of Figure 18-b). In particular, the output buffer size matches that of the WOLA hop size. Given the above, a direct comparison of the recordings made at the output of the baseline pedals with those resulting from simulation of real-time implementations of the candidate solutions is possible, including informal listening tests in which the pitch-shifted signal is mixed with the dry signal as is the intended use of such effect pedals.

End-to-End Candidate Latency Specification Preliminary measurements for the baseline pedals' latencies (100% wet signal) suggests a maximum value of 20 ms for candidate solutions (DSP platform and octave doubling method combined). This figure corresponds to the latency measured for the Micro POG model, which ranked as presenting the highest latency amongst the baseline pedals, see Figure 22.

6.4 Constant-ERB-Bandwidth Filter Bank Setup

The novel solutions of Figure 14 form the two first candidates tested in this study. Their tested implementations comprise an identical analysis bank, the magnitude response of which is pictured in Figure 20. Each complex analysis filters is derived from a real-valued bi-quadratic IIR filter prototype as described in Section 5.2.1. The choice of filter design parameter values is summarised in Table 6 and discussed hereafter.

ERB-PS2 As discussed in Section 5.2.2, an appropriate compromise between the downward chirp artefact and the overall perceived roughness can be found by adjusting the value of the bandwidth design parameter q_{ERB}. In the case of the ERB-PS2 candidate, preliminary tests suggests that setting the modulated sub-bands to a sixth of the target critical band width (i.e. $q_{\mathrm{ERB}} = 6$) yields results comparable to the Micro POG baseline pedal. Moreover, inspection of the effective output bandwidths of the baseline pedals (after octave doubling, see Figure 24) reveals it is acceptable to limit the output band to 160Hz-8kHz. This corresponds approximately to the 5-33 ERB band. Thus $\mathrm{z}_{\mathrm{left}} = 5$ and $\mathrm{BW}_{\mathrm{ERB}} = 28$. The analysis

[34] i.e. ERB-PS2 and ERB-SSM2, see Section 6.4.

[35] i.e. PV2 and OCEAN2, see Section 6.5.

crossover attenuation is set to -12 dB as suggested by Juillerat (see Section 5.1.1) and the design parameter q_C is set accordingly by trial and error.

ERB-SSM2 The filter bank implementation leads to sharper modulated subbands in the case of ERB-SSM2, given that the frequency scaling effect produced by PS2 ($q_{PS} = 2$) does not occur with SSM2 ($q_{PS} = 1$). Filter bank design parameters values for ERB-SSM2 are adapted accordingly in Table 6. Increased sharpness also leads to increased attenuation at crossover.

Method	q_{ERB}	q_C	z_{left}	BW_{ERB}	K	G_a	G_t
ERB-PS2 ($q_{PS} = 2$)	6	4	5	28	43	-12 dB	-12 dB
ERB-SSM2 ($q_{PS} = 1$)	12	8	5	28	43	-12 dB	-24 dB

Table 6: Implementation setup values for ERB-PS2 and ERB-SSM2 candidate solutions. Resulting analysis crossover attenuation G_a, target crossover attenuation G_t and total count K of required analysis complex passband filters.

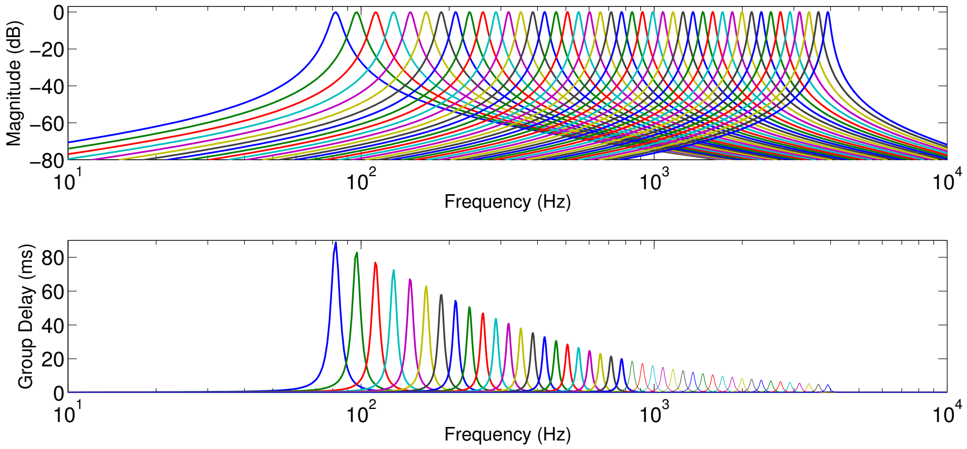


Figure 20: Passband magnitude responses (top) and group delays (bottom) of the constant-ERB-bandwidth filter bank implementation of ERB-PS2 and ERB-SSM2 candidates.

6.5 Uniform DFT Filter Bank Setup

PV2 Two additional candidates are implemented according to the WOLA scheme. Firstly, a phase vocoder (PV2) implementation is carried-out using the time-stretch & resample method pictured in Figure 9 and following Equation (31). The implemented WOLA scheme is pictured in Figure 21. Compared to the WOLA scheme of Section 3.3.2, it comprises additional steps at the analysis and synthesis stages where a circular shift operation is applied. This ensures that the phase of the frequency bins of the FFT are given with a time origin corresponding to the center of the analysis window, and not its origin. While not always a requirement in analysis-synthesis methods, it can have some practical implications, particularly in phase vocoder applications [11]. This was found to be the case in the current PV2 implementation where comparative preliminary tests demonstrated that integration of the circular shift steps improved transient preservation. In the absence of the circular shift steps, the response to an input impulse signal consisted in two successive impulses. In the presence of the circular shift steps, the response corresponded to a single impulse as desired (see Figure 23, next section).

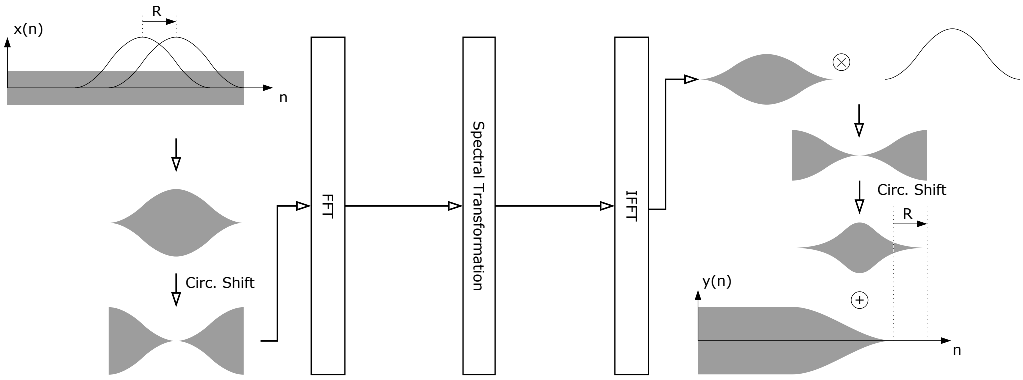


Figure 21: Implemented WOLA scheme for the PV2 candidate.

OCEAN2 Secondly, an implementation (OCEAN2) of Juillerat's OCEAN algorithm is carried out following the description made in Section 4.2.1. The implemented WOLA scheme is almost identical to that of Figure 6 (see Section 3.3.2). Unlike in this figure however, spectral transformation is conducted directly on the complex passband signals as formulated by Equation (17).

Hop Size Specification In the case of PV2, the hop size[36] is set to correspond to the the input/output buffer blocksize of the simulated DSP platform, i.e. to 16 samples (see Section 6.3). In the case of OCEAN2, the transient duplication artefact is minimised by using as little overlap as possible between short time Fourier transform frames, as discussed in Section 4.2.1. Here, 75% overlap is used to ensure the weak COLA requirement is met using a Hann window upon analysis and synthesis.

