Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing
Espoo 2004 Report 73

Model-Based Analysis of Noisy Musical Recordings with Application to Audio Restoration

Paulo A. A. Esquef

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission for public examination and debate in Auditorium S4, Department of Electrical and Communications Engineering, Helsinki University of Technology, Espoo, Finland, on the 2nd of April, 2004, at 12 o'clock noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Laboratory of Acoustics and Audio Signal Processing

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Akustiikan ja äänenkäsittelytekniikan laboratorio

Helsinki University of Technology
P.O.Box 3000
FIN-02015 HUT
Tel. +358 9 4511
Fax. +358 9 460 224
E-mail lea.soderman@hut.fi

ISBN 951-22-6945-7
ISSN 1456-6303

Otamedia Oy
Espoo, Finland 2004

HELSINKI UNIVERSITY OF TECHNOLOGY
P.O. BOX 1000, FIN-02015 HUT
http://www.hut.fi

ABSTRACT OF DOCTORAL DISSERTATION

Author Paulo Antonio Andrade Esquef

Name of the dissertation

Model-Based Analysis of Noisy Musical Recordings with Application to Audio Restoration

Date of manuscript March 1st, 2004	Date of the dissertation April 2nd, 2004
☐ Monograph	☑ Article dissertation (summary + original articles)

Department	Electrical and Communications Engineering
Laboratory	Laboratory of Acoustics and Audio Signal Processing
Field of research	Audio Signal Processing
Opponent(s)	Dr. Patrick J. Wolfe
Supervisor (Instructor)	Prof. Vesa Välimäki

Abstract

This thesis proposes digital signal processing algorithms for noise reduction and enhancement of audio signals. Approximately half of the work concerns signal modeling techniques for suppression of localized disturbances in audio signals, such as impulsive noise and low-frequency pulses. In this regard, novel algorithms and modifications to previous propositions are introduced with the aim of achieving a better balance between computational complexity and qualitative performance, in comparison with other schemes presented in the literature. The main contributions related to this set of articles are: an efficient algorithm for suppression of low-frequency pulses in audio signals; a scheme for impulsive noise detection that uses frequency-warped linear prediction; and two methods for reconstruction of audio signals within long gaps of missing samples.

The remaining part of the work discusses applications of sound source modeling (SSM) techniques to audio restoration. It comprises application examples, such as a method for bandwidth extension of guitar tones, and discusses the challenge of model calibration based on noisy recorded sources. Regarding this matter, a frequency-selective spectral analysis technique called frequency-zooming ARMA (FZ-ARMA) modeling is proposed as an effective way to estimate the frequency and decay time of resonance modes associated with the partials of a given tone, despite the presence of corrupting noise in the observable signal.

Keywords acoustic signal processing, computer music, audio restoration, spectral analysis

UDC	Number of pages 153
ISBN (printed) 951-22-6945-7	ISBN (pdf) 951-22-6950-3
ISBN (others)	ISSN 1456-6303

Publisher Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing

Print distribution Report 73 / Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland

☑ The dissertation can be read at http://lib.hut.fi/Diss/

Preface

This thesis represents the research work I carried out in the Laboratory of Acoustics and Audio Signal Processing, at Helsinki University of Technology (HUT), Espoo, Finland, from Fall 2000 to Fall 2003. My academic work in Finland since 2000 has been supported by a scholarship from the Brazilian National Council for Scientific and Technological Development (CNPq-Brazil). Partial support for the research was received from Academy of Finland from 2001 to 2003 and from European project ALMA (IST-2001-33059) in 2003 and 2004.

I wish, in the first place, to thank Prof. Matti Karjalainen for granting me the privilege of working in the Laboratory of Acoustics and Audio Signal Processing at HUT, and for the subsequent support throughout my stay there. Matti's wisdom and large experience in the field of audio and acoustics not only were priceless during the discussions that permeated the research conduction, but also gave me confidence regarding the high standard of the work done.

I am deeply grateful to my supervisor Prof. Vesa Välimäki for the continuous encouragement and guidance he provided me. Vesa's enthusiasm and commitment to efficiency were crucial to the quality and quantity of the accomplishments of this thesis. Working with Vesa also help me to develop my writing skills in English a great deal, due to his always careful proof-reading of our papers and numerous tips on how to convey ideas concisely and clearly.

Mrs Lea Söderman handled brilliantly all bureaucratic arrangements of the work. She was also magnificent in taking care of the paperwork related to a wrist injury I had in 2003. I thank Lea and Vesa from the bottom of my heart for all help offered in that difficult period. I would like also to thank Mr Jussi Hynninen for taking care of computer-related problems.

I am grateful to the colleagues that co-authored some of the papers in this thesis for their fruitful and stimulating discussions as well as relevant suggestions on the text of the manuscripts. I acknowledge the pre-examiners of this thesis, Dr. Riitta Niemistö and Dr. Simon Godsill for their invaluable comments and suggestions on the content and structure of the manuscript.

I have been introduced to audio restoration by Dr. Luiz Wagner P. Biscainho, who, together with Dr. Paulo S. R. Diniz supervised my Master's thesis at Federal University of Rio de Janeiro. Prof. Diniz gave me substantial encouragement for the continuation of my postgraduate studies in Finland, even if that meant loosing me as a member of the research group he leads. His generosity and friendship are very much appreciated. Dr. Biscainho, besides being an estimated personal friend, accompanied closely the research I did at HUT. He was always present in the difficult and joyful moments. His invaluable and unrestrictive support, both work- and life-wise, was fundamental to keep me going ahead with the thesis work.

The Aku-lab is full of creative and witty people, and working in such environment was certainly stimulating. I had amusing moments whenever we got together in social and sport activities. I thank in particular Dr. Cumhur Erkut, Dr. Hanna Järveläinen, Dr. Lutz Trautmann, Dr. Tom Bäckström, Henkka, Klaus, Mairas, Mara, Miikka, Poju, and Tuomas for sharing work, life, and music experiences with me.

I have always had a good time with my friends Dr. Stefan Werner, Dr. Pasi Lassila, Dr. Are Hj$\varnothing$rungnes, João Pedro, Ricardo (Boca), and Cássio during many lunch and coffee breaks, among other occasions. These guys deserve special thanks for enduring my endless complaints. They were always all ears and ready to give me wise advises, which counted a lot for my rescue from the darkness.

In Finland I lived in a lovely flat attached to the house of Mr Pekka and Mrs Anneli Välimäki. They have received me warmly and, together with Vesa, were kind enough to put a piano at my disposal. I am grateful to them for letting me take part in some of their family celebrations, in which I had cheerful times as well as got to know better the Finnish culture. Moreover, the visits to their countryside cottage were always relaxing and funny events. *Kiitos paljon!*

Finally, and most importantly, I would like to thank my parents Paulo and Rosa, as well as my sister Gisele and my brother Israel, for their love and unconditional support to whatever (and sometimes suspicious) life decisions I have made, such as coming to pursue a PhD in Finland.

Paulo A. A. Esquef Espoo, 23th of February 2004.

List of Publications

This thesis is composed of the following publications, referred to as [P1]-[P7]:

[P1] P. A. A. Esquef, L. W. P. Biscainho, and V. Välimäki, "An Efficient Algorithm for the Restoration of Audio Signals Corrupted with Low-Frequency Pulses," *J. Audio Eng. Soc.*, vol. 51, no. 6, pp. 502–517, June 2003.

[P2] P. A. A. Esquef, M. Karjalainen, and V. Välimäki, "Detection of Clicks in Audio Signals Using Warped Linear Prediction," in *Proc. 14th IEEE Int. Conf. on Digital Signal Processing (DSP2002)*, Santorini, Greece, July 2002, vol. 2, pp. 1085–1088.

[P3] P. A. A. Esquef, V. Välimäki, K. Roth, and I. Kauppinen, "Interpolation of Long Gaps in Audio Signals Using the Warped Burg's Method," in *Proc. 6th Int. Conf. Digital Audio Effects (DAFx-03)*, London, UK, Sept. 2003, pp. 18–23.

[P4] P. A. A. Esquef, "Interpolation of Long Gaps in Audio Signals Using Line Spectrum Pair Polynomials," Tech. Rep. 72, Lab. of Acoustics and Audio Signal Processing, Helsinki University of Technology, Feb. 2004, (Submitted to IEEE Trans. Speech and Audio Processing, May 2003.).

[P5] P. A. A. Esquef, V. Välimäki, and M. Karjalainen, "Restoration and Enhancement of Solo Guitar Recordings Based on Sound Source Modeling," *J. Audio Eng. Soc.*, vol. 50, no. 4, pp. 227–236, 2002.

[P6] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "Frequency-Zooming ARMA Modeling of Resonant and Reverberant Systems," *J. Audio Eng. Soc.*, vol. 50, no. 12, pp. 1012–1029, Dec. 2002.

[P7] P. A. A. Esquef, M. Karjalainen, and V. Välimäki, "Frequency-Zooming ARMA Modeling for Analysis of Noisy String Instrument Tones," *EURASIP Journal on Applied Signal Processing – Special Issue on Digital Audio for Multimedia Communications*, vol. 2003, no. 10, pp. 953–967, Sept. 2003.

Contents

List of Abbreviations and Acronyms

AR	Auto-Regressive
ARMA	Auto-Regressive Moving-Average
CASA	Computational Auditory Scene Analysis
CD	Compact Disc
DSP	Digital Signal Processing
DWG	Digital Waveguide
DWT	Discrete Wavelet Transform
EMSR	Ephraim-Malah Suppression Rule
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques
FIR	Finite Impulse Response
FZ-ARMA	Frequency-Zooming ARMA
IIR	Infinite Impulse Response
LP	Long Playing
LPC	Linear Predictive Coding
LS	Least Squares
LSF	Line Spectrum Frequency
LSP	Line Spectrum Pair
LTI	Linear Time-Invariant
MA	Moving-Average (model)
PAQM	Perceptual Audio Quality Measure
RPM	Revolutions Per Minute
SD	Spectral Distortion
SNR	Signal to Noise Ratio
SSM	Sound Source Modeling
STFT	Short-Time Fourier Transform
STSA	Short-Time Spectral Attenuation
TPSW	Two-Pass Split-Window
WLSP	Weighted Line Spectrum Pair
w.r.t.	with respect to

List of Symbols

a_i	AR model coefficients
$e(n)$	Modeling error, prediction error, or residual
$E[\cdot]$	Expected value operator
p	Model order of an AR model (or the AR part of an ARMA model)
q	Model order of the MA part of an ARMA model
$P(z)$	Symmetric LSP polynomial
$Q(z)$	Antisymmetric LSP polynomial
λ	Frequency warping factor
f_s	Sampling rate in Hz
η	Weight factor in the WLSP model

Chapter 1

Introduction

1.1 A Brief History of the Recording Technology

The first reproducible recording of human voice was made in 1877 on a tin-foil cylinder phonograph devised by Thomas A. Edison. Since then, much effort has been expended to find better ways to record and reproduce sounds. By the mid-twenties, the first electric recordings appeared and gradually took over purely acoustic recordings. The development of electronic computers—from the transistorized generation in the mid-fifties to the micro-processed generation in the early seventies, in conjunction with the ability to record data into magnetic or optical media—culminated in the standardization of compact disc format in 1980. Nowadays, digital technology has been applied to several audio applications. For example, it can be used to improve the quality of modern and old recording/reproduction techniques but also, in a somewhat opposite direction, to trade sound quality for less storage space and less taxing transmission capacity requirements. For a comprehensive time-line and description of the most prominent events regarding the recording technology history see [1] and [2].

