

UNIVERSITÀ
POLITECNICA
DELLE MARCHE

DII - Department of Information Engineering
PhD in Biomedical, Electronics and Telecommunication Engineering

Metodologie Innovative di Digital Signal Processing per l'Identificazione e l'Analisi di Sistemi Sonori Reali

Innovative Digital Signal Processing Methodologies for Identification and Analysis of Real Audio Systems

Thesis Advisor
Prof. Stefania Cecchi

Candidate
Alessandro Terenzi

Academic Year 2019-2020 (XXXIV cycle)

Alla mia famiglia

Abstract

Many real word audio systems exist, each has its own characteristics but almost all of them can be identified from the fact that they are able to generate or modify a sound. If a natural or artificial system can be defined as a sound system, then it is possible to apply the techniques of digital signal processing for the studying and the emulation of the system. In this thesis, innovative methodologies for digital signal processing applied to real audio systems will be discussed. In particular, three different audio systems will be considered: the world of vacuum-based non linear audio devices with particular attention to guitar and hi-fi amplifiers; the room acoustic environment and its effect on the sound propagation; and finally the sound emitted by honey bees in a beehive. Regarding the first system, innovative approaches for the identification of the Volterra series and Hammerstein models will be proposed, in particular an approach to overcome some limitation of Volterra series identification. The application of a sub-band structure to reduce the computational cost and increase the convergence speed of an adaptive Hammerstein model identification will be proposed as well. Finally, an innovative approach for the measurement of several distortion parameters using a single measure, exploiting a generalized Hammerstein model, will be presented. For the second system, the results of the application of a multi-point equalizer to two different situations will be exposed. In particular, in the first case, it will be shown how a multi-point equalization can be used not only to compensate the acoustical anomalies of a room, but also to improve the frequency response of vibrating transducers mounted on a rigid surface. The second contribution will show how a sub-band approach can be used to improve the computational cost and the speed of an adaptive algorithm for a multi-point and multi channel equalizer. At the end, the focus will be on a natural sound system, i.e., a honey bees colony. In this case, an innovative acquisition system for honey bees sound monitoring will be presented. Then, the approaches developed for sound analysis will be exposed and applied to the recorded sounds in two different situations. Finally, the obtained results, achieved with the application of classification algorithms, will be exposed. In the final part of the work some minor contributions still related to signal processing applied to real sound systems are presented. In particular, an implementation of an active noise control system is discussed, and two algorithms for digital effects where the former improves the sound performances of compact loudspeakers and the latter generates a stereophonic effect for electric guitars are exposed.

Abstract

Esistono molti sistemi audio reali e ciascuno ha le proprie caratteristiche ma tutti sono accomunati dal fatto che sono sistemi in grado di generare o modificare un suono. Se un sistema naturale o artificiale può essere definito come sistema sonoro, allora è possibile applicare le tecniche del digital signal processing per studiare il sistema ed emularne il comportamento. In questo lavoro di tesi si propone di introdurre delle metodologie innovative di processamento del segnale applicate ad alcuni sistemi sonori reali. In particolare, vengono studiati e discussi tre diversi sistemi: il mondo dei dispositivi non lineari basati su valvole, con particolare attenzione agli amplificatori per chitarra e hi-fi, l'ambiente acustico di una stanza ed il suo effetto sulla propagazione del suono ed infine il suono emesso dalle api in un alveare. Per quanto riguarda il primo sistema, vengono proposti dei contributi innovativi per l'identificazione di modelli come la serie di Volterra ed il modello di Hammerstein; in particolare viene discusso un approccio per superare alcune limitazioni dell'identificazione tramite serie di Volterra e l'applicazione di una struttura in sottobande per ridurre il costo computazionale e incrementare la velocità di convergenza di un algoritmo adattativo per l'identificazione del modello di Hammerstein. In ultima analisi, viene proposto un approccio innovativo in grado di stimare con una singola misura vari parametri di distorsione sfruttando un modello di Hammerstein generallizato. Per quanto riguarda il secondo ambito, vengono proposti i risultati relativi a due applicazioni di equalizzazione multipunto: nel primo caso si mostrerà come l'equalizzazione può essere usata non solo per compensare le anomalie sonore generate all'interno di una stanza, ma anche per migliorare la risposta in frequenza di particolari trasduttori a vibrazione ancorati ad un pannello rigido; nel secondo caso si illustra come un approccio in sottobande può migliorare l'efficienza computazionale e la velocità di un algoritmo di equalizzazione adattativo multipunto e multicanale. Infine, viene presentato un sistema sonoro naturale, ovvero quello generato da un alveare. In questo caso si illustrerà un sistema di acquisizione innovativo sviluppato per monitorare gli alveari con particolare attenzione al suono; succesivamente si mostrano gli approcci messi a punto per analizzare il suono registrato in due condizioni reali ed infine verranno si illustrano i risultati ottenuti grazie allo studio del suono usando algoritmi di classificazione. Inoltre, nella parte finale dell'elaborato sono presenti dei contributi secondari ma che hanno comunque come focus principale il signal processing applicato ad ambienti acustici reali, in particolare si discute di un'implementazione di un algoritmo di cancellazione attiva del rumore e di due algoritmi per effetti digitali in cui il primo migliora le performance sonore di altoparlanti compatti, ed il secondo genera un effetto stereofonico per chitarra elettrica.

Acknowledgements

Il dottorato è stata una bellissima esperienza, ma piena di sfide e difficoltà, per questo motivo credo che non sarei stato in grado di arrivare in fondo senza l'aiuto ed il supporto della famiglia, degli amici e dei colleghi.
Il primo ringraziamento va senza dubbio a mia moglie Vanessa, che mi è sempre stata accanto e mi ha supportato oltre che sopportato (ammetto che in certi momenti avrei forse meritato di finire a dormire in giardino o in garage).
Il secondo pensiero va alle famiglie, non solo la mia (grazie Giorgio, Oriella, Valentina, nonna Gina e tutti gli zii) ma anche a quella di Vanessa che ormai da un po' di anni mi considera a tutti gli effetti un figlio.
Devo poi ringraziare gli scout, che ormai da più di 20 anni fanno parte della mia vita, il servizio coi ragazzi e la fede in Dio che mi hanno insegnato sono una parte fondamentale di me.
Grazie alla Prof.ssa Stefania Cecchi. Grazie per la sua fiducia e per avermi dato questa possibilità; nonostante a volte io sia un po' "goffo" mi ha sempre incoraggiato cercando di farmi crescere e migliorare.
Grazie al Prof. Francesco Piazza e a tutti gli altri professori e ricercatori che in questi anni mi hanno permesso di collaborare con loro facendomi crescere e apprendere nuove conoscenze.
Grazie poi ai colleghi dell'ufficio: Stefano, Valeria e Nicola, perché prima ancora che validi colleghi sono stati dei buoni amici in grado di alleggerire tante giornate passate insieme al lavoro.
Non posso poi dimenticare tutti gli altri amici conosciuti sia durante gli anni dell'università che durante il dottorato, senza di loro non avrei il bellissimo ricordo che ho degli anni di studio e di questi anni passati in facoltà come dottorando e assegnista.
Un ringraziamento speciale va anche alla legge di Murphy, sempre presente in questi anni ed in particolare in questo ultimo periodo: senza di essa mancherebbe qualcosa nella mia vita.
In ultimo, non posso non ringraziare anche le api, il progetto che le riguarda è senza dubbio quello che mi porto più nel cuore. Ho imparato moltissimo da loro e poterle osservare da vicino non era certo qualcosa che mi sarei aspettato di fare durante il dottorato, ma che mi sento fortunato ad aver fatto. Inoltre, mi piace inoltre pensare che il nostro lavoro con loro possa essere un piccolissimo contributo che diamo al pianeta, con la speranza di lasciarlo un po' migliore di come lo abbiamo trovato.

Contents

List of Figures

List of Tables

1 Introduction **1**
- 1.1 Thesis Contributions . . . 2
- 1.2 Thesis outline . . . 3

2 Non linear audio system identification, modeling and measurement **4**
- 2.1 Background . . . 4
- 2.2 Example of non linear audio systems . . . 6
- 2.3 Non linear system modeling . . . 8
  - 2.3.1 LIP non linear filters . . . 8
  - 2.3.2 Volterra series . . . 9
  - 2.3.3 Wiener-Hammerstein model . . . 9
- 2.4 Non linear system identification . . . 10
  - 2.4.1 Multivariance Volterra series identification . . . 10
  - 2.4.2 Sub-band Hammerstein adaptive identification . . . 24
- 2.5 Non linear measurement . . . 31
  - 2.5.1 Non linear parameters . . . 31
  - 2.5.2 Proposed approach . . . 33
- 2.6 Conclusions . . . 47

3 Room response identification and equalization **48**
- 3.1 Background . . . 48
- 3.2 Room impulse response identification . . . 50
  - 3.2.1 Maximum length sequences . . . 50
  - 3.2.2 Sine sweep . . . 51
- 3.3 Equalization algorithm . . . 52
  - 3.3.1 Multi-point equalizer . . . 52
- 3.4 Proposed approaches . . . 53
  - 3.4.1 Applying sound equalization to vibrating sound transducers mounted on rigid panels . . . 54
  - 3.4.2 Sub-band adaptive multi-point multi-channel equalizer . . . 57
- 3.5 Conclusions . . . 70

4 Honey bee hive as real audio system: acquisition and analysis **71**
- 4.1 Background . . . 71
- 4.2 The sound acquisition system . . . 81
  - 4.2.1 Hardware . . . 81
  - 4.2.2 Sound visualization and management . . . 83
- 4.3 Honey bees sound analysis . . . 83

4.3.1 Features extraction techniques . . . 83
4.3.2 Classifiers . . . 87
4.3.3 Experimental results . . . 90
4.3.4 Orphaned colony . . . 90
4.3.5 Swarming . . . 100
4.4 Conclusions . . . 105

5 Other contributions **106**
5.1 Active noise control for snoring reduction . . . 106
5.1.1 Adaptive Feedforward Active Noise Control for Snooring Cancellation . . . 107
5.1.2 Algorithm description . . . 108
5.1.3 Experimental results . . . 110
5.1.4 conclusions . . . 112
5.2 Stereo reproduction enhancement . . . 112
5.2.1 Algorithm Description . . . 112
5.2.2 Algorithm Implementation . . . 115
5.2.3 Algorithm testing . . . 116
5.2.4 Subjective results . . . 117
5.2.5 Conclusions . . . 119
5.3 Applying decorrelation to a guitar vibrato effect . . . 120
5.3.1 Proposed algorithm . . . 120
5.3.2 Algorithm implementation . . . 122
5.3.3 Experimental results . . . 123
5.3.4 Algorithm testing . . . 123
5.3.5 Conclusions . . . 125

6 Conclusions **127**
6.1 List of pubblications . . . 129

Bibliography **131**

List of Figures

2.1 Physical construction and electrical schematic of triode valve, image from [163] . . . 6
2.2 Basic schematic of a vacuum tube amplifier. Red lines represent the power supply, while dashed lines represent some environment effect which are present only on guitar amplifiers, image adapted from [64] . . . 7
2.3 Wiener-Hammerstein models for non linear system representation. . . . 9
2.4 Input signals used (a) for the identification (b) for testing, expressed in procedure defined units (p.d.u.) [157]. . . . 12
2.5 Roma 27AC: first order Volterra kernel, (a) in time domain ($h_1(\tau)$) (b) in frequency domain ($|H_1(f)|^2$). . . . 15
2.6 NMSE of complete Volterra models of Roma 27AC. (a) WGN input, (b) music input. 16
2.7 NMSE of reduced Volterra models of Roma 27AC. (a) WGN input, (b) music input. 16
2.8 Roma 27AC: second order Volterra kernel, $h_2(\tau_1, \tau_2)$. . . . 17
2.9 Roma 27AC: third order Volterra kernel $h_3(\tau_1, \tau_2, \tau_3)$. . . . 17
2.10 OTL509/25: first order Volterra kernel (a) in time domain ($h_1(\tau)$) (b) in frequency domain ($|H_1(f)|^2$). . . . 18
2.11 NMSE of complete Volterra models of OTL509/25. (a) WGN input, (b) music input. 18
2.12 NMSE of reduced Volterra models of OTL509/25. (a) WGN input, (b) music input. 19
2.13 OTL509/25: second order Volterra kernel ($h_2(\tau_1, \tau_2)$). . . . 19
2.14 OTL509/25: third order Volterra kernel ($h_3(\tau_1, \tau_2, \tau_3)$). . . . 20
2.15 TubePre: first order Volterra kernel (a) in time domain ($h_1(\tau)$) (b) in frequency domain ($|H_1(f)|^2$). . . . 21
2.16 NMSE of complete Volterra models of TubePre for (a) WGN input, (b) music input. 22
2.17 NMSE of complete Volterra models of TubePre for (a) WGN input, (b) music input. 22
2.18 TubePre: second order Volterra kernel, $h_2(\tau_1, \tau_2)$. . . . 23
2.19 TubePre: third order Volterra kernel $h_3(\tau_1, \tau_2, \tau_3)$. . . . 23
2.20 A multi-band structure (MSAF) for Hammerstein system identification. Identical fullband adaptive filters $W(z)$ and polynomial coefficients p are placed in all the sub-bands. Synthesis and analysis filters are generated trough WPT and DCT. . . . 26
2.21 Two level wavelet decomposition filterbanks, in particular (a) shows a non uniform discrete wavelet transform decomposition three, while (b) shows a uniform wavelet packet transform decomposition three. . . . 28
2.22 (a) Vacuum tube synthetic model block diagram, and (b) measurement setup used for PreSonus tube pre identification process. . . . 29
2.23 Comparison in terms of MSE on experiment of Equation (2.42) between (a) fullband algorithm, (b) DCT approach and (c) wavelet approach. . . . 30
2.24 Comparison in terms of MSE on experiments on the tube pre-amplifier PreSonus TubePre with gain of 0dB between (a) fullband algorithm, (b) DCT approach and (c) wavelet approach. . . . 31
2.25 Comparison in terms of MSE on experiments on the tube pre-amplifier PreSonus TubePre with gain of 15dB between (a) fullband algorithm, (b) DCT approach and (c) wavelet approach. . . . 31

2.26 (a) Generalized polynomial Hammerstein model scheme and (b) exponential chirp signal in a non-linear device: time behavior of the fundamental frequency and of the first two harmonics. The time distance between the frequency of a harmonic and the same frequency seen as fundamental frequency is independent from the considered frequency. 35
2.27 Trend of the $o(t)$ signal in the simple case of a quadratic non-linearity: when the exponential chirp $x^A_{exp}(t)$ is input (black curve); when the exponential chirp $x^B_{exp}(t)$ is input (red curve); when the sum of the two exponential chirp signals ($x^A_{exp}(t) + x^B_{exp}(t)$) is input (green curve). The peaks associated to the sum and difference intermodulation components are evidenced in the green curve. 39
2.28 IMD estimation for different Signal To Noise Ratio (SNR) values, in case the case of the non-linear polynomial function defined by (2.63). 41
2.29 Results in terms of percent in the case of non-linear function defined by 2.63 (a) IMD, (b) THD, (c) HD_2, and (d) HD_3 considering the proposed approach and the IEC 60268 measurements. 42
2.30 Results in terms of percent for the synthetic experiment described in Figure 2.22(a) in terms of percent (a) IMD, (b) THD, (c) HD_2, and (d) HD_3 considering the proposed approach and the IEC 60268 measurements. 42
2.31 (a) Hardware connection for the real experiment considering the PreSonus TubePre preamplifier, and (b) Synthesis Roma 27AC amplifier. 44
2.32 Results in terms of percent for the PreSonus tube preamplifier experiment (a) IMD, (a) THD, (c) HD_2, and (d) HD_3 considering the proposed approach and the IEC 60268 measurements. 44
2.33 Results in terms of percent for the Roma Hi-Fi amplifier with Auna loudspeakers experiment, (a) IMD, (b) THD, (c) HD_2, and (d) HD_3 considering the proposed approach and the IEC 60268 measurements. 45

3.1 Schematic of impulse response measurement setup, the input signal $x(n)$ is generated and then reproduced by means of an amplifier and a loudspeaker, the signal travel through the environment and is modified by its impulse response $h(n)$, finally the output signal $y(n)$ is acquired with a microphone. 50
3.2 Real room impulse response (a), and its frequency response (b), the impulse response gives important information of the room reflections and the time of propagation of the sound, while the frequency response shows how the frequencies are attenuated from the environment. 50
3.3 Schematic of impulse response measurement using a MLS (a), and a sine sweep (b) . 51
3.4 Scheme of multi-point equalization algorithm 52
3.5 (a),(b) shaker mounted on a wood panel, and (c) multi-point equalization results taking into consideration one of the measured impulse response, the prototype function and the final equalizer. 54
3.6 Multi-point equalization results: (a) Original sweep, (b) Recorded sweep, (c) Recorded sweep considering the equalization procedure. 55
3.7 Results for subjective listening tests taking into consideration three attributes and the tracks of Table 3.1(a). 56
3.8 Block diagram of the proposed solution. 57
3.9 Multi-channel decorrelation procedure. 57
3.10 Sub-band IRs identification procedure with $L = 2$ and $M = 1$. 60
3.11 Loudspeakers and microphones positions for ((a)) experiments 1 and 2 and ((a)) experiments 3 and 4. 62
3.12 MSE for D=1, 64, 128, 256, 512, 1024, considering white noise as input signal. 64

3.13 Experiment 1: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to the Right channel. 65
3.14 Experiment 2: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to Left channel. 65
3.15 Experiment 3: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to the Right channel. 66
3.16 Experiment 4: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to Left channel. 66
3.17 Experiment 1: ((a)) difference between the real room magnitude responses and the identified room magnitude responses for ((a)) $D = 1$ and for ((b)) $D = 256$ considering one microphone with reference to the Right channel. 67
3.18 Experiment 2: difference between the real room magnitude responses and the identified room magnitude responses for ((a)) $D = 1$ and for ((b)) $D = 256$ considering one microphone with reference to Left channel. 67
3.19 Experiment 3: difference between the real room magnitude responses and the identified room magnitude responses for ((a)) $D = 1$ and for ((b)) $D = 256$ considering one microphone with reference to Left channel. 67
3.20 Experiment 4: difference between the real room magnitude responses and the identified room magnitude responses for ((a)) $D = 1$ and for ((b)) $D = 256$ considering one microphone with reference to Left channel. 68
3.21 Experiment 1: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for sub-band identification (h) prototype response considering a sub-band identification, with respectively ((a)) Left channel and ((b)) Right channel with a smoothing factor of 1/12. 68
3.22 Experiment 2: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for sub-band identification (h) prototype response considering a sub-band identification, with respectively ((a)) Left channel and ((b)) Right channel with a smoothing factor of 1/12. 68
3.23 Experiment 3: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for sub-band identification (h) prototype response considering a sub-band identification, with respectively ((a)) Left channel and ((b)) Right channel with a smoothing factor of 1/12. 69
3.24 Experiment 4: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for sub-band identification (h) prototype response considering a sub-band identification, with respectively ((a)) Left channel and ((b)) Right channel with a smoothing factor of 1/12. 69

4.1 Different microphones and accelerometers placement inside the colonies, based on different approaches. In particular: on Figure 4.1(a) from [212], B is the microphone used and it is placed upper the queen cage. Figure 4.1(b) shows the microphone placement of [87]: the sensors are placed upon the brood frames inside a cage to protect them from propolisation. Figure 4.1(c) shows the solution adopted in [59], where a custom frame with the sensors inside has been chosen. Figure 4.1(d) refers to the approach proposed in [124], with microphones placed outside the hive, so that a cage to protect against propolisation is not necessary. On Figure 4.1(e), accelerometers placement proposed in [174] is presented: the sensors exploit vibrations and do not suffer from propolisation problems. Figure 4.1(f) belongs to approaches presented in [53, 54, 151, 200]: the microphones are hidden inside the hive walls, and a grid is used to protect them. Figure 4.1(g) from [180] shows a custom brood frame. Finally, Figure 4.1(h) from [175] shows accelerometers positioning similar to the one proposed in [174]. 80
4.2 The acquisition system installed in the university campus on figure 4.2(a), while the connection between the Bee board and the Queen board modules with the remote server is visible on figure 4.2(b). 81
4.3 Schematic of electrical connections for the sound acquisition. The signal of ADMP401 microphone is acquired by means of Behringer UCA222 USB audio card which is connected to the Raspberry Pi board on figure 4.3(a), while on figure 4.3(b) is visible the flow chart of the acquisition routine. 82
4.4 Placement of two microphones inside the hive, in particular microphone 1 is placed near the hive entrance, while microphone 2 is place near the hive brood. 82
4.5 Graphical user interfaces of the software developed for sound acquisition and preliminary analysis. In particular: the sound emitted by a selected hive is shown, in time (top screen) and frequency (bottom screen) domain, in arbitrary units (a.u.). 83
4.6 Schematic of spectrograms generation. 84
4.7 Schematic of mel spectrograms and MFCC algorithms, in particular if the output is saved before the logarithm the mel spectrograms obtained, otherwise the MFCC are obtained. 84
4.8 Schematic of Hilbert Huang transform algorithm. 85
4.9 Application of the HHT algorithm to a test signal. The input signal is a sine wave which change its frequency with time combined with a triangular wave. The EMD decompose the signal into its components plus a residue, then the Hilbert transform is applied to the obtained IMFs and the Hilbert spectrum of each basis function is obtained. 86
4.10 A 3-Level asymmetric dyadic analysis filter Bank. $G(z)$ and $H(z)$ are respectively the low-pass and the high-pass filters derived from the mother wavelet. 87
4.11 Schematic of the proposed convolutional neural network, containing four convolution layers each one followed by max-pooling layers. The last three layers are dense layers, with 256, 32, 1 units respectively. 89
4.12 Spectrogram of the sound generated in the beehive considering the two conditions. . 91
4.13 Features extraction exploiting mel frequency cepstral coefficients approach. 91
4.14 Features extraction exploiting Hilbert-Huang transform in terms of decomposition level (i.e., intrinsic mode function) representing a different spectral content. 92
4.15 Features extraction exploiting continuous wavelet transform with Morlet as mother wavelet.The number of scale values has been determined automatically in order to obtain a complete signal representation. 92
4.16 Features extraction exploiting discrete wavelet transform with a 10-level dyadic filter bank. The 11st level corresponds to the approximation coefficient. 92
4.19 Mel frequency cepstrum coefficients feature extractions considering (a) normal colony, and (b) a orphaned colony. 94

4.17 Spectrograms considering (a) normal colony, and (b) a orphaned colony. 94
4.18 Mel spectrograms considering (a) normal colony, and (b) a orphaned colony. 94
4.20 Hilbert-Huang transform feature extractions considering (a) normal colony, and (b) a orphaned colony. 95
4.21 Discrete wavelet transform feature extractions considering (a) normal colony, and (b) a orphaned colony. 95
4.22 Continuous wavelet transform feature extractions considering (a) normal colony, and (b) a orphaned colony. 95
4.23 The ⋆ represents the AUC score on the test set for each fold of the hive-independent setup and CNN experiment. The ○ represents the average AUC score on test sets over the two folds and the ● reports the average AUC score of the test sets across the two folds of the random split setup. 97
4.24 The ⋆ represents the AUC score on the test set for each fold of the hive-independent setup and SVM experiment. The ○ and ○ represent the average AUC score over the two folds in both train and test sets, respectively. The ● reports the average AUC score of the test sets across the two folds of the random split setup. 98
4.25 Measurement during the swarming event in colony 1 during a day in April 2019 considering sound variations (a) and the related weight variations (b). 101
4.26 Measurement during the swarming event in colony 2 during a day in June 2019 considering considering sound variations (a) and the related weight variations (b). . . 101
4.27 Spectrograms produced during swarming, considering the swarming of colony 1 (a) and colony 2 (b). 102
4.28 MFCC produced during swarming, considering the swarming of colony 1 (a) and colony 2 (b). 102
4.29 Second IMF during swarming, considering the swarming of colony 1 (a) and colony 2 (b). 103
4.30 Discrete wavelet coefficients during swarming, considering the swarming of colony 1 (a) and colony 2 (b). 103
4.31 Scalograms during swarming, considering the swarming of colony 1 (a) and colony 2 (b). 104

5.1 (a) Overall scheme of adaptive feedforward FXLMS-ANC systems and (b) its block implementation. 108
5.2 DSP board used for the real time implementation (a), and setup for real experiments in a semi-anechoic chamber(b). 110
5.3 Time domain behaviour of selected snooring signals. 111
5.4 Frequency domain behaviour of the selected snooring signals. 111
5.5 Obtained results of the ANC system in terms of MSE at the error microphone. . . . 111
5.6 Diagram of the proposed structure 112
5.7 Virtual Bass Enhancer structure 113
5.8 NLD structure - from method 1 described in [16] 113
5.9 VST plugin hosted in a DAW (Ableton) 115
5.10 Virtual bass results (a) in time and (b) frequency domain. 116
5.11 Stereo enhancer results (a) in time and (b) frequency domain. 117
5.12 Decorrelator results in time domain with (a) 0 iteration, (b) 100 iterations and (c) 1000 iterations. 117
5.13 Results for subjective listening tests taking into consideration the four attributes and the tracks of Table 5.3. In particular blue bars are related to spatial impression, orange bars are related to stereo impression, yellow bars are related to freedom from noise and distortion and grey bars are related to main impression. 119

5.14 Complete algorithm structure. The mono input of the guitar is fed to the decorrelator which transform the mono signal into a stereo signal, then the vibrato effect is applied to both the channel, finally the processed signal is mixed with the dry signal thanks to the dry/wet control **u**. 121
5.16 Spectrograms of the left channel (first figure) and right channel(second figure) considering (a) a superposition of two sine waves at 2kHz and 4kHz as input signal, (b) the results of the vibrato application, and (c) the final results of the whole algorithm. 122
5.15 (a) Main scheme of the vibrato effect, and (b) NU-Tech environment running the proposed algorithm. 122
5.17 Results of the listening tests evaluating (a) Sound Quality and (b) Impairment, with (A) the signal elaborated only by the vibrato effect and (B) by the full proposed algorithm. 125

List of Tables

2.1 Experiments Summary . . . 13
2.2 Variances used in the multiple-variance estimation method. . . . 15
2.3 Parameters of reduced Volterra models of Roma 27AC. . . . 17
2.4 Variances used in the multiple-variance estimation method. . . . 19
2.5 Parameters of reduced Volterra models of OTL509/25. . . . 20
2.6 Variances used in the multiple-variance estimation method. . . . 21
2.7 Parameters of reduced Volterra models of TubePre. . . . 23
2.8 Comparison between the number of kernel elements required by the complete Volterra model and by the reduced one, for the identified amplifiers. . . . 24
2.9 Number of operations required for a filter bank implementation of Volterra filters. . . 24
2.10 Mean percent difference between the proposed approach and IEC 60268 measurements, for each distortion parameter considered for synthetic experiments. . . . 43
2.11 Mean percent difference between the proposed approach and IEC 60268 measurements, for each distortion parameter considered for real devices experiments. . . . 46

3.1 (a) List of sound tracks used for the listening tests, and (b) bipolar scale values adopted for the subjective experiments. . . . 56
3.2 SD evaluation considering the single band identification ($D = 1$) and the sub-band identification ($D = 256$) for both experiments with a frequency range of $10 - 20000$ Hz. . . . 64
3.3 SD evaluation considering the single band identification ($D = 1$) and the sub-band identification ($D = 256$) for both experiments with a frequency range of $10 - 200$ Hz. 64

4.1 Summary of the state-of-the-art of approaches discussed in this chapter. . . . 77
4.2 Proposed Network Architecture with two different inputs: (a) MFCC coefficients with 20 frequency and 62 time frames and (b) a 256x256 pixels RGB images. . . . 88
4.3 AUCs comparison between the six feature extraction techniques. . . . 100
4.4 Precision, recall, F1 measures comparison for the no-queen-bee class. . . . 100
4.5 Precision, recall, F1 measures comparison for the queen-bee class. . . . 100

5.1 Algorithm parameters used for objective experiments. . . . 116
5.2 Algorithm parameters used for subjective experiments. . . . 118
5.3 List of sound tracks used for the listening tests. . . . 118
5.4 Proposed algorithm tunable parameters. . . . 122
5.5 Parameters values used for the subjective listening tests. . . . 125
5.6 Unipolar scales values used for subjective tests. . . . 125
5.7 List of sound tracks used for the listening tests. . . . 125

Chapter 1

Introduction

When we speak of sound systems, the first thought is related to audio amplifiers, loudspeakers and more in general to any device able to reproduce a sound experience. However, within this definition, it is possible to include many other natural and artificial systems, i.e., anything that can generate or modify a sound can be considered as a sound system.

Starting from this definition, any sound system can be studied applying digital signal processing (DSP) techniques. In particular, it is possible to analyze and measure its characteristics, and to find a mathematical model which approximates the real system behaviour and its structure. This model has several purposes, i.e., it can be used to study and better understand the real system, it can be used for the virtual emulation of the system or in conjunction with the real object to modify some of its characteristics.

In this work, different audio systems have been analyzed, modeled and identified with DSP approaches showing that these techniques can be applied in many different research fields to solve different problems. In particular, three main sound systems have been analyzed and will be discussed: the problem of measurement and identification of non linear audio devices, the problem of room response characterization and finally, the problem of honey bee sound analysis.
Considering non linear audio devices, the focus is on vacuum tube devices used as audio equipment which show a non linear behaviour with a particular attention on amplifiers and guitar pedals. For this class of systems, the derivation, of a complete mathematical model is possible and it can be used for many different applications such as distortion emulation, measurement and its compensation.
Taking into consideration the room environment, it represents another important audio system where sound is propagated and listened. The psychical characteristics of the environment can significantly affect the sound and its perception, but again, it is possible to measure and model this system. This modeling is obtained through the impulse response measurement that can be used to estimate its acoustical characteristic and used for an equalization process to compensate the negative effects introduced on the sound.
Honey bees colonies sound is the last discussed system. Even if it seems a quite peculiar choice, it is still a sound system with its own acoustical characteristics that can be measured and analyzed by means of DSP techniques.
Finally, some minor contributions will be presented showing that algorithms based on active noise control or psychoacoustic phenomena could be still used to improve the sound quality.

1.1 Thesis Contributions

The contributions of this thesis are focused on the three systems described above plus some minor contributions which are also related to the application of digital signal processing techniques to real audio systems.

- For the non linear audio systems the first approach involves an algorithm to identify a reliable non linear model based on Volterra series. In comparison with the previous known state of the art, the main goal has been the improvement of the experimental results, considering many different non linear audio devices. Moreover, an important aspect analyzed is the memory requirements for an implementation of the model on a DSP board. The other non linear model discussed has been analyzed focusing on computational cost. In this case the known non linear model has been improved using a sub-band structure. Finally for the distortion parameters estimation, the approach presented is totally innovative and developed thinking to the limitations of known procedures adopted in the international standards.

- For the room environment, two works are presented: the former exploits a multi-point equalization algorithm to improve the performance of vibrating transducers, in this case, the focus was an application to a real world scenario. The latter exploits sub-band structure in order to reduce the computational cost of a multi-point adaptive equalizer. In this case, the proposed work is an improvement of a previous known algorithm.

- For the honey bees sound analysis, several contributions are presented: an innovative sound acquisition system, has been developed trying to solve some of the critical problems of the other systems proposed in the state of the art. Indeed, many of the proposed systems has two main problems, i.e., a limited number of sensors and a limited acquisition time, after which the systems proposed were usually dismantled. The proposed system, has been developed considering many different type of sensors and has been built for a long term acquisition (until now over two years of continuous acquisition). Focusing on the acquired data, a complete analysis of the recorded sounds in different colony health situations is discussed. In particular, innovative features extraction procedures have been proposed and then different classifier algorithms have been used to analyze the sound and make a classification.

- Other minor contributions, are presented: an active noise control algorithm for snoring reduction, an algorithm capable of improving the sound performances of compact loudspeakers and an algorithm derived from the previous one and used as an electric guitar effect. These contributions are improvements of previous known algorithms, and have been developed focusing mainly on real world implementations.

1.2 Thesis outline

The aim of this thesis has already been discussed in this previous Section. The structure of this thesis is the following.

Chapter 2 Outlines the non linear audio systems. First, an introduction on non linear systems with particular attention to vacuum tube amplifier is given, then an overview on the non linear modeling techniques adopted is presented and finally the proposed approaches with experimental results are discussed.

Chapter 3 Discusses the problem of equalization and acoustical environments. First, an introduction on room impulse response behaviour and its measurement is given, then the multi-point equalizer adopted in the proposed approaches is presented, and finally the two contributions are presented and discussed with their own experimental results.

Chapter 4 The last chapter is dedicated to the analysis of the sound emitted by honey bees. First a deep analysis on the state of the art regarding honey bee sound analysis is presented, then the acquisition system is discussed. The next Section involves the explanation of the features extraction techniques and the classifier algorithms used for the sound analysis, and finally the two case studies analyzed are presented with their experimental results.

Chapter 5 Illustrates the other minor contributions:

- An active noise control system for snoring reduction.
- An algorithm for stereo reproduction enhancement.
- Application of decorrelation to a guitar vibrato effect.

Chapter 2

Non linear audio system identification, modeling and measurement

The main focus of this work is the application of different DSP techniques on non linear audio systems that represent one of the most critical aspect of real audio systems. Non linearity can influence negatively the sound reproduction but in the case of vacuum tube devices, it is used to add a particular effect to the sound reproduction. In the next chapter, several techniques for non linear systems modeling identification and measurements in terms of system distortion on real world audio devices will be presented. Several experiments have been carried on and results will be presented and discussed.

2.1 Background

In the last years, several different approaches have been investigated in order to develop a technique suitable for an efficient modelling of tube audio devices [163]. These approaches can be divided in three main categories: the white box approaches where the system structure is known (e.g., the schematic of an amplifier is available), the black box approaches where the system is totally unknown and only the inputs/outputs are available, the gray box approaches where the structure is not completely known but some features are available and could be used to improve the model. Typically, white box approaches are used to approximate the amplifier circuit gain stage exploiting differential equations [115] or the wave digital filters [72, 116], obtaining a complete mathematical model of real devices. On the other hand, the gray or black box approach are based on well known models, such as Hammerstein model [139, 184, 187], Wiener model [139, 184], Wiener Hammerstein model [75, 139], and the Volterra series, that are identified exploiting particular input signals (e.g., exponential sine sweep). Starting from the input/output characteristic and a block diagram of the analog circuit, an iterative parameter optimization has been also considered in [73, 93], to model a dynamic range compression system. To the gray box approaches belongs also the Significance Aware Filtering concept [102–104, 110] introduced in the field of non linear acoustic echo control, where the long memory non linear acoustic system is decomposed into synergetic subsystems, separating the long memory linear system representing the acoustic echo path from a non linear part composed by a very low number of parameters. An example of black box approach can be found in [152] where a Volterra series with kernels expressed as products of one-dimensional functions has been used for

modeling a surround amplifier.
Considering the non linear identification of other audio devices, several techniques have been applied for the identification of guitar pedal distortion effects [10, 74, 105, 163, 215] or non linear guitar speaker cabinet [186]. Regarding the pedal distortion emulation, a methodology based on Volterra series has been introduced in [10], exploiting its expansion to approximate the circuit non linear part, e.g., the asymmetric distortion produced by a triode valve or by a saturated operational amplifier. In a similar way, a parametric Wiener Hammerstein model exploiting an iterative algorithm has been used in [74] for guitar stompbox identification. In case of simple circuits such as the overdrive or distortion guitar pedal, approaches based on physical modeling have also been studied. In [105] the circuit is approximated exploiting a non linear state-space model, while in [215] the circuit elements are approximated by means of numerical solutions of ordinary differential equations. Taking into consideration the speaker cabinet of electric guitars, the Volterra kernels have been used for the cabinet approximations in [186] exploiting a logarithmic sine sweep technique, which provides a way to separate the different orders of Volterra kernels from the linear part. With relation to non linear digital effects, in [100] it is shown how the formalism of the Volterra series can be used to represent the non linear Moog ladder filter. In this case using a truncated version of Volterra series allows to develop a model for an extended range of input signals. Then, non linear system identification can be applied also to loudspeaker development: the non linear model can be used to design an equalization filter capable of reducing the device non linearity. In [19], a Volterra series has been applied to estimate the non linearity of a horn loudspeaker, while in [117] it is shown how to reduce the identified non linearity developing an equalization filter. A different technique is the non linear autoregressive moving average with exogenous input (NARMAX) model, originally proposed in [132], that has been used in [70, 71] to model loudspeaker non linearities. Taking into consideration the overall scenario of digital music, in [185] a method for band-limited discrete-time identification and modeling of analog non linear audio effects, like tube amps, exciters etc., has been presented using off-time digital cross correlation measurements. Finally considering more classical musical instruments, the Volterra series representation has also been used to obtain a model of non linear behavior of instruments such as an ocarina in [205] or a more conventional flute in [204].
However, it is not always necessary to have a complete model of the non-linear system: it may be useful to have numerical estimators available that give a synthetic and meaningful measure of the non-linearity degree of the device, whose values can be measured in a relatively simple and repeatable way. These estimators are also useful to verify the quality and stability of operation of non-linear system models. These evaluation parameters must be associated with input signals of defined characteristics. In addition, they must also be representative of the signals with which the system operates in reality. Sinusoidal signals, combinations of sinusoidal signals, random noise with a speech-shaped spectrum are among the signals used as inputs to identify quantities such as the Harmonic Distortion (HD), the Total Random Distortion (TRD), the Inter-Modulation Distortion (IMD) which, for different aspects, provide indications on the degree and type of non-linearity introduced by the considered system. Indeed, procedures for an efficient evaluation of these synthetic estimators of devices' non-linearity are essential to verify the quality of the models and to control their non-linear behavior. Among the synthetic estimators, the total harmonic distortion (THD), the Harmonic Distortion of order k (HD_k) and the IMD are the most widely used for non-linearity estimation [177, 209]. Different measurement techniques have been defined for both HD and IMD,

based on the use of gliding tones (linear chirp signals), or on step-by-step sinusoidal tones, choosing the speed of variation (in the former case) and the frequencies and duration of the tones to obtain the necessary measurement accuracy, and measurement times compatible with practical needs. For IMD, techniques have been defined that consider a fixed tone and a variable tone (modulation distortion of the second and third order) as input signals, or techniques that consider both the tones variable (difference frequency distortion).

2.2 Example of non linear audio systems

The non linear audio devices discussed in this work are vacuum tube devices. These audio systems, are still widely adopted in many fields such as hi-fi, studio equipment and musical instruments (guitar amplifiers and effects). In the next Section, a brief introduction on vacuum tube technology will be given.

Vacuum tube devices

Although semiconductor technologies have displaced vacuum tube devices in nearly all fields of electronics, vacuum tubes are still widely used in professional audio amplifiers. In comparison with solid state components, vacuum tube suffers of many problems like a considerable background noise and a remarkable non linear behaviour which tends to produce distortion. However, for many professional audio devices like guitar amplifiers or hi-end amplifiers, vacuum tubes are still one of the most suitable choice. The reason is related to the type of distortion produced by vacuum tube which is different from the one produced by solid state components, and human ear tends to prefer the vacuum tube one. Vacuum tubes, or thermionic valves, were invented in the early 1900s for amplifying low-level voltage signals. The structure of vacuum tubes consists of two or more electrodes inside a glass or a metal enclosure. One of the most used tube is the triode valve, of which the schematic is visible on Figure 2.1

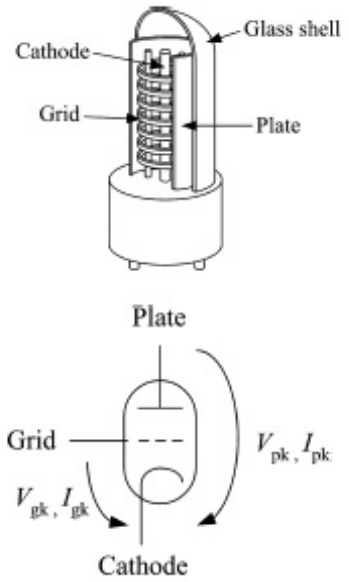


Figure 2.1: Physical construction and electrical schematic of triode valve, image from [163]

On a real working tube, the cathode is heated with a filament and starts to emit electrons thanks to the thermoionic effect. When a potential difference is applied between the heated cathode and the anode (or plate), a flow of electrons is established between the two terminals. Grid is placed between anode and cathode and it allows a control on the flux of electrons, just like the base of a BJT or the gate of a MOSFET. The Vpk applied between anode and cathode could reach hundreds of volts and in comparison with solid state transistors, this is one of the most critical aspects of vacuum tube. While triodes are the most common valve for the preamplifers and the tone Section in many amplifiers, for the power stage a more common choices are the tetrodes and pentodes. These tubes have respectively four and five terminals, the additional terminals are essentially other grids which improves the performances of the valve, reducing the distortion and the capacitance between the anode and the control grid. The anode-to-cathode current Ipk is a non linear function of both the grid to cathode and anode to cathode voltages: $Ipk = f(Vgk, Vpk)$. For this reason a change in the voltage of the grid causes a change in current flow between the cathode and anode. Amplification occurs when the change in current is converted to a change in voltage by a large valued load resistor. Although amplification is nominally linear around a central operating voltage known as the *bias*, at extreme signal levels, the amplified output will saturate. When the anode to cathode voltage Vgk is very small, current flow cut-off sharply. Very large Vgk causes the plate voltage to approach that of the cathode again, limiting the current and resulting in a non linearly saturating characteristic. To find the full non linear transfer characteristic from input to output requires the solution of a non linear system of implicit equations, because in a typical amplifier circuit, Vpk depends on Ipk and vice-versa.

Vacuum tube amplifiers

Once the basic principles of vacuum tube are given, in Figure 2.2 a schematic of the principal component of vacuum tube amplifier are presented.

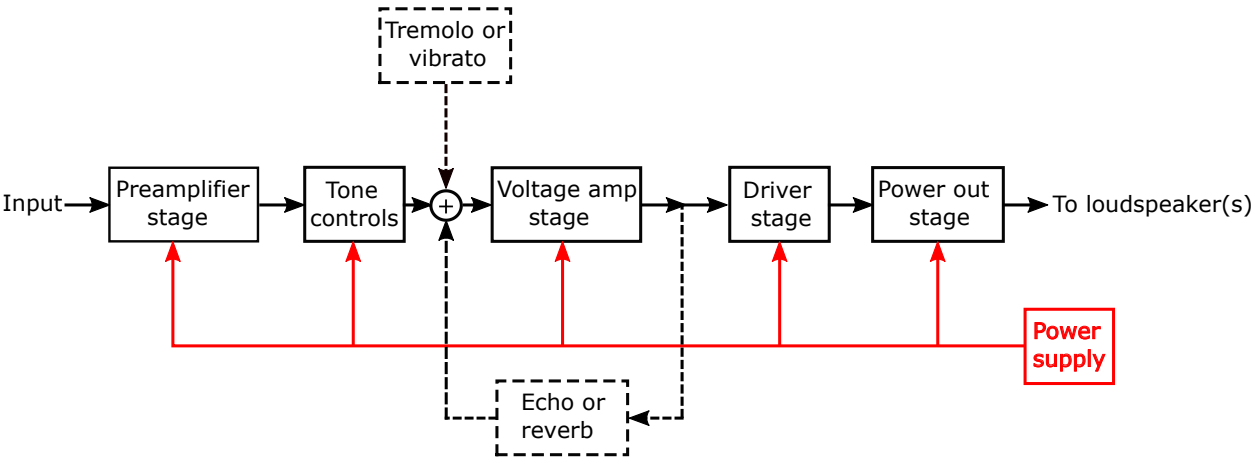


Figure 2.2: Basic schematic of a vacuum tube amplifier. Red lines represent the power supply, while dashed lines represent some environment effect which are present only on guitar amplifiers, image adapted from [64]

First, the signal is processed by the preamplifier Section. This Section has two main task: give a preliminary gain to the signal and making an adaptation of the impedance (this is critical especially

with high impedance signal like the guitar pickups signal). Then, a tone control is present, this Section could be both active (i.e., it could cut or boost the frequencies) or passive (i.e., it could cut the frequencies) some guitar and hi-fi amps don't have it. Then, the voltage-amplifier stage provides the voltage gain and also give the full output voltage swing. The driver stage amplifier role is to provide adequate input power, to drive the power amplifier. Finally, the power amplifier generates enough power to drive the loudspeakers, it is generally based on pentode tubes and it is one of the most critical parts since it requires a perfect impedance match with the loudspeakers to avoid damage on the tubes. For this reason many amplifiers adopt a transformer to adapt the tube impedance to the loudspeaker impedance. Dashed blocks represent some environment effects which are present only in guitar amplifiers and are not present in hi-fi devices, these effects could be implemented mechanically inside the amplifier (e.g., a with a spring based reverb) or with digital algorithms, in particular vibrato effect is deeply analyzed and discussed on chapter 5.3.

2.3 Non linear system modeling

Once the general characteristic of the analyzed systems is given, on the next Section the structures used in the modeling process will be presented. These models allow the mathematical representation of many different real word audio devices, with different accuracy and computational costs.

2.3.1 LIP non linear filters

The family of linear in parameters (LIP) non linear filters is a wide family of filters which can be used to approximate arbitrarily well any any discrete-time, time-invariant, finite-memory, causal, continuous non linear system as the one of Eq. 2.1.

$$y(n) = f[x(n), x(n-1), ..., x(n-N+1)] \tag{2.1}$$

Where $y(n)$ is the system output, $x(n)$ in the input and N is the system memory. The name of these filters is related to the fact that the parameters used to adapt the filters to the system are linear, but it must underline the fact that the output of this model is non linear just like the real system that they model. The function f in Eq. 2.1 can be interpreted as a multidimensional function over the space $\mathcal{R}_1^N$. Each delayed input sample corresponds to a different dimension. Taking into account this notation, it is possible to expand the function f in Eq. 2.1 with a series of basis functions fi according to:

$$f[x(n), x(n-1), ..., x(n-N+1)] = \sum_{i=1}^{\infty} c_i f_i[x(n), x(n-1), ..., x(n-N+1)], (2.2) \tag{2.2}$$

where $c_i \in \mathcal{R}$ and each f_i is a continuous function from $\mathcal{R}_1^N$ to $\mathcal{R}$. Every choice of the set of basis functions f_i leads to a different family of LIP non linear filters, which can be used to identify or to approximate the non linear systems in Eq. 2.1.

2.3.2 Volterra series

One of the most known and used families of LIP filters are the Volterra filters. To understand these filters, the first step is the introduction of the double truncated Volterra series that model a discrete time, non linear, finite memory system

$$y(n) = h_0 + \sum_{i=1}^{P} (H_i x)(n), \tag{2.3}$$

$$(H_i x)(n) = \sum_{\tau_1=0}^{M} \sum_{\tau_2=\tau_1}^{M} \cdots \sum_{\tau_i=\tau_{i-1}}^{M} h_i(\tau_1, , \tau_i) \prod_{j=1}^{i} x(n-\tau_j),$$

where $y(n)$ is the output of the system, $x(n)$ its input, and $h_i(\tau_1, , \tau_i)$ is the Volterra kernel of order i. The output $y(n)$ is a linear combination of product of delayed input samples, where the delays τ_j are also the indexes of the kernel matrices $h_i(\tau_1, , \tau_i)$. In order to allow system identification by means of a cross-correlation method the Volterra series must be rearranged in terms of orthogonal operators:

$$y(n) = \sum_{i} (H_i x)(n) \equiv \sum_{i} (G_i x)(n). \tag{2.4}$$

These operators, commonly known as Wiener G–functionals, are orthogonal to each other and also to Volterra operators of lower orders if the input $x(n)$ is a stationary white noise with zero mean and variance σ_x^2. The Wiener G–functionals, unlike the Volterra ones, are non-homogeneous functionals, being linear combination of polynomial terms of different orders. The Wiener kernels can be identified by means of the multiple variance method with the formulas presented in [159] and will be discussed in Section 2.4.

2.3.3 Wiener-Hammerstein model

A particular subclass of LIP filters are the Wiener-Hammerstein non linear models. Despite these three models (i.e., the Wiener, the Hammerstein and the Wiener-Hammerstein model) are have a very simple and easy to understand structure, they have a significant limitation, i.e., they cannot represent systems with memory. However, even with this limitation, many real system can be still represented with these models, moreover in comparison with Volterra filters they have a reduced computational cost. These structures are visible on Figure 2.3.

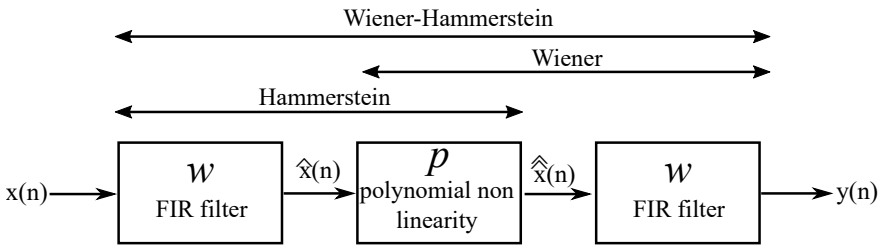


Figure 2.3: Wiener-Hammerstein models for non linear system representation.

In particular, as visible on Figure, the Hammerstein model is composed of a FIR filter fallowed by a static non linearity, the Wiener model is composed of a static non linearity fallowed by a

FIR filter and finally the Wiener Hammerstein model involves two FIR filters, one placed before the static non linearity and the other placed after. The static non-linearity is often chosen as a polynomial (power series), i.e.,

$$\hat{x}(n) = p_1 x(n) + p_2 x(n)^2 + ... + p_M x(n)^M, \tag{2.5}$$

where M and p_m represent the non-linearity order and the m-th order coefficient respectively. The non-linearity input vector is defined as $x_p(n) = [x(n), x^2(n), ..., x^M(n)]^T$, so, for the Hammerstein model, an input matrix can be characterized as:

$$X(n) = [x_p(n), x_p(n-1), ..., x_p(n-L+1)]^T, \tag{2.6}$$

with L denoting the memory size of the FIR filter. The input-output relationship of the complete Hammerstein filter is:

$$y(n) = w^T X(n)^T p, \tag{2.7}$$

where $w = [w_0, w_1, ..., w_{N-1}]^T$ and $p = [p_1, p_2, ..., p_M]^T$ are the linear and non-linear coefficient vector respectively. It must underline that, the Hammerstein model has been taken as example, but the same mathematical expression can be adapted easily to both the Wiener model, and the Wiener-Hammerstein model. The choice of one structure over the others is strictly related to the real system which has to be modeled and the right choice has a strong influence on the identification process and on the accuracy of the results. For this reason, a priori knowledge of the real system could be useful for making the right decision, but since this is not always possible some techniques are also available in literature for selecting the right model, e.g., in [172] where to select the model is proposed an approach based on the observation of the frequency response of the real system.

2.4 Non linear system identification

In the next Section, several innovative approaches for the identification of the models presented in Section 2.2 will be discussed. Experimental results on real word scenario will be discussed making several comparison between the proposed approaches and the state of the art.

2.4.1 Multivariance Volterra series identification

One of the main drawbacks when a non linear system is identified with Volterra series is called the locality of the solution, i.e., the solution obtained is accurate only for input signal with an amplitude similar to the one used for the identification process.The Volterra series, tested with different power inputs, shows an increasing error as the input power increases or decreases with respect to that used in the identification. In [159] the author shows a strong dependence between the power of the white noise (σ_x^2), used to identify a Volterra series, and the approximation capability of model itself, independently from the algorithm used in the identification. This happens because several errors affecting the identification of the Volterra series depend on the input power. The cause of these errors can be related to the input non idealities and to the approximation errors due to the double truncation of the series, with respect to memory and order of non linearity [161]. In traditional algorithms, using high σ_x inputs has the advantage of stimulating higher order non linearities, in

order to achieve more accurate higher order kernel identification. As a drawback, the use of high σ_x values causes high identification errors in lower order kernels, since the higher order Wiener kernels, multiplied by the variances of lower order kernels, are summed to the lower order Volterra kernels [159]. The rationale should be to use a low σ_x value for the lower order kernels and to gradually increase it for higher order ones. This is the core of the multiple-variance method, where inputs with different variances are used for Wiener kernel identification, allowing to reach a lower error with respect the traditional identification for inputs with wide dynamics. The Wiener kernels explained in Section 2.3.2 can be identified by means of the multiple variance method with the formulas presented in [159], that are explicitly reported in the following up to the fourth order,

$$k_0^{(0)} = E[y^{(0)}(n)], \tag{2.8}$$

$$k_1^{(1)}(\tau_1) = \frac{1}{A_1} E\left[y^{(1)}(n)\, x^{(1)}(n-\tau_1)\right] , \tag{2.9}$$

$$\begin{aligned} k_2^{(2)}(\tau_1,\tau_2) \;&= \tfrac{1}{2!A_2^2}\left\{E\left[y^{(2)}(n)\,\textstyle\prod_{i=1}^{2} x^{(2)}(n-\tau_i)\right]\right. \\ &\quad \left. -A_2 k_0^{(2)}\delta_{\tau_1\tau_2}\right\} , \end{aligned} \tag{2.10}$$

$$\begin{aligned} k_3^{(3)}(\tau_1,\tau_2,\tau_3) \;&= \tfrac{1}{3!A_3^3}\left\{E\left[y^{(3)}(n)\,\textstyle\prod_{i=1}^{3} x^{(3)}(n-\tau_i)\right]\right. \\ &\quad - A_3^2\left[k_1^{(3)}(\tau_1)\delta_{\tau_2\tau_3} + k_1^{(3)}(\tau_2)\delta_{\tau_1\tau_3}\right. \\ &\quad \left.\left. + k_1^{(3)}(\tau_3)\delta_{\tau_1\tau_2}\right]\right\} , \end{aligned} \tag{2.11}$$

$$\begin{aligned} k_4^{(4)}(\tau_1,\tau_2,\tau_3,\tau_3) &= \tfrac{1}{4!A_4^4}\left\{E\left[y^{(4)}(n)\,\textstyle\prod_{i=1}^{4} x^{(4)}(n-\tau_i)\right]\right. \\ &\quad - A_4^2\left[k_2^{(4)}(\tau_1,\tau_2)\delta_{\tau_3\tau_4} + k_2^{(4)}(\tau_1,\tau_3)\delta_{\tau_2\tau_4} + \right. \\ &\quad k_2^{(4)}(\tau_1,\tau_4)\delta_{\tau_2\tau_3} + k_2^{(4)}(\tau_2,\tau_3)\delta_{\tau_1\tau_4} \\ &\quad \left.\left. k_2^{(4)}(\tau_2,\tau_4)\delta_{\tau_1\tau_3} + k_2^{(4)}(\tau_3,\tau_4)\delta_{\tau_1\tau_2}\right]\right\} , \end{aligned} \tag{2.12}$$

where $k_p^{(j)}$ and $y^{(j)}(n)$ are respectively the Wiener kernel of order p and the system output obtained with the input $x^{(j)}(n)$ of variance $A_j = \sigma^2_{x^{(j)}}$. In the cross-correlation method the expectations in (2.8)-(2.12) are estimated using time averages over the entire duration of the input and output signal.

Wiener kernels can now be rearranged to obtain directly the expression of the Volterra kernels and so the Volterra series. Appropriate normalization is needed in Wiener to Volterra conversion formulas, unlike in the usual ones, for taking into account the use of multiple variances. The new formulas for the changeover from Wiener to Volterra are reported in the following equations,

$$h_4 = k_4^{(4)}, \tag{2.13}$$

$$h_3 = k_3^{(3)}, \tag{2.14}$$

$$h_2 = k_2^{(2)} - 6A_2 \sum_{\tau_3=0}^{M-1} k_4^{(4)}(\tau_1,\tau_2,\tau_3,\tau_3), \tag{2.15}$$

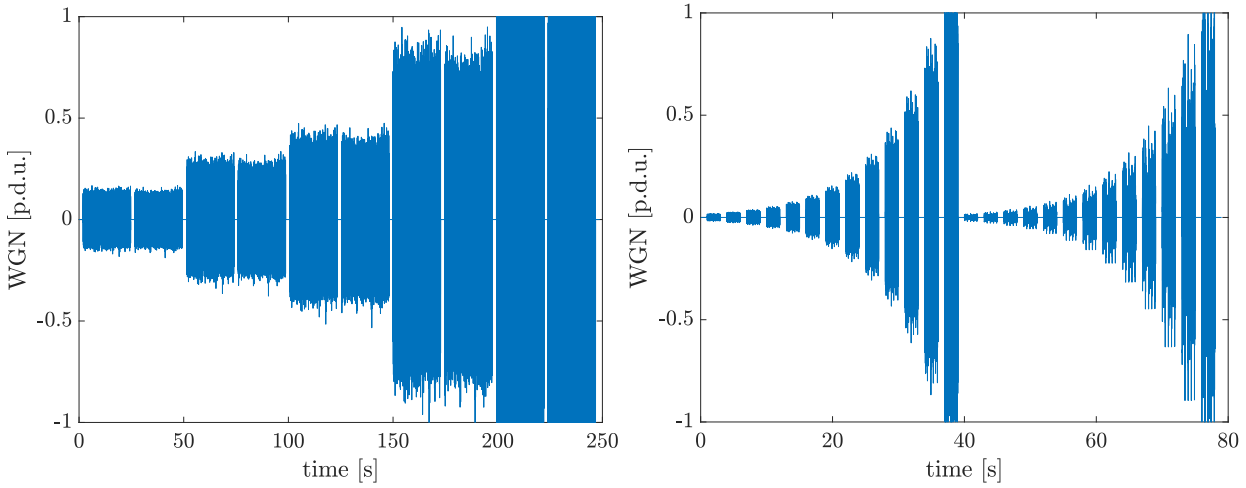


Figure 2.4: Input signals used (a) for the identification (b) for testing, expressed in procedure defined units (p.d.u.) [157].

$$h_1 = k_1^{(1)} - 3A_1 \sum_{\tau_2=0}^{M-1} k_3^{(3)}(\tau_1, \tau_2, \tau_2), \tag{2.16}$$

$$\begin{aligned} h_0 = \quad & k_0^{(0)} - A_0 \textstyle\sum_{\tau_1=0}^{M-1} k_2^{(2)}(\tau_1, \tau_1) \\ & +3A_0^2 \textstyle\sum_{\tau_1=0}^{M-1} \sum_{\tau_2=0}^{M-1} k_4^{(4)}(\tau_1, \tau_1, \tau_2, \tau_2), \end{aligned} \tag{2.17}$$

where all kernels have memory equal to M, and the system has a maximum order $P = 3$.

The following formula

$$N = \binom{M+P}{P}, \tag{2.18}$$

gives the total number of coefficients of a Wiener or Volterra truncated series, when all kernels have the same memory. Equation (2.18) provides the well known exponential relationship, used by all papers in support of the curse of dimensionality of Volterra series.

The curse can be broken by simply avoiding the kernels to have the same memory, and above all avoiding to have the same memory of the first order one. A procedure for estimating the kernels of non linear audio devices with a symmetric distribution of the coefficients around the main peak, as in audio amplifiers, is discussed in the following. The procedure could be easily modified to address also asymmetric distributions and it consists of the following parts:

- The identification of the first order kernel is carried out using (2.8) with a large memory, since it is not computationally expensive. Then M_1 is chosen by discarding all negligible kernel values, i.e., those under the noise floor.
- From the first order impulse response the delay of the main peak, P_1, is estimated.
- Higher order kernels are identified with (2.8)–(2.12), starting by kernels with an memory M_p equal to a fraction of M_1 and centered at P_p equal to P_1.
- M_p can be now increased (or, if needed, reduced) by a steepest descent procedure, until the performance of the Volterra model is not satisfactory.