Analysis and Synthesis Window Duration The window lengths for both STFT solutions are set according to the maximum latency requirement of 20 ms. In

[36] Or equivalently the downsampling factor.

Method	N	R & IO buffer size	Overlap
PV2	1169	16	98.6%
OCEAN2	897	224	75%

Table 7: Implementation setup for the WOLA-based candidate methods.

the case of the PV2 solution, the window length can be derived from equation (32):

$$N \sim \frac{d_{\mathrm{max}} f_{\mathrm{s}}}{\frac{1}{2} + \frac{1}{2\gamma}} - R, \tag{41}$$

where d_{max} denotes the maximum latency requirement (i.e. 20 ms in our case) and f_{s} denotes the sampling frequency (i.e. 44100 Hz in our case). The subtracted R term allows to take into account of the output buffer latency following the real-time model pictured in Figure 18-b).

In the case of the OCEAN2 solution, the window length can be set according to the worst case offline latency pictured in the top left graph of Figure 8 in the specific case of a 75% overlap. This leads to the following empirical expression:

$$N \sim d_{\mathrm{max}} f_{\mathrm{s}}. \tag{42}$$

The specification for the WOLA-based candidate methods are summarized in Table 7. The overlap for the PV2 solution satisfies the requirement of setting the downsampling factor to a lower value than half that specified by the strong COLA requirement in the case of combined analysis and synthesis windows (see Sections 3.3.3 and 4.2.2), that is:

$$\begin{aligned} R &\leq \frac{(N-1)/6}{2}, \\ &\leq 97.3 \text{ samples.} \end{aligned} \tag{43}$$

The above expression being verified for the selected hop size value, i.e. $R = 16$, aliasing is not expected to occur in the sub-bands of the PV2 solution.

7 Results

7.1 Latency and Transient Preservation

Baseline Solutions The processed signal outputs generated by the baseline pedals in response to an impulse test signal were recorded following the measurement setup described in Section 6.2. A time-frequency representation of the recorded signals is given in Figure 22. The time origin of each spectrogram[37] of this Figure was shifted to coincide with the impulse's timing in the loopback recording test signal (top left graph of the Figure). Consequently, readings of the time scale are representative of the total latency introduced by each pedal independently from the measurement setup.

Increased ringing of the output towards the lower frequencies seem[38] to indicate that the Micro POG and original POG share a multi-resolution filter bank design similar to that of Juillerat (see Section 4.3). In particular, the corresponding spectrograms uncovers a downwards chirp signature that is present in Juillerat's design, i.e. the group delay and ringing duration of the system increases towards the low frequencies. Unlike Juillerat's design however, the original POG pedal's response exhibits two distinct phases. In a first phase a rising edge appears around 14 ms, formed by a straight vertical line on the spectrogram (see Figure 22 bottom left). This focused impulse sound component is followed by a second phase formed by the downward chirp signature mentioned above. It is still unclear how this first phase is generated. It is possible that the original POG pedal is equipped with an onset detector and implements a transient-dedicated processing stage that is executed in parallel to the main pitch-shifting stage following detected onsets. The Micro POG does not exhibit a similar early impulse phase. It exhibits a high-frequency latency of 20 ms. The Slammi pedal's response is structured differently with equal ringing (smearing) across the frequencies. This indicates the use of equal bandwidth passband filters across the signal's band. This pedal exhibits a latency slightly above 20 ms for all frequencies.

Candidate Solutions An equivalent time-frequency representation is provided in Figure 23 for the candidate methods. As above, readings of the time scale corresponds to the total latency introduced by each candidate[39].

The spectrogram of the ERB-PS2 presents a downwards chirps signature similar to that of the two POG pedals, and especially to that the Micro POG. The time-frequency representation of the ERB-SSM2 method's response suggests that the pitch shift frequency error it introduces breaks the vertical phase coherence between the bands, which results in a blurred-out version of the downwards chirp signature.

Both ERB candidate method exhibits greatly reduced latency with regards to the baseline pedals. In particular the ERB-PS2 method yields a latency lower than 3 ms for the shifted frequency components located above 3 kHz. This can seem to

[37] 64 samples window and 32 samples hop size at 44100 Hz sampling rate.

[38] The conditional is used purposefully here as pitch shifting is a non-LTI process.

[39] Real-time DSP platform model and candidate method combined.

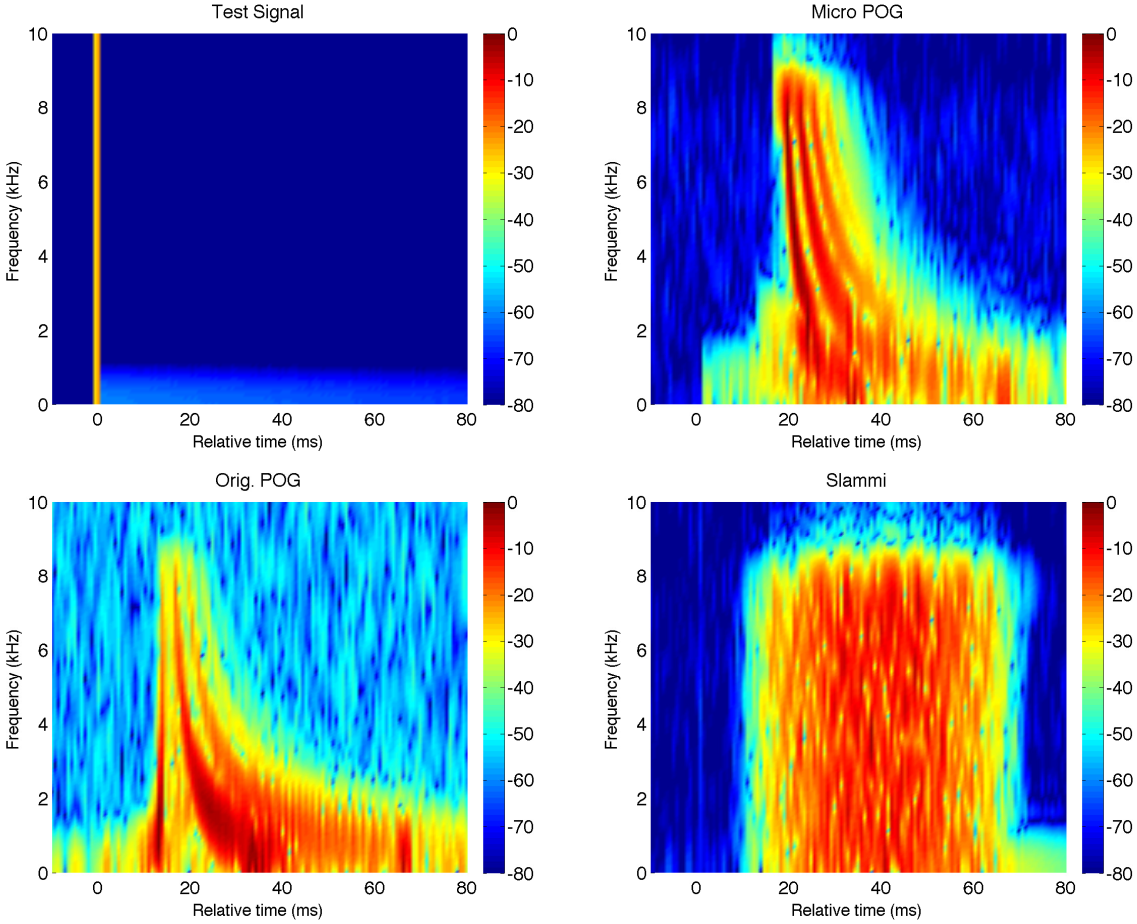


Figure 22: Spectrograms of test impulse signal (top left) and of the resulting outputs measured for the baseline pedals: Micro POG (top right), original POG (bottom left) and Slammi (bottom right).

contradict the analysis filters' calculated group delay curves represented in Figure 20. Indeed, all filters exhibit theoretical group delay values higher than 5 ms at their center frequencies. The quicker response demonstrated by the measurement results from signal energy being distributed left and right of these center frequencies where the group delay is lower (see Figure 20).