1.2 Traditional Audio Restoration

Audio restoration basically aims at improving the sound of old recordings. The primary goal is to reduce spurious noise artefacts, which are usually introduced by the recording/playback mechanisms, while preserving, as much as possible, the original recorded sound.

The first step in a typical audio restoration procedure consists of transferring the sound from old matrices to a digital form. This task involves practical matters such as locating the original matrices or the best sounding copies; finding the best way or equipment to play back a given matrix; and dealing with the usual lack of standardization associated with obsolescent recording/playback systems.

Sound transferring from old media to more modern ones, for instance from 78 RPM to LP disks, was a common practice even before the digital audio era. The idea was to benefit from the improved features of more advanced recording systems. Among other matters, these features relate to a larger data storage capacity, longer life-time, easier handling and care, easier replication of the material, and greater accessibility to a wider public. Even in the analog era, attempts have been made to improve the sonic quality of old recordings within the sound transferring process. The same issues were in question, perhaps in a more exacerbated way, when digital media came about.

As highlighted earlier, audio restoration goes beyond merely digitizing analog audio. Thanks to the progressive increases in the computational power of digital processors, more sophisticated and powerful processing of digitized data has became feasible in practice. Nowadays, audio restoration is accomplished by devising DSP algorithms devoted to reducing or suppressing from the recorded material spurious noises and disturbances introduced by the old recording/playback systems.

The most commonly found types of degradation associated with old recordings can be roughly classified into localized and global disturbances [3]. For example, short impulsive noises (clicks, crackles, and pops) as well as low-frequency long pulses (thumps) belong to the former class while continuous background disturbances or interferences such as broadband noise (hiss), buzz, and hum are usually classified as global disturbances. Other examples of audio degradations include non-linear distortions, e.g., clipping, surface noise, and slow or fast frequency modulations, wow or flutter respectively.

A review of the main techniques available for treating localized disturbances, such as long pulses and clicks is given in Chapter 3. Methods and algorithms for dealing with other types of degradations can be found elsewhere [3].

As audio-related DSP techniques evolve, it may be possible to overcome the intrinsic limitations of old recording systems. As a result, the sonic quality of the restored audio material could be improved beyond that of the original matrices. It is already possible, for example, to artificially extend the bandwidth of a given audio signal based on the observable band-limited one [4], [5], [6], [7].

1.3 Towards Sound Objects and Content

Currently available audio restoration tools allow automating some tasks, e.g., click removal, that would be tiresome if carried out manually. However, attaining satisfactory restoration results is still quite dependent on wise choices of processing parameters. Moreover, the actual restoration methods lack *a priori* knowledge on the contents of the signal to be restored. For example, noise-like events, such as a whip sound or a drum brushing, may be mistakenly treated as click or hiss, respectively. Therefore, the need of a reasonable personal judgment on the final sonic quality of a restored sound becomes crucial.

It seems reasonable to believe that, in the future, audio restoration techniques can benefit from incorporating extra information into the sound being processed. For example, knowledge about the contents of the signal and the mechanisms of sound production and perception could be taken into account in the restoration procedures. To the author's knowledge, attempts in this direction have been restricted to the use of psychoacoustic criteria embedded in de-hissing methods only [8], [9], [10], [11], [12], [13].

Recently, much research has focused on finding ways to represent audio signals in a structured organization through high-level constituent sound objects [14], [15], [16], [17]. In the future, one can foresee the possibility of carrying out audio restoration in an object-based framework. A rather idealized goal in this case would be first to analyze the audio content from the sound of old recordings to obtain control data for appropriate sound synthesis models and then, to generate completely new and noiseless audio signals, based on these models. Of course, realizing audio restoration via resynthesis is a rather formidable challenge from the engineering point of view. It entails interdisciplinary research areas such as Computational Auditory Scene Analysis (CASA) [18], [19], automatic transcription of music [20], [21], [22], object-based audio coding [23], [24], [25], and sound synthesis [26], [27], [28], [29], [30], [31], [32]. Apart from technical matters, from the subjective point of view, performing audio restoration through an object-based sound resynthesis may raise questions regarding the preservation of artistic integrity in the reconstructed signals. However, one should bear in mind that audio restoration is of a non-substitutive character. Thus, whatever results an object-based approach may offer in the future, it will constitute just another way of listening to old recordings.

1.4 Scope of this Thesis

This thesis deals with audio signal processing and its applications to restoration of old recordings. All sound manipulations presented here take place in the digital domain, after transferring and digitizing the audio from analog sources. Two distinct modeling approaches are used in this thesis: conventional signal modeling and sound source modeling techniques. Within the former category, the propositions are restricted to methods for suppression of localized disturbances in audio signals, such as long pulses and clicks. As regards the latter approach, the work is limited to analysis and synthesis of tones from plucked string instruments. Within this scenario, propositions for bandwidth extension of guitar tones in connection with signal de-hissing are presented.

1.5 Contents of this thesis

The work in the first four publications deals with signal modeling techniques for audio restoration and concentrates on the detection and suppression of localized disturbances in audio signals, such as impulsive noise and low-frequency pulses. In this context the proposed algorithms include: an efficient algorithm for suppression of low-frequency pulses in audio signals; a model-based scheme for impulsive noise detection that uses frequency-warped linear prediction; and two methods for reconstruction of audio signals in long fragments of degraded samples.

The remaining publications elaborate on the application of SSM techniques to audio restoration and on the problems involved in such a framework. A case study featuring bandwidth extension of guitar tones is presented. Moreover, the problems associated with the calibration of the sound source models from noisy measurements are outlined. In dealing with this matter, frequency-selective model-based spectral analysis tools are proposed as a robust means to extract the desired model parameters from noisy sources.

1.6 Structure of this Thesis

In addition to the introduction, Chapter 2 reviews basic concepts that permeate the contents of this thesis. Chapter 3 provides a brief overview of standard techniques used for treating the most commonly found types of degradations associated with old recordings. Moreover, it relates prior works in the field with the contributions offered in this thesis. Such contributions are summarized in Chapter 4. Finally, conclusions and future directions are given in Chapter 5. The publications related to this thesis are included as annexes.

Chapter 2

Basic Concepts and Tools

This chapter briefly reviews some concepts and tools that are used throughout the publications included in this manuscript. It covers issues related to AR modeling, frequency-warping, and digital waveguide synthesis.

2.1 AR and ARMA Modeling

The resonant nature associated with most of the sound vibrations makes AR and ARMA processes suitable tools to model short fragments of audio signals. In fact, AR and ARMA models find use in several audio applications, such as speech and audio coding, sound synthesis, and spectral analysis.

2.1.1 Basic Definitions

An $\mathrm{ARMA}(p, q)$ process $x(n)$ can be generated by filtering white noise $e(n)$ through a causal linear shift-invariant and stable filter with transfer function [33]

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}}. \tag{2.1}$$

Considering a flat power spectrum for the input, i.e., $P_e(z) = \sigma_e^2$, the resulting output $x(n)$ has a generalized power spectrum function given by

$$P_x(z) = \sigma_e^2 \frac{B(z)B^*(1/z^*)}{A(z)A^*(1/z^*)}, \tag{2.2}$$

where the symbol $*$ stands for complex-conjugate. For real-valued filter coefficients, $P_x(z)$ has $2p$ poles and $2q$ zeros. The power spectrum associated with an $\mathrm{ARMA}(p,q)$ process, which is attained by evaluating $P_x(z)$ for $z = e^{\mathrm{j}\omega}$, is then given by

$$P_x(e^{\mathrm{j}\omega}) = \sigma_e^2 \frac{|B(e^{\mathrm{j}\omega})|^2}{|A(e^{\mathrm{j}\omega})|^2}. \tag{2.3}$$

In the time domain, an $\mathrm{ARMA}(p,q)$ process $x(n)$ relates to excitation $e(n)$ through the following difference equation

$$x(n) + \sum_{k=1}^{p} a_k x(n-k) = \sum_{k=0}^{q} b_k e(n-k). \tag{2.4}$$

An AR process is a particular case of an ARMA process when $q = 0$. Thus, the generator filter assumes the form

$$H(z) = \frac{b_0}{1 + \sum_{k=1}^{p} a_k z^{-k}}, \tag{2.5}$$

which is usually referred to as the transfer function of an all-pole filter.

2.1.2 Parameter Estimation of AR and ARMA Processes

Estimation of AR and ARMA models is a well researched topic with a vast literature available [34], [33]. A compact review regarding the estimation of AR and ARMA models can be found in sections 1 and 2 of [P6]. AR model estimation consists of a linear optimization problem in which a certain cost function is minimized w.r.t. the AR model parameters. Different definitions of the cost function lead to different solutions. For example, the popular autocorrelation and covariance methods use the sum of the squared magnitude of the modeling error as cost functions. The difference lies in the time range over which the error is considered [33]. Other approaches, such as Burg's method, explore the modularity of lattice structures to provide a solution for the model parameters in an iterative form [35], [36]. If the stability of the models is a crucial issue to the application at hand, then the autocorrelation and Burg's methods are suitable choices, since they guarantee stable AR estimates.

Parameter estimation of ARMA processes is a more involved task. The difficulty arises from solving the normal equations which are no longer linear in the ARMA coefficients. Thus, the usual solutions rely on non-linear and iterative optimization procedures, e.g., Prony's method [33] and the Steiglitz-McBride iteration [37]. One drawback of these methods is that the estimated models cannot be guaranteed to be minimum-phase, thus raising obstacles to inverse filtering problems. In addition, and especially for high-order models, the model estimates may become unstable.