Table 2.1: Experiments Summary

Device Type	Number of input variances		
	Identification	Test	
	Noise	Noise	Music
Synthesis	4	13	13
SI Audio	5	13	13
Presonus	4	13	13

With this procedure the delay of the first kernel value different from zero is

$$D_p = P_p - M_P/2 \ . \tag{2.19}$$

The delay D_p and the memory M_p changes for each kernel. In the experimental part it will show that with the proposed technique it is possible to obtain a complexity reduction of one order of magnitude for a third order system and of two orders of magnitude for a fourth order system.

Experimental results

The proposed methodology has been tested with three different vacuum-tube amplifiers, considering the following two experiments for each devices:

1. Identification and test of one complete multiple-variance and four complete single-variance Volterra models in order to prove that the multiple-variance model performs better, for a wide range of input powers, than the single-variance models;

2. Identification and test of reduced-kernel models to prove that the reduced models perform comparably to the complete ones and that also in this case the multiple-variance performs better than the single-variance ones.

Regarding the measurement two different setups have been considered, i.e., the first for the two tube amplifiers and the latter for the tube preamplifier. Focusing on the first setup both the amplifiers have been connected to a pairs of Auna 501 passive 2-way speakers with a RMS rated power of 100 W, an impedance of 8 Ω, and a frequency range of 92 Hz – 20 kHz. The volume has been set at the highest value and after half an hour of warm up, the input signal has been applied with a 6 dB of attenuation of the D/A converter, and the output signal has been measured across the speaker terminals. Even though these two amplifiers are based on tube technology, they show a different circuit: the SI Audio amplifier is an all triode OTL Amplifier (output transformerless), while the Synthesis amplifier shows a more conventional architecture, with an output transformer. In the second setup, a microphone tube preamplifier, i.e., Presonus TubePre, has been used connecting it in loop with the measurement setup with an unbalanced cable. The TubePre uses a triode 12AX7 valve, and it allows two main controls: a drive potentiometer which acts on the valve saturation and a gain control which acts on the overall gain of the preamplifier adding more volume and distortion. The drive has been set at 4dB (within a range from 0 to 20dB), the gain at 12dB (within a range from 3 to 48dB) and the A/D converter attenuation of 20 dB; this setting has low values of drive and gain, in order to avoid a too heavy distortion. For all measurement setup, a National Instruments Compact Rio chassis (cRIO-9024) equipped with a 2-channel voltage analog output NI-9260 and

with a 3-channel voltage analog input NI-9232, connected to a desktop PC have been considered. Both generation and acquisition board have been configured to use the same sample clock with a sample rate of 44100 Hz. The National Instruments software LabView was used to generate the test signals and to acquire the measurements. Two types of sound tracks have been considered for the measurement. The first, used for the identification procedure, is composed by the concatenation of white noise signals with different power values, interleaved with a 1 s of silence, as shown in Figure 2.4(a). The second, used for testing the identified model, is a combination of music and white noise, as shown in Figure 2.4(b).In order to exploit all the resolution of the D/A converter, used to drive the amplifier, the maximum variance in the identification signal was set equal to 1/12, $\sigma_5^2 = A_5 = 1/12$. With this value only a negligible percentage (0.05%) of noise values saturates the D/A converter, exceeding 1. In the following, the non linear devices are modeled with third order or fourth order Volterra systems. Thus, five different input variances were considered to allow the identification of each kernel of a fourth order system at a different input variance. A subset of these input variances will be used for the identification of third order systems.

The set of variances of the five white noises $\{\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2, \}$ was chosen equal to

$$\{A_5/64,\ A_5/16,\ A_5/8,\ A_5/2,\ A_5\}. \tag{2.20}$$

Accordingly, the corresponding numeric values of standard deviations $\{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \}$ are

$$\{0.036,\ 0.0722,\ 0.102,\ 0.204,\ 0.288\}. \tag{2.21}$$

To test the model performance, a normalized mean square error (NMSE), defined over the output spectrogram has been used. To define it formally, let us write

$$Y_{\mathrm{A}}(\tau, \omega) = \mathrm{STFT}[y_{\mathrm{A}}(n)] \tag{2.22}$$

$$Y_{\mathrm{M}}(\tau, \omega) = \mathrm{STFT}[y_{\mathrm{M}}(n)] \tag{2.23}$$

where $y_{\mathrm{A}}(n)$ and $y_{\mathrm{M}}(n)$ are the output of the amplifier and of the model, respectively, and the STFT is the Short Time Fourier Transform, calculated using a Hamming window such that the signal is divided into eight segments with 50% overlapping samples. The mean square error between two matrices can be defined as

$$\mathrm{MSE}[Y_{\mathrm{A}}, Y_{\mathrm{M}}] = \frac{1}{RC} \sum_i \sum_j [Y_{\mathrm{A}}(i,j) - Y_{\mathrm{M}}(i,j)]^2 \tag{2.24}$$

where RC is the product between the number of rows and columns of one matrix. Now we are able to define the NMSE as

$$\mathrm{NMSE} = \mathrm{MSE}(|Y_{\mathrm{A}}|, |Y_{\mathrm{M}}|)/\mathrm{MSE}(|Y_{\mathrm{A}}|, 0), \tag{2.25}$$

where $|.|$ denotes the magnitude applied to each element of the matrix. The use of MSE in the frequency domain allows a comparison of the two signals from the harmonic content point of view.

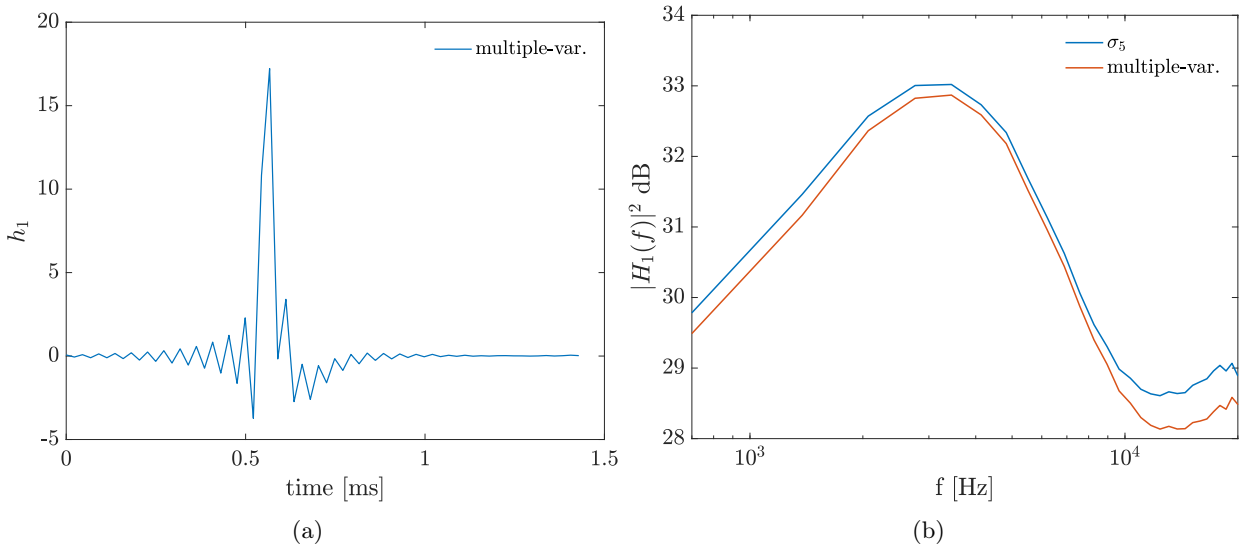


Figure 2.5: Roma 27AC: first order Volterra kernel, (a) in time domain ($h_1(\tau)$) (b) in frequency domain ($|H_1(f)|^2$).

Table 2.2: Variances used in the multiple-variance estimation method.

kernel	input variance
k_0	σ_1^2
k_1	σ_2^2
k_2	σ_4^2
k_3	σ_5^2

Synthesis Roma 27AC identification results In the first experiment the Roma 27AC vacuum-tube amplifier has been identified. Four Volterra models were identified with the cross-correlation algorithm, each model using an input with a different variance belonging to the set in (2.20). Then, a single Volterra model was identified with the multiple-variance method [159], estimating each Wiener kernel at a different variance as reported in Tab. 2.2.

The memory spans of the kernels of all these Volterra models were

$$\begin{aligned} M_1 &= 64, \\ M_2 &= 32, \\ M_3 &= 32, \end{aligned} \tag{2.26}$$

where M_j is the memory of the Wiener and Volterra kernels of order j.

Figure 2.5 shows the first order kernel of the multiple-variance Volterra model, in time (Figure 2.5(a)) and frequency domain (Figure 2.5(b)) respectively. The range of time values in the x-axis of Figure 2.5(a) corresponds to the memory value, $M_1 = 64$, multiplied for the sampling time $\tau_{\max} = 64/44100 = 1.45$ ms. Figure 2.5(b) also shows the power spectrum of the first order kernel of the single-variance Volterra model, identified with the maximum variance, σ_5^2.

Figure 2.6 shows the NMSE of the single-variance Volterra models and that of the multiple-variance Volterra model, as a function of input standard deviation. The Figure was obtained by using as testing input of the Volterra models each of the white noise signals belonging to the testing

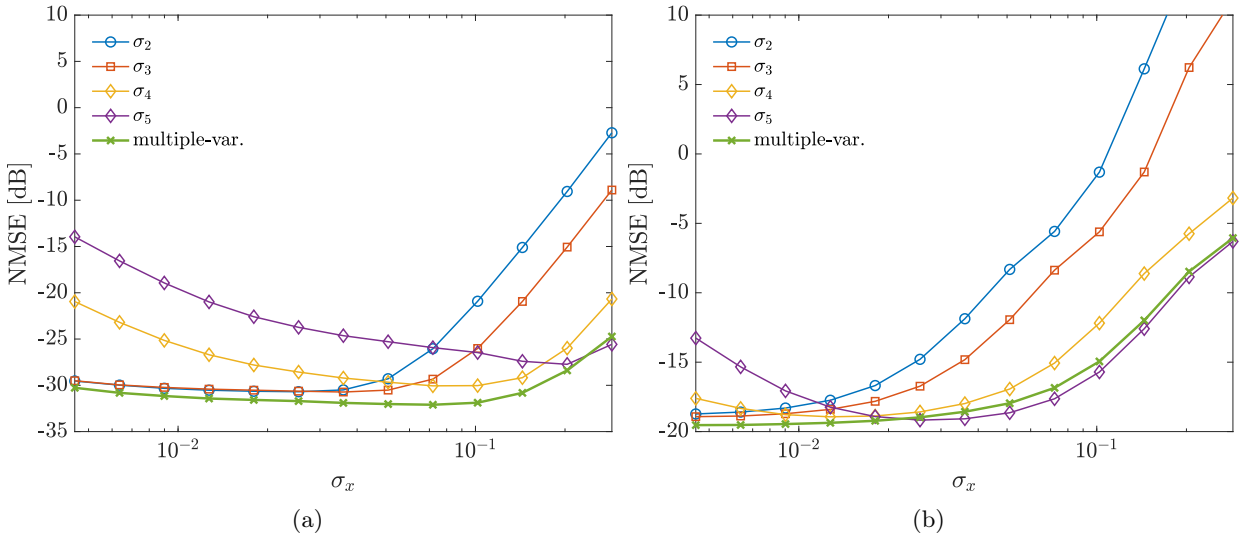


Figure 2.6: NMSE of complete Volterra models of Roma 27AC. (a) WGN input, (b) music input.

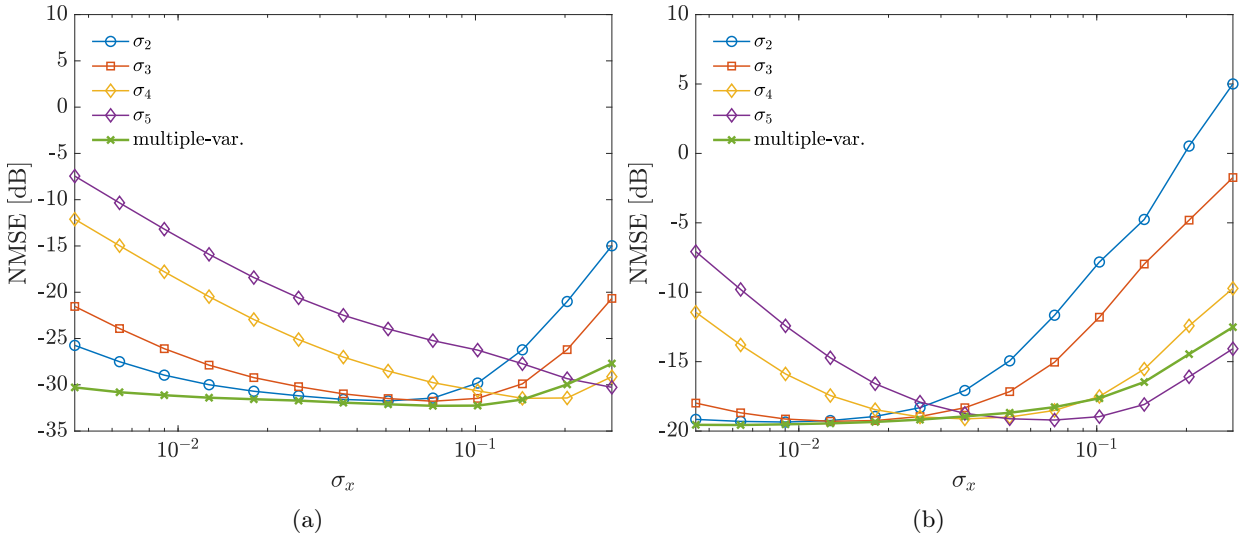


Figure 2.7: NMSE of reduced Volterra models of Roma 27AC. (a) WGN input, (b) music input.

set shown in Figure 2.4(b), and calculating the NMSE for each of them. In the Figure you can count 13 NMSE points for each curve, one for each of the white noise input signals.

In Figure 2.6(a) we can see that the error curve related to the multiple-variance method interpolates the inferior values of the curves for single-variances, as the multiple-variance method was meant to do.

The single-variance Volterra models have performance similarly to the multiple-variance one only in a short interval around the variance used for the measurement. Outside this interval the error increases up to 15 dB. It should be noticed that the proposed model has a NMSE always less that -25 dB.

Figure 2.6(b) shows similar curves, but obtained with the music inputs reported in Figure 2.4(b). Also in this Figure the multiple-variance curve approaches the inferior values of the curves for single-variances, although the single-variance model identified with σ_5 goes better for some inputs, reaching the best performance difference of -0.79 dB, but for low variances it performs worse until 6.3 dB.

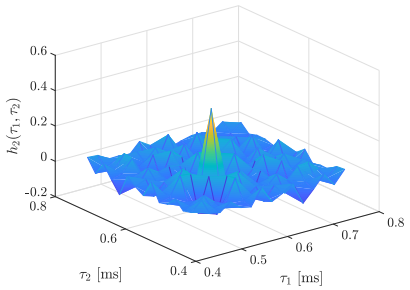


Figure 2.8: Roma 27AC: second order Volterra kernel, $h_2(\tau_1, \tau_2)$.

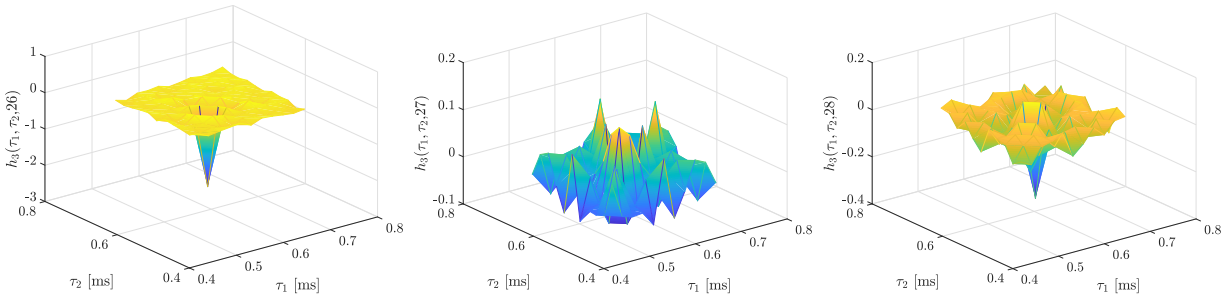


Figure 2.9: Roma 27AC: third order Volterra kernel $h_3(\tau_1, \tau_2, \tau_3)$.

Each of the Volterra models with the kernel memories reported in (2.26) has a total of 6577 kernel elements. This can make very difficult their real-time implementation. With the aim of reducing the complexity of Volterra models, the memory extent of the kernel has been reduced, keeping a reduced number of kernel values around the peak of the same kernel. The reduced Volterra models, that has been found to have comparable performance to the complete model, have the parameters reported in Tab. 2.3, where the D_j are kernel delays, as defined in (5.18). The reduced models have now just 454 kernel elements and can thus be easily implemented in real-time.

Figure 2.7(a) shows the performance of the reduced-memory Volterra models, obtained using white noise inputs. In comparison with Figure 2.6(a), it is possible to see an improvement of single-variance model performance for high input variances and a worsening for small input variances. This can be explained by considering that, in the reduce kernel models, even though all the neglected elements of higher order kernels could contain some missing information about the kernel, in most cases they contain identification noise. This higher order kernel noise becomes relevant for high power inputs (so avoiding it improves the performance), but it is less relevant for low variance inputs, where the missing information prevail. The multiple-variance method, having much less identification error, is much less affected from the kernel truncation, maintaining the same performance at low input variance and performing a little better only for high input variances.

Table 2.3: Parameters of reduced Volterra models of Roma 27AC.

memory	delay
$M_1 = 64$	$D_1 = 0$
$M_2 = 15$	$D_2 = 19$
$M_3 = 11$	$D_3 = 21$

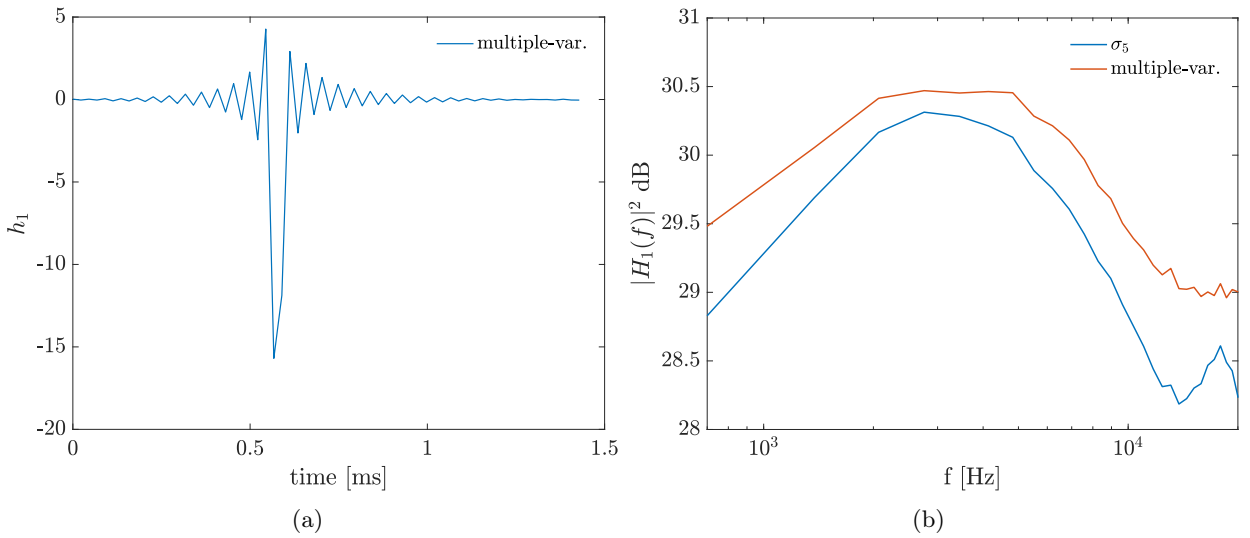


Figure 2.10: OTL509/25: first order Volterra kernel (a) in time domain ($h_1(\tau)$) (b) in frequency domain ($|H_1(f)|^2$).

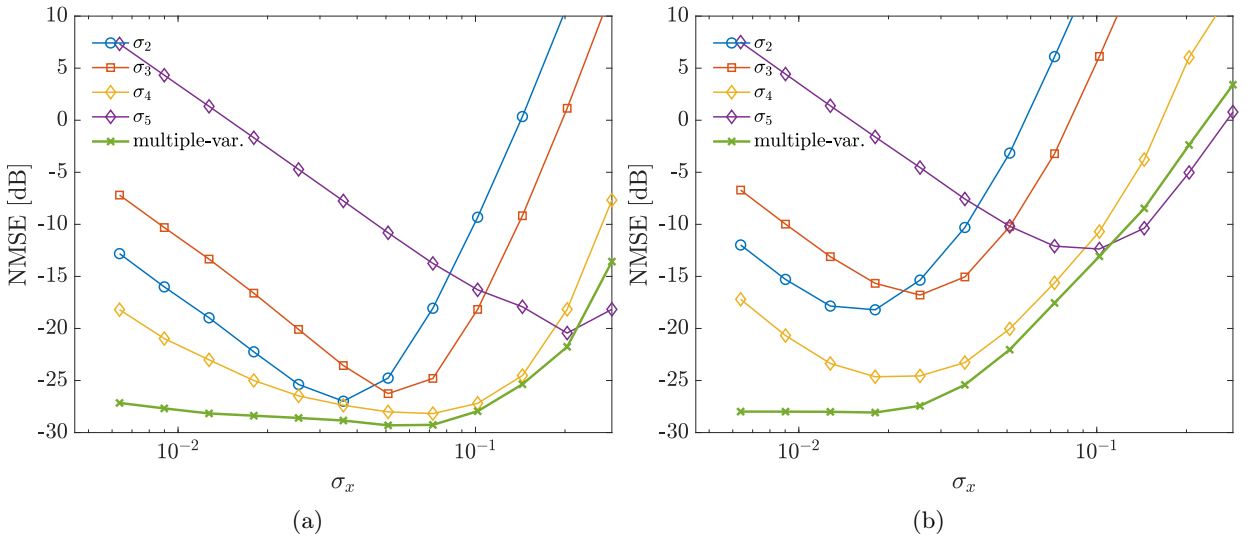


Figure 2.11: NMSE of complete Volterra models of OTL509/25. (a) WGN input, (b) music input.

Similar results are also shown in Figure 2.7(b) that reports the NMSE obtained with music input. Here, the single-variance model identified with σ_5, achieved a better performance difference of 1.6 dB, with respect the multiple-variance one, but with an overall worse performance difference of 12.5 dB.

Figure 2.8 shows the values of the second-order Volterra kernel of the multiple-variance method that have been estimated in the reduced model. It can be noticed that the kernel elements shown in the Figure start from 4.3 ms, corresponding to $D_2 = 19$ samples.

Figure 2.9 shows three slices of the third order kernel of the multiple-variance Volterra model. The third order kernel of the reduced model has a delay similar to the second order one ($D_3 = 21$ samples), but a smaller memory, $M_3 = 11$.

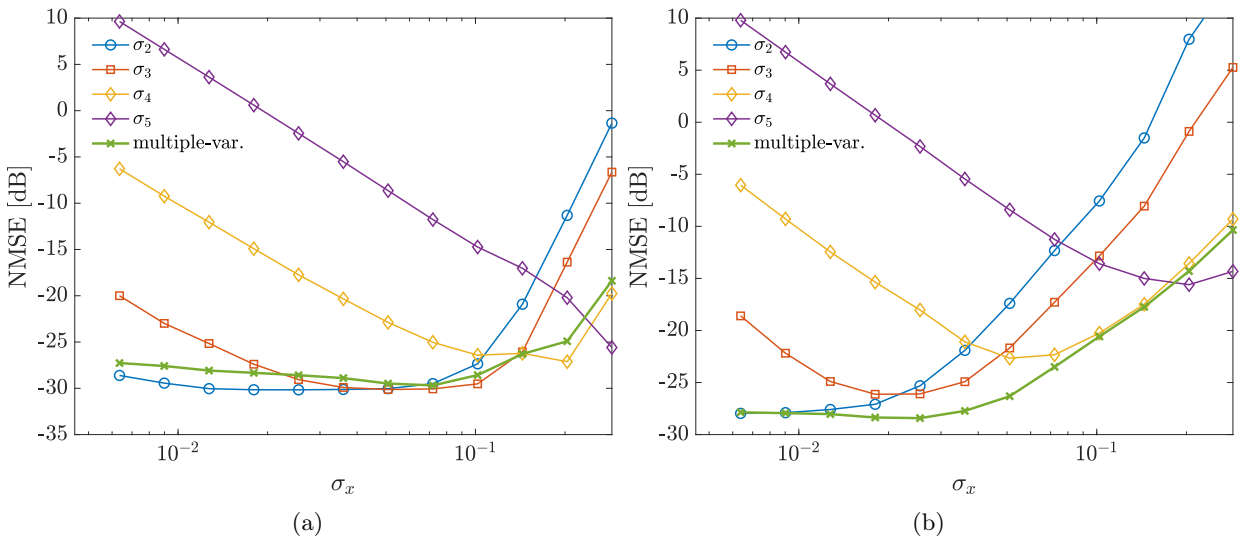


Figure 2.12: NMSE of reduced Volterra models of OTL509/25. (a) WGN input, (b) music input.

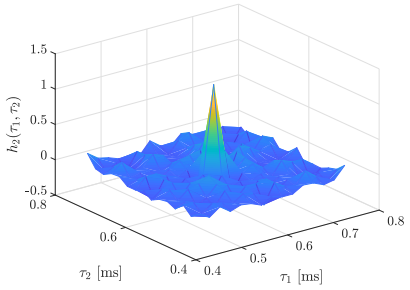


Figure 2.13: OTL509/25: second order Volterra kernel ($h_2(\tau_1, \tau_2)$).

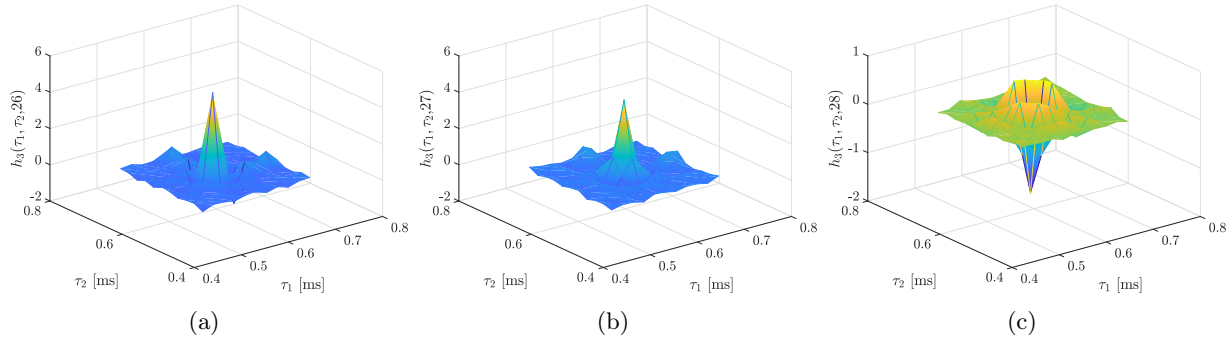
SI Audio OTL509/25 identification results In the second experiment the SI audio vacuum-tube amplifier was identified. Also in this case, four single-variance and one multiple-variance Volterra models were identified. The models were of the fourth order. In the multiple-variance method each kernel was identified with the input variances reported in Tab. 2.4.

The memory of the kernels of all the identified Volterra models were

$$\begin{aligned} M_1 &= 64, \\ M_2 &= 34, \\ M_3 &= 32, \\ M_4 &= 29. \end{aligned} \tag{2.27}$$

Table 2.4: Variances used in the multiple-variance estimation method.

kernel	input variance
k_0	σ_1^2
k_1	σ_2^2
k_2	σ_3^2
k_3	σ_4^2
k_4	σ_5^2



(a) (b) (c)

Figure 2.14: OTL509/25: third order Volterra kernel ($h_3(\tau_1, \tau_2, \tau_3)$).

Table 2.5: Parameters of reduced Volterra models of OTL509/25.

memory	delay
$M_1 = 64$	$D_1 = 0$
$M_2 = 15$	$D_2 = 19$
$M_3 = 11$	$D_3 = 21$
$M_4 = 5$	$D_4 = 24$

Figures 2.10(a) and 2.10(b) show the Volterra model of the first order, in time and frequency domain, respectively. Figure 2.10(b) also shows the power spectrum of the first order model obtained by the single-variance Volterra model, identified at the maximum variance, σ_5^2. It can be noticed a flat response, a magnitude of $H_1(f)$ of about 30 dB that, considering the attenuation of the D/A converter, corresponds to a gain of 36 dB.

Figure 2.11 shows the NMSE of the single-variance Volterra models and of the multiple-variance Volterra model, as a function of input standard deviation. The Figure was obtained by using as input of the Volterra models each of the white noise signals belonging to the testing set, shown in Figure 2.4(b), and calculating the NMSE for each of them.

Figure 2.11(a) shows that, also in this experiment, the error curve of the multiple-variance Volterra models interpolates the inferior values of the errors of the single-variance models, apart for the single-variance models with σ_5, which goes better at the point of maximum variance, but performs worse up to 30 dB at low variances.

Figure 2.11(b) shows similar curves, but obtained with the music inputs of Figure 2.4(b). Also in this Figure the multiple-variance method interpolates the inferior values of the single-variance one, although the single-variance model identified with σ_5, goes better for some inputs, with the best performance difference of -2.6 dB, but for low variances it performs worse until 35.5 dB of difference.

Each of these Volterra models, with the kernel memories reported in (2.27), has a total of 42604 kernel elements. Consequently, also in this experiment the procedure for the reduction of kernel element has been applied. The reduced Volterra models, that has been found to have performance comparable to the complete models, have parameters reported in Tab. 2.5.

The reduced models have now just 541 kernel elements and can be easily implemented in real-time.

Figures 2.12(a) and 2.12(b) show the performance of the reduced-memory Volterra models,

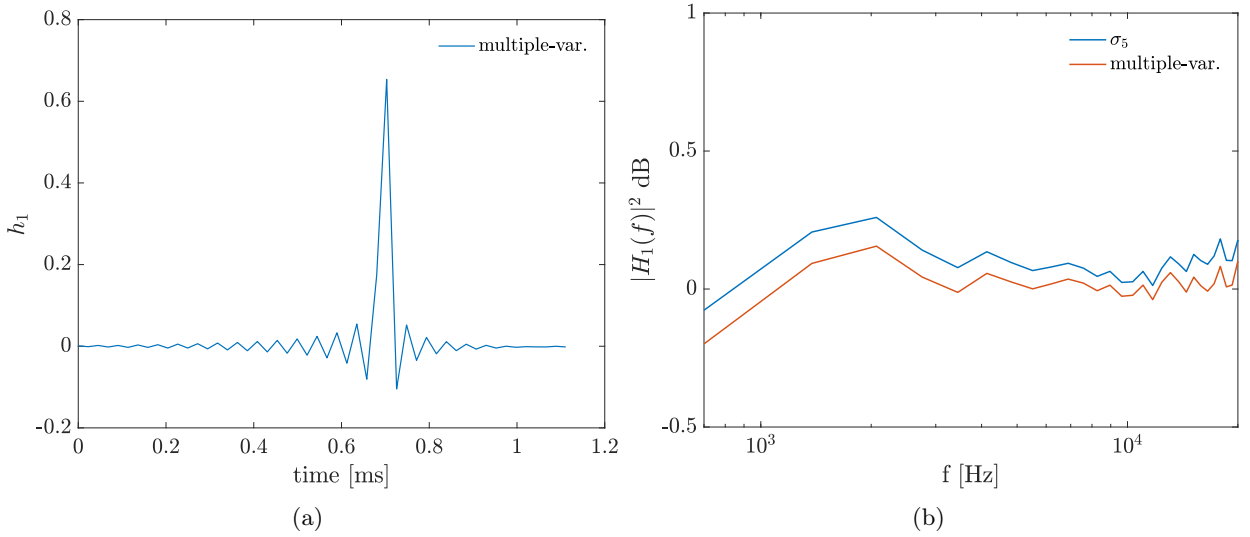


Figure 2.15: TubePre: first order Volterra kernel (a) in time domain ($h_1(\tau)$) (b) in frequency domain ($|H_1(f)|^2$).

obtained using white noise and music inputs, respectively. Also in this experiment, in comparison with Figure 2.11 we can see a performance improvement of the single-variance Volterra model σ_5 for high input variances and a worsening for small input variances. In Figures 2.12(a) we can see that in some small variance intervals the single-variance Volterra models perform slightly better than the multiple-variance one, but perform definitely worse for other variance values. In Figure 2.12(b) the multiple-variance Volterra model approaches better the inferior values of single-variance models.

Figure 2.13 shows the second-order kernel that is used in the reduced multiple-variance Volterra model. The elements shown in Figure start from 4.3 ms ($D_2 = 19$).

Figure 2.14 shows three slices of the third order kernel of the reduced multiple-variance Volterra model. The third order kernel has a delay similar to the second order one, but a smaller memory, $M_3 = 11$.

- Presonus TubePre identification results

In the last experiment the Presonus TubePre microphone pre-amplifier was identified. Four single-variance Volterra models and a multiple-variance model of the third order were first identified. Each kernel of the multiple-variance model was identified with a different variance as shown in Tab. 2.6.

Table 2.6: Variances used in the multiple-variance estimation method.

kernel	input variance
k_0	σ_1^2
k_1	σ_2^2
k_2	σ_4^2
k_3	σ_5^2

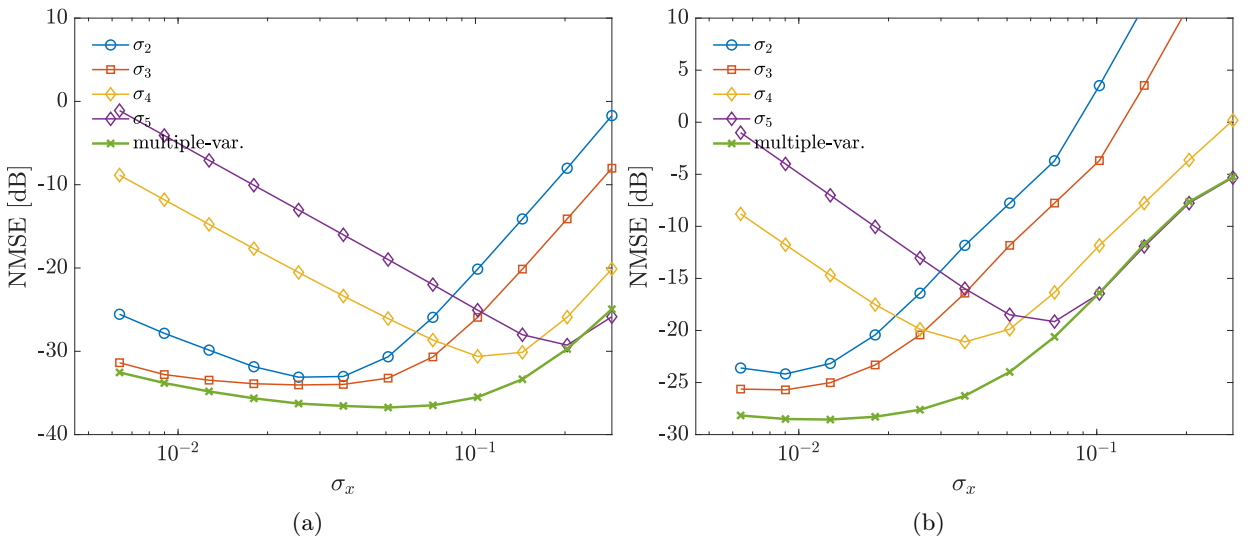


Figure 2.16: NMSE of complete Volterra models of TubePre for (a) WGN input, (b) music input.

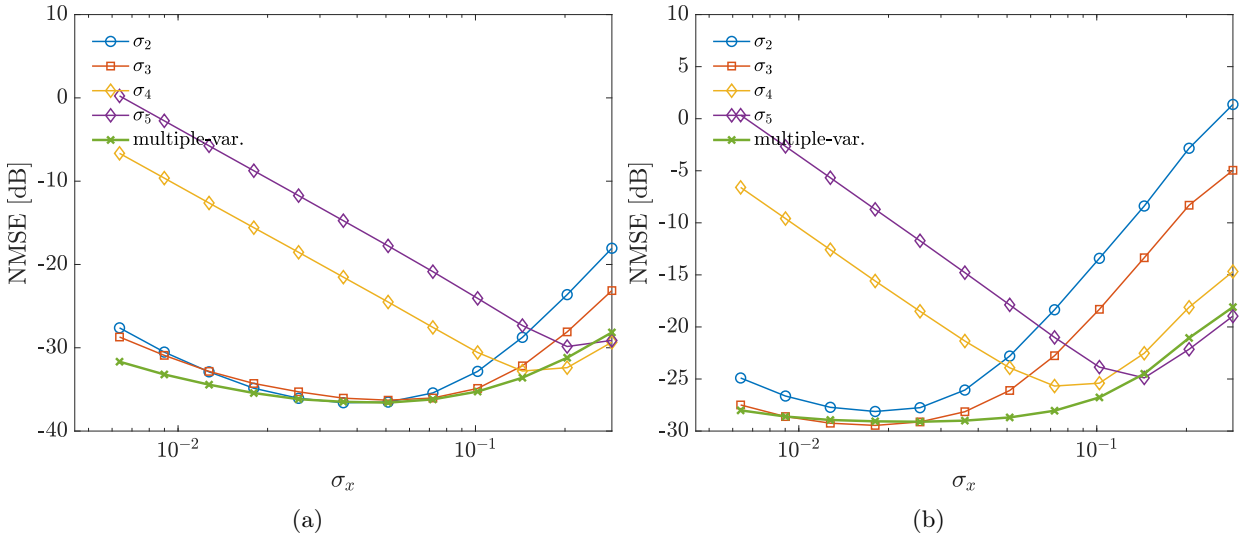


Figure 2.17: NMSE of complete Volterra models of TubePre for (a) WGN input, (b) music input.

The memory of the kernels of all the Wiener and Volterra models are

$$\begin{aligned} M_1 &= 50, \\ M_2 &= 35, \\ M_3 &= 35. \end{aligned} \tag{2.28}$$

Figures 2.15(a) and 2.15(b) show the Volterra model of the first order, in time and frequency domain, respectively. Figure 2.15(b) also shows the power spectrum of the first order model obtained by the single-variance Volterra model, identified at the maximum variance, σ_5^2. It can be noticed that, while Figures 2.5(b) and 2.10(b) correspond to amplifier gains of about 36 dB, the magnitude of TubePre, considering the attenuation of the A/D converter, corresponds to a 20 dB gain.

Figure 2.16 shows the NMSE of single-variance Volterra models and that of the multiple-variance Volterra model, as a function of input standard deviation. Figure 2.16 shows the best performance

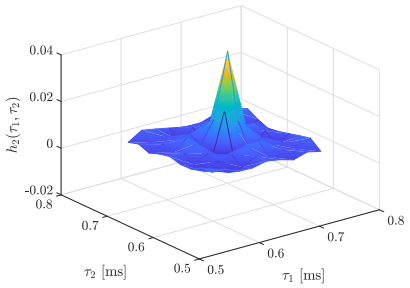
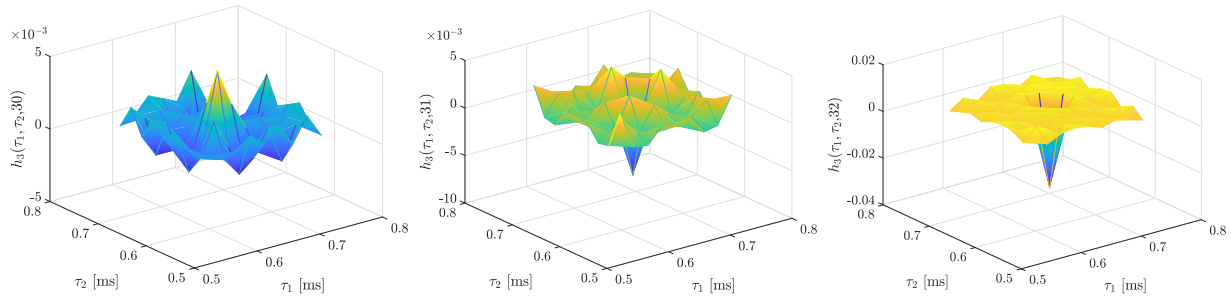


Figure 2.18: TubePre: second order Volterra kernel, $h_2(\tau_1, \tau_2)$.



(a) (b) (c)

Figure 2.19: TubePre: third order Volterra kernel $h_3(\tau_1, \tau_2, \tau_3)$.

of the multiple-variance methodology in the three amplifier identifications, in fact the NMSE of the multiple-variance Volterra model is always smaller than that of the single-variance Volterra models.

Each of these Volterra models, with the kernel memories reported in (2.28), has a total of 8451 kernel elements. Also in this case the procedure for the reduction of the number of kernel elements has been applied. The reduced Volterra models, that we have found to have performance comparable to the complete ones, have the parameters in Tab. 2.7. In this case the reduced models have just 261 kernel elements.

Figures 2.17(a) and 2.17(b) show the performance of the reduced-memory Volterra models, obtained using white noise and music inputs, respectively. The best performance shown in Figure 2.16 is maintained also for the reduced multiple-variance Volterra models.

Figure 2.18 shows the values of second-order kernel identified in the reduced multiple-variance Volterra model. The elements shown in the Figure start from 4.3 ms ($D_2 = 19$).

Figure 2.19 shows three slices of the third order kernel of the reduced multiple-variance Volterra model. The third order kernel has a delay and memory equal to the second order one.

Table 2.7: Parameters of reduced Volterra models of TubePre.

memory	delay
$M_1 = 50$	$D_1 = 0$
$M_2 = 9$	$D_2 = 26$
$M_3 = 9$	$D_3 = 26$

Table 2.8: Comparison between the number of kernel elements required by the complete Volterra model and by the reduced one, for the identified amplifiers.

Amplifier	Model order	**Number of kernel elements**	
		Complete model	Reduced model
Synthesis	3	6577	471
SI Audio	4	42604	541
Presonus	3	8451	261

Table 2.9: Number of operations required for a filter bank implementation of Volterra filters.

Amplifier	Model order	**Complete model**		**Reduced Model**	
		MAC˜per sample	MMACS	MAC˜per sample	MMACS
Synthesis	3	7664	338.0	617	27.2
SI Audio	4	57178	2521.5	792	34.9
Presonus	3	9745	429.8	359	15.8

Table 2.8 shows the number of kernel elements of the complete models and of the reduced ones, for all the three experiments. A reduction in the number of elements of one order of magnitude for the third order models and of two order of magnitude for the fourth order one can be achieved. As it can be seen by comparing Figures 2.6 and 2.7, Figures 2.11 and 2.12, and finally Figures 2.16 and 2.17, if the multiple-variance methodology is used this reduction in complexity can be realized without worsening the identification performance. This complexity reduction makes it possible to do a real time implementation of the proposed models. Using a filter bank implementation for the Volterra filter [190], a kernel of order P and memory M requires $\binom{M+P-2}{P-1}$ multiplications for computing the input samples cross-products and $\binom{M+P-1}{P}$ multiply–accumulate (MAC) operations for the filter bank operations. Many attempts have been made to reduce this number, by a proper combination of the operations, either in time [17, 79] or in frequency domain [25, 143, 178]. Regardless these techniques, Table 2.9 reports the number of MAC required by each proposed model per each sample and the million of MAC per second (MMACS), by using a filter bank implementation. Each multiplication used for computing the input samples cross-products is counted as a MAC. It can be noticed that the reduced models, which provide the same modeling accuracy of the complete ones, are affordable by most of the DSPs on the market. On the contrary, the complete models provide an unacceptable computational burden.

2.4.2 Sub-band Hammerstein adaptive identification

The previous reported approach allows a very accurate identification of a non linear devices, but on the other hand, they are quite complex and require a considerable computational power. For this reasons, in many applications more simpler models like those exposed in Section 2.3.3 are used. In this Section, a computationally efficient procedure for the Hammerstein model identification is exposed. It must be underlined that the same algorithm could be easily adapted also to the Wiener model with some minor changes. Starting from the notation of Section 2.3.3, the output signal $y(n)$

of the Hammerstein model is:

$$y(n) = w^T X(n)^T p, \tag{2.29}$$

where $w = [w_0, w_1, ..., w_{N-1}]^T$ and $p = [p_1, p_2, ..., p_M]^T$ are the linear and non-linear coefficient vector respectively, and $X(n)$ is the input signal rearranged in a matrix form. For the estimation of the coefficients of the filter w and of the polynomial p, an adaptive procedure based on normalized least mean square (NLMS) algorithm could be adapted [20], using the following formulas:

$$\begin{aligned} p(n+1) &= p(n) + \frac{\mu_p}{|X(n)w(n)|^2 + \epsilon} e(n) X(n) w(n) \\ w(n+1) &= w(n) + \frac{\mu_w}{|X(n)^T p(n)|^2 + \epsilon} e(n) X(n)^T p(n), \end{aligned} \tag{2.30}$$

where ϵ is a small arbitrary constant, μ_p, μ_w are respectively the polynomial and the filter step size, with the residual error $e(n)$ can be estimated according to:

$$e(n) = x(n) - w^T X(n)^T p. \tag{2.31}$$

In this case, the adaptation concerns both the linear FIR filter w and the power series coefficients p. Thus, since NLMS is a gradient descent algorithm, the performance of Hammerstein filter depends strongly on the initialization of adaptive filters. In fact, the cost function now is no longer a paraboloid and it shows local minima which should to be avoided. According to [20], in order to improve the convergence rate, the vector p has been initialized to all zeros with only the first coefficient set to one, while all the w coefficients have been initialized to zeros.
Fullband NLMS of Equations (2.31) and (2.30) could be modified in a sub-band structure, in order to improve the convergence rate and the performance in the case of non stationary signals such as speech [131], introducing a minimum amount of computational cost. In this paper, two MSAF (multi-band-structured sub-band adaptive filter) structures are employed. Differently from the sub-band adaptive filters (SAF) algorithm, where a different adaptive subfilter is used for each sub-band, in the MSAF technique the same fullband adaptive filter (according to Equation (2.30)) is considered for each sub-band. In this way, identical polynomial coefficients p and adaptive filters $W(z)$ are placed before the decimators, so they operate on the set of sub-band signals at the original sampling rate, as shown in Figure 2.20. Although MSAF structure has an higher computational cost than SAF, according to [126] MSAF has been chosen because it allows to avoid aliasing problems encountered in the critical sampled SAF. In particular, the proposed work aims at adapting the Hammerstein model to the MSAF structure, using wavelet and DCT filterbanks.

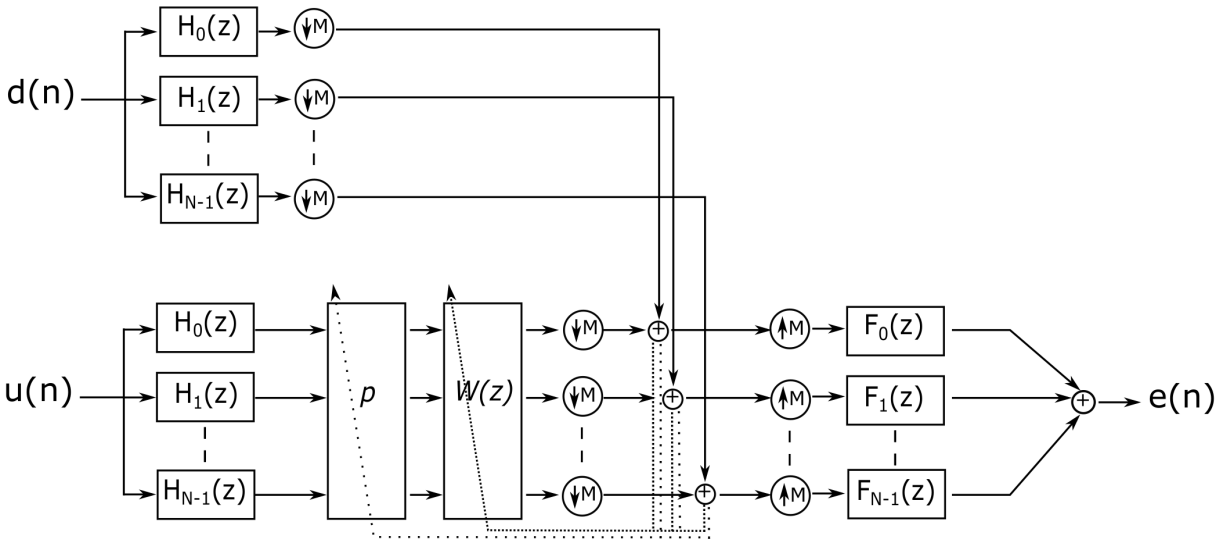


Figure 2.20: A multi-band structure (MSAF) for Hammerstein system identification. Identical fullband adaptive filters $W(z)$ and polynomial coefficients p are placed in all the sub-bands. Synthesis and analysis filters are generated trough WPT and DCT.

Wavelet filterbank

One of the proposed approaches involves the use of wavelet filterbank for sub-band model identification. Wavelet transform is a well known mathematical instrument for time-frequency signal analysis [65, 136]. In particular, in the discrete-time domain, wavelet transform is strongly tied to the concept of multiresolution [137]. Multiresolution in signal analysis means that the signal of interest can be analyzed at different scales in time with different analysis functions usually derived from one properly scaled prototype function. This allows the extrapolation of different features of the signal that will strongly depend on the used analysis function. The basic idea behind multiresolution is to expand the signal in terms of those analysis functions by successive approximation. Discrete wavelet transform (DWT) is defined by the following formula:

$$X_{DWT}\{p,q\} = \sum_{n=-\infty}^{\infty} x(n)\psi^*_{p,q}(n), \tag{2.32}$$

where ψ is called mother wavelet and symbol $*$ denotes the complex conjugate operator. The wavelet transform reconstructs the original signals by means of mother wavelet as basis functions. The choice of ψ is strongly related to the type of signal which has to be analyzed. In the literature, many different types of mother wavelet are available with different characteristics [65, 137]. For this work a fourth order Daubechies wavelet has been used, finding the best trade-off between performance and computational cost of the proposed approach. Computation of Equation (2.32) can be implemented through a recursive approach based on filterbanks. In that case, the DWT is equivalent to a perfect reconstruction filterbank. The only difference is in the used filters. In the wavelet case, the filters can be obtained from the mother wavelet and they are the same in each level of the decomposition. The filterbank structure for DWT transform is visible in Figure 2.21(a), where $G(z)$ denotes low-pass filters and $H(z)$ indicates high-pass filters. If the mother wavelet is

orthogonal (like the Haar wavelets and related Daubechies wavelets, Coiflets, and some developed by Mallat), each branch of the wavelet filterbank satisfies a quadrature mirror filter relationship. Thus, the analysis filters $H(z)$, $G(z)$ and the synthesis filters $\tilde{H}(z)$,$\tilde{G}(z)$ can be derived with the following relationships:

$$\begin{aligned} G(z) &= H(-z), \\ \tilde{H}(z) &= G(-z) = H(z), \\ \tilde{G}(z) &= -G(z) = -H(-z). \end{aligned} \tag{2.33}$$

Coefficients of the H filter can be derived from the ψ function of the chosen mother wavelet according to:

$$h_j[n] = \frac{1}{\sqrt{2^j}} \psi\Big[\frac{-n}{2^j}\Big]. \tag{2.34}$$

The filters coefficients of each branch are the same because of the dyadic sampling. In fact, dilating the wavelet filters by a j-th value, it is the same as filtering by those two filters and then subsampling by 2^j. In this way, an extremely computationally efficient structure is obtained and it has an O(N) complexity regardless of the depth of the decomposition, with N being the length of $x\,[n]$. In this paper, the wavelet packet transform (WPT) has been employed for the filterbank project. WPT is an extension of the discrete wavelet transform (DWT) in which the time-scale plane is sampled dyadically and, again, critically but in an uniform manner. It offers more flexibility than the DWT and richer signal analysis. Respect to DWT, at each decomposition level WPT analyzes in the same way both approximation and detail coefficients. In this way, a tree-like structure is obtained, and the number of coefficients grows exponentially with the number of levels as $2^{Nlevels}$. WPT filterbank structure is visible in Figure 2.21(b). For this work, a three level WPT filterbank has been chosen generating an eight branches structure with uniform sampling. It is possible to write the time-domain filterbank analysis matrix $\mathbf{H}_0$ (suppose 4-taps analysis filters for sake of simplicity), as follows:

$$\mathbf{H}_0 = \begin{bmatrix} h(3) & h(2) & h(1) & h(0) & & & \\ & & h(3) & h(2) & h(1) & h(0) & \\ & & & & . & . & .. \\ g(3) & g(2) & g(1) & g(0) & & & \\ & & g(3) & g(2) & g(1) & g(0) & \\ & & & & . & . & .. \end{bmatrix}, \tag{2.35}$$

where downsampling is already included. The synthesis matrix $\mathbf{F}_0$ can be generated, according to Equations (2.33), as follows:

$$\mathbf{F}_0 = \mathbf{H}_0^T. \tag{2.36}$$

Now the fullband structure described by Equations (2.31) and (2.30) can be adapted to WPT filterbank. First the $x_p(n)$ signal is divided in sub-band with the discrete wavelet packet transform

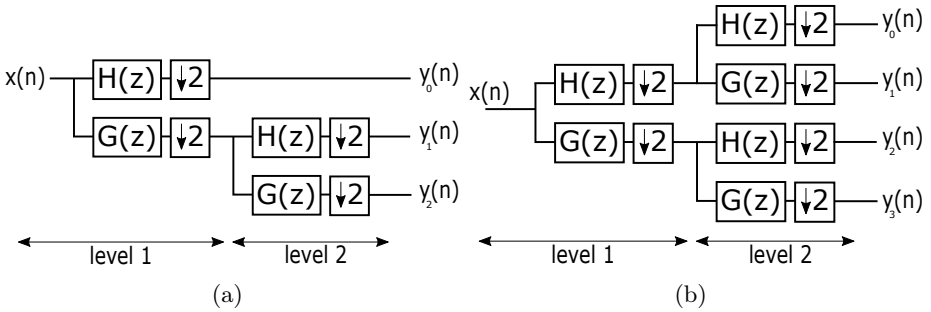


Figure 2.21: Two level wavelet decomposition filterbanks, in particular (a) shows a non uniform discrete wavelet transform decomposition three, while (b) shows a uniform wavelet packet transform decomposition three.

in the matrix form of (2.35). Then, the non-linear transformation p is applied to this new input signal in wavelet-domain and subsequently it is filtered by w. Both the linear filter and the polynomial coefficients are adapted at the decimated rate with the sub-band error signals, according to the fullband formula of Equation (2.30). A complete MSAF structure is reported in Figure 2.20, where the analysis and synthesis filterbanks are generated according to the Equations (2.35) and (2.36).

DCT Filterbank

In this paper also a DCT filterbank has been employed for the first time and it has been tested comparing its performances to the fullband-Hammerstein structure. Considering N sub-bands, the cosine-modulated analysis filters $H_i(z)$ are derived from a prototype low-pass filter $P(z)$ with a cut-off frequency of $\pi/2N$ [131], i.e.,

$$H_i(z) = \alpha_i P\big[zW_{2N}^{(i+0.5)}\big] + \alpha_i^* P\big[zW_{2N}^{-(i+0.5)}\big], \tag{2.37}$$

where $i = 0, 1, ..., N-1$, $W_{2N} = e^{-j\pi/N}$ and α_i is given as:

$$\alpha_i = \exp\left\{j\left[\theta_i - \frac{\pi}{N}(i+0.5)\left(\frac{L-1}{2}\right)\right]\right\}, \tag{2.38}$$

where L is the length of the prototype filter and $\theta_i = (-1)^i\frac{\pi}{4}$. The synthesis filters $F_i(z)$ are the time-reversed version of the analysis filters, i.e.,

$$F_i(z) = z^{-L+1}H_i(z^{-1}). \tag{2.39}$$

The real-valued impulse responses of the analysis and synthesis filters are derived from the Equation (2.37) as follows:

$$\begin{aligned} h_i(n) &= 2p(n)\cos\left[\frac{\pi}{N}(i+0.5)\left(n-\frac{L-1}{2}\right)+\theta_i\right], \\ f_i(n) &= 2p(n)\cos\left[\frac{\pi}{N}(i+0.5)\left(n-\frac{L-1}{2}\right)-\theta_i\right]. \end{aligned} \tag{2.40}$$

Perfect reconstruction of the system can be achieved by imposing the flatness constraint to the prototype filter $P(z)$, i.e.,

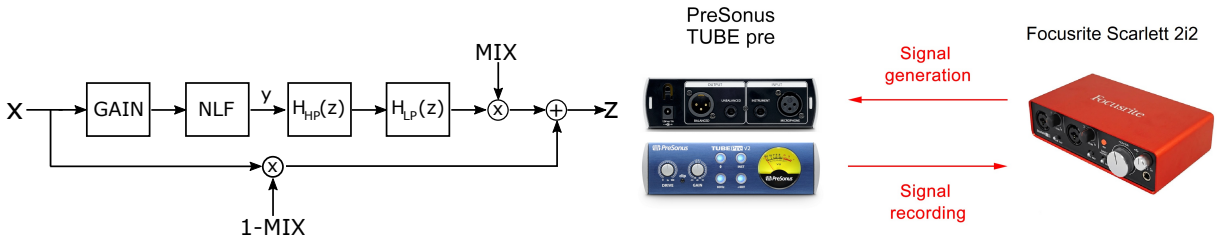
$$|P(e^{j\omega})|^2 + |P(e^{j(\omega-\pi/N)})|^2 \cong 1, \quad for \;\; 0 < \omega < \pi/N. \tag{2.41}$$

The analysis and syntesis filterbanks of the Figure 2.20 are generated according to the Equation (2.40).

Experimental results

Several experiments were carried out to test the proposed approaches, in particular a simulated system and a real world audio device have been identified. The fullband approach of [20] has been compared with DCT and Wavelet proposed algorithms. For the identification procedure a male recording from TIMIT database [92] has been used. The track has been resampled from 8 kHz to 48 kHz in order to allow its reproduction with the sound card used for the real measurements.

Synthetic experiment For the synthetic experiment, a non-linear model that represents the behavior of a vacuum tube has been selected [218]. The schematic block of the model is reported in Figure 2.22(a).



(a) (b)

Figure 2.22: (a) Vacuum tube synthetic model block diagram, and (b) measurement setup used for PreSonus tube pre identification process.

The first step is related to the application of the following non-linear function (NLF):

$$f(x) = \frac{x-\beta}{1-e^{-\alpha\cdot(x-\beta)}} + \frac{\beta}{1-e^{\alpha\cdot\beta}}, \quad \alpha \neq 0, x \neq \beta, \tag{2.42}$$

where x is the input and α, β are the parameters used to modify the distortion values. Then, both a low-pass and a high-pass filtering of the output signal y are performed to remove the DC component and to shape the higher frequencies.

