Signal processing and methods for advanced acoustic applications

Alessio Izzo

Centre for Signal and Image Processing
Department of Electronic and Electrical Engineering
University of Strathclyde
Glasgow

This dissertation is submitted for the degree of
Doctor of Philosophy

20 April 2020

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Alessio Izzo
20 April 2020

Abstract

From a loudspeaker manufacturer perspective, the intelligibility of sound can be significantly affected on both downstream and upstream side of a loudspeaker production chain, differently. On the downstream side, the sound intelligibility can be affected by characteristics of typical acoustic environments: sound waves are partially reflected by the physical boundaries of the environment leading to reverberation, echo and feedback problems. On the upstream side, the quality of the sound is directly correlated to the quality of the speaker: inspections and quality control protocols are conducted during pre-production, production, and pre-shipment stage to reduce the amount of damaged drivers.
The research presented in this thesis deals with signal processing algorithms in order to develop both robust downstream and upstream solutions of a loudspeaker production chain, providing increased performance and sound quality for advanced acoustic systems in realistic conditions.

On the downstream side of a loudspeaker production chain, the acoustic feedback problem is considered and a novel algorithm for the adaptive feedback cancellation in a single acoustic MIMO array is proposed. When a microphone is too close to the loudspeaker or the amplification is too large, acoustic feedback can occur where acoustic effects are perceived as howling and ringing, degrading sound intelligibility and sound quality. While the canonical methods (automatic gain control, notch filtering, phase modulation) provide a reactive solution with limited performance, gain and high computational complexity, the new framework namely, Partitioned Block Frequency Domain (PBFD) based Adaptive Feedback Cancellation (AFC) method, is able to tackle the acoustic feedback problem in large acoustic spaces. The results of the proposed framework is compared with the state of the art using real acoustic data showing superior performance with up to 18dB of Maximum Stable Gain (MSG) and 30 seconds less convergence time.

On the upstream side of a loudspeaker production chain, the use of radar micro-

Doppler for loudspeaker analysis is introduced for the first time. This approach offers the potential benefits to characterize the mechanical motion of a loudspeaker and identify defects. Increasing quality checks at various stages of production (with limited costs) can provide substantial benefits to loudspeaker manufacturers. Compared with acoustic based approaches, the use of a radar allows reliable measurements in an acoustically noisy end of production line. In addition, when compared to a laser vibrometric approach the use of radar micro-Doppler reduces the number of measurements required and provides direct access to the information of the metallic components of the loudspeaker.

Following the modelling of the radar return from the loudspeaker, a procedure to extract mechanically impaired features of the loudspeaker motion is introduced. Results show the ability to detect the linear and harmonic frequency responses of both good and defected speakers. These are used as features of a Bidirectional Long Short-Term Memory (BiLSTM) Recurrent Neural Network (RNN) classifier, leading to classification accuracy above the 98% on real data.

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

$\mathbf{A}\left[q,n\right]$	Prediction error filter coefficients matrix of the time varying source signal model
A	Direct path attenuation in the 3 parameters covariance matrix model of the acoustic feedback path impulse response
$A\left[q,n\right]$	Prediction error filter of the time varying source signal model
B	Bandwidth
c	Propagation speed of electromagnetic wave ($c \approx 3 \times 10^8 m/s$)
$C(\cdot,\cdot)$	Cohen class
(X', Y', Z')	Reference coordinate system
(X, Y, Z)	Space fixed coordinate system
(x, y, z)	Local of body fixed coordinate system
d	Initial delay in the 3 parameters covariance matrix model of the acoustic feedback path impulse response
$D_H\left[n\right]$	Set of NHS design parameters
A_{Dr}	Amplitude of the driving force
D_v	Amplitude of the vibration
e	Mathematical constant $e \simeq 2.71\ldots$
$e\left[n\right]$	A posteriori error signal
$\hat{F}\left[e^{j\omega}, n\right]$	Time varying frequency spectrum of the estimated feedback path

$\hat{F}\left[q,n\right]$	Time varying estimate of the feedback path model
$\mathbf{F}\left[q,n\right]$	Time varying feedback path polynomial matrix in MIMO scenario
$F\left[e^{j\omega},n\right]$	Time varying frequency spectrum of feedback path
$f_{ij}^{0}\left[n\right]\ldots f_{ij}^{n_F}\left[n\right]$	Time varying feedback path impulse response coefficients for loudspeaker-microphone (i,j) pair in MIMO scenario
$F_{ij}\left[q,n\right]$	Time varying feedback path model for the loudspeaker-microphone (i,j) pair
f_0	Carrier frequency of the radar signal
f_D	Doppler frequency
f_R	Frequency of the received radar signal
f_v	Vibration frequency of the target
f_{Dr}	Frequency of the driving force
f_{mD}	Micro-Doppler frequency shift
f_{trans}	Doppler frequency shift due to target translation
$\mathbf{G}\left[q,n\right]$	Time varying electro acoustic forward path polynomial matrix in MIMO scenario
$G\left[e^{j\omega},n\right]$	Time varying frequency spectrum of electro acoustic forward path
$g_{ij}^{0}\left[n\right]\ldots g_{ij}^{n_G}\left[n\right]$	Time varying electro acoustic forward path impulse response coefficients for loudspeaker-microphone (i,j) pair in MIMO scenario
$G_{ij}\left[q,n\right]$	Time varying forward path model for the loudspeaker-microphone (i,j) pair
$\mathcal{H}[e^{j\omega},k]$	DFT of time varying phase modulation filter response
$H\left(e^{j\omega},n\right)$	Time varying phase modulation filter frequency response
$h\left[n,\tau\right]$	Time varying phase modulation filter impulse response
$H\left[q,n\right]$	Time varying source signal model
$H_N\left[q,n\right]$	Adaptive notch filter banks

$I\left(t\right)$	In-phase component of the radar signal
I	Moment of inertia
$J\left[q,n\right]$	Electro acoustic forward path before the amplification
$K\left[n\right]$	Broadband gain of forward path
n	Discrete time variable
$PWVD(\cdot,\cdot)$	Pseudo Wigner-Ville Distribution
$\mathbf{q}$	Delay operator
$Q\left(t\right)$	Quadrature component of the radar signal
$\hat{\mathbf{R}}_{\mathbf{f}init}$	Acoustic feedback path impulse response covariance matrix estimate based on initial measurement
$\hat{\mathbf{R}}_{\mathbf{f},3}$	Acoustic feedback path impulse response covariance matrix estimate based on 3 parameters Sabine model
$\mathbf{R}_{\mathbf{f}}$	Acoustic feedback path impulse response covariance matrix
$\mathbf{R}_{\mathbf{v}}$	Source signal covariance matrix
$s_D\left(t\right)$	In-phase and quadrature component version of the received radar signal
$s_r\left(t\right)$	Received radar signal
$s_t\left(t\right)$	Transmitted radar signal
$s_t^{90\,\mathrm{deg}}\left(t\right)$	Phase shifted transmitted radar signal
$\mathcal{R}_t$	Time varying rotation matrix
$\mathcal{R}_X$	Elemental rotation matrix on x-axis
$\mathcal{R}_Y$	Elemental rotation matrix on y-axis
$\mathcal{R}_Z$	Elemental rotation matrix on z-axis
$\mathcal{R}_{X-Y-Z}$	General rotation matrix of roll-pitch-yaw convention
$\mathcal{R}_{Z-Y-Z}$	General rotation matrix of x-sequence convention

$SPWVD(\cdot,\cdot)$	Smoothed Pseudo Wigner-Ville Distribution
$STFT(\cdot,\cdot)$	Short-Time Fourier Transform
t	Time variable
T_{60}	Room reverberation time
$\bar{\mathbf{u}}[n]$	Loudspeaker signal vector in MIMO scenario
$\tilde{\mathbf{U}}$	Prefiltered loudspeaker signal Henkel matrix
$\tilde{\mathbf{u}}$	Prefiltered loudspeaker signal vector
$U\left[e^{j\omega},n\right]$	Time varying frequency spectrum on loudspeaker signal
$\mathbf{I}$	Unit matrix
$\bar{\mathbf{v}}[n]$	Source signal vector in MIMO scenario
$V\left[e^{j\omega},n\right]$	Time varying frequency spectrum on source signal signal
$v_1\left[n\right]\ldots v_S\left[n\right]$	Source signal in MIMO scenario
$\hat{\mathbf{w}}$	Skew symmetric matrix associated to the angular velocity vector $\mathbf{w}$
$\mathbf{V}$	Translation velocity of the center of mass of the rigid body
$\mathbf{w}'$	Angular velocity unit vector
$\mathbf{w}$	Angular velocity vector
v_t	Target velocity
w_x	Angular velocity of the body rotation on x-axis
w_y	Angular velocity of the body rotation on y-axis
w_z	Angular velocity of the body rotation on z-axis
$\mathbf{W}$	Weighting matrix fro regularization purpose
$w\left(t\right)$	Time window function
$w_G\left(t\right)$	Gaussian window function

$WVD(\cdot,\cdot)$	Wigner-Ville Distribution
$\hat{x}_f\left[n\right]$	Estimate of the acoustic feedback signal
$x_f\left[n\right]$	Acoustic feedback signal
$x[n]$	Discrete time domain signal
$x(t)$	Continuous time domain signal
$\bar{\mathbf{y}}[n]$	Microphone signal vector in MIMO scenario
$\tilde{\mathbf{y}}$	Prefiltered microphone signal vector
$u_1\left[n\right]\ldots u_L\left[n\right]$	Loudspeaker signal in MIMO scenario
$y_1\left[n\right]\ldots y_S\left[n\right]$	Microphone Signal in MIMO scenario