2.2 Frequency Warping

The Fourier transform is one of the most common ways to transform a signal from its time-domain representation into a frequency-domain representation. In this case, the frequency-resolution of the resulting spectrum is uniform along the frequency axis. Frequency warping techniques are primarily concerned with the design of transformations in which non-uniform frequency-resolutions can be attained. Typically, a frequency mapping operator is involved in this task.

In this thesis, frequency warping is restricted to a conformal bilinear mapping. This choice is convenient, since it implies substituting the unit delays z^{-1} with first-order allpass filters $D(z)$ in the filter structures and definitions used [38]. The allpass filter works as a frequency-dependent delay element and its transfer function is defined by

$$\tilde{z}^{-1} \triangleq D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}. \tag{2.6}$$

While z^{-1} has a linear phase response, the phase response of $D(z)$ can be made non-linear by adjusting the warping factor λ. Indeed, the mapping from the uniform to the warped frequency scale is governed by the phase response of $D(z)$, which is given by [38]

$$\tilde{\omega} = -\omega + 2\arctan\left\{\frac{-\lambda\sin(\omega)}{1 - \lambda\cos(\omega)}\right\}, \tag{2.7}$$

where $\omega = 2\pi f / f_\mathrm{s}$ and f_s is the sampling frequency. Figure 2.1 shows the attained mapping for several values of λ. For positive values of λ, the resolution at low frequencies is increased whereas negative values of λ yield a higher resolution at high frequencies. Suitable values of λ can be chosen depending on the application at hand. For instance, in [39] Smith and Abel provide a closed-form formula that, for a given sampling rate, provides the optimum value of λ that best approximates the frequency resolution of the Bark scale.

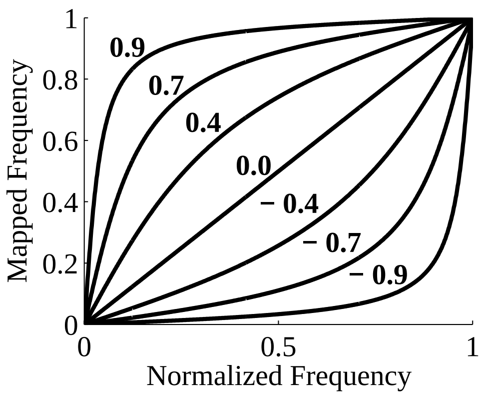

Figure 2.1: Frequency mapping for several values of λ.

As regards filtering on a frequency-warped scale, it is possible to realize frequency-warped versions of FIR and IIR filters in either the direct-form or lattice structure implementations [40], [41]. Moreover, frequency-warped linear prediction and AR modeling estimation can be formulated similarly to standard methods, such as the autocorrelation and Burg's methods [38], [41]. For a comprehensive view on frequency-warping techniques and its applications to audio signal processing see [42] and [43].

2.3 Digital Waveguide Synthesis

Digital waveguides are an efficient means to implement physics-based sound synthesis of musical instruments. For a lossy vibrating string, the direct implementation of the discretized solution of its wave equation leads to a structure composed of distributed delays and loss elements [29]. The key idea behind the computational savings associated with DWG lies in lumping these delay and loss elements into single components. For a comprehensive review on the topic see [29]. The resulting DWG structure can be seen as an extended version of the Karplus-Strong algorithm [44], [45], [46].

Digital waveguide filters account mainly for modeling the resonator part of a musical instrument. Modeling of the remaining functional parts, such as the excitation mechanism and the radiator system (instrument body), and their interactions should also be considered in the synthesizer design [45], [46], [47]. It should be pointed out that implementing accurate parametric models for the radiator element may be computationally expensive [47]. A highly efficient means to overcome this drawback is to resort to the commuted DWG synthesis [48], [49]. Its basic principle consists of commuting the resonator and radiator parts, which are assumed to be LTI systems, and gathering the excitation and the radiator parts into a wavetable.

2.3.1 String Model

A simple and efficient model of an isolated vibrating string is illustrated in Fig. 2.2. The model's transfer function is given by

$$S(z) = \frac{Y(z)}{E_{\mathrm{CWS}}(z)} = \frac{1}{1 - z^{-L_{\mathrm{i}}} F(z) H_{\mathrm{loss}}(z)}, \tag{2.8}$$

where $z^{-L_{\mathrm{i}}}$ and $F(z)$ are, respectively, the integer and fractional parts of the delay-line associated with the length of the string L, and $H_{\mathrm{loss}}(z)$ is the transfer function of a loss filter, which governs the frequency dependent losses of the harmonic modes.

The length of the string is given by $L = f_{\mathrm{s}}/\hat{f}_0$, where $\hat{f}_0$ is an estimate of the fundamental frequency of the tone and f_{s} is the adopted sampling rate. In general, L is a real number and can be decomposed into $L = L_{\mathrm{i}} + \delta$, with $L_{\mathrm{i}} = \lfloor L \rfloor$, where operator $\lfloor \cdot \rfloor$ stands for *the greatest integer less than or equal to*. The fractional part, which is specified by δ, can be implemented through fractional delay filters [50]. For instance, the Lagrange interpolator [50] is a straightforward choice for the fractional delay filter. In such a case, a closed-form formula for computing the filter coefficients exists for given specifications of filter order and delay.

The loss filter $H_{\mathrm{loss}}(z)$ is usually of lowpass characteristic. In addition, its magnitude response must be less than unity to guarantee the stability of $S(z)$. A suitable choice for the loss filter is

$$H_{\mathrm{loss}}(z) = g\frac{1+a}{1+az^{-1}}, \tag{2.9}$$

where $0 < g < 1$ and $-1 < a < 0$. The parameters g and a of this one-pole filter are associated, respectively, with perceptual features such as the overall decay time and the brightness of the tone. Optimum design in a weighted LS sense of one-pole filters, for a given set of specifications for the decay time of the partial modes, is proposed in [51], [52], [53]. Design of higher-order IIR loss filters is presented in [54].

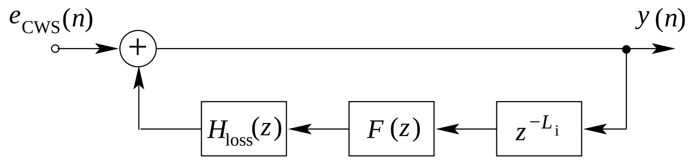

Figure 2.2: Block diagram of a simple one-polarization string model.

2.3.2 Model Calibration

As for the decay time specifications, they can be derived analytically from the physical model [55]. Alternatively, they can be obtained via STFT-based analysis of recorded tones [51], [52], [56]. Other means include non-linear optimization methods [57], [58] and parametric techniques [59], [60], [61].

STFT-based methods assume exponentially decaying partial envelopes. However, due to differences between the transversal (horizontal and vertical) polarizations of vibration, partial envelopes may exhibit amplitude beating and two-stage decay [62]. See [63] and [53] for examples of methods for estimating the frequency difference between partial modes from the extracted partial envelopes. Notwithstanding, accurate decay time estimation of partial modes via envelope fitting can be prevented by the presence of multi-modal partials and corrupting background noise in the recorded tone. In publication [P5], an SSM-based method for extending the bandwidth of a guitar tone whose high frequencies have been lost due to aggressive de-hissing is offered. In this case, as the high-frequency partials were either immersed in noise or absent, the calibration of the waveguide model was carried out with the aid of a similar fullband and clean guitar tone. However, in publications [P6] and [P7] the authors address the estimation of the decay time of partial modes under noisy conditions. Moreover, they demonstrate the effectiveness of the FZ-ARMA modeling [64] in accomplishing this task.

Naturally, the string model depicted in Fig. 2.2 is rather simplified. In order to attain more realistic sonic results, string models should account for other features related to the vibrating string phenomenon. Among them, one can mention: dispersion in stiff strings [45], [65], [66], [67], [68]; vertical and horizontal polarizations of vibration [46], [53], [69], [70]; sympathetic string coupling [45], [46]; and non-linear effects [71], [72]. Usually, the more sophisticated the synthesis model, the less computationally efficient it is. In this regard, psychoacoustics can help to achieve an optimum balance between computational cost and sound quality. For example, one may enable or disable parts of a synthesizer model according to the salience of the perceptual attributes of the tone to be synthesized [73].

Chapter 3

Overview of Digital Audio Restoration

Digital audio restoration takes place after the sound transfer process from analog sources to digital domain. Thumps, pops, clicks, crackles, and hiss are common onomatopoeias used to characterize the sound produced by spurious noises that usually corrupt old recordings. The following sections review standard procedures for reducing these noises and follow the order of precedence in which the problems are usually tackled.

3.1 De-thumping

Thumps are produced by long pulses of low-frequency content that additively corrupt an audio signal. These pulses are related to the mechanical response of the playback mechanisms to an abnormal excitation, e.g., the response of the stylus-arm apparatus to large discontinuities in the groove walls of a disk. Examples of long pulses are depicted in Fig. 3.1.

Apart from crude techniques, such as highpass filtering of the signal, a few methods, with various degrees of sophistication, exist for treating long pulses. The template matching method introduced by Vaseghi [74], [75] figures among one of the first propositions. The basic assumption behind this method is that long pulses are identical in shape varying only in amplitude. Thus, given a template of the corrupting pulse, one can locate other pulse occurrences through high values of the cross-correlation coefficient, measured between the template and the corrupted signal. Then, noise suppression is carried out by subtracting an

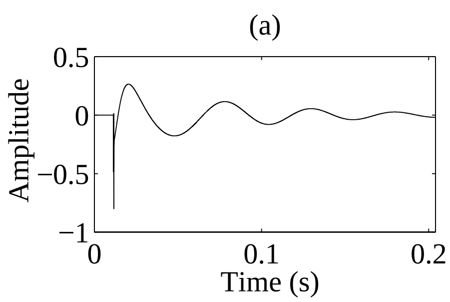

Figure 3.1: Examples of long pulses: (a) a single pulse and (b) superimposed pulses.

amplitude-scaled template from degraded portions of the signal, after an adequate time synchronization.

The high-amplitude click that usually appears at the beginning of a long pulse is to be removed by a de-clicking algorithm (to be discussed in Section 3.2). Figure 3.2 shows a block diagram that illustrates the functional stages of the template-matching de-thumping method. The main limitation of the template matching method is the lack of robustness in detecting and suppressing pulses whose shape varies over time or pulses that are superimposed [3].