The following filters are used:

$$H_{\mathrm{HP}}(z) = \frac{1-2z^{-1}+z^{-2}}{1-2r_{\mathrm{h}}z^{-1}+r_{\mathrm{h}}^2z^{-2}} \quad H_{\mathrm{LP}}(z) = \frac{1-r_{\mathrm{l}}}{1-r_{\mathrm{l}}z^{-1}}, \tag{2.43}$$

where H_{HP} and H_{LP} are the high-pass and low-pass filters, respectively, and r_l, r_h are their control parameters. A final control is used to mix the distorted and the original signals managing the amount of resulting distortion through the use of the MIX variable (see Figure 2.22(a)).

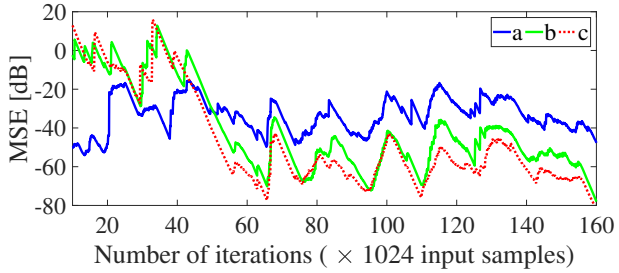


Figure 2.23: Comparison in terms of MSE on experiment of Equation (2.42) between (a) fullband algorithm, (b) DCT approach and (c) wavelet approach.

For the experiment, the following parameters have been set:

- $\beta = -0.2$;
- $r_{\mathrm{l}} = 0.1$, $r_{\mathrm{h}} = 0.99$;
- $\alpha = 10$;
- GAIN $= 1$;
- MIX $= 1$,

where GAIN and MIX refer to the variables of Figure 2.22(a) and the selected MIX value determines the presence of the distorted signal only.
The results are visible in Figure 2.23. For each method, the best MSE curve has been considered, by varying μ_p, μ_w and ϵ values. For the results the best performances have been achieved with a $\mu_p = 0.5$, $\mu_w = 0.8$. After the first iteration, the comparison between the fullband NLMS approach of [20] and the proposed approaches, shows that the DCT and the Wavelet algorithms reach a lower MSE, in particular the wavelet-based approach shows the best performance. Despite the reduced gain in computational cost of MSAF strucutres, the sub-band approaches still improve their performances with non stationary signals. Finally, it is worth noting that the best results obtained by wavelet filterbank are due to the wavelet capability of decomposing non stationary signals.

Real device experiment Regarding real world experiment, a tube based studio preamplifier has been considered, i.e., the PreSonus Tube Pre. The preamplifier is based on the 12AX7 triode valve and has two main controls: the gain, which acts on the overall volume, and the drive, which controls the amount of signal routed to the 12AX7 vacuum tube. Two experiments have been carried out varying the drive value and setting the gain to the lowest value (i.e., 3dB). For the first experiment the drive has been set at the lowest value of 0dB, while for the second experiment has been set to 15dB. A Focusrite Scarlett 2i2 sound card with the NU-Tech software [26] has been used for signal generation and acquisition. A scheme of measurement setup is visible in Figure 2.22(b).

The results of the two experiments are reported in Figures 2.24 and 2.25. In particular, Figure 2.24 shows the results for the lowest value of drive, which corresponds to a limited non-linear behavior of the device under test. As expected from synthetic experiment, sub-band approaches have better performances than the fullband. In particular, wavelet based structure has the best

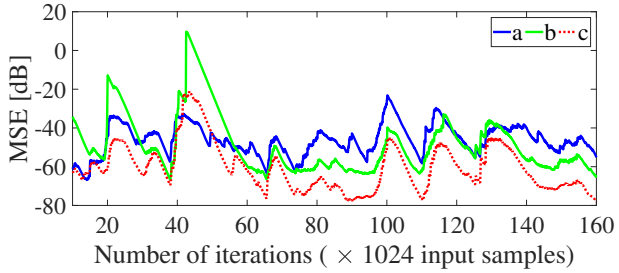


Figure 2.24: Comparison in terms of MSE on experiments on the tube pre-amplifier PreSonus TubePre with gain of 0dB between (a) fullband algorithm, (b) DCT approach and (c) wavelet approach.

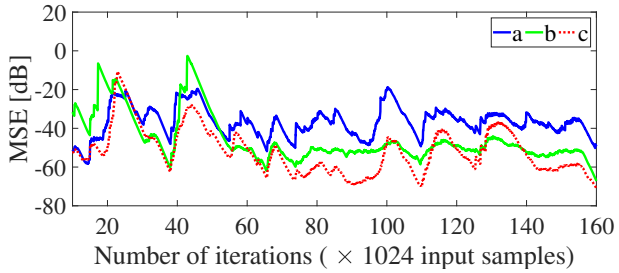


Figure 2.25: Comparison in terms of MSE on experiments on the tube pre-amplifier PreSonus TubePre with gain of 15dB between (a) fullband algorithm, (b) DCT approach and (c) wavelet approach.

performance reaching the lowest MSE. Figure 2.25 shows the results with an high level of drive, which corresponds to stronger non-linear behaviour. Also in this case the better identification is reached with the sub-band approaches with the better performance obtained by wavelet-transform filterbank.

2.5 Non linear measurement

As explained in Section 2.1, a complete identification of the non linear device is not necessary for many applications. For this reason, it is also clear the necessity of accurate and straightforward methods to monitor distortion levels, also outside the testing laboratories. In this context, efficient estimators of total harmonic distortion, harmonic distortion of order n, and intermodulation distortion are needed, which are among the most widely used parameters for non-linearity estimation [176]. In this Section, several approaches for the measurement of these indicator will be proposed with a particular attention on intermodulation distortion. It is a robust feature indicator within the whole systems operating bandwidth, and the measurement is based on the effects produced by the non linear system with a composite signal as input.

2.5.1 Non linear parameters

Intermodulation distortion

The intermodulation distortion (IMD) is produced by the non linear behaviour of a real system when the input signal contains two or more harmonics. In this case the system will generate new harmonics that are not present in the original signal and which are a linear combination of the

harmonics in the input signal. Thinking to an audio system, this is a critical aspect since IMD could degrade significantly the signal intelligibility. The standard technique for IMD estimation is detailed in IEC 60268 [4]. This is based on the observation that if two sinusoidal tones at angular frequencies ω_A and ω_B, with $\omega_A > \omega_B$, are input to a non-linear system, the output signal will contain, at each time t, harmonic components at integer multiples of ω_A and ω_B of the two tones. Moreover, the output signal will contain also harmonic components at angular frequencies that are linear combinations (sum or difference) of the individual harmonic tones $\omega = \pm n_A\omega_A \pm n_B\omega_B$ (with n_A and n_B integer values). As the system is band limited, the largest part of the harmonics produced by sum fall outside the useful band. Consequently, the harmonics produced by difference are considered, particularly those for which $n_A = n_B = k$, for $\{k = 1, 2\}$, being the most relevant as they show the highest energy content. According to IEC 60268, a quantitative evaluation of the intermodulation produced by the non-linear characteristics of the system can be measured by considering the frequency behavior of the component $\omega_A - \omega_B$, the difference frequency distortion of the second order (i.e. , when $k = 1$), when two signals of equal amplitude are input to the device. A narrow bandpass filter is adjusted for the angular frequency $\omega_A - \omega_B$ to measure the Root Mean Square (RMS) value of the component at that frequency. The value $\omega_A - \omega_B = 2\pi \times 80$ [rad/s] is suggested for the standard procedure. Denoting as P_{f_i} the sound pressure at frequency f_i, the quantity:

$$\text{IMD}\% = \frac{P_{f_A - f_B}}{P_{f_A} + P_{f_B}} \times 100 \tag{2.44}$$

is evaluated as the second order difference frequency IMD and it is assigned to the frequency value of $[\omega_A + \omega_B]/2$. The same measurement is repeated for different frequencies and amplitudes of the input tones to produce a result that is a function of applied voltages and frequency. In the more general case, for non-audio/speech applications, (2.44) can be formulated equivalently by replacing the pressure P_{f_i} with the amplitude of the harmonic component of the signal measured at the output of the non-linear system at the specified frequency.

Total harmonic distortion

The total harmonic distortion is a different type of distortion, it is generated when the non linearity of the system generates higher order harmonics of a given tone which are not present in the original signal. In comparison with IMD, this parameter is more crucial for hi-fi devices, but in some application like hi-gain guitar amplifiers is a desirable effect since is responsible of rock guitars sound. The estimation of THD is detailed in IEC 60268 [4], and is based on the observation of the output when the system is fed with a singular sinusoidal tone at angular frequency ω_A. The non linearity will produce in the output signal, at each time t, harmonic components at integer multiples of ω_A (i.e., higher order harmonics). According to IEC 60268, a quantitative evaluation of the harmonic distortion could be given generating a pure tone ω_A at different frequencies and measuring N_H higher order harmonics which are present in the output signal with the following formula:

$$\text{THD}\% = \sqrt{\frac{\sum_{n=2}^{N_H} |P_{f_n}|^2}{\sum_{n=1}^{N_H} |P_{f_n}|^2}} \times 100, \tag{2.45}$$

where P_{f_i} is the sound pressure at frequency f_i. The estimation is made making a ratio between the magnitude squared sum of all the harmonics except the fundamental, and the magnitude squared sum of all the harmonics considering also the fundamental. This indicator gives an idea of the relevance of the higher order harmonics with respect to the fundamental. The number N_H of harmonics which has to be considered, is not specified by the IEC 60268, and a preliminary spectral analysis could be useful to define the right number of higher order harmonics which has to be considered. Finally, as reported for IMD, in the more general case, for non-audio/speech applications, (2.46) can be formulated equivalently by replacing the pressure P_{f_i} with the amplitude of the harmonic component of the signal measured at the output of the non-linear system at the specified frequency.

Harmonic distortion of order k

The harmonic distortion of order k (HD_k) is similar to THD. The same type of distortion is considered but in this case HD_k allow a more specific analysis in order to better understand which higher order harmonic has the most significant impact on the overall distortion. For this reason the HD_k is estimated with the following formula:

$$\mathrm{HD}_k\% = \frac{|P_{f_k}|}{\sqrt{\sum_{n=1}^{N_H} |P_{fn}|^2}} \times 100. \tag{2.46}$$

The same denominator of THD fomrula is present, while on the numerator only the k-th harmonic considered is present. Typical real word applications estimate the HD_k for $k = 2, 3$, i.e., for the second and the third harmonic. Again as for IMD and THD the pressure P_{f_i} can be replaced with the amplitude of the harmonic component.

2.5.2 Proposed approach

The parameters explained in the previous Section are extremely time consuming, i.e., each measurement has a limited validity and is strictly related to the amplitude and frequency used for the test signal. For this reason, in the next Section, an innovative approach for the estimation of these parameters on a wide range of frequencies will be explained and then applied to the measurement of a real word audio devices. The approach is based on chirp based identification of generalized Hammerstein model.

Chirp based identification of Hammerstein model

The model identified is the generalized Hammerstein model, whose structure is represented in figure 2.26(a) and is quite similar to the Hammerstein model discussed in Section 2.2. A powerful method for its identification has been proposed, which makes use of an exponential swept-sine signal to excite the non-linear system [84, 176]. In (2.47), a general expression for an exponential swept-sine signal is given:

$$x(t) = \sin(\Phi(t)) \times [u(t) - u(t-T)] = C_{exp}(t, f_{MIN}, f_{MAX}, \Phi_0) \tag{2.47}$$

where $\Phi(t)$ is the instantaneous phase, T is the time duration of the swept-sine, f_{MIN} and f_{MAX} are the lower and the upper limits respectively of the swept frequency range, i.e., the bandwidth

B is equal to $B = f_{MAX} - f_{MIN}$, Φ_0 is the initial phase, and $u(t)$ is the Heaviside step function. The exponential swept-sine instantaneous frequency f_{inst} is defined as $f_{inst}(t) = \frac{1}{2\pi}\frac{d\Phi(t)}{dt}$ and $\Phi(t)$ is defined such that its time derivative follows the exponential law $f_{inst}(t) = k_0 e^{\frac{t}{L}}$. The constant L is defined as $L = \frac{T}{ln(f_{MAX}/f_{MIN})}$ and describes the time variation rapidity of f_{inst}. It follows that if $f_{inst}(t)|_{t=0} = f_{MIN}$ is wanted, then $k_0 = f_{MIN}$. Thus, the instantaneous frequency expression $f_{inst}(t) = f_{MIN}e^{\frac{t}{L}}$ is obtained after having set both L and f_{MIN}. The function $x(t) = \sin(2\pi f_{MIN}L[e^{\frac{t}{L}} - 1] + \Phi_0)$ implements the exponential chirp C_{exp}. The corresponding instantaneous frequency function $f_{inst}(t)$ is strictly monotonic, thus its inverse can be used to derive the time instant at which $f_{inst}(t)$ equals the generic frequency value f^*:

$$t^* \ni f_{inst}(t^*) = f^* \rightarrow t^* = L\ln(\tfrac{f^*}{f_{MIN}}). \tag{2.48}$$

Consider now the filter matched with the excitation, whose impulse response $\Psi(t, f_{MIN}, f_{MAX}, \Phi_0)$ is such that its convolution with the exponential swept-sine signal C_{exp} outputs an approximated Dirac's delta function $\delta(t)$, as for (2.49):

$$C_{exp}(t, f_{MIN}, f_{MAX}, \Phi_0) * \Psi(t, f_{MIN}, f_{MAX}, \Phi_0) \simeq \delta(t). \tag{2.49}$$

It has been shown in [84, 154, 176] that, if the non-linear system response to the exponential swept-sine signal is filtered through the matched filter, and the swept-sine signal fulfills the phase alignment conditions shown in [83, 153], then this filter works in an equivalent way on both the signal and its harmonics. In [153], authors show that the phase alignment condition is fulfilled if the exponential swept-sine signal $x_{exp}(t) = sin[2\pi f_{MIN}L(e^{\frac{t}{L}} - 1) + \Phi_0]$ can be written in the form $x_{exp}(t) = sin[2\pi f_{MIN}Le^{\frac{t}{L}}]$. In this case the phase alignment condition is fulfilled independently of the extreme frequencies f_{MIN} and f_{MAX} of the chirp signal and of its duration T. The initial phase can be constrained by appropriately choosing the value of the product $f_{MIN}L$. Within the validity limits of the Hammerstein model, the fulfillment of these conditions implies that the output of the matched filter is in the form of a sequence of impulsive functions $g_k(t)$ related to the different harmonics. This is because there is a close relation between the harmonics of the frequency f^* reached by the signal at time t^* (the harmonics whose frequencies are kf^*) and the same frequencies kf^* present in the undistorted exponential swept-sine signal at a time distance ΔT_k from t^*. The time shift ΔT_k is equal in value to $\Delta T_k = L\ln(k)$ and it is related solely to the order of the harmonic. Consequently, the k-th harmonic of the signal at time t equals the fundamental of the same signal at time $t + \Delta T_k$:

$$C_{exp}(t, kf_{MIN}, kf_{MAX}, k\Phi_0) = C_{exp}(t + \Delta T_k, f_{MIN}, f_{MAX}, \Phi_0). \tag{2.50}$$

This property is depicted in Figure 2.26(b) and it is the basis of the identification procedure of the Hammerstein model described in [84, 154, 176]: each one of the $g_k(t)$ functions is associated with a specific harmonic generated by the non-linear system, and the time positions of each $g_k(t)$ in the response of the matched filter show the order of the corresponding harmonic. In other words, the $g_k(t)$ function associated with the k-th harmonic is shifted earlier in time by the amount ΔT_k with respect to the $g_0(t)$ function, associated with the fundamental.
The $g_k(t)$ functions can be thus obtained by simply windowing the matched filter response in a time

interval associated with the k-th harmonic. Finally, the $h_k(t)$ functions identifying the Hammerstein model can be retrieved from the $g_k(t)$ through a linear transformation.

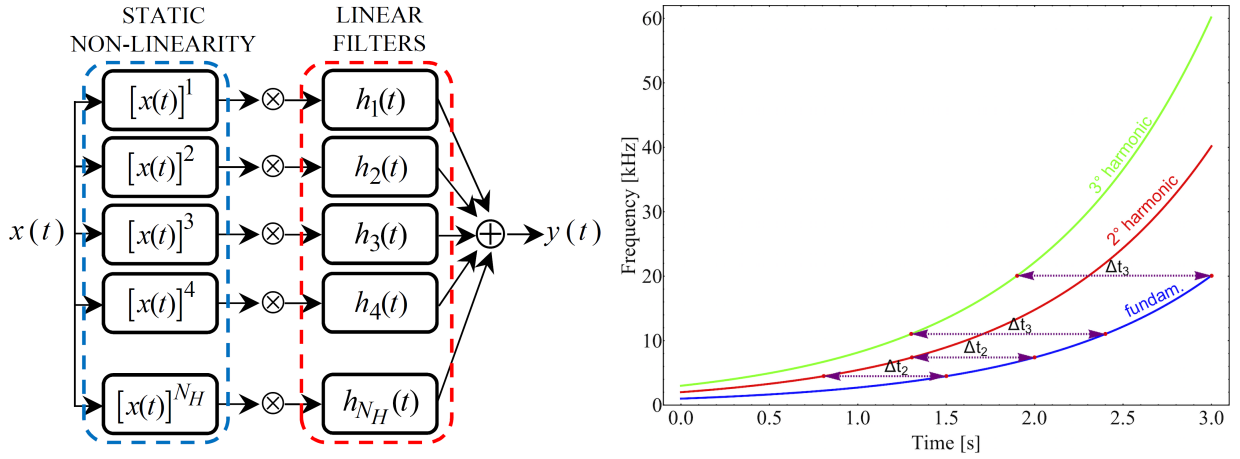


Figure 2.26: (a) Generalized polynomial Hammerstein model scheme and (b) exponential chirp signal in a non-linear device: time behavior of the fundamental frequency and of the first two harmonics. The time distance between the frequency of a harmonic and the same frequency seen as fundamental frequency is independent from the considered frequency.

Distortion parameters estimation

Starting from the above explained model, it is possible to estimate the frequency behavior of THD and HD_k through $g_k(t)$ [176, 209]. In fact the response of the Hammerstein model can be written in the time domain as:

$$y(t) = \sum_{k=1}^{N_H} x(t)^k * h_k(t). \tag{2.51}$$

The input $x(t)$ is a sine swept signal, and its powers can be reformulated in terms of its harmonics by using the Chebyshev polynomials of the first kind $T_n[\cdot]$. We can write:

$$\begin{aligned} y(t) &= \sum_{k=1}^{N_H} [C_{exp}(t, f_{MIN}, f_{MAX}, \Phi_0)]^k * h_k(t) = \\ &= \sum_{k=1}^{N_H} [\sum_{i=1}^{N_H} ([A_c]^{-1})_{i,k}\, C_{exp}(t, if_{MIN}, if_{MAX}, i\,\Phi_0)] * h_k(t) = \\ &= \sum_{k=1}^{N_H} [C_{exp}(t, kf_{MIN}, kf_{MAX}, k\,\Phi_0)] * g_k(t) \end{aligned} \tag{2.52}$$

where the $g_k(t)$ are related to the kernels $h_k(t)$ of the Hammerstein model through the linear transformation

$$g_k(t) = \sum_{i=1}^{N_H} ([A_c]^{-1})_{i,k}\; h_i(t) \tag{2.53}$$

and $[A_c]$ is the matrix of coefficients of Chebyshev polynomials of the first type [176] and the indices, in the adopted notation, indicate the trasposition of the matrix. Equation (2.52) can be written in

the frequency domain as:

$$Y(\omega) = \mathcal{F}\{y(t)\} = \sum_{k=1}^{N_H} G_K(\omega) \;\; \mathcal{F}\{C_{exp}(t, kf_{MIN}, kf_{MAX}, k\Phi_0\} \tag{2.54}$$

where $\mathcal{F}\{\cdot\}$ is the Fourier transform of the k-th harmonic of the input signal and $G_K(\omega)$ are the Fourier Transforms of the impulsive functions $g_k(t)$. As (2.54) shows, if the input is a sinusoidal signal, i.e., a special case of an exponential swept-sine signal, the output signal consists of the superposition of its harmonics; the amplitude of the k-th harmonic is given by $|G_k(\omega)|$ for each angular frequency ω_0. Thus, the $HD_k(\omega)$ and the $THD(\omega)$ can be estimated from the retrieved $G_k(\omega)$:

$$\mathrm{HD}_k(\omega) = \frac{|G_k(k\omega)|}{\sqrt{\sum_{n=1}^{N_H} |G_n(n\omega)|^2}} \tag{2.55}$$

$$\mathrm{THD}(\omega) = \sqrt{\frac{\sum_{n=2}^{N_H} |G_n(n\omega)|^2}{\sum_{n=1}^{N_H} |G_n(n\omega)|^2}}. \tag{2.56}$$

Also, IMD, which is one of the most important parameter can be derived with a similar procedure. In fact, the IMD can be measured by feeding the input of the non-linear system with a signal $x_D(t, \Delta T, f_{MIN}, f_{MAX}, T)$ consisting of the sum of two exponential swept-sine signals $x^A_{exp}(t)$ and $x^B_{exp}(t)$, sharing the same rate of exponential increase in frequency L, but shifted in time by the quantity ΔT (double chirp signal):

$$\begin{aligned} x_D(t, T, \Delta T, f_{MIN}, f_{MAX}, L, \Phi_0) = & \\ &= x^A_{exp}(t) + x^B_{exp}(t) = \\ &= x^A_{exp}(t) + x^A_{exp}(t + \Delta T) = \\ &= \Big\{ \sin(2\pi f_{MIN} L[e^{\frac{t}{L}} - 1] + \Phi_0) + \\ &+ \sin(2\pi f_{MIN} L[e^{\frac{t+\Delta T}{L}} - 1] + \Phi_0) \Big\} \cdot \\ &\cdot [u(t) - u(t - (T - \Delta T))]. \end{aligned} \tag{2.57}$$

The ratio between the instantaneous frequencies of the two exponential swept-sine signals summed at the input is constant, thus not dependent from a given time instant t:

$$\frac{f^B_{inst}(t)}{f^A_{inst}(t)} = \frac{f_{MIN} e^{\frac{(t+\Delta T)}{L}}}{f_{MIN} e^{\frac{t}{L}}} = e^{\frac{\Delta T}{L}}. \tag{2.58}$$

The non-linearity of the system will generate harmonic components in the output signal. Thus, there will be harmonics of the instantaneous frequency of each of the two exponential swept-sine signals and linear combinations (sum and difference) of these harmonics. The instantaneous frequency of the components generated by the non-linearity will vary over time following an exponential law, with the same growth rate L of the two chirps summed at the input. We verify this aspect for the intermodulation difference component of the second order, a signal that we indicate hereinafter

with C_{exp}^{DIFF}, whose instantaneous frequency is the difference of the two fundamentals. It is simple to show that the argument of the signal has a component at angular frequency $\omega_1(t) - \omega_2(t)$ which can be expressed in the form $2\pi f_{MIN} L e^{\frac{t}{L}}(e^{\frac{\Delta T}{L}} - 1)$, i.e., it is proportional to $e^{\frac{t}{L}}$. It follows that the time trend of its instantaneous frequency $f_{inst}^{A}(t) - f_{inst}^{B}(t)$ is given by:

$$f_{inst}^{A}(t) - f_{inst}^{B}(t) = f_{MIN}\left(e^{(t+\Delta T)/L} - e^{t/L}\right) = f_{MIN} e^{t/L}\left(e^{\Delta T/L} - 1\right). \tag{2.59}$$

As shown in (2.59), C_{exp}^{DIFF} is a chirp signal whose instantaneous frequency, produced by the difference frequency distortion component of the second order, evolves in time following an exponential law whose growth rate is L, i.e. , the same of each one of the two exponential swept-sine signals summed at the input. In particular, the instantaneous frequency equals the one of the first exponential swept-sine multiplied by the factor $(e^{\frac{\Delta T}{L}} - 1)$. This factor can be expressed in terms of a delay τ, imposing that $(e^{\frac{\Delta T}{L}} - 1) = e^{\frac{\tau}{L}}$, which gives:

$$\tau = L \times ln(e^{\frac{\Delta T}{L}} - 1), \tag{2.60}$$

where τ is the time difference between the current instant and the one at which a frequency equal to the difference $f_{inst}^{A}(t) - f_{inst}^{B}(t)$ is found in the undistorted exponential swept-sine signal, and ΔT is the time shift between the two exponential swept-sine signals at the input. The intermodulation component $f_{inst}^{A}(t) - f_{inst}^{B}(t)$ grows exponentially with the same rate L of the two exponential chirps summed at the input. Thus an impulsive function will be produced at the output of the matched filter if the phase alignment condition is met as well. As already pointed out, the pulse compression method can be in fact successfully applied when the instantaneous phase of the exponential swept-sine signal and the one of its harmonics are aligned. It is simple to show that the same condition is satisfied for the intermodulation signal.

In fact, the instantaneous phase of the difference frequency swept-sine signal can be written as:

$$\begin{aligned} \phi_{A-B}(t) &= \phi_A(t) - \phi_B(t) = \\ &= (2\pi f_{MIN} L e^{\frac{(t+\Delta T)}{L}}) - (2\pi f_{MIN} L e^{\frac{t}{L}}) = \\ &= 2\pi[f_{MIN}(e^{\frac{\Delta T}{L}} - 1)] L e^{\frac{t}{L}}. \end{aligned} \tag{2.61}$$

The difference frequency swept-sine signal has thus an instantaneous phase $\phi_{A-B}(t)$ which can be written in the form that guarantees the fulfillment of the phase alignment condition, and whose minimum frequency is $f_{MIN}(e^{\frac{\Delta T}{L}} - 1)$. Consequently, the conditions for phase alignment for each individual exponential chirp also guarantees phase alignment for the frequency-difference chirp signal. If the above conditions hold, when the signal $y(t)$ at the output of the non-linear system is processed through the matched filter, a signal $o(t)$ consisting of a sequence of short duration waveforms is obtained. The peaks deriving from the compression of the fundamental and from the harmonics of each one of the two exponential swept-sine signals can be identified in $o(t)$. Moreover, the peak associated to the difference of the two exponential swept-sine signals is found at a time delay τ from the reference time; the corresponding waveform is denoted as $g_{A-B}(t)$. Figure 2.27 shows the trend of the $o(t)$ signal in the simple case of a quadratic non-linearity. The two plots at the top of Figure 2.27 show the $o(t)$ signal if the exponential chirp $x_{exp}^{A}(t)$ (black curve) and

the $x^B_{exp}(t)$ (red curve) are placed separately at the input to the non-linear system: in this case the fundamental component is not present in the signal raised to the second power, the continuous component falls out of the chirp band. Consequently, in the $o(t)$ signal there is only the peak associated to the second harmonic, at a time distance ΔT_2 (or $\Delta T_2 + \Delta T$) from the origin of time.

In the case in which the two exponential chirp signals, temporally distant ΔT, are summed and put at the input of the quadratic non-linearity, the signal $o(t)$ at the output of the matched filter (represented in green) will contain, besides the two peaks relative to the components of second harmonic (already present for the single chirp), also a component associated to the instantaneous sum frequency and a component associated to the instantaneous difference frequency; this latter peak will be located with a delay τ with respect to the origin of time. Starting from the above waveforms obtained after compression with the matched filter, the IMD can be estimated as:

$$\mathrm{IMD}(f)_{-chirp-} = \frac{|G_{A-B}(f)|}{|G^A_1(f)+G^B_1(f)|} \times 100 \tag{2.62}$$

where $G_{A-B}(f)$, $G^A_1(f)$ and $G^B_1(f)$ are the Fourier Transforms of the kernels $g_{A-B}(t)$, $g^A_1(t)$ and $g^B_1(t)$ respectively, where $g^A_1(t)$ and $g^B_1(t)$ are the impulsive functions associated to the fundamental component of the two swept-sine signals summed at the input. This formulation, allows the estimation of the entire frequency trend of the intermodulation distortion. That can be obtained starting from the response to a single signal by which the non-linear system is energized. To properly select the $g_k(t)$, windows were adopted whose width is comparable to ΔT. We have adopted windows based on Reactance Transformation [36], with a very low value of the control parameter, in order to guarantee both a consistent reduction of the side lobes and a minimum alteration of the spectrum of the signals obtained.

Note that the IMD estimation based on the proposed method makes use of the Fourier transforms of three $g_k(t)$ among the functions obtained at the output of the matched filter. Therefore, the proposed method does not depend on the choice of the order N_H of the Hammerstein model to which the estimation is implicitly linked.

In fact, the proposed IMD estimation procedure is implicitly based on the hypothesis that the non-linear system can be represented by a parallel Hammerstein model. The validity of this hypothesis has been verified indirectly by comparing the estimates of the IMD obtained with the proposed method and the one provided by the standard procedure.

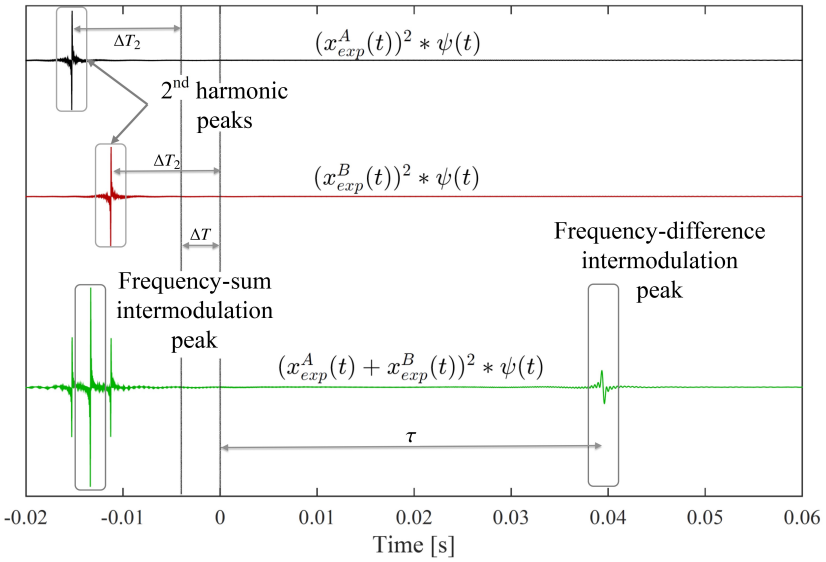


Figure 2.27: Trend of the $o(t)$ signal in the simple case of a quadratic non-linearity: when the exponential chirp $x^A_{exp}(t)$ is input (black curve); when the exponential chirp $x^B_{exp}(t)$ is input (red curve); when the sum of the two exponential chirp signals $(x^A_{exp}(t) + x^B_{exp}(t))$ is input (green curve). The peaks associated to the sum and difference intermodulation components are evidenced in the green curve.

Experimental results

Several experiments were carried out to test the proposed approach to measure the IMD, THD and HD_k, analyzing several scenarios. In particular, both a simulated system and real-world audio devices were tested to analyse the results obtained by means of the proposed method, compared to a well-known measurement system, i.e. , the National Instruments sound and vibration toolkit [7]. The latter provides the possibility of measuring IMD, THD and HD_k according to the international standard [4].A National Instruments PXI chassis (PXIe-1073) equipped with the PXI-4461 dynamic signal acquisition module [6], and connected to a desktop PC were used for the signals generation and acquisition. A LabVIEW script has been developed to manage the whole procedure. All the measurements were performed by setting a sampling frequency of 196 kHz, and repeated at various frequencies in order to obtain an estimation of the distortion parameters in the frequency range from 1.5 kHz to 15 kHz. The IEC standard-compliant measurement procedure for the estimate of harmonic distortion and intermodulation distortion was carried out separately for each frequency. For the harmonic distortion measurements, the frequency range of interest was divided into frequency intervals of 100 Hz width; for each of the 136 frequencies identified, a sinusoidal tone of 10 ms duration was sent, and the response signal of the non-linear system was analysed, detecting the amplitudes of the fundamental component and the harmonics up to the seventh one. From this data the values of THD, HD_2 and HD_3 were calculated for each of the 136 frequencies. About the measurement of the intermodulation distortion, for each of the 136 frequencies, a signal was sent consisting of the superposition of two sinusoidal tones of frequencies adjacent to the measurement frequency, separated by 80 Hz from each other. In the case of the chirp measurement, the signal used for the experiments consisting of the superposition of two exponential up-chirps, was defined taking into account the frequency band in which the estimate of distortion was to be made: the

frequency range of each of the two overlapping chirps was defined as a function of the band to be covered, extending it beyond the measurement frequencies to avoid the oscillations that occur at the ends of the band in the chirp signal spectrum. These aspects have been here deeply investigated and a way of selecting the parameters of the input signal has been defined that allows an accurate control of the quality of the estimate throughout the band of interest. In the work [34], in fact, limitations in the accuracy of the estimation had emerged at the extremes of the frequency band. This effect was analyzed in detail, verifying that it can be associated with two different aspects. A first aspect is related to the frequency band covered by the difference chirp, generated by the intermodulation of two chirps whose frequency ratio is $\frac{f^B_{inst}(t)}{f^A_{inst}(t)} = e^{\Delta T/L}$. If each of the incoming chirps covers a frequency range from f_{MIN} to f_{MAX}, the difference chirp C^{DIFF}_{exp} will extend from the minimum frequency $(e^{\Delta T/L} - 1)f_{MIN}$ to $(e^{\Delta T/L} - 1)f_{MAX}$. Consequently, it is necessary that the matched filter covers a frequency band that includes this extended range to get a correct estimate of the IMD in the entire frequency band. A second aspect is related to the value of the $T * B$ product associated to each one of the two chirps summed at the input: for the IMD estimation to be accurate, both frequencies on which the estimation is based must fall into a range where the chirp spectrum is appreciably flat. Meeting this condition requires (i) that the $T * B$ product is high enough; (ii) that the entity of the $f^B_{inst}(t) - f^A_{inst}(t)$ frequency difference is considered, in particular at the upper end of the covered band. The IMD is calculated between two harmonic components that differ, instant by instant, by a quantity that varies between $f_{MIN}(e^{\frac{\Delta T}{L}} - 1)$ and $f_{MAX}(e^{\frac{\Delta T}{L}} - 1)$ as t increases. The estimated IMD is associated at each instant with the intermediate frequency between the two. An excessive increase in the difference between the two frequencies on which the IMD is calculated can take one of the two components out of the useful band of the matched filter, and thus alter the IMD estimate for the corresponding frequency. Taking into account both these conditions, it was possible to define input signals that allow to predictably control the accuracy of the IMD estimation, as shown by the synthetic and real examples below. The time distance ΔT between the two chirps was defined in such a way that the difference between the instantaneous frequencies of the two overlapping chirps would be kept at contained values. The results shown in the following figures were obtained by sending a single input signal to the system, whose total duration was of 2.3 s; the two chirp components were superimposed at a time distance equal to 7.45 ms. The signal thus defined was sent only once to the system, to obtain the estimate of the frequency trend of the distortion. A number of experiments, not reported here, were performed, and similar results were obtained, with different input signals, for other frequency bands. The description of the measurement process provides a clear indication of the considerable difference in complexity and duration of the measurement operations carried out according to the two procedures, the one standardized by IEC and the one herein proposed.

Figure 2.29 shows the results obtained with the polynomial function introduced in Section 2.5.2. The proposed approach exhibits excellent results in comparison to both the theoretical values and to the method of IEC 60268, for all the measured distortion parameters, showing only some minor difference in the estimation, ranging from 0.1% for HD_2 to 0.02% for HD_3, on average. Figure 2.30 shows the results obtained with the synthetic tube model of Section 2.5.2: a good accordance with the IEC measured parameters is obtained, showing only some minor difference in the estimation, ranging from 0.02% for HD_3 to 0.9% for IMD, on average.

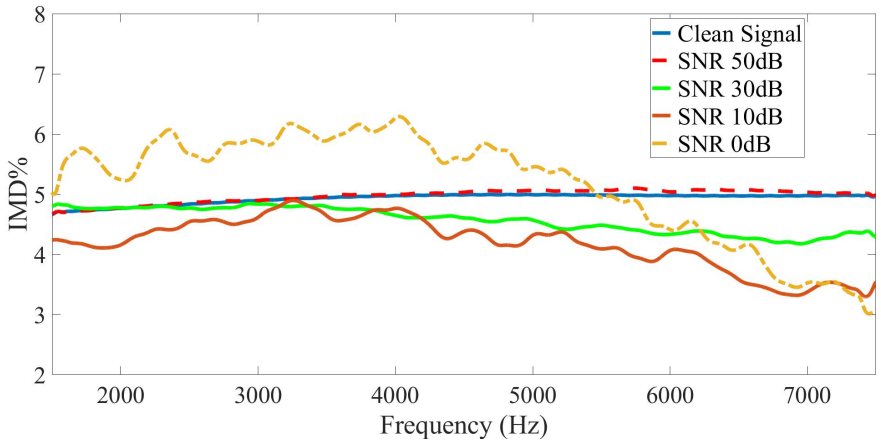


Figure 2.28: IMD estimation for different Signal To Noise Ratio (SNR) values, in case the case of the non-linear polynomial function defined by (2.63).

.

Experiments with simulated systems For the simulated experiments, the distortion parameters produced by a non-linear polynomial function and by the synthetic model of a vacuum tube have been measured.
The first synthetic experiment was performed considering the following function:

$$f(x) = A \cdot x^2 + x^3, \tag{2.63}$$

where the parameter A allows to vary the amount of quadratic non-linearity. Since the non-linearity is defined in analytical form, it is possible to calculate the theoretical value of IMD, THD and HD_k according to the value assigned to the constant A: in this case its value was set to A=0.1125, with an input signal of 2 V peak-to-peak amplitude (V_{pp}), leading to the theoretical values IMD=5%, THD=32.33%, HD_2=7.10%, HD_3=31.54%.

The robustness of the pulse compression techniques with respect to environmental noise is well known. So, one of the idea that has been verified was if this feature is also proper to the method of estimating the intermodulation distortion. This characteristic has been verified through the following experiment: a zero-mean white Gaussian noise has been added to the signal at the output of the non-linear system. The variance of the additive noise has been modified to adjust the resulting signal-to-noise ratio (SNR). It should be noted that the SNR is evaluated on the signal input to the matched filter. At the output of this filter, the signal will be affected by a noise that will overlap the sequence of functions $g_k(t)$: since the amplitude of the different impulsive functions varies significantly between the different harmonics, the signal associated with the intermodulation component, typically much smaller than the amplitude of the fundamental and the first harmonics, will be greatly disadvantaged, buried in the noise even for not very low SNR values. Despite this, the IMD(f) estimate, although worsening as the SNR value decreases, maintains trends close to the ideal one even for very low SNR values, as clearly shown in Figure 2.28.
A further simulated experiment was carried out considering a non-linear behavior that represents the effect of a vacuum tube [218], for this experiment the same model exposed in Section 2.4.2 has been used with an input signal of 2 V_{pp} and with the following parameters: $\beta = -0.2$, $r_l = 0.1$, $r_h = 0.99$, $\alpha = 10$, $GAIN = 1$ and $MIX = 1$.

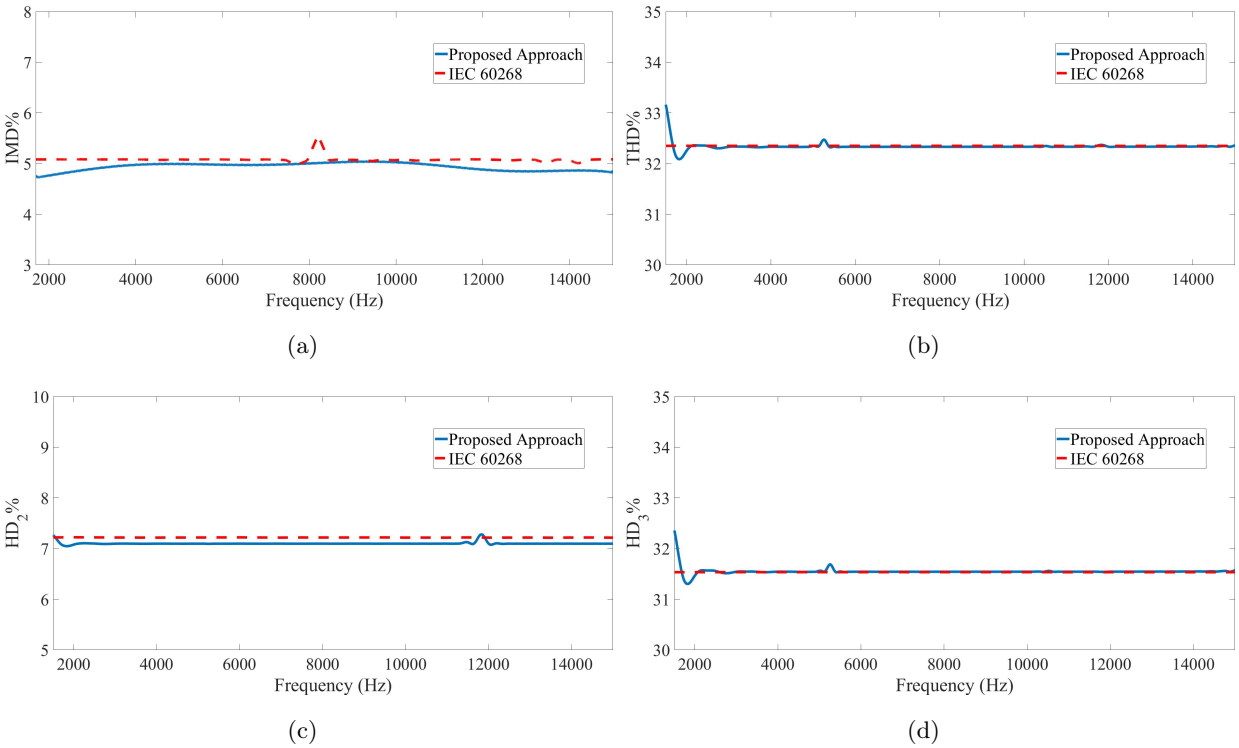


Figure 2.29: Results in terms of percent in the case of non-linear function defined by 2.63 (a) IMD, (b) THD, (c) HD_2, and (d) HD_3 considering the proposed approach and the IEC 60268 measurements.

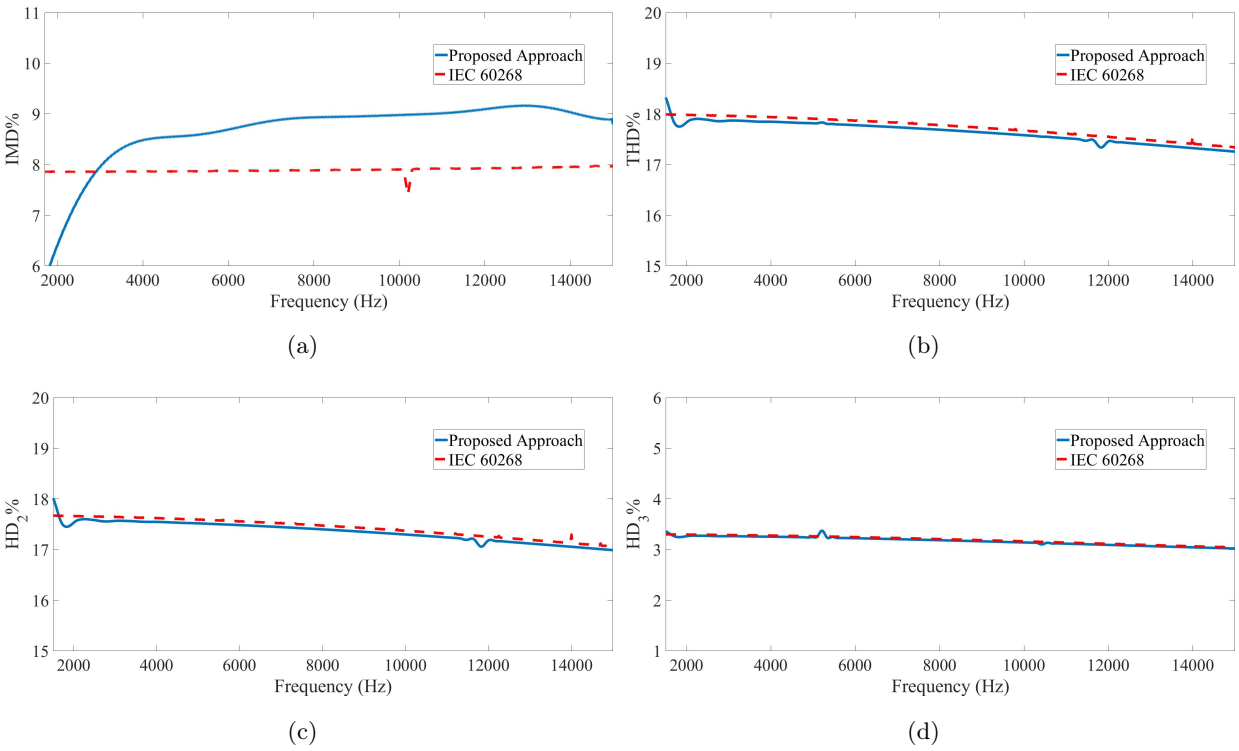


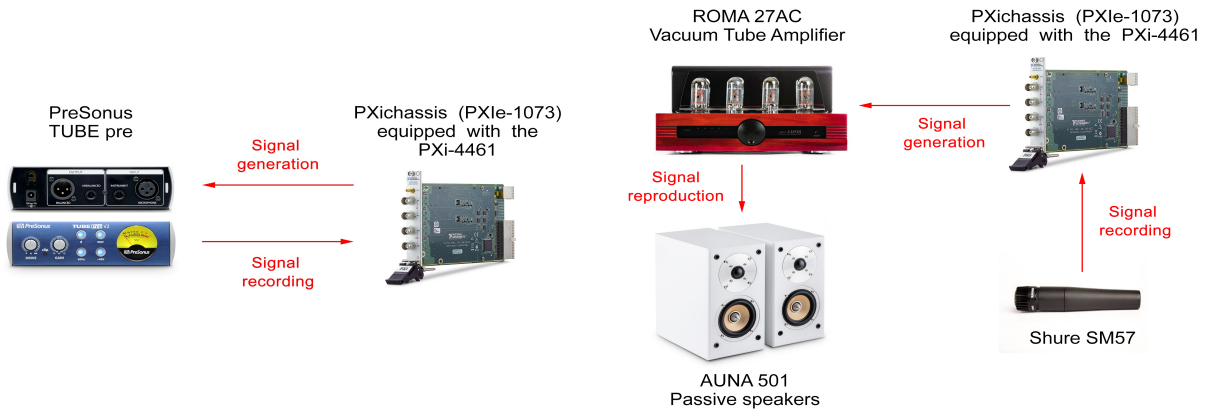
Figure 2.30: Results in terms of percent for the synthetic experiment described in Figure 2.22(a) in terms of percent (a) IMD, (b) THD, (c) HD_2, and (d) HD_3 considering the proposed approach and the IEC 60268 measurements.

Table 2.10: Mean percent difference between the proposed approach and IEC 60268 measurements, for each distortion parameter considered for synthetic experiments.

Experiment	Δ_{IMD}	Δ_{THD}	Δ_{HD_2}	Δ_{HD_3}
Polynomial model	0.1%	0.03%	0.1%	0.02%
Vacuum Tube model	0.9%	0.09%	0.08%	0.02%

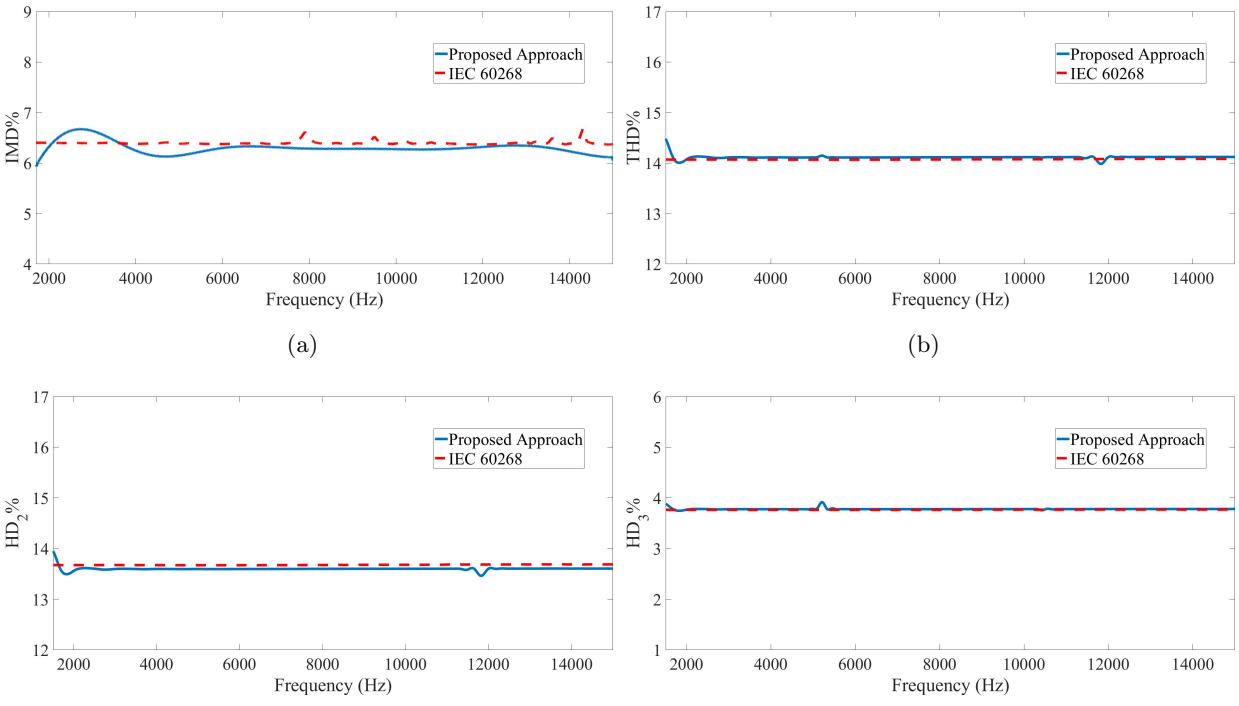
Table 2.10 shows the mean percent difference between the two estimation for synthetic experiments, and the results shows only a limited difference between the two measurements.

Experiments with real devices For the real experiments, the distortion parameters produced by two real devices based on vacuum-tube technology, were measured. The former device was a preamplifier, i.e. , the PreSonus TubePre, a microphone and instruments preamplifier equipped with a 12AX7 triode valve. The device has been connected directly to the PXI-4461 acquisition board, as reported in Figure 2.31(a). For this experiment, the two main controls of the preamplifier have been used, i.e. the drive that acts on the valve saturation and the gain control which acts on the overall gain. They have been set to 3 dB and to 0 dB respectively, to obtain a low level of non-linearity, as reported in the experiments. Measurements have been made using a 2 V_{pp} input signal. The latter device was a Hi-Fi amplifier, i.e. the Synthesis Roma 27AC, a 25 Watt AB Class tube amplifier. The hardware connection is reported in Figure 2.31(b). The amplifier was connected to the PXI-4461 acquisition board for the signal generation and to a pair of loudspeakers that are used to load the amplifier and for the signal reproduction. In particular, an Auna 501 passive 2-way speaker has been used featuring an RMS rated power of 100 W, an impedance of 8 Ω, and a frequency range of 92 Hz - 20 kHz. The volume of the amplifier has been set to the highest value and an input signal of 1 V_{pp} has been used. After half an hour of warm up, the input signal was applied and the output signal has been measured by means of a dynamic microphone, i.e. a Shure SM57, placed in front of one of the speakers. The choice of a dynamic microphone instead of a more precise measurement microphone is related to a specific reason, i.e., the SM57 is a widely adopted microphone used in many real word scenarios for miking loudspeakers and guitar amplifiers. In comparison with the amplifier and the loudspeaker, it should introduce a negligible distortion, which in every case it is measured from the system and considered in the setup.



(a) (b)

Figure 2.31: (a) Hardware connection for the real experiment considering the PreSonus TubePre preamplifier, and (b) Synthesis Roma 27AC amplifier.



(a) (b)

(c) (d)

Figure 2.32: Results in terms of percent for the PreSonus tube preamplifier experiment (a) IMD, (a) THD, (c) HD_2, and (d) HD_3 considering the proposed approach and the IEC 60268 measurements.

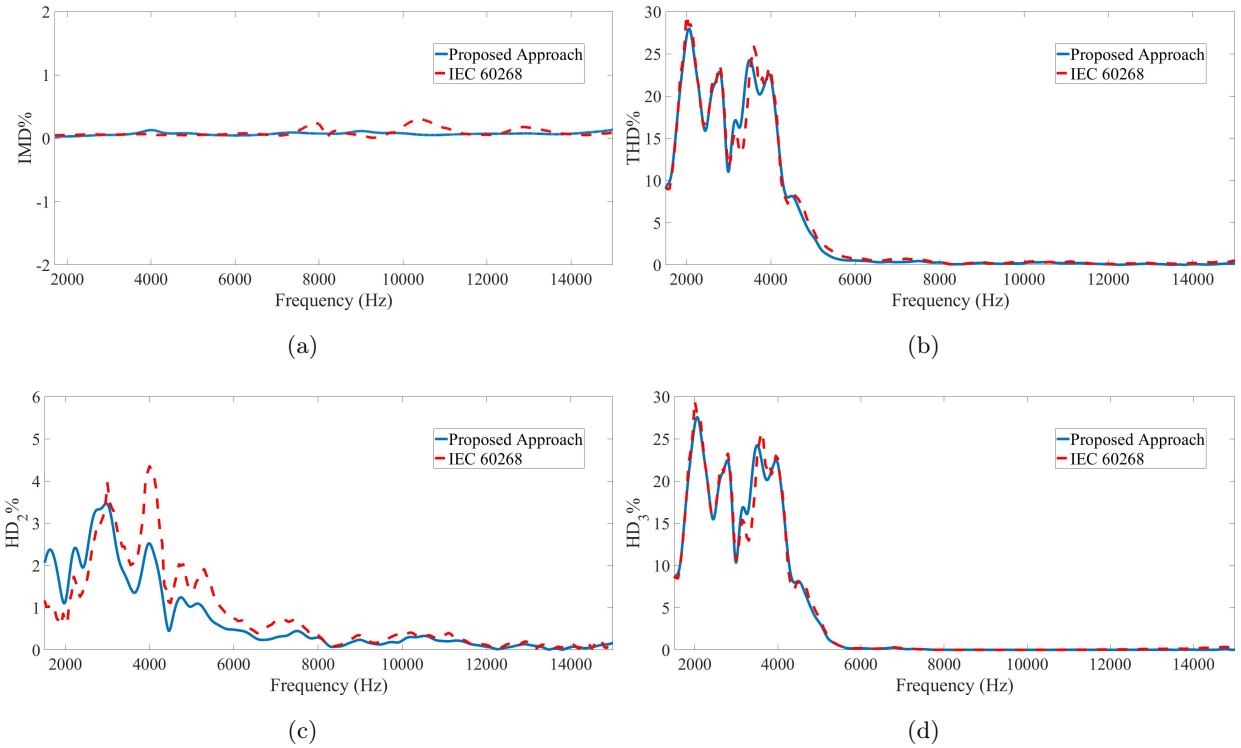


Figure 2.33: Results in terms of percent for the Roma Hi-Fi amplifier with Auna loudspeakers experiment, (a) IMD, (b) THD, (c) HD_2, and (d) HD_3 considering the proposed approach and the IEC 60268 measurements.

Figure 2.32 shows the obtained results for the first experiment on real devices, i.e. , the PreSonus TubePre, considering a low level of non-linearity. Taking into consideration the HD_k and THD measurement, the proposed approach shows good results in comparison to the IEC 60268 methodology. This fact is also confirmed by the overall distortion measurement, that presents a very small difference at lower frequencies, ranging from 0.1% for IMD to 0.01% for HD_3, on average. Figure 2.33 shows the results obtained for the Hi-Fi amplifier plus the loudspeakers that present a high level of distortion. In this case, from the point of view of harmonic distortion measurement, there is a difference in the second harmonic, that seems to be underestimated. However, the THD measurement is less affected and the result accurately follow the one obtained according to the IEC 60268 methodology, also in terms of IMD values. In particular, the difference on the measurement varies from 2% for HD_2, to 0.1% for IMD, on average. Moreover the underestimation is limited to the frequency range from 1 to 8 kHz, where the system seems to have a strong non-linear behavior. Future work will be oriented to a deeper investigation of this aspect performing further experiments with different loudspeakers and amplifiers. In conclusion, in a real scenario the proposed approach can be a valid alternative to support the IEC procedure for preliminary, fast and reliable distortion parameters estimation. In the case of very complex frequency trends and high distortion levels, if an estimate with a high degree of accuracy is required, the standard IEC procedure in an acoustically controlled environment should be used.

Table 2.11: Mean percent difference between the proposed approach and IEC 60268 measurements, for each distortion parameter considered for real devices experiments.

Experiment	Δ_{IMD}	Δ_{THD}	Δ_{HD_2}	Δ_{HD_3}
Preamplifier	0.1%	0.05%	0.6%	0.01%
Hi-Fi Amplifier	0.05%	0.4%	0.3%	0.3%

Table 2.11 shows the mean percent difference between the two estimation for real devices experiments, and the results shows only a limited difference between the two measurements.

2.6 Conclusions

In conclusion, in this chapter several aspects of non linear modeling and identification have been analyzed. First, an introduction to non linear devices and tube amplifiers has been given. Then, several innovative and state of the art models have been explained and discussed. The first has been the family of LIP filters, then the complex but reliable models based on the Volterra filters and finally the simpler but effective Wiener and Hammerstein models. In the identification Section, an explanation of how each approach has been applied is given. In particular the Volterra filters have been identified using the multivariance approach to overcome the locality of the solution, making several experiments on real word audio devices, moreover it has been shown that it is possible to reduce the dimensions of the identified model without affecting too much the accuracy of the identification allowing a more easier real time implementation. Then, speaking of the Hammerstein model, it has been investigated the possibility of using a sub-band approach to reduce the computational cost of the NLMS algorithm used for the identification. Finally, it has shown that the full model identification is not always necessary, but it is possible to adapt a non linear model to extract some important parameters such as THD, HD_k and IMD. A final observation is also related to the experimental setups adopted for all the experiments. Indeed, one of the most complex aspect was related to measurement setup, i.e., as explained in Section 2.2 vacuum tube amplifiers has several technical characteristics that must be taken into consideration. They require a precise impedance adaptation between the amplifier output and load, and since the output voltage is in the range of tens of volts the acquisition system could be damaged during the measurement. For these reasons several experiments have been carried out to find the best setup for each experiment.

Chapter 3

Room response identification and equalization

Another important real audio system is the room environment equipped with a reproduction system. In particular, the environment can modify the listening experience due to things like furniture, carpets, type and position of the walls that can introduce undesired artifacts (e.g., standing waves, frequency band extensions, non linearities) to the sound reproduction. For these reasons equalization algorithms could be used to compensate this unwanted behaviour allowing a better sound quality reproduction. Each equalization process has two fundamental steps: the room impulse response (RIR) measurement and the room response equalization (RRE). The first procedure allows a precise measurement of the acoustical behaviour of the environment, while the latter allows to compensate the unwanted phenomena introducing a digital filter in the reproduction chain. In the next chapter, two different equalization procedure will be presented and discussed.