Given the above, the novel ERB-based candidate solutions achieve significant latency reduction with regards to the baseline pedals, specifically: 11 ms compared to the original POG pedal, and 17 ms compared to the Micro POG and Slammi pedals. The ringing profiles of these candidate methods are comparable to that of the two POG pedals. It is thus possible that the overhead latency exhibited in the POG models result from a platform integration choice and not directly from the pitch shifting method employed.

By contrast the short-time Fourier based methods yield responses that are more focused in time. In particular, the PV2 method produces an essentially undistorted output with an overall latency matching exactly the design specification (i.e. 20 ms see Section 6.5) as set by Equation (41). Echo images of the impulse are generated with faint amplitudes (about -40 dB with regard to peak impulse) approximately 13 ms before and after the main impulse[40]. The author found no explanations for these low-amplitude artefacts. As expected (see Section 4.2.1), OCEAN2 produces a group of main impulse response duplicates distributed according to the OLA hop size length (i.e. $R/f_s \sim 5.5$ ms) and centred, with a noticeable jitter in this case, around the specified latency (i.e. 20 ms see Section 6.5).

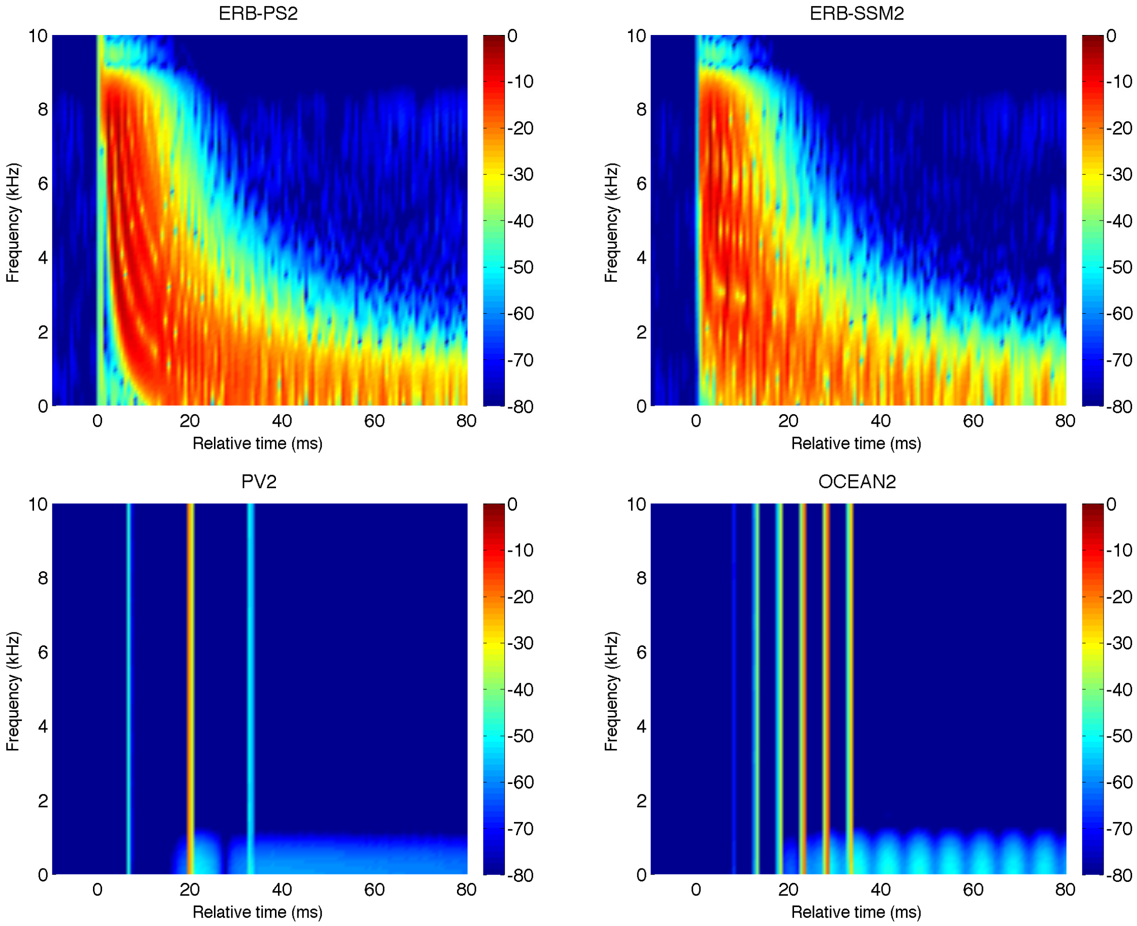


Figure 23: Spectrograms of the output responses to the test impulse signal of Figure 22 as measured for the candidate methods: ERB-PS2 (top left), ERB-SSM2 (top right), PV2 (bottom left) and OCEAN2 (bottom right).

Transient Preservation: Subjective Evaluation

A subjective listening evaluation was carried-out by the author to assess the comparative performance of each candidate method with regards to the baseline pedals

[40] This corresponds to half the duration of the analysis STFT window.

in terms of transient preservation. A recorded guitar performance consisting in a succession of various plucking, rhythmic ornamentation and scrapping techniques performed on muted guitar strings was used as the test signal. Resulting pitch-shifted signals were recorded for all baseline pedals and candidate methods are provided online on a companion web page[41].

As suggested by the spectrogram of Figure 23 (bottom left), the PV2 method preserves the transient crispness of the input signal and provides the highest performance amongst the tested solutions, baseline methods and candidates combined.

According to the author, the Micro POG, the original POG, the ERB-PS2 and ERB-SSM2 method show comparable level of performance. The downwards chirp artefact is relatively faint in all cases. It is challenging to provide a ranking betweens these solutions. It seems however that ERB-SSM2 provides a slightly better performance than the other three. This could be attributed to a time-frequency blurring of the downwards chirp effect resulting from using a fixed frequency translation for each band, thus rendering the downwards chirp less detectable by the human ear.

The transient duplication artefact can be clearly heard for the OCEAN2 solution, resulting in significant performance degradation compared to the PV2 method. However, the author found the transient duplicates artefact from OCEAN2 to be less detrimental to transient crispness than the downward's chirp. Finally, the transients at the Slammi pedal's output are found to be heavily smeared.

7.2 Steady-State Sound Quality

7.2.1 Incurred Out-of-Band Distortion

Baseline Solutions The processed signal outputs generated by the baseline pedals in response to a sine sweep signal described as test signal were recorded following the measurement setup described in Section 6.2. A time-frequency representation of the recorded signals is given in Figure 24.

The spectrograms[42] reveal that both the original POG and Micro POG baseline pedals generate by-product harmonic distortion. Specifically, the resulting distortion is formed by harmonics of the sine sweep frequency starting from the third harmonic (although a faint first harmonic component is also present). The by-product harmonic distortion indicates the presence of a nonlinear process acting on the sine sweep. The level of each distortion ray remains at all time lower to approximately 25 dB below the pitch-shifted sine sweep signal for both POG baseline pedals. The Slammi pedal does not exhibit such harmonic distortion.