Greek Symbols

α	Azimuth angle of the target respect to the radar
α_p	Azimuth angle of the target in the reference coordinates system
α_r	Regularization parameter
β	Elevation angle of the target respect to the radar
β_n	Noise floor in the 3 parameters covariance matrix model of the acoustic feedback path impulse response
β_p	Elevation angle of the target in the reference coordinates system
β_{wn}	Angular wave number
$\epsilon\left[n\right]$	A priori error signal
γ	Damping constant
$\Psi\left(\cdot,\cdot\right)$	Kernel function of the Cohen Class
λ	Wavelength of the transmitter radar signal
$\mathbf{\Omega}$	Angular velocity vector of the body rotation
Ω	Scalar angular velocity of the body rotation

ω	Radial frequency variable
$\mathbf{\Phi}_r$	Weighting matrix fro regularization purpose
$\Phi\left(t\right)$	Phase of radar signal
φ	Euler angle:counter clockwise rotation around the z axis. Yaw angle in roll-pitch-yaw convention
π	Mathematical constant $\pi \simeq 3.14\ldots$
ψ	Euler angle:counter clockwise rotation around the x axis. Roll angle in roll-pitch-yaw convention
$\sigma\left(x,y,z\right)$	Reflectivity function of the point scatterer in the body fixed coordinate system
$\Pi\left(t\right)$	Rectangular function
$\mathbf{\Sigma}$	Diagonal matrix in the inverse covariance source signal matrix
$\sigma_r^2\left[n\right]$	Time varying source excitation signal variance
$\chi\left(\cdot,\cdot\right)$	Spectrogram
$\boldsymbol{\tau}$	Vector torque
τ	Applied torque on the mass
τ_{ex}	Excitation time
τ_{RIR}	Exponential decay time constant in 3-parameters covariance matrix model of the feedback path
θ	Euler angle:counter clockwise rotation around the y axis. pitch angle in roll-pitch-yaw convention
$\boldsymbol{\xi}$	Reference value for regularization purpose

Superscripts

$(\cdot)^T$	Transpose operator

Other Symbols

$\mathcal{P}$	Set of critical frequency

Acronyms / Abbreviations

ADAS	Automatic Drive Assistance System
AEC	Adaptive Echo Cancellation
AEQ	Automatic Equalization
AFC	Adaptive Feedback Cancellation
AGC	Automatic Gain Control
AIF	Adaptive Inverse Filtering
AM	Amplitude Modulation
ANF	Adaptive Notch Filtering
ANN	Artificial Neural Network
ATR	Automatic Target Recognition
BLST	Bidirectional Long Short-Term Memory
BPTT	BackPropagation Through Time
BRNN	Bidirectional Recurrent Neural Network
CM	Center of Mass
CNN	Convolutional Neural Network
DFT	Discrete Fourier Transform
DM	Delay Modulation
DNN	Deep Neural Network
DoA	Direction of Arrival
EM	Electromagnetic
FDM	Finite Difference Method
FEM	Finite Element Method
FFT	Fast Fourier Transform

FIR	Finite Impulse Response
FM	Frequency Modulation
FS	Frequency Shift
FT	Fast Fourier Transform
IIR	Infinite Impulse Response
LMR	Levenberg-Marquardt Regularization
LOS	Line of Sight
LPTV	Linear Periodically Time Varying system
LS	Least Squares
LSTM	Long Short-Term Memory
LTI	Linear Time Invariant system
LTV	Linear Time Varying system
MIMO	Multi Input Multi Output
MLP	Multi-Layer Perceptron
MSG	Maximum Stable Gain
NER	Name Entity Recognition
NHS	Notch filter based Howling Suppression
NMLS	Normalised Minimum Least Squares
NLP	Natural Language Processing
NN	Neural Network
PA	Public Address
PEM	Prediction Error Method
PFC	Phase modulation Feedback Control
PM	Phase Modulation

PWVD	Pseudo Wigner-Ville Distribution
RCS	Radar Cross Section
ReLU	Rectified Linear Unit
RF	Radio Frequency
RIR	Room Impulse Response
RLS	Recursive Least Squares
RNN	Recurrent Neural Network
SAR	Synthetic Aperture Radar
SGD	Stochastic Gradient Descent
SISO	Single Input Single Output
SNR	Signal-to-Noise Ratio
SPWVD	Smoothed Pseudo Wigner-Ville Distribution
STFT	Short Time Fourier Transform
STFT	Short Time Fourier Transform
TFD	Time-Frequency Distribution
TR	Tikhonov Regularization
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
URLS	Under-determined Recursive Least Squares
WVD	Wigner-Ville Distribution

Chapter 1

Introduction

1.1 Preface

"The importance of communication in human society has been recognized for thousands of years, far longer than we can demonstrate through recorded history" [1]. With this statement, the authors Richmond and McCroskey wanted to emphasize how communication plays a key factor on the development of the human beings, making them different from the other animal species. Thanks to all sound signals, not only the information content is carried between talker and listener, but even the emotional expression allowing other people to understand us. A clear example how emotional expression is carried from sounds is music, where different factors influence the emotional valence of a piece, as tempo, mode and loudness. From acoustic perspective, sounds quality has an important role to let listeners perceive and interpret sounds properly.

In order to let listeners perceive correctly the sound, the preservation and improvement of the quality of the sound is tackled from two opposite points of view of loudspeaker production chain by acoustic transducers manufacturers. On the downstream side, the sound intelligibility can be affected by characteristics of typical acoustic environments: sound waves are partially reflected by the physical boundaries of the environment leading to reverberation, echo and feedback problems. On the upstream side, the quality of the sound is directly correlated to the quality of the speaker: inspections and quality control protocols are conducted during pre-production, production, and pre-shipment stage to reduce the amount of damaged drivers. Thus, since the intelligibility of sound can be significantly affected by both characteristics of typical acoustic environments and physical defect of a speaker, downstream and upstream solutions of the loudspeakers production chain could be found.

Acoustic Feedback problem

On the downstream side of the production chain, the acoustic feedback problem is considered. When loudspeakers are used to reproduce speech or music signals in an acoustic environment, while at the same time microphones are present in the same room to capture local acoustic signals, it is unavoidable (most of the time) that loudspeaker sounds are captured by microphones, in addition to the local sounds. This problem is generally known as acoustic echo problem. The scenario becomes even more complicated when loudspeakers are used to reproduce local sound signals captured by microphones. Since, in this case, the echoes signals are highly correlated with the local sound signals, constructive interference of these signals at the microphones may lead to oscillations that are perceived as ringing and howling effects. This problem is known as acoustic feedback problem and it can be considered as one of the most long standing problems in acoustic signal processing. While acoustic echoes may degrade speech intelligibility and disturb a normal course of the speech, acoustic feedback (also referred to as the Larsen Effect) may distort speech and audio signals through howling, ringing, echoes and excessive reverberation.

Condition monitoring on production chain

All types of speakers, from the simple USB speakers to home theatre speakers up to the large professional speakers used in large concert halls may be affected by physical defect during the assembly on the production line. For this reason, loudspeakers condition monitoring is an important topic in audio manufacturing in order to both fulfil the customer expectations and reduce the manufacturer costs to replace the damaged driver. The quality of these speakers is ensured by conducting inspection and quality control protocols during pre-production, production, and pre-shipment stage. In this domain, laser based analysis tools have been shown to yield significantly better results compared to traditional acoustic ones. The former approach is more frequently used in advanced markets like automotive audio components and systems, while the latter is widely used in RD and manufacturing of acoustic transducers and consumer products (e.g. loudspeakers or audio products).

1.2 Motivation

One of the main challenging applications of acoustic feedback control is related with Public Address (PA) systems. Sound reinforcement systems for PA system are used in many different areas, such as concert hall, auditorium, information broadcasting in airports, train stations and other public environments such as conference systems for large scale meetings. Due to the acoustic modelling complexity of large venues, acoustic feedback control in PA systems is often limited to the howling suppression, without modelling the room acoustics. In the past 50 years several solutions have been proposed, modelled and implemented using both software and hardware. A common method consists to place a notch filter into the signal path. Although the loop gain is correctly decreased at the critical frequencies, this approach has a main drawback: the system can only react after howling or self-oscillations have occurred. For this reason, notch filtering cannot prevent the audience from experiencing howling and ringing effects. To keep high the quality of the sound, a proactive approaches is required, such as the Adaptive Feedback Cancellation (AFC) technique. It is aimed to predict the feedback signal component and then subtract its prediction from the microphone signal. Considering large acoustic rooms (with a reverberation time longer than 1 second), standard AFC algorithm will lead to a biased solution and high computational complexity. In this thesis, a new technique for an unbiased estimation of long acoustic feedback paths is investigated and proposed as downstream solution of the loudspeakers production chain.