3.1 Background

A wide variety of RRE methods have been studied over the last 40 years with the aim of improving the listening experience by compensating the room transfer function (RTF) from the loudspeaker system to the listener [27]. Different RRE approaches have been proposed in the literature: many of them consider fixed equalizers, designed a priori before the filtering operation [27, 38, 99, 144, 148].

However, the room can be considered as a time-varying environment (a "weakly non-stationary" system as defined in [99]), so the room response varies with position [144] and with time [99]. In audio processing, this is a very challenging task since many issues have to be examined [47]. Therefore, adaptive solutions should be adopted in order to track and correct slow variations in the room response due to temperature changes or movement of people or other obstacles. In particular, two types of room response (RR) adaptive equalizers can be found: single-point (single-input/single-output SISO, multiple-input/single-output MISO) [148] and multi-point (single-input/multiple-output SIMO, multiple-input/multiple-output MIMO) [27, 31] room equalizers.

The single-point room equalization filter is obtained from a measurement of the room response in a single position, so the room equalization is achieved only in a reduced zone around the measured point. Differently from single position case, the multi-point room equalizer is designed starting from several measurements of the RTF at different locations in order to expand the area of equalization.

While there is a huge amount of approaches to RRE [43], in what follows we review only the most relevant ones related to the proposed solution.

A first adaptive RRE system was proposed by Elliot et al. in [77], where the equalizer is designed by adaptively minimizing the sum of the squared errors between the equalized responses and the delayed input signal. A biased adaptive algorithm has been lately proposed in [91], in order to increase the convergence speed and improve the robustness of the adaptive identification algorithm in the case of a low signal to noise ratio. Unfortunately, this method requires an a-priori estimation of the impulse responses and it is not easy to obtain in many practical applications. In [47, 50], a technique suitable for RRE in current commercial hi-fi audio products is discussed. This method is derived from the technique presented in [88], where a single position RRE approach in the frequency domain is considered. The system of [88] allows the equalization by dividing the loudspeaker signal and the microphone signal in sub-bands and updating the filter weights in these sub-bands. This approach is interesting because it combines the robustness towards peaks and notches of the room responses, the simplicity of the implementation, and the capability of tracking RR variations. In [42, 46], the method of [88] was refined by first developing a multiple position solution and then by considering a warped frequency domain in order to expand equalization to the lower frequencies. However, the adaptive room response equalization approaches proposed in [42, 46, 88] investigate the equalization of a single sound source, i.e., of a single audio channel. Contrariwise, the technique in [50] presents a multi-channel solution that contemplates also the non-uniqueness problem. This problem happens in the stereophonic acoustic echo cancellation [22] because of the ill-conditioned covariance matrix. To solve the non-uniqueness problem, a method to minimize the interchannel coherence is usually applied [23]. For example, in [196], a simple method that uses positive and negative half-wave rectifiers is employed to reduce the coherence between the channels.Unfortunately, many of the techniques used to weaken the channel cross-correlation often introduce considerable distortions, which are unacceptable in the case of high quality sound reproduction systems [181]. Therefore, an appropriate technique, capable of decorrelating the input signals without modifying the audio quality, must be examined.

In [50], an adaptive multi-channel and multiple position RRE system has been briefly introduced. In [47] the same method has been completely detailed and expanded in order to provide a real-time implementation on commercial hi-fi products suitable for a stereo reproduction. The approach of [47] allows us to estimate the room responses with great accuracy by reducing the inter-channel coherence through a technique that produces only a modest degradation of the sound quality. The RR equalizer of [47] is designed starting from a prototype that is derived in the warped frequency domain to increase the equalization in the low frequency spectrum and to simultaneously reduce the computational cost of the filters design. Unlike other techniques presented in the literature, the system discussed in [47] is able to solve the problem of multi-channel and multiple position room response equalization for time varying environments by applying a simple and computationally efficient solution and without introducing artefacts. In [48], an adaptive multi-channel and multiple position room response equalization system is reported. In particular, an adaptive and accurate estimation of the room responses is provided introducing a normalized least mean square optimization approach with a variable step-size, and taking advantage of a multi-channel interchannel coherence reduction technique based on the missing fundamental phenomenon. Then, the equalizer is designed with the introduction of a novel prototype function derived from the combination of quasi-anechoic

impulse responses with the impulse responses recorded in the real environment to be equalized.

3.2 Room impulse response identification

The first step for equalization is the measurement of the room impulse response $h(n)$. This function contains all the acoustical parameters of the environment which has to be analyzed. For this reason is clear the necessity of a reliable and accurate procedure able to measure these responses. A typical setup for impulse response measurement is visible in Figure 3.1.

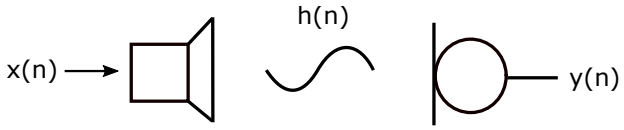
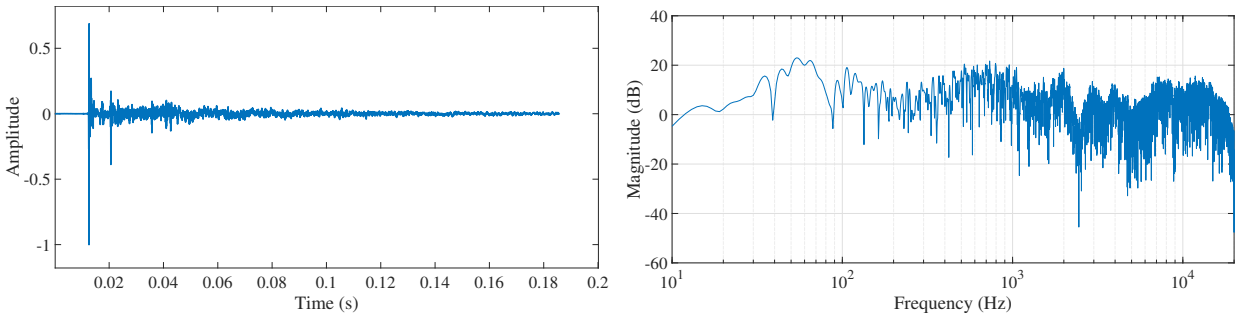


Figure 3.1: Schematic of impulse response measurement setup, the input signal $x(n)$ is generated and then reproduced by means of an amplifier and a loudspeaker, the signal travel through the environment and is modified by its impulse response $h(n)$, finally the output signal $y(n)$ is acquired with a microphone.

A source (i.e., an amplifier and a loudspeaker) generates the test signal $x(n)$, then the signal travels within the environment and finally, the signal $y(n)$ is acquired through a microphone. From the knowledge of the original signal $x(n)$ and the output $y(n)$, it is possible to estimate the room impulse response $h(n)$.



(a) (b)

Figure 3.2: Real room impulse response (a), and its frequency response (b), the impulse response gives important information of the room reflections and the time of propagation of the sound, while the frequency response shows how the frequencies are attenuated from the environment.

During years, many algorithms have been proposed for an efficient and accurate estimation of $h(n)$, among these, two of the most widely used approaches are based on maximum length sequence (MLS) and sine sweep (SS) identification. Both these techniques allow an accurate and reliable RIR estimation, but require that the system is linear, otherwise the estimation could be significantly degraded. For this reason, if one of the system elements shows a non linear behaviour, more advanced techniques are necessary [39]. However, in many real acoustic environment, the linear hypothesis is satisfied and MLS and SS could be successfully adopted for RIR measurements.

3.2.1 Maximum length sequences

Maximum length sequences are a class of binary pseudo random sequences composed of $+1$ and -1 values. These sequences could be easily generated with a shift register and they allow an identifi-

cation of the impulse response through a simple convolution operation [29]. The cross correlation between the input $x(n)$ and the output $y(n)$ for the linear systems is related to the auto correlation of the input with the system impulse response $h(n)$, i.e.,

$$R_{xy}(n) = R_{xx}(n) * h(n). \tag{3.1}$$

An important property of the MLS is that the autocorrelation function $R_{xx}(n)$ is an impulse represented by the Dirac delta function:

$$R_{xx}(n) = \delta(n). \tag{3.2}$$

Combining equation 3.1 and 3.2 the system impulse response $h(n)$ can be estimated from the cross correlation between $x(n)$ and $y(n)$:

$$R_{xy}(n) = \delta(n) * h(n) = h(n). \tag{3.3}$$

For MLS signal, the cross correlation can be calculated efficiently by means of the the fast Hadamard transform, making MLS an easily and accurate method for RIR estimation.

3.2.2 Sine sweep

The RIR estimation through sine sweep signals is a notable alternative to MLS measurements. With respect to MLS, the SS identification is more robust against time variance and distortions [145]. Sine sweeps are time varying signals composed of a pure tone which changes its frequency as a function of time. Depending on how the frequency changes, different sweeps are available, e.g., the linear sweep (the sweep frequency changes as a linear function of t) or logarithmic sine sweep (the sweep frequency changes as a logarithmic function of t). As reported in [145] with a sine sweep it is possible to estimate the RIR using different approaches, e.g., the ratio technique which is based on the ratio between the Fourier transform of sweep input signal and the Fourier transform of the output signal, or by using more advanced technique like the time delay spectrometry (TDS) [101]. However, one of the most adopted algorithm is based on the application of an inverse filter with a procedure similar to the one described in Chapter 2 for the identification of generalized Hammerstein system. The schematic of this algorithm is visible in Figure 3.3(b).

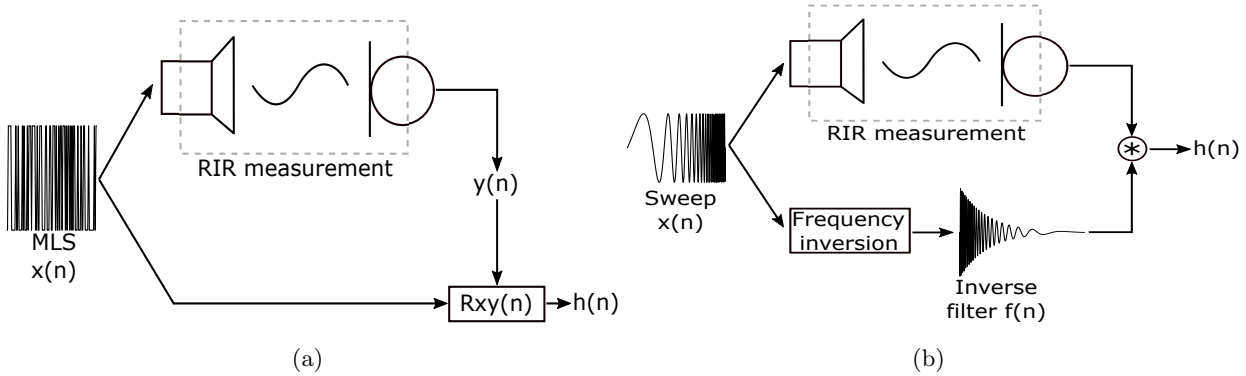


Figure 3.3: Schematic of impulse response measurement using a MLS (a), and a sine sweep (b)

.

Considering a logarithmic sine sweep as input signal, the sweep can be defined according to:

$$x(n) = sin\left[\frac{T\omega_1}{ln\left(\frac{\omega_1}{\omega_2}\right)}\left(e^{\frac{n}{T}ln\left(\frac{\omega_1}{\omega_2}\right)} - 1\right)\right], \tag{3.4}$$

where ω_1 and ω_2 are respectively the start and stop angular frequencies of the sweep and T is the duration. For the impulse response estimation it is necessary the derivation of an inverse filter $f(n)$ whose convolution with the sine sweep is a Dirac delta, i.e., $x(n) * f(n) = \delta(n - K)$, where K is a delay. $f(n)$ can be derived from the sine sweep with a simple inversion in the frequency domain [145]. Now, the impulse response can be estimated by a simple convolution:

$$h(n) = y(n) * f(n). \tag{3.5}$$

3.3 Equalization algorithm

Once the impulse response has been measured, the next step is the equalization. According to Section 3.1, this process can consider a single-point equalization (i.e., a singe RIR is measured and then used for the equalization procedure) or a more advanced multi-point approach (i.e., multiple RIR are measured and then used within the equalization process). In the next Section, the multi-point algorithm used in this work will be explained and discussed.

3.3.1 Multi-point equalizer

The adopted equalization technique is presented in [38], and has been considered due to its capability of enlarging the listening sweet spot taking into consideration different microphones positions. The considered multi-point technique takes into consideration a quasi-anechoic approach [18], thus producing also a general equalization of the used loudspeakers. The prototype function is derived from the combination of quasi-anechoic IRs, derived from a gated version (up to the first reflection) of the responses, with the IRs recorded in the real environment.

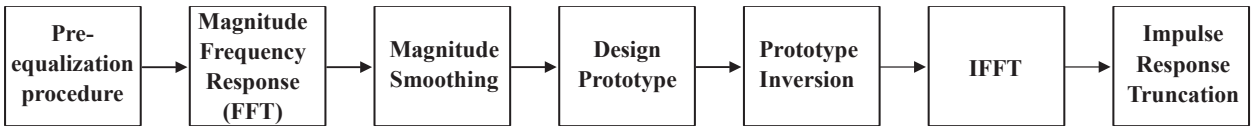


Figure 3.4: Scheme of multi-point equalization algorithm

Figure 3.4 shows the equalization approach used in the presented work that follows these steps: 1. Starting from a set of impulse responses derived in the zone to-be-equalized, a pre-processing is considered exploiting the quasi-anechoic IR spectrum for frequency greater than a certain transition frequency and the original (ungated) IR spectrum below the same transition frequency. For simplicity of representation in this sub-Section we work in the DFT domain and the operation is performed by applying the following equation to each RIR:

$$H_{m,l}(\kappa) = \widehat{H}_{m,l}(\kappa) \cdot w_{\mathrm{lf}}(\kappa) + \widetilde{H}_{m,l}(\kappa) \cdot w_{\mathrm{hf}}(\kappa), \tag{3.6}$$

where $\kappa = 0, \cdots, K-1$ indicates the frequency bin, $\widehat{H}_{m,l}(\kappa)$ is the frequency response of the original

estimated RIR, $\widetilde{H}_{m,l}(\kappa)$ is the frequency response of the gated RIR, $w_{\mathrm{lf}}(\kappa)$ and $w_{\mathrm{hf}}(\kappa)$ are the half Hann windows used for selecting the low-pass and high-pass frequency bands, respectively. The linear combination in Eq. (3.6) is used to equalize only the direct sound in mid-high frequency range, which determines localization and most of timbre perception, while full equalization is applied in the modal frequency range [50].

2. Then, starting from the magnitude spectra of the IRs, a smoothing operation is applied, simulating the poorer frequency resolution at higher frequencies of the human auditory system. The approach of [98] has been used to obtain a non uniform frequency magnitude spectrum smoothing of the frequency response $H_{m,k}(\kappa)$ for all m and l. In this way, a broader equalized zone is achieved exploiting a less precise equalization at higher frequencies resulting from the non-uniform resolution, which decreases increasing the frequency.

3. At this point, a representative response of the considered acoustic environment is derived taking into account all smoothed IRs. The prototype frequency response is obtained using an arithmetic mean of the zero-phase smoothed frequency responses, as follows:

$$H_{\mathrm{pr}_l}(\kappa) = \frac{1}{M}\sum_{m=1}^{M} H_{\mathrm{sm}_{m,l}}(\kappa), \tag{3.7}$$

with $\kappa = 0, \cdots, K-1$ and $H_{\mathrm{sm}_{m,l}}(\kappa)$ the smoothed frequency response from loudspeaker l to microphone m.

4. Finally, a frequency domain inverse filter is obtained through the use of a frequency deconvolution with regularization technique [122] applied to the prototype as follows:

$$H_{\mathrm{inv}_l}(\kappa) = \frac{H^*_{\mathrm{pr}_l}(\kappa)}{\left|H_{\mathrm{pr}_l}(\kappa)\right|^2 + \beta}, \tag{3.8}$$

being $\kappa = 0, \cdots, K-1$, $H^*_{\mathrm{pr}_l}(\kappa)$ the complex conjugate of $H_{\mathrm{pr}_l}(\kappa)$, and β the regularization factor, which allows us to avoid excessive gains often appearing at high frequencies. In the experimental results presented in Section 5.3.3, a small regularization factor with value 10^{-5} is considered. The equalization filter frequency response of length W is computed in the unwarped domain by interpolating with a cubic spline [42] the K values of $H_{\mathrm{inv}_l}(\kappa)$.

5. After that, the inverse FFT of the interpolated frequency response is performed.

6. Finally, the resulting sequence is truncated in order to determine the equalization filter.

Therefore, it results a valid approach for a real-time application since its computational complexity is mostly determined by the inverse FFT, that is an $O(W \log W)$ algorithm [122].

3.4 Proposed approaches

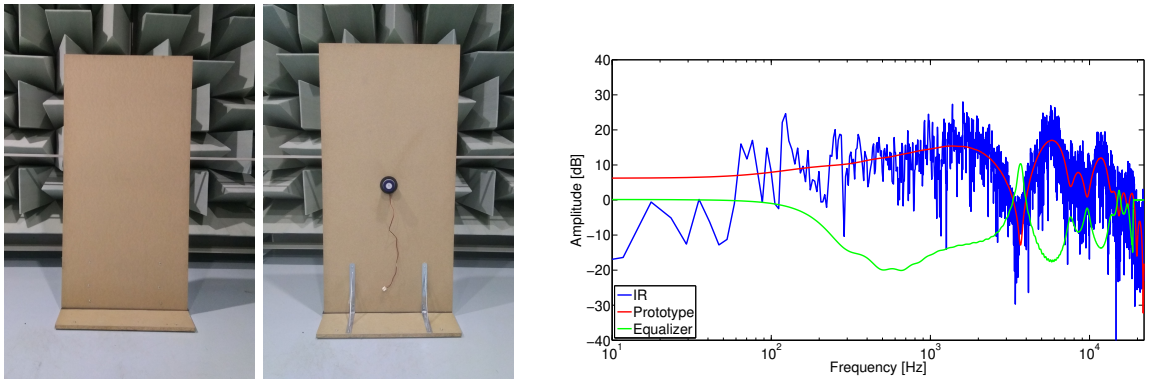
Focusing on equalization, two works are presented in this chapter. The first involves the application of the previous exposed multi-point equalization algorithm to the problem of reproduction with vibrating transducers, the latter involves the application of the equalizer in a more complex structure which realizes an adaptive multi-point and multi-channel equalization using a computational efficient sub-band structure.

3.4.1 Applying sound equalization to vibrating sound transducers mounted on rigid panels

During recent years, loudspeakers manufacturers have proposed to the market vibrating transducers that can be installed on a surface to be transformed in a speaker capable of delivering music or speech. These systems are very appreciated by the market and by interior designers for the possibility of hiding transducers and eliminating any impact on aesthetics [213]. For this reason, they are commonly called "invisible sound speakers" and they are used in several applications from basic public address to home entertainment [210]. However, due to several issues, such as surface attributes (e.g., stiffness, size), surface anchoring structure, position and number of the transducers, the audio frequency response of the final system is typically not flat. In this scenario, the use of audio equalization can improve the performance of these systems starting from their characterization in terms of impulse responses measurement.

Experimental results

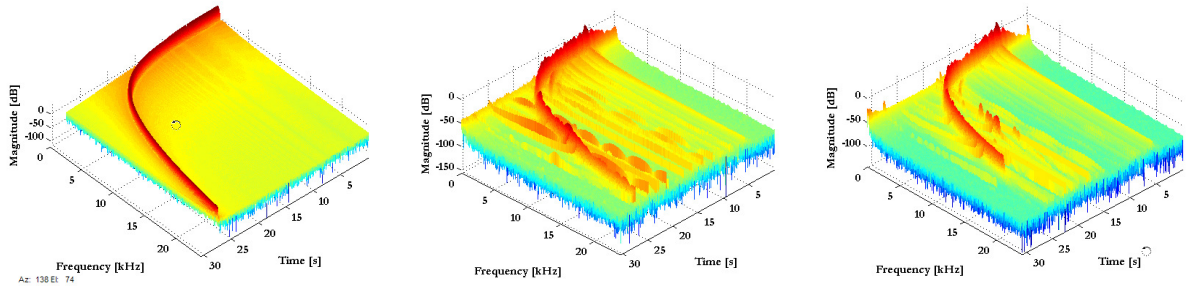
For the experiments, the device has been mounted mounted on a wood surface of dimension $150x75$ cm as visible in Figure 3.5(a). Four impulse responses have been acquired at a distance of 1.5 m. The multi-point equalization has been applied in the range of $100 - 16000$ Hz and the results are reported in Figure 3.5(c) showing one of the impulse responses measured, the prototype and the final equalization curve.



(a) (b) (c)

Figure 3.5: (a),(b) shaker mounted on a wood panel, and (c) multi-point equalization results taking into consideration one of the measured impulse response, the prototype function and the final equalizer.

Objective results A time-frequency analysis in terms of Energy Decay Relief (EDR) has been reported in Figure 3.6. The EDR is defined as the time-frequency representation of the signal energy decay and it was calculated as reported in [113, 114]. A sweep sine signal was used as input of the reproduction system (Figure 3.6(a))). Then, measurements with (Figure 3.6(b)) and without (Figure 3.6(c)) the equalization procedure have been acquired. It is evident that the equalization procedure is capable of reducing most of the panel resonances. However, a small contribution of new resonances appears probably due to some non-linearities of the panel that the equalization is not able to compensate.



(a) (b) (c)

Figure 3.6: Multi-point equalization results: (a) Original sweep, (b) Recorded sweep, (c) Recorded sweep considering the equalization procedure.

Subjective results To assess the overall audio quality perception of the equalization procedure, subjective listening tests were performed according to the ITU-R BS.1284-1 [5] which provides guidelines for the general assessment of sound quality. Following the recommendations, the subjects involved in the listening tests were 10 expert listeners (5 males and 5 females, ages from 21 to 35) with a technical background in acoustics. The subjective tests have been performed considering the setup of Fig. 3.5. Before the listening test, a training test was presented to the listeners to get familiar with the test procedure, the test materials, and the test environment, as suggested in [5]. All the subjects had the possibility to listen to each audio item in all conditions under evaluation.For the proposed test session, a signal equalized with multi-point procedure was compared with the original signal without equalization (i.e., the reference signal). For the listening test, the subjects were instructed to make a paired comparison between the original sound track (i.e., the reference signal) and the equalized stimulus under test Table 4.2(b) shows the bipolar scale value used for the comparison. The selected attributes suggested by [5] to be evaluated are the following:

Main impression : "The integrity of the total sound image and the interaction between the other parameters;"

Spatial impression : "The performance appears to take place in an appropriate spatial environment;"

Transparency : "All details of performance can be clearly perceived."

The most important attribute is the basic audio quality based on the Main Impression that includes all aspects of the sound quality to be assessed. According to [5], the score was given using the continuous quality scale divided into six equal intervals (with a recommended resolution of 1 decimal place) [38]. Each stimulus was 20 s long, as suggested in the recommendation, and the listeners had the possibility to repeat the presentation. The same listening level has been used for all processed items, setting the SPL of reproduction to 70dB. The selected signals for the subjective evaluation are reported in Table 3.1(a). The same professional audio equipment used for the measurement of the impulse responses exploited for multi-point equalization has been used for the reproduction of each stimulus. As suggested in [5], the subjective data have been processed to derive the mean values and the confidence intervals. Figure 3.7 summarizes the obtained results taking into consideration the four different musical genres shown in Table 3.1(a): the mean grade

Table 3.1: (a) List of sound tracks used for the listening tests, and (b) bipolar scale values adopted for the subjective experiments.

(a)

Genre	Author	Sound Track
Pop	M.Biondi	I Know It's Over
Jazz	H. Hancock	Edith and the Kingpin
Country	J.Cash	Cry, Cry, Cry
Soul	Seal	A change is gonna come

(b)

Value	Meaning
3	Much better
2	Better
1	Slightly better
0	The same
-1	Slightly worse
-2	Worse
-3	Much worse

values, computed over all involved subjects are plotted with vertical colored blocks, while the confidence intervals calculated with a significance level of 0.05 are given by vertical black lines. The results show an improvement of the performance produced by the proposed approach confirming the necessity of a multi-point equalization in case of vibrating sound transducers. In particular, for "Main impression" attribute, best results have been achieved with the soul track, while for "Spatial Impression" attribute very similar results have been recorded for all tracks. "Transparency" attribute has been obtained very good results in the country track.

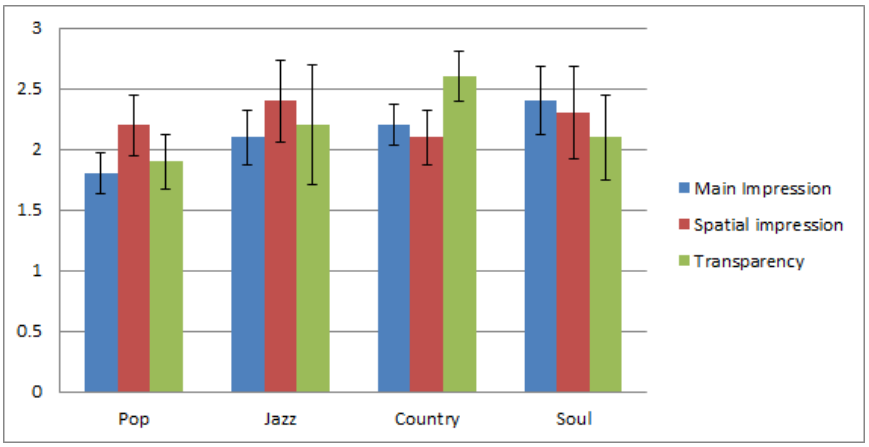


Figure 3.7: Results for subjective listening tests taking into consideration three attributes and the tracks of Table 3.1(a).

3.4.2 Sub-band adaptive multi-point multi-channel equalizer

In this Section, an evolution of the technique proposed in [47, 48] is presented, improving the identification procedure. In particular, the proposed system is designed for providing an iterative estimation of RIRs and, at the same time, a multi-point equalization. The system repeats the approach presented in [48] with a variation in the room response identification that is performed considering a multirate sub-band approach in order to reduce the computational cost and to improve the convergence speed. The block diagram of the proposed approach is shown in Figure 3.8.

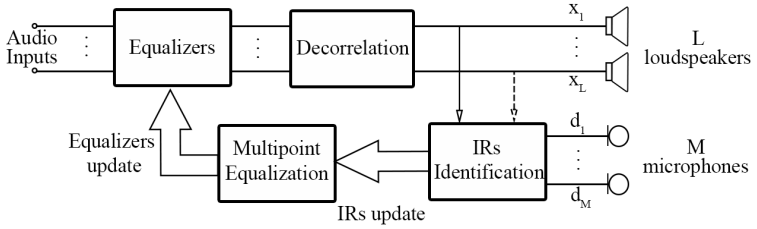


Figure 3.8: Block diagram of the proposed solution.

In the case of more than one loudspeaker, convergence problems can occur due to the channels correlation. These problems are very common in multi-channel acoustic echo cancellation because of the possible errors in the identification of the acoustic paths [189, 195]. In this context, a method to reduce the inter-channel coherence must be exploited, as described in [24]. Furthermore, the identification of room responses is achieved by a sub-band adaptive filtering using the structure presented in [166]. In this way, $L \cdot M$ room responses between the L loudspeakers and the M microphones are estimated and then exploited for the equalizer design. The input signal decorrelation and the sub-band room response identification are described in the following Section, while the multi-point equalizer has already been described in Section 3.3.

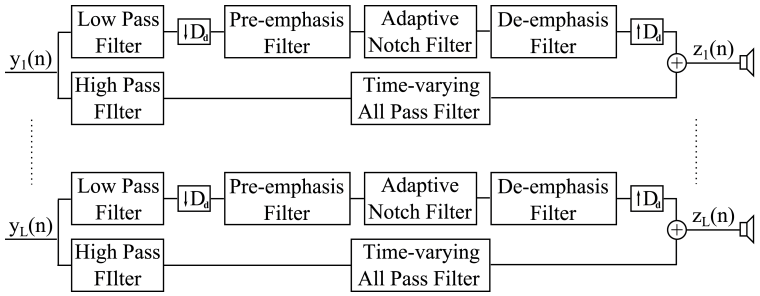


Figure 3.9: Multi-channel decorrelation procedure.

Input decorrelation The multi-channel input signals decorrelation is obtained by the psychoacoustic criteria of the missing fundamental [128], as reported in [182]. This psychoacoustic phenomenon is associated with the human capability of perceiving the fundamental frequency although it is not actually present in the signal. Moreover, second-order time-varying all-pass filters are included in this approach so as to expand the solution to the entire frequency spectrum [182]. The scheme of the algorithm is reported in Figure 3.9, in which each input channel $y_p(n)$ of the decorrelation block, being $p = 1, \cdots, L$, is divided into two sub-bands [48]. Hence, after a decimation operation of D_d, an adaptive notch filter $H_p(z, n)$ is applied in the lower frequencies while a second-order time-varying all-pass filter $F_p(z, n)$ is applied in the high-frequency spectrum caus-

ing an alteration of the signal phase. This method allows us to accurately identify each processed channel $z_p(n)$. In the low-frequency band, the notch filters are created by L second-order lattice structures in order to remove the L fundamental frequencies. The p-th notch filter is described as follows:

$$H_p(z,n) = \frac{1 + 2k_p(n)z^{-1} + z^{-2}}{1 + k_p(n)[1 + \alpha_p(n)]z^{-1} + \alpha_p(n)z^{-2}}, \tag{3.9}$$

where $p = 1, \cdots, L$ represents the channel index, $k_p(n)$ is the adaptive coefficient connected to the tracked frequency $f_p(n)$ and $\alpha_p(n)$ is the pole-zero contraction factor that controls the bandwidth of the filter [130]. The cut-off frequency of the notch filter controlled by $\alpha_p(n)$ must change in time so as to identify and delete the fundamental frequency. Thus, decorrelation is guaranteed in the whole low-frequency band providing results similar to a set of time-varying all-pass filters [12]. The contraction factor $\alpha_p(n)$ is related to each channel and the time-varying vector $\boldsymbol{\alpha}(n) = [\alpha_1(n), \cdots, \alpha_L(n)]$ provides disparity among channels even if the fundamental frequency is the same. In particular, a right circular shift of one sample $s(\cdot, 1)$ is applied to the vector $\boldsymbol{\alpha}(n)$ every Q samples [182], as explained in the following equation:

$$\boldsymbol{\alpha}(n) = \begin{cases} s[\boldsymbol{\alpha}(n-1), 1] & \text{if } \left(n - Q\left\lfloor \frac{n}{Q} \right\rfloor\right) = 0 \\ \boldsymbol{\alpha}_p(n-1) & \text{otherwise.} \end{cases} \tag{3.10}$$

The adaptive coefficient $k_p(n)$ is included in the interval $(-1, 1)$ to avoid the divergence of the filter and represented by the following sigmoid function:

$$k_p(n) = \frac{2}{1 + e^{-\gamma_p(n)}} - 1, \tag{3.11}$$

being $\gamma_p(n) \in R$. The tracking of the fundamental frequency is obtained by finding $\gamma_p(n)$ that minimize the output energy of the filter in (3.9), as described in [49]. In this way, the filter of (3.9) is completely determined and it is capable of removing the fundamental frequency. It is possible to derive this fundamental frequency $f_p(n)$ with the knowledge of the sampling frequency f_s and of the down-sampling factor D_d, considering the following equation:

$$f_p(n) = \frac{f_\mathrm{s}}{D_\mathrm{d}} \cdot \frac{1}{2\pi} \cos^{-1}[-k_p(n)]. \tag{3.12}$$

Moreover, the low-frequency band could contain also some harmonics, so a pre-emphasis filter $H_\mathrm{pre}(z)$ is used to emphasize the low-frequency in order to improve the method and its effect is cancelled by the de-emphasis filter $H_\mathrm{de}(z)$. The filters are described by the following mathematical functions:

$$H_\mathrm{pre}(z) = \frac{1}{1 - \nu z^{-1}} \tag{3.13}$$

$$H_\mathrm{de}(z) = 1 - \nu z^{-1}, \tag{3.14}$$

being $0 < \nu < 1$. Considering the high-frequency range, the phase of the input channels is changed by the application in each channel of L second-order time-varying all-pass filters. The transfer

function of the p-th all-pass filter is described by the following equation [158]:

$$F_p(z,n) = \frac{k_p^2(n) - 2k_p(n)z^{-1} + z^{-2}}{1 - 2k_p(n)z^{-1} + k_p^2(n)z^{-2}}, \tag{3.15}$$

so, it is identified by a pole with a multiplicity of 2 connected to the coefficient $k_p(n)$ of Eq. (3.11). This characterization of the all-pass filter allows us to maintain the spatial perception of the speech [12] because the restriction $|k_p(n)| < 1$ guarantees stability and causality of the filter and ensures that the inter-aural time delay difference between the two ears is lower than the well-known "just noticeable inter-aural delay" [219]. As described in [182], the alteration in sound direction is negligible as the maximum variation in the time of arrival is about $40\,\mu$s for all frequencies.

Sub-band Room Response Identification The architecture used for the room response identification is based on a sub-band adaptive filtering structure with critical sampling [166]. Considering D the number of sub-bands and $h_{\mathrm{p}}(n)$ the impulse response of a prototype filter of order N_h that guarantees perfect reconstruction when used in a cosine modulated analysis-synthesis system [166], the analysis and synthesis filters are obtained as follows,

$$g_k(n) = 2h_{\mathrm{p}}(n)\cos\left[\frac{\pi}{D}(k+0.5)\Big(n - \frac{N_h}{2}\Big) + \vartheta_k\right], \tag{3.16}$$

$$f_k(n) = 2h_{\mathrm{p}}(n)\cos\left[\frac{\pi}{D}(k+0.5)\Big(n - \frac{N_h}{2}\Big) - \vartheta_k\right], \tag{3.17}$$

where $\vartheta_k = (-1)^k\frac{\pi}{4}$, for $0 \le k \le D-1$ and $0 \le n \le N_h$. Figure 3.10 shows the sub-band structure considering $L = 2$ loudspeakers and one of the microphones. In Figure 3.10, **GG** is the double analysis filter-bank, which is derived from the uniform cosine modulated filter-bank **G**, and **F** is the corresponding synthesis filter-bank. The double analysis filter-bank **GG** is composed by the D filters $G_k(z)G_k(z)$ for $k = 0,\ldots,D-1$, and the $D-1$ filters $G_k(z)G_{k+1}(z)$ for $k = 0,\ldots,D-2$, with the following impulse responses [167]:

$$g_k(n) * g_k(n) \approx 2[h_{\mathrm{p}}(n) * h_{\mathrm{p}}(n)]\cos\left[\frac{\pi}{D}(k+0.5)\Big(n - \frac{N_h}{2}\Big) + 2\vartheta_k\right], \tag{3.18}$$

$$g_k(n) * g_{k+1}(n) \approx 2q_0(n)\cos\left[\frac{\pi}{D}(k+0.5)\Big(n - \frac{N_h}{2}\Big)\right], \tag{3.19}$$

where

$$q_0(n) = \left[h_{\mathrm{p}}(n)e^{j\frac{\pi}{2D}n}\right] * \left[h_{\mathrm{p}}(n)e^{-j\frac{\pi}{2D}n}\right]. \tag{3.20}$$

As a consequence, considering an input signal $x_l(n)$ with $l = 1, ..., L$ the the loudspeaker index, the outputs of the filter-bank after the downsampling operation are $2D-1$ signals derived as follows in Z-domain:

$$X_{l,k,k}(z) = X_l(z^{\frac{1}{D}})G_k(z^{\frac{1}{D}})G_k(z^{\frac{1}{D}}) \tag{3.21}$$

for $k = 1,\ldots,D-1$, and

$$X_{l,k,k+1}(z) = X_l(z^{\frac{1}{D}})G_k(z^{\frac{1}{D}})G_{k+1}(z^{\frac{1}{D}}) \tag{3.22}$$

for $k = 1,\ldots,D-2$. These signals constitute the inputs of a bank of adaptive filters. Let us denote

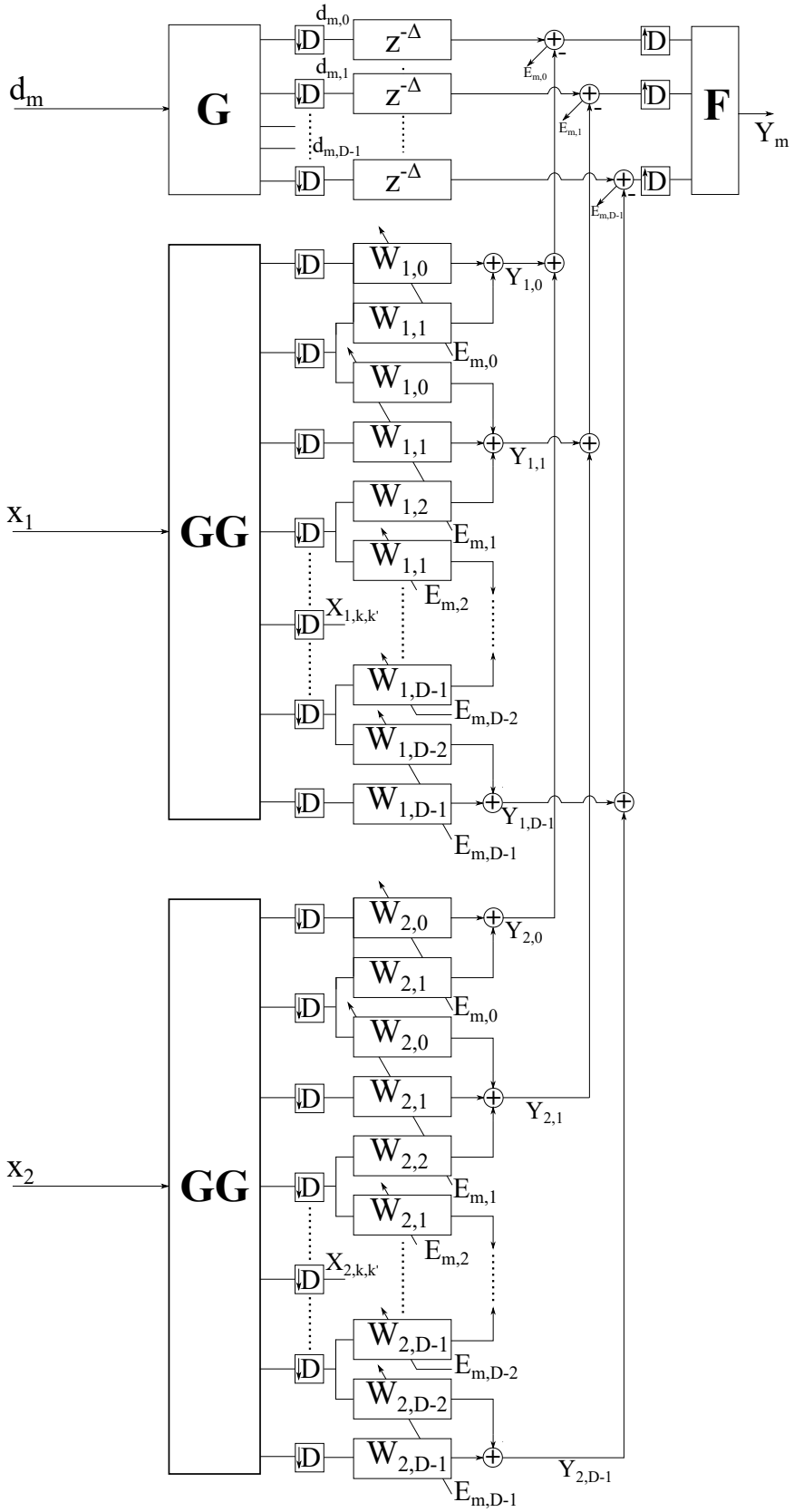


Figure 3.10: Sub-band IRs identification procedure with $L = 2$ and $M = 1$.

with ξ the downsampled time index, with $\mathbf{x}_{l,k,k}(\xi)$ and $\mathbf{x}_{l,k,k+1}(\xi)$ the input vectors formed by the samples of signals in (3.21) and (3.22), respectively, and with $\mathbf{w}_{m,l,k}(\xi)$ the vectors collecting the

coefficients of the adaptive filters from input l to microphone m. Each adaptive sub-filter should have at least $K = (\frac{N+N_h+2}{D} + 1)$ coefficients [166] with a uniform frequency bandwidth of π/D and a center frequency of $\pi/(2D)$. The sub-filters are obtained through the minimization of the sum of the instantaneous sub-band squared-errors, represented by:

$$J_m(\xi) = \sum_{k=0}^{D-1} e_{m,k}^2(\xi), \tag{3.23}$$

where $m = 1, ..., M$ and M is the number of microphones. The error signals are obtained by the following relation

$$e_{m,k}(\xi) = d_{m,k}(\xi - \Delta) - \sum_{l=1}^{L} y_{m,l,k}(\xi), \tag{3.24}$$

where $d(\cdot)$ is the desired signal, $\Delta = \frac{N_h+1}{D}$ is the delay introduced by the filter bank $\mathbf{G}$, and

$$y_{m,l,k}(\xi) = \mathbf{x}_{l,k,k}^T(\xi)\mathbf{w}_{m,l,k}(\xi) + \mathbf{x}_{l,k-1,k}^T(\xi)\mathbf{w}_{l,m,k-1}(\xi) + \mathbf{x}_{l,k,k+1}^T(\xi)\mathbf{w}_{l,m,k+1}(\xi). \tag{3.25}$$

The filters $\mathbf{w}_{m,l,k}$ are adapted according to the following formula:

$$\begin{aligned}\mathbf{w}_{m,l,k}(\xi+1) = \mathbf{w}_{m,l,k}(\xi) + \mu_{l,k}(\xi)[&\mathbf{x}_{l,k,k}(\xi)e_{m,k}(\xi) \\ &+ \mathbf{x}_{l,k-1,k}(\xi)e_{m,k-1}(\xi) + \mathbf{x}_{l,k,k+1}(\xi)e_{m,k+1}(\xi)].\end{aligned} \tag{3.26}$$

The step sizes $\mu_{l,k}(\xi)$ are normalized by the sum of instantaneous powers of the signals involved in the coefficients adaptation, i.e.,

$$\mu_{l,k}(\xi) = \frac{\mu}{\delta + P_{l,k,k}(\xi) + P_{l,k-1,k}(\xi) + P_{l,k,k+1}(\xi)}, \tag{3.27}$$

where δ is a small positive constant to avoid division by zero and the power is estimated as follows:

$$P_{l,k,k'}(\xi+1) = \beta P_{l,k,k'}(\xi) + (1-\beta)X_{l,k,k'}^2(\xi), \tag{3.28}$$

with β a constant in the range $(0, 1)$. Finally, the reconstructed frequency response between the l-th loudspeaker and the m-th microphone $\widehat{H}_{m,l}(z)$ is computed as the sum of the interpolated sub-filters $W_{m,l,k}(z^D)$, filtered by the synthesis filter-bank $\mathbf{F}$, i.e.,

$$\widehat{H}_{m,l}(z) = \sum_{k=0}^{D-1} W_{m,l,k}(z^D)F_k(z). \tag{3.29}$$

The computational complexity of the sub-band identification algorithm is lower than a full band LMS approach as reported in [166]. In particular, the overall number of multiplications per input sample required for the filtering and adaptation of the sub-filters $W_{m,l,k}(z^D)$ is computed as follows:

$$\frac{2(3D-2)N_p}{D^2} + \frac{2(3D-2)N_h}{D^2}, \tag{3.30}$$

where the first term corresponds to the filtering operations and the second term corresponds to the adaptation procedure with N_p the length of the full band system. For high-order adaptive filters,

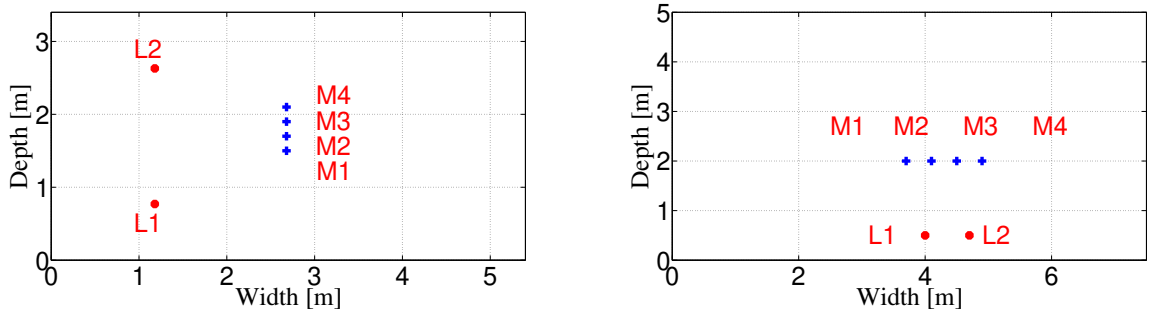
the dominant term in the above expression is $6N_p/D$, which is about $D/3$ times smaller than the number of multiplications required by the fullband LMS algorithm ($2N_p$) as reported in [166].

Experimental results

Some experimental results are reported in this Section to show the effectiveness of the proposed approach. The performance of the decorrelation algorithm has not been included for space limitations and also because its effectiveness was fully demonstrated in [182]. Also the effectiveness of the multi-point approach has been demonstrated in several papers of the same authors, e.g., [43, 47, 48]. Several tests have been done considering impulse responses measured two real rooms. Rooms size with loudspeakers and microphones positions are shown in Figures 3.11 and 3.11(b): the microphones distance was set to 30 cm and they were placed at 1.2 m height. Professional equipment was used following the procedure described in [48]. More in details, measurements have been performed using a professional ASIO sound card and microphones with an omnidirectional response. A personal computer running NU-Tech platform has been used to manage all I/Os [129]. The impulse responses have been derived using a logarithmic sweep signal excitation [45] at 48 kHz sampling frequency. These responses are then used as terms of comparison in the identification procedure. For the adaptation procedure a filter length L_f of 4096 samples has been considered, working on blocks of $N = 8192$ samples, with $D = 256$ and $f_s = 48$ kHz. The equalizer is designed in the warped domain considering $W = 8192$ frequency points and the final length of the equalizer is 1024 samples. The equalized frequency range goes from 10 Hz to 20 kHz with the same sampling frequency of 48 kHz. The stereo input signals used for the presented results are the following sound tracks:

- "International Geophysical Year" from Donald Fagen,
- "I Sat by the Ocean" from Queens Of The Stone Age,

corresponding to Experiment 1/3 and 2/4 respectively. Two songs have been chosen as input signals in order to show the algorithm performance considering a real scenario and variable inputs. This is a very important aspect since the identification algorithm must work during the sound reproduction without altering the sound perception. Some audio samples of these experiments can be downloaded at [44].



(a) (b)

Figure 3.11: Loudspeakers and microphones positions for ((a)) experiments 1 and 2 and ((a)) experiments 3 and 4.

Results on sub-band Identification procedure Figure 3.13 shows the four room transfer functions relative to the right loudspeaker channel considering the first sound track and room 1 (e.g., Experiment 1), Figure 3.14 shows the four room transfer functions relative to the left loudspeaker channel for the second sound track and room 1(e.g., Experiment 2), Figure 3.15 shows the four room transfer functions relative to the right loudspeaker channel considering the first sound track and room 2 (e.g., Experiment 3), and finally Figure 3.16 shows the four room transfer functions relative to the left loudspeaker channel for the second sound track and room 2 (e.g., Experiment 4). The impulse responses identified with a logarithmic sweep signal procedure are compared with the responses obtained with the sub-band identification procedure considering $D = 1$ and $D = 256$ sub-bands. The good performance obtained in terms of identification are also confirmed by the results reported in Figures 3.17, 3.18, 3.19, 3.20 where a frequency difference of the impulse responses identified with the sub-band procedure and the impulse responses obtained with the sweep identification are reported. It is evident that the sub-band structure is capable of identifying the impulse responses and, increasing the number of sub-bands, it is possible to obtain a very accurate estimation of the responses. This is evident for all experiments thus considering different inputs. Furthermore, the use of the sub-band structure allows us to obtain a quicker convergence increasing the number of sub-bands D, as shown in Figure 3.12. This is due to the fact that increasing the number of sub-bands, the signal is divided in small frequencies parts where it is more stationary and where it is possible to use a more suitable step size exploiting Eqs. (3.27) and (3.28). However, increasing the number of sub-bands could lead to an increase of the memory usage of the hardware system, increasing the parallel computational load. A compromise between identification performance and hardware capability should be evaluated.

Results on Equalization procedure Figure 3.21, 3.22, 3.23, and 3.24 show the results for the four experiments in terms of smoothed real room magnitude responses identified with the sub-band procedure, the prototype responses and the equalization curves obtained from the single band identification and the sub-band identification. It is evident that there are several differences due to a different accuracy in the identification of the responses considering the sub-band procedure. This implies also a different resolution of the equalization curve that can increase the quality of final equalization procedure. This is also demonstrated considering the spectral deviation (SD) measure as reported in Table 3.2 and 3.3. The SD gives a measure of the deviation of the magnitude response from a flat one [27], considering each IR before and after the equalization as also reported in [43]. It is evident that using the sub-band identification procedure, it is still possible to obtain a reduction of the SD, especially considering the results obtained in the low frequency range, i.e., in Table 3.3.

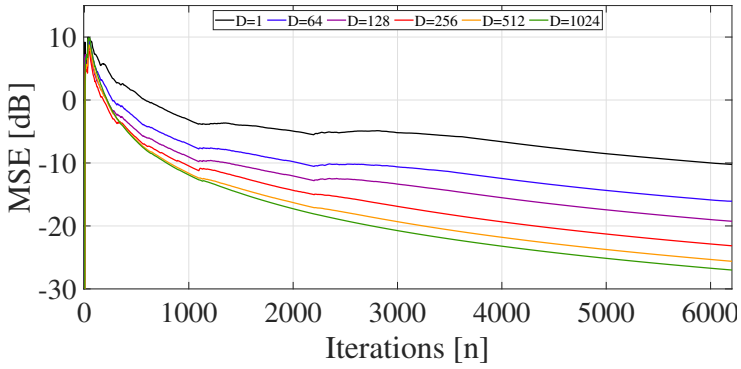


Figure 3.12: MSE for D=1, 64, 128, 256, 512, 1024, considering white noise as input signal.

Table 3.2: SD evaluation considering the single band identification ($D = 1$) and the sub-band identification ($D = 256$) for both experiments with a frequency range of $10 - 20000$ Hz.

Experiments sub-bands Number Frequency Range	Initial SD	Final SD $D = 1$ 10Hz-20kHz	Final SD $D = 256$ 10Hz-20kHz
EX1 - right channel	10.96	2.60	**2.59**
EX1 - left channel	9.50	2.66	**2.65**
EX2 - right channel	10.96	2.60	**2.59**
EX2 - left channel	9.50	2.66	**2.65**
EX3 - right channel	10.41	2.65	**2.64**
EX3 - left channel	12.17	2.59	**2.58**
EX4 - right channel	10.41	2.65	**2.64**
EX4 - left channel	12.17	2.59	**2.58**

Table 3.3: SD evaluation considering the single band identification ($D = 1$) and the sub-band identification ($D = 256$) for both experiments with a frequency range of $10 - 200$ Hz.

Experiments sub-bands Number Frequency Range	Initial SD	Final SD $D = 1$ 10Hz-200Hz	Final SD $D = 256$ 10Hz-200Hz
EX1 - right channel	4.09	3.98	**3.79**
EX1 - left channel	3.91	2.81	**2.64**
EX2 - right channel	4.09	3.90	**3.79**
EX2 - left channel	3.91	2.65	**2.63**
EX3 - right channel	3.80	3.43	**3.20**
EX3 - left channel	3.78	3.25	**2.95**
EX4 - right channel	3.80	3.36	**3.20**
EX4 - left channel	3.78	3.33	**3.00**

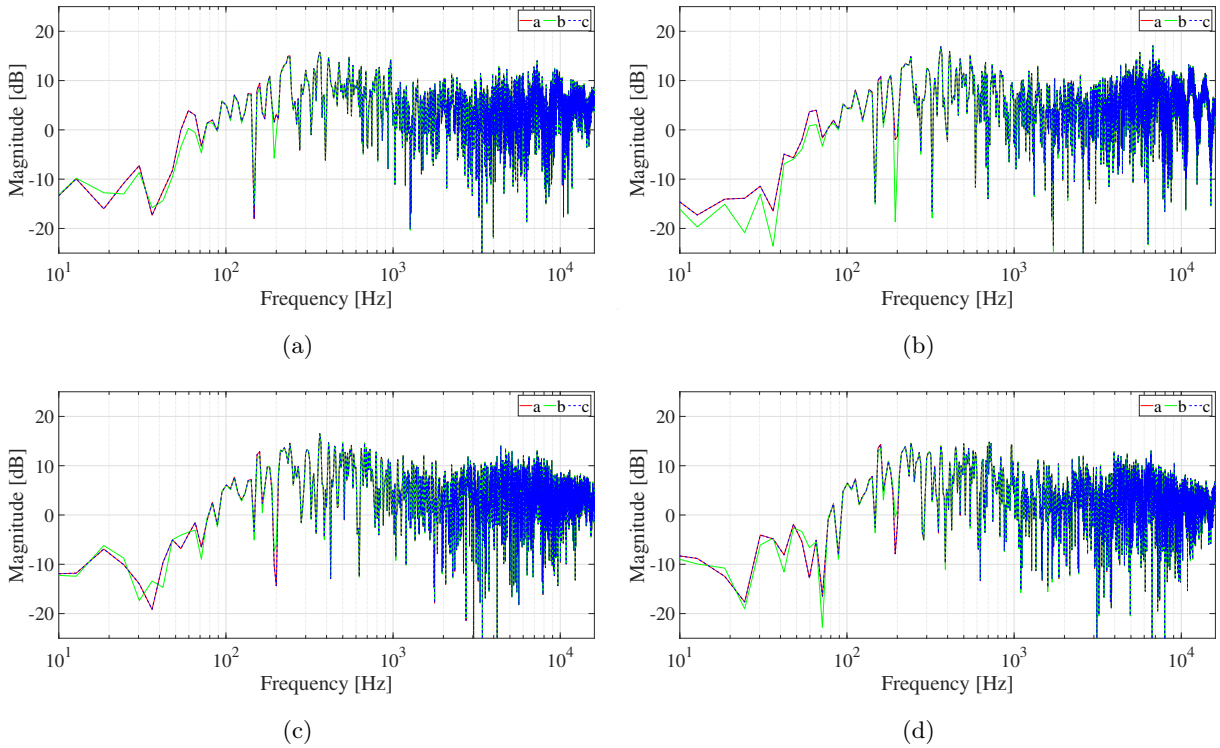


Figure 3.13: Experiment 1: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to the Right channel.

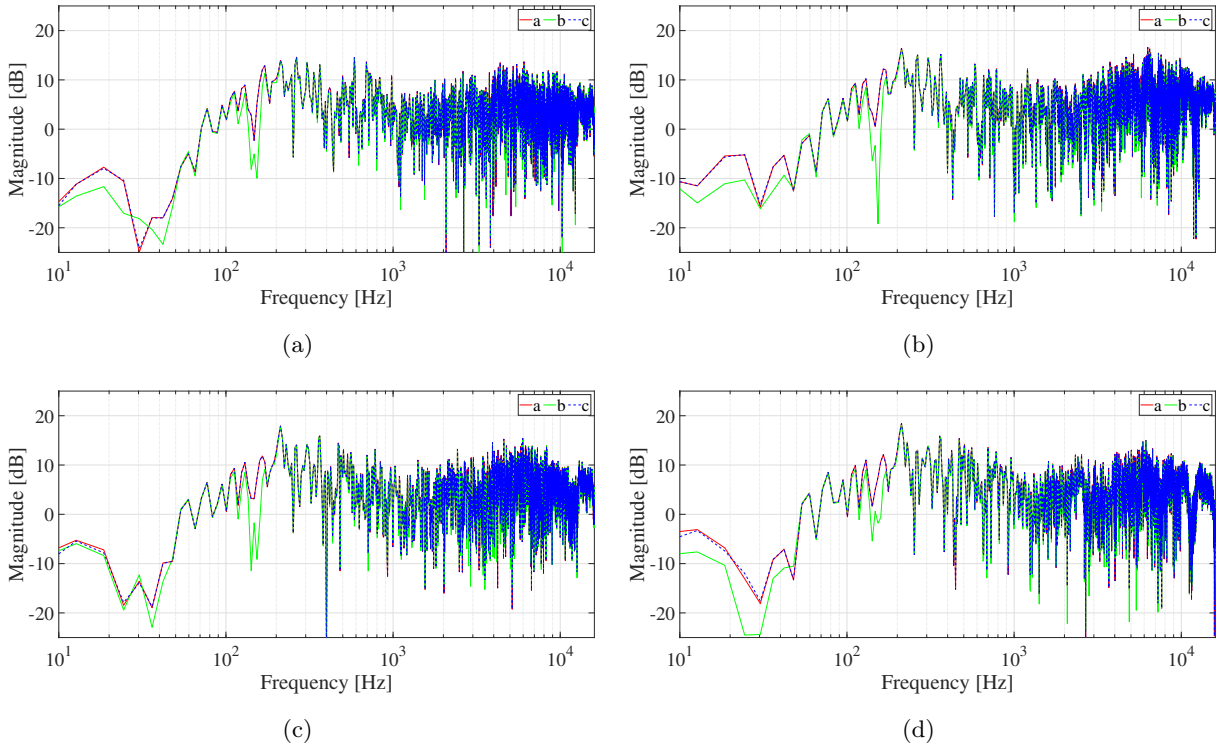


Figure 3.14: Experiment 2: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to Left channel.

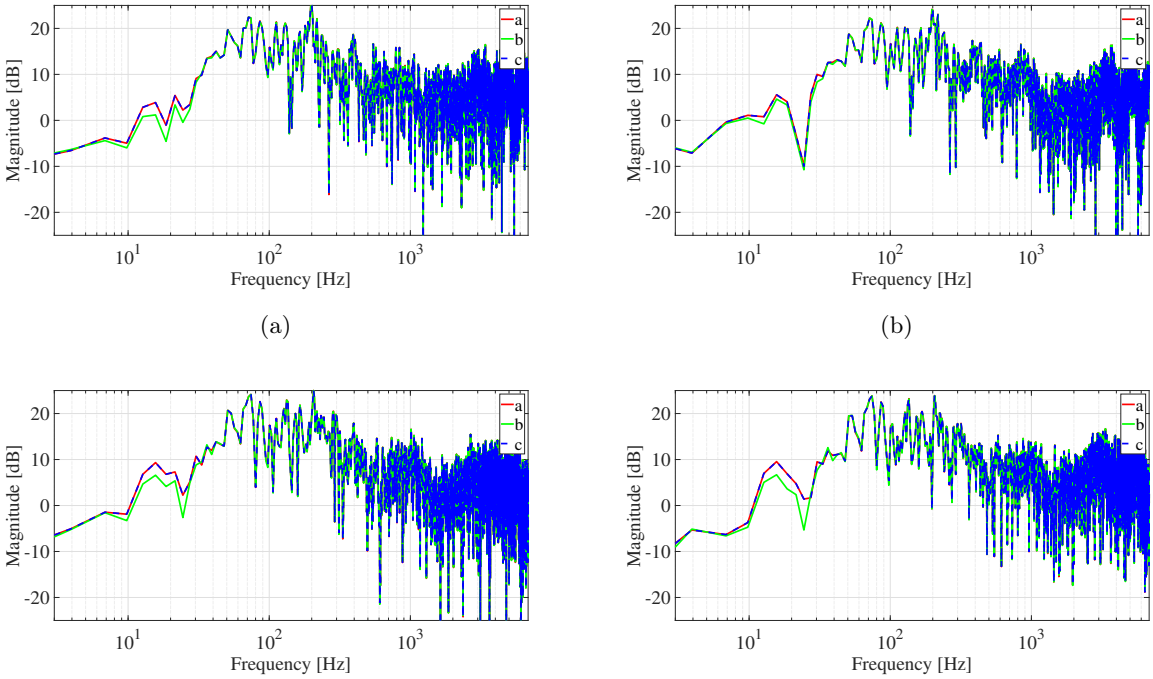


Figure 3.15: Experiment 3: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to the Right channel.