Candidate Solutions The responses of the candidate solutions to the sine sweep test signal are pictured in Figure 25. The ERB-PS2 method is the only one to exhibit out-of-band distortion. The distortion components are distributed in the same fashion as that of the two POG pedals. They exhibit a lower magnitude than

[41] http://research.spa.aalto.fi/publications/theses/thuillier_mst/

[42] A 2048-sample window and a 512-sample hop size at 44100 Hz sampling rate were used.

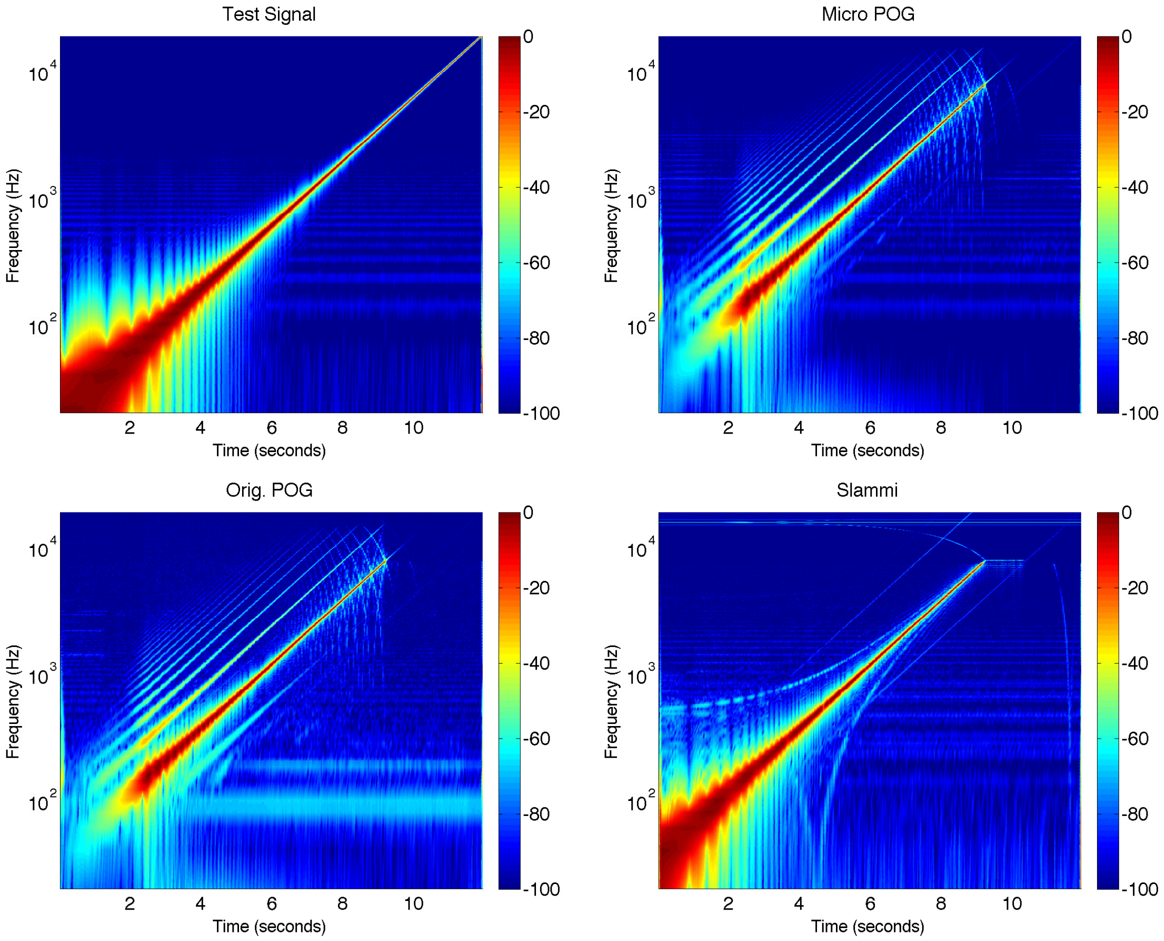


Figure 24: Spectograms of the test sine sweep signal (top left) and of the resulting outputs measured for the baseline pedals: Micro POG (top right), original POG (bottom left) and Slammi (bottom right).

that of these baseline pedals however, being maintained approximately 40 dB below the shifted sine sweep signal.

Occurrence of such out-of-band distortion contradicts our theoretical prediction for the phase scaling method (see Table 4). It is yet unclear how this distortion occurs. Interestingly the PV2 method exhibits no such distortion. This could probably be explained by the relatively different process by which pitch shifting is obtained in this case, that is by downsampling of a time-stretched version of the input signal. Further study would be required to provide an explanation to this phenomenon.

As expected neither SSM-based candidate solutions[43] exhibit out-of-band distortion. The frequency translation errors inherent to this sub-band modulation method however appear on the spectrograms, and is especially clear for OCEAN2.

[43] i.e. ERB-SSM2 and OCEAN2.

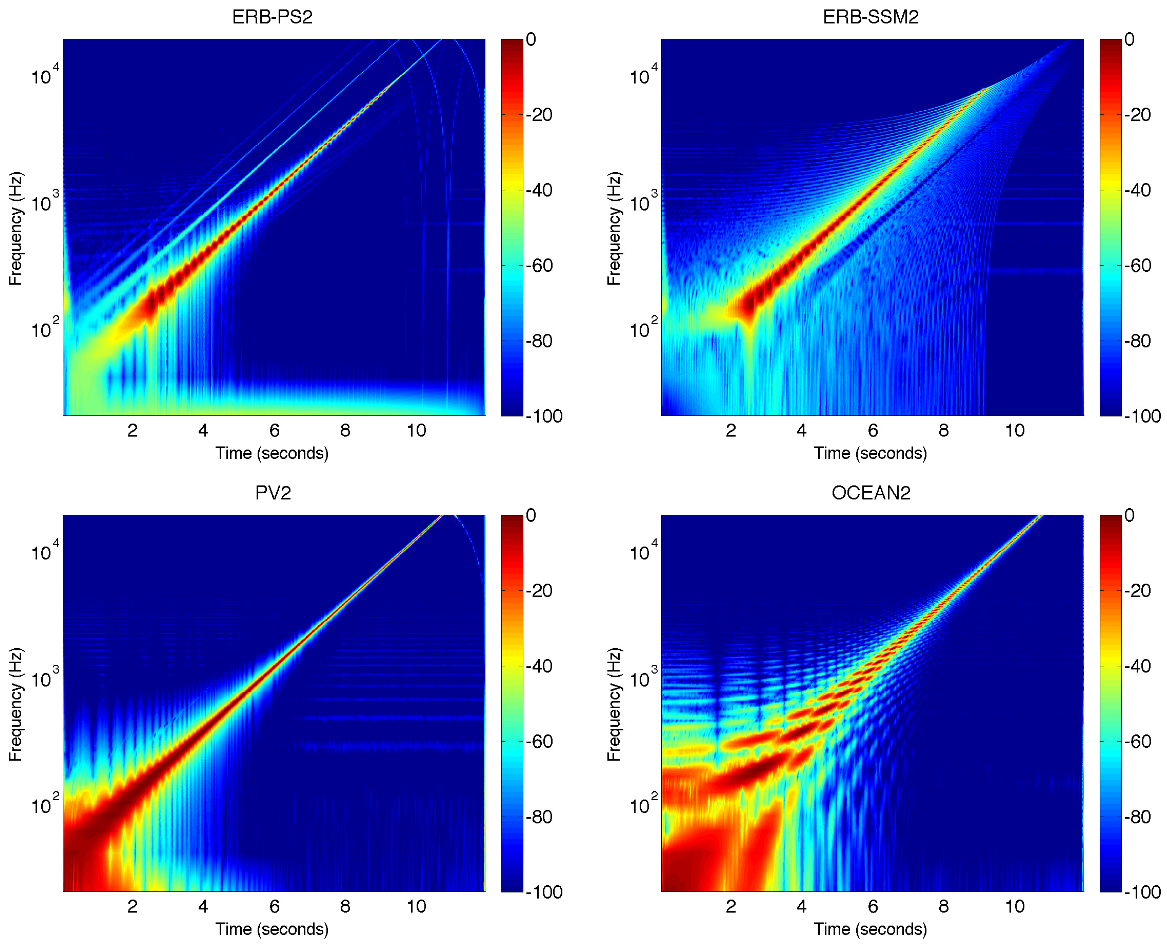


Figure 25: Spectrograms of the output responses to the test sine sweep signal of Figure 24 as measured for the candidate methods: ERB-PS2 (top left), ERB-SSM2 (top right), PV2 (bottom left) and OCEAN2 (bottom right).