A model-based approach to long pulse removal has been proposed in [76]. In this method the corrupted audio signal is modeled as a mixture of two distinct AR processes. Thus, signal restoration is accomplished by separating these two processes. High-order AR models are used to fit the underlying audio signal. As for the pulse, low-order AR models are employed and the initial click is incorporated as part of the pulse. This is done by setting a high-variance model excitation during the click portion. The main drawback of the AR-separation approach is its high computational cost. The method has been further developed into a more efficient implementation that utilizes Kalman filtering techniques [77].

In [78], Esquef *et al.* have proposed an algorithm for long pulse removal based upon non-linear and polynomial smoothing filters. Extensions of this preliminary work were reported later in publication [P1]. In the proposed pulse removal method, a first estimate of the pulse's waveform is obtained through a modified version of a filtering technique called two-pass split-window (TPSW) [79], [80]. The key purpose in this filtering scheme is to mitigate the spurious effects of the initial high-amplitude click on the pulse estimate. Moreover, the authors point out the need and also offer two propositions to vary the length of the TPSW filter over

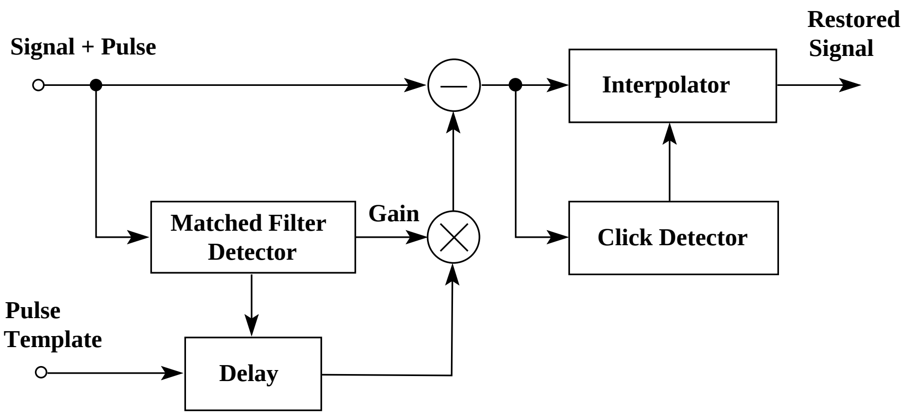

Figure 3.2: Processing stages of the template-matching method for de-thumping. Adapted from [75].

time, in order to cope with typical waveform evolutions of long pulses.

The final pulse estimate, which is to be suppressed from the corrupted signal, results from a piecewise polynomial fitting scheme applied to the TPSW-based pulse estimate. The polynomial fitting technique, which is meant to further smooth the pulse estimate, resembles Savitzky-Golay smoothing filters [81], [82]. For appropriate values of processing parameters, the restoration performance of both the TPSW- and the AR-based pulse removal methods are comparable. However, the computational complexity of the TPSW-based algorithm is two orders of magnitude lower than that of the AR-separation method.

3.2 De-clicking

Impulsive-like disturbances contaminating audio signals are heard as clicks. Such disturbances are usually caused by small-scale imperfections on the groove walls of a disk. Among other causes, these imperfections can be due to the intrinsic porosity of the disk material, solid particles adhering to the groove walls, or the presence of superficial scratches on the disk surface [83].

Systems for click removal were attempted even in the analog domain [84], [85], [86], [87]. Intriguingly, some of the rationales used in these ancient systems, such as a two-stage task procedure, viz. click detection and signal reconstruction, have been similarly employed later in digital approaches to audio de-clicking.

In model-based approaches for audio de-clicking, short frames of the under-

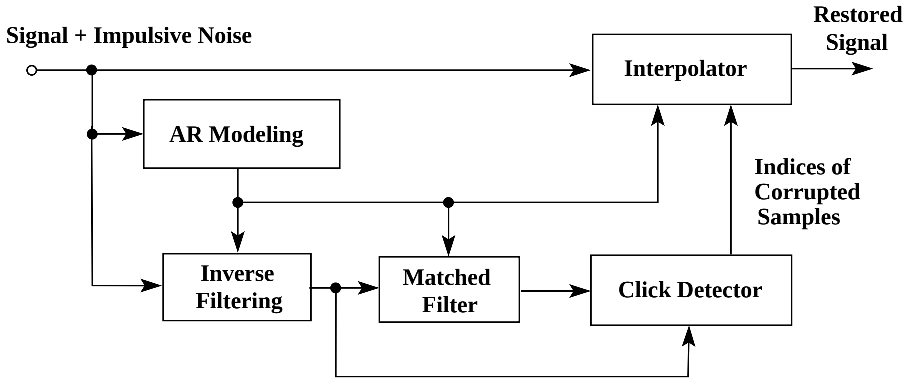

Figure 3.3: Model-based de-clicking method. Adapted from [75].

lying audio signal $x(n)$ are modeled as p^{th}-order AR processes. Moreover, the degradation mechanism assumes the additive form

$$y(n) = x(n) + i(n)d(n), \tag{3.1}$$

where the term $i(n)d(n)$ corresponds to the noise component, with $i(n)$ being a binary sequence that indicates the presence of noise ($i(n) = 1$).

The goal of click detection algorithms is to estimate the sequence $i(n)$ that minimizes the percentage of both miss detection and false alarm of clicks. In the AR-based click detection scheme proposed in [74] and [88], the magnitude of the AR model excitation, which is obtained via inverse filtering, is compared against an adaptive threshold. Samples of the excitation with magnitude values exceeding that of the threshold reflect corrupted samples in the signal. Further improvements in click detection performance can be achieved by using matched filtering [74], two-sided linear prediction [89], and more elaborated thresholding schemes [90].

Noise suppression is attained by replacing the corrupted samples with others that resemble the underlying signal. This task can be satisfactorily accomplished through AR-based interpolation methods [91], [74], [92]. Figure 3.3 depicts a block diagram with the functional stages of a model-based de-clicking scheme.

Audio de-clicking can be also carried out within a statistical Bayesian framework [93], [76], [94], [95], [3]. In the Bayesian approach, models are used for both the underlying signal and the noise processes. Moreover, click detection and suppression are accomplished within a single processing stage. Other means for audio de-clicking include neural-network algorithms [96] and model-based adaptive filtering techniques [97], [98].

In publication [P2], the authors have investigated the use of frequency-warped linear prediction within a model-based click detection scheme. The results obtained show that focusing the modeling efforts on high frequencies, i.e., by employing negative values for the warping factor, favors the click detection performance.

Various kinds of approaches have been used to tackle the reconstruction of discrete-time signals across gaps of missing samples. Among the available techniques one can mention: bandlimited reconstruction [99], [100]; sinusoidal modeling interpolation [101]; waveform substitution [102], [103]; AR-based interpolation [91], [74], [92]; and multirate signal reconstruction [104], [105], [106].

In AR-based schemes, the signal reconstruction is attained through the minimization of the variance of the model excitation w.r.t. the unknown samples. Due to constraints related to the non-stationarity of audio signals, AR-based interpolation methods are usually suitable for reconstructing relatively short-duration (up to about 20 ms) portions of audio signals. Interpolation across longer gaps may lead to poor signal reconstruction, which is characterized by excessive evenness and decaying power of the signal toward the middle of the gap. The simplest way to circumvent these shortcomings is to increase the order of the AR model. Other solutions include: interpolators that use AR models augmented with cosine basis functions [3]; interpolators with constant excitation energy [107], [108]; interpolators based on random sampling [109], [110]; and interpolators that employ two different AR models, one to model the segment that immediately precedes a gap and another for the fragment that succeeds the gap [111], [112].

In publication [P3] the authors compare the performance (computational cost versus perceptual quality) of the interpolation scheme proposed in [112] against a modified version of it that employs frequency-warped AR models. Carrying out AR model estimation and signal reconstruction in a frequency-warped scale is computationally more expensive than when done in the conventional way. However, in the warped case, the warping factor can be tuned to focus the modeling efforts on the more perceptually prominent low-frequency resonance modes of the signal. As a consequence, better models in the perceptual sense can be attained. Simulation results show that, at a given cost level and for low-order models, the warped-based scheme performs better than the conventional scheme. The perceptual quality of the reconstructed signals was measured with the perceptual audio quality measure (PAQM) [113]. Briefly, the PAQM is an objective distortion measure that takes into account psychoacoustic phenomena. The PAQM represents a dissimilarity index that is obtained by comparing the inner-ear representations of a reference and a processed signal. Additional discussion on the usage of the

PAQM in the context of audio restoration can be found in section 6.4 of [P1].

Publication [P4] also addresses the interpolation of audio signals over long gaps of missing samples. This time, however, instead of using frequency-warped AR models, the author investigates the use of modified AR models within the interpolation scheme of [112]. The modification consists of first estimating an AR model, for instance, using Burg's method, and then straightforwardly constructing a modified AR model via a weighted sum of its corresponding line spectrum pair polynomials. The weight parameter controls the pole location of the modified model. Moreover, it can be tuned to place the poles close to or on the unit circle. This resource is shown to improve the performance of the interpolation scheme. The quality of the processed signals was assessed using the PAQM [113].

3.3 De-hissing

Signal contamination with additive broadband noise is perceived as hiss. The noise component is usually due to thermic measurement noise, circuitry noise, and intrinsic characteristics of the storage medium. For example, 78 RPM disks and analog tape recordings are usually associated with hiss noise. For the latter medium, however, the noise characteristics are more stationary than the former [114].

Audio and speech de-hissing has been approached through several means. Perhaps the most common is via short-time spectral attenuation (STSA) methods, which have their origin in speech de-noising [115], [116], [117]. The basic assumption behind STSA methods is that the corrupting noise is a zero-mean white Gaussian process uncorrelated to the underlying signal. In practice, due to the non-stationarity of general audio signals, STSA methods employ a block-based overlap-and-add processing scheme. Moreover, the transformations from time into frequency and vice versa are carried out through the discrete Fourier transform.

In STSA methods, noise reduction is accomplished by subtracting an estimate of the power spectrum of the noise realization from the observed power spectrum of the corrupted signal. This strategy is known as the power spectrum subtraction suppression rule. Other related suppression rules include the magnitude spectral subtraction [116] and Wiener filtering [74]. All these suppression rules can be formulated in terms of a time-varying filter that attenuates the magnitude spectrum of the noisy signal differently, according to the current SNR measured at a given frequency bin. For comparisons among different suppression rules see [118], [114].