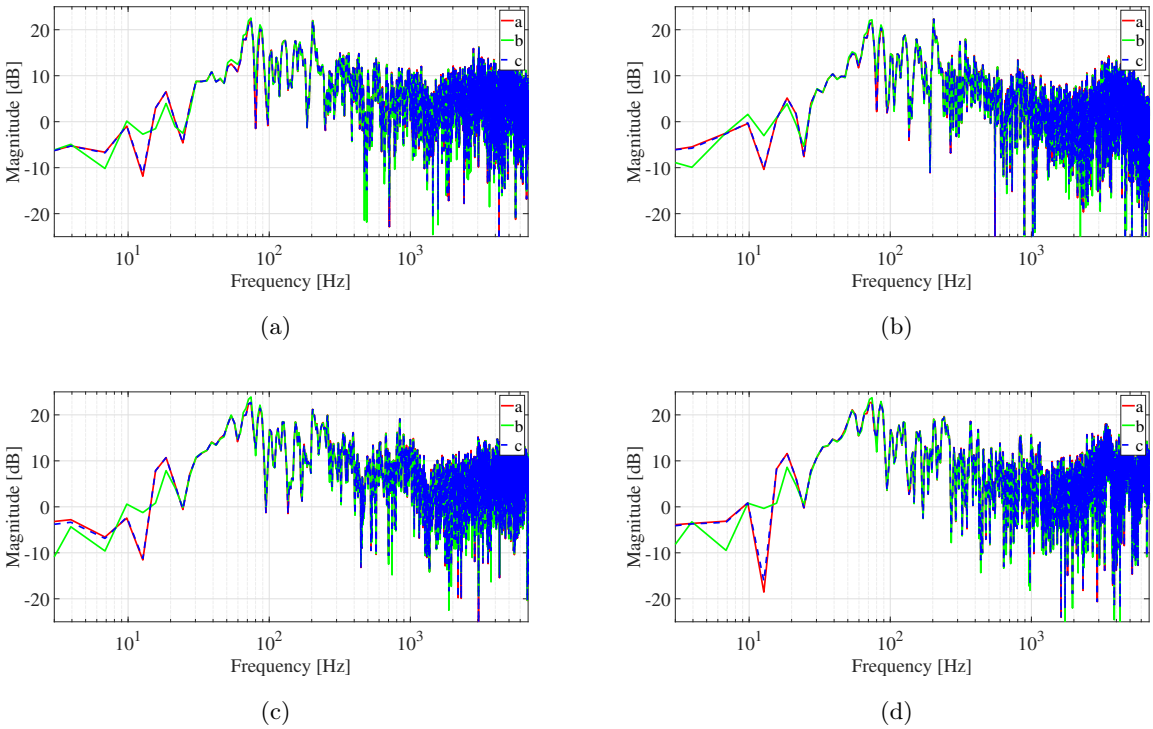
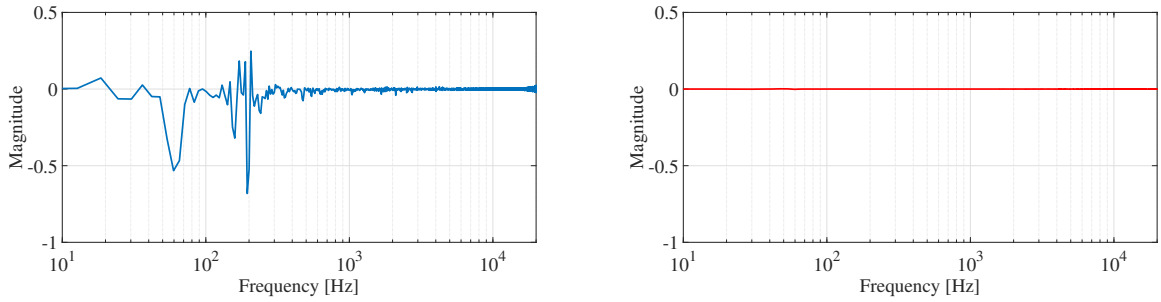
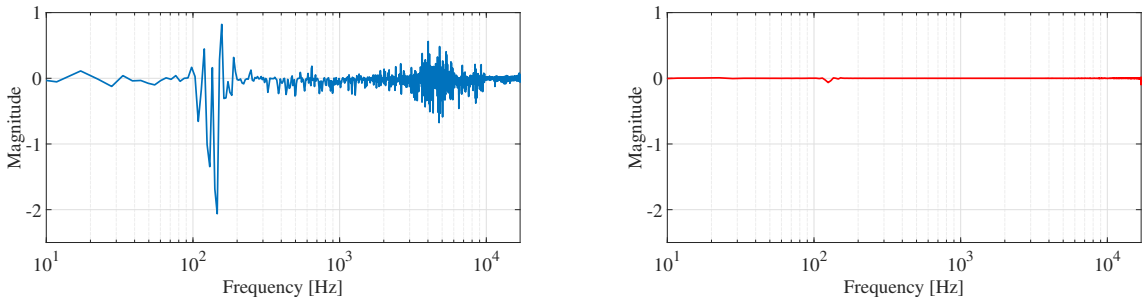


Figure 3.16: Experiment 4: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to Left channel.



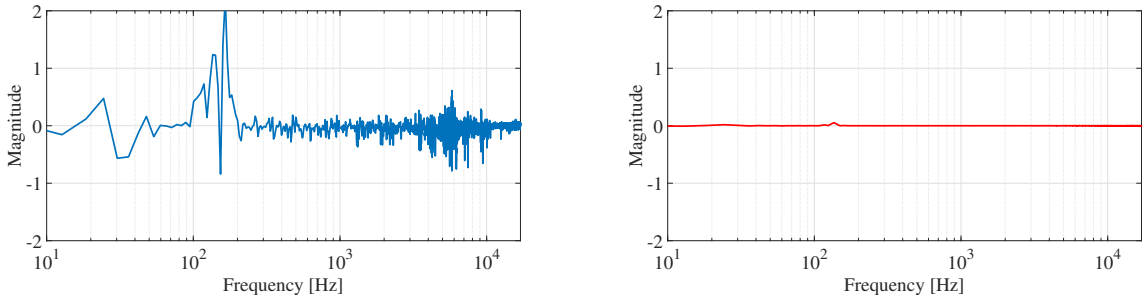
(a) (b)

Figure 3.17: Experiment 1: ((a)) difference between the real room magnitude responses and the identified room magnitude responses for ((a)) $D = 1$ and for ((b)) $D = 256$ considering one microphone with reference to the Right channel.



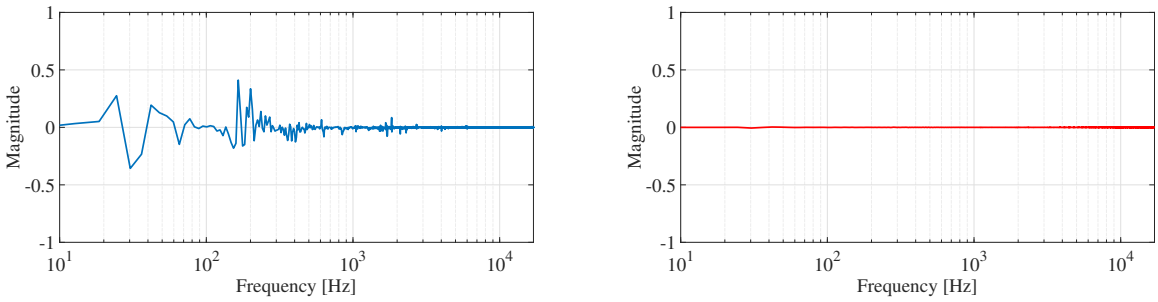
(a) (b)

Figure 3.18: Experiment 2: difference between the real room magnitude responses and the identified room magnitude responses for ((a)) $D = 1$ and for ((b)) $D = 256$ considering one microphone with reference to Left channel.



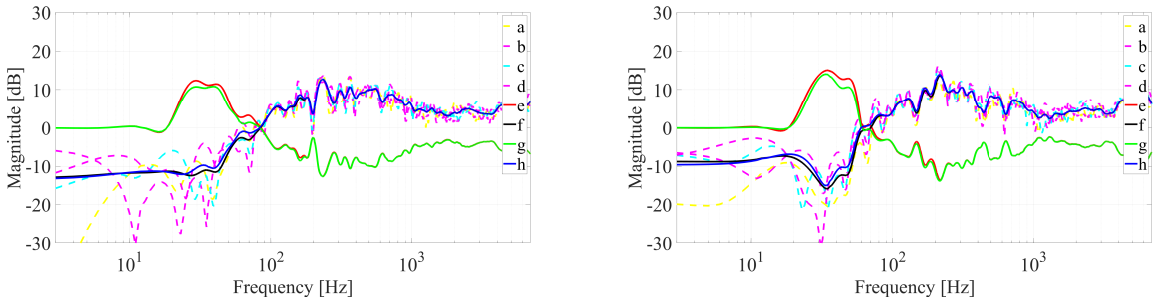
(a) (b)

Figure 3.19: Experiment 3: difference between the real room magnitude responses and the identified room magnitude responses for ((a)) $D = 1$ and for ((b)) $D = 256$ considering one microphone with reference to Left channel.



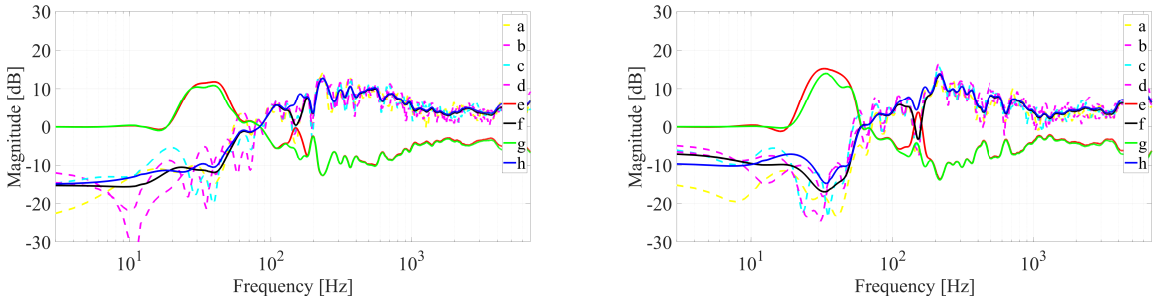
(a) (b)

Figure 3.20: Experiment 4: difference between the real room magnitude responses and the identified room magnitude responses for ((a)) $D = 1$ and for ((b)) $D = 256$ considering one microphone with reference to Left channel.



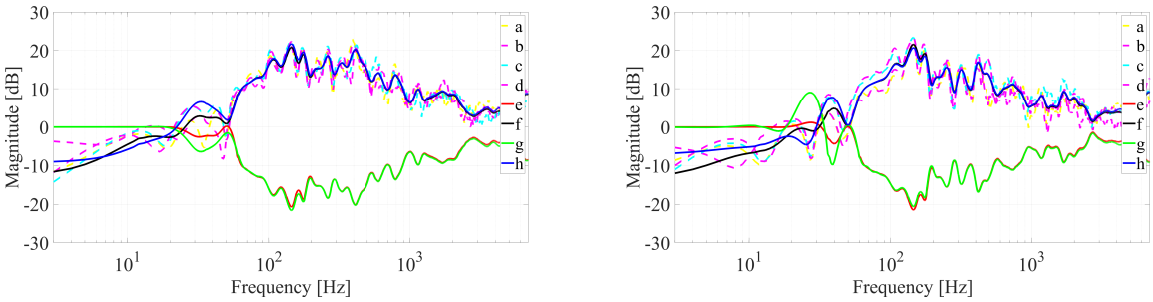
(a) (b)

Figure 3.21: Experiment 1: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for sub-band identification (h) prototype response considering a sub-band identification, with respectively ((a)) Left channel and ((b)) Right channel with a smoothing factor of 1/12.



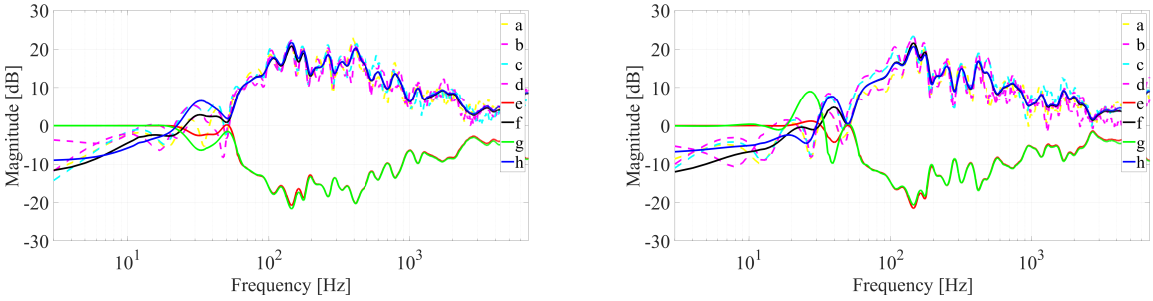
(a) (b)

Figure 3.22: Experiment 2: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for sub-band identification (h) prototype response considering a sub-band identification, with respectively ((a)) Left channel and ((b)) Right channel with a smoothing factor of 1/12.



(a) (b)

Figure 3.23: Experiment 3: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for sub-band identification (h) prototype response considering a sub-band identification, with respectively ((a)) Left channel and ((b)) Right channel with a smoothing factor of 1/12.



(a) (b)

Figure 3.24: Experiment 4: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for sub-band identification (h) prototype response considering a sub-band identification, with respectively ((a)) Left channel and ((b)) Right channel with a smoothing factor of 1/12.

3.5 Conclusions

In this chapter, the room response identification and equalization have been discussed, since real acoustic environment is a critical part for any sound reproduction system. The problem of RIR measurement has been firstly discussed focusing on the importance of a reliable impulse response estimation procedure considering, two state of the art approaches for RIR measurement. Taking into consideration equalization, an overview of the multi-point equalizer adopted has been explained and discussed. Focusing on results, the first approach presented involves the application of the multi-point equalizer to vibrating transducers. These devices generate sound making into a sound sources any surface on which they are installed. This is great advantage for interior design solution, but the installation choices like the type and the material of the surface, the anchors could affect the sound reproduction performance heavily. For these reasons, an equalization algorithm could enhance the sound performances of this systems. Experimental results show that with a multi-point equalization, a better frequency response is achievable and in the meantime it is also possible to compensate the installation problems as well as the environment behaviour from both a subjective and an objective point of view. The second approach discussed is more focused on a computation efficient structure for a multi-point adaptive equalizer. The work proposed rely on the same multi-point equalizer but involves an adaptive procedure for room impulse response estimation, i.e., the RIR are not estimated a priori, but are updated constantly by means of an adaptive algorithm. In order to improve convergence speed and reducing the computational cost, a multirate sub-band approach has been applied, using a perfect reconstruction filter bank. Several different experiments considering impulse response acquired in two different rooms have been considered showing that the proposed approach allows a precise RIR estimation, at a reduced cost.

Chapter 4

Honey bee hive as real audio system: acquisition and analysis

Among insects, honey bees (*Apis mellifera L.*) are well known for their positive effects. They do not produce only honey, beeswax, royal jelly, and propolis but they are at the basis of the plants pollination, playing a key role in the proliferation of both spontaneous and cultivated flora. Colony collapse disorder (CCD) is a recent, pervasive syndrome affecting honey bee colonies in Europe and the rest of the world, which is characterized by a sudden disappearance of honey bees from the hive [85, 156, 207]. The consensus among bee scientists is that honey bee colonies decline is the result of multiple stressors, working independently, in combination, or synergistically to impact honey bee health [156, 208]. CDD has emphasized the significance of a continuous and extensive monitoring to investigate factors that may negatively affect the life cycle of bees. Large bee mortality can result in loss of pollination services which have important negative ecological and economic impacts that could significantly affect the maintenance of wild plant diversity, wider ecosystem stability and crop production. Solutions such as robotic pollination were proposed in recent years, but it was already shown to be a technically and economically unacceptable solution, currently posing substantial ecological and ethical risks [171]. In this context, the necessity of a continuous and intensive monitoring of bee hives' status emerges clearly, in order to safeguard and protect these important insects. A non-invasive technique to analyze the bees' state is the sound analysis [155]. In fact, vibration and sound signals are used by bees to communicate within the colony [89, 142] and they can be used to understand the behaviour of the honey bees. In this context, an acquisition system and a procedure to analyze the sound have been proposed and then will be described in this chapter considering the obtained results.

4.1 Background

The early works

The topic of analysing the sound emitted by honey bees has always raised the interest of many scientists, from different fields. Among the first ones to make some observations on honey bees' sound was Aristotle who lived in the third century BC. In one of his writings [15], he observed that when a swarming is imminent a particular sound is produced continuously by the bees, several days before the swarming event. Later on, one of the early works came from the $XVII$ century

AD, in particular one of the first descriptions of queen piping was given by Charles Butler [183], who reported in his book that the bees produce two different sounds. A century later, Francis Huber [109] studying the birth of a new queen, discovered that the first young queens born emit a sound called tooting, and that other young queens, which are still sitting in their cells, respond with another sound called quacking.

The XX century and the first modern study

During the second half of the XX century, thanks to modern electronics, deeper and more accurate studies on the sound emitted by honey bees were carried out. In 1957 Frings et al. [89], published an article showing that bees react to different sounds. In particular, it was demonstrated that signals of certain frequencies and amplitude produce an almost total cessation of movement of workers bee and drones. Some years later, in 1964, Wenner et al. [212] performed one of the first spectral analyses on the honey bee's sound. Using a spectrograph and a microphone placed inside the colony, he discovered information about the sound produced during the waggle dance and after the birth of a new queen. The waggle dance is used by bees to communicate within the colony the information about the distance and position of flowers. The work shows that during the dance, bees emit a specific pulsed sound, in which the number of pulses is proportional to the distance between the colony and the flowers. Wenner also deepened the study about tooting, showing that this sound is composed of two different pulses, the first is a long pulse of 1 second duration followed by shorter pulses, with a fundamental frequency of 500 Hz and some overtones. The quacking has a lower fundamental frequency and starts with shorter pulses. In 1974, Simpson et al. [193] performed a study to better understand the queen bee piping sound [192]. Using a modified speaker attached to a hive, a sinusoidal 650 Hz vibration signal was generated, in order to obtain a signal similar to the one produced by queen bees. When the signal was applied to four hives containing very small colonies with unmated queens, all the four colonies swarmed. The second experiment was carried out with sixteen bigger colonies, each with a young queen bee. This time the sound was applied only to eight colonies which swarmed, leaving a portion of their bees in the hive. The other eight colonies which did not receive the sound didn't swarm. Despite the simplicity of the generated sound, the experiments show that bees react to sound, even if the limited technology available at the time didn't allow the authors to generate more complex signals, that could work also for colonies with an older queen bee.

In 1985 Dietlein et al. [68] performed a longer study acquiring and analyzing the sound from three different colonies for over a year. The setup used involved two different microphones: one placed inside the hive (for the recording of bees' sound) and the other one placed between the inner cover and the telescoping outer cover of the hive (to capture the background noise). The microphones were both connected to an analog circuit which, thanks to a differential amplifier, subtracts the background noise, acquired from the outer microphone, from the honey bees' sounds. The amplitude of the signal was recorded hourly for a year, showing that there are some significant variations in the amplitude and frequencies of the sound produced in different seasons. Moreover, a frequency analysis has shown that also the spectral content of the signal changes with the seasons and the hour of the day. This study shows an interesting prospective on long-term sound analysis inside the colony, but one of the main drawbacks is probably the analog circuitry used to process the signal, i.e. the analog subtraction between the signals generated by the microphones and the

amplification could introduce an additive noise on the signal. One year later, in 1986, Michelsen et al. [142] performed a study more focused on the sound emitted during the waggle dance, analyzing both the vibration and the airborne sound produced. The idea was to better understand which mechanisms are used to communicate during the waggle dance: since it is performed inside the hive in the dark, bees cannot rely on their sight to see the dance, and they must use other communication means, such as vibration, airborne sound, and chemical transmission. Using both a laser vibrometer and microphones placed inside the colony, Michelsen discovered that no vibration is produced during the dance, and that the whole information is passed through the airborne sounds. In particular, as already discovered by Wenner [212], the waggle sound is a pulsed sound, with 20 ms - long pulses at the frequency of 250-300 Hz. It was also discovered that the waggle sound at 2 cm from the bee who is dancing, produces a sound wave with a Sound Pressure Level (SPL) of 73 dB. The second part of Michelsen's work is more focused on the mechanical aspects of the propagation of sound and vibration inside the colony. In this case the experiment involves the generation of synthetic comb vibration to better understand the freezing phenomena already described in [89]. Again, as already discovered, bees seem to freeze at specific frequencies, and this led Michelsen to speculate that these particular signals are used by bees to quiet the colony, in order to allow a better transmission of important messages. About ten years later, Eren et al. [80] carried out a study in order to better understand and try to emulate the communication between worker bees and the queen bee. From this study it results that worker bees are not able to produce sound with frequencies higher than 500 Hz, while queen bees are able to produce almost all workers frequencies and many higher ones. Researchers tried to acquire the sound from over 150 queen bees confined in standard queen bee shipment cases, arranged in a matrix form. The acquired sound has been analyzed and then synthesized thanks to a computer. However, the results seem quite inconsistent, exhibiting a weak response from both queen bee and workers to the synthesized sounds, that are a more complex than the one proposed in [142, 212]. This research topic seems quite promising and interesting: the inconsistent results could be mainly due to the experimental setup used for the queen bee sound acquisition. In fact, putting together over than 150 queen bees is not a normal situation, and the sound produced could be very different from the real one generated inside the colony. To achieve better results it is desirable to acquire the signals in a more natural situation, trying to avoid stressors acting on the bees.

The XXI century and the technological advancement

Within the last twenty years the technological progress has allowed the application of new solutions such as digital signal processing, ML and low cost smart sensors, enabling new discoveries. In 2005, Ferrari et al. [86] focused their studies on the swarming phenomena. In particular, researchers acquired signals from three hives for over 270 hours, observing nine swarming events. Each hive was equipped with an omnidirectional microphone, placed on top of hive's looms under the cover, and one humidity and temperature sensor placed inside the hive in between the looms. In the paper, authors do not provide specific motivation for the positioning of the sensors, but we can speculate about the need for keeping the number of microphones used as low as possible - thus adopting a single omnidirectional one - and for monitoring in-the-hive ambient conditions which could as well be correlated to a swarming event. Acquired data were synchronized and then manually labeled in order to identify the swarming events, but authors do not specify which of the three hives was involved in

the process. Each events has been analyzed both in frequency and time domain, in order to find any correlation between sound, temperature and humidity during swarming. Sound spectrograms have shown that there is a quick change in the frequency content during swarming. Before the swarming, most of the energy content of the signal is around 150 Hz, while during swarming, it jumps to 500 Hz. Moreover, the joint analysis of sound, temperature and humidity has shown that during swarming there is an increase in the sound amplitude, and meanwhile temperature and humidity decrease. According to the authors, this behaviour could be due to the ventilation produced by bees ready for swarming.

In 2008, Papachristoforou et al. [164] made an interesting discovery, showing that the sound emitted by honey bees could reach high frequencies, even up to 15-16 kHz. The experiment carried out was focused on sound emitted by the bees during a hornet attack. A hornet was artificially introduced inside the hive and the bees' reaction was recorded. A frequency analysis based on spectrograms revealed that during a hornet attack guard bees produce hissing sounds with a fundamental frequency at around 5 kHz, with several high order harmonics which can reach 15 and 16 kHz. The generated sound seems to have a precise structure and this led to think that this is a true signal used for communication and not a simple noise produced under stress. Some studies seem to show that honey bees are capable of picking up sound at over 10 kHz with the sub-genual organ, a chordotonal sensor localized in the proximal part of the tibia of each leg [121]. Such a high-frequency structure of the signal could be motivated by the fact that it makes the sound very distinctive from the background noise of the colony, allowing a clear transmission of the alarm message to the whole colony. This study has an important consequence, since it seems to show that the sound produced by the bees reaches frequencies that many other studies seem to ignore completely. In 2011, Eskov et al. [81] proposed a totally different approach, based on statistical sound analysis. The authors considered the recorded signal as a noise with a specific spectral content. As other types of colored noise, also the one produced from the bees has a precise statistical behaviour which could be measured. Data from twelve hives were acquired every night from midnight to 3 A.M. within the spring and summer periods. Each recording was divided into one second long fragments which could be analyzed by estimating the relative fluctuations of the signal. Each fragment was first normalized then smoothed using the Procedure of Optimal Linear Smoothing (POLS) [3] with Gaussian kernel, applied to the integrated sequence of original data. Finally, the sequences of ranked amplitudes of relative fluctuations have been estimated giving a statistical indicator of sound behaviour. The paper shows that when bees are going to swarm, the fluctuation changes and this could be used to predict the swarming event several days before it could happen. This work is quite interesting as it proposes an original point of view and could be used for swarming detection. The main drawback is related to the fact that the authors assume that most of the recorded signal is only noise, and this assumption could be a limit in some cases. In 2013, Howard et al., [106] proposed a classification algorithm for queenless colonies, based on the Stockwell Transform (S-Transform) [197] as features extraction tool. Four colonies of two different species have been monitored, with two normal colonies and two orphaned ones. Data were first analyzed using the S-Transform, which is a time-frequency representation of the signal derived from the wavelet transform with a modification in the phase of the mother wavelet. A qualitative analysis was carried out, comparing the S-Transform, with spectrograms and Fourier transform of the signal. For the classification and representation of the signals, a Self-Organising Maps (SOM) [123] was used. The SOM is trained with two neural net-

works and allows a clear representation and classification of datasets featuring high dimensionality. Results show that the SOM approach is able to classify the two states with some issues, probably related to the use of two different bee species, or to the classification algorithm. In 2014, Quandour et al. [173] went beyond the simple sound analysis, using the recorded sound to automatically detect if the colony was healthy or infected by the varroa mite. Varroa destructor is one of the most dangerous honey bees parasites and is considered as one of the factors which is contributing to the higher levels of bee losses around the world [97]. The proposed approach was based on the extraction of four statistical indicators from the sound: peak frequency, spectral centroid, bandwidth, and root variance frequency. These four features were then passed to a Principal Component Analysis (PCA) algorithm to reduce the data dimensionality, and finally the data were input to a classifier to detect if the colony was healthy or infected. Two types of classifiers were used, the former based on SVM and the latter on LDA. Both shows good performances allowing the discrimination of the two colony statuses. A notable aspect is the reduced computational costs of the proposed algorithm which was implemented on a low-cost single board computer, but on the other side the dataset used in the experiments was too limited to give a reliable benchmark on algorithm performances. One year later, Murphy et. al [146] proposed a complete platform for honey bees' monitoring. In this work the microphone is only one of the sensors installed within the hive, and the other parameters acquired include CO_2, temperature, humidity, acceleration data. Additionally, an infrared camera inside the hive and a thermal camera outside the hive are installed. Sound is used to detect possible swarming events in a very simple way: the system filters and then estimates the envelope of the recorded sound, and if the amplitude of the signal rises quickly, the system sends a message to the beekeeper. This simple monitoring system could be used for swarming detection on a very simple hardware with also a reduced power consumption, but due to its simplicity, it could easily generate false positive warnings. In 2017, Ramsey et al. [174] focused their work on the whooping signal, trying to detect this particular signal in different conditions. Vibracoustic signals of three colonies were acquired in different periods of the year and in different geographical locations, with two colonies placed in the UK and one in France. Each colony was equipped with a high precision accelerometer placed in the center of a brood frame. The whooping signal was detected with a two-step process. First, the spectrogram of the recorded signal is matched to the spectrogram of a template pulse, then the ratio of the cross-correlation product and the Euclidean distance among the two is used to find the pulsed signal. Then, in order to discriminate between whooping signal and non whooping signal, PCA and LDA are used. The number of whooping signals recorded shows variations which seem related to weather conditions, geographical positions and time of the day. The whooping sound monitoring could be used as an indicator for hive monitoring, however the type of sensor used and its precise positioning inside the colony limit this approach, avoiding a possible large scale implementation of the system. In 2018, Kulyukin et al. [124] published an article where they exploit ML techniques to analyze the sound of the hives. In particular, the main goal was to distinguish the honey bee sound, from the background noise and the cricket chirping noise. Four microphones were placed outside the entrance of six Langstroth bee hives. A sound frame of 30 s was recorded every hour, from May to July. Bee hives were placed in different locations, with many different background noises. Data have been manually labeled into three categories: honey been, cricket, background noise. With the obtained dataset, several approaches were tested to classify the data, in particular an ML approach based on a Convolutional Neural Network (CNN), has been

compared to traditional classifiers such as Logistic Regression, k-Nearest Neighbors (k-NN), SVM with a linear kernel, one vs. rest classification and (M4) Random Forests. The results show that for this type of problems, an ML approach could be used with success, but since it wasn't directly applied to the honey bees' sound, a similar algorithm could be used only as a preprocessing technique to remove unwanted sounds. In the same year, Cejrowski et al. [59] proposed an algorithm to detect the presence of the queen bee based on sound analysis. The system involves a custom brood frame placed inside the colony, with a microphone, and a temperature and humidity sensor. Data have been acquired from a single hive from February 2017 to August 2017, forcing a critical situation for the bees, by removing the queen bee from the hive. Data acquired with the normal colony and the orphaned colony have been analyzed using a Linear Predictive Coding (LPC) 0for features extraction. This algorithm is based on the source-filter model of speech signal, i.e., first the input signal is used to produce a linear model with a number of poles, then the resulting set of coefficients is used to model the unknown system. Following the LPC coefficients estimation, the t-distributed stochastic neighbor embedding (t-SNE) [135] algorithm has been used to reduce the data dimensionality and finally, an SVM algorithm was used to classify the results. The algorithm seems to work well with a good ability of detecting changes in the colony sound, however the data used are limited, with a single colony and a single orphaned event analyzed. Moreover, with the new queen bee the algorithm seems to need a new training process while a similar approach should be able to detect orphaned and normal colony situations with no need of such a step.

Patents

Several patents have been also issued during years. One of the first was presented in 1957 proposing a particular instrument called the Apidictor [214]. The system presented is composed of a microphone, a vu-meter and a series of tunable analog filters. The idea is to adjust the filters to the specific frequency bands at which bees emit specific sounds (i.e., the piping sound) and then to monitor the activity. In 2007, Bromenshenk et al. [32] submitted a patent to exploit the sound emitted by honey bees to monitor air pollution. The main idea is that when bees are exposed to sub-lethal concentrations of various airborne toxicants, the generated sounds are different and specific for each pollutant. These sounds can be acquired and stored into a database. Then, when a new sound is acquired inside a colony, it is possible to make a comparison with the database, in order to detect pollutants near the colony. For the comparison, first the spectrograms of various sounds are extracted and then a classification by means of linear discriminant functions is carried on. In 2012, Brundage et al. [33] proposed a system to detect bees productivity based on sound analysis. The idea was to analyze the fundamental harmonic produced by the bees' flight, which is in the range between 180 Hz to 260 Hz. The presence of a downward frequency shift in the fundamental frequency corresponds to a flying bee launching from locations around the bee hive entrance. By counting the number of frequency shifts it is possible to estimate the number of bees which have left the hive during a day, and then it is possible to use these data to monitor the hive productivity. One of the most recent patent is from Bencsik et al. [21] who proposed, in 2015, a solution based on previous research on sound emitted by honey bees [174]. The system involves the use of one or more accelerometers, placed at the center of the brood frame. The acquired signal is then transformed into the frequency domain producing a spectrogram. The spectrogram is then processed with PCA to reduce the data dimension. Finally, a linear discriminant function is used to analyze the signal

Table 4.1: Summary of the state-of-the-art of approaches discussed in this chapter.

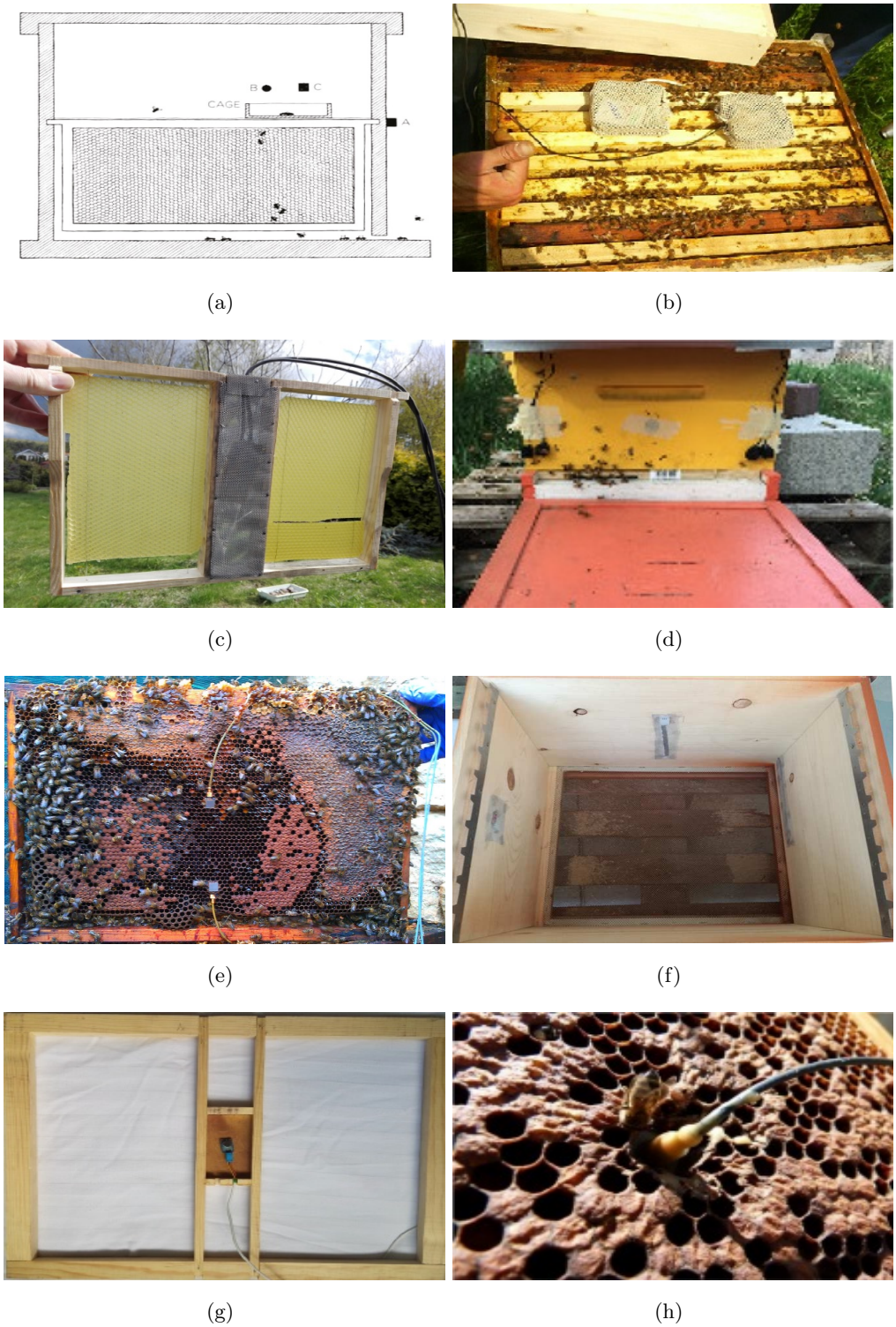
Approach	Description	Applications	References
Spectrograms	The sound is recorded and then analyzed using spectrograms, searching for changes in the harmonic content of the signal.	Analysis of waggle dance, analysis of piping an tooting. Swarming detection. Measuring bees reaction to hornet attack.	[68, 86, 142, 164, 212]
Tone based sound synthesis	A loudspeaker or a shaker is placed inside the hive, different tones at different amplitudes and frequencies are generated and bees reaction is monitored.	Find the frequency at which bees react with movement cessation. Reproduce the harmonic generated by the queen bee, to stimulate a swarming.	[89, 142, 193]
Amplitude monitoring	Amplitude and envelope of the recorded signal is used to detect different behaviours.	Changes of the amplitude in different seasons and conditions. Measuring of SPL during waggle dance. Swarming detection.	[68, 142, 146]
Bees sound synthesis	The bees sound is firstly recorded and then analyzed and synthesized by means of a computer. The synthesized sound is then reproduced inside the colony and bees reaction is monitored.	Measure the response of worker bees to the synthesized queen bee sound.	[80]
Noise analysis	The recorded sound inside the colony is considered as a noise with a specific statistical behaviour. Some statistical indicators are extracted from the sound, changes in the statistical indicators are related to specific colony behaviours.	Swarming detection and prediction.	[81]
Statistical indicator analysis	From the recorded sound, peak frequency, spectral centroid, bandwidth and root variance frequency are extracted. PCA is used to reduce the dimensionality of the indicators and finally SVM or LDA is used to classify the signals.	Detect the presence of Varroa destructor inside the colony.	[173]
Whooping detection	Precision accelerometer inside the colony are used to record the bees vibrations. Spectrograms of vibrations are cross correlated with a pulse signal to detect pulsed signals, LDA and PCA are then used to isolate whooping signals.	Measuring the variation of the whooping signal during different seasons and geographical locations.	[174]
Bees sound detection	The sound is acquired at the hive entrance. Spectrograms of the recorded sound are classified using different algorithms such as, CNNs, logistic regression, SVM, k-NN, one vs. rest and random forest.	Distinguishing the honey bee sound, from the background noise and the cricket chirping noise	[124]
LPC sound analysis	Sound acquired inside the hive is analyzed using LPC as features extraction algorithm. T-SNE algorithm is then used to reduce dimensionality, and finally SVM is used to classify the signals.	Queen bee presence detection.	[59]
HHT and MFCC analysis	Recorded sound inside the colony is analyzed using MFCCs and HHT as features. CNNs and SVM are then applied to classify the signals.	Queen bee presence detection, swarming detection.	[53, 151, 200]
MFCC analysis	MFCCs are estimated from the recorded signal. Lasso regularization is then used for dimensionality reduction and finally logistic regression algorithm is used for classification.	Queen bee presence detection.	[180]
Wavelet analisys	Wavelet transform is applied to the recorded signal to analyze the sound and detect different behaviour.	Queen bee presence detection, swarming detection	[53, 200]
MFCC and LPC analysis	MFCC and LPC are used as features, HMM and GMM are used as classifier.	Swarming detection.	[217]
Multimensional FFT	Two and three dimensional spectrograms are generated starting from the sound recorded using accelerometers placed inside the colony. A discriminant function is then used to classify the signals and detect specific events using two different algorithms.	Swarming detection and swarming prediction.	[175]

at a reduced computational cost in order to detect event as swarming or a Varroa mite infection.

Recent results

In the 2019-2020 years, several works have been published on bee sound analysis. Nolasco et al. [151] applied ML techniques to analyze honey bees' sound and detect the queen bee presence. Data from the Nu-Hive project [54] were used, analyzing sound from two different colonies in normal and orphaned situations. Two features extraction techniques were used: Mel Frequency Cepstral Coefficients (MFCC) and the Hilbert-Huang Transform (HHT) [108]. MFCCs is a widely known technique for signal representation, in which the coefficients are generated starting from the square of signal spectrogram, then a triangular filterbank is applied to the signal, finally the Discrete Cosine Transform (DCT) is applied to the logarithmic output of the filterbank. Regarding HHT, the features extraction technique is based on a combination of two algorithms, i.e., the Empirical Mode Decomposition (EMD), which decomposes the signal into a set of basis functions, and the Hilbert Transform (HT) which transforms each basis function into a time-frequency representation of the signal. Using the SVM classifier, a comparison of the performance of both approaches was presented, showing that the best results were achieved with a combination of MFCCs and HHT coefficients. Some experiments with CNNs were carried out as well, using MFCCs as input data and providing good performances. In the mentioned work, authors exploit ML for orphaned colony analysis, however the most innovative part i.e., the HHT, has not been used in combination with ML, leaving this aspect to a future development. Within the same year another work was presented involving the queen bee presence detection: A. Robles-Guerrero, et al. [180] acquired sensor data from five hives, each hive was equipped with a single microphone placed a the center of a brood frame. Data were acquired from March to April, recording 30 s - long frames each ten minutes with a Raspberry Pi 2. MFFCs were first computed and then statistical moments for each mel coefficient were calculated involving: mean, standard deviation, variance, skewness, median and kurtosis. Lasso regularization was then used to reduce again the dimensionality and finally a Logistic Regression algorithm was used for classification. Some experiments were also proposed using Singular Value Decomposition (SVD) and scatter plots to analyze the behavior and separability of the datasets. Among the five colonies, one lost naturally its queen, while in other two families the queen bee was removed intentionally. Two experiments were carried out, the former comparing one orphaned colony with the four healthy ones, and the latter removing two queen bees and comparing the data with two normal colonies, and the queenless one. The classifier was able to separate orphaned colony from the healthy one with a good accuracy. In the same year, another study on sound based swarming detection was published [217]. Data from the open source bee hive project [8] were taken, using MFCCs and LPC as features. As classifiers, two different algorithms were used, the former based on Hidden Markov Model (HMM) and the latter on Gaussian mixture model (GMM). Several experiments with different combinations of classifiers and features have been carried on, showing that the best solution is MFCCs in combination with HMM classifier. The work shows also some experiments focusing on the performance degradation due to high level of background noise in the recorded signals. Again, as in other works, the main drawback is the limited dataset size, since only three swarming events were available and were used for the experiments with only 90 minutes of recordings for the training phase. Two other studies were published in the same year, which more deeply analyze another point, i.e., how to extract the information from the recorded

sound inside the colony. In [53, 200] the authors proposed some innovative approaches for the features extraction problem. Many previous researches focused their work on the classifier, using only spectrograms or at least MFCCs to extract important information from the sound. In these two works, some innovative approaches for features extraction are discussed, focusing on orphaned colony situations [200] and swarming [53]. Well known approaches as spectrograms and MFCCs are compared with innovative algorithms, based on Hilbert-Huang transform, and wavelet transform [65]. The results presented seem to show that these approaches are able to better distinguish different signals, and this will be more deeply investigated in a future work where these approaches will be combined with classification algorithms. Another study proposed in 2019 is from [112]. Here, the authors proposed a Raspberry Pi-based acquisition system, able to collect the following parameters: images from the hive entrance, sound, weight, external temperature and humidity, internal and external light level, internal temperature and air quality. Data are analyzed to detect changes in different periods; weight is also used for swarming detection. A simple analysis of recorded sound is performed by means of a Fourier transform. One of the most recent contributions in this field is from Ramsey et al. [175]. In this work the authors improve the results of their previous work [174]. In particular, using the same dataset they propose two different methodologies to predict swarming. One is based on the extraction of one hour logarithmic averaged spectra, then the cross-correlation product between the spectra and three discriminant functions is calculated. Finally, a linear discriminant function is used to detect swarming. The discriminant functions are estimated by a specific algorithm which analyzes spectrograms of swarming events to estimate the discriminant curves. The second approach is based on a long-term estimation: the algorithm analyzes ten days of vibrational data, acquired from midnight to 5:00 A.M. Ten spectrograms are firstly estimated and then the Fast Fourier Transform (FFT) of each spectrogram is performed on the time course of the magnitude of all the uploaded spectral frequencies for each day, yielding two dimensional Fourier transform images (2DFT). The FFT is then further calculated over each pixel of the series of 2DFTs found in the preceding days, obtaining a three dimensional Fourier transform (3DFT). The 3DFT spectrograms are cross correlated with specific discriminant functions and eventually the swarming is predicted using a linear discriminant function. The results seem to show that the first approach is more reliable, providing less false positive swarming events. The second part of the article is focused on the structure of tooting and quacking signals. The results are quite interesting, opening to the possibility of having a reliable swarming prediction. The only problem in the proposed approach is the type of sensor used, which has a high cost and is not suitable for a large scale application. Finally, another recent work is presented in [51], where a combined study is considered by crossing the results of the analysis of the sound generated inside the hive by the honey bees, with other common parameters such as temperature, humidity, CO_2 and hive weight, and weather conditions, acquired with several sensors. In this work the changes in the weight produced during the swarming phenomena are related to the changes in the sound amplitude. Moreover, the other sensors' variations are analyzed in a normal day and during longer periods, showing trends related to the honey production and the colony activity.



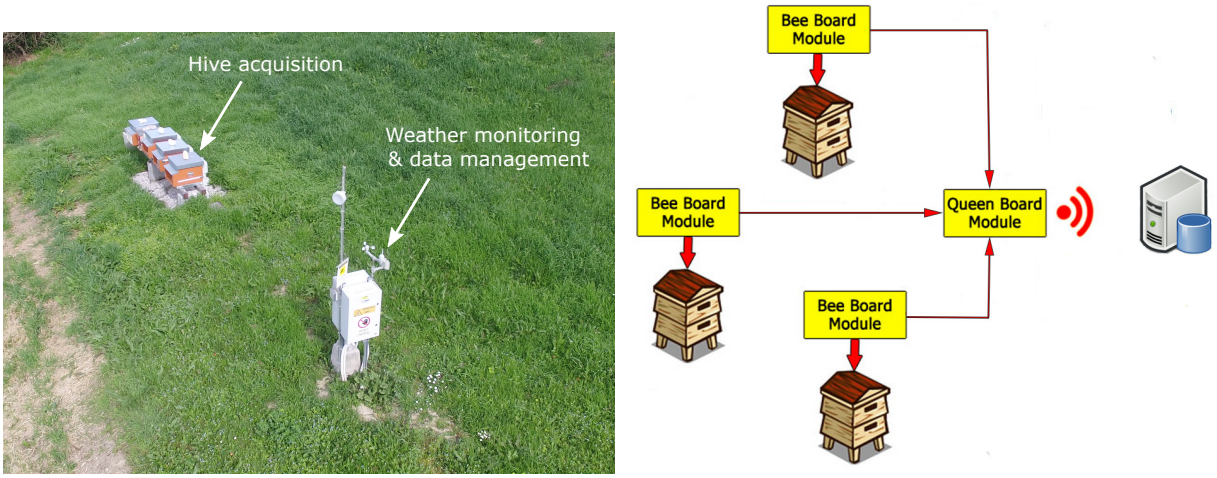
(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 4.1: Different microphones and accelerometers placement inside the colonies, based on different approaches. In particular: on Figure 4.1(a) from [212], B is the microphone used and it is placed upper the queen cage. Figure 4.1(b) shows the microphone placement of [87]: the sensors are placed upon the brood frames inside a cage to protect them from propolisation. Figure 4.1(c) shows the solution adopted in [59], where a custom frame with the sensors inside has been chosen. Figure 4.1(d) refers to the approach proposed in [124], with microphones placed outside the hive, so that a cage to protect against propolisation is not necessary. On Figure 4.1(e), accelerometers placement proposed in [174] is presented: the sensors exploit vibrations and do not suffer from propolisation problems. Figure 4.1(f) belongs to approaches presented in [53, 54, 151, 200]: the microphones are hidden inside the hive walls, and a grid is used to protect them. Figure 4.1(g) from [180] shows a custom brood frame. Finally, Figure 4.1(h) from [175] shows accelerometers positioning similar to the one proposed in [174].



(a) (b)

Figure 4.2: The acquisition system installed in the university campus on figure 4.2(a), while the connection between the Bee board and the Queen board modules with the remote server is visible on figure 4.2(b).

4.2 The sound acquisition system

The acquisition systems developed and used for the data analyzed in this work, is part of the strategic project entitled "Nu-Hive: Innovative technologies for the hive monitoring". The main idea behind the project was to develop a reliable system for the acquisition of multiple sensors installed inside a hive with particular attention to the sound acquisition system. The whole acquisition system is visible on Figure 4.2(a). The system is modular and it is composed of two main modules. The former, named Bee board, is installed in each hive, and the latter, named Queen board, is the main module installed near the colony and it is capable of gathering the data from all the Bee board modules installed in each hive. The Bee board module consists of a RaspberryPi 3B equipped with a Behringer UCA22 sound card and two ADMP401 MEMS microphones. There are also other sensors such as two DHT22 humidity and temperature sensors, a properly designed weight scale, and a CO_2 sensor. The Queen board module consists of a RaspberryPi 3B equipped with several sensors to acquire weather parameters near the hives and an Ethernet switch that, through a 5 GHz wireless bridge, allows the communication to a remote server where the acquired data is verified and stored the schematic of the system is visible on Figure 4.2(b). The system has acquired the data continuously for over two years and half, collecting over 18TB of data, which are mainly uncompressed .wav files of the recorded sound.

4.2.1 Hardware

In the following Section, a brief description of the sound acquisition system will be discussed. A complete description of the other sensors acquisition system can be found in [51].

For sound measurement, two microphones are installed in each wood hive in two different positions. The former is in the center position of the front side of the hive, while the latter is in the center of the back side, housed in a dug into the wood. In order to have a full integration of the system, MEMS (Micro Electro-Mechanical Systems) microphones have been chosen. This way it is possible to keep the whole system as hidden and smaller as possible. The selected microphone is

the Analog Devices ADMP401 [1] with a frequency bandwidth between 100 Hz and 15 kHz. The microphones signals are acquired by a 24 bit USB sound card connected to the Raspberry Pi with a sample rate of 32 kHz. Electrical connections of the system are visible on Figure 4.3(a). Each microphone has its own board that is fed with a 3.3V power supply obtained by the Raspberry Pi GPIO. The analog signal is obtained from the USB sound card. The advanced Linux sound architecture (ALSA) driver have been used with pyaduio library to develop the acquisition procedure. Figure 4.3(b) shows the acquisition routine: each frame is saved inside an acquisition buffer that is saved as a .wav file every ten minutes.

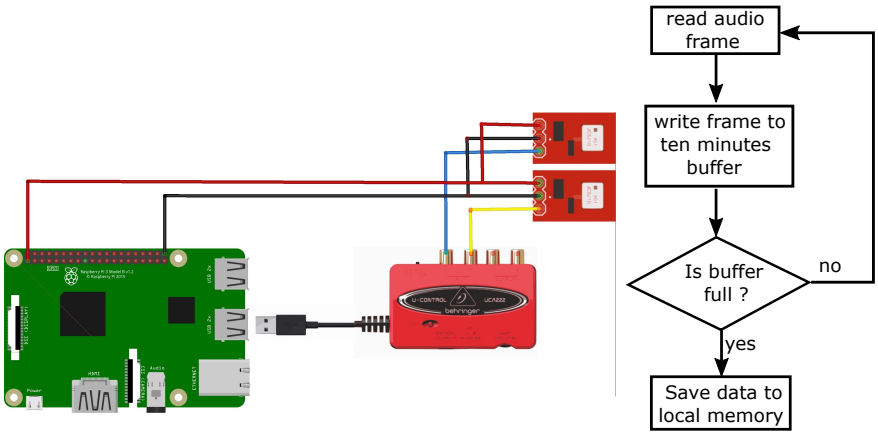


Figure 4.3: Schematic of electrical connections for the sound acquisition. The signal of ADMP401 microphone is acquired by means of Behringer UCA222 USB audio card which is connected to the Raspberry Pi board on figure 4.3(a), while on figure 4.3(b) is visible the flow chart of the acquisition routine.

The placement of the two microphones has been chose according to entomological indication in order to have one microphone near the brood and the latter near the hive entrance. Figure 4.4 shows the final results. Sensors have been covered with a natural material derived from the silk, to avoid propolisation, which could degrade the sensors performances.

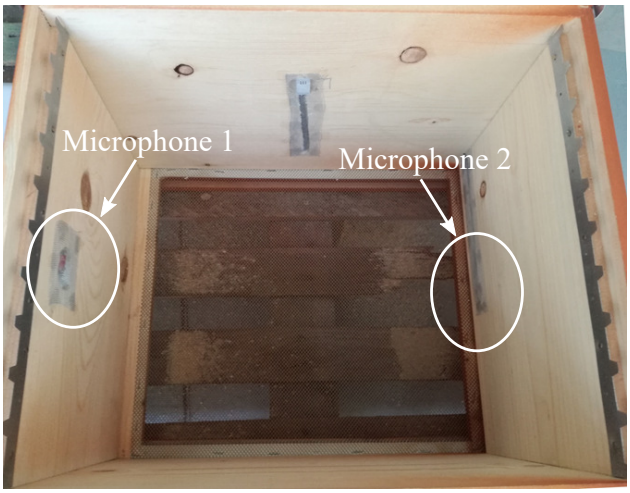


Figure 4.4: Placement of two microphones inside the hive, in particular microphone 1 is placed near the hive entrance, while microphone 2 is place near the hive brood.

4.2.2 Sound visualization and management

The software components necessary for the data acquisition platform have been developed using two programming languages: the Raspberry Pi acquisition code has been written in Pyhton 2.7, while the server tasks for data visualization and data storage have been implemented using LabView 2016. Focusing on the server side, different LabView Virtual Instruments (VIs) have been developed, in particular on Figure 4.5 is visible part of the VI for data visualization that allows a preliminary analysis of the recorded sound by means of a fast Fourier transform. Another VI has been developed for sound storage management, this program exploits the samba protocol to move the recordings from the Raspberry Pis local memories to the remote network attached storage, this VI is also able to make an integrity check on the recorded data and in case of problems (i.e., data are not present or corrupted) is able to send an alert mail.



Figure 4.5: Graphical user interfaces of the software developed for sound acquisition and preliminary analysis. In particular: the sound emitted by a selected hive is shown, in time (top screen) and frequency (bottom screen) domain, in arbitrary units (a.u.).

4.3 Honey bees sound analysis

Honey bees sound analysis can be achieved in two steps. First of all, the extraction of particular features capable of notify the main characteristics of the sound is needed. Then, a procedure to automatic detect this features has to be included.

4.3.1 Features extraction techniques

According to the state of the art exposed in Section 4.1 many works have been proposed for honey bees sound analysis. However, most of these works have analyzed the honey bee sound with techniques which have been developed thinking to the human voice, which have a totally different behaviour. With respect to human voice, which uses phonemes and has an harmonic content within

the range between 300 Hz and 4 kHz, honey bees communicate only with frequencies, within the range between 100 Hz and 3.5 kHz [173] but with some notable exception such as some warning signals which could reach much higher frequencies like 15-16 kHz [164]. For these reasons, in the next Section the different techniques exploited for features extraction will be explained. First, three well known approaches widely used in speech analysis (i.e., MFCC, Mel spectrograms and spectrograms) will be briefly explained, and then the innovative approaches introduced in this work will be discussed. These approaches, which are based on Hilbert Huang transform and wavelet transform, have shown to be more reliable and able to underline the different harmonic content of the bees signals.

Spectrograms

The first exploited technique is **spectrograms** following [158]. It allows a visual representation of the signal, showing its spectral contents as it varies with time. Spectrogram is based on short time Fourier transform (STFT), i.e., the signal is first windowed by means of a sliding time window then each frame is frequency transformed through discrete Fourier transform (DFT). A simple schematic of spectrogram algorithm is visible on Figure 4.6.



Figure 4.6: Schematic of spectrograms generation.

Mel spectrograms & MFCC

Mel spectrograms and mel frequency cepstrum coefficients (MFCC) are two well known features extraction techniques which are very similar, according to Figure 4.7 the MFCC add some further steps to the mel spectrogram algorithm. For mel spectrograms, the audio input is first buffered into frames. The frames are overlapped, then a Hamming Window is applied to each frame, and then the frame is converted to frequency-domain representation with FFT. Each frame of the frequency-domain representation passes through a mel filter bank. The spectral values output from the mel filter bank are summed, and then the channels are concatenated. For MFCC generation, the logarithm is applied to the output of the filterbank, then the discrete cosine transform (DCT) is applied and finally a normalization is applied.

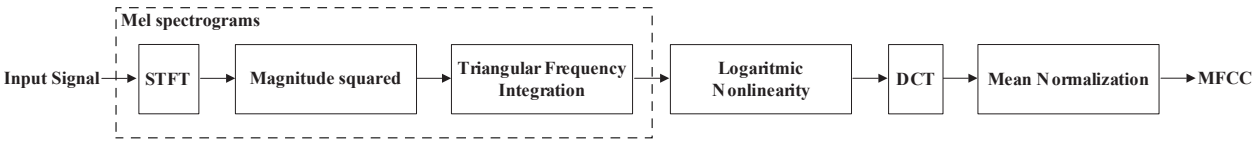


Figure 4.7: Schematic of mel spectrograms and MFCC algorithms, in particular if the output is saved before the logarithm the mel spectrograms obtained, otherwise the MFCC are obtained.

Hilbert Huang Transform

Hilbert Huang transform (HHT) is an algorithm for time-frequency analysis based on empirical mode decomposition (EMD) and Hilbert transform (HT). The schematic of this extraction procedure can be found in 4.8.

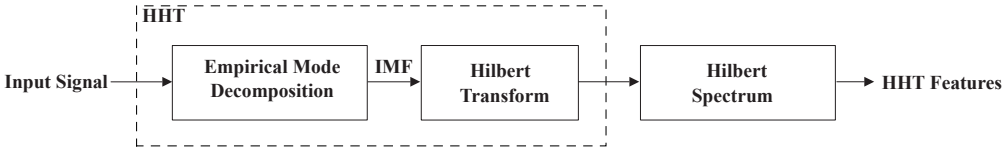


Figure 4.8: Schematic of Hilbert Huang transform algorithm.

EMD decomposes the original signal generating a series of basis functions called intrinsic mode functions (IMF). The IMFs are obtained through an adaptive procedure directly from the analyzed signal. Each IMF must satisfy the following properties: (A) The number of extrema and the number of zero-crossings must be either equal or differ at most by one extrema. (B) The mean value of the envelope defined by the local maxima and the local minima is zero.
Once the properties of the IMFs have been defined, for a given $x(n)$ signal with $n = 1, \ldots, N$, an iterative procedure for IMF estimation can be derived with the following steps:

1. Identify all local extrema.
2. Connect all the local maxima by a cubic spline.
3. Repeat the procedure to produce the lower envelope.
4. Estimate their mean $m_1(n)$.
5. The first estimation of the IMF can now be written as
   $h_1(n) = x(n) - m_1(n)$.
6. Repeat the procedure up to k times until the function
   $h_{1k}(n) = h_{1(k-1)}(n) - m_{1k}(n)$
   does not satisfy the IMF properties.
7. Now the first IMF component is equal to $c_1(n) = h_{1k}(n)$.
8. Remove from the original signal the component $c_1(n)$ obtaining the first residue $r_1(n) = x(n) - c_1(n)$.
9. Treat $r_1(n)$ as the new signal for the decomposition procedure.