7.2.2 Added Roughness: Objective Evaluation

As discussed in Section 6.1.1, an objective evaluation of performance in terms of added roughness is provided using an input test signal composed of a fixed 1 kHz pure tone ($\sim$ 15.5 ERB) mixed with a slowly swept sinusoid of equal amplitude. The swept sinusoid's frequency is made to cover the range from 700 Hz to 1370 Hz in a logarithmic fashion within a time frame of 47 seconds. The bottom graph of Figure 26 represents the spectrogram of a corresponding ideally transposed output signal, obtained by synthesising the sinusoid pair with doubled frequency values. A roughness target curve is generated from this ideal output signal using the web-accessible implementation of Vassilaki's auditory model for roughness estimation[44]. It should theoretically exhibit five successive phases:

- An initial phase and a final phase where the sinusoides reside in separate

[44] See Section 6.1.1.

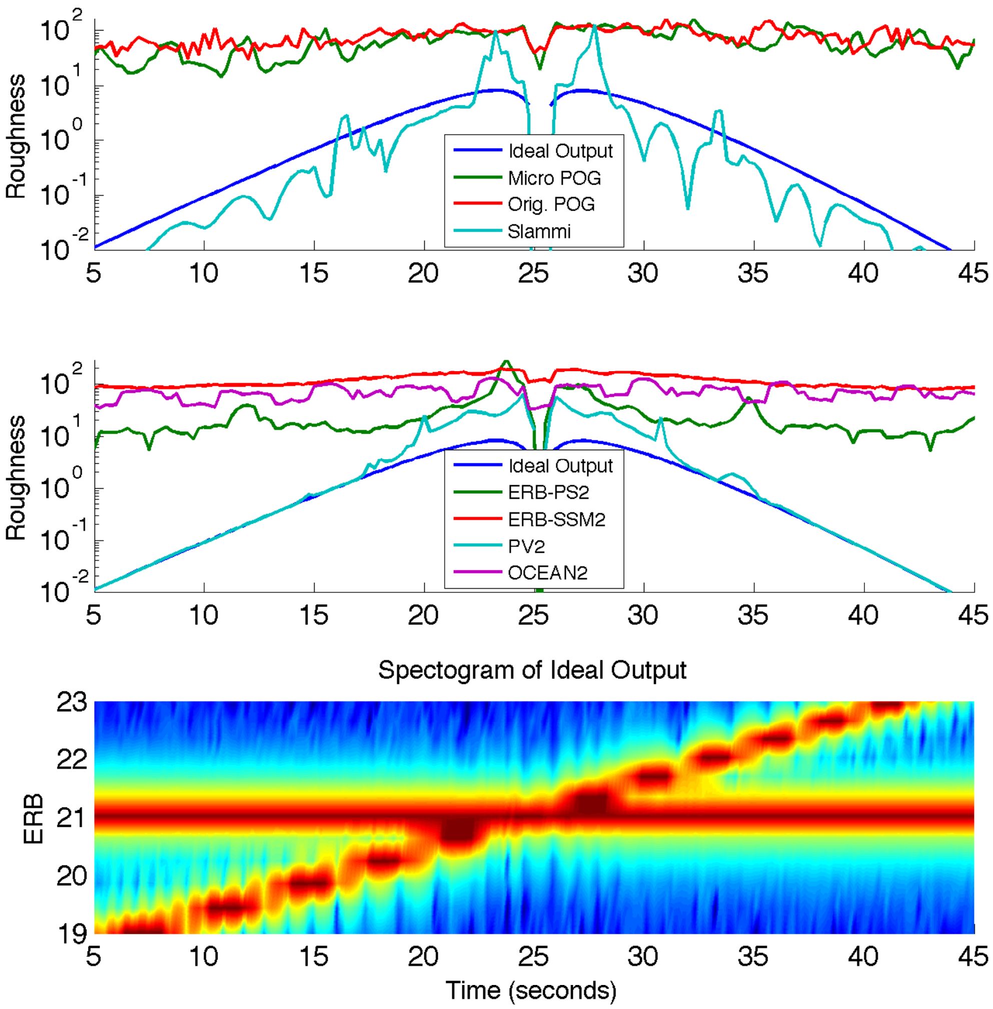


Figure 26: Overview of estimated output roughness curves for all baseline (top) and candidate (middle) solutions in response to an interfering sine sweep test signal. Spectrogram of an ideal octave doubled output (below).

critical bands. The roughness target curve should essentially be of minimal value during these two phases.

- A median (third) phase directly before, during and directly after the time of the sinusoid pair's crossing, when the beatings generated by mistunings result in an amplitude fluctuation percept. The roughness target curve should drop to minimal values during that phase.

- A second phase and a fourth phase directly before and after the median (third) phase, when the sinusoid are close enough in frequency to be located within the same critical band but still sufficiently mistuned to generate a roughness percept. The roughness target curve should rise to maximum values during these two phases.

The target curve is pictured in both top and middle graphs of Figure 26 under the label "Ideal Output". Inspection of the spectrogram reveals that the second, third and fourth phases mentioned above should be located between 15 and 33 seconds approximately. As expected, the roughness curve for the "Ideal Output" comprises maxima in this time period. Moreover, as expected, a minimum is also observed at the time of the sinusoid pair's crossing (around 25 seconds). It should be noted that, for presentation convenience, the roughness values are plotted in an unorthodox fashion in Figure 26, using a logarithmic scale. This allows for providing a qualitative performance comparison between baseline pedals and candidate solutions, which follows.

As pictured in Figure 26, all baseline and candidate solutions generate added roughness compared to the ideally transposed synthesized output[45]. In particular, Micro POG, original POG, ERB-SSM2 and OCEAN2 consistently produce a relatively high level of roughness throughout: contrary to expectations, no drop occurs before 15 seconds and after 33 seconds. In the case of OCEAN2 and ERB-SSM2, this phenomenon is probably caused by the generation of beating pairs within the transition bands as the sine is swept through, given the inherent mistuning errors generated by these solutions (see Section 5.1.1). In the case of the Micro POG and original POG, the sustained level of roughness could originate from the relatively high level of generated out-of-band distortion observed above (see Figure 24).

By contrast, ERB-PS2, PV2 and Slammi generate globally lower-valued roughness curves with maxima shortly before and after the crossing of the sinusoid pair. The author believes the maxima to be caused by intermodulation distortion occurring within the sub-bands as discussed in Section 5.1.1.

The roughness curves of ERB-PS2 and Micro POG are reproduced on a linear scale in Figure 27 so as to provide a quantitative performance comparison in term of output roughness. As pictured, ERB-PS2 outperforms the Micro POG consistently with the exception of a significant roughness spike around 24 seconds, where roughness level is found to be nearly three times that of the Micro POG. Further work could be conducted to identify the cause for this spurious increase and lead, the case being, to further improvement of the ERB-PS2 candidate.