If the sound intelligibility can be affected by the acoustic characteristics of the environments on the downstream side of loudspeaker production chain, it can be affected by physical defects of the driver on the upstream side: the quality of the speaker could be compromised during the production stage. Increasing the quality check on the production line could provide potential benefits to loudspeakers manufacturers, with limited costs. By characterizing the mechanical motion of a loudspeaker, defects could be identified and issues could be addressed. Since traditional acoustic and laser analysis have technical and practical limitations, the effectiveness of acoustic End-Of-Line tests (EOL) or acoustic measurements are limited by the surrounding environment. Normally, these techniques require specifically designed insulated booths or silent areas for the signal-to-noise ratio of audio data to be meaningful. There are two main limitations when laser-based scanner vibrometer systems (Scanning Vibrometer System (SCN)) are used in place of the traditional acoustic approach. The first is the requirement of a very large sets of measurements (up to almost 3000 points) to fully characterize a loudspeaker and its non linearities, thus being a serious time consuming activity. The second is the limitation due to the presence of any physical obstacle in the line of sight between the laser source and the membrane (or acoustic source) under test.

In this thesis, a novel approach based on radar micro-Doppler is investigated to analyse and measure the return from loudspeaker. This approach is motivated by the potential cost effectiveness and operational advantages that a radar based approach could introduce over acoustic and laser based ones. With respect to the traditional acoustic measurement, a radar based approach is not affected by the acoustic environmental factors, allowing its use for End of Line (EoL) test. Unlike the SCN system, the radar has the ability to cope with visual occlusion due to plastic parts and the capability of separation metallic components of a loudspeaker from non metallic ones through the use of the back-scattering intensity.

1.3 Original Contributions

The original contributions contained in this thesis are in the fields of signal processing and methods for advanced acoustic applications. The novel contributions are the following:

- In Chapter 5, an innovative method of using Adaptive Feedback Cancellation algorithm in large acoustic spaces is shown: the proposed Partitioned Block approach consists of slicing the feedback path in p segments of length P each (e.g the impulse response of the system) to improve the algorithm performance. It can be applied either in the time domain or in the frequency domain, where the latter, called Partitioned Block Frequency Domain, shows faster convergence, lower computational cost and higher estimation accuracy. The results of the proposed framework is compared with the state of the art using real acoustic data showing superior performance with up to 18dB of Maximum Stable Gain (MSG) and 30 seconds less convergence time.

- In Chapter 6, a model for the radar return from a loudspeaker based on the Thiele&Small parameters is developed: in order to analyse the radar signal echoes of loudspeakers and its micro-Doppler signatures using low cost radar systems, voice coil displacement models for single tone and sine sweep stimulus are needed.

- In Chapter 6, an innovative method to measure mechanical frequency response of loudspeakers with low cost radar is also proposed: in order to characterise the speaker with a single radar measurement, a methodology is required. Using matched filtering approach in a real scenario, both linear and non linear impulse responses of the speaker are obtained. Since the device under test is never linear, the non linear impulse responses will appear in correspondence to the harmonics of the input signal. Thanks to the exponential sine sweep, non linear products do not contaminate the linear impulse response. These occur at very precise anticipatory times before the linear response; in this way linear and non linear impulse responses can be isolated.

- In Chapter 7, a deep learning based system is designed in order to detect and classify faulty speakers: the system's response is affected in varying ways by different irregular defects, making the non-linear behaviour of the loudspeaker a powerful indicator of possible manufacturing problems. Applying Fourier Transform to the matched filter outputs, mechanical frequency responses are obtained and used as input vectors of the Bidirectional Long Short-Term Memory (BiLSTM) Recurrent Neural Network (RNN). Performance analysis shows an accuracy above 98% on real measurements.

1.4 Publications

Journal Publications

- Loudspeaker Analysis: A Radar Based Approach, A. Izzo, L. Ausiello, C. Clemente, J. J. Soraghan, IEEE Sensors Journal, doi=10.1109/JSEN.2019.2946987, ISSN=2379-9153.

- Multimodel CFAR detection in Foliage Penetrating SAR Images, A. Izzo, M. Liguori, C. Clemente, C. Galdi, M. Di Bisceglie, J.J. Soraghan, IEEE Transaction on Aerospace and Electronic System, Year 2017, vol. 53, issue:4, Pages:1769-1780.

Conference Papers

- Radar micro-Doppler for loudspeaker analysis: an industrial process application, A. Izzo, L. Ausiello, C. Clemente, J.J. Soraghan, International Radar Conference "Radar 2017", Belfast, 23rd-26th October 2017.

- Efficient Micro-Doppler based pedestrian activity classification for ADAS systems using Krawtchouk moments, A. Amann, A. Izzo, C. Clemente, 11th International Conference on Mathematics in Signal Processing "IMA"; Birmingham, 12th-14th December 2016.

- Partitioned Block Frequency Domain Prediction Error Method based Acoustic Feedback Cancellation for long feedback path, A. Izzo, L. Ausiello, C. Clemente, J.J. Soraghan, 11th International Conference on Mathematics in Signal Processing "IMA"; Birmingham, 12th-14th December 2016.

- A Location Scale Based CFAR Detection Framework for FOPEN SAR Images, M. Liguori, A. Izzo, C. Clemente, C. Galdi, M. Di Bisceglie, J.J. Soraghan; Sensor Signal Processing for Defence "SSPD 2015"; Edinburgh, 7th-8th September 2015.

1.5 Thesis Organization

The remainder of the Thesis is divided into seven chapters organised as follows:

- Chapter 2 provides an overview on the acoustic feedback problem, describing the reason why the system becomes unstable, leading to ringing and howling effects. In the second part of the chapter, a review of existing acoustic feedback control techniques are discussed, showing the difference between feedforward suppression and feedback cancellation techniques. Although the feedforward suppression technique is effective for feedback control, it has significant limitations. Due to the reactive nature of this technique, distortions are introduced in the loudspeaker signal affecting the sound quality. With the aim to find a better solution, the attention is moved on feedback cancellation techniques, with particular interest on adaptive feedback cancellation method.

- Chapter 3 introduces the basic concept of micro-Doppler effect in radar. Before to review the uses and models of micro-Doppler, the working principle of a radar, namely the Doppler effect, and the canonical form of a received radar signal are introduced. Having introduced the key concepts, an overview of the effect of the micro-Doppler in radar are provided. In order to understand how to extract the micro-Doppler signature, the concept of time-frequency analysis is also introduced, with a detailed description of the commonly used time-frequency analysis tools. To understand the effect of translation and rotation of a target, the basic principle of rigid body motion in the context of radar signal processing is analysed. For a good interpretation of the received radar echoes from a vibrating surface such as a loudspeaker and a better understanding of the effect of non linear motion dynamic, two related canonical cases are also taken in consideration, namely micro-Doppler induced by a vibrating point and pendulum oscillation. Finally, with the aim to exploit the concept of radar micro-Doppler for condition monitoring of loudspeakers, some aspects of the electrodynamic transducer motion, and how to acoustically characterize the behaviour of a speaker are introduced.

- Chapter 4 deals with deep learning algorithms. Motivated by the recent advances arising from deep learning application in different fields, the goal of this chapter is to give a brief overview of different deep learning techniques. The most general architecture is introduced, together with Convolutional Neural Network (CNN) and how these can be used in the radar domain. A particular emphasis is given to Recurrent Neural Network (RNN). Finally, to cope the vanishing gradient problem that affect the performance of RNNs, the Long Short-Time Memory (LSTM) is introduced.

- Chapter 5 presents a new framework for an unbiased estimation of long acoustic feedback paths, improving the sound intelligibility in scenarios such as public address systems. The Partitioned Block (PB) version of the traditional Prediction Error Method (PEM) based Levenberg-Marquardt Regularization (LMR)-Normalised Least Mean Square (NLMS) algorithm is introduced. The Partitioned Block approach consists of slicing the feedback path in p segments of length P each (e.g the impulse response of the system) to improve the algorithm performance. It can be applied either in the time domain or in the frequency domain, where the latter, called Partitioned Block Frequency Domain, shows faster convergence, lower computational cost and higher estimation accuracy. The results of the proposed framework is compared with the state of the art using real acoustic data showing superior performance in terms of Misalignment (MSL) and Maximum Stable Gain with less convergence time.

- Chapter 6 presents a novel use of radar micro-Doppler for loudspeaker analysis. The approach offers the potential benefits of characterising the mechanical motion of a loudspeaker in order to identify defects and design issues. Compared to acoustic based approaches, the use of a radar allows reliable measurements in an acoustically noisy end of production line. In addition, when compared with a laser vibrometric approach the use of radar micro-Doppler reduces the number of measurements required and provides direct access to the information of the metallic components of the loudspeaker. In this chapter experimental results and analysis of the micro-Doppler signatures of loudspeakers using low cost radar systems are presented. Based on Thiele&Small parameters, the voice coil displacement is modelled and micro-Doppler signatures for a single tone and a sine sweep stimulus are presented. Furthermore, in order to characterise the speaker with a single radar measurement, a methodology to measure mechanical frequency response of loudspeakers is also shown.

- Chapter 7 shows the ability of the radar technology to automatically classify faulty speakers, where a framework based on the mechanical impulse response computation is proposed. Although Convolutional Neural Networks (CNN) are mostly preferred in radar domain, here Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) is taken in consideration: handling the mechanical frequency response of the Device Under Test (DUT) as time series or sequence data makes LSTM-RNN suitable for classification purpose. In order to avoid the traditional problems in deep learning (namely overfitting, vanishing and exploding gradient problems) some solutions are embedded in the proposed deep learning based classifier architecture. Finally, performance analysis are also shown. The proposed architecture outperforms the traditional k-NN classifier, used for benchmark purposes.

- Chapter 8 presents a summary and conclusions of the Thesis, providing an overview of possible future directions of this research work.