Now the original signal can be reconstructed as a superposition of a certain number M of IMF plus a residue $r(n)$ i.e.,

$$x(n) = \sum_{j=1}^{M} c_j(n) + r(n). \tag{4.1}$$

Where $c_j(n)$ is the jth IMF. Once the signal has been decomposed, the Hilbert transform [138] is applied to each IMF and used for the estimation of the analytic signal $a_j(n)$ as follows:

$$a_j(n) = c_j(n) + j\mathcal{H}\{c_j(n)\} \tag{4.2}$$

where $\mathcal{H}$ indicates the Hilbert transform with $j = 1, \ldots, M$. Then, 4.2 can be expressed in polar coordinates, i.e., $a_j(n) = A_j(n)e^{i\phi_j(n)}$ where $A_j(n)$ is the instantaneous amplitude of the signal,

and $\phi_j(n)$ is the phase from which the instantaneous frequency can be derived according to:

$$f_j(n) = \frac{f_s}{2\pi}\big[\phi_j(n+1) - \phi_j(n)\big]. \tag{4.3}$$

Where fs is the sampling frequency. The result is a time-frequency distribution of a signal that has its own amplitude, designated as the Hilbert spectrum $H_j(w,n)$, where w is the frequency index. An example of the HHT algorithm is visible on Figure 4.9.

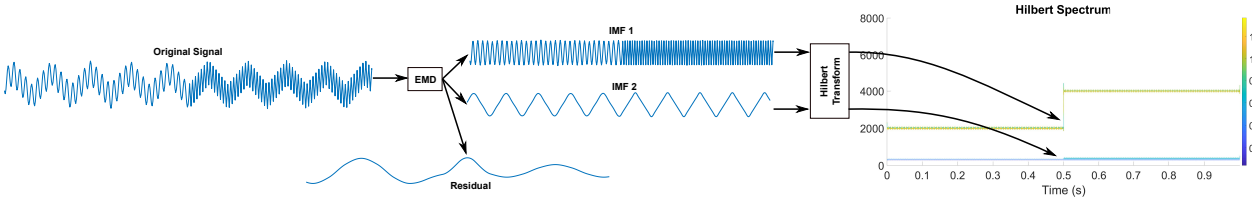


Figure 4.9: Application of the HHT algorithm to a test signal. The input signal is a sine wave which change its frequency with time combined with a triangular wave. The EMD decompose the signal into its components plus a residue, then the Hilbert transform is applied to the obtained IMFs and the Hilbert spectrum of each basis function is obtained.

Finally the Hilbert marginal spectrum is computed for each $H_j(w,n)$ according to:

$$H_j(w) = \sum_{n=1}^{N} H_j(w,n). \tag{4.4}$$

The results is a frequency amplitude representation for each Hilbert spectrum which can be represented in a three dimensional map where for each IMF the spectral content is clearly visible.

Wavelet transform

Another approach here proposed is based on wavelet transform (WT). Wavelets are mathematical functions that decompose the signal by mean of properly amplitude and time shifted basis functions called mother wavelet. Wavelets allow a time-frequency analysis of non stationary signals, and wavelet based approaches have been already applied successfully to animal sound analysis as reported in [67, 133]. For a given signal $x(t)$ wavelet transform is defined by the following formula:

$$X(s,u) = \frac{1}{\sqrt{\mid s \mid}} \int_{-\infty}^{+\infty} x(t)\psi^* \left(\frac{t-u}{s}\right) dt, \tag{4.5}$$

where s is the scale index, u the wavelet temporal index and ψ is the mother wavelet. Variation of scale index is related to the signal frequencies, i.e., an higher scale value means a dilated wavelet which approximate well low frequencies, while low scale value means a compressed wavelet which is closer to high frequencies. The choice of the mother wavelet is strictly related to the type of signal which has to be analyzed, and a complete description of the most important mother wavelets can be found in [66]. When a signal is analyzed by means of WT, a graphic time-frequency representation can be derived using the square module of the transformed signal, i.e.,

$$S(s,u) = \frac{1}{\mid s \mid} \left| \int_{-\infty}^{+\infty} x(t)\psi^* \left(\frac{t-u}{s}\right) dt \right|^2, \tag{4.6}$$

where $S(u, s)$ is defined as *scalogram*. In 4.5 and 4.6 index u and s change continuously realizing the **continuous wavelet transform (CWT)**. On CWT, scale values and time index change continuously, but can be discretized generating the **discrete wavelet transform (DWT)**, limiting the s and u values according to:

$$\begin{cases} s = s_0^{-m}, & \text{with } s_0 > 1,\ m \in \mathbf{Z} \\ u = nu_0 s_0^{-m}, & \text{with } u_0 > 0,\ n \in \mathbf{Z}. \end{cases} \tag{4.7}$$

DWT as already been used for features extractions and animal sound analysis [133, 188]. One of the main advantages of DWT is its reduced computational costs due to the fact that the transform can be easily implemented with a three structure filter bank as the one reported on Figure 4.10.

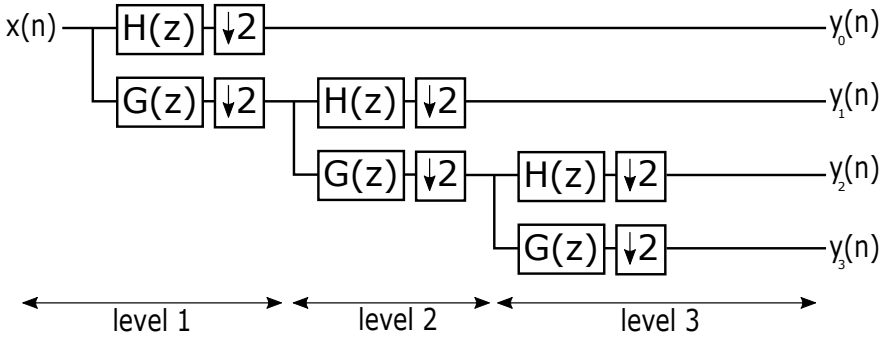


Figure 4.10: A 3-Level asymmetric dyadic analysis filter Bank. $G(z)$ and $H(z)$ are respectively the low-pass and the high-pass filters derived from the mother wavelet.

The filterbank implement a multiresolution analysis of the signal, i.e., the lower frequencies are analyzed with an higher resolution. Each branch of the filterbank is composed of an highpass filter $H(z)$ and a lowpass filter $G(z)$ and two decimators. Each level generates two distinct outputs: the detailed coefficients, i.e., the output of the high-pass filter, and the approximation coefficients, i.e., the output of the low-pass filter. The details ed coefficients are stored, while the approximation coefficients are used as input for the next decomposition level. Since the taps for the low-pass and high-pass filter are generated from the mother wavelet [206], the output of the filter bank correspond to the output of the discrete wavelet transform. Due to non uniform decimation, each wavelet coefficient vector has a different length. In order to obtain a more clear graphic representation, each DWT coefficients have been properly interpolated replicating its samples and thus obtaining a set of uniform length vectors with the same length of the input signal. For the mother wavelet many different wavelet families have been tested and then the discrete Meyer [141] has been selected due to its results with the analyzed signals.

4.3.2 Classifiers

After the features extractions, a classification algorithm is necessary in order to automatically analyze the honey bees sound. For this reason, in this work two different classifier algorithms have been exploited: the first based on support vector machine (SVM) and the latter on convolutional neural network (CNN). For this part, a collaboration with the Queen Mary institute of London has been carried out. In the next Sections, the parameters of SVM classifier and the architecture of the

neural network adopted will be presented.

Support vector machine

SVM is a rather simple but powerful method that discriminates data in its classes based by finding a boundary that maximizes the distance between samples from the different classes. An important aspect of SVMs is the possibility of performing this procedure for data that is not linearly separable, through means of a kernel. As indicated in [150], SVMs are good classifiers for this problem when used with a radial basis function (RBF) kernel. Here, all SVMs are computed with the RBF kernel, with penalty parameter (C) of 1 and the gamma parameter of 1/(number of features).

Convolutional neural network

The network architecture, is visible on Figure 4.11, and is derived from the one presented in [151]. Tables 4.2(a), 4.2(b) illustrates the detailed network architecture, with a focus on the output sizes and the parameters values used in each layer.

Table 4.2: Proposed Network Architecture with two different inputs: (a) MFCC coefficients with 20 frequency and 62 time frames and (b) a 256x256 pixels RGB images.

(a)

Layer	Output	Description
input	(20, 62)	MFCCs
convolution	(20, 62, 16)	3×3 kernel size, 16 filters
max-pooling	(10, 31, 16)	3×3 pool size, 2×2 strides
dropout	(10, 31, 16)	0.5 rate
convolution	(10, 31, 16)	3×3 kernel size, 16 filters
max-pooling	(5, 16, 16)	3×3 pool size, 2×2 strides
dropout	(5, 16, 16)	0.5 rate
convolution	(5, 16, 16)	3×1 kernel size, 16 filters
max-pooling	(3, 8, 16)	3×1 pool size, 2×2 strides
dropout	(3, 8, 16)	0.5 rate
convolution	(3, 8, 16)	3×1 kernel size, 16 filters
max-pooling	(2, 4, 16)	3×1 pool size, 2×2 strides
dropout	(2, 4, 16)	0.5 rate
dense	(256)	256 units
dropout	(256)	0.5 rate
dense	(32)	32 units
dropout	(32)	0.5 rate
dense	(1)	1 unit

(b)

Layer	Output	Description
input	(256, 256, 3)	RGB Image
convolution	(256, 256, 16)	3×3 kernel size, 16 filters
max-pooling	(128, 128, 16)	3×3 pool size, 2×2 strides
dropout	(128, 128, 16)	0.5 rate
convolution	(128, 128, 16)	3×3 kernel size, 16 filters
max-pooling	(64, 64, 16)	3×3 pool size, 2×2 strides
dropout	(64, 64, 16)	0.5 rate
convolution	(64, 64, 16)	3×1 kernel size, 16 filters
max-pooling	(32, 32, 16)	3×1 pool size, 2×2 strides
dropout	(32, 32, 16)	0.5 rate
convolution	(32, 32, 16)	3×1 kernel size, 16 filters
max-pooling	(16, 16, 16)	3×1 pool size, 2×2 strides
dropout	(16, 16, 16)	0.5 rate
dense	(256)	256 units
dropout	(256)	0.5 rate
dense	(32)	32 units
dropout	(32)	0.5 rate
dense	(1)	1 unit

The output dimensions, of Table 4.2(a) refer to an input obtained after applying the MFCCs technique for the feature extraction, with 20 frequency coefficients and 62 time frames, while Table 4.2(b) refers to an input considering a RGB image of 256x256 pixels. A change in the feature extraction method only reflects in a change of the input dimensions and it doesn't influence the remaining network structure. As it can be seen from 4.2(a), there are four convolutional layers composed of 16 filters, each one followed by a max-pool layer and a dropout layer; the first two convolutional layers have a kernel size of 3×3 and a max-pool size of 3×3 with a 2×2 stride, whereas the last two convolutional layers have a 3×1 kernel size and a pool size of the same dimensions

(2×2). The last block, instead, has a three dense layers composed of 256, 32, 1 units respectively. To regularize the training procedure, dropout layers with a rate of 0.5 were used after each max-pooling layers and after the first two dense layers. A leaky rectifier linear unit (LReLU) [134] was employed for all convolutional and dense layers except the last one, which uses a sigmoid function for the classification. The network training was performed by optimizing the binary crossentropy between predictions and targets; root mean square propagation (RMSProp) [96] was used as optimizer with a learning rate of 0.001 and the batch size was set to 145. The training was stopped when the validation loss did not decrease for more than five epochs. As regards the initialization of the network weights, a uniform distribution with zero biases [95] was used for both the convolutional and dense layers. 4.11 shows a block schematic of the proposed architecture.

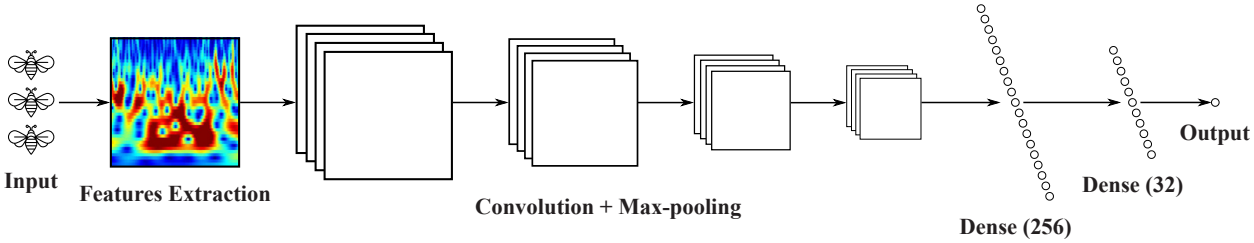


Figure 4.11: Schematic of the proposed convolutional neural network, containing four convolution layers each one followed by max-pooling layers. The last three layers are dense layers, with 256, 32, 1 units respectively.

Dataset split

For all the proposed experiments, the interest is constructing a system for a real-world scenario. For this reason it has been evaluated how well the classifiers are able to generalize to unseen hives. Thus, besides randomly splitting the dataset between train and test sets, it has been also implemented a "hive-independent" splitting scheme. This means having training samples belonging only to certain hives, and testing using samples from other, unseen hives. For the random scheme, a test size of 5% of the total amount of data is used and, when applying the SVM classifier, all remaining data (95%) is used in a single training set. For the CNN implementation, the remaining data is further split in half between the training and validation sets. For the hive-independent scheme, the data is split in two, according to which hive they belong to; one is kept for training and the other for testing. In the case of the CNNs, a validation set, used for early stopping purposes, is obtained by randomly selecting 10% of the training set data.

Evaluation metrics

To evaluate the performances of the two classifiers with different algorithms, several metrics have been considered. The first one is the area under the curve (AUC) score [140], which is a metric that describes the performance of a binary classifier across all possible classification thresholds. In particular, a model whose predictions are 100% correct has an AUC score of 1.0; on the contrary, a model whose predictions are completely wrong has an AUC score of 0.0 Precision, recall and F1 measure, which are evaluation metrics widely used in the audio recognition task [30, 90], have also

been computed. They are defined respectively as:

$$P_l = \frac{tp_l}{tp_l + fp_l} \tag{4.8}$$

$$R_l = \frac{tp_l}{tp_l + fn_l} \tag{4.9}$$

$$F1_l = \frac{2P_l R_l}{P_l + R_l} \tag{4.10}$$

where tp_l is the number of the true positives, fp_l the number of false positives and fn_l the number of false negatives for each of the two considered labels l (i.e., queen-bee presence or absence). In other words, precision quantifies how many of the predicted results are relevant, while recall measures how many truly relevant results are returned. F1 measure quantifies instead the overall performance of the network, and it is computed as the harmonic mean of precision and recall.

4.3.3 Experimental results

The experimental results obtained will be presented in the next Sections focusing on different colony health status. Many of these results have been already published in [53, 151, 200] other will be presented here and published in a future work. In particular two case study have been analyzed, i.e., the orphaned colony situation and the swarming phenomena.

4.3.4 Orphaned colony

Most of the experiments have been focused on the orphaned colony situation. In this case, the queen bee is missing and if a new queen bee is not introduced quickly the whole colony is going to die within 3-4 weeks. Data analyzed came from two different colonies each acquired both in a normal situation, i.e., with the queen bee and in an orphaned situation i.e., without the queen bee.

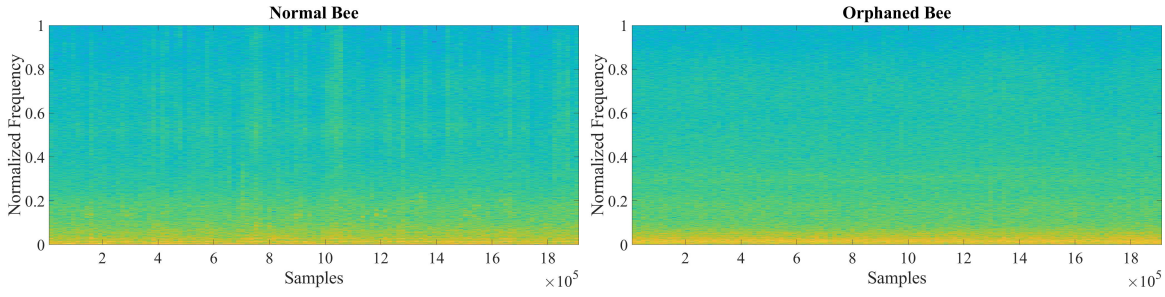
Features extractions performances

In the next Section, the first case study will be considered focusing on features extraction. During years, the approaches have been improved, starting from the first simpler experiments with HHT to the last results where more complex algorithm based on wavelet have been developed. In the next Sections the evolution of these techniques will be presented and discussed.

First features extraction experiments The first experiments have been presented in [151], they involves the application of both SVM and CNN and will be resumed in the next Section. In this early work, the HHT features have been used in a simpler way than how it was exposed in the Section 4.3.1. Indeed, for each ten minutes recording the signal has been divided into 1 s frames, each frame has been decomposed with EMD using a ten level decomposition. Each IMF has been transformed using the Hilbert transform and then the mean frequency and its amplitude has been estimated for each IMF obtaining ten mean frequencies and ten mean amplitude for each frame. These HHT based features have been used and compared with a 20 coefficients MFCC algorithm, and a mel spectrograms representation with 120 frequency bands.

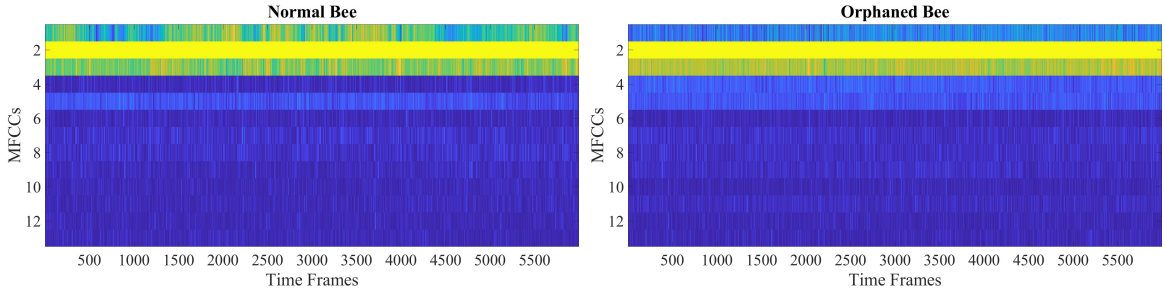
Second features extraction experiments Starting from the results presented in [151], the work has been focused on the development of innovative features extraction approaches involving the technique presented in the Section 4.3.1. The results here presented have been published in the following article [200]. For these experiments, no classifiers have been applied, focusing on a qualitative evaluations of the performances of different features extractions approaches. Data came from the same dataset previous discussed considering one minutes time frames, in particular data from the normal colony came from the 8 May 2018 at 12.30, while orphaned colony recording is from 13 July 2017 at 12.30, the same hour of the day has been chosen to avoid different behaviour related to different time of the day. STFT and MFCCs features extraction approaches have been considered as reference. The analysis has been performed on MATLAB, using the algorithms exposed on Section 4.3.1 except for mel spectrograms, with the following parameters for image generations:

- Spectrograms → Han window, 50% overlap, FFT size of 640 samples.
- MFCC → hop length 512 samples, FFT size 2048 samples, 20 coefficients.
- HHT → 5 level of decomposition.
- CWT → Morlet as mother wavelet.
- DWT → 10 level filterbank, discrete Meyer as mother wavelet.



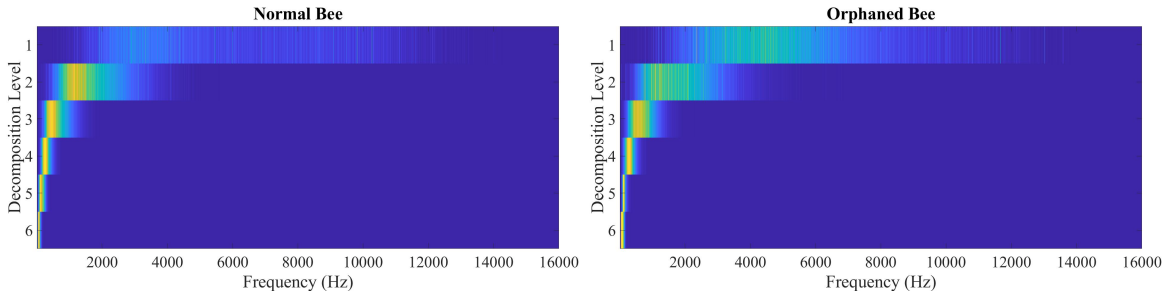
(a) (b)

Figure 4.12: Spectrogram of the sound generated in the beehive considering the two conditions.



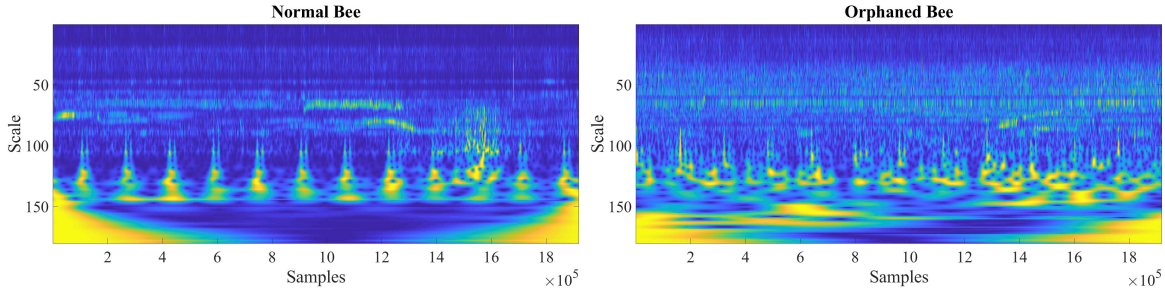
(a) (b)

Figure 4.13: Features extraction exploiting mel frequency cepstral coefficients approach.



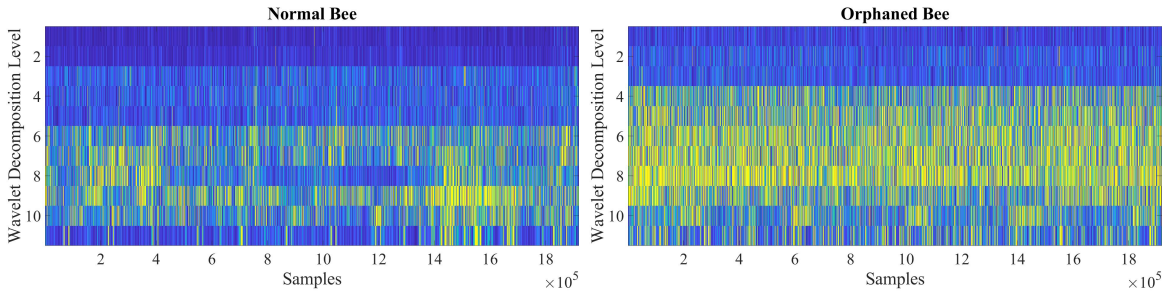
(a) (b)

Figure 4.14: Features extraction exploiting Hilbert-Huang transform in terms of decomposition level (i.e., intrinsic mode function) representing a different spectral content.



(a) (b)

Figure 4.15: Features extraction exploiting continuous wavelet transform with Morlet as mother wavelet.The number of scale values has been determined automatically in order to obtain a complete signal representation.



(a) (b)

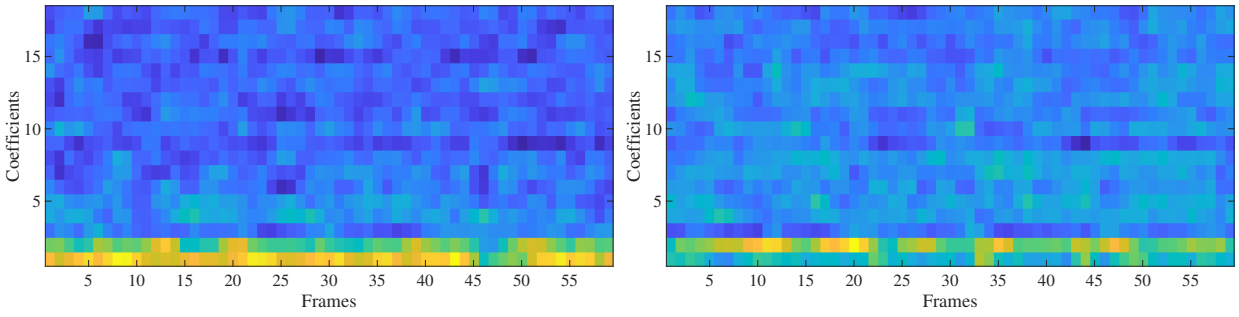
Figure 4.16: Features extraction exploiting discrete wavelet transform with a 10-level dyadic filter bank. The 11st level corresponds to the approximation coefficient.

Figure 4.12 shows the spectrograms obtained from the analyzed data. Comparing the normal situation with the orphaned bees some differences are visible especially in the higher frequency regions. In particular, some lines are evident on Figure 4.12(a) that are not present on Figure 4.12(b). Regarding the lower part of the spectrum, there is more energy on Figure 4.12(b) than on Figure 4.12(a) also in this case. These differences in the spectral content are more clearly visible in the MFCC coefficients of Figure 4.13. In particular, focusing on the first five coefficients that are related to the lower frequencies, the first coefficient seems to have less energy in the orphaned

colony, but the third, fourth and fifth coefficients seem to have more energy respect to the normal colony. The other coefficients have the same energy in both situations. Figure 4.14 shows the obtained results for HHT-based approach in terms of decomposition level and frequency range. Comparing the two situations, the harmonic content is different since the orphaned colony shows more harmonics at the higher and lower frequencies in contrast with the normal colony that has more energy in the middle frequencies, i.e., second level of IMF. With respect to other techniques, this approach gives a better estimation of the fundamental frequencies of the original signal. The shifting process used for empirical mode decomposition [108] removes riding waves and smoothes uneven amplitudes, leaving only the mode of oscillation contained in the original signal. In comparison with spectrogram, the HHT shows differences in the harmonic content with a better resolution, while with respect to MFCC, the HHT is more clear representing the harmonics actually included in the analyzed signal. Figure 4.15(a) shows the WT-analysis from a normal colony, while 4.15(b) shows scalograms from orphaned colony, and it is evident that we have obtained different spectral contents in the two situations. Orphaned colony Figure shows a behaviour with more harmonics at lower frequencies, while at the higher frequencies the energy spread looks more uniform. On the contrary, the normal bee situation seems to have less and more defined frequencies. It should be noted that the highest energy region at the border of scales values must not be considered since is due to border effect, as also reported in the state of the art [136].

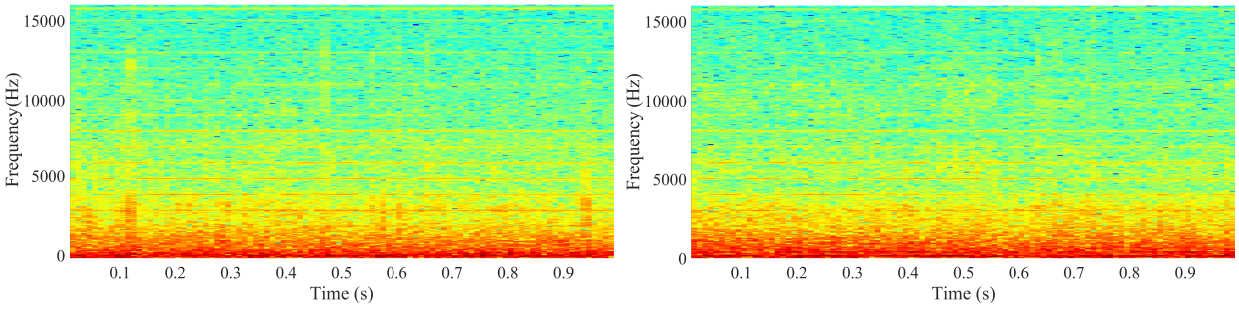
Recent features extraction experiments Starting from the results presented in the previous Section a further step has been taken in order to allow an easy application of the proposed features to the above mentioned classifier algorithms. For this reason, some changes have been implemented, i.e., one second recordings have been analyzed, mel spectrograms have been added with the following parameters for images generation:

- Spectrograms → Han window, 50% overlap, FFT size of 640 samples.
- Mel Spectrograms → Hamming window, window length 960 samples, FFT size of 960 samples, overlap 640 samples, 32 bands filterbank.
- MFCC → hop length 512 samples, FFT size 2048 samples, 18 coefficients (discharged the first and second coefficient, i.e. the coefficient from 2 to 20 have been considered).
- HHT → 5 level of decomposition.
- CWT → Bump as mother wavelet.
- DWT → 10 level filterbank, discrete Meyer as mother wavelet.



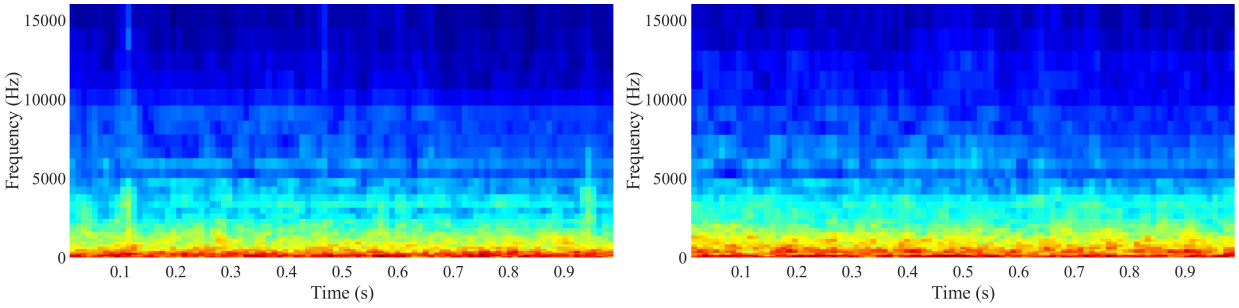
(a) (b)

Figure 4.19: Mel frequency cepstrum coefficients feature extractions considering (a) normal colony, and (b) a orphaned colony.



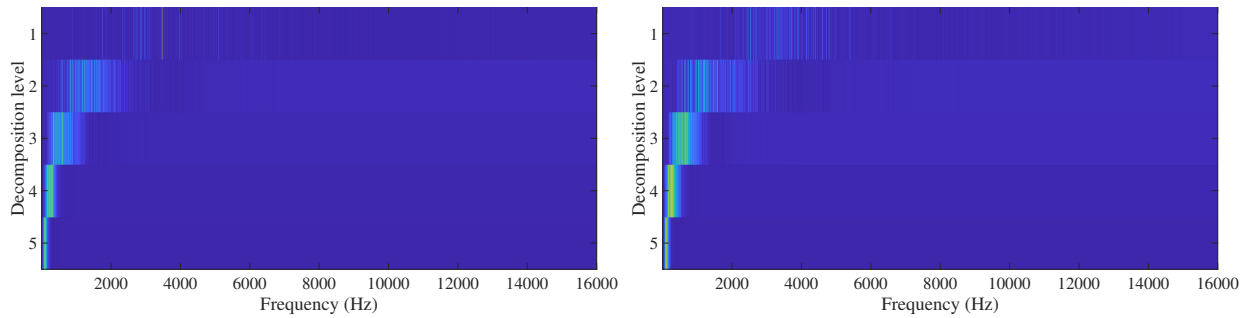
(a) (b)

Figure 4.17: Spectrograms considering (a) normal colony, and (b) a orphaned colony.



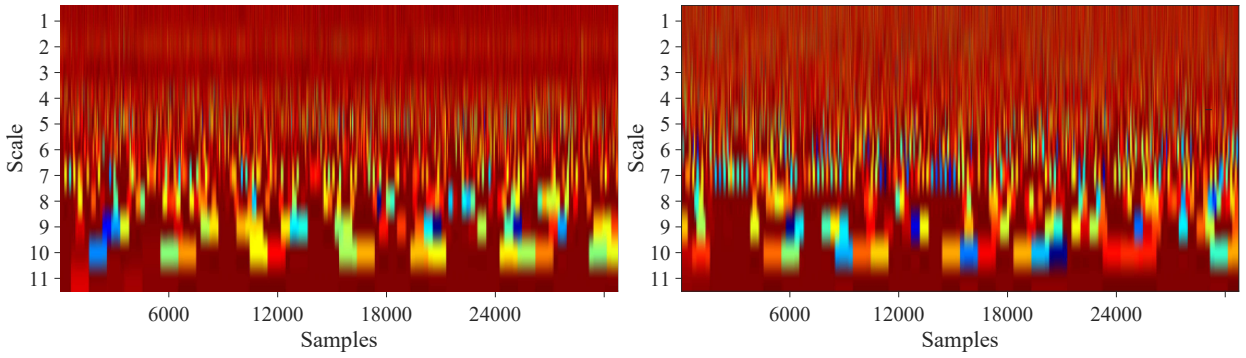
(a) (b)

Figure 4.18: Mel spectrograms considering (a) normal colony, and (b) a orphaned colony.



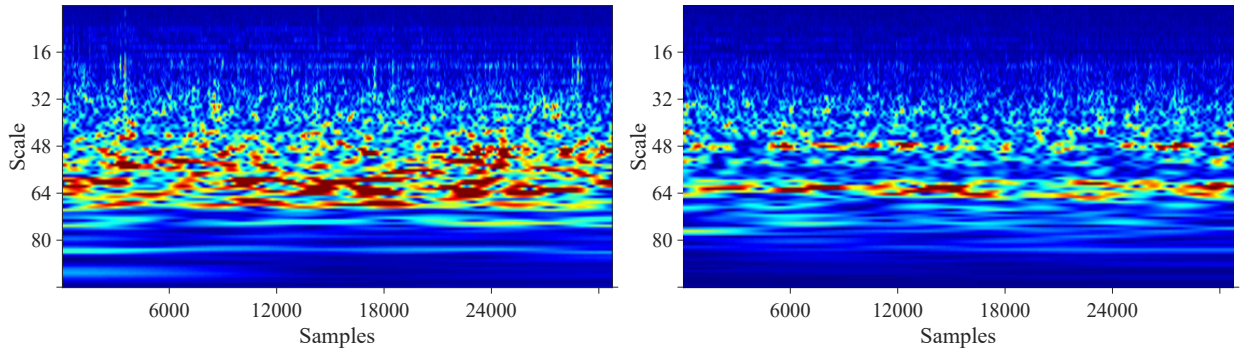
(a) (b)

Figure 4.20: Hilbert-Huang transform feature extractions considering (a) normal colony, and (b) a orphaned colony.



(a) (b)

Figure 4.21: Discrete wavelet transform feature extractions considering (a) normal colony, and (b) a orphaned colony.



(a) (b)

Figure 4.22: Continuous wavelet transform feature extractions considering (a) normal colony, and (b) a orphaned colony.

Figures 4.17 shows the spectrograms obtained from the analyzed data. Comparing the normal situation with the orphaned bees some differences are visible especially in the higher frequency regions. In particular, some lines are evident in Fig. 4.17(a) that are not present in Fig. 4.17(b), lower frequency shows also a different energy spread, in particular orphaned colony seems to have more energy and a more uniform energy distribution.

Taking into consideration mel spectrograms of Figures 4.18, the same difference between normal colony on Figure 4.18(a) and orphaned 4.18(b) of spectrograms are visible. However, the non uniform scale resolution of mel spectrograms allow a clearer analysis of the different harmonic content of the two signals.
Figures 4.19 shows the MFCC results. According to the experimental results the first two coefficients have been discharged. Comparing the two Figures the first two coefficients (the third and the fourth considering the two discharged) shows the major difference. In particular on Figure 4.19(a), with the normal colony recording, the first and coefficient has more energy while on the orphaned colony, on Figure 4.19(b) is the second the most energetic. Figures 4.20 shows the obtained results for HHT-based approach in terms of decomposition level and frequency range. Comparing the two situations, the harmonic content is different since the orphaned colony on Figure 4.20(b), shows more harmonics at the higher and lower frequencies in contrast with the normal colony, on Figure 4.20(a) that has more energy in the middle frequencies, i.e., second level of IMF. With respect to other techniques, this approach gives a better estimation of the fundamental frequencies of the original signal. The shifting process used for empirical mode decomposition [108] removes riding waves and smoothes uneven amplitudes, leaving only the mode of oscillation contained in the original signal. In comparison with spectrograms and mel mel spectrograms, the HHT shows differences in the harmonic content with a better resolution, while with respect to MFCC and wavelet transform, the HHT is more clear representing the harmonics actually included in the analyzed signal.
Focusing on wavelet transform, both discrete and continuous case have been considered and applied. Figure 4.22 shows the results for the CWT case. Figure 4.22(a) shows the WT-analysis from a normal colony, while Figure 4.22(b) shows scalograms from orphaned colony, and it is evident that we have obtained different spectral contents in the two situations. Orphaned colony Figure shows a behaviour with less and more limited frequencies, while normal situation shows more harmonics. Finally DWT approach shows different harmonic content content, in particular the difference between normal colony on Figure 4.21(a) and the orphaned one on Figure 4.21(b) are more visible on medium-high frequency, i.e., on lower scale coefficients, where the normal colony seem to have less energy on the first and third coefficient.

Classifiers performances

The above mentioned features extraction techniques have been used in combination with the classifiers of Section 4.3.2, results will be presented and discussed in the next Section.

First classifier results Early classifier results have adopted the first features presented, using both SVM and CNN. However, due to its simplicity, the HHT features extracted in these experiments, have been used only with SVM, leaving to MFCC and Mel spectrograms all CNN based experiments. When using **CNNs** it is important to consider the amount of data needed to train such large networks. To meet these constraints, the original 10 min audio samples have been segmented into 1 min segments. Feature extraction is performed on this new set of shorter samples. To mitigate the loss of temporal context that the shorter segments bring and which was deemed important for representing beehive sounds, it has been adapted a procedure to obtain long-term features as the one introduced in [28], where each spectrogram is transformed in a stack of averaged

slices over time. To increase the generalization of the models, the data have been carried out with the augmentation on the dataset by creating 3 versions of each sample with a random pitch shift between -1 and 1 semitone. The normalization of the data has been performed frequency-wise with z-score normalization along the training set samples. The normalization parameters computed for the training set are then used to apply the same transformation to both validation and test sets. The experiments carried out are the following:

CNN_MFCCs20: Uses as input data 20 MFCCs.

CNN_MEL120: Input samples are obtained from Mel spectra with 120 frequency bands.

CNN_LOG_MEL120: This configuration uses the log-Mel spectra computed with 120 frequency bands.

The results are visible on Figure 4.23

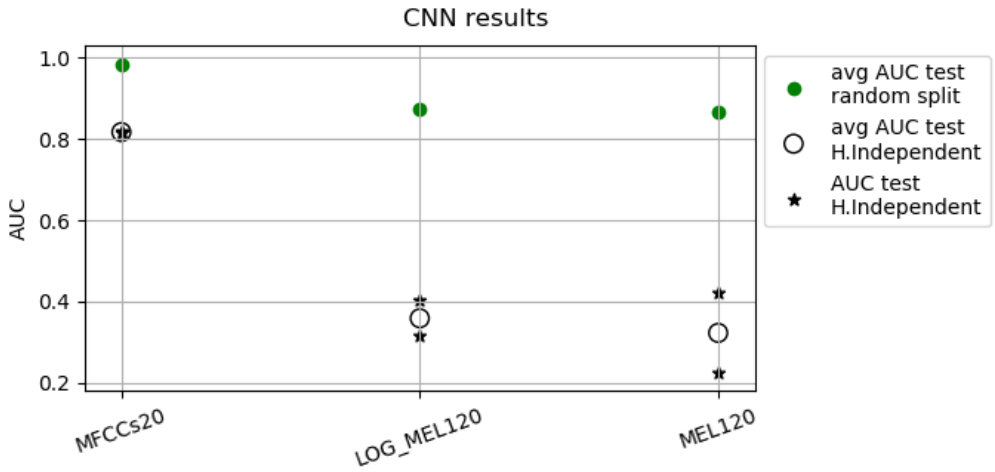


Figure 4.23: The ⋆ represents the AUC score on the test set for each fold of the hive-independent setup and CNN experiment. The ∘ represents the average AUC score on test sets over the two folds and the • reports the average AUC score of the test sets across the two folds of the random split setup.

The resulting AUC scores for the experiments with CNNs are shown on Figure 4.23 For the three CNN experiments carried out in a random split setup, the averaged AUC scores (•) show the ability of the procedure in this classification task for both MFCCs and Mel spectra. The challenge in predicting the beehive state in hives different than the ones where the model was trained is evident when observing the resulting averaged test AUC score in a "hive-independent" setup, (∘); overall the results in this setup decrease when comparing with the random split. Once again, as Figure 4.23 **[MFCCs20]** shows, the use of MFCCs as feature is especially useful in this problem, while Mel spectra when used as features do not appear to generalize well to unseen hives given the resulting AUC lower than 0.5. Focusing on **SVM experiments**, several different approaches have been tested. using also HHT as previous described. The features combination used in this experiments are the following:

SVM_MFCCs20: each input sample is a vector of 20 MFCCs resulting from averaging the 10min MFCC coefficients over time.

SVM_HHTdwns20: the HHT spectrogram obtained for a 10min audio recording is aggregated over time and its maximum frequency limited to 6000 Hz. This process results in a vector with 6000 frequency bands representing one 10min recording. The frequency bands are further downsampled into 20 HHT bands, in order to reduce feature dimension.

SVM_MFCCs20_HHTdwns20: a combination of both representations described above is used. In total, a sample corresponding to a 10 min audio recording is represented by a vector of size 40 which is the concatenation of both feature vectors.

SVM_MEL120dwns20: each Mel spectrogram of a 10 min audio recording is averaged over time resulting in a 120-dimensional vector which is further downsampled into 20 bands.

SVM_LOG_MEL120dwns20: the log-Mel spectrograms of the 10 min audio recordings are averaged over time, and the frequency bands downsampled into 20 bands.

Results with these approaches are visible on Figure 4.24.

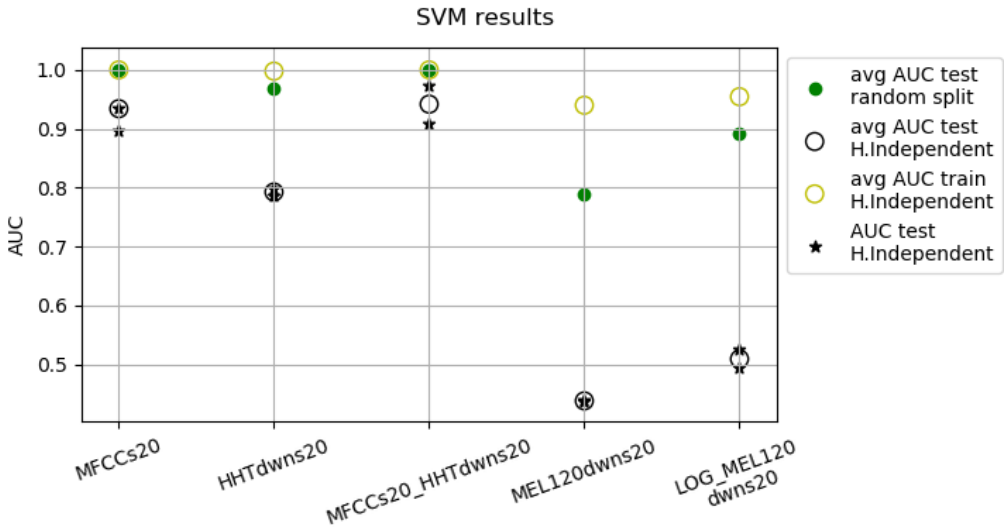


Figure 4.24: The ⋆ represents the AUC score on the test set for each fold of the hive-independent setup and SVM experiment. The ○ and ○ represent the average AUC score over the two folds in both train and test sets, respectively. The ● reports the average AUC score of the test sets across the two folds of the random split setup.

Observing the results for the averaged test AUC score in the random split setup (●), MFCC and HHTs yield similar classification results in this setup. The main challenge however is to generalize allowing good classification performances even with data unseen before from the network, in this case performance of the classifier could be worsen. This aspect is confirmed here with the reported lower values of the averaged test AUC scores in the “hive-independent" setup (○) for every experiment conducted. Despite the lower scores for the “hive-independent" setup when compared with random splits, results indicate that the SVMs are successful in generalizing to unseen hives for most feature combinations. An interesting result is presented in for the features combination **[MFCCs20_HHTdwns20]** where the averaged test AUC score (~ 0.94) in this setup is better than the one reported in **[MFCCs20]** (~ 0.91). These results indicate that the combination of these two features results in better predictions in a “hive-independent" setup than the predictions carried out with classifiers learning from each feature individually. It is also of notice in **[MEL120dwns20]** and

[LOG_MEL120dwns20] the inadequacy of using these representations together with SVMs for this classification task on the “hive-independent" setup.

Recent classifier results The work here presented will be published in the future. In comparison with [151] several differences are present. First of all the more recent features extraction techniques of the previous Sections have been applied, then also the following changes have been applied in order to improve the system performances:

- Only the CNN has considered for the experiments.
- The preprocessing Section has been performed removing the pitch-shift data augmentation, since according to [87, 107, 142] most of the information is carried by the signal frequencies, and a pitch shifting in these signals could be misleading.
- One second frames has been considered in order to increase the number of training/testing samples (approximately 345 000 files in total), this choices is related to some experimental results and previous work [59].
- the data obtained from the feature extraction has been used as image instead of data matrix [14, 120, 147, 179] to avoid the direct management of the features coefficients that are huge for some of the proposed approaches;
- Since the dataset includes recordings from two microphones in a different position inside the hive with a precise entomological meaning, data from both microphones have been analyzed independently.

The six feature extraction techniques have been evaluated comparing the AUC scores obtained by analyzing, individually, microphone 1 and 2. As it can be seen from 4.3, for the random split case the spectrograms have the highest AUC score, even if the CWT and the Mel-spectrograms are quite similar; for the hive-independent setup, on the contrary, the CWT has the highest score on channel 1, while on channel 2 seems to perform better the Mel Spectrograms. As regards the HHT, for both the random split and the hive-independent setups the performances are worse than the other two techniques, symptom that this particular method is not well suited for this classification task, this approach probably should work better with other colony health status (i.e., the swarming as reported in [53]). Since the hive-independent setup, as said, is the most interesting for a real-world implementation, CWT as a feature extraction method applied to honey bees is definitely to be taken into consideration. Another interesting thing that can be seen from 4.3 is that, for the random split experiments, all six approaches have shown better performances on channel 2, while for the random split this is not true. Focusing on precision, recall and F1 the scores are presented in table4.4 for the orphaned colony, and on table 4.5 for the normal colony. The result for the random split experiment are positive for almost all the technique proposed, again as for the AUC scores the best performances are reached by spectrograms and CWT. It is interesting to underline that comparing the two tables for the the hive independent experiment the network tends to classify most of the images on the normal colony class. However, while spectrograms and Mel spectrograms seems to suffer more this problem, the approaches based on HHT and wavelet seems to have less problems.

Table 4.3: AUCs comparison between the six feature extraction techniques.

	AUC			
	Random Split		Hive-Independent	
	Mic.1	Mic.2	Mic.1	Mic.2
Spectrograms	**0.9967**	**0.9997**	0.3247	0.7082
Mel-spectrograms	0.9866	0.9950	0.4429	**0.7858**
MFCCs	0.9757	0.9906	0.5060	0.6555
HHT	0.9195	0.9514	0.6301	0.6290
DWT	0.9189	0.9480	0.7589	0.4892
CWT	0.9814	0.9968	**0.7648**	0.5937

Table 4.4: Precision, recall, F1 measures comparison for the no-queen-bee class.

	No-queen-bee class											
	Random split						Hive-Independent					
	Mic. 1			Mic. 2			Mic. 1			Mic. 2		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
STFT spectrograms	**0.9864**	0.9439	**0.9647**	**0.9967**	**0.9928**	**0.9948**	0.3025	0.2378	0.2663	0.6163	**0.8719**	**0.7221**
Mel spectrograms	0.8811	**0.9799**	0.9279	0.9902	0.9904	0.9699	0.3243	0.2085	0.2538	**0.7043**	0.7272	0.7155
MFCCs	0.9429	0.8990	0.9204	0.9620	0.9572	0.9596	0.4857	0.3553	0.4104	0.6248	0.2254	0.3313
HHT	0.8570	0.7892	0.8218	0.8736	0.9014	0.8873	0.6015	0.4724	0.5292	0.6012	0.5875	0.5821
DWT	0.8046	0.8967	0.8481	0.8256	0.9353	0.8770	0.7155	**0.6445**	**0.6781**	0.4856	0.2680	0.3454
CWT	0.9525	0.9020	0.9266	0.9700	0.9795	0.9747	**0.7353**	0.4911	0.5889	0.5540	0.5238	0.5385

Table 4.5: Precision, recall, F1 measures comparison for the queen-bee class.

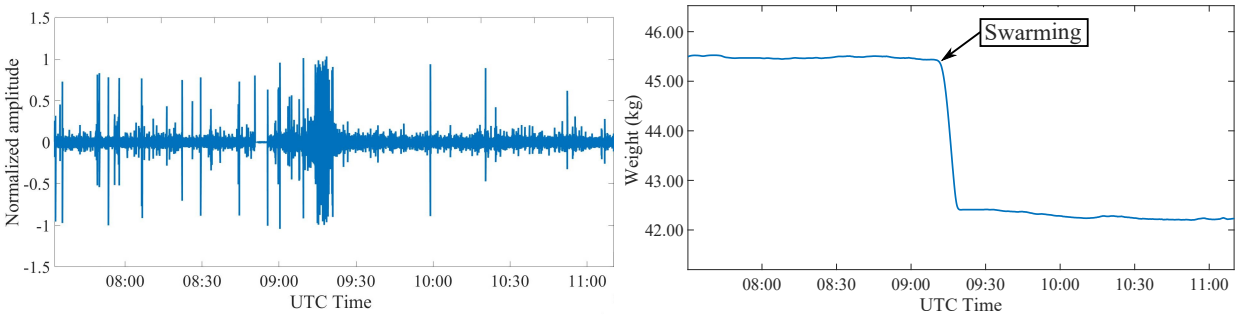
	Queen-bee class											
	Random split						Hive-Independent					
	Mic. 1			Mic. 2			Mic. 1			Mic. 2		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
STFT spectrograms	0.9462	**0.9869**	**0.9662**	**0.9929**	**0.9968**	**0.9948**	0.3694	0.4488	0.4052	**0.7791**	0.4543	0.5739
Mel spectrograms	**0.9774**	0.8678	0.9193	0.9524	0.9906	0.9711	0.4145	0.5633	0.4776	0.7165	0.6930	**0.7046**
MFCCs	0.9035	0.9455	0.9240	0.9574	0.9621	0.9598	0.4897	0.6218	0.5479	0.5260	**0.8640**	0.6539
HHT	0.8025	0.8668	0.8334	0.8968	0.8680	0.8822	0.5638	0.6854	0.6187	0.5875	0.6239	0.6052
DWT	0.8830	0.7817	0.8293	0.9254	0.8024	0.8595	**0.6751**	0.7424	**0.7071**	0.4927	0.7146	0.5832
CWT	0.9069	0.9551	0.9304	0.9793	0.9696	0.9744	0.6165	**0.8223**	0.7046	0.5462	0.5760	0.5607

4.3.5 Swarming

Another important event which has been analyzed is swarming. Swarming takes place when the queen bee leaves the hive alone with at least an half of the whole colony and starts searching for a new place to live. Several causes could trigger swarming, but as reported in [191] it results in a loss of honey production, and sometimes could lead to the death of the whole colony. For this reasons, swarming monitoring is an important beekeeper's activity and an early detection of this phenomena could help their work. One of the most important parameter which can be acquired for swarming monitoring is weight: indeed swarming is revealed by a slightly decrease in weight due to the bees leaving the hive. Despite weight is a reliable way to detect swarming, it requires a specific measurement system that has to be very precise (i.e., an accuracy of few grams) and which must have a range in the order of tens of kilograms, this requirements are necessary since the weight monitoring is used not only for swarming, but also for the colony monitoring and must be able to detect also small changes related to normal bees operations. Furthermore it should resist

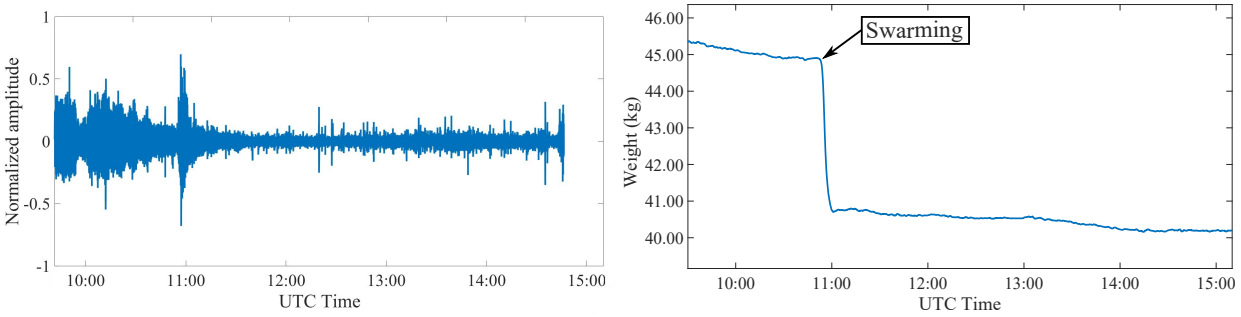
to atmospheric agent and must not interfere with bees activity. However weight is not the only parameter which suddenly change during swarming, as reported in [87]. Sound can also be a good indicator of swarming, respect to weight, requires only a microphone which can be installed easier in the colony. For this reasons several experiments have been carried out to monitor the swarming, in particular two swarming events have been detected and analyzed using both sound and weight.

Weight and sound envelope swarming detection A simple, but effective analysis could be done by analyzing directly the amplitude of the recorded signals and by making a comparison between the weight variations and the sound variations. These experiments are visible on Figures 4.25, 4.26.



(a) (b)

Figure 4.25: Measurement during the swarming event in colony 1 during a day in April 2019 considering sound variations (a) and the related weight variations (b).



(a) (b)

Figure 4.26: Measurement during the swarming event in colony 2 during a day in June 2019 considering considering sound variations (a) and the related weight variations (b).

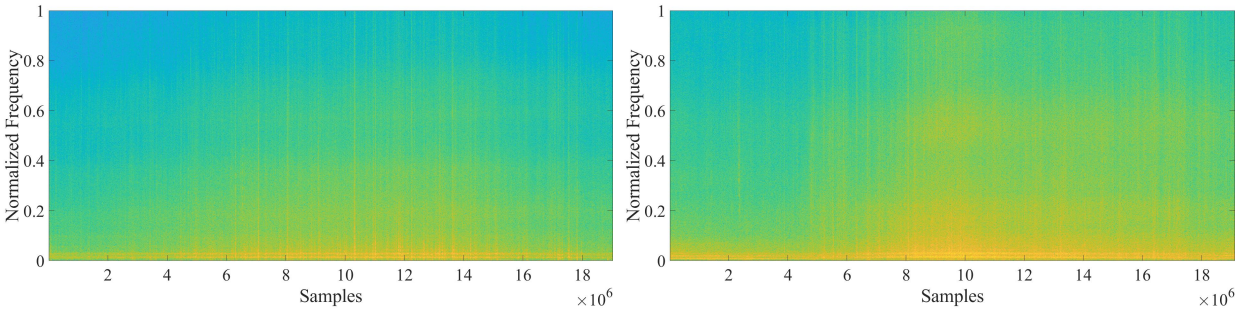
From the previous exposed Figures is clearly visible that the quickly drop in the weight is precisely synchronized with the change in the sound amplitude. Despite this simple amplitude analysis has already proposed as swarming detection mechanism [146] its simplicity could easily produce false positive (i.e. the sound of a car horn), especially when sound is used as a standalone system without the weight monitoring.

Features extraction swarming detection The problems exposed in the previous Section, shows that a sound based swarming detection, could not work properly using only an envelope detection,

more accurate algorithms are necessary. In the next Section, the two swarming events have been analyzed with the algorithms explained in Section 4.3.1. For these experiments the ten minutes recording during swarming have been analyzed with the following parameters:

- Spectrograms → Hamming window, 50% overlap.
- MFCC → hop length 640 samples, FFT size 9600 samples, 13 coefficients.
- HHT → only second IMF analyzed.
- CWT → Morlet as mother wavelet.
- DWT → 10 level filterbank, fourth level fourth level Daubechies mother wavelet [65].

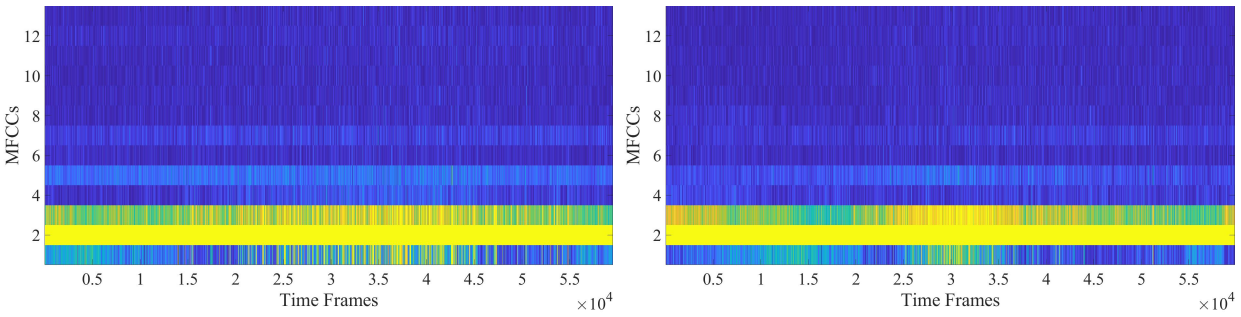
The first approach is the spectrogram and the results are visible on Figures 4.27.



(a) (b)

Figure 4.27: Spectrograms produced during swarming, considering the swarming of colony 1 (a) and colony 2 (b).

The increase in medium and low frequencies during the swarming is clearly visible. As visible on figures 4.28 also the first coefficients of MFCC report an increase in the energy during the swarming.

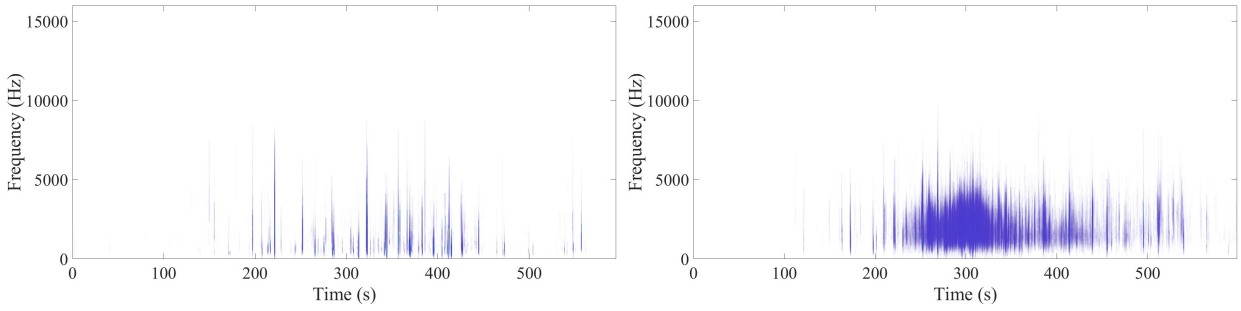


(a) (b)

Figure 4.28: MFCC produced during swarming, considering the swarming of colony 1 (a) and colony 2 (b).