7.2.3 Added Roughness: Subjective Evaluation

Subjective evaluation of sound output roughness was conducted for all solutions from a recorded interpretation of the chord sequence of the song "Where Is My Mind" by the Pixies [53]. As above, the output sound from each candidate solution

[45] An exception to this is the Slammi pedal before 22 seconds and after 49 seconds, which shows lower estimated roughness values than the ideal output. An analysis of Vassilaki's auditory roughness model would be required in order to provide an explanation for this phenomenon, which is beyond the scope of this work.

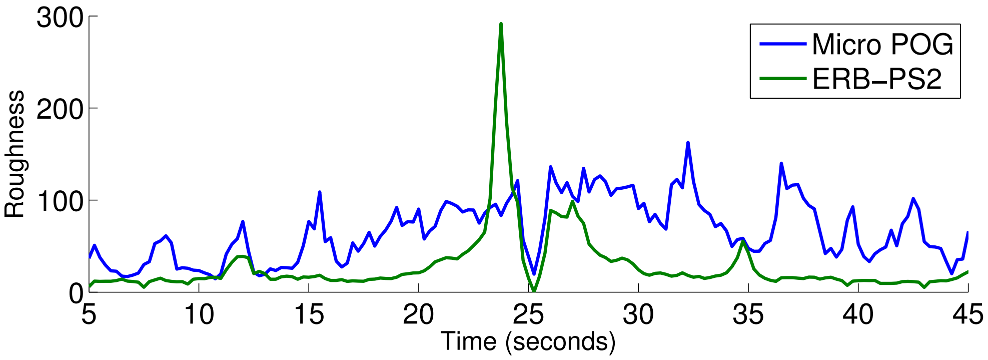


Figure 27: Objective roughness output comparison between the Micro POG baseline pedal and the ERB-PS2 candidate.

and baseline pedal is provided online on the companion web page[46] (see "100% Wet" recordings).

The Slammi and the original POG seem to generate the least roughness in the author's opinion. In the case of the original POG pedal however this observation should probably be discarded as inconclusive given that this pedal's lowpass post-filter was not properly bypassed during measurement. The Micro POG and ERB-PS2 generate very similar roughness according to the author. It would prove difficult to rank one above the other.

The ERB-SSM2 seem to introduce slightly more roughness than the Micro POG pedal and the ERB-PS2 candidate. This can come as a surprise as this method is not subject to intermodulation distortion as mentioned above. Both STFT-based methods exhibit considerably more roughness than the other tested solutions. In addition, both OCEAN2 and ERB-SSM2 exhibit significant mistuning.

[46] http://research.spa.aalto.fi/publications/theses/thuillier_mst/

	Latency	Transient Alteration	Out-of-Band Distortion	Roughness: Pair of pure tones (objective)	Roughness: Guitar chord progression (subjective)	Mistuning
Micro POG	20 ms	chirp ++	++	+++	++	no
Orig. POG	14 ms	chirp ++	++	+++	+	no
Slammi	20 ms	smearing +++	no	-	+	no
ERB-PS2	∼ 3 ms	chirp ++	+	++	++	no
ERB-SSM2	∼ 3 ms	chirp +	no	++++	+++	yes
PV2	20 ms	no -	no	-	++++	no
OCEAN2	20 ms	duplicates +	no	+++	++++	yes

Table 8: Comparative summary of results. Red, orange and green colors respectively indicate poor, average and good performance ranking relative to other candidates and baselines. Minus sign indicates the smallest level observed amongst candidates and baselines. Plus signs express increasingly higher observed levels above said smallest observed level.

8 Summary

In this thesis, four candidate solutions to the problem of real-time polyphonic octave doubling for the guitar were tested. These solutions are individually sumarized hereafter.

PV2: A simplified variant of the time-stretching and resampling Phase Vocoder technique specifically applicable to integer pitch shifting factors as suggested by Laroche [35]. This technique ensures vertical phase coherence without relying on phase locking. The transients are thus preserved with essentially no alterations, in particular without the transient smearing and “phasiness” artifacts that are found in general-purpose implementations of the Phase Vocoder. Moreover the latency introduced by this solution is modest for an analysis-synthesis method. It corresponds to about 75% of that resulting from the perfect reconstruction uniform DFT filter bank on which it is implemented. The author found this fact to be relatively understated in the literature.

OCEAN2: An application of Juillerat's OCEAN method, in which the center frequencies of the STFT sub-bands are relocated on the frequency scale by shifting the bins of the transform to a doubled index position [28]. Each STFT sub-band is thus translated to a doubled frequency position.

ERB-SSM2: A slightly modified implementation of Juillerat's Rollers method [29]. As in OCEAN2, sub-bands of the input signal are translated to a doubled value on the frequency scale in a piecewise manner. The translation of each sub-band is carried out using single-sideband modulation. An appropriately specified multi-resolution filter bank allows to achieve low latency for high frequencies while preserving sufficient frequency resolution at lower frequencies to minimise mistuning errors. Moreover, the filter-bank is devoid of synthesis filters. This significantly reduces latency, e.g. by a factor of approximately two when compared to a solution with identical filters at both the analysis and synthesis stages. The tested implementation was designed using a constant-ERB-bandwidth filter bank specification instead of Juillerat's constant-Q specification. Moreover, in-phase and quadrature sub-band components were extracted using complex-valued analysis filters instead of sub-band Hilbert transform estimators. This simplification contributes to lowering the latency of the system compared to that of Juillerat.

ERB-PS2: A novel solution similar to ERB-SSM2, in which a doubling of the instantaneous sub-band phases is carried out instead of single-sideband modulation. In this case, the purpose of using a multi-resolution design lies in allowing, for a given latency constraint, lower intermodulation distortion within the low frequency sub-bands. As discussed, higher intermodulation results in added roughness at the output. The perceptually grounded constant-ERB-bandwidth specification is here particularly relevant as it acknowledges the relation between sub-band bandwidth, intermodulation distortion and generated roughness within the corresponding critical band. The author found that it allowed to tune the filter-bank's time-frequency sub-band resolutions towards a better balance between the chirp artefact and added roughness tradeoffs than constant-Q.

Evaluation Method Candidate solution performance were compared to a baseline consisting in three selected state of the art guitar effect pedals of the market, namely the original POG, the Micro POG and the Slammi from Electro-Harmonix. Subjective and objective evaluations were carried out to assess the level of added roughness and transient alterations introduced by each of the candidate solutions and baseline pedals.

Candidate Setup PV2 and OCEAN2 were set up to match the worst latency found in the baseline pedals, that is 20 ms. Specifically, a Hann window of 1161 samples[47] was employed at the analysis and synthesis stage of the PV2 method. A shorter Hann window of 897 samples was used in the case of OCEAN2 upon analysis and synthesis stages. The filter-banks specification of ERB-SSM2 and ERB-PS2 were set by trial an error to approximately match the transient chirp effect found in the Mircro POG and original POG baseline pedals.

Results The ERB-PS2 developped in this work provided the best performance amongst the candidates. Most importantly it provides greatly reduced latency compared to the baseline solutions with comparable, and in some case improved, sound quality. In particular, ERB-PS2 solution provided a maximum latency of 3 ms in the 3 kHz - 8 kHz bands of the pitch shifted signal. This constitutes an improvement of 11 ms with regard to the original POG pedal and 17 ms with regards to the Micro POG pedal. The ERB-PS2 solution thus meets high standards of low latency requirement, often deemed impossible for real-time polyphonic pitch shifting applications. Further, it opens possibility for a reduction of the added roughness artefact under relaxed latency requirement. In particular, the analysis filter bandwidths could be reduced by an identical bandwidth value (in Hz) so as to improve the frequency resolution to the limit of the relaxed latency requirement without significantly modifying the group delay differences across the sub-bands, and thus without perceptible aggravation of the downwards chirp artefact.