Chapter 2

Acoustic feedback control for public address system

2.1 Introduction

During the travelling time, sound signals are often distorted from the source (e.g. the speaker) to the receiver (e.g. the listener). Distortion may appear due to several reasons depending on the scenario. In a train station, for example, a train approaching can be a source of distortion, as the background noise in a telephone communication or an echo in an auditorium. In particular, the echo problem in a scenario such as an auditorium is a topic tackled in a broad field, known as room acoustics [2]. Room acoustics is the broad term that describes how sound waves interact with a room. Each room, and all the objects contained in it, will react differently to different frequencies of the sound.

When a sound wave propagates in an enclosed space, the wave is partially reflected by the physical boundary of the environment (e.g. walls, floor, ceiling of the room). The listener will receive not only the direct component of the sound wave, but also a multitude of delayed and attenuated replicas of the source signal, denoted as indirect components. This effect is known as *reverberation* and, depending on the particular application, it is viewed either as a desired or undesired effect.

In the case reverberation is viewed as desired effect, as for example improving the acoustics in an auditorium, it is referred as *artificial enhancement of the reverberation effect*.

Contrariwise, when the aim is to improve the intelligibility of the speech, reverberation effects are viewed as undesired effect. In this case, it is referred as *dereverberation* where the aim is to cancel or suppress the indirect sound components, retaining the direct one.

In a typical dereverberation scenario, the source signal is captured using one or more microphones positioned relatively far from the source. For this reason, it is not possible to consider the source signal as reference signal, hence most of dereverberation algorithms perform a blind identification of the room acoustics, operating only on the available microphone signal.

In a scenario where loudspeakers are used to reproduce speech or sounds in acoustic environment while, at the same time, microphones capture local sound signals in the same room, the undesired *acoustic echo problem* appears. This problem differs from dereverberation mainly for two reasons:

- in acoustic echo scenario generally, the loudspeaker signal is typically available as reference signal, thus the room acoustic can be identified using non blind system identification technique, while in a dereverberation scenario the room acoustics can be performed only with blind identification techniques;
- both direct and indirect sound components of the echo signal have to be cancelled from the microphone signals, while for dereverberation purpose the direct sound component is retained.

The scenario can be further complicated in case loudspeakers are used to reproduce local sound signals captured by microphone, leading at the *acoustic feedback* problem. Due to the high correlation between the echo signals and the local sound signals, constructive interference of these signals at the microphone may lead to oscillations perceived as ringing and howling effects. Both, echo and feedback problems influence the quality of the sound: as acoustic echoes degrade the speech intelligibility and disturb the normal course of conversations, acoustic feedback problems affect speech and sound with distortions audible through howling, ringing, echoes and high reverberation.

Acoustic echo and feedback control play a crucial role in several applications [3, 4]. In-vehicle communications is an emerging application for acoustic signal processing, that receives a lot of interest from the automotive industry. The aim is to improve the comfort of speech conversations as well as the quality of audio playback within the vehicle, despite the high levels of, e.g., road, wind, and traffic noise. Since an in-vehicle communications system includes loudspeakers that reproduce the sound signals captured by nearby microphones, acoustic feedback is usually unavoidable. In addition, acoustic echoes may result from audio playback being picked up by the microphones. Some example applications are in-car communications systems, cockpit communications systems in aircrafts, and on-board passenger information systems in trains.

Acoustic echo and feedback control have been successfully applied even in the hearing aids domain [5–7]. Since the limited dimension of ears cavity, microphone and loudspeakers are placed close each other, making this device prone to the feedback and howling problems. The success of acoustic feedback control in hearing aids, suggested to the scientific community to use this technique in other similar applications. Acoustic feedback becomes a challenging problem when it is related to Public Address (PA) systems. Since extremely high model orders are required for modelling the acoustics of large places, acoustic feedback control in PA applications is often limited to the howling suppression. Aim of this research is instead to find a technique able to ideally suppress completely the feedback signal components, without affect the sound quality of the source signal.

After introducing the acoustic feedback problem in Section 2.2, a review of existing acoustic feedback control techniques are discussed in Section 2.3. In Section 2.4, the feedforward suppression techniques are presented. Although the feedforward suppression techniques are effective for feedback control, they have significant limitations: while Notch filter based Howling Suppression (NHS) methods are reactive, in the sense that howling can usually be perceived before being detected, other methods as Automatic Gain Control (AGC) or frequency modulation methods achieve limited gain or introduce distortions in the loudspeaker signals, affecting the sound quality. With the aim to find a better solution, the attention is moved on Feedback cancellation techniques in Section 2.5, with particular interest on adaptive feedback cancellation methods.

2.2 Acoustic Feedback Problem

Among the early pioneers of sound reproduction and amplification system, surely two engineers from the American electronics company Magnavox, Edwin Jensen and Peter Pridham, stand out from the others [8]. During a series of tests in their laboratory from 1911 to 1915, they connected a microphone and loudspeaker to a 12 volt battery, resulting in the first ever occurrence of acoustic feedback. The term acoustic feedback has been used to refer to the undesired acoustic coupling between a loudspeaker and a microphone. Every time that a microphone capture a desired sound signal which is then processed (e.g. amplified) and played back by a loudspeaker in the same environment, as in the case of a PA system, the

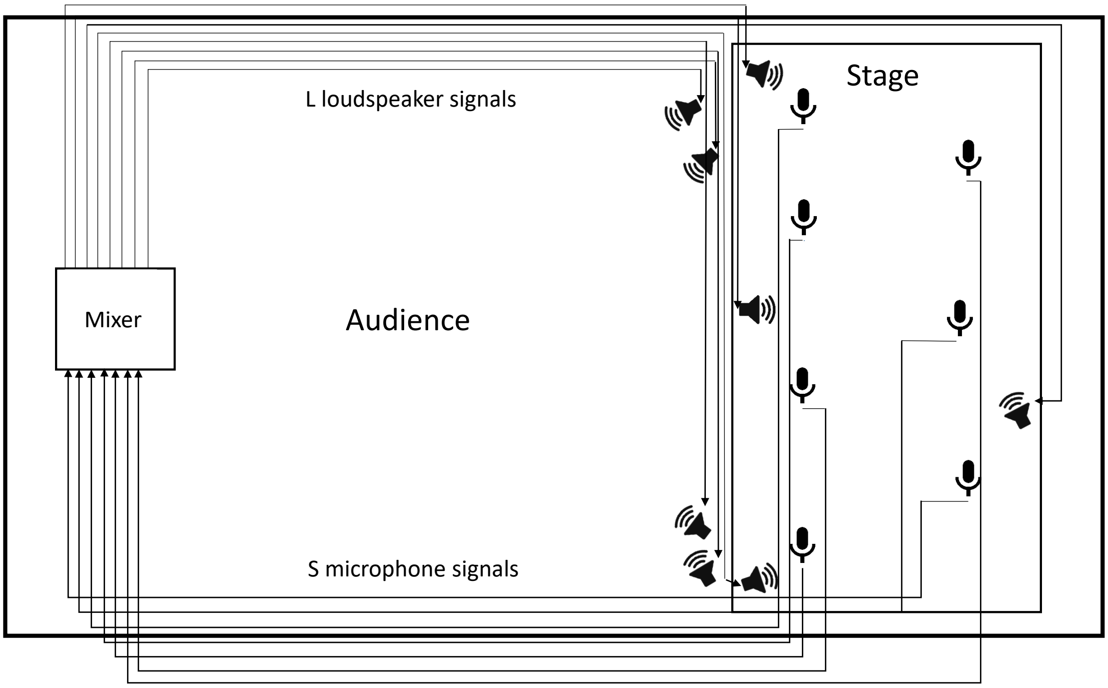


Fig. 2.1 Illustration of typical Public Address (PA) scenario: 7 microphones, 4 on-stage loudspeakers, 4 loudspeaker pointing at the audience and a mixing/signal processing/amplification console.

loudspeaker signal is unavoidably fed back into the microphone. In this event, a closed signal loop is created and as soon as the loop gain rises above a threshold, several undesired signals start affecting the system performance and degrading both sound intelligibility and sound quality, limiting the achievable amplification. Perceived result of this acoustic coupling is the characteristic howling effect. One of the first researchers to investigate the acoustic feedback problem was a Danish physicist Søren Absalon Larsen [9], reason why both the acoustic coupling and howling effect are sometimes also referred to as the *Larsen effect.* Of particular interest in this thesis are PA systems. A PA system is an electronic system comprising microphones, amplifiers, loudspeakers, and related equipment. It increases the apparent volume (loudness) of a human voice, musical instrument, or other acoustic sound source or recorded sound or music. PA systems are used in any public venue that requires that an announcer or performer be sufficiently audible at a distance or over a large area. A PA system may include multiple microphones or other sound sources, a mixing console to combine and modify multiple sources, and multiple amplifiers and loudspeakers for louder volume or wider distribution. In Figure 2.1 a typical PA system scenario is shown. In this scenario the sound of possibly multiple sound sources of interest are picked up by microphones. The microphone signals are then sent to the mixer console where additional processing may be applied. At this stage, not only the amplification is included, but even digital audio effects such as compression and equalization may be applied. The amplified mixed signals are then sent to the loudspeakers. Usually, microphones and loudspeakers are positioned in such a way that, considering their directivity, the loudspeakers sound does not directly hit the microphone. This is done in order to avoid direct acoustic coupling. However, in all sound reinforcement applications, loudspeaker sounds are inevitably reflected by the boundary of the acoustic environment, generating indirect acoustic coupling.