Using more advanced techniques such as HHT the sound produced during swarming is even more clear. Figures 4.29 shows the results. In this case, since the signal has a quickly time frequency variation, the Hilbert marginal spectrum has not been used. Figures 4.29(a) and 4.29(b) shows the Hilbert spectrum of the second IMF which is related to medium low frequencies. With respect to

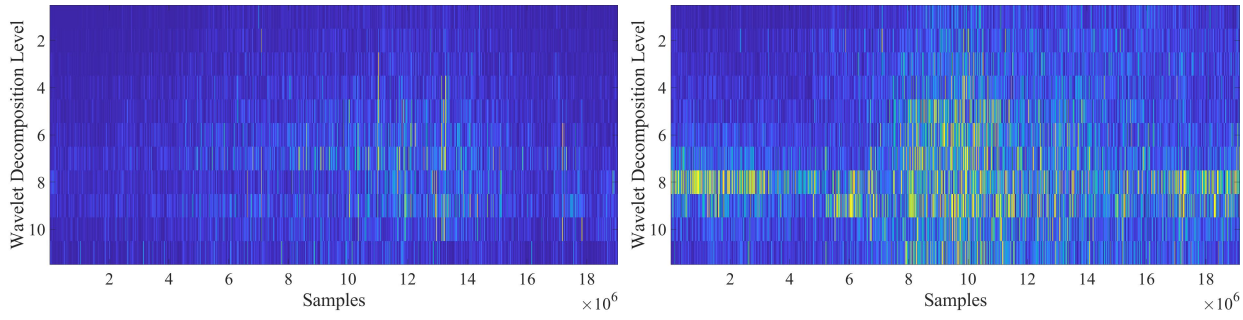
other techniques, HHT gives a better estimation of the fundamental frequencies of the original signal. The shifting process used for empirical mode decomposition [108] removes peaks and smooth uneven amplitudes, leaving only the mode of oscillation contained in the original signal. In comparison with spectrogram, the HHT shows the difference in the harmonic content with a higher resolution, than MFCC, directly representing the harmonics actually included in the analyzed signal.



(a) (b)

Figure 4.29: Second IMF during swarming, considering the swarming of colony 1 (a) and colony 2 (b).

Figures 4.30(a) and 4.30(b) show features maps extracted with discrete wavelet transform. Different harmonics content during swarming is clearly visible. In particular coefficients between level three and level eleven, which are related to medium low frequencies, show cleanly changes in sound during swarming. In comparison with spectrogram, DWT improves the frequency resolution making more distinguishable the swarming. With respect to MFCCs, where only the first three coefficients have remarkable differences, almost all DWT coefficients show differences and they could be used for swarming detection. Moreover, the DWT implementation has a very low computational cost, requiring only a filter bank with a limited number of taps. For example, the taps required for a fourth order Daubechies-based filter banks, as the one used in this work, are only eight for each filter.

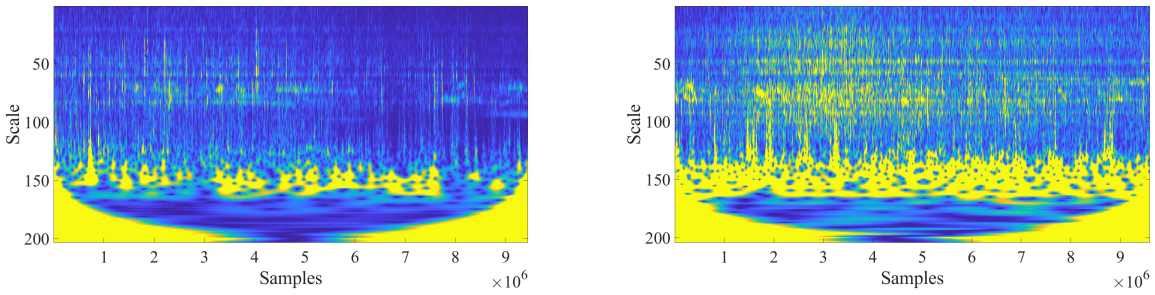


(a) (b)

Figure 4.30: Discrete wavelet coefficients during swarming, considering the swarming of colony 1 (a) and colony 2 (b).

Finally the CWT has been also considered, in this case five minutes recordings, centered on the swarming has been used. This is related to the high computational cost of CWT which makes quite complex the generation of a ten minutes scalograms. Figures 4.31 show scalograms obtained by means of continuous wavelet transform. Despite length limitation scalograms adequately show

well change in the harmonic content during swarming. In particular an increase of the energy of the coefficients with medium and high scales values is clearly visible . It should be noted that the high energy region at the border of highest scales values (i.e., the lower frequencies) must not be considered due to border effect, as also reported in the state of the art [136].



(a) (b)

Figure 4.31: Scalograms during swarming, considering the swarming of colony 1 (a) and colony 2 (b).

Considering all the proposed techniques, the two most promising are based on HHT and wavelet transform. In particular HHT and DWT show better performances with clear signal representation. CWT shows a good signal representation but has a higher computational cost which limit the length of analyzed signals. The best trade-off between performance and computational cost is achieved by discrete wavelet transform: thanks to its filterbank structure it can be be implemented with very low computational cost.

Classifier results For the swarming analysis, no classifier algorithm have been applied, this is due to the fact that only two swarming have been recorded, and since it is a very time limited phenomena, the collected data are not enough for the training procedure. However, since this phenomena has a very specific time frequency behaviour, a swarming detection algorithm could be developed using directly the features coefficients (e.g., the DWT coefficients) to detect the event without the necessity of other classifiers. Future work will be analyzed this aspect.

4.4 Conclusions

In this chapter the analysis of honey bees sound as a real audio system has been discussed. A deep analysis of the state of the art for this type of problem has been discussed showing that a sound based honey bees health monitoring is possible and requires three important steps: an acquisition system, an efficient algorithm for features extractions, and a classifier able to detect different features structures. For these reasons the first step has been the development of a reliable acquisition system, able to detect several different parameters with particular attention to the sound. Three hives have been equipped with the system and have acquired data continuously for over 2 years. Then, from the recorded data two different situations have been analyzed, i.e., the orphaned colony situation and the swarming phenomena. For these situations many experiments have been carried out in order to find the best features extraction algorithm, well known approaches such as spectrograms, MFCC and Mel spectrogram have been compared with innovative approaches based on Hilbert Huang transform and wavelet transform. For the orphaned colony situation, several experiments have been carried out using different features extractions and different classifiers. The early experiments have adopted well known approaches with like spectrograms and MFCC with CNN using HHT only with SVM showing good performances but only when used in combination with MFCC. Starting from these results a deeper study on features extractions have been carried on focusing CNN as classifiers. The most recent results have shown that innovative approaches like the ones based on wavelet transform have the best performance especially in the hive independent situations. Results have also shown that microphone placement is an important aspect which needs to be further investigated in future works. Focusing on swarming, two events have been analyzed, the lack of data has avoided the application of a classifier algorithm to this situation, however, some interesting results have been achieved. A preliminary analysis has shown that in the moment of swarming, there is a weight drop synchronized with a slightly increase in the amplitude of the recorded sound. This simple amplitude detection could be used but suffer of many false positive warnings, for this reason features extraction algorithms have been applied, and results have shown that HHT and DWT could detect well the swarming. Future works will analyze the possibility of using these features to detect swarming, even without a classifiers using only some statistical indicator. Finally future work will deeper analyze another important aspect, i.e., the development of a complete and reliable honey bee sound analysis system, able to acquired the sound, generate the features and make a classification of the recorded sound.

Chapter 5

Other contributions

In the next chapter some minor contribution will be exposed. In particular three works are reported: the first shows a real time application of an active noise control system, the second proposes a real time algorithm for sound enhancement of compact loudspeaker and the last one shows the development of a stereophonic guitar effect. Each Section will be presented with its own conclusion and experimental results.

5.1 Active noise control for snoring reduction

Active noise control (ANC) is a widely adopted technique for noise reduction, based on the generation of an artificial anti-noise signal for the cancellation of the primary noise. In the recent years, several studies have been focused on the application of ANC systems for noise reduction of snoring activity. The well-known problem of snoring activity produces serious consequences from a clinical and social point of view. The volume produced by such noise can reach 90dB and it affects not only the snoring person but also the people that are around him. Some common problems include loss of productivity, reduction of attention and unsafety drive [60, 61]. Moreover, some studies has shown that snoring could be a symptom of illnesses related to sleep-disordered breathing (SDB), such as sleep apnoea-hypopnea [169]. The snoring noise is generated by the soft palate vibration. Recent study has shown a link between talking and snoring activity, both being generated in the vocal tract. Hence, similarly to vocal signal, the acoustical analysis of snoring activity shows a fundamental frequency followed by some high order harmonics [169]. The snoring signal is composed by two fundamental parts: inspiration and expiration. In the frequency domain, the signal spectrum is characterized by some harmonics. Most part of the power is contained in lower part of the spectrum; in fact, the fundamental frequency that has to be removed is generally positioned between 100Hz and 300Hz. In particular, considering a whole snoring period, the majority of the power is contained in the range between 100Hz and 200Hz during the inspiration interval, and between 200Hz and 300Hz during expiration interval. In this respect the development of ANC systems results an interesting challenge for this frequencies range. Regarding the state of the art, many solutions have been presented for reducing snore. Considering passive systems, earplugs or specific pillow, equipped with electrical valve and pumps to modify the pillow position near the neck have been developed [211]. However, these approaches could be very annoying for the user, e.g., sleeping with the earplugs could be very uncomfortable. Moreover, considering the low fre-

quency spectrum of snoring, such passive systems result of little or no effect. An approach based on active noise control system is presented in [125, 216]. It is based on a feedforward structure using the filtered-X least mean square (FX-LMS) algorithm, for the adaptation of finite impulse response (FIR). In [125] both single and multiple channel structure are presented. In the first case, the system is composed by a reference microphone capturing the original noise, two loudspeakers for the anti-noise signal generation, and two error microphones capturing the same residual noise for the adaptive algorithm. In multiple channel configurations, the system is composed by a single reference microphone, two loudspeakers and two error microphones capturing residual noise in two different positions to achieve two different quiet zones. In [60] six different structures of ANC systems using three different FX-LMS algorithms are presented. In this scenario, the next chapter shows a real-time implementation of an active noise control for snoring reduction developed through a DSP platform and an innovative headboard equipped with two loudspeakers and four microphones. The applied ANC algorithm is an evolution of the approach proposed in [165] and it consists of feedforward FX-LMS ANC. Several experiments have been performed considering a real setup installed in a semi-anechoic chamber. The obtained results show the effectiveness of the proposed approach in terms of computational complexity and noise reduction. The presented work has been published in [56].

5.1.1 Adaptive Feedforward Active Noise Control for Snooring Cancellation

Noise reduction through active techniques is based on the introduction of a signal generated by a secondary source that produces a destructive interference in a desired area controlled by one or more microphones [78, 126, 127, 149]. In case of adaptive control, the signal reproduced by the secondary source is adaptively modified depending on the noise recorded at the error and reference microphones.

ANC systems can be classified into two categories, i.e., feedforward and feedback systems. The former are characterized by an adaptation processing based on the error signal recorded with a microphone positioned in the area to be quiet and a reference signal obtained by a microphone that picks up the noise near to the primary source. The latter are characterized by an adaptation processing only based on the error signal since the primary noise source is not available.

Focusing on FX-LMS feedforward approach [11], the mono-channel scheme is represented in Fig. 5.1(a). It is composed of a reference microphone that picks up the noise source $x(n)$, an error microphone that picks up the noise in the area to be quiet $e(n)$ and a loudspeaker that generates the interference signal $y(n)$. This signal is generated by $x(n)$ filtered with the adaptive filter $W(z)$, whose coefficients are produced by the LMS block on the basis of $e(n)$ and $x(n)$ filtered with the estimation of the path between the loudspeaker and the error microphone (secondary path). It is worth noting that the feedback control can be obtained from the feedforward version by adding a primary noise estimation to the error signal $e(n)$.

Active noise cancellation based on FX-LMS algorithm can require high orders adaptive filters involving a high computational cost implementation. Furthermore, a high order filter can be necessary for the estimation of secondary path, depending on the room characteristics. Thus, computational complexity is a considerable issue these systems deals with. Therefore, a block implementation of FX-LMS should be considered for performance improving. In the following, a block feedforward ANC for snooring cancellation is presented. The algorithm dissertation for feedforward scheme is

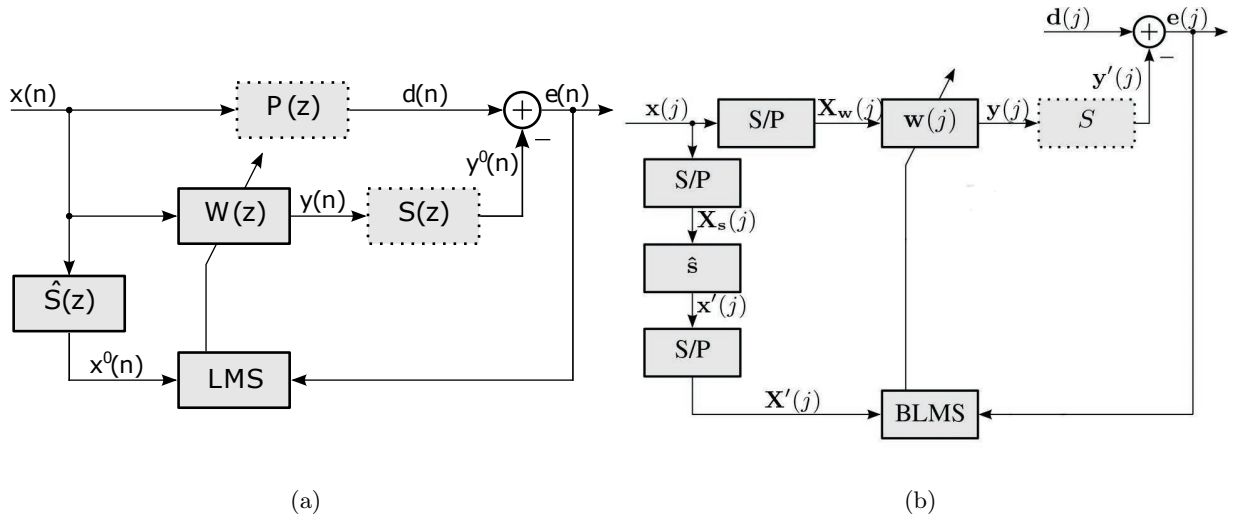


Figure 5.1: (a) Overall scheme of adaptive feedforward FXLMS-ANC systems and (b) its block implementation.

derived from the feedback scheme presented in [165].

5.1.2 Algorithm description

Fig. 5.1(b) synthesizes the scheme of the proposed block FX-LMS algorithm, relative to adaptive feedforward ANC, obtained by extending the scheme of Fig. 5.1(a)[1]. In this block diagram, the current sample n is replaced with the current block j of F_s samples and serial to parallel converters have been added. In a practical situation, the converters and convolutions with FIR filters are implemented in an optimized way by using the overlap and save (OLS) technique [62, 158].

Considering Fig. 5.1(b) and a block size of F_s samples, the j-th block of the noise signal $\mathbf{d}(j)$ is given by

$$\mathbf{d}(j) = \left[\begin{array}{cccc} d(jF_s) & d(jF_s-1) & \dots & d(jF_s-F_s+1) \end{array} \right]^T. \tag{5.1}$$

The input of the ANC system is the error signal, recorded by a microphone positioned in the area to be quiet, i.e.,

$$\mathbf{e}(j) = \left[\begin{array}{cccc} e(jF_s) & e(jF_s-1) & \dots & e(jF_s-F_s+1) \end{array} \right]^T, \tag{5.2}$$

where $e(n)$ is given by

$$e(n) = d(n) - y'(n), \tag{5.3}$$

and $y'(n)$ is the output of the algorithm at error microphone position. Considering an adaptive filter of N samples, the matrix $\mathbf{X_w}(j)$ can be created as

$$\mathbf{X_w}(j) = \left[\begin{array}{cccc} \mathbf{x}(j) & \mathbf{x}(j-1) & \dots & \mathbf{x}(j-N+1) \end{array} \right] \tag{5.4}$$

and it is composed of F_s rows and N columns. Therefore, taking into account the coefficients vector of the adaptive filter

$$\mathbf{w}(j) = \left[\begin{array}{cccc} w_1(j) & w_2(j) & \dots & w_N(j) \end{array} \right]^T, \tag{5.5}$$

[1]Samples are denoted with lowercase italic letters (e.g., a) vector of samples are denoted with lowercase bold letters (e.g., $\mathbf{a}$) and matrix of samples are denoted with uppercase bold letters (e.g., $\mathbf{A}$).

the output $\mathbf{y}(j)$ of the system (i.e. a vector of F_s samples) is given by the following matrix multiplication

$$\mathbf{y}(j) = \mathbf{X_w}(j)\mathbf{w}(j). \tag{5.6}$$

The result of (5.6) is the convolution between $x(n)$ and $w(n)$. The updating of the adaptive filter coefficients is done following the block LMS approach [62]. However, in this approach, a filtered version of x (i.e., x'), is considered as follows [78]:

$$\mathbf{x'}(j) = \mathbf{X_s}(j)\hat{\mathbf{s}}, \tag{5.7}$$

where

$$\mathbf{X_s}(j) = \left[\begin{array}{cccc} \mathbf{x}(j) & \mathbf{x}(j-1) & \dots & \mathbf{x}(j-M+1) \end{array} \right], \tag{5.8}$$

$$\hat{\mathbf{s}} = \left[\begin{array}{cccc} \hat{s}_1 & \hat{s}_2 & \dots & \hat{s}_M \end{array} \right]^T, \tag{5.9}$$

and the matrix $\mathbf{X'}$ is created as $\mathbf{X_w}$

$$\mathbf{X'}(j) = \left[\begin{array}{cccc} \mathbf{x'}(j) & \mathbf{x'}(j-1) & \dots & \mathbf{x'}(j-N+1) \end{array} \right]. \tag{5.10}$$

The correlation matrix is obtained with a convolution between the filtered-x block $\mathbf{x'}(j)$ and the error vector $\mathbf{e}(j)$

$$\mathbf{c}(j) = \mathbf{X'}^T(j)\mathbf{e}(j), \tag{5.11}$$

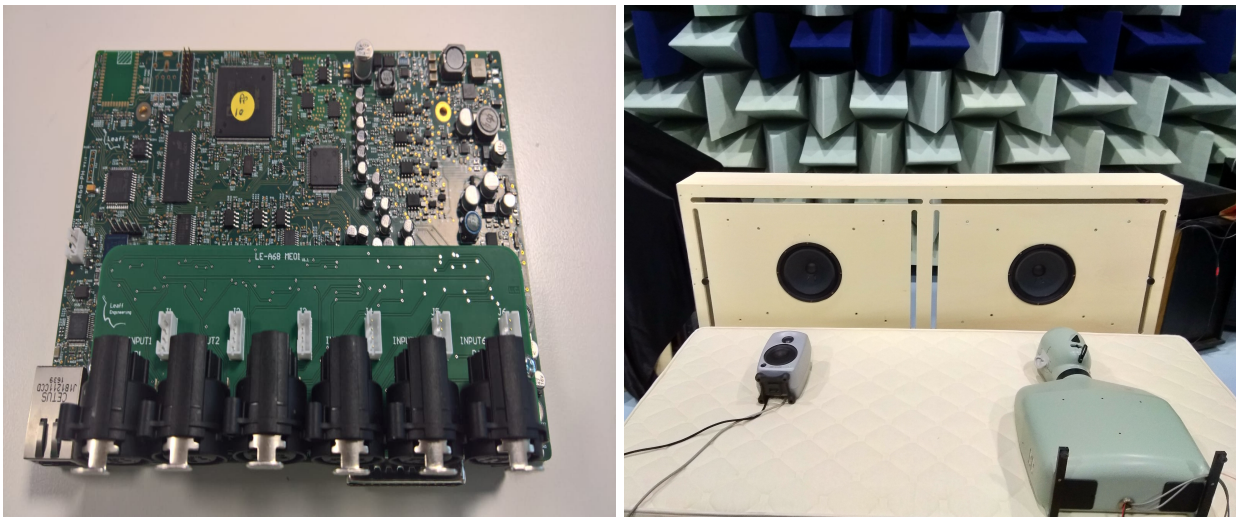
whose dimensions are $N \times 1$. Therefore, $\mathbf{c}(j)$ is used in the updating of the adaptive filter coefficients

$$\mathbf{w}(j+1) = \mathbf{w}(j) + 2\mu\mathbf{c}(j). \tag{5.12}$$

DSP implementation

The implementation of the ANC, described in Section 5.1.2, has been accomplished through a DSP platform of Figure 5.2(a). This system is composed by two modules. The main board is realized with a ADSP-21489 SHARC DSP, 256 Mbit SDRAM PC133 and 16 Mbit Serial FLASH. It allows high performance with a 450 MHz DSP core speed and audio clock frequency up to 96kHz. The second board is used as microphone interface: it allows to have up to 6 balanced microphone inputs with phantom, with $0dB/+20dB$ gain switch for each channel. The minimization of the hardware delay represents an important aspect in ANC implementations in order to maximize the cancellation frequency band: the more the noise frequency band to be cancelled is wide, the more the delay has a bigger impact on the performance results [76, 127]. Furthermore, taking into account analyzed noise frequency range, downsample and upsample operations can be done before and after the execution of ANC, respectively. In this way, also a relevant decreasing of computational complexity can be obtained. However, it is worth noting the particular importance assumed by the anti-aliasing filter design. In fact, the filter order affects the overall delay, hence it has to be as small as possible. For all the tests, a downsample of factor 8 has been considered. In this case, the Nyquist frequency becomes

3 kHz and by designing the filter with Parks-McClellan algorithm with a very wide transition band, (i.e., 100 - 2900 Hz), a passband ripple of −60 dB, and a stopband ripple of −80 dB, the minimum obtainable order is 40, involving an acceptable delay of 20 samples at 48 kHz (corresponding to 0.4 ms) [158].

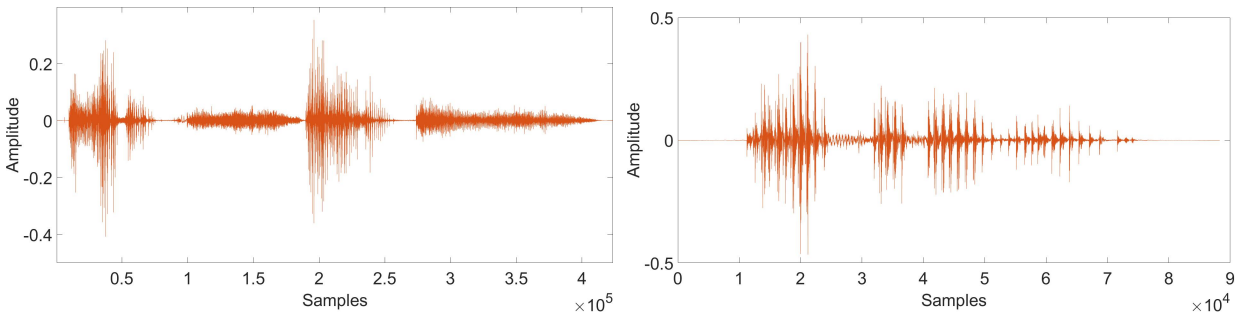


(a) (b)

Figure 5.2: DSP board used for the real time implementation (a), and setup for real experiments in a semi-anechoic chamber(b).

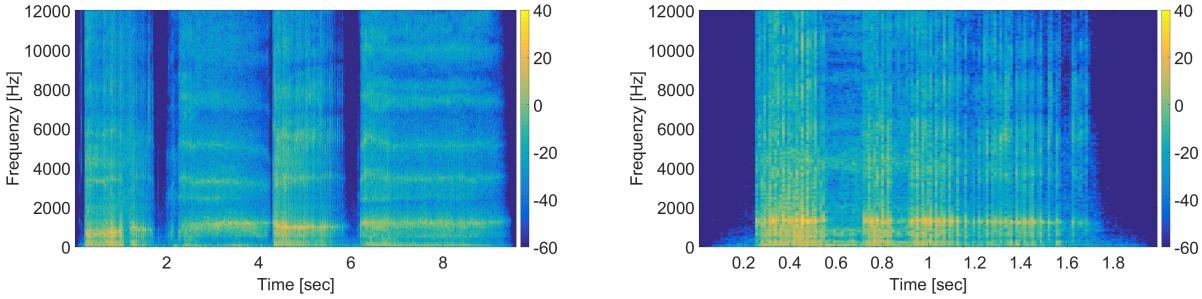
5.1.3 Experimental results

To test the system, several experiments have been performed in a semi-anechoic chamber. In particular, an innovative headboard equipped with two loudspeakers and four microphones has been developed. Figure 5.2(b) shows the real setup, and it is possible to note that headboard is equipped with rails for adjusting microphones in several positions around the loudspeaker positions. The tests have been conducted with recorded snooring signals using a GENELEC loudspeaker to simulate a subject who is potentially snooring. The DSP board has been connected to the microphones positioned on the headboard and to a power amplifier driving the loudspeakers. The loudspeakers are two 8" woofers with a frequency range of 60-5000Hz while the microphones are Audio-Technica ES945 with an omnidirectional polar pattern and a frequency range of 40-18000Hz. For the test, one loudspeaker and two microphones (reference and error) have been considered. The loudspeaker and the error microphone are located in the opposite position with respect to the snooring subject, i.e., to simulate his/her partner, while the reference microphone is near the snooring subject.



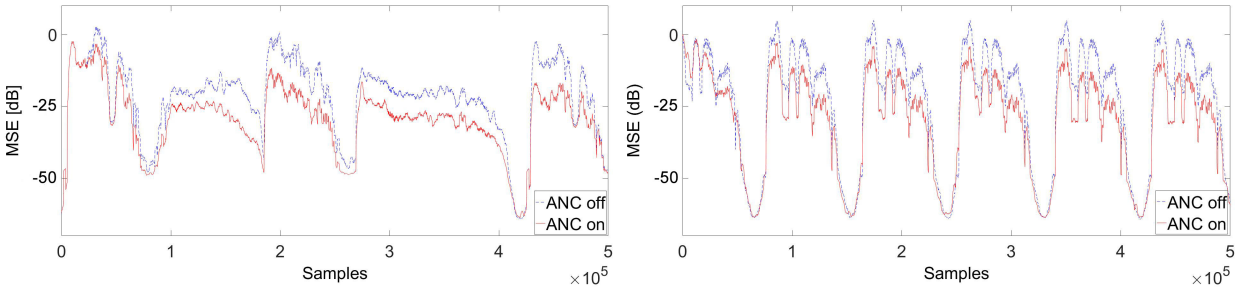
(a) (b)

Figure 5.3: Time domain behaviour of selected snooring signals.



(a) (b)

Figure 5.4: Frequency domain behaviour of the selected snooring signals.



(a) (b)

Figure 5.5: Obtained results of the ANC system in terms of MSE at the error microphone.

Figure 5.5 shows the obtained results considering two snooring signals. The time domain behaviour of the snooring signals is reported in Figures 5.3(a) and 5.3(b) while their frequency domain behaviour is reported in Figures 5.4(a) and 5.4(b). It is clear that the main contribution of the noise is positioned under 2000 Hz and that the signals have a non stationary behaviour. Figures 5.5(a) and 5.5(b) show the obtained results of the snooring cancellation in terms of mean square error (MSE) considering the activation of the ANC algorithm. It is evident that in both cases the system is capable of reducing the snooring signals with a reduction of 8-10 dB. Future works will be oriented on the investigation on the microphones positions exploiting the headboard rails for performance validation.

5.1.4 conclusions

A real-time implementation of an active noise control for snoring reduction developed through a DSP platform has been presented. The applied ANC algorithm consists of feedforward FX-LMS ANC and several experiments have been performed considering a real setup installed in a semi-anechoic chamber, exploiting an innovative headboard equipped with two loudspeakers and four microphones. Some results have been presented considering two recorded snooring signal, showing the effectiveness of the proposed approach in terms of noise reduction. Future works will be oriented on subjective listening tests to validate the performance of the implemented algorithm considering a real scenario.

5.2 Stereo reproduction enhancement

In the next Section will be presented an algorithm developed for the Matlab AES Plugin competition at the 144th AES Convention in New York from the 16 to 19 October 2019. The idea was to develop an algorithm to improve the sound experience with small and compact loudspeakers which can be found in many multimedia devices and which suffer of poor sound quality (e.g., laptops and phones). [58]. The aim of the proposed structure is to improve the sound experience on these devices exploiting and combining three different computational efficient algorithms which work together on different frequency bands in order to improve the sound quality on the whole audio spectrum.

5.2.1 Algorithm Description

The main purpose of the proposed algorithm is to improve the sound experience in the entire audible frequency range. Based on the working frequencies, the whole processing can be split into three algorithms: a virtual bass enhancer for low frequencies, a stereo enhancer for mid frequencies and a decorrelator for high frequencies. Fig. 5.6 shows the block diagram of the whole process.

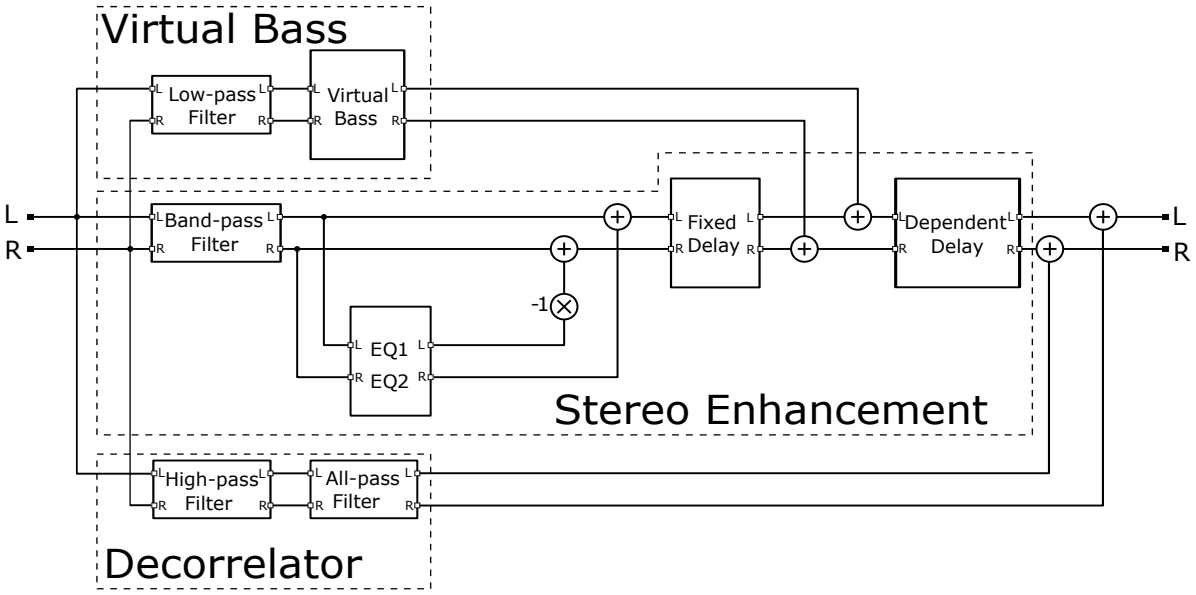


Figure 5.6: Diagram of the proposed structure

Virtual Bass Enhancer

Low frequency range [20 Hz - 260 Hz] is processed by the virtual bass enhancer, whose structure is visible in Fig. 5.7.

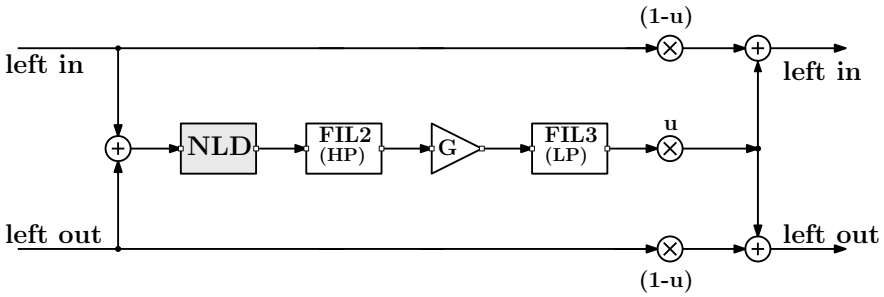


Figure 5.7: Virtual Bass Enhancer structure

The algorithm is derived from the one described in [16] with an approach which is computational efficient. First left and right channels are merged, then the non linear distortion (NLD) is applied to the signal. The resulting signal is high-pass filtered by FIL2 to remove the distorted fundamental. Then, a gain G is applied to the signal before being low-pass filtered by FIL3 to remove the undesired high frequency components. The resulting processed signal is added on both left and right channels, where the u parameter enables wet/dry tuning for this effect. Focusing on NLD, the block schematic of the algorithm is shown in Fig. 5.8.

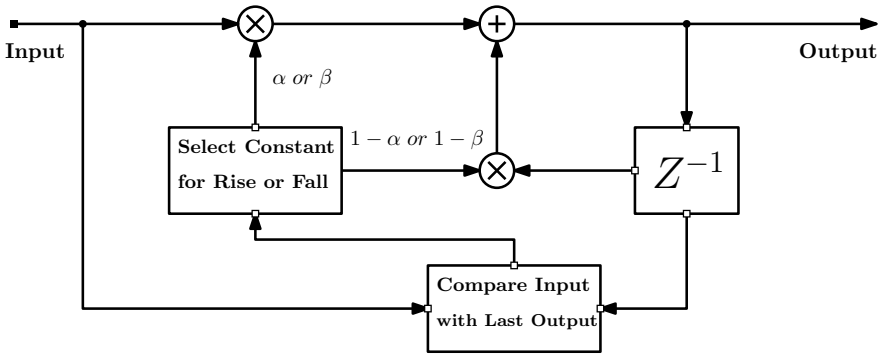


Figure 5.8: NLD structure - from method 1 described in [16]

In order to guarantee that the correct fundamental is reconstructed by human ears, both even and odd harmonics must be generated by the NLD, otherwise the perceived frequency would be different from the correct fundamental, e.g., with only even harmonics the perceived frequency would be double compared to the fundamental [16]. To achieve this, a modified envelope detector scheme has been used. When a signal is passed through the algorithm, the zero crossings with positive slope are preserved. On the other side the points at which the signal goes from positive to negative (negative slope) can be shifted by varying the fall time of the circuit. The algorithm is realized according to Fig. 5.8 with a comparator. The comparison between current input sample and the last output sample sets which constant must be multiplied with the input. In particular, it will select a fast time constant when the input is greater than last output, otherwise it will select a slow time constant. According to schematic of Fig. 5.8, α is the rise constant and β is the fall

constant, both this parameters can be configured by the user.

Stereo Enhancer

Mid frequency range [260 Hz - 1260 Hz] is processed by the stereo enhancer. The algorithm purpose is to extend the apparent source width (ASW) given by compact stereo loudspeaker systems on such frequency range and the approach is derived from the one described in [82]. The directivity of a one-way loudspeaker increases with frequency. Let f_{eff} be the frequency at which a compact stereo loudspeaker system with side-firing loudspeakers reaches optimal directivity for a desired spatial effect. In this case it is desirable to limit directivity above that frequency [82]. To achieve such effect, a mix of direct and dipole reproduction can be used to drive the transducers in a stereo loudspeaker system. Algorithm processing on left and right channels, respectively, is described as follows:

$$s_l(t) = u \cdot l(t) + (1-u) \cdot e(l(t-\tau) + r(t-\tau)) \tag{5.13}$$

$$s_r(t) = u \cdot r(t) + (1-u) \cdot e(l(t-\tau) - r(t-\tau)) \tag{5.14}$$

where s_l and s_r are the output signal, l and r the input signal, $u = \frac{3}{4}$ defines the direct-dipole reproduction mix, $\tau = 6\,\text{ms}$ is a delay for changing the dipole's phase, and $e(\ )$ is the EQ function of a peaking filter with $f_{eff} = 300$ Hz center frequency, a quality factor $Q = 1.2$, a gain of -7.1 dB and peak gain of 5.8 dB. The direct-dipole mixing is controlled by the u parameter: in particular, u is the amplitude of direct reproduction and $(1-u)$ is the amplitude of dipole reproduction. The phase change, obtained with τ delay, is needed to prevent cancellation of the right signal with the dipole. The algorithm block diagram is shown in Fig. 5.6. Both left and right channels are equalized by the peaking filter; then they are swapped and applied back to the signal: right equalizer output EQ2 is added to the left channel, and left equalizer output EQ1 is subtracted from the right channel. Such approach, in which the middle band reproduces the sound partially directly and partially as dipole, makes the directivity pattern of stereo loudspeakers uniform at mid frequency range.

Decorrelator

High frequency range [1260 Hz - 20000 Hz] is processed by the decorrelator [119]. To enlarge the ASW, the algorithm reduces the correlation between left and right channels. This result is achieved using a cascade of randomly generated second order IIR all-pass filters. The all-pass filters modify the signal phase without affecting its amplitude [158]. All-pass biquad IIR filters with real output have the following transfer function:

$$H(z) = \frac{z^{-2} - 2\Re(\rho)z^{-1} + \rho^2}{1 - 2\Re\rho z^{-1} + \rho^2 z^{-2}}, \tag{5.15}$$

where $\rho = \alpha e^{2\pi\omega j}$, α controls the distance of the poles from the unit circle and ω is the angle. These all-pass filters exhibit conjugate symmetry, which causes their output to be real. This is important for processing real signals as it maintains the magnitude at all real frequencies. While the pole angle controls the frequency at which the largest phase change occurs, the distance between the pole and the unit circle controls the amount of the phase distortion. The approach described in [119] locates

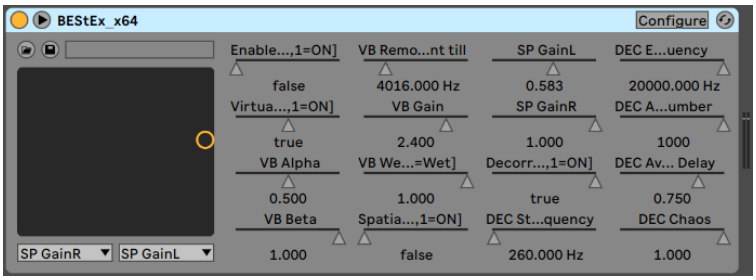


Figure 5.9: VST plugin hosted in a DAW (Ableton)

the poles for all-pass filters randomly along the entire audio frequency range, thus nothing prevents multiple decorrelation filters from having poles placed in the same locations. Considering such issue, the algorithm proposed here imposes angle constraints on every pole depending on the number of all-pass iterations used. In the frequency domain, this corresponds to splitting the working bandwidth into a number of sub-bands equal to the number of all-pass iterations, and every all-pass can have a pole in only one sub-band. The i-th all-pass iteration will thus have its target frequency described by the following equation:

$$f_i = f_{start} + (i + x_i) \cdot \frac{f_{stop} - f_{start}}{N + 1} \tag{5.16}$$

where f_{stop} and f_{start} define the working bandwidth, N is the number of iterations, and x_i is a random value with uniform probability distribution in the interval $(-0.5, 0.5)$. This approach prevents multiple poles from sharing the same location.

5.2.2 Algorithm Implementation

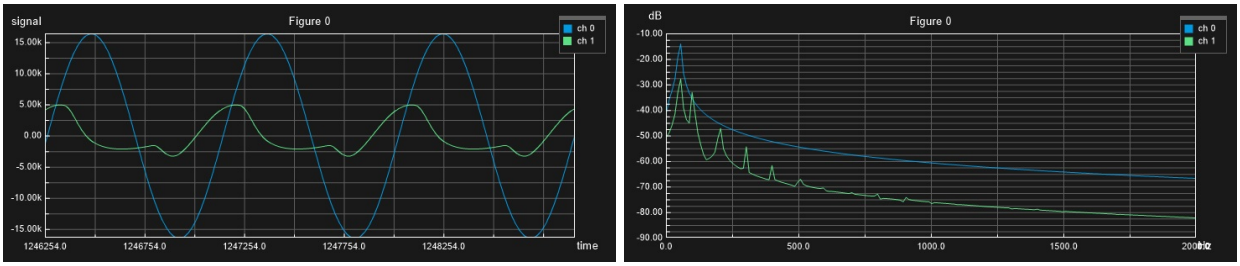
The proposed approach has been implemented developing a VST plugin via Matlab software [9]. This Section focus is to underline notable features of such implementation. The input signal is split by a three band crossover with cut-off frequencies of 260Hz and 1260Hz, resulting in three frequency ranges each processed by its own algorithm, as described in Section 5.2.1. The crossover is composed by four 16th order Linkwitz-Riley filters. It is essential to have a crossover with steep slopes, otherwise the algorithm performance will be reduced dramatically by harmonic contributions from undesired frequency ranges.

The decorrelator all-pass filters coefficients are randomly generated based on two different statistical distributions. In particular, random values for pole angles - determining the frequency at which the phase distortion is applied - follow a uniformly distributed statistic, while the random values for pole radius - determining the steepness of the phase distortion - follow a gaussian distribution. The produced VST audio plugin user interface (UI) gives control over the chaos level and mean value of pole radii, while the pole angles depend on the number of all-pass iterations set. The VST plugin UI, hosted in Ableton, is shown in Fig. 5.9.

Two types of delay blocks are needed to sync the audio stream on the three processing branches. A fixed delay is applied on the equalizer output to sync it with the virtual bass output. The result is once again delayed, depending on the number of all-pass iterations set for the decorrelator, to sync it with the high frequency processing branch output.

Table 5.1: Algorithm parameters used for objective experiments.

Algorithm	Parameters	Values
Virtual Bass Enhancer	α	0.8
	β	0.2
	Gain	0.6
	$\text{FIL2}_{\text{cut-off}}$	50 Hz
	Wet/Dry	1
Stereo Enhancer	SPGainL	0.5
	SPGainR	0.5
Decorrelator	Start frequency	1240
	Stop frequency	20000 Hz
	Delay	115200 Hz
	Random seed	1



(a) (b)

Figure 5.10: Virtual bass results (a) in time and (b) frequency domain.

5.2.3 Algorithm testing

Before starting with the subjective experiments, each part of the whole algorithm has been tested separately with specific experiments, NU-Tech software [2] has been used for running the developed VST.

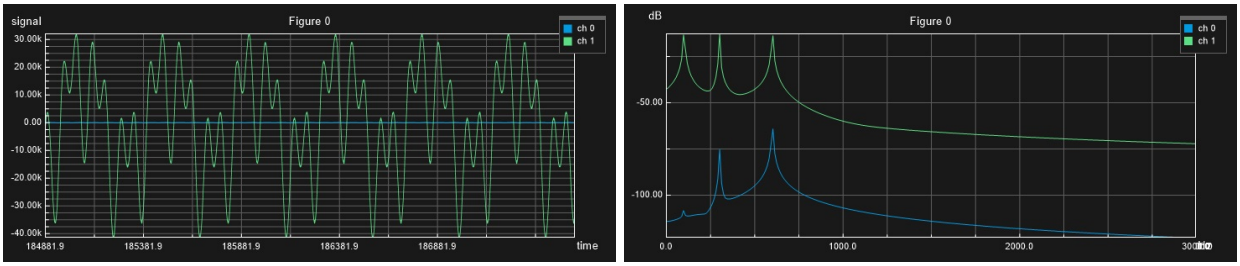
The following experiments have been carried out:

1. for the virtual bass enhancer, the higher harmonics generation is analyzed;
2. for the stereo enhancer, the obtained results on each channel is considered;
3. for the decorrelator, the decorrelation operation between the left and right channel is evaluated.

Parameters used for the experiments are reported in Table 5.1.

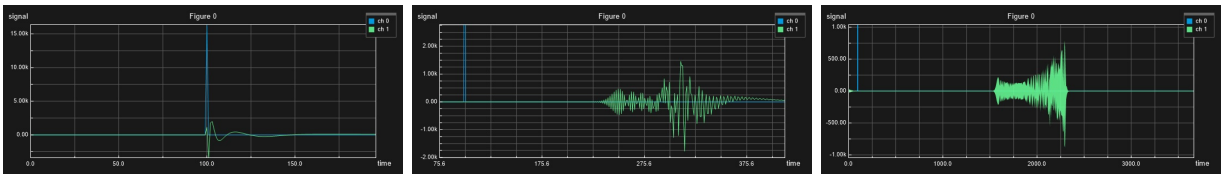
Focusing on first experiment, Fig. 5.10 shows the results in both time and frequency domain of the virtual bass enhancer. In particular, green lines represent the input signal, i.e., a 50 Hz sine wave, while blue lines are the output signal. In the frequency domain representation, it is clearly visible the reduction of the fundamental harmonics due to high-pass filter and the presence of higher order harmonics to create the effect of the missing fundamental.

Fig. 5.11 shows the results for the stereo enhancer. In particular in Fig. 5.11(a), green lines represent the left input signal generated as a superposition of three different sine waves at 100 Hz,



(a) (b)

Figure 5.11: Stereo enhancer results (a) in time and (b) frequency domain.



(a) (b) (c)

Figure 5.12: Decorrelator results in time domain with (a) 0 iteration, (b) 100 iterations and (c) 1000 iterations.

300 Hz and 500 Hz while blue lines represent the right channel that is fed with silence. Fig. 5.11(b) reports the frequency domain after the procedure. According to the stereo enhancer procedure, the right channel is modified reporting some components of the left channel.

Finally, the last algorithm is the decorrelator, and results are visible in Fig. 5.12. Blue lines represent the input, i.e., a single pulse. Green lines represent the same pulse after the algorithm processing where each Figures show a different numbers of iterations. Fig. 5.12(a) shows the output for zero iteration, Fig. 5.12(b) shows the result for 100 iterations and finally Fig. 5.12(c) shows the output for 1000 iterations.

5.2.4 Subjective results

To assess the overall audio quality perception of the entire procedure, subjective listening tests were performed according to the ITU-R BS.1284-1 [5] which provides guidelines for the general assessment of sound quality. Following the recommendations, the subjects involved in the listening tests were 10 expert listeners (5 males and 5 females, ages from 21 to 35) with a technical background in acoustics. The subjective tests have been performed using a pairs of compact Trust loudspeakers connected to a laptop computer running Nu-Tech software for sound reproduction. Before the listening test, a training test was presented to the listeners to get familiar with the test procedure, the test materials, and the test environment, as suggested in [5]. All the subjects had the possibility to listen to each audio item in all conditions under evaluation. For the proposed test session, a signal improved with the proposed algorithm was compared with the original signal without processing (i.e., the reference signal).

For the listening test, the subjects were instructed to make a paired comparison between the original sound track (i.e., the reference signal) and the processed one, scoring some of the attributes as suggested by [5]. The selected attributes to be evaluated are the following:

Table 5.2: Algorithm parameters used for subjective experiments.

Algorithm	Parameters	Values
Virtual Bass Enhancer	α	0.8
	β	0.2
	Gain	0.6
	$\text{FIL2}_{\text{cut-off}}$	150 Hz
	Wet/Dry	0.7
Stereo Enhancer	SPGainL	0.5
	SPGainR	0.5
Decorrelator	Start frequency	1240 Hz
	Stop frequency	20000 Hz
	Delay	115200
	Random seed	1
	Iterations	200

Table 5.3: List of sound tracks used for the listening tests.

Genre	Author	Sound Track
Pop	M.Jackson	Billie Jean
Jazz	H. Hancock	Edith and the Kingpin
Rock	Queen	Bohemian rhapsody
Soul	Marvin Gaye	I heard it through the grapevine

Main impression : "The integrity of the total sound image and the interaction between the other parameters;"

Spatial impression : "The performance appears to take place in an appropriate spatial environment;"

Stereo impression : "The sound image appears to have the correct and appropriate direction distribution of sound sources;"

Freedom from noise and distortions : "Absence of various disturbing phenomena such as electrical noise, acoustic noise, public noise, bit errors, distortions, etc..".

The most important attribute is the basic audio quality based on the main impression that includes all aspects of the sound quality to be assessed. According to [5], the score was given using the continuous quality scale divided into six equal intervals (with a recommended resolution of 1 decimal place) [38]. Bipolar scale values adopted for the comparison between the processed and unprocessed tracks are reported in Table 4.2(b).

Each stimulus was 20 s long, as suggested in the recommendation, and the listeners had the possibility to repeat the presentation. The same listening level has been used for all processed items. Table 5.2 shows the algorithm parameters used for the tests. In comparison with parameters of Table 5.1, the decorrelator number of iterations has been fixed and several other parameters have been changed to tune up the algorithm with the loudspeakers used for the experiments. The selected signals for the subjective evaluation are reported in Table 5.3. Results are visible on Fig.

5.13, which summarizes the obtained results taking into consideration the four different musical

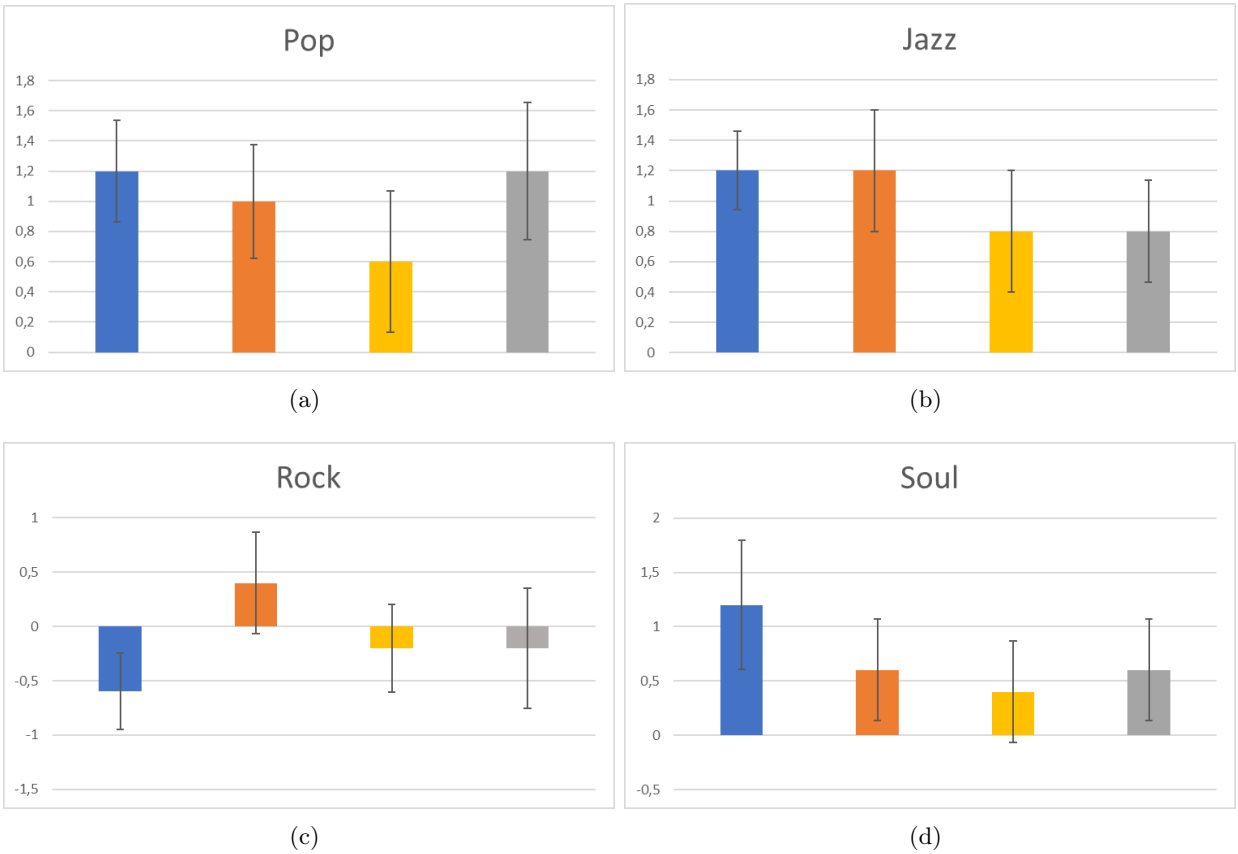


Figure 5.13: Results for subjective listening tests taking into consideration the four attributes and the tracks of Table 5.3. In particular blue bars are related to spatial impression, orange bars are related to stereo impression, yellow bars are related to freedom from noise and distortion and grey bars are related to main impression.

genres shown in Table 5.3. The mean grade values, computed over all involved subjects are plotted with vertical colored blocks, while the confidence intervals calculated with a significance level of 0.05 are given by vertical black lines. The results show an improvement of the performance produced by the proposed approach for pop 5.13(a), jazz 5.13(b) and soul 5.13(b) genres. Focusing on results of Fig. 5.13(c), it is visible that for this type of music the algorithm requires some improvement, this could be due to the different harmonic content of this type of music. This aspect will be deeper analyzed in future works. Best performance on main impression parameter are achieved by pop music, while while jazz reaches the best performance for the other parameters.

5.2.5 Conclusions

In this Section an approach for the improvement of sound experience for small low-cost loudspeakers systems has been presented. Three algorithms have been combined together to improve the sound experience on the whole audible spectrum. Both objective and subjective test have been carried out to validate the proposed algorithm. Results show that for many different music genres the proposed approach seems to work well while for rock music the results show that the approach has to be improved. Future works will be oriented to a deep analysis of this aspect, adding more subjective tests with different music tracks and considering several listeners in order to better understand the

obtained results.

5.3 Applying decorrelation to a guitar vibrato effect

5.3.1 Proposed algorithm

As explained in chapter 2 speaking about an amplifier structure as the one of figure 2.2 some guitar amplifiers have some environment effect built in, the most common are echo, reverb and vibrato, which belongs to the family of delay-based audio effects [218]. These effects are very used and they emulate the behaviour of sound reflections in an acoustic environment. Vibrato effect, in particular, can be defined as a periodically pitch variation of a sound signal due to a periodically variation of a time delay. It is possible to say that the vibrato is a musical effect that allows to obtain a warmer sound and it could be produced by singers voice [13] or by musicians playing instruments, like guitar [111]. The typical implementation of this effect includes a delay line and a low-frequency oscillator to drive the delay time parameter as reported in [63, 69, 218]. For most acoustic instruments the value of the average delay time is typically between 5 and 10 ms, while the rate of the low-frequency oscillator may be between 5 and 14 Hz as described in [194, 218]. In [69], the combination of a vibrato effect with a single sideband modulation and demodulation approach is reported in order to emulate the mechanical vibrato bar of an electric guitar. Vibrato effect is usually employed in a mono signal even if the reproduction of the guitar performance is done considering two or more loudspeakers. In this context, the aim of this paper is the extension of this effect considering a stereophonic reproduction, exploiting decorrelation effect to obtain a spatialization of the reproduced sound. This effect is capable of improving the apparent source width of the audio signal reducing the phase coherence of a given signal while maintaining perceptual transparency [119]. In the literature, several decorrelation techniques can be found. In [118], an all-pass filter technique based on discrete Fourier transform (DFT) is exploited in order to obtain two stereophonic decorrelated signals starting from a monophonic input signal. Although this process allows to create a filter with unit magnitude in the DFT bins, it does not ensure a flat magnitude response. Moreover, this technique could produce audible artifacts because of the periodicity of the DFT transform. In [170], a filterbank combined with a series of delays is used. However, the main drawbacks of this method are the computational complexity and the frequency cancellation at the band edges. In [37], the signal decorrelation is achieved by sinusoidal frequency and amplitude modulation, but this method suffers from latency introduced by the signal analysis. The approach presented in [119] exploits a cascade of IIR all-pass filters in order to not modify the signal amplitude. The main focus of this Section is the generation of a stereophonic effect applied to guitar sounds and based on the use of decorrelation and vibrato effects to evaluate the quality of the final reproduction performance. The decorrelation is based on the technique proposed in [119], with the introduction of constraints on the IIR pole-zero definition as introduced by the authors in [94] and thus obtaining a stereo signal starting from a mono signal. The vibrato effect is achieved by designing a fractional delay and applying a linear interpolation algorithm, as described in [218]. Finally, the processed signal is mixed with the unprocessed one through a dry/wet control. Several experiments will be reported in order to prove the effectiveness of the proposed structure considering objective and subjective analysis. The proposed algorithm aims at reproducing a guitar modulation effect through a stereophonic system. Figure 5.14 shows the complete structure of the

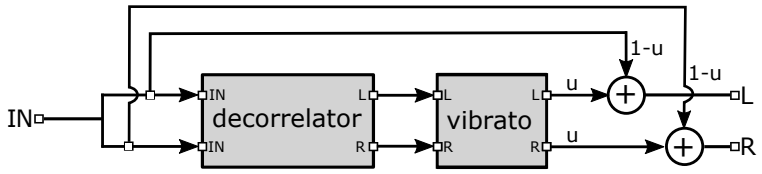


Figure 5.14: Complete algorithm structure. The mono input of the guitar is fed to the decorrelator which transform the mono signal into a stereo signal, then the vibrato effect is applied to both the channel, finally the processed signal is mixed with the dry signal thanks to the dry/wet control **u**.

implemented algorithm. Firstly, the decorrelation transforms the monophonic input signal to a stereophonic signal allowing to widen the sound spatial image. Then, the application of the vibrato effect creates a periodical pitch variation of the decorrelated signals on each channel. Finally, the elaborated signal is mixed with the dry signal using the dry/wet control **u**.

Decorrelator

The first stage of the proposed algorithm consists of the decorrelator [94]. It allows to obtain a stereo signal with a limited correlation between left and right channels. This part of the algorithm is strictly related to the one described in 5.2.1 and follows the same mathematical notations and expressions.

Vibrato

Many audio effects are based on the modulation technique. In particular, vibrato effect is achieved by varying periodically the time delay of the input signal $x(n)$, producing a periodical pitch variation as reported in [218]. The time delay is driven by a low-frequency modulating signal $m(n)$ as shown in Figure 5.15(a), i.e.,

$$m(n) = \alpha \sin(\omega_c n T), \tag{5.17}$$

where α is the modulation width, ω_c is the modulation frequency and T is the period. Defining D as the basic delay in samples, the varying time delay $M(n)$ is calculated as follows:

$$M(n) = \Big\lfloor D + m(n) \Big\rfloor. \tag{5.18}$$

Therefore, the output signal $y(n)$ is obtained though the linear interpolation of the fractional delayed input signal [218], i.e.:

$$y(n) = x(n - [M(n) + 1])d_{\text{frac}}(n) + x(n - M(n))(1 - d_{\text{frac}}(n)), \tag{5.19}$$

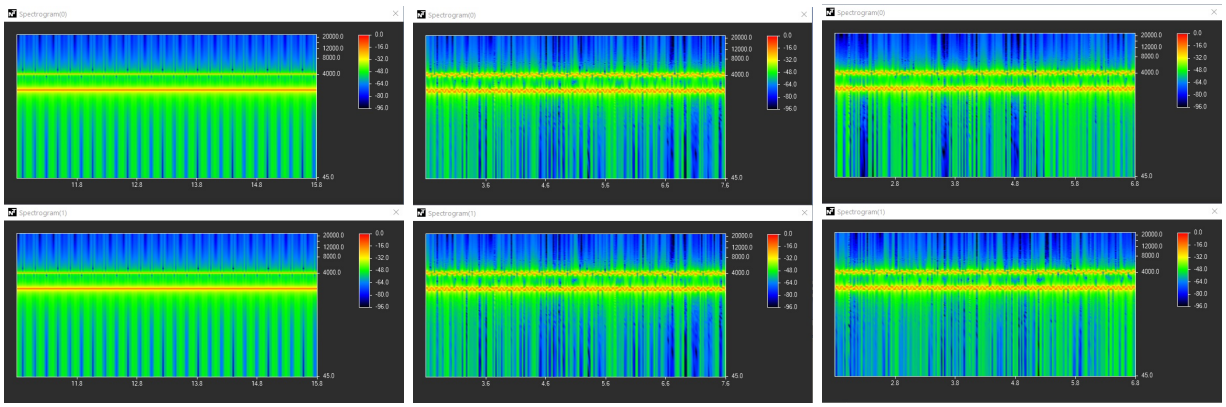
where $d_{\text{frac}}(n)$ is the fractional delay, so a fraction of the normalized sampling interval obtained as follows:

$$d_{\text{frac}}(n) = D + m(n) - M(n). \tag{5.20}$$

It is important to note that $d_{\text{frac}}(n)$ is a non-integer value, so $0 \leq d_{\text{frac}}(n) \leq 1$.