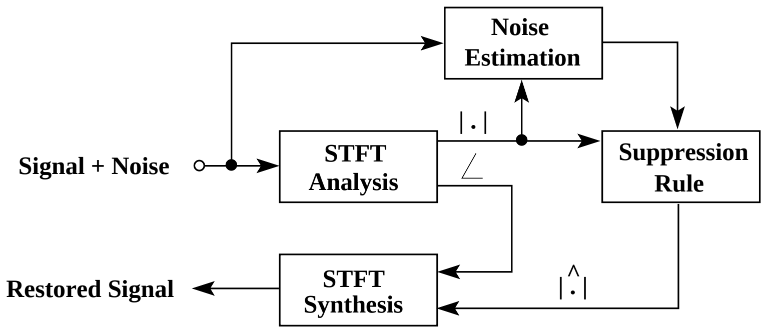

Figure 3.4: STSA-based hiss reduction system. Adapted from [119].

Figure 3.4 illustrates the typical processing stages of STSA de-hissing methods.

A typical side-effect of STSA methods is the presence of random tonal noise in the processed signals. Also known as musical noise, this phenomenon originates from the random characteristic of the power spectrum of single noise realizations [118]. Thus, at random frequency values, the power of some noise components may lie above the estimated noise power spectrum. These components are not suppressed and, as a result, one hears brief tonal bursts with random frequencies popping up in the signal along time.

There exist several means to reduce the musical noise phenomenon. Perhaps the simplest consists in overestimating the noise power spectrum. In this case, musical noise reduction is a trade-off for the loss of valuable signal content, especially on the high-frequency range. Other ways to reduce musical noise involve: spectral averaging over successive frames [116]; applying heuristic rules to the spectral attenuation factor [120], [75]; leaving a minimum noise floor to mask the musical noise components [121], [122], [3]. Recently, some interest in incorporating psychoacoustic-related phenomena into the suppression rules used in STSA methods has been aroused [8], [9], [10], [11], [12], [123], [13].

The suppression rule proposed by Ephraim and Malah [124], [125], [126], [127] is known to be effective in reducing the musical noise residual. The EMSR corresponds to the optimum (in the minimum mean square error sense) short-time amplitude estimator of a sinusoid buried in additive noise. Moreover, the EMSR can be formulated in terms of the *a posteriori* and the *a priori* SNRs associated with a given frequency bin. In practice, the *a priori* SNR is estimated from previously processed frames via a decision-directed approach. Simpler alternatives to the EMSR have been also proposed in [128] and [119].

Still within the class of non-parametric de-hissing techniques, the wavelet-based shrinkage has received a lot of attention in the pertinent literature [129], [130], [131], [132], [133], [134]. In this method the noisy signal is mapped into a multi-resolution time-frequency representation via a discrete wavelet transform. In such representation, the underlying signal is supposed to be coded by a few high-amplitude coefficients. Conversely, the noise process is represented by a large number of low-level coefficients. Thus, de-noising is accomplished simply by discarding these coefficients before transforming the signal back to the original domain.

In addition to non-parametric techniques, more sophisticated model-based approaches have been proposed for audio de-hissing. For example, AR modeling of degraded speech and audio signals has been tackled in [135]. Statistical-based techniques that employ AR and ARMA modeling have also been developed for joint treatment of broadband and impulsive noise [76], [95], [136], [97]. Adaptive filtering schemes for audio de-hissing are presented in [74], [75], and [97]. Hybrid methods that combine either different non-parametric schemes [137] or mix non-parametric and parametric strategies [138] can also be found in the relevant literature.

Chapter 4

Contributions of this Thesis

4.1 [P1]

Summary

In this paper an efficient algorithm for long pulse removal is presented. The method uses a modified version of a non-linear filtering scheme called two-pass split-window, which has been borrowed from sonar signal processing. The key idea is to perform a simple moving-average filtering over a given segment corrupted by a long pulse. However, the filtering element used is a square window with a gap in the middle. If a square window were used, the pulse estimate would be biased due to the presence of the high-amplitude click that drives the pulse. By employing a split-window, the bias is eliminated at the click location, but remains around the click occurrence. The second pass aims at correcting this lateral bias. After the TPSW filtering stage the pulse estimate is further smoothed out by means of a piecewise polynomial fitting scheme.

The pulse estimation method is two orders of magnitude less expensive computationally than the method based on separation of AR processes and performs comparably to the latter. Moreover, the calibration of the processing parameters can be easily and intuitively carried out through a graphic user interface.

Author's Contribution

The author initiated the ideas presented and was responsible for all the simulations carried out in this work. He wrote the greater part of the text and prepared companion webpage. The co-authors contributed section 1.4 and the appendix as well as helpful comments and suggestions on the manuscript.

4.2 [P2]

Summary

This paper investigates the use of frequency-warped linear prediction for click detection in audio signals. The results show that using warped linear prediction, while focusing the resolution on higher frequencies, favors the click detection task. According to the experimental results, the change in the frequency resolution *per se* does not play a major role in improving the click detection performance. The benefit comes from a spectral tilt that appears in the prediction error when using warped linear prediction. When the warping factor is tuned to increase resolution at high frequencies, the spectral tilt assumes a highpass filter characteristic, facilitating the click detection. Although the spectral tilt could be easily compensated for, doing so is not desirable in this case. A side-effect of using warped linear prediction for click detection is a more accentuated click spread in the excitation. To overcome this problem a scheme for click detection that uses the backward and forward prediction errors is proposed in the paper. Objective measures taken over de-clicked signals reveal that employing warped linear prediction for click detection can offer a better balance between miss detection and false alarm of clicks.

Author’s Contribution

The author proposed the research, carried out all the experiments reported in the paper, wrote most of the text of the manuscript, and prepared companion webpage. The co-authors contributed valuable discussions on the subject as well as pertinent comments and suggestions on the manuscript, which they also proofread.

4.3 [P3]

Summary

In this paper, a model-based scheme for interpolation of audio signals over extensive portions of degraded samples is presented. The proposed algorithm carries

out the interpolation by means of a model-based signal extrapolation scheme that uses frequency-warped AR models, computed via a frequency-warped version of Burg's method. The warped-based interpolation scheme is about 1.7 times more costly than the conventional procedure. However, at the same cost levels and for low-order models, the results indicate that using frequency warping to focus modeling on the low-frequency range favors the interpolation task, in that the energy of the interpolated signal is better preserved across the gaps. The performance of the two interpolation methods was compared via PAQM and SNR measures.

Author's Contribution

Given the collaborative character of this work, the author's contributions to the proposed idea and its development were comparable to those of the co-authors. With the exception of sections 3.1 and 3.2, the author wrote most of the manuscript and conducted all the experiments reported. The co-authors also implemented the frequency-warped version of Burg's method, contributed some of the figures, as well as provided suggestions and comments on the text readability. They proofread the manuscript as well.

4.4 [P4]

Summary

This paper addresses the reconstruction of audio signals over long fragments of missing samples through AR-based interpolation schemes. In particular, the novel interpolation scheme uses modified versions of conventionally estimated AR models that are computed via a weighted sum of their corresponding line spectrum pair (LSP) polynomials. A single weighting factor can be tuned to place the poles of the model close to or on the unit circle. Performance comparisons using the PAQM show that, for a given low-order model, using the WLSP-modified AR models within the interpolation method yields reconstructed signals with higher perceptual quality than that of the signals reconstructed via the conventional interpolation scheme.

Author's Contribution

The author was responsible for the research conducted and the results reported. The author also designed the companion webpage.

4.5 [P5]

Summary

This paper discusses the use of sound source modeling (SSM) techniques for enhancement and restoration of guitar tones. Contrary to traditional signal modeling techniques, which aim at modeling signal waveforms, SSM techniques attempt to model the physical phenomena behind the sound production. Thus, SSM techniques should be placed within the framework of structured audio and object-based processing. The authors outline the challenges and limitations of SSM techniques as regards practical applications to audio enhancement. As a case study, they present an SSM-based method for bandwidth extension of lowpass filtered guitar tones. Moreover, the proposed bandwidth extension scheme is applied to strongly de-hissed guitar tones as a means to recover the lost high-frequency content.

Author's Contribution

The author composed the article and was responsible for the experiments and results reported as well as for the elaboration of the companion webpage. The co-authors participated through fruitful discussions on the ideas that permeate the paper and provided comments and suggestions to improve the text readability.

4.6 [P6]

Summary

This paper addresses the analysis and modeling of resonant systems by means of a frequency-selective model-based approach, called frequency-zooming ARMA modeling. Similar to other subband analysis methods presented in the literature [59], [139], [140], [141], in the FZ-ARMA analysis technique a certain subband of interest is selected from the signal spectrum and an ARMA model is fitted to the corresponding decimated subband signal. The subband model can be used to estimate the frequencies and decay times of the resonance modes present in the subband. Other applications comprise modeling of signals immersed in noise, modeling and equalization of room impulse responses, as well as modeling and synthesis of musical instrument sounds.

Author's Contribution

The author was responsible for the text and experimental results shown in sections 3.2, 3.3, 3.4, and 4.3. In addition to contributing to the overall polishing of the paper, the author prepared the companion webpage and the sound examples offered therein.

4.7 [P7]

Summary

This paper investigates the use of a frequency-selective spectral analysis technique, called FZ-ARMA modeling, for analysis of noisy guitar tones. Frequency zooming is first employed to isolate a given partial of the tone. Then, an ARMA model is fitted to the corresponding complex-valued subband signal. From the poles of the ARMA model one obtains the frequencies and decay times of the resonance modes that compose the partial. The authors show that FZ-ARMA analysis is particularly suitable when estimating modal parameters of partials whose amplitude envelope exhibits two-stage decay and beating. Moreover, comparisons with the ESPRIT method were provided. They revealed that, under noisy conditions, using ARMA models to approximate the subband signals offers a more robust means to estimate the decay time of the resonance modes present in a given partial. As a case study, a noisy guitar tone was first analyzed through the FZ-ARMA modeling scheme. Then, based on the set of extracted modal parameters, a DWG filter was designed.

Author's Contribution

The author wrote the article and was responsible for the experiments and results reported. The co-authors contributed extensive discussions on the matters related to the topic and provided invaluable comments and suggestions to improve the readability of the manuscript.

Chapter 5

Conclusions

5.1 Summary

This thesis addressed digital audio restoration through both signal modeling techniques and sound source modeling methods. Within the former category, novel methods for restoration of audio signals degraded by localized defects, such as long low-frequency pulses, clicks, and crackles, were proposed. More specifically, an efficient algorithm for long pulse removal was presented in [P1] and a frequency-warped AR-based method for click detection was introduced in [P2]. The task of signal reconstruction over long portions of missing samples was investigated in [P3] and [P4].