A Linear Time Invariant (LTI) system can be described and analytically modelled by its impulse response $h\left(t\right)$ (in time domain) and by its frequency response $H\left(e^{j\omega}\right)$ (in frequency domain). If the impulse response is known, for any input signal $x\left(t\right)$, the output signal $y\left(t\right)$ can be computed through the convolution operation:

$$y\left(t\right) = h\left(t\right) * x\left(t\right). \tag{2.1}$$

The frequency response $H\left(e^{j\omega}\right)$ can be obtained through the convolution theorem:

$$H\left(e^{j\omega}\right) = \frac{Y\left(e^{j\omega}\right)}{X\left(e^{j\omega}\right)} \tag{2.2}$$

with $X\left(e^{j\omega}\right)$ and $Y\left(e^{j\omega}\right)$ the Fourier transform of the input and output signals, respectively.

In case of a closed loop system, as in the case of PA system model in Figure 2.2, it can be described in discrete time domain as:

$$\begin{cases} \bar{\mathbf{y}}\left[n\right] & = \mathbf{F}\left[q,n\right]\bar{\mathbf{u}}\left[n\right] + \bar{\mathbf{v}}\left[n\right] \\ \bar{\mathbf{u}}\left[n\right] & = \mathbf{G}\left[\bar{\mathbf{y}}\left[n\right], n\right] \end{cases} \tag{2.3}$$

where:

$$\bar{\mathbf{v}}\left[n\right] = \left[v_1[n] \dots v_S[n]\right]^T \tag{2.4}$$

$$\bar{\mathbf{y}}\left[n\right] = \left[y_1[n] \dots y_S[n]\right]^T \tag{2.5}$$

$$\bar{\mathbf{u}}\left[n\right] = \left[u_1[n] \dots u_L[n]\right]^T \tag{2.6}$$

represent respectively the vectors of the S source signals $v_i[n]$, S microphone signals $y_i[n]$ and L loudspeaker signals $u_j[n]$, with $i = 1 \dots S$ and $j = 1 \dots L$.

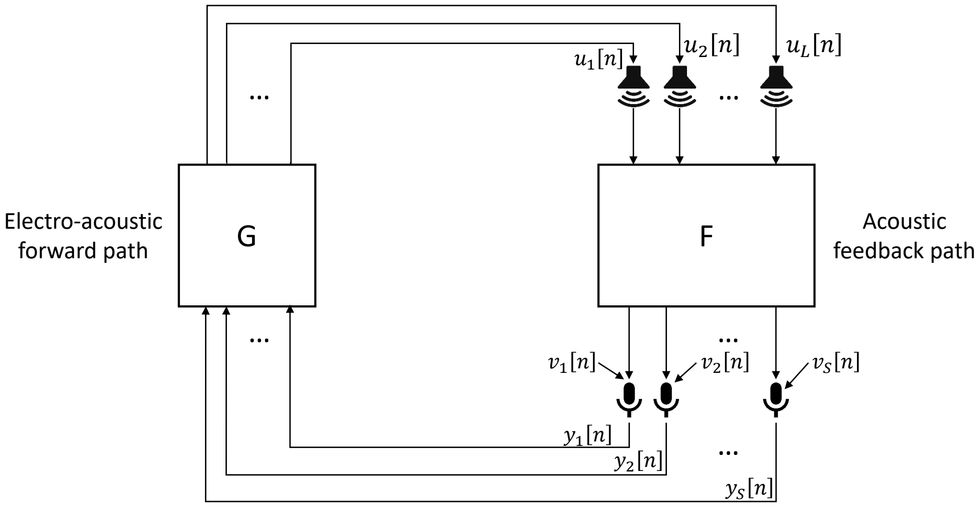


Fig. 2.2 PA system with S microphone and L loudspeaker in discrete time domain.

The acoustic coupling between the (j,i)th loudspeaker-microphone pair can be modelled by acoustic feedback path transfer function [4]:

$$F_{ij}\left[q,n\right] = f_{ij}^{0}\left[n\right] + f_{ij}^{1}\left[n\right]q^{-1}\dots f_{ij}^{n_F}\left[n\right]q^{-n_F} = \mathbf{f}_{ij}^{T}\left(n\right)\mathbf{q}, \tag{2.7}$$

with $\mathbf{q} = \left[1\, q^{-1}\dots\, q^{-n_F}\right]$ the delay operator[1], and $\mathbf{f}_{ij}\left(n\right)$ the FIR filter coefficients vector at the discrete time index n. In (2.3), the multi channel feedback path matrix $\mathbf{F}\left[q,n\right]$ is then defined as an $S \times L$ polynomial matrix:

$$\mathbf{F}\left[q,n\right] = \begin{bmatrix} F_{11}\left[q,n\right] & \cdots & F_{1L}\left[q,n\right] \\ \vdots & \ddots & \vdots \\ F_{S1}\left[q,n\right] & \cdots & F_{SL}\left[q,n\right] \end{bmatrix} \tag{2.8}$$

Thus, the generic element of the $S \times L$ polynomial matrix represents the acoustic feedback path between the jth loudspeaker and ith microphone, modelled as a linear time varying system of finite order n_F. The assumption made on the linearity of the acoustic feedback path is generally considered to be reasonable, since the effects of the sound propagation and reflections in acoustic environment are quasi independent. The finite order assumption can be justified by the observation that a typical room impulse response (RIR) has an exponentially decay envelope such

[1] $q^{-k}x\left[n\right] = x\left[n-k\right]$

that it can be truncated to have $n_F + 1 < \infty$ [10].
In the electro acoustic forward path, the S microphone signals are mixed and amplified to obtain the L loudspeaker signals, where some additional signal processing is usually performed. Since non linear dynamic processing is involved in this stage, the forward mapping $G_{ji}\left[q,n\right]$ between the $(i,j)\,th$ microphone-loudspeaker pair should be modelled as a non linear time varying filter. In order to be able to perform a stability analysis of the closed loop system, the assumption of linear time varying filter is made on the forward path as well, such that the $G_{ji}\left[q,n\right]$ can be modelled as the transfer function:

$$G_{ji}\left[q,n\right] = g_{ji}^{0}\left[n\right] + g_{ji}^{1}\left[n\right]q^{-1}\dots g_{ji}^{n_G}\left[n\right]q^{-n_G}. \tag{2.9}$$

So the multi channel forward matrix is then defined as a $L \times S$ polynomial matrix:

$$\mathbf{G}\left[q,n\right] = \begin{bmatrix} G_{11}\left[q,n\right] & \cdots & G_{1S}\left[q,n\right] \\ \vdots & \ddots & \vdots \\ G_{L1}\left[q,n\right] & \cdots & G_{LS}\left[q,n\right] \end{bmatrix} \tag{2.10}$$

Furthermore, it has been assumed that sound sources have sufficient directivity and are close enough to the respective microphones, such that the acoustic transfer function matrix from the sources to the microphones is an identity matrix.

While many sound reinforcement systems comprise multiple loudspeakers and microphones, most acoustic feedback control methods have been proposed in a single-channel context (i.e., for one loudspeaker and one microphone), without a framework for an extension to multi-channel systems being explicitly provided. For this reason, the acoustic feedback problem will be addressed only for a Single Input Single Output (SISO) scenario, without providing any framework for a Multi Input Multi Output (MIMO) scenario. For this reason the subscript (i,j) referred to microphone-loudspeaker pair will be omitted.

In a closed loop system as the single channel sound reinforcement system in Figure 2.3, the closed-loop frequency response from the source signal to the loudspeaker signal can be expressed as follows:

$$\frac{U\left[e^{j\omega},n\right]}{V\left[e^{j\omega},n\right]} = \frac{G\left[e^{j\omega},n\right]}{1 - G\left[e^{j\omega},n\right]F\left[e^{j\omega},n\right]} \tag{2.11}$$

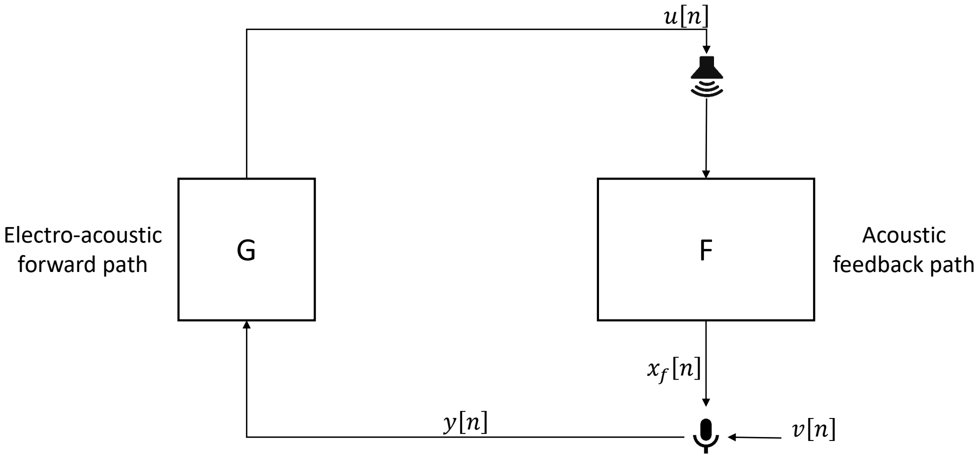


Fig. 2.3 Generic PA system in Single Input Single Output (SISO) scenario.