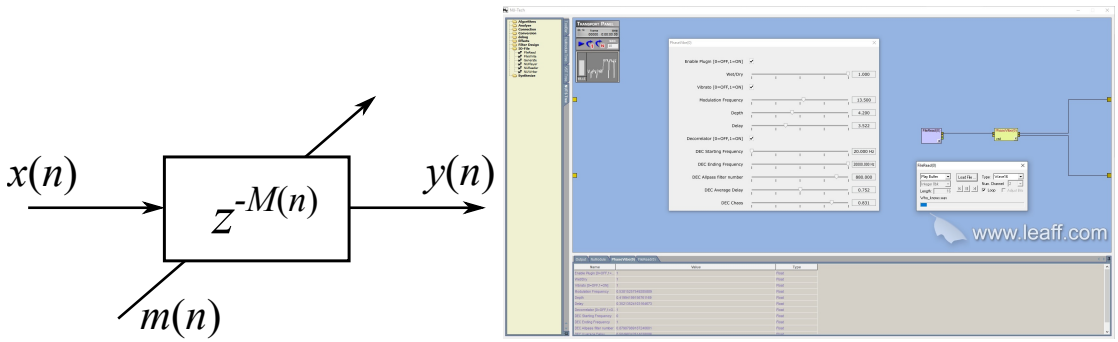
Table 5.4: Proposed algorithm tunable parameters.

Algorithm	Name	Description
Whole	dry/wet	Control the mix of processed and unprocessed signal
Decorrelator	Starting frequency (Hz)	Frequency at which the decorrelator start to put IIR filters
	Stop frequency (Hz)	Frequency at which the decorrelator stop to put IIR filters
	Chaos	A random seed used for IIR generation
	Delay	Delay which replace the unitary delay of IIR filters
	Iterations	Number of IIR filters generated and used
Vibrato	Frequency (Hz)	Frequency of the modulation
	Depth	Modulation width
	Delay	Delay in ms



(a) (b) (c)

Figure 5.16: Spectrograms of the left channel (first figure) and right channel(second figure) considering (a) a superposition of two sine waves at 2kHz and 4kHz as input signal, (b) the results of the vibrato application, and (c) the final results of the whole algorithm.



(a) (b)

Figure 5.15: (a) Main scheme of the vibrato effect, and (b) NU-Tech environment running the proposed algorithm.

5.3.2 Algorithm implementation

The proposed algorithm has been implemented as a VST plug-in using the Matlab software [9] and following the main scheme of Figure 5.14. The final plug-in could be used on every DAW software and it has several parameters which could be adjusted by the users. Figure 5.15(b) shows the use

of the plug-in within the software NU-Tech [2] that is a software for real-time signal processing as described in [129]. The plug-in has several parameters, as reported in Table 5.5, that can be linked to the previous formulation as follows:

- Dry/wet → parameter u of Figure 5.14;
- Starting frequency → the frequency f_{start} of Equation (5.16);
- Stop frequency → the frequency f_{stop} of Equation (5.16);
- Chaos → random seed x_i;
- Decorrelator delay → the delay of IIR filters z^{-n};
- Iterations → number of iterations N;
- Frequency → the modulation frequency in pulsation ω_c of Equation (5.17);
- Depth → the modulation width α of Equation (5.17);
- Vibrato delay → the basic delay D of Equation (5.18).

5.3.3 Experimental results

The evaluation of the proposed system has been performed considering both objective and subjective analysis.

5.3.4 Algorithm testing

Before starting with the subjective experiments, several experiments have been carried out to validate the proposed algorithm. Both the decorrelator and vibrato effect have been tested separately and in the proposed configuration of Figure 5.14. Since the decorrelator is the same of 5.2.1 the same experimental results have been obtained and are present in the previous Section. Regarding the vibrato effect, its results are visible on spectrograms of Figure 5.16(b). For these experiments the algorithm has been fed with a signal which is a superposition of two sine waves at 2kHz and 4kHz. According to the proposed approach both channels have been fed with the same signal whose spectrograms are visible on Figure 5.16(a). Figure 5.16(b) shows the spectrograms considering only the vibrato effect and it is evident that the pitch of the original signal change as a function of the time, according to the vibrato implementation [218] while the left and right channel are still identical since the decorrelation approach is not applied. Finally, Figure 5.16(c) shows the spectrograms of left and right channel with the overall algorithm. The pitch modulation is still clearly visible, moreover the spectral content of the two channels are now different showing that the decorrelator is acting on both channels.

Subjective evaluation

To assess the overall audio quality perception of the proposed approach, subjective listening tests were performed according to the ITU-R BS.1284-1 [5] which provides guidelines for the general assessment of sound quality. According to Table 5.6, two indicators have been used to estimate the

processed signals. The first one is the quality that is related to both the overall sound quality of the processed signal and the evaluation of the effect itself. The latter, i.e., impairment is related to the presence of audible artifacts or undesirable distortion which could be present in the processed signals. For subjective results, four guitar recordings have been prepared. Each stimulus was 20 s long, as suggested in the recommendation, and the listeners had the possibility to repeat the presentation. The same listening level has been used for all processed items. The stimuli used on this work are guitar chord progression inspired from the songs reported in Table 5.7. Clean recordings have been acquired with a Fender Stratocaster connected to a Maudio Firewire 410, using Logic Pro X and Amplitube VST. Then NuTech software [2] has been used to process the recordings with the proposed algorithm setting the algorithm parameters as reported in Table 5.5. The choice of the parameters has been performed differently for each track, trying to keep intact all the original sound characteristics. For each recorded track, two listening stimuli have been prepared, i.e., the first stimulus using the whole algorithm while the second one using only the vibrato effect. Figure 5.15(b) shows the Nu-Tech environment running the VST of the proposed algorithm.

For the subjective tests, the listening panel was composed of 11 listeners, 9 men and 2 women aged between 26 and 49, in which 8 were expert listeners and 3 not. The listeners were asked to evaluate the signals elaborated by:

A. only the vibrato effect;

B. the full proposed algorithm displayed in Figure 5.14.

The subjective tests were carried out using headphones as playback devices. Figure 5.17 shows the results of the listening tests. The resulting scores were elaborated and analyzed by mean with a 95% confidence interval. Figure 5.17 shows the subjective results. In particular, as reported in Figure 5.17(a), the proposed algorithm reports a better sound quality except for the blues genre, in which the application of the only vibrato effect reaches the same score. Finally, Figure 5.17(b) shows the results related to the impairment. Also in this case the scores are higher when the proposed algorithm is applied except for the blues, in which the use of vibrato seems to produce better results. To sum up, the general perception of the vibrato effect appears to be improved in combination with the decorrelator, focusing on the blues track performances the bad result with the impairment parameter could be related to the specific song adopted for the experiments, since it has many moments of silence where the artifacts produced by the decorrelator could be more evident. In any case more subjective results are necessary to deeper investigate this aspect.

Table 5.5: Parameters values used for the subjective listening tests.

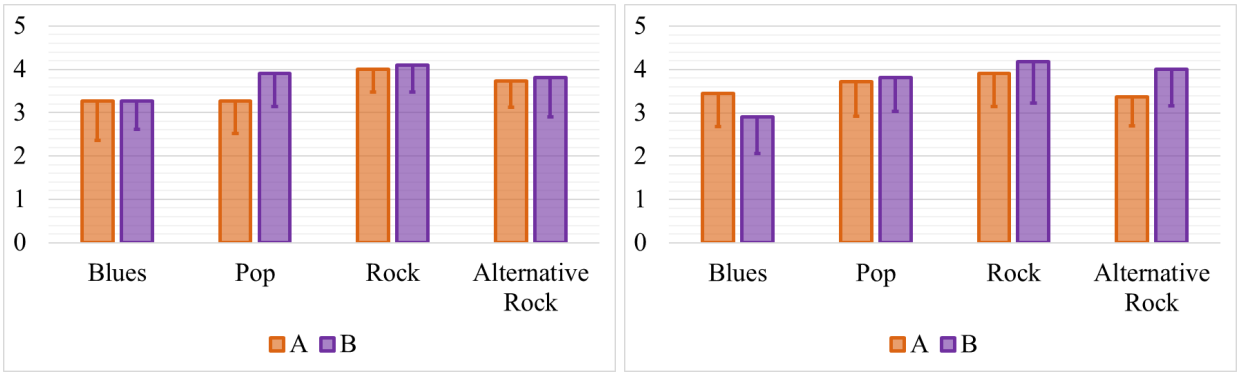
	Music genre			
Parameter	Blues	Pop	Rock	Alternative
Dry/Wet	1	1	0.8	1
Start freq.	20	20	20	20
Stop freq.	20000	20000	20000	20000
Chaos	0.831	0.811	0.811	0.811
Dec. delay	0.752	0.752	0.752	0.752
Iterations	831	880	880	880
Freq.	13.5	6.0	5.0	5.0
Depth	4.2	5.1	4.9	10.0
Vib. delay	3.5	4.0	5.6	7.0

Value	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Table 5.6: Unipolar scales values used for subjective tests.

Genre	Author	Sound Track
Blues	Jimi Hendrix	Who Knows
Pop	Nancy Sinatra	Bang Bang
Rock	Chris Isaak	Wicked Game
Alternative Rock	Radiohead	Planet Telex

Table 5.7: List of sound tracks used for the listening tests.



(a) (b)

Figure 5.17: Results of the listening tests evaluating (a) Sound Quality and (b) Impairment, with (A) the signal elaborated only by the vibrato effect and (B) by the full proposed algorithm.

5.3.5 Conclusions

In this paper, a novel method based on the application of signal decorrelation to a guitar modulation effect has been presented. The decorrelation technique is based on a previous approach that employs a cascade of IIR all-pass filters and introduces angle constraints on each pole. The decorrelator allows to obtain a stereophonic signal starting from a monophonic input. Then, the vibrato effect is applied to both channels in order to create a periodical pitch variation. Finally, a dry/wet control is utilized to mix the elaborated signal with the unprocessed one. The algorithm has been tested through objective and subjective tests. Regarding the objective results, the output signals of the decorrelator have been analyzed in order to verify the effectiveness of the algorithm with known input signal.

For the vibrato effect and for the entire algorithm, spectrograms have been considered in order to evaluate the effect of the procedure in frequency domain starting from a sinusoidal input. Finally, the sound quality and the impairment have been evaluated through subjective tests comparing the signal elaborated only by vibrato and by the proposed algorithm. The proposed algorithm have demonstrated a general improvement of the vibrato effect in both evaluated subjective attributes, confirming a better perception of the reproduced sound. Future works will be oriented to a deep analysis of the experimental results, adding more subjective tests with different music tracks and considering several listeners in order to better understand the obtained results. Objective results could be extended considering measurements such as signal to noise ratio (SNR).

Chapter 6

Conclusions

In this work the application of advanced DSP techniques to real word audio systems has been presented. In particular, three different real world audio systems have been considered.

The first system involves the problem of identification and measure of non linear audio devices with particular attention to vacuum tube audio amplifiers. For this family of sound system, two identification approaches have been discussed and a procedure for the measure of several distortion parameters has been considered. It has been shown how the modern DSP techniques made possible problems overcoming like the computational cost or the memory request, and how it is possible to obtain a more efficient procedure for the distortion parameters measurement. Focusing on the presented results, one of the most important aspects analyzed is the computational cost. It has been shown that complex algorithms for non linear system identification can be improved using innovative digital signal processing techniques. Future works will be carried on. In particular, for the Volterra model, it will be analyzed the implementation of the reduced model in a commercial DSP board. The subband model will be improved considering a more efficient structure as the one presented in [168] instead of the MSAF. Finally the distortion estimation technique will be improved considering the measurement of other distortion parameters such as the third order intermodulation distortion.

For the second system, the equalization problem has been discussed. For this field, it has been shown how the proposed techniques can be used to improve the sound experience with real devices and how it is possible to improve the performances and computational cost of an advanced equalization procedure. The results presented shows that equalization algorithms are crucial for some real word scenarios, moreover it has been shown that innovative techniques can be applied successfully in order to improve the performances of pre existent algorithms. Future work will deeper discuss important aspects such a real time implementation of the presented algorithms and a more accurate subjective test campaign. Finally, innovative techniques, developed for different systems, like the wavelet transform and HHT will be taken into consideration also for the RRE problem.

The last system considered involves the analysis of sound emitted by honey bees in a beehive. Even if it seems a particular system, it has been shown how several techniques developed (like the wavelet transform) can be also applied to this biological sound system, allowing its identification and analysis. The attained results seems very positive, showing that sound could be a concrete solution for a smart and low cost honey bee health monitoring system able to detect and analyze different

situations. For this system, many future works will be considered in the next future. First, as already discussed in Chapter 4, the main focus will be the dataset improvement and the application of the developed techniques to other critical situations like parasite control and the airborne toxic monitoring. Talking about the acquisition system, other sensors will be added (e.g., a microphone for background noise) and other will be improved.

6.1 List of pubblications

1. S. Orcioni, A. Terenzi, S. Cecchi, F. Piazza, and A. Carini. Identification of volterra models of tube audio devices using multiple-variance method. *Journal of the Audio Engineering Society*, 66(10):823–838, 2018

2. S. Cecchi, A. Terenzi, P. Peretti, and F. Bettarelli. Real time implementation of an active noise control for snoring reduction. In *Proc. 144th Audio Engineering Society Convention*. Audio Engineering Society, 2018

3. S. Cecchi, A. Terenzi, S. Orcioni, P. Riolo, S. Ruschioni, and N. Isidoro. A preliminary study of sounds emitted by honey bees in a beehive. In *Proc. 144th Audio Engineering Society Convention*, May 2018

4. S. Orcioni, A. Carini, S. Cecchi, A. Terenzi, and F. Piazza. Identification of nonlinear audio devices exploiting multiple-variance method and perfect sequences. In *Proc. 144th Audio Engineering Society Convention*. Audio Engineering Society, 2018

5. A. Carini, S. Orcioni, A. Terenzi, and S. Cecchi. On Room Impulse Response Measurement Using Perfect Sequences for Wiener Nonlinear Filters. In *Proc. 26th European Signal Processing Conference*, Rome, Italy, 2018

6. A. Carini, S. Orcioni, A. Terenzi, and S. Cecchi. Nonlinear system identification using wiener basis functions and multiple-variance perfect sequences. *Signal Processing*, 160:137–149, 2019

7. P. Burrascano, S. Laureti, M. Ricci, A. Terenzi, S. Cecchi, S. Spinsante, and F. Piazza. A swept-sine pulse compression procedure for an effective measurement of intermodulation distortion. *IEEE Transactions on Instrumentation and Measurement*, 69(4):1708–1719, 2019

8. S. Cecchi, A. Terenzi, F. Piazza, and F. Bettarelli. Applying sound equalization to vibrating sound transducers mounted on rigid panels. In *Proc. 147th Audio Engineering Society Convention*. Audio Engineering Society, 2019

9. S. Cecchi, A. Terenzi, S. Orcioni, and F. Piazza. Analysis of the sound emitted by honey bees in a beehive. In *Proc. 147th Audio Engineering Society Convention*. Audio Engineering Society, 2019

10. I. Nolasco, A. Terenzi, S. Cecchi, S. Orcioni, H. L. Bear, and E. Benetos. Audio-based identification of beehive states. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8256–8260. IEEE, 2019

11. S. Cecchi, A. Terenzi, S. Orcioni, S. Spinsante, V. M. Primiani, F. Moglie, S. Ruschioni, C. Mattei, P. Riolo, and N. Isidoro. Multi-sensor platform for real time measurements of honey bee hive parameters. In *IOP Conference Series: Earth and Environmental Science*, volume 275, page 012016. IOP Publishing, 2019

12. A. Terenzi, S. Cecchi, S. Orcioni, and F. Piazza. Features extraction applied to the analysis of the sounds emitted by honey bees in a beehive. In *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 03–08. IEEE, 2019

13. A. Terenzi, V. Bruschi, S. Cornell, A. Castellani, and S. Cecchi. A multiband structure based on hammerstein model for nonlinear audio system identification. In *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 9–14. IEEE, 2019

14. A. Terenzi, S. Cecchi, S. Spinsante, S. Orcioni, and F. Piazza. Real-time system implementation for bee hives weight measurement. In *2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, pages 231–236. IEEE, 2019

15. S. Cecchi, A. Terenzi, V. Bruschi, A. Carini, and S. Orcioni. A subband implementation of a multichannel and multiple position adaptive room response equalizer. *Applied Acoustics*, 173:107702

16. A. Terenzi, S. Cecchi, and S. Spinsante. On the importance of the sound emitted by honey bee hives. *Veterinary Sciences*, 7(4):168, 2020

17. S. Cecchi, S. Spinsante, A. Terenzi, and S. Orcioni. A smart sensor-based measurement system for advanced bee hive monitoring. *Sensors*, 20(9):2726, 2020

18. A. Carini, S. Orcioni, A. Terenzi, and S. Cecchi. Orthogonal periodic sequences for the identification of functional link polynomial filters. *IEEE Transactions on Signal Processing*, 68:5308–5321, 2020

19. D. A. Giliberti, F. Iseini, N. Pelagalli, A. Terenzi, and S. Cecchi. An advanced audio system for stereo reproduction enhancement. In *Proc. of 148th Audio Engineering Society Convention*, Vienna, Austria, May 2020

20. A. Terenzi, V. Bruschi, S. Nobili, and S. Cecchi. Applying signal decorrelation to a guitar modulation effect. In *Proc. 149th Audio Engineering Society Convention*. Audio Engineering Society, 2020

21. A. Terenzi, S. Spinsante, and S. Cecchi. Review on electric vehicles exterior noise generation and evaluations. In *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive*. IEEE, 2020

Bibliography

[1] Admp401 mems microphone, Analog Devices. https://www.analog.com/media/en/technical-documentation/obsolete-data-sheets/ADMP401.pdf.

[2] NU-Tech, Leaff Engineering. http://www.nu-tech-dsp.com/content.php.

[3] Chapter 8 Optimal Smoothing. In *Stochastic Models, Estimation, and Control: Volume 2*, volume 141 of *Mathematics in Science and Engineering*, pages 1 – 22. Elsevier, 1982.

[4] IEC 60268-5, Sound system equipment –part 5: Loudspeakers. International Electrotechnical Commission, 2003.

[5] ITU-R BS. 1284-1 General methods for the subjective assessment of sound quality, International Telecommunication Union, Geneva, 2003.

[6] National Instruments, PXIe-4461 DSA Dynamic Signal Acquisition Device. http://www.ni.com/pdf/manuals/373770j.pdf, 2008.

[7] National Instruments, Sound and vibration toolkit. https://www.ni.com/trysv/, 2017.

[8] Open source beehives project. https://www.osbeehives.com/pages/about-us, 2019.

[9] Matlab-audio plugin creation and hosting. https://www.mathworks.com/help/audio/audio-plugin-creation-and-hosting.html, Mathworks Inc.

[10] F. T. Agerkvist. Volterra series based distortion effect. In *Proc. 129th Audio Engineering Society Convention*, Nov 2010.

[11] M. T. Akhtar, M. Abe, and M. Kawamata. New Structure for Feedforward Active Noise Control Systems With Improved Online Secondary Path Modeling. *IEEE Trans. Speech Audio Process.*, 13(5):1082–1088, Sep. 2005.

[12] M. Ali. Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-pass Filtering for Signal Decorrelation. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 6, pages 3689–3692, Seattle, WA, USA, May 1998.

[13] N. Amir, T. Erlich, N. Grabstein, and J. Fainguelernt. Automated evaluation of singers' vibrato through time and frequency analysis of the pitch contour using the dsk6713. In *2009 16th International Conference on Digital Signal Processing*, pages 1–5, 2009.

[14] S. Amiriparian, M. Gerczuk, S. Ottl, N. Cummins, M. Freitag, S. Pugachevskiy, A. Baird, and B. W. Schuller. Snore sound classification using image-based deep spectrum features. In *INTERSPEECH*, volume 434, pages 3512–3516, 2017.

[15] Aristoteles. *The Works of Aristotle, the Famous Philosopher; Containing His Complete Master-Piece, Displaying the Secrets of Nature in the Generation of Man... to which are Added, the Family Physician... and the Experienced Midwife... Aloe, His Book of Problems...: A Treatise of Physiognomy. His Last Legacy...* Booksellers (by H. Norman), 1870.

[16] M. Arora, S. Jang, and H. Moon. Low complexity virtual bass enhancement algorithm for portable multimedia device. In *Audio Engineering Society Conference: 29th International Conference: Audio for Mobile and Handheld Devices*. Audio Engineering Society, 2006.

[17] M. M. Banat. Pipelined volterra filter. *Electronics Letters*, 28(13):1276–1278, June 1992.

[18] B. Bank. Combined quasi-anechoic and in-room equalization of loudspeaker responses. In *Proc. 134th Audio Engineering Society Convention*, May 2013.

[19] D. Bard. Horn loudspeakers nonlinearity comparison and linearization using volterra series. In *Proc. 124th Audio Engineering Society Convention*, May 2008.

[20] E. L. O. Batista and R. Seara. A new perspective on the convergence and stability of nlms hammerstein filters. In *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 343–348, Sep. 2013. doi: 10.1109/ISPA.2013.6703764.

[21] M. Bencsik. Beehive monitoring, 2015. Patent, GB2015/051580.

[22] J. Benesty, D. R. Morgan, and M. M. Sondhi. A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation. *IEEE Trans. Speech Audio Process.*, 6(2):156–165, Mar. 1998.

[23] J. Benesty, C. Paleologu, T. Gansler, and S. Ciochina. *A perspective on stereophonic acoustic echo cancellation*, volume 4. Springer Science & Business Media, 2011.

[24] J. Benesty, M. M. Sondhi, and Y. Huang. *Springer Handbook of Speech Processing*. Springer, 2008.

[25] R. Bernardini. A fast algorithm for general volterra filtering. *IEEE Transactions on Communications*, 48(11):1853–1864, Nov 2000.

[26] F. Bettarelli, S. Cecchi, and A. Lattanzi. NU-Tech: The entry tool of the HArtes toolchain for algorithms design. In *Proc. 124th Audio Engineering Society Convention*, Amsterdam, The Netherlands, May 2008.

[27] S. Bharitkar and C. Kyriakakis. *Immersive Audio Signal Processing*. Springer, New York, 2006.

[28] V. Bisot, R. Serizel, and S. Essid. Feature learning with matrix factorization applied to acoustic scene classification. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 25(6):1216–1229, 2017.

[29] J. Borish and J. B. Angell. An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise. *J. Audio Eng. Soc.*, 31(7):478–488, Jul. 1983.

[30] J. J. Bosch, J. Janer, F. Fuhrmann, and P. Herrera. A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals. In *ISMIR*, pages 559–564, 2012.

[31] L.-J. Brannmark, A. Bahne, and A. Ahlen. Compensation of loudspeaker–room responses in a robust mimo control framework. *IEEE transactions on audio, speech, and language processing*, 21(6):1201–1216, 2013.

[32] J. J. Bromenshenk. Honey bee acoustic recording and analysis system for monitoring hive health, 2007. US Patent 7549907.

[33] T. J. Brundage. Acoustic sensor for beehive monitoring, 10 2012. US Patent 8,152,590.

[34] P. Burrascano, S. Laureti, M. Ricci, A. Terenzi, and S. Cecchi. Pulse Compression Procedure for an Effective Measurement of Intermodulation Distortion. In *Proc. 10th Int'l Symposium on Image and Signal Processing and Analysis*, Ljubljana, Slovenia, Sep. 2017.

[35] P. Burrascano, S. Laureti, M. Ricci, A. Terenzi, S. Cecchi, S. Spinsante, and F. Piazza. A swept-sine pulse compression procedure for an effective measurement of intermodulation distortion. *IEEE Transactions on Instrumentation and Measurement*, 69(4):1708–1719, 2019.

[36] P. Burrascano, S. Laureti, L. Senni, and M. Ricci. Pulse compression in nondestructive testing applications: Reduction of near sidelobes exploiting reactance transformation. *IEEE Transactions on Circuits and Systems I: Regular Papers*, Available on line at `10.1109/TCSI.2018.2862868`:11, Aug. 2018.

[37] A. Cabrera and G. Kendall. Multichannel control of spatial extent through sinusoidal partial modulation (spm). In *Proceedings of the Sound and Music Computing Conference 2013*, pages 532–537, 2013.

[38] A. Carini, S. Cecchi, I. Omiciuolo, F. Piazza, and G. L. Sicuranza. Multiple Position Room Response Equalization in Frequency Domain. *IEEE Trans. Speech and Audio Processing*, 2(1):122–135, Jan. 2012.

[39] A. Carini, S. Orcioni, A. Terenzi, and S. Cecchi. On Room Impulse Response Measurement Using Perfect Sequences for Wiener Nonlinear Filters. In *Proc. 26th European Signal Processing Conference*, Rome, Italy, 2018.

[40] A. Carini, S. Orcioni, A. Terenzi, and S. Cecchi. Nonlinear system identification using wiener basis functions and multiple-variance perfect sequences. *Signal Processing*, 160:137–149, 2019.

[41] A. Carini, S. Orcioni, A. Terenzi, and S. Cecchi. Orthogonal periodic sequences for the identification of functional link polynomial filters. *IEEE Transactions on Signal Processing*, 68:5308–5321, 2020.

[42] S. Cecchi, A. Carini, A. Primavera, and F. Piazza. An Adaptive Multiple Position Room Response Equalizer in Warped Domain. In *Proc. EUSIPCO 2012*, pages 1955–1959, Bucharest, Romania, 2012.

[43] S. Cecchi, A. Carini, and S. Spors. Room Response Equalization - A Review. *Applied Sciences (Switzerland)*, 8(1), 2018.

[44] S. Cecchi and et al. Audio samples. `http://www2.units.it/ipl/research/AudEq/AppliedAcoustics20/data.htm`, 2020. [Online].

[45] S. Cecchi, L. Palestini, P. Peretti, L. Romoli, F. Piazza, and A. Carini. Evaluation of a Multipoint Equalization System based on Impulse Responses Prototype Extraction. In *Proc. 127th Audio Engineering Society Convention*, New York, NY, USA, Oct. 2009.

[46] S. Cecchi, A. Primavera, F. Piazza, and A. Carini. An Adaptive Multiple Position Room Response Equalizer. In *Proc. EUSIPCO 2011*, pages 1274–1278, Barcellona, Spain, May 2011.

[47] S. Cecchi, L. Romoli, A. Carini, and F. Piazza. A Multichannel and Multiple Position Adaptive Room Response Equalizer in Warped Domain: Real-time Implementation and Performance Evaluation. *Applied Acoustics*, 82:28–37, Aug. 2014.

[48] S. Cecchi, L. Romoli, M. Gasparini, A. Carini, and F. Bettarelli. An Adaptive Multichannel Identification System for Room Response Equalization. In *Proc. Int. Conf. Electronics, Computers and Artificial Intelligence*, Bucharest, România, Jun. 2015.

[49] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza. Low-complexity Implementation of a Real-time Decorrelation Algorithm for Stereophonic Acoustic Echo Cancellation. *Signal Processing*, 92(11):2668–2675, Nov. 2012.

[50] S. Cecchi, L. Romoli, F. Piazza, and A. Carini. A Multichannel and Multiple Position Adaptive Room Response Equalizer in Warped Domain, Sep. 2013. Proc. 8th Int'l Symposium on Image and Signal Processing and Analysis.

[51] S. Cecchi, S. Spinsante, A. Terenzi, and S. Orcioni. A smart sensor-based measurement system for advanced bee hive monitoring. *Sensors*, 20(9):2726, 2020.

[52] S. Cecchi, A. Terenzi, V. Bruschi, A. Carini, and S. Orcioni. A subband implementation of a multichannel and multiple position adaptive room response equalizer. *Applied Acoustics*, 173:107702.

[53] S. Cecchi, A. Terenzi, S. Orcioni, and F. Piazza. Analysis of the sound emitted by honey bees in a beehive. In *Proc. 147th Audio Engineering Society Convention*. Audio Engineering Society, 2019.

[54] S. Cecchi, A. Terenzi, S. Orcioni, P. Riolo, S. Ruschioni, and N. Isidoro. A preliminary study of sounds emitted by honey bees in a beehive. In *Proc. 144th Audio Engineering Society Convention*, May 2018.

[55] S. Cecchi, A. Terenzi, S. Orcioni, S. Spinsante, V. M. Primiani, F. Moglie, S. Ruschioni, C. Mattei, P. Riolo, and N. Isidoro. Multi-sensor platform for real time measurements of honey bee hive parameters. In *IOP Conference Series: Earth and Environmental Science*, volume 275, page 012016. IOP Publishing, 2019.

[56] S. Cecchi, A. Terenzi, P. Peretti, and F. Bettarelli. Real time implementation of an active noise control for snoring reduction. In *Proc. 144th Audio Engineering Society Convention*. Audio Engineering Society, 2018.

[57] S. Cecchi, A. Terenzi, F. Piazza, and F. Bettarelli. Applying sound equalization to vibrating sound transducers mounted on rigid panels. In *Proc. 147th Audio Engineering Society Convention*. Audio Engineering Society, 2019.

[58] S. Cecchi, M. Virgulti, A. Primavera, F. Piazza, F. Bettarelli, and J. Li. Investigation on audio algorithms architecture for stereo portable devices. *Journal of the Audio Engineering Society*, 64(1/2):75–88, 2016.

[59] T. Cejrowski, J. Szymański, H. Mora, and D. Gil. Detection of the bee queen presence using sound analysis. In *Asian Conference on Intelligent Information and Database Systems*, pages 297–306. Springer, 2018.

[60] S. Chakravarthy and S. Kuo. Application of active noise control for reducing snore. In *IEEE Int. Conf. on Acoustics Speech and Signal Processing*, volume 5, pages V–V, May 2006.

[61] C.-Y. Chang, S.-T. Pan, and K.-C. Liao. Active noise control and its application to snore noise cancellation. *Asian Journal of Control*, 15(6):1648–1654, 2013.

[62] G. A. Clark, S. K. Mitra, and S. R. Parker. Block Implementation of Adaptive Digital Filters. *IEEE Trans. Acoust., Speech, Signal Process.*, 29(3):744–752, Jun. 1981.

[63] A. Clarke. Digital implementation of vibrato. 2012.

[64] J. Darr and W. Oliver. *Electric guitar amplifier handbook*. Bobbs-Merrill Company, 1965.

[65] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.

[66] I. Daubechies. Ten lectures on wavelets. volume 61. Siam, 1992.

[67] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrésani. Asymptotic wavelet and gabor analysis: Extraction of instantaneous frequencies. *Information Theory, IEEE Transactions on*, 38:644 – 664, 04 1992.

[68] D. Dietlein. A method for remote monitoring of activity of honeybee colonies by sound analysis. *Journal of Apicultural Research*, 24(2):176–183, 1985.

[69] S. Disch and U. Zölzer. Modulation and delay line based digital audio effects. In *Proceeding of the 2nd COST-G6 Workshop on Digital Audio Effects (DAFx'99)*, pages 5–8, 1999.

[70] A. Dobrucki and P. Pruchnicki. Application of the NARMAX-method for modelling of the nonlinearity of dynamic loudspeakers. In *Proc. 106th Audio Engineering Society Convention*, May 1999.

[71] A. Dobrucki and P. Pruchnicki. Application of the NARMAX method to the modelling of the nonlinearity of dynamic loudspeakers. *Archives of Acoustics*, 26(4), 2001.

[72] W. R. Dunkel, M. Rest, K. J. Werner, M. J. Olsen, and J. O. Smith. The fender bassman 5f6-a family of preamplifier circuits–a wave digital filter case study. In *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16)*, Brno, CZ, September 5–9 2016.

[73] F. Eichas, E. Gerat, and U. Zölzer. Virtual analog modeling of dynamic range compression systems. In *Proc. 142th Audio Engineering Society Convention*, May 2017.

[74] F. Eichas, S. Möller, and U. Zölzer. Block-oriented modeling of distortion audio effects using iterative minimization. In *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-15)*, Trondheim, Norway, Nov 30 Dec 3 2015.

[75] F. Eichas, S. Möller, and U. Zölzer. Block-oriented gray box modeling of guitar amplifiers. In *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, pages 184–191, Edinburgh, UK, September 5–9 2017.

[76] S. Elliott. Dsp in the active control of sound and vibration. In *Audio Engineering Society Conference: UK 7th Conference: Digital Signal Processing (DSP)*, Sep 1992.

[77] S. J. Elliott and P. A. Nelson. Multiple-Point Equalization in a Room Using Adaptive Digital Filters. *J. Audio Eng. Soc.*, 37(11):899–907, Nov. 1989.

[78] S. J. Elliott and P. A. Nelson. Active Noise Control. *IEEE Signal Process. Mag.*, 10(4):12–35, Oct. 1993.

[79] H. Enzinger, K. Freiberger, G. Kubin, and C. Vogel. Fast time-domain Volterra filtering. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 225–228, Nov 2016.

[80] H. Eren and et al. Electronic sensing and identification of queen bees in honeybee colonies. In *IEEE Instrumentation and Measurement Technology Conference*, May 19-21 1997.

[81] E. Eskov and V. Toboev. Changes in the structure of sounds generated by bee colonies during sociotomy. *Entomological review*, 91(3):347, 2011.

[82] C. Faller. Compact stereo loudspeakers with dipole processing. In *Proc. 146th Audio Engineering Society Convention*. Audio Engineering Society, 2019.

[83] A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. *Journal of the Audio Engineering Society*, (5093), February 2000.

[84] A. Farina, A. Bellini, and E. Armelloni. Non-linear convolution: a new approach for the auralization of distorting systems. In *Proc. 110th Audio Engineering Society Convention*, Amsterdam, NL, May 2001.

[85] J. Faucon, L. Mathieu, M. Ribière, A. Martel, P. Drajnudel, S. Zeggane, and et al. Honey bee winter mortality in france in 1999 and 2000. *Bee World*, 83:14–23, 2002.

[86] S. Ferrari, M. Silva, M. Guarino, and D. Berckmans. Monitoring of swarming sounds in bee hives for prevention of honey loss. In *International Workshop on Smart Sensors in Livestock Monitoring*, 2006.

[87] S. Ferrari, M. Silva, M. Guarino, and D. Berckmans. Monitoring of swarming sounds in beehives for early detection of the swarming period. *Computers and Electronics in Agriculture*, 65:72–77, 2008.

[88] A. J. S. Ferreira and A. Leite. An Improved Adaptive Room Equalization in the Frequency Domain. In *Proc. 118th Audio Engineering Society Convention*, Barcelona, Spain, May 2005.

[89] H. Frings and F. Little. Reactions of honey bees in the hive to simple sounds. *Science*, pages 122–125, 1957.

[90] F. Fuhrmann and P. Herrera. Polyphonic instrument recognition for exploring semantic similarities in music. In *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, 2010.

[91] L. Fuster, M. de Diego, M. Ferrer, A. Gonzalez, and G. Pinero. A biased multichannel adaptive algorithm for room equalization. In *Proc. 20th European Signal Processing Conference*, pages 1344–1348, Bucharest, Romania, Aug. 2012.

[92] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren. Timit acoustic-phonetic continuous speech corpus.

[93] E. Gerat, F. Eichas, and U. Zölzer. Virtual analog modeling of a urei 1176ln dynamic range control system. In *Proc. 143th Audio Engineering Society Convention*, Oct 2017.

[94] D. A. Giliberti, F. Iseini, N. Pelagalli, A. Terenzi, and S. Cecchi. An advanced audio system for stereo reproduction enhancement. In *Proc. of 148th Audio Engineering Society Convention*, Vienna, Austria, May 2020.

[95] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256. PMLR, May 2010.

[96] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[97] D. Goulson, E. Nicholls, C. Botías, and E. L. Rotheray. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science*, 347(6229), 2015.

[98] P. D. Hatziantoniou and J. N. Mourjopoulos. Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses. *J. Audio Eng. Soc.*, 48:259–280, Apr. 2000.

[99] P. D. Hatziantoniou and J. N. Mourjopoulos. Errors in real-time room acoustics dereverberation. *J. Audio Eng. Soc.*, 52(9):883–899, Sep. 2004.

[100] T. Hélie. On the use of Volterra series for real-time simulations of weakly nonlinear analog audio devices: application to the moog ladder filter. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx06)*, pages 7–12, 2006.

[101] R. C. Heyser. Acoustical measurements by time delay spectrometry. *J. Audio. Eng. Soc.*, 15:370–382, Oct. 1967.

[102] C. Hofmann, M. Guenther, C. Huemmer, and W. Kellermann. Efficient nonlinear acoustic echo cancellation by partitioned-block significance-aware hammerstein group models. In *Proc. 24th European Signal Processing Conference*, pages 1783–1787, Budapest, Hungary, Aug 2016.

[103] C. Hofmann, C. Huemmer, M. Guenther, and W. Kellermann. Significance-aware filtering for nonlinear acoustic echo cancellation. *EURASIP Journal on Advances in Signal Processing*, 2016(1):113, Nov 2016.

[104] C. Hofmann, C. Huemmer, and W. Kellermann. Significance-aware hammerstein group models for nonlinear acoustic echo cancellation. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5934–5938, May 2014.

[105] M. Holters, K. Dempwolf, and U. Zölzer. A digital emulation of the boss sd-1 super overdrive pedal based on physical modeling. In *Proc. 131th Audio Engineering Society Convention*, Oct 2011.

[106] D. Howard, O. Duran, G. Hunter, and K. Stebel. Signal processing the acoustics of honeybees (apis mellifera) to identify the" queenless" state in hives. *Proceedings of the Institute of Acoustics*, 35(1):290–297, 2013.

[107] M. Hrncir, F. G. Barth, and J. Tautz. Vibratory and airborne sound-signals in bee communication. In S. Drosopoulos and M. Claridge, editors, *In Insect Sounds and Communication: Physiology, Behaviour, Ecology, and Evolution*, pages 421–436. CRC Press, 2006.

[108] N. Huang. Introduction to the Hilbert-Huang transform and its related mathematical problems. In *Hilbert-Huang Transform and Its Applications*, pages 1–26. World Scientific, 2005.

[109] F. Huber. New observations on bees, i and ii. *Translated by CP Dadant. Dadant & Sons, Hamilton*, 111:230, 1792.

[110] C. Huemmer, C. Hofmann, R. Maas, and W. Kellermann. The significance-aware epfes to estimate a memoryless preprocessor for nonlinear acoustic echo cancellation. In *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 557–561, Dec 2014.

[111] D. Hunter. *Guitar Effects Pedals-The Practical Handbook*. Hal Leonard Corporation, 2004.

[112] G. Hunter, D. Howard, S. Gauvreau, O. Duran, and R. Busquets. Processing of multimodal environmental signals recorded from a" smart" beehive. *Proceedings of the Institute of Acoustics*, 41(1):339–348, 2019.

[113] J.-M. Jot. An Analysis/Synthesis Approach to Real-Time Artificial Reverberation. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 221–224, San Francisco, CA, Mar 1992.

[114] J.-M. Jot, L. Cerveau, and O. Warusfel. Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model. In *Proc. 103rd Audio Engineering Society Convention*, New York, NY, Sep 1997.

[115] T. I. Karimov, D. N. Butusov, and A. I. Karimov. Computer simulation of audio circuits with vacuum tubes. In *2016 XIX IEEE International Conference on Soft Computing and Measurements (SCM)*, pages 114–116, May 2016.

[116] M. Karjalainen and J. Pakarinen. Wave digital simulation of a vacuum-tube amplifier. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, pages V–V, Toulouse, France, 2006.

[117] T. Katayama and M. Serikawa. Reduction of second order non-linear distortion of a horn loudspeaker by a Volterra filter-real-time implementation. In *Proc. 103th Audio Engineering Society Convention*, Sep 1997.

[118] G. S. Kendall. The decorrelation of audio signals and its impact on spatial imagery. *Computer Music Journal*, 19(4):71–87, 1995.

[119] E. Kermit-Canfield and J. Abel. Signal decorrelation using perceptually informed allpass filters. In *Proceedings of the 19th International Conference on Digital Audio Effects*, pages 225–31, 2016.

[120] A. Khamparia, D. Gupta, N. G. Nguyen, A. Khanna, B. Pandey, and P. Tiwari. Sound classification using convolutional neural network and tensor deep stacking network. *IEEE Access*, 7:7717–7727, 2019.

[121] O. Kilpinen and J. Storm. Biophysics of the subgenual organ of the honeybee, apis mellifera. *Journal of Comparative Physiology A*, 181(4):309–318, 1997.

[122] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante. Fast Deconvolution of Multichannel Systems using Regularization. *IEEE Trans. Speech Audio Process.*, 6(2):189–194, Mar. 1998.

[123] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990. cited By 5107.

[124] V. Kulyukin, S. Mukherjee, and P. Amlathe. Toward audio beehive monitoring: Deep learning vs. standard machine learning in classifying beehive audio samples. *Applied Sciences*, 8:1573, 09 2018.

[125] S. Kuo and R. Gireddy. Real-time experiment of snore active noise control. In *IEEE Int. Conf. on Control Applications*, pages 1342–1346, Oct 2007.

[126] S. M. Kuo and D. R. Morgan. *Active noise control Systems: Algorithms and DSP implementation*. Wiley, New York, NY, USA, 1996.

[127] Kuo, S. M. and Morgan, D. R. Active Noise Control: a Tutorial Review. *Proc. IEEE*, 87(6):943–973, Jun. 1999.

[128] E. Larsen and R. M. Aarts. *Audio bandwidth extension.* J. Wiley & Sons, 2004.

[129] A. Lattanzi, F. Bettarelli, and S. Cecchi. NU-Tech: The entry tool of the hartes toolchain for algorithms design. In *Proc. 124th Audio Engineering Society Convention*, pages 1–8, Amsterdam, The Netherlands, May 2008.

[130] J. Lee, E. Song, Y. Park, and D. Youn. Effective Bass Enhancement Using Second-Order Adaptive Notch Filter. *IEEE Trans. Consum. Electron.*, 54:663–668, May 2008.

[131] K.-A. Lee, W.-S. Gan, and S. M. Kuo. *Subband Adaptive Filtering: Theory and Implementation.* Wiley Publishing, 2009.

[132] I. J. Leontaritis and S. A. Billings. Input-output parametric models for non-linear systems part i: deterministic non-linear systems. *International Journal of Control*, 41(2):303–328, 1985.

[133] H. T. Lubis MZ, Pujiyati S and W. PD. Haar wavelet method to spectral analysis continuous wavelet transform 1d using whistle sound to position of dolphins (tursiops aduncus). *Journal of Applied & Computational Mathematics*, 5(2):1–8, 2016. doi: 10.4172/2168-9679.1000305.

[134] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.

[135] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[136] S. Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way.* Academic Press, Inc., Orlando, FL, USA, 3rd edition, 2008.

[137] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), July 1989. doi: 10.1109/34.192463.

[138] L. Marple. Computing the discrete-time "analytic" signal via fft. *IEEE Transactions on Signal Processing*, 47(9):2600–2603, Sept 1999.

[139] V. J. Mathews and G. L. Sicuranza. *Polynomial Signal Processing.* Wiley, New York, 2000.

[140] A. Mesaros, T. Heittola, and D. Ellis. Datasets and evaluation. In T. Virtanen, M. D. Plumbley, and D. Ellis, editors, *Computational Analysis of Sound Scenes and Events*, pages 147–179. Springer, 2018.

[141] Y. Meyer. *Ondelettes et opérateurs.* Number v. 3 in Actualités Mathématiques. Hermann, 1991.

[142] A. Michelsen, W. H. Kirchner, and M. Lindauer. Sound and vibrational signals in the dance language of the honeybee, apis mellifera. *Behavioral Ecology and Sociobiology*, 18(3):207–212, 1986.

[143] M. Morhac. A fast algorithm of nonlinear volterra filtering. *IEEE Transactions on Signal Processing*, 39(10):2353–2356, Oct 1991.

[144] J. Mourjopoulos. On the variation and invertibility of room impulse response functions. *J. Sound Vibr.*, 102(2):217–228, Sep. 1985.

[145] S. Müller and P. Massarani. Transfer-function measurement with sweeps. *J. Audio Eng. Soc.*, 49(6):443–471, Jun. 2001.

[146] F. E. Murphy, E. Popovici, P. Whelan, and M. Magno. Development of an heterogeneous wireless sensor network for instrumentation and analysis of beehives. In *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, pages 346–351, May 2015.

[147] Z. Mushtaq, S.-F. Su, and Q.-V. Tran. Spectral images based environmental sound classification using cnn with meaningful data augmentation. *Applied Acoustics*, 172:107581, 2020.

[148] S. T. Neely and J. B. Allen. Invertibility of a room impulse response. *J. Acoust. Soc. Amer.*, 66(1):165–169, Jul. 1979.

[149] P. A. Nelson and S. J. Elliott. *Active Control of Sound*. Academic Press, 1993.

[150] I. Nolasco. Audio-based beehive state recognition. Master's thesis, Queen Mary University of London, 2018.

[151] I. Nolasco, A. Terenzi, S. Cecchi, S. Orcioni, H. L. Bear, and E. Benetos. Audio-based identification of beehive states. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8256–8260. IEEE, 2019.

[152] A. Novak. Identification of nonlinear systems: Volterra series simplification. *Acta Polytechnica*, 47(4):72–75, May 2007.

[153] A. Novak, P. Lotton, and L. Simon. Synchronized swept-sine: Theory, application, and implementation. *Journal of the Audio Engineering Society*, 63(10):786–798, 2015.

[154] A. Novak, L. Simon, F. Kadlec, and P. Lotton. Nonlinear system identification using exponential swept-sine signal. *IEEE Trans. Instrum. Meas.*, 59(8):2220–2229, Aug. 2010.

[155] S. Ntalampiras, I. Potamitis, and N. Fakotakis. Acoustic detection of human activities in natural environments. *Journal of the Audio Engineering Society*, 60(9):686–695, 2012.

[156] B. Oldroyd. What's killing american honey bees? *PLOS Biol*, 5, 2007.

[157] H. Olesen. Properties and units in the clinical laboratory sciences. i. syntax and semantic rules iupac-ifcc recommendations 1995. *Clinica Chimica Acta*, 245(2):S5 – S21, 1996.

[158] A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-Time Signal Processing*. Prentice Hall International Inc., 1999.

[159] S. Orcioni. Improving the approximation ability of Volterra series identified with a cross-correlation method. *Nonlinear Dynamics*, 78(4):2861–2869, Sep. 2014.

[160] S. Orcioni, A. Carini, S. Cecchi, A. Terenzi, and F. Piazza. Identification of nonlinear audio devices exploiting multiple-variance method and perfect sequences. In *Proc. 144th Audio Engineering Society Convention*. Audio Engineering Society, 2018.

[161] S. Orcioni, M. Pirani, and C. Turchetti. Advances in Lee-Schetzen method for Volterra filter identification. *Multidimensional Systems and Signal Processing*, 16(3):265–284, 2005.

[162] S. Orcioni, A. Terenzi, S. Cecchi, F. Piazza, and A. Carini. Identification of volterra models of tube audio devices using multiple-variance method. *Journal of the Audio Engineering Society*, 66(10):823–838, 2018.

[163] J. Pakarinen and D. T. Yeh. A review of digital techniques for modeling vacuum-tube guitar amplifiers. *Computer Music Journal*, 33(2):85–100, 2009.

[164] A. Papachristoforou, J. Sueur, A. Rortais, S. Angelopoulos, A. Thrasyvoulou, and G. Arnold. High frequency sounds produced by cyprian honeybees apis mellifera cypria when confronting their predator, the oriental hornet vespa orientalis. *Apidologie*, 39(4):468–474, 2008.

[165] P. Peretti, S. Cecchi, L. Romoli, and F. Piazza. Adaptive Feedback Active Noise Control for Yacht Environments. *IEEE Trans. Control Syst. Technol.*, 22(2):737–744, March 2014.

[166] M. Petraglia, R. Alves, and P. Diniz. New Structures for Adaptive Filtering in Subbands with Critical Sampling. *IEEE Trans. Signal Process.*, 48(12):3316–3327, Dec. 2000.

[167] M. Petraglia and P. Batalheiro. Prototype Filter Design for Subband Adaptive Filtering Structures with Critical Sampling. In *Proc. IEEE International Symposium on Circuits and Systems*, volume 1, pages 543–546, Geneva, Switzerland, May 2000.

[168] M. R. Petraglia and S. K. Mitra. Performance Analysis Of Adaptive Filter Structures Based On Subband Decomposition. In *Proc. International Symposium on Circuits and Systems*, pages 60–63, Chicago, IL, USA, May 1993.

[169] D. Pevernagie, R. Aarts, and M. De Meyer. The acoustics of snoring. *Sleep Medicine Reviews*, 14(2):131 – 144, 2010.

[170] G. Potard and I. Burnett. Decorrelation techniques for the rendering of apparent sound source width in 3d audio displays. In *Proc. Int. Conf. on Digital Audio Effects (DAFx'04)*, 2004.

[171] S. G. Potts, P. Neumann, B. Vaissière, and N. J. Vereecken. Robotic bees for crop pollination: Why drones cannot replace biodiversity. *Science of the total environment*, 642:665–667, 2018.

[172] A. Primavera, M. Gasparini, S. Cecchi, W. Hariya, S. Murai, K. Oishi, and F. Piazza. A novel measurement procedure for Wiener/Hammerstein classification of nonlinear audio systems. In *Proc. 144th Audio Engineering Society Convention*, May 2018.

[173] A. Qandour, I. Ahmad, D. Habibi, and M. Leppard. Remote beehive monitoring using acoustic signals. *Acoustics Australia / Australian Acoustical Society*, 42(3):204–209, 2014.

[174] M. Ramsey, M. Bencsik, and M. I. Newton. Long-term trends in the honeybee 'whooping signal'revealed by automated detection. *PloS one*, 12(2):e0171162, 2017.

[175] M.-T. Ramsey, M. Bencsik, M. I. Newton, M. Reyes, M. Pioz, D. Crauser, N. S. Delso, and Y. Le Conte. the prediction of swarming in honeybee colonies using vibrational spectra. *Scientific reports*, 10(1):1–17, 2020.

[176] M. Rébillat, R. Hennequin, E. Corteel, and B. F. Katz. Identification of cascade of Hammerstein models for the description of nonlinearities in vibrating devices. *Journal of Sound and Vibration*, 330(5):1018–1038, Feb. 2011.

[177] M. Rébillat, R. Hennequin, C. E., and B. F. G. Katz. Prediction of harmonic distortion generated by electro-dynamic loudspeakers using cascade of hammerstein models. In *Proc. 128th Audio Engineering Society Convention*, London, UK, May 2010.

[178] M. J. Reed and M. O. Hawksford. Efficient implementation of the volterra filter. *IEE Proceedings - Vision, Image and Signal Processing*, 147(2):109–114, Apr 2000.

[179] Z. Ren, N. Cummins, V. Pandit, J. Han, K. Qian, and B. Schuller. Learning image-based representations for heart sound classification. In *Proceedings of the 2018 International Conference on Digital Health*, pages 143–147, 2018.

[180] A. Robles-Guerrero, T. Saucedo-Anaya, E. González-Ramírez, and J. I. De la Rosa-Vargas. Analysis of a multiclass classification problem by lasso logistic regression and singular value decomposition to identify sound patterns in queenless bee colonies. *Computers and Electronics in Agriculture*, 159:69–74, 2019.

[181] L. Romoli, S. Cecchi, P. Peretti, and F. Piazza. A Mixed Decorrelation Approach for Stereo Acoustic Echo Cancellation Based on the Estimation of the Fundamental Frequency. 20(2):690–698, Feb. 2012.

[182] L. Romoli, S. Cecchi, and F. Piazza. A Novel Decorrelation Approach for Multichannel System Identification. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6652–6656, Florence, Italy, May 2014.

[183] G. Sarton. The feminine monarchie of charles butler, 1609. *Isis*, 34(6):469–472, 1943.

[184] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini. Comparison of hammerstein and wiener systems for nonlinear acoustic echo cancelers in reverberant environments. In *17th International Conference on Digital Signal Processing (DSP)*, pages 1–6, July 2011.

[185] J. Schattschneider and U. Zölzer. Discrete-time models for nonlinear audio systems. In *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)*, December 1999.

[186] T. Schmitz and J. J. Embrechts. Nonlinear guitar loudspeaker simulation. In *Proc. 134th Audio Engineering Society Convention*, May 2013.

[187] T. Schmitz and J.-J. Embrechts. Hammerstein kernels identification by means of a sine sweep technique applied to nonlinear audio devices emulation. *Journal of the Audio Engineering Society*, 65(9):696–710, 2017.

[188] A. Selin, J. Turunen, and J. T. Tanttu. Wavelets in recognition of bird sounds. *EURASIP Journal on Advances in Signal Processing*, 2007(1):051806, Dec 2006.

[189] S. Shimauchi and S. Makino. Stereo Projection Echo Canceller with True Echo Path Estimation. In *1995 International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 3059–3062. IEEE, 1995.

[190] G. L. Sicuranza and A. Carini. On a class of nonlinear filters. In M. G. I. Tabus, K. Egiazarian, editor, *Festschrift in Honor of Jaakko Astola on the Occasion of his 60th Birthday*, volume TICSP Report #47, pages 115–144, 2009.

[191] J. Simpson. The problem of swarming in beekeeping practice. *Bee World*, 39(8):193–202, 1958.

[192] J. Simpson. The mechanism of honey-bee queen piping. *Zeitschrift für vergleichende Physiologie*, 48(3):277–282, 1964.

[193] J. Simpson and S. P. Greenwood. Influence of artificial queen-piping sound on the tendency of honeybee, apis mellifera, colonies to swarm. *Insectes Sociaux*, 21(3):283–287, 1974.

[194] J. O. Smith. *Physical audio signal processing: For virtual musical instruments and audio effects*. W3K publishing, 2010.

[195] M. M. Sondhi, D. R. Morgan, and J. L. Hall. Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem. *IEEE Signal processing letters*, 2(8):148–151, 1995.

[196] C. Stanciu, C. Paleologu, J. Benesty, S. Ciochina, and F. Albu. Variable-forgetting factor rls for stereophonic acoustic echo cancellation with widely linear model. In *Proc. 20th European Signal Processing Conference*, Bucharest, Romania, 2012.

[197] R. Stockwell. A basis for efficient representation of the s-transform. *Digital Signal Processing*, 17(1):371 – 393, 2007.

[198] A. Terenzi, V. Bruschi, S. Cornell, A. Castellani, and S. Cecchi. A multiband structure based on hammerstein model for nonlinear audio system identification. In *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 9–14. IEEE, 2019.

[199] A. Terenzi, V. Bruschi, S. Nobili, and S. Cecchi. Applying signal decorrelation to a guitar modulation effect. In *Proc. 149th Audio Engineering Society Convention*. Audio Engineering Society, 2020.

[200] A. Terenzi, S. Cecchi, S. Orcioni, and F. Piazza. Features extraction applied to the analysis of the sounds emitted by honey bees in a beehive. In *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 03–08. IEEE, 2019.

[201] A. Terenzi, S. Cecchi, and S. Spinsante. On the importance of the sound emitted by honey bee hives. *Veterinary Sciences*, 7(4):168, 2020.

[202] A. Terenzi, S. Cecchi, S. Spinsante, S. Orcioni, and F. Piazza. Real-time system implementation for bee hives weight measurement. In *2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, pages 231–236. IEEE, 2019.

[203] A. Terenzi, S. Spinsante, and S. Cecchi. Review on electric vehicles exterior noise generation and evaluations. In *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive*. IEEE, 2020.

[204] L. Tronchin. Non linear convolution and its application to audio effects. In *Proc. 131th Audio Engineering Society Convention*, Oct 2011.

[205] L. Tronchin, V. L. Coli, and F. F. Gionfalo. Modelling nonlinearities on musical instruments by means of volterra series. In *Proc. 142th Audio Engineering Society Convention*, May 2017.

[206] P. P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[207] R. Van der Zee, L. Pisa, S. Andonov, R. Brodschneider, R. Chlebo, M. Coffey, and et al. Managed honey bee colony losses in canada, china, europe, israel and turkey, for the winters of 2008–9 and 2009–10. *J Apic Res.*, 51:100–114, 2012.

[208] D. Van Engelsdorp, J. J. Hayes, R. Underwood, and P. J.S. A survey of honey bee colony losses in the united states, fall 2008 to spring 2009. *J Apic Res.*, 49:7–14, 2010.

[209] J. Vanderkooy and S. P. I. Thomson. Harmonic distortion measurement for nonlinear system identification. In *Proc. 140th Audio Engineering Society Convention*, Paris, France, Jun. 2016.

[210] A. Watson. Unobtrusive audio: Current trends and design considerations. In *Proceedings of 21st Audio Engineering Society Conference: Audio at Home*, Apr 2006.

[211] R. Wei, H. S. Kim, X. Li, J. J. Im, and H. J. Kim. A development of mechanism for reducing snoring. In *Int. Conf. on Electronics and Information Engineering*, volume 2, pages V2–242–V2–245, Aug 2010.

[212] A. M. Wenner. Sound communication in honeybees. *Scientific American*, 210(4):116–125, 1964.

[213] G. Wersényi. Evaluation of vibrating sound transducers with glass membrane based on measurements and numerical simulations. In *Proceedings of 132nd Audio Engineering Society Convention*, Gyor, Hungary, Apr. 2012.

[214] E. F. Woods. Means for detecting and indicating the activities of bees and conditions in beehives, 1957. US Patent 2806082A.

[215] D. T. Yeh, J. Abel, and J. O. Smith. Simulation of the diode limiter in guitar distortion circuits by numerical solution of ordinary differential equations. In *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France, September 10-15 2007.

[216] R. K. Yenduri, S. R. Chakravarthy, and S. M. Kuo. Quiet comfort beds with electronic noise reduction system. In *IEEE Int. Conf. on Industrial Technology*, pages 2593–2597, Dec 2006.

[217] A. Zgank. Bee swarm activity acoustic classification for an iot-based farm service. *Sensors*, 20(1):21, 2020.

[218] U. Zölzer. *DAFX: digital audio effects*. John Wiley & Sons, 2011.

[219] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. Springer-Verlag, 1990.