Future Work This possible improvement could be explored in a future study. Future work could also include:

- Implementation of a hybrid ERB-PS2-SSM2 solution. Such a solution could allow to blur-out the lower end of the downwards chirp artefact similarly to what seems to be occurring in the Micro POG pedal (see Figure 22).

[47] For a sampling rate of 44100 Hz.

- Identifying the source of the harmonic distortion observed at the output of the ERB-PS2 solution in response to the sine sweep. This would provide valuable insights for improving future designs.

References

[1] Jont B. Allen and Lawrence R. Rabiner. A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE*, 65(11):1558–1564, 1977.

[2] D. Arfib, F. Keiler, U. Zölzer, V. Verfaille, and J. Bonada. *Time-Segment Processing*, volume DAFX: Digital Audio Effects, chapter 7, pages 219–278. John Wiley & Sons, Ltd, Chichester, UK, second edition, 2011.

[3] J. Blauert and P. Laws. Group delay distortions in electroacoustical systems. *The Journal of the Acoustical Society of America*, 63(5):1478–1483, 1978.

[4] J. Bonada. Audio time-scale modification in the context of professional audio post-production. *Research work for PhD program, Universitat Pompeu Fabra, Barcelona, Spain*, 2002.

[5] Russell Bradford, Richard Dobson, et al. Real-time sliding phase vocoder using a commodity GPU. In *Proc. of the International Computer Music Conference (ICMC)*, pages 587–590, Huddersfield, UK, 2011.

[6] Russell Bradford, Richard Dobson, and John Ffitch. The sliding phase vocoder, August 2007. Presentation Slides for the 2007 International Computer Music Conference (ICMC), Copenhagen, Denmark.

[7] Judith C. Brown and Miller S. Puckette. A high resolution fundamental frequency determination based on phase changes of the Fourier transform. *Journal of the Acoustical Society of America*, 94(2):662–667, 1993.

[8] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

[9] Roland Corporation. The history of Boss compact pedals. http://www.roland.co.uk/blog/the-history-of-boss-compact-pedals/, 2013. Accessed: 04-08-2015.

[10] Ronald E Crochiere. A weighted overlap-add method of short-time Fourier analysis/synthesis. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 28(1):99–102, 1980.

[11] Amalia De Götzen, Nicola Bernardini, and Daniel Arfib. Traditional implementations of a phase vocoder: the tricks of the trade. In *Proc. Workshop on Digital Audio Effects (DAFx-00), Verona, Italy*, 2000.

[12] Mark Dolson. The phase vocoder: A tutorial. *Computer Music Journal*, pages 14–27, 1986.

[13] David Dorran. *Audio time-scale modification*. PhD thesis, Dublin Institute of Technology, Dublin, Ireland, 2005.

[14] David Dorran and Robert Lawlor. An efficient audio time-scale modification algorithm for use in a subband implementation. In *Proc. of the International Conference on Digital Audio Effects*, pages 339–343, London, UK, 2003.

[15] David Dorran and Robert Lawlor. Time-scale modification of music using a synchronized subband/time-domain approach. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages iv–225, Montreal, Canada, 2004.

[16] David Dorran, Robert Lawlor, and Eugene Coyle. A comparison of time-domain time-scale modification algorithms. In *Audio Engineering Society Convention 120*, Paris, France, 2006.

[17] Jonathan Driedger and Meinard Müller. TSM toolbox: Matlab implementations of time-scale modification algorithms. In *Proc. of the 17th Int. Conference on Digital Audio Effects (DAFx-14)*, pages 249–256, Erlangen, Germany, September 2014.

[18] P. Dutilleux, G. De Poli, A. von dem Knesebeck, and U. Zölzer. *Time-Segment Processing*, volume DAFX: Digital Audio Effects, chapter 6, pages 185–217. John Wiley & Sons, Ltd, Chichester, UK, second edition, 2011.

[19] P. Dutilleux, K. Dempwolf, M. Holters, and U. Zölzer. *Nonlinear Processing*, volume DAFX: Digital Audio Effects, chapter 4, pages 101–138. John Wiley & Sons, Ltd, Chichester, UK, second edition, 2011.

[20] P. Dutilleux, M. Holters, S. Disch, and U. Zölzer. *Filters and Delays*, volume DAFX: Digital Audio Effects, chapter 2, pages 47–81. John Wiley & Sons, Ltd, Chichester, UK, second edition, 2011.

[21] P. Dutilleux, M. Holters, S. Disch, and U. Zölzer. *Modulators and Demodulators*, volume DAFX: Digital Audio Effects, chapter 3, pages 83–99. John Wiley & Sons, Ltd, Chichester, UK, second edition, 2011.

[22] Electro Harmonix. POG2 : Polyphonic Octave Generator. http://www.ehx.com/products/pog2. Accessed: 04-08-2015.

[23] J. L. Flanagan and R. M. Golden. Phase vocoder. *Bell System Technical Journal*, 45(9):1493–1509, 1966.

[24] Azadeh Haghparast, Henri Penttinen, and Vesa Välimäki. Real-time pitch-shifting of musical signals by a time-varying factor using normalized filtered correlation time-scale modification (NFC-TSM). In *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, pages 7–13, Bordeaux, France, September 2007.

[25] Christian Hamon, Eric Moulines, and Francis Charpentier. A diphone synthesis system based on time-domain prosodic modifications of speech. In *Proc.*

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pages 238–241, Glasgow, Scotland, May 1989.

[26] Fredric J Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

[27] Eric Jacobsen and Richard Lyons. The sliding DFT. *IEEE Signal Processing Magazine*, 20(2):74–80, 2003.

[28] Nicolas Juillerat and Béat Hirsbrunner. Low latency audio pitch shifting in the frequency domain. In *Proc. International Conference on Audio Language and Image Processing (ICALIP)*, pages 16–24, Shanghai, China, 2010.

[29] Nicolas Juillerat, Simon Schubiger-Banz, and Stefan Müller Arisona. Low latency audio pitch shifting in the time domain. In *International Conference on Audio, Language and Image Processing (ICALIP)*, pages 29–35, Shangai, China, 2008.

[30] Thorsten Karrer, Eric Lee, and Jan Borchers. PhaVoRIT: A phase vocoder for real-time interactive time-stretching. In *Proc. International Computer Music Conference (ICMC)*, pages 708–715, New Orleans, USA, 2006.

[31] Timo I. Laakso, Vesa Välimäki, Matti Karjalainen, and Unto K. Laine. Splitting the unit delay - tools for fractional delay filter design. *IEEE Signal Processing Magazine*, 13(1):30–60, 1996.

[32] Nelson Posse Lago and Fabio Kon. The quest for low latency. In *Proc. International Computer Music Conference (ICMC)*, pages 33–36, Miami, USA, 2004.

[33] Jean Laroche. Autocorrelation method for high quality time/pitch scaling. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 131–134, New Paltz, NY, USA, October 1993.

[34] Jean Laroche and Mark Dolson. Phase-vocoder: About this phasiness business. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (ASSP)*, page 4, New Paltz, NY, USA, October 1997.

[35] Jean Laroche and Mark Dolson. Improved phase vocoder time-scale modification of audio. *IEEE Transactions on Speech and Audio Processing*, 7(3):323–332, 1999.

[36] Jean Laroche and Mark Dolson. New phase-vocoder techniques for pitch-shifting, harmonizing and other exotic effects. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 91–94, New Paltz, New York, October 1999.

[37] Eric Lee, Thorsten Karrer, and Jan Borchers. An analysis of startup and dynamic latency in phase vocoder-based time-stretching algorithms. In *Proc. International Computer Music Conference (ICMC)*, volume 2, pages 73–80, Copenhagen, Denmark, August 2007.