Here, $\omega \in [0, 2\pi]$ represents the radial frequency variable, $U\left[e^{j\omega}, n\right]$ and $V\left[e^{j\omega}, n\right]$ denote the short term frequency spectrum of the time varying loudspeaker and source signals, while $G\left[e^{j\omega}, n\right]$ and $F\left[e^{j\omega}, n\right]$ the short term frequency responses of the acoustic feedback and electroacoustic forward paths respectively, which can be calculated using the short-time discrete Fourier transform. The term $G\left[e^{j\omega}, n\right] F\left[e^{j\omega}, n\right]$ appearing to the denominator of the equation (2.11) is often referred to as "loop response" (the corresponding magnitude response $\left|G\left[e^{j\omega}, n\right] F\left[e^{j\omega}, n\right]\right|$ is referred to as loop gain, while $\angle G\left[e^{j\omega}, n\right] F\left[e^{j\omega}, n\right]$ is the phase response referred to as loop phase). When the closed loop system exhibits instability, oscillations may appear and the characteristic howling sound is perceived. For linear time-varying system, the proper way to evaluate the stability of a system is through the so-called circle criterion [11]. However, in order to achieve consistency with the literature on acoustic feedback control, the slowly time varying assumption of electro-acoustic forward path and feedback path characteristics is made. With this assumption, the closed loop system stability analysis can be done with the classical Nyquist stability criterion [12]. According to the Nyquist's criterion, the closed loop system becomes unstable if there exists a radial frequency $\omega = 2\pi f/fs$ for which:

$$\begin{cases} \left|G\left[e^{j\omega}, n\right] F\left[e^{j\omega}, n\right]\right| \geq 1 \\ \angle G\left[e^{j\omega}, n\right] F\left[e^{j\omega}, n\right] = 2n\pi, \qquad n \in \mathbb{Z} \end{cases} \tag{2.12}$$

If the unstable system is further excited at the critical frequency *f*, then an oscillation at this frequency will occur. This howling will be very annoying for the audience and the system gain generally has to be reduced. As a consequence, the maximum stable gain of the PA system has an upper limit due to the acoustic feedback [4, 13].

With the aim of quantifying the achievable amplification in a sound reinforcement system, it is usual to define a broadband gain $K[n]$ as the average magnitude of the forward path frequency response $G[e^{j\omega}, n]$ [4]:

$$K[n] = \frac{1}{2\pi} \int_0^{2\pi} \left| G\left[e^{j\omega}, n\right] \right| d\omega \tag{2.13}$$

and to extract it from

$$G[q, n] = K[n] \, J[q, n] \, . \tag{2.14}$$

Assuming that the electro acoustic forward path before the amplification $J[q, n]$ is known and $K[n]$ can be varied, the Maximum Stable Gain (MSG) of the PA system is defined as:

$$MSG[n] \ (d\mathrm{B}) = 20 \log_{10} K[n] \tag{2.15}$$

such that

$$\max_{\omega \in \mathcal{P}} \left| G\left[e^{j\omega}, n\right] F\left[e^{j\omega}, n\right] \right| = 1, \tag{2.16}$$

resulting in

$$MSG[n] \ (d\mathrm{B}) = -20 \log_{10} \left[\max_{\omega \in \mathcal{P}} \left| J\left[e^{j\omega}, n\right] F\left[e^{j\omega}, n\right] \right| \right], \tag{2.17}$$

where $\mathcal{P}$ denotes the set of frequencies that fulfil the phase condition in (2.12), also called critical frequencies of the PA system, such that

$$\mathcal{P} = \{\omega \angle G\left[e^{j\omega}, n\right] F\left[e^{j\omega}, n\right] = 2k\pi, \ k \in \mathbb{Z}\}. \tag{2.18}$$

Considering a flat frequency response of the electro-acoustic forward path $J[e^{j\omega}, n]$, the MSG can be defined as well as:

$$MSG[n] \ (d\mathrm{B}) = -20 \log_{10} \left[\max_{\omega \in \mathcal{P}} \left| F\left[e^{j\omega}, n\right] \right| \right]. \tag{2.19}$$

Thus, the MSG represents the maximum gain of the system that can be achieved without audible feedback oscillation. In [14] the MSG is computed through a statistical analysis of room acoustics. The author concluded that for a sound reinforcement system without feedback control, the MSG can be calculated as:

$$MSG\left[n\right]\ (d\mathrm{B}) = -10\log_{10}\left[\log_{10}\left(BT_{60}/22\right)\right] - 3.8 \tag{2.20}$$

with B and T_{60} representing the bandwidth and the reverberation time of the room, respectively. The gain margin is defined as the difference between the MSG and the actual gain of the system. From a sound point of view, a margin of $2/3d\mathrm{B}$ is recommended to avoid audible ringing effect. In order to control the Larsen effect and thus increase the MSG, several methods have been developed over the past decades; these are reviewed in the next section.

2.3 State of the Art in Acoustic Feedback Control

Most of the time, audio feedback is initiated when microphones are placed in the direction of the output speakers. Using a professional acoustic set-up is the best way to prevent such audio feedback problems. By strategically placing the main speakers in front of the microphone and using a number of small speakers, or monitors, which are pointing back to the place where the orator or each band member is positioned, audio feedback can be substantially reduced and even eliminated. These approaches based on the microphone and loudspeaker selections and positioning, suppression of discrete room modes using notch filters, and equalization of the room response using equalizer filters belong to the class of manual acoustic feedback control[15]. Although some of these approaches are still preferred among musicians, in this thesis only automatic acoustic feedback control methods are discussed. As reported in [16], to reduce the effects introduced by acoustic feedback, several techniques have been proposed in literature. They can be broadly classified into feedforward suppression and feedback cancellation techniques (Figure 2.4). In Figure 2.5 is illustrated how these two types of techniques control the acoustic feedback. While the feedforward suppression techniques modify the electro acoustic forward path $G\left[q,n\right]$ from the microphone signal $y\left[n\right]$ to the loudspeaker signal $u\left[n\right]$ for suppressing the feedback effect, the feedback cancellation techniques make an estimation $\hat{F}\left[q,n\right]$ of the acoustic feedback path $F\left[q,n\right]$ to create a signal $\hat{x}_f\left[n\right]$ to cancel the feedback signal $x_f\left[n\right]$. The motivation for both categories is to ensure that the conditions of the Nyquist stability criterion in equation (2.12) are not satisfied in order to avoid

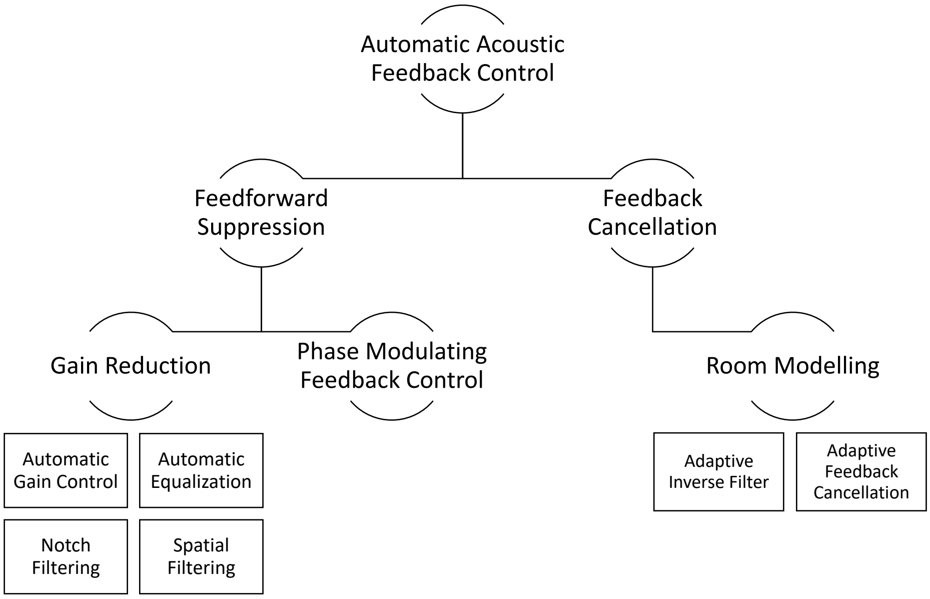


Fig. 2.4 Categorized automatic acoustic feedback control methods.

instability. The feedforward suppression techniques modify the electro acoustic forward path transfer function $G\left[e^{j\omega},n\right]$ to avoid that loop response fulfils the conditions of the Nyquist stability criterion by for example carrying out a gain reduction. The feedback cancellation techniques minimize the contribution of $F\left[e^{j\omega},n\right]-\hat{F}\left[e^{j\omega},n\right]$. Clearly, an ideal feedback cancellation system is better than a feedforward suppression system, since it removes the feedback contribution of $F\left[e^{j\omega},n\right]$ completely and provides an unmodified electro acoustic forward path transfer function $G\left[e^{j\omega},n\right]$. Before introducing feedback cancellation techniques, a brief overview on feedforward suppression techniques is given in the next section.

2.4 Feedforward Suppression Techniques

In feedforward suppression techniques, the electro acoustic forward path $G\left[q,n\right]$ is modified in such a way that the system is stable in conjunction with the feedback path $F\left[q,n\right]$. It can be further divided in two categories: gain reduction and phase modulation methods.