As regards realizing audio restoration within the SSM framework, a general discussion on the related advantages and drawbacks was provided in [P5]. In addition, case studies that feature SSM-based bandwidth extension and de-hissing of single guitar tones were presented. Analysis and modeling of resonant systems based on subband AR and ARMA modeling were covered in [P6], where a technique called FZ-ARMA modeling was described. The FZ-ARMA modeling scheme finds applications in various tasks related to audio signal processing, e.g., modeling and equalization of room impulse responses as well as modeling of musical instrument sounds. In [P7], the FZ-ARMA analysis technique was employed to calibrate a DWG model of a vibrating string based on an observed noisy measurement of a plucked guitar tone.

5.2 Future Directions

Most of the audio restoration tools that are currently available offer a certain degree of automatism for some de-noising tasks. Nonetheless, the importance of a discerning judgment on the attained sonic improvements cannot be neglected. The state-of-the-art restoration methods do not yet take into account aspects related to the content of the signals to be restored. Thus, it is up to an expert listener to decide whether impulse-like and hiss-like events should be preserved, as being part of the original signal, or are to be removed as undesired noises.

It is the opinion of the author that digital audio restoration techniques can benefit from the inclusion of high-level content information in their formulation. Such high-level information may comprise features such as the genre of the audio program being processed and the types of generative sources involved in the sound production. In this context, the work that has been carried out in the field of CASA, automatic transcription of music, object-based audio coding, and sound synthesis is of significant importance. Other relevant factors concern the characteristics of the environment in which the sound has been produced as well as psychoacoustic properties of the human auditory system. To some extent, the latter have already been incorporated into algorithms devoted to audio de-hissing.

Taking into account high-level information on the audio content may potentially bring several advantages to audio restoration methods. However, as pointed out in [P5], such strategy implies working in an object-based audio representation, which is less general than processing approaches based on conventional signal modeling techniques. Nevertheless, *a priori* knowledge on the audio content could be used to guide the choice of more appropriate processing parameters within a certain restoration procedure. Moreover, it could also help to select more adequate restoration algorithms, depending on the type of audio program at hand. A further and more involved step would be embedding high-level content information within a model-based framework for audio restoration.

Bibliography

[1] R. T. Beyer, *Sounds of Our Times – Two Hundred Years of Acoustics*, Springer-Verlag, New York, 1999.

[2] S. Schoenherr, “Recording Technology History,” URL: `http://history.acusd.edu/gen/recording/notes.html`, Retrieved Dec. 2003.

[3] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration – A Statistical Model Based Approach*, Springer-Verlag, London, UK, 1998.

[4] Y. M. Cheng, D. O’Shaughnessy, and P. Mermelstein, “Statistical Recovery of Wideband Speech from Narrowband Speech,” *IEEE Trans. Speech and Audio Processing*, vol. 2, no. 4, pp. 544–548, Oct. 1994.

[5] C. Cheng, “High Frequency Compensation of Low Sample-Rate Audio Files: A Wavelet-Based Excitation Algorithm,” in *Proc. Int. Computer Music Conf.*, Thessaloniki, Greece, Sept. 1997.

[6] J. Epps and W. H. Holmes, “A New Technique for Wideband Enhancement of Coded Narrowband Speech,” in *Proc. IEEE Workshop Speech Coding*, Porvoo, Finland, 1999, pp. 174–176.

[7] E. Larsen, R. M. Aarts, and M. Danessis, “Efficient High-Frequency Bandwidth Extension of Music and Speech,” Presented at the 112th Convention of the AES, Munich, Germany, May 2002, Preprint 5627.

[8] D. Tsoukalas, M. Paraskevas, and J. Mourjopoulos, “Speech Enhancement Using Psychoacoustic Criteria,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1993, vol. 2, pp. 359–362.

[9] D. Tsoukalas, J. Mourjopoulos, and G. Kokkinakis, “Perceptual Filters for Audio Signal Enhancement,” *J. Audio Eng. Soc.*, vol. 45, no. 1/2, pp. 22–36, Jan./Feb. 1997.

[10] M. Lorber and R. Hoeldrich, “A Combined Approach for Broadband Noise Reduction,” in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1997.

[11] P. J. Wolfe and S. J. Godsill, “Towards a Perceptually Optimal Spectral Amplitude Estimator for Audio Signal Enhancement,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, June 2000, vol. 2, pp. 821–824.

[12] P. J. Wolfe and S. J. Godsill, “Perceptually Motivated Approaches to Music Restoration,” *J. New Music Research*, vol. 30, no. 1, pp. 83–92, Mar. 2001, Special Issue: Music and Mathematics.

[13] I. Kauppinen and K. Roth, “Adaptive Psychoacoustic Filter for Broadband Noise Reduction in Audio Signals,” in *Proc. 14th Int. Conf. Digital Signal Processing*, Santorini, Greece, July 2002, vol. 2, pp. 962–966.

[14] B. L. Vercoe, W. G. Gardner, and E. D. Scheirer, “Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations,” *Proc. IEEE*, vol. 86, no. 5, pp. 922–940, May 1998.

[15] D. Ellis, *Prediction-Driven Computational Auditory Scene Analysis*, Ph.D. thesis, MIT, Department of Electrical Engineering and Computer Science, June 1996.

[16] E. Scheirer, “The MPEG-4 Structured Audio Standard,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Seattle, USA, May 1998, vol. 6, pp. 3801–3804.

[17] E. D. Scheirer, “Structured Audio and Effects Processing in the MPEG-4 Multimedia Standard,” *Multimedia Systems*, vol. 7, no. 1, pp. 11–22, 1999.

[18] A. S. Bregman, *Auditory Scene Analysis*, MIT Press, Cambridge, 1990.

[19] D. Rosenthal and H. G. Okuno, Eds., *Computational Auditory Scene Analysis*, Lawrence Erlbaum Associates, Mahwah, New Jersey, USA, 1998.

[20] K. D. Martin, “Automatic Transcription of Simple Polyphonic Music: Robust Front End Processing,” Tech. Rep. 399, Massachusetts Institute of Technology, The Media Laboratory, Dec. 1996.

[21] A. Klapuri, “Automatic Transcription of Music,” M.S. thesis, Tampere Univ. of Technology, Tampere, Finland, 1998, URL: `http://www.cs.tut.fi/~klap/iiro/`.

[22] J. Bello, G. Monti, and M. Sandler, “Techniques for Automatic Music Transcription,” in *Proc. 1st Int. Symposium on Music Information Retrieval*, Plymouth, Massachusetts, USA, Oct. 2000.

[23] E. Scheirer and Y. Kim, “Generalized Audio Coding with MPEG4 Structured Audio,” in *Proc. AES 17th Int. Conf. – High-Quality Audio Coding*, Florence, Italy, Sept. 1999, pp. 189–204.

[24] H. Purnhagen, “Advances in Parametric Audio Coding,” in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1999, pp. 31–34.

[25] T. Tolonen, *Object-Based Sound Source Modeling*, Ph.D. thesis, Helsinki University of Technology, Espoo, Finland, 2000, URL: `http://lib.hut.fi/Diss/2000/isbn9512251965/`.

[26] J. O. Smith, “Viewpoints on the History of Digital Synthesis,” in *Proc. Int. Computer Music Conf.*, Montreal, Canada, Oct. 1991, pp. 1–10.

[27] P. R. Cook, *Real Sound Synthesis for Interactive Applications*, A K Peters, 2002.

[28] A. Chaigne, “On the Use of Finite Differences for Musical Synthesis. Application to Plucked Stringed Instruments,” *J. Acoustique*, vol. 5, no. 2, pp. 181–211, Apr. 1992.

[29] J. O. Smith, “Physical Modeling Using Digital Waveguides,” *Computer Music J.*, vol. 16, no. 4, pp. 74–91, 1992.

[30] X. Serra, “Musical Sound Modeling with Sinusoids plus Noise,” in *Musical Signal Processing*, C. Roads, S. Pope, A. Picialli, and G. De Poli, Eds. Swets & Zeitlinger, 1997.

[31] T. F. Quatieri and R. J. McAulay, "Audio Signal Processing Based on Sinusoidal Analysis and Synthesis," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., pp. 314–416. Kluwer Academic Publishers, Boston, USA, 1998.

[32] E. B. George and M. J. T. Smith, "Analysis-by-Synthesis/Overlap-Add Sinusoidal Modeling Applied to the Analysis and Synthesis of Musical Tones," *J. Audio Eng. Soc.*, vol. 40, no. 6, pp. 497–516, June 1992.

[33] M. H. Hayes, *Statistical Signal Processing and Modeling*, John Wiley & Sons, Inc., 1996.

[34] S. M. Kay, *Modern Spectral Estimation*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1988.

[35] J. Burg, "Maximum Entropy Spectral Analysis," in *Proc. 37th Meeting of Soc. of Exploration Geophysics*, 1967, (reprinted in *Modern Spectrum Estimation*, D. G. Childers, Ed., IEEE Press, New York).

[36] J. Makhoul, "Stable and Efficient Lattice Methods for Linear Prediction," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-25, no. 5, pp. 423–428, Oct. 1977.

[37] K. Steiglitz and L. E. McBride, "A Technique for the Identification of Linear Systems," *IEEE Trans. Automatic Control*, vol. AC-10, pp. 461–464, Oct. 1965.

[38] H. W. Strube, "Linear Prediction on a Warped Frequency Scale," *J. Acoust. Soc. Am.*, vol. 68, no. 4, pp. 1071–1076, Oct. 1980.

[39] J. O. Smith and J. S. Abel, "Bark and ERB Bilinear Transforms," *IEEE Trans. Speech and Audio Processing*, vol. 7, no. 6, pp. 697–708, Nov. 1999.

[40] M. Karjalainen, A. Härmä, and U. K. Laine, "Realizable Warped IIR Filters and their Properties," in *Proc. IEEE Int. Conf. Audio, Speech, and Signal Processing*, Munich, Germany, Apr. 1997, vol. 3, pp. 2205–2209.

[41] A. Härmä, "Implementation of Recursive Filters Having Delay Free Loops," in *Proc. IEEE Int. Conf. Audio, Speech, and Signal Processing*, Seattle, USA, May 1998, vol. 3, pp. 1261–1264.