[38] Francis F. Lee. Time compression and expansion of speech by the sampling method. *J. Audio Eng. Soc*, 20(9):738–742, 1972.

[39] Sungjoo Lee, Hee Dong Kim, and Hyung Soon Kim. Variable time-scale modification of speech using transient information. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 1319–1322, Munich, April 1997.

[40] Heinrich W. Löllmann and Peter Vary. *Low Delay Filter-Banks for Speech and Audio Processing*, volume Speech and Audio Processing in Adverse Environments of *Signals and Communication Technology*, chapter 2, pages 13–61. Springer Berlin Heidelberg, 2008.

[41] John Makhoul and Amro El-Jaroudi. Time-scale modification in medium to low rate speech coding. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 11, pages 1705–1708, Tokyo, Japan, 1986.

[42] Michael R. Portnoff. Implementation of the digital phase vocoder using the fast Fourier transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(3):243–248, 1976.

[43] Miller Puckette. Phase-locked vocoder. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (ASSP)*, pages 222–225, 1995.

[44] Miller Puckette. *The Theory and Technique of Electronic Music*. World Scientific Publishing Company Inc., River Edge, NJ, USA, 2007.

[45] Miller S. Puckette and Judith C. Brown. Accuracy of frequency estimates using the phase vocoder. *IEEE Transactions on Speech and Audio Processing*, 6(2):166–176, 1998.

[46] Ville Pulkki and Matti Karjalainen. *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*. Wiley, 2015.

[47] Axel Röbel. *Signal modifications using the STFT*. http://recherche.ircam.fr/anasyn/roebel/amt_lecture.html, August 2006. Accessed: 01-08-2015. Presentation Slides.

[48] Salim Roucos and Alexander M. Wilgus. High quality time-scale modification for speech. In *Proc. IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP)*, volume 10, pages 493–496, Tampa, Florida, USA, 1985.

[49] Julius O. Smith. *Introduction to Digital Filters with Audio Applications*. http://ccrma.stanford.edu/~jos/filters/, Accessed: 01-08-2015. Online book.

[50] Julius O. Smith. *Mathematics of the Discrete Fourier Transform (DFT)*. http://ccrma.stanford.edu/~jos/mdft/, Accessed: 01-08-2015. Online book, 2007 edition.

[51] Julius O. Smith. *Spectral Audio Signal Processing*. http://ccrma.stanford.edu/~jos/sasp/, Accessed: 01-08-2015. Online book, 2011 edition.

[52] Ernst Terhardt. On the perception of periodic sound fluctuations (roughness). *Acta Acustica united with Acustica*, 30(4):201–213, 1974.

[53] Charles Thompson. Where is my mind? [Recorded by The Pixies], 1988. On Surfer Rosa [CD], Elektra, 2003.

[54] Parishwad P. Vaidyanathan. *Multirate Systems and Filter Banks*. Pearson Education, India, 1993.

[55] Pantelis N. Vassilakis and K. Fitz. SRA: A web-based research tool for spectral and roughness analysis of sound signals. In *Proc. of the 4th Sound and Music Computing (SMC) Conference*, pages 319–325, Athens, Greece, 2007.

A Bandlimited Full Wave Rectifying

The absolute value function can be approximated by Taylor series expansion so as to provide a modulator with bandlimited output in response to a band-limited input. Specifically, the function can rewritten into the form $(1 + x_b)^\alpha$ where $x_b = x^2 - 1$ and $\alpha = \frac{1}{2}$. This corresponds to the binomial case of the Taylor series expansion. Thus:

$$|x| = \sum_{k=0}^{\infty} \binom{\frac{1}{2}}{k} (x^2 - 1)^k. \tag{44}$$

The third order approximation is:

$$|x| \approx \frac{5}{16} + \frac{15}{16}x^2 - \frac{5}{16}x^4 + \frac{1}{16}x^6. \tag{45}$$

The above approximation has a nonzero value at the origin. An offset can nevertheless be applied for proper repositioning at the y-axis origin. This however requires scaling the x-axis so as to ensure that the approximation function of |x| has maximum value 1 which is reached at $x = \pm 1$. The scaled an offseted approximation meeting these requirements is:

$$|x(t)|_3 = \mathrm{c}_1 x^2 + \mathrm{c}_2 x^4 + \mathrm{c}_3 x^6, \tag{46}$$

where $\mathrm{c}_1 = 1.60$, $\mathrm{c}_2 = -0.91$ and $\mathrm{c}_3 = 0.31$.

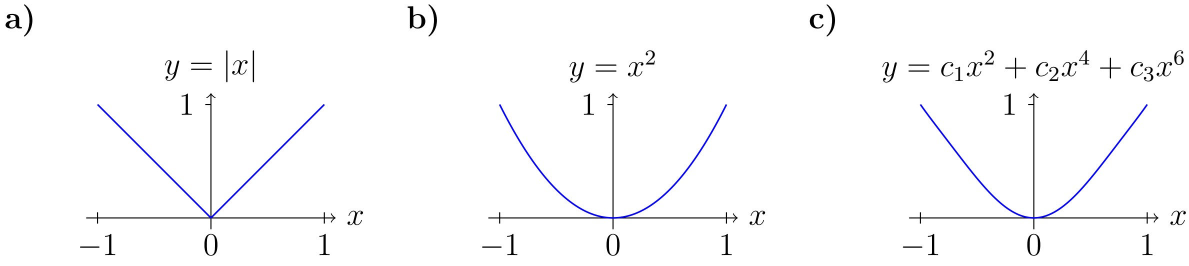


Figure 28: **a)** Full-wave rectifying **b)** Power of two **c)** Band-limited full-wave rectifying

As represented in Figure 28, the approximated waveshape curve comprise a smooth transition around $x = 0$, unlike the full-wave rectifier. Consequently, if the input is ban limited, the output signal is band-limited also. More specifically, the input signal's bandwidth is expanded by a factor corresponding to the degree of the Taylor series approximation. Figure 29 represents the waveshape response curves in decibel for the power-of-two, full-wave rectifier and its approximation. As pictured, the relatively low-order approximation of the full-wave rectifier only provides a slight improvement of a couple of decibels over the power-of-two case. It is doubtful that such modest improvement would justify the increased computational cost. In alternative, post-processing solutions for compensating the nonlinear response of the power-of-two waveshaper could prove more advantageous. For example, a dynamic compressor could be employed at waveshaper's output. Such a solution however incurs added latency.

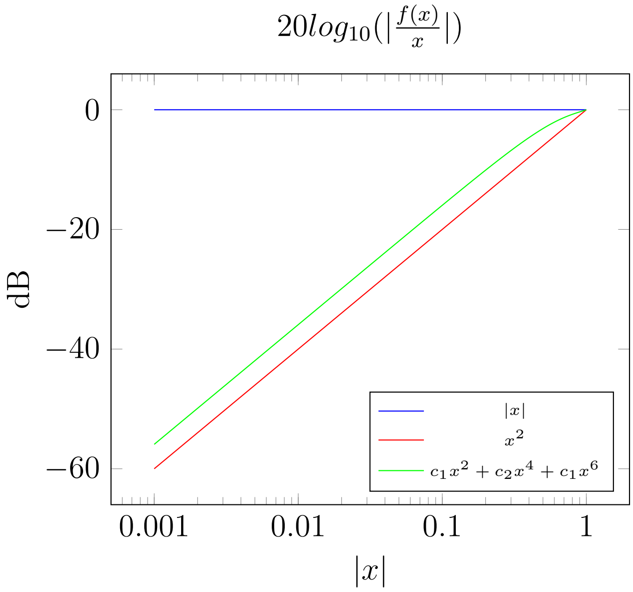


Figure 29: Dynamic response of the full-wave rectifier, power-of-two and band-limited full-wave rectifier waveshapers.