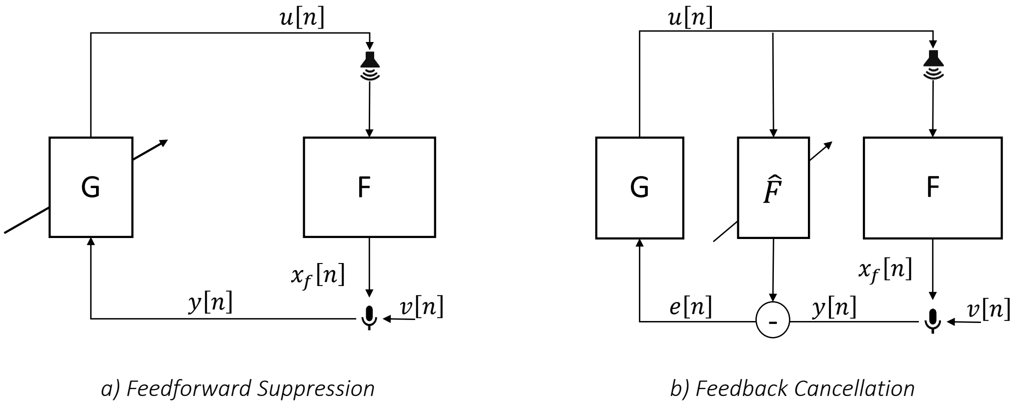


Fig. 2.5 A single channel acoustic feedback control system. The arrows through blocks indicate modifications made specifically for feedback control. a) Feedforward suppression scheme. b) Adaptive Feedback Cancellation (AFC) scheme.

2.4.1 Gain Reduction Methods

The gain reduction can be simply and effectively performed by the users of audio systems using a volume control to reduce $G\left[e^{j\omega}, n\right]$. To automate the action of a human operator for preventing or eliminating howling effects in a sound reinforcement system, more sophisticated automatic gain reduction methods exist. Depending on the bandwidth to which gain reduction is applied, three methods can be differentiated:

- *Automatic Gain Control* (AGC) method [17, 18]: the gain is reduced equally in the entire frequency bandwidth, by decreasing the amplification in the electro acoustic forward path;
- *Automatic Equalization* (AEQ) method [18, 19]: the gain reduction is applied in critical sub-bands of the entire frequency range, in particular in the sub-bands where the loop gain is close to the unit;
- *Notch filter based Howling Suppression* (NHS) methods [20, 21]: the gain is reduced in narrow frequency bands around critical frequencies.

In [17], the first AGC method was proposed: in case howling is detected, the broadband gain is immediately reduced, and after a time interval the gain is restored to the initial condition. Tonal components are discriminated from howling frequencies components by assuming that the latter one persists in time for several seconds. In [22], the gain of the signal path is adjusted by a gain selection logic: in the presence of howling, the gain is set to zero and the output frame is with all zero valued samples. Of course, AGC methods do not increase the MSG since the spectral shape of the loop gain is not altered. On the other hand, AGC methods have shown a good reliability: if the gain is sufficiently reduced, an unstable system is guaranteed to be stabilized. Consequently, most of the acoustic feedback control methods include the AGC as a rescue procedure in the case all the other methods fail.

A sub-band implementation of the AGC method lead the way to the AEQ methods, as proposed in [18]. If howling detection is performed in frequency sub-bands, then the gain reduction can be limited to these sub-bands in which howling is detected. In [19], the howling detection is performed firstly in wide sub-bands, and subsequently the most critical sub-bands are further divided in narrow sub-bands where howling is repeated. In [23] automatic equalization filter for in-car communication (ICC) systems is presented, where second adaptive filter is introduced that aims at equalizing the sound, radiated from one or more loudspeakers, at the listener's position in order to achieve a linear frequency response.

Depending on how the howling detection and notch filtering are performed, jointly or separately, the NHS methods can be divided in one or two stages. The one stage method requires a proactive approach to instability detection, while the latter is reactive in the sense that notch filters are activated only after howling is detected. In [20], one stage NHS method is proposed, where suppression of acoustic howling is developed based on Regularized Adaptive Notch Filters (RANF). More popular and widespread gain reduction method for acoustic feedback control present in the market are the two stages NHS methods. In [21], a combination of correlation and power based features are used to detect howling, while adaptive notch filters are used for the actual suppression of howling. Due to the reactive nature of this approach, it tries to stabilize the system and suppress the howling after that oscillations have occurred. In both one and two stage approach, the most critical part is the howling detection stage. Since howling is known to consist of sinusoidal signal components, the detection of howling is based on the frequency analysis of the microphone signal. Howling components usually show a frequency spectrum with large magnitude. However also voice speech and tonal music components have the same property. Fortunately, howling shows some distinct temporal and spectral features, useful to discriminate it from voice and tonal source signal components. Spectral criteria for discriminating between howling and tonal components are based on one or more of the following features: the power ratio of the candidate howling component and the entire spectrum, the power ratio of the candidate howling component and its harmonics, and the power ratio of the candidate howling component and its neighbouring frequency components. On the other hand, temporal criteria for howling detection rely on the observation that howling components typically persist for a longer time than tonal components and exhibit an exponentially increasing magnitude until the sound reinforcement system saturates [24].

Spatial filtering, also known as *beamforming* technique, can be employed for acoustic feedback control with the aim to smooth the loop response of the closed loop system $G\left[e^{j\omega}, n\right] F\left[e^{j\omega}, n\right]$ by using microphone and loudspeaker arrays which the received or emitted signals are processed with beamforming filters. The task of a beamformer is to selectively pick up signals coming from a predefined direction, the so-called *steering direction*, in a way that the designed microphone array has its main lobe in the direction of the source and null side lobes in the direction of the loudspeakers. Alternatively, beamforming filters may be applied on the loudspeaker array, but with the main lobe pointing to the audience, and null side lobes in the direction of the microphone [25]. In the context of hearing aids, beamforming algorithm is successfully applied to acoustic feedback cancellation problem in [5].

In all these methods, the general problem is that large values of the loop gain must be detected and the gain reduction may therefore be applied at times or frequencies where no instability is present, leading to sound quality degradations for the user. Furthermore, with gain reductions the audio system might only provide a less-than-desired amplification in $G\left[e^{j\omega}, n\right]$.

2.4.2 Phase Modulation Methods

In [14],[26] and [27] for the first time Frequency Shifting (FS) method for acoustic feedback control of PA system was introduced. It consists in the frequency shift of the microphone signals before these are amplified and sent to the loudspeaker. By applying the FS, the loop gain can be smoothed, such that, ideally, the MSG is determined by the average magnitude response rather than the peak magnitude response [27]. The author demonstrated that the stability of the system with FS and a diffuse field transfer function can be analysed using statistical model. In this way a gain improvement of 12dB could theoretically be achieved in a typical room by employing FS. This model also showed that this result should theoretically depend on the reverberation time. In [28], a digital FS implementation using a truncated FIR Hilbert filter was proposed. Recently, FS based feedback control was proposed also for acoustic guitar [29]. Generally, main drawback of the FS method is that it does not preserve the harmonic relations between the tonal components in the voice speech and music signals.

First approach of Phase Modulation (PM) feedback control method was introduced in [30], while Frequency Modulation (FM) was proposed in [31]. In both, the aim was to find the best modulation index in order to bypass the phase condition of the Nyquist's criterion. In [32], PM method was used in combination with adaptive filters. Although the strength of these methods is its simplicity, on the conceptual and computational point of view, this feedback control method shows three major drawbacks. First of all, the achievable MSG is limited. A MSG enhancement of 12dB has been found to be the theoretical upper bound using a FS technique in a typical room acoustic sound reinforcement system. To avoid the audible effect of the FS, the system should operate 6dB below the upper bound, limiting the practically realizable MSG to 6dB [24]. The second drawback is that applying a PM filter in the electro acoustic forward path, signal distortion in unavoidable. The last disadvantage is related to the fact that in a multi channel system the stability of the PFC is shown to decrease as the number of channels increase, hence the practical use of the PFC for PA system is expected to be limited [33].

In conclusion, although the feedforward suppression techniques is effective for feedback control, it has significant limitations. The gain reduction techniques limit the amplification in the forward path $G\left[q,n\right]$, which is contradicting the main purpose of PA system and audio reinforcement systems including hearing aids. Phase modification techniques, instead, can lead to severe sound quality distortions in loudspeaker signals $u\left[n\right]$. For this reason, in the next section an overview of feedback cancellation techniques will be given which, generally, allow a higher forward path gain $G\left[e^{j\omega},n\right]$ and better sound quality in $u\left[n\right]$.

2.5 Feedback Cancellation Techniques

In contrast to the feedforward suppression approaches, in room modelling methods a model of the acoustic feedback path is identified either off-line (e.g initialization of the sound reinforcement system) or on line. Two room modelling methods can be distinguished, depending on how the model is applied to accomplish feedback cancellation.

In Adaptive Feedback Cancellation (AFC), the acoustic feedback path model is used to predict the feedback signal component in the microphone signal. The predicted feedback signal is then subtracted from the microphone signal. The output will be a feedback compensated signal: if an accurate model of the acoustic feedback path is found, a nearly complete suppression of the acoustic coupling can be applied. In this event the feedback compensated signal will be an estimation of the source signal and high value of MSG can be achieved [4, 24].

The second room modelling approach is known as Adaptive Inverse Filtering (AIF). The inverse of the acoustic feedback path can be modelled and identified, equalizing the microphone signal by inserting the inverse model in the closed signal loop [4, 24]. Although this technique is not used in acoustic feedback control, a hybrid AIF-AFC approach was proposed in [34], in which the inverse model coefficients are adjusted based on the acoustic feedback model identified in the AFC algorithm. In [35], AIF was proposed for speech dereverberation.