[42] A. Härmä, "Audio Coding with Warped Predictive Methods," Licentiate thesis, Helsinki University of Technology, Espoo, Finland, 1998, URL: `http://www.acoustics.hut.fi/~aqi/papers/LT.ps.gz`.

[43] A. Härmä, M. Karjalainen, V. Välimäki, L. Savioja, U. Laine, and J. Huopaniemi, "Frequency-Warped Signal Processing for Audio Applications," *J. Audio Eng. Soc.*, vol. 48, no. 11, pp. 1011–1031, Nov. 2000.

[44] K. Karplus and A. Strong, "Digital Synthesis of Plucked String and Drums Timbres," *Computer Music J.*, vol. 7, no. 2, pp. 43–55, 1983.

[45] D. Jaffe and J. O. Smith, "Extension of the Karplus-Strong Plucked String Algorithm," *Computer Music J.*, vol. 7, no. 2, pp. 56–69, 1983.

[46] M. Karjalainen, V. Välimäki, and T. Tolonen, "Plucked-String Models: From the Karplus-Strong Algorithm to Digital Waveguides and Beyond," *Computer Music J.*, vol. 22, no. 3, pp. 17–32, 1998.

[47] M. Karjalainen and J. O. Smith, "Body Modeling Techniques for String Instrument Synthesis," in *Proc. Int. Computer Music Conf.*, San Francisco, California, USA, Aug. 1996, pp. 232–239.

[48] J. O. Smith, "Efficient Synthesis of Stringed Musical Instruments," in *Proc. Int. Computer Music Conf.*, Tokyo, Japan, Sept. 1993, pp. 64–71.

[49] M Karjalainen, V. Välimäki, and Z. Jánosy, "Towards High-Quality Sound Synthesis of the Guitar and String Instruments," in *Proc. Int. Computer Music Conf.*, Tokyo, Japan, Sept. 1993, pp. 56–63.

[50] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the Unit Delay – Tools for Fractional Delay Filter Design," *IEEE Signal Processing Magazine*, vol. 13, no. 1, pp. 30–60, Jan. 1996.

[51] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Jánosy, "Physical Modeling of Plucked String Instruments with Application to Real-Time Sound Synthesis," *J. Audio Eng. Soc.*, vol. 44, no. 5, pp. 331–353, May 1996.

[52] V. Välimäki and T. Tolonen, "Development and Calibration of a Guitar Synthesizer," *J. Audio Eng. Soc.*, vol. 46, no. 9, pp. 766–778, Sept. 1998.

[53] B. Bank, "Physics-Based Sound Synthesis of the Piano," M.S. thesis, Helsinki University of Technology, Lab. of Acoustics and Audio Signal Processing, June 2000, Report no. 54, URL:
`http://www.acoustics.hut.fi/publications/`.

[54] B. Bank and V. Välimäki, "Robust Loss Filter Design for Digital Waveguide Synthesis of String Tones," *IEEE Signal Processing Letters*, vol. 10, no. 1, pp. 18–20, Jan. 2003.

[55] L. Trautmann, B. Bank, V. Välimäki, and R. Rabenstein, "Combining Digital Waveguide and Functional Transformation Methods for Physical Modeling of Musical Instruments," in *Proc. AES 22th Int. Conf. – Virtual, Synthetic, and Entertainment Audio*, Espoo, Finland, June 2002, pp. 307–316.

[56] C. Erkut, V. Välimäki, M. Karjalainen, and M. Laurson, "Extraction of Physical and Expressive Parameters for Model-Based Sound Synthesis of the Classical Guitar," Presented at the 108th Convention of the AES, Paris, France, Feb. 2000, Preprint 5114, URL:
`http://lib.hut.fi/Diss/2002/isbn9512261901/`.

[57] A. W. Y. Su and S.-F. Liang, "A New Automatic IIR Analysis/Synthesis Technique for Plucked-String Instruments," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 7, pp. 747–754, Oct. 2001.

[58] J. Riionheimo and V. Välimäki, "Parameter Estimation of a Plucked String Synthesis Model Using a Genetic Algorithm with Perceptual Fitness Calculation," *EURASIP J. Applied Signal Processing*, vol. 2003, no. 8, pp. 791–805, July 2003.

[59] J. Laroche, "A New Analysis/Synthesis System of Musical Signals Using Prony's Method – Application to Heavily Damped Percussive Sounds," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 1989, vol. 3, pp. 2053–2056.

[60] M. Sandler, "New Synthesis Techniques for High-Order All-Pole Filters," *IEE Proceedings I on Communications, Speech and Vision*, vol. 139, no. 1, pp. 9–16, Feb. 1991.

[61] J. Mackenzie, I. Kale, and G. D. Cain, "Applying Balanced Model Truncation to Sound Analysis/Synthesis Models," in *Proc. Int. Computer Music Conf.*, Sept. 1995, pp. 400–403.

[62] G. Weinreich, “Coupled Piano Strings,” *J. Acoust. Soc. Am.*, vol. 62, no. 6, pp. 1474–1484, Dec. 1977.

[63] V. Välimäki, M. Karjalainen, T. Tolonen, and C. Erkut, “Nonlinear Modeling and Synthesis of the Kantele – a Traditional Finnish String Instrument,” in *Proc. Int. Computer Music Conf.*, Beijing, China, Oct. 1999, pp. 220–223.

[64] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, “AR/ARMA Analysis and Modeling of Modes in Resonant and Reverberant Systems,” Presented at the 112th Convention of the AES, Munich, Germany, May 2002, Preprint 5590.

[65] S. A. Van Duyne and J. O. Smith, “A Simplified Approach to Modeling Dispersion Caused by Stiffness in Strings and Plates,” in *Proc. Int. Computer Music Conf.*, Sept. 1994, pp. 407–414.

[66] A. Paladin and D. Rocchesso, “A Dispersive Resonator in Real Time on MARS Workstation,” in *Proc. Int. Computer Music Conf.*, Oct. 1992, pp. 146–149.

[67] D. Rocchesso and F. Scalcon, “Accurate Dispersion Simulation for Piano Strings,” in *Proc. Nordic Acoustical Meeting*, Helsinki, Finland, June 1996, pp. 407–414.

[68] G. Essl and P. R. Cook, “Banded Waveguides: Towards Physical Modeling of Bar Percussion Instruments,” in *Proc. Int. Computer Music Conf.*, Beijing, China, Oct. 1999, pp. 321–324.

[69] B. Bank, V. Välimäki, L. Sujbert, and M. Karjalainen, “Efficient Physics-Based Sound Synthesis of the Piano Using DSP Methods,” in *Proc. European Signal Processing Conf.*, Tampere, Finland, Sept. 2000, vol. 4, pp. 2225–2228.

[70] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kronland-Martinet, “Resynthesis of Coupled Piano String Vibrations Based on Physical Modeling,” *J. New Music Research – Special Issue: Musical Applications of Digital Signal Processing*, vol. 30, no. 5, pp. 313–326, Sept. 2001.

[71] J. O. Smith, “Nonlinear Commuted Synthesis of Bowed Strings,” in *Proc. Computer Music Conf.*, Thessaloniki, Greece, Sept. 1997, pp. 264–267.

[72] T. Tolonen, V. Välimäki, and M. Karjalainen, "Modeling of Tension Modulation Nonlinearity in Plucked Strings," *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 3, pp. 300–310, May 2000.

[73] H. Järveläinen, *Perception of Attributes in Real and Synthetic String Instrument Sounds*, Ph.D. thesis, Helsinki University of Technology - Lab. of Acoustics and Audio Signal Processing, Espoo, Finland, Feb. 2003, URL: `http://lib.hut.fi/Diss/2003/isbn9512263149/`.

[74] S. V. Vaseghi, *Algorithms for Restoration of Archived Gramophone Recordings*, Ph.D. thesis, Engineering Department, Cambridge University, Cambridge, England, 1988.

[75] S. V. Vaseghi and R. Frayling-Cork, "Restoration of Old Gramophone Recordings," *J. Audio Eng. Soc.*, vol. 40, no. 10, pp. 791–801, Oct. 1992.

[76] S. J. Godsill, *The Restoration of Degraded Audio Signals*, Ph.D. thesis, Engineering Department, Cambridge University, Cambridge, England, Dec. 1993, URL: `http://www-sigproc.eng.cam.ac.uk/~sjg/thesis/` .

[77] S. J. Godsill and C. H. Tan, "Removal of Low Frequency Transient Noise from Old Recordings Using Model-Based Signal Separation Techniques," in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, Oct. 1997.

[78] P. A. A. Esquef, L. W. P. Biscainho, V. Välimäki, and M. Karjalainen, "Removal of Long Pulses from Audio Signals Using Two-Pass Split-Window Filtering," Presented at the 112th Convention of the AES, Munich, Germany, May 2002, Preprint 5535.

[79] W. A. Struzinski, "A New Normalization Algorithm for Detection Systems," *J. Acoust. Soc. Am., Suppl.*, vol. 75, no. S1, pp. S43, 15 Apr. 1984.

[80] W. A. Struzinski and E. D. Lowe, "A Performance Comparison of Four Noise Background Normalization Schemes Proposed for Signal Detection Systems," *J. Acoust. Soc. Am.*, vol. 76, no. 6, pp. 1738–1742, Dec. 1984.

[81] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, July 1964.

[82] W. H. Press and S. A. Teukolsky, "Savitzky-Golay Smoothing Filters," *Computers in Physics*, vol. 4, no. 6, pp. 669–672, 1990.

[83] P. Wilson, "Record Contamination: Causes and Cure," *J. Audio Eng. Soc.*, vol. 13, no. 2, pp. 166–176, Apr. 1965.

[84] G. R. Kinzie, Jr. and D. W. Gravereaux, "Automatic Detection of Impulsive Noise," Presented at the 41st Convention of the AES, Oct. 1971, Preprint 824.

[85] G. R. Kinzie, Jr. and D. W. Gravereaux, "Automatic Detection of Impulsive Noise," *J. Audio Eng. Soc.*, vol. 21, no. 3, pp. 331–336, Mar. 1973.

[86] R. S. Burwen, "Suppression of Low Level Impulsive Noise," Presented at the 61st Convention of the AES, New York, NY, USA, Nov. 1978, Preprint 1388.

[87] P. G. Craven and M. A. Gerzon, "The Elimination of Scratch Noise from 78 RPM Records," Presented at the 50th Convention of the AES, Mar. 1975, Preprint L-37.

[88] S. V. Vaseghi and P. J. W. Rayner, "A New Application of Adaptive Filters for Restoration of Archived Gramophone Recordings," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1988, vol. 5, pp. 2548–2551.

[89] S. V. Vaseghi and P. J. W. Rayner, "Detection and Suppression of Impulsive Noise in Speech Communication Systems," *IEE Proceedings*, vol. 137, no. 1, pp. 38–46, Feb. 1990.

[90] P. A. A. Esquef, L. W. P. Biscainho, P. S. R. Diniz, and F. P. Freeland, "A Double-Threshold-Based Approach to Impulsive Noise Detection in Audio Signals," in *Proc. X European Signal Processing Conf.*, Tampere, Finland, Sept. 2000, vol. 4, pp. 2041–2044.

[91] A. J. E. M. Janssen, R. N. J. Veldhuis, and L. B. Vries, "Adaptive Interpolation of Discrete-Time Signals that Can Be Modeled as Autoregressive Processes," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-34, no. 2, pp. 317–330, Apr. 1986.

[92] R. Veldhuis, *Restoration of Lost Samples in Digital Signals*, Prentice-Hall, 1990.

[93] S. J. Godsill and P. J. W. Rayner, "A Bayesian Approach to the Detection and Correction of Error Bursts in Audio Signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Mar. 1992, vol. 2, pp. 261–264.

[94] S. J. Godsill and P. J. W. Rayner, "A Bayesian Approach to the Restoration of Degraded Audio Signals," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 4, pp. 276–278, July 1995.

[95] S. J. Godsill and P. J. W. Rayner, "Robust Noise Reduction for Speech and Audio Signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 1996, vol. 2, pp. 625–628.

[96] A. Czyzewski, "Learning Algorithms for Audio Signal Enhancement, Part I: Neural Network Implementation for the Removal of Impulsive Distortions," *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 815–831, Oct. 1997.

[97] M. Niedźwiecki and K. Cisowski, "Adaptive Scheme for Elimination of Broadband Noise and Impulsive Disturbances from AR and ARMA Signals," *IEEE Trans. Signal Processing*, vol. 44, no. 3, pp. 528–537, Mar. 1996.

[98] A. Bari, S. Canazza, G. De Poli, and G. A. Mian, "Some Key Points on Restoration of Audio Documents by the Extended Kalman Filter," in *Proc. Diderot Forum on Mathematics and Music*, Vienna, Austria, Dec. 1999, pp. 37–47.

[99] A. Papoulis, "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," *IEEE Trans. Circuits Syst.*, vol. 22, no. 9, pp. 735–742, Sept. 1975.

[100] P. J. S. G. Ferreira, "Iterative and Noniterative Recovery of Missing Samples for 1-D Band-Limited Signals," in *Nonuniform Sampling – Theory and Practice*, F. Marvasti, Ed., chapter 5, pp. 235–281. Kluwer Academic/Plenum Publishers, New York, N.Y., USA, 2001.

[101] R. C. Maher, "A Method for Extrapolation of Missing Digital Audio Data," *J. Audio Eng. Soc.*, vol. 42, no. 5, pp. 350–357, May 1994.

[102] D. J. Goodman, G. B. Lockhart, O. J. Waen, and W. C. Wong, "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 6, pp. 1440–1448, Dec. 1986.

[103] M. Niedźwiecki and K. Cisowski, "Smart Copying – A New Approach to Reconstruction of Audio Signals," *IEEE Trans. Signal Processing*, vol. 49, no. 10, pp. 2272–2282, Oct. 2001.

[104] S. Montresor, J. C. Valiere, J. F. Allard, and M. Baudry, "Evaluation of Two Interpolation Methods Applied to Old Recordings Restoration," Presented at the 90th Convention of the AES, Paris, France, Feb. 1991, Preprint 3022.

[105] B.-S. Chen and Y.-L. Chen, "Multirate Modeling of AR/ARMA Stocastic Signals and Its Application to the Combined Estimation-Interpolation Problem," *IEEE Trans. Signal Processing*, vol. 43, no. 10, pp. 2302–2312, Oct. 1995.

[106] G. Cocchi and A. Uncini, "Subbands Audio Signal Recovering Using Neural Nonlinear Prediction," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2001, vol. 2, pp. 1289–1292.

[107] P. J. W. Rayner and S. J. Godsill, "The Detection and Correction of Artefacts in Degraded Gramophone Recordings," in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1991, pp. 151–152.

[108] M. Niedźwiecki, "Statistical Reconstruction of Multivariate Time Series," *IEEE Trans. Signal Processing*, vol. 41, no. 1, pp. 451–457, Jan. 1993.

[109] J. J. Ó Ruanaidh and W. J. Fitzgerald, "Interpolation of Missing Samples for Audio Restoration," *Electronics Letters*, vol. 30, no. 8, pp. 622–623, Apr. 1994.

[110] S. J. Godsill and P. J. W. Rayner, "Statistical Reconstruction and Analysis of Autoregressive Signals in Impulsive Noise Using the Gibbs Sampler," *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 4, pp. 352–372, July 1998.

[111] W. Etter, "Restoration of a Discrete-Time Signal Segment by Interpolation Based on the Left-Sided and Right-Sided Autoregressive Parameters," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1124–1135, May 1996.

[112] I. Kauppinen, J. Kauppinen, and P. Saarinen, "A Method for Long Extrapolation of Audio Signals," *J. Audio Eng. Soc.*, vol. 49, no. 12, pp. 1167–1180, Dec. 2001.

[113] J. G. Beerends, "Audio Quality Determination Based on Perceptual Measurement Techniques," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., chapter 1, pp. 1–38. Kluwer Academic Publishers, Boston, USA, 1998.

[114] S. J. Godsill, P. J. W. Rayner, and O. Cappé, "Digital Audio Restoration," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., chapter 4, pp. 133–194. Kluwer Academic Publishers, Boston, USA, 1998.

[115] J. Lim, "Evaluation of a Correlation Subtraction Method for Enhancing Speech Degraded by Additive White Noise," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 26, no. 5, pp. 471–472, Oct. 1978.

[116] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.

[117] J. Lim, "Speech Enhancement," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1986, vol. 11, pp. 3135–3142.

[118] O. Cappé and J. Laroche, "Evaluation of Short-Time Spectral Attenuation Techniques for the Restoration of Musical Recordings," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 1, pp. 84–93, Jan. 1995.

[119] P. J. Wolfe and S. J. Godsill, "Efficient Alternatives to the Ephraim and Malah Suppression Rule for Audio Signal Enhancement," *EURASIP Journal on Applied Signal Processing – Special Issue on Digital Audio for Multimedia Communications*, vol. 2003, no. 10, pp. 1043–1051, Sept. 2003.

[120] S. F. Boll, "A Spectral Subtraction Algorithm for Suppression of Acoustic Noise in Speech," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1979, vol. 4, pp. 200–203.

[121] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1979, vol. 4, pp. 208–211.

[122] J. A. Moorer and M. Berger, "Linear-Phase Bandsplitting: Theory and Applications," *J. Audio Eng. Soc.*, vol. 34, no. 3, pp. 143–152, Mar. 1986.

[123] A. Bari, S. Canazza, G. De Poli, and G. A. Mian, "Toward a Methodology for the Restoration of Electroacoustic Music," *J. New Music Research – Special Issue: Preserving Electroacoustic Music*, vol. 30, no. 4, Dec. 2001.

[124] Y. Ephraim and D. Malah, "Speech Enhancement Using Optimal Non-Linear Spectral Amplitude Estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1983, vol. 8, pp. 1118–1121.

[125] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[126] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[127] O. Cappé, "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor," *IEEE Trans. Speech and Audio Processing*, vol. 2, no. 2, pp. 345–349, Apr. 1994.

[128] P. J. Wolfe and S. J. Godsill, "Simple Alternatives to the Ephraim and Malah Suppression Rule for Speech Enhancement," in *Proc. 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Aug. 2001, pp. 496–499.

[129] J. Berger, R. R. Coifman, and M. J. Goldberg, "Removing Noise from Music Using Local Trigonometric Bases and Wavelet Packets," *J. Audio Eng. Soc.*, vol. 42, no. 10, pp. 808–818, Oct. 1994.

[130] A. Teolis and J. J. Benedetto, "Noise Suppression Using a Wavelet Model," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1994, vol. 1, pp. 17–20.

[131] D. L. Donoho, "De-Noising by Soft-Thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[132] H. K. S. Mallat, D. L. Donoho, and A. S. Willsky, "Best Basis Algorithm for Signal Enhancement," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 1995, vol. 3, pp. 1561 –1564.

[133] N. A. Whitmal, J. C. Rutledge, and J. Cohen, "Wavelet-Based Noise Reduction," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 1995, vol. 5, pp. 3003–3006.

[134] P. K. Ramarapu and R. C. Maher, "Methods for Reducing Audible Artifacts in a Wavelet-Based Broad-Band Denoising System," *J. Audio Eng. Soc.*, vol. 46, no. 3, pp. 178–190, Mar. 1998.

[135] J. Lim and A. V. Oppenheim, "All-Pole Modeling of Degraded Speech," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 26, no. 3, pp. 197–210, Jun. 1978.

[136] M. Niedźwiecki, "Recursive Algorithms for Elimination of Measurement Noise and Impulsive Disturbances From ARMA Signals," *Signal Processing VII: Theories and Applications*, pp. 1289–1292, 1994.

[137] S. B. Jebara, A. Benazza-Benyahia, and A. B. Khelifa, "Reduction of Musical Noise Generated by Spectral Subtraction by Combining Wavelet Packet Transform and Wiener Filtering," in *Proc. X European Signal Processing Conf.*, Tampere, Finland, Sept. 2000, vol. 2.

[138] I. Kauppinen and K. Roth, "Audio Signal Extrapolation – Theory and Applications," in *Proc. 5th Int. Conf. on Digital Audio Effects*, Hamburg, Germany, Sept. 2002, pp. 105–110, URL: `http://www.unibw-hamburg.de/EWEB/ANT/dafx2002/papers.html`.

[139] J. Laroche and J.-L. Meillier, "A Simplified Source/Filter Model for Percussive Sounds," in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1993, pp. 173–176.

[140] J. Laroche, "The Use of The Matrix Pencil Method for the Spectrum Analysis of Musical Signals," *J. Acoust. Soc. Am.*, vol. 94, no. 4, pp. 1958–1965, Oct. 1993.

[141] R. B. Sussman and M. Kahrs, “Analysis and Resynthesis of Musical Instrument Sounds Using Energy Separation,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Atlanta, Georgia, USA, May 1996, vol. 2, pp. 997–1000.