Similarly to the Acoustic Echo Cancellation (AEC), in the AFC approach an adaptive filter is used to model, identify and track the impulse response of the acoustic feedback path. As in the adaptive microphone array beamforming approach, the AFC will suffer of biased solution due to the correlation problem. Unlikely from the case of AEC, the input signal of the adaptive filter and the disturb signal, namely the loudspeaker and source signals respectively, are correlated. When standard adaptive filtering theory is applied to AFC problem, the estimation of the impulse response of the acoustic feedback path will be biased, leading as consequence to a partial cancellation of the source signal into the microphone signal [6, 36]. For this reason, in any kind of AFC scheme is usually inserted a decorrelation method, applied in the closed loop or in the adaptive filtering circuit [10].

One of the earliest approach used in AFC to decorrelate signals in closed loop systems was introduced in [37] by injecting white noise signal, in non continuously mode, in the closed signal loop to identify the low frequency response of the acoustic feedback path. Continuous white noise injection was even proposed in [34, 38, 39]. Due to the sound degradation originated by the noise addition, several decorrelation techniques have been proposed, with the aim to keep high the sound quality of the source signal. In [39] a shaped spectrum noise signal was used in order to make the noise less perceptible. Non linear or time varying signal operations in the electro acoustic forward path for decorrelation duty have been proposed as well, as in [40], where frequency shifter, periodic phase or delay modulator are investigated. In [34] a half wave rectified version of the loudspeaker signal is used as non linear decorrelation technique. Finally, a processing delay in the electro acoustic forward path has been investigated in [6, 41] in the context of hearing aids.

While all these decorrelation techniques are rather effective when applied in the closed signal loop, their effect still influence the sound quality, making them not suitable in applications where delivering sound quality is the main task. Because of this, decorrelation techniques to be applied in the adaptive filter received a grown interest. The most promising technique is based on decorrelating prefilters, designed to whiten the source signal components in the microphone signal. Adopted in hearing aids in [42, 7, 43], the same approach was adapted in[36, 44] for PA system as well. A Frequency-Domain Adaptive Filter (FDAF) was proposed in [45] in combination with decorrelating prefilters for Double-Talk-Robust Acoustic Echo Cancellation. In [46] and [47], AFC using a Frequency-Domain Kalman Filter Approach with decorrelating prefilters was proposed. Although AFC can provide best result in terms of MSG and sound quality, here the main difficulty lies in the simultaneous identification of the optimal prefilter and the acoustic feedback path model from the closed loop signals.

2.5.1 Adaptive Feedback Cancellation

Adaptive feedback cancellers can be either applied in continuous or non continuous mode. Of course in the case of non continuous adaptation, widly used in hearing aids domain, it will adapt the coefficients of the filter only when instability is detected or when the input signal level is low [48–50]. Since the main aim of this thesis is to preserve the quality of the sounds, reactive methods (such as non continuous AFC) will not be treated. For this reason, it will be referred to the AFC algorithm just the meaning of continuous adaptation mode.

In a sound reinforcement system, the microphone signal $y\left[n\right]$ is a superposition of the source signal $v\left[n\right]$ and a feedback signal component $x_f\left[n\right]$, fed back from the loudspeaker to the microphone. The AFC approach to acoustic feedback control is aimed to predict the feedback component and subtract it from the microphone signal. So, when a FIR filter $\hat{F}\left[q,n\right]$ is placed in parallel with the acoustic feedback path, as in Figure 2.6, its input is the loudspeaker signal $u\left[n\right]$, and its desired output is the feedback signal. Accordingly, the feedback signal $x_f\left[n\right]$ can be predicted by the adaptive filter output signal $\hat{x}\left[n\right] = \hat{F}\left[q,n\right]u\left[n\right]$, which represents the estimated feedback signal. In this way, if the RIR estimate $\hat{F}\left[q,n\right]$

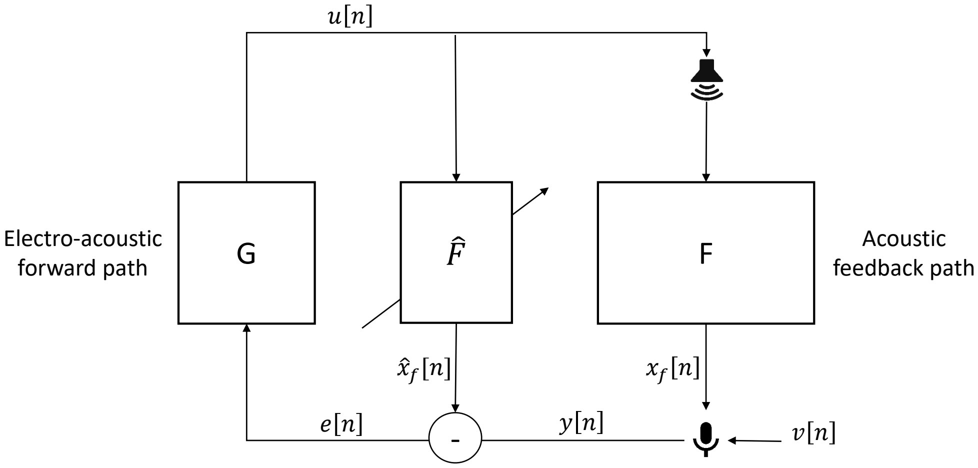


Fig. 2.6 Adaptive Feedback Cancellation (AFC) scheme in SISO scenario: the estimated feedback signal component $\hat{x}_f[n]$, obtained from $\hat{F}[q,n]$, is subtracted from the microphone signal $y[n]$.

is available at time n, then a feedback-compensated signal can be calculated as:

$$e[n] = y[n] - \hat{F}[q,n]\, u[n]\,, \tag{2.21}$$

where the error signal $e[n]$ approximates the source signal $v[n]$. The loudspeaker signal $u[n]$, finally, can be computed amplifying the error signal $e[n]$ with the forward path gain $G[q,n]$:

$$u[n] = G[q,n]\, e[n]\,. \tag{2.22}$$

For the system in Figure 2.6, the closed loop frequency response is given by:

$$\frac{U[e^{i\omega},n]}{V[e^{j\omega},n]} = \frac{G[e^{j\omega},n]}{1 - G[e^{j\omega},n]\left(F[e^{j\omega},n] - \hat{F}[e^{j\omega},n]\right)} \tag{2.23}$$

As a consequence, the Nyquist stability criterion in (2.12) is no longer valid, and the loop response become:

$$\begin{cases} \left|G[e^{j\omega},n]\left(F[e^{j\omega},n] - \hat{F}[e^{j\omega},n]\right)\right| \geq 1 \\ \angle G[e^{j\omega},n]\left(F[e^{j\omega},n] - \hat{F}[e^{j\omega},n]\right) = 2n\pi, \qquad n \in \mathbb{Z} \end{cases} \tag{2.24}$$

leading the MSG expression in (2.17) to be equal to:

$$MSG\left[n\right](d\mathrm{B}) = -20\log_{10}\left[\max_{\omega\in\mathcal{P}}\left|J\left[e^{j\omega},n\right]\left(F\left[e^{j\omega},n\right]-\hat{F}\left[e^{j\omega},n\right]\right)\right|\right]. \quad (2.25)$$

When the estimated RIR $\hat{F}\left[q,n\right]$ converges more to the real RIR $F\left[q,n\right]$, the feedback compensated signal $e\left[n\right]$ will get closer to the near end signal $v\left[n\right]$, thus leading a better audio quality. From (2.25), it is easy to understand that a better estimation of the RIR leads a larger achievable MSG. Theoretically, in the ideal hypothesis of $\hat{F}\left[q,n\right]\equiv F\left[q,n\right]$, the system would no longer exhibit a closed signal loop and hence the MSG would be infinitely large.

Considering a SISO scenario with the data model for discrete time range $[1,n]$ defined as:

$$\mathbf{y}\left[n\right] = \mathbf{U}\left[n\right]\mathbf{f}\left[n\right]+\mathbf{v}\left[n\right], \quad (2.26)$$

with the microphone signal $\mathbf{y}$ of dimension $N\times 1$, the loudspeaker signal $\mathbf{U}$ in the matrix form with dimension $N\times n_F$ and source signals $\mathbf{v}$ with dimension $N\times 1$, respectively defined as[2]:

$$\mathbf{y}\left[n\right] = \left[y\left[n\right]\ y\left[n-1\right]\ \ldots\ y\left[1\right]\right]^T, \quad (2.27)$$

$$\mathbf{U}\left[n\right] = \left[\mathbf{u}\left[n\right]\ \mathbf{u}\left[n-1\right]\ \ldots\ \mathbf{u}\left[1\right]\right]^T, \quad (2.28)$$

$$\mathbf{u}\left[n\right] = \left[u\left[n\right]\ u\left[n-1\right]\ \ldots\ u\left[n-n_F\right]\right]^T, \quad (2.29)$$

$$\mathbf{v}\left[n\right] = \left[v\left[n\right]\ v\left[n-1\right]\ \ldots\ v\left[1\right]\right]^T, \quad (2.30)$$

the estimation of the RIR can be found by minimizing the Least Square (LS) criterion:

$$\min_{\hat{\mathbf{f}}[n]}\left\{\mathbf{e}^2\left[n\right]\right\} = \min_{\hat{\mathbf{f}}[n]}\left\{\left[\mathbf{y}\left[n\right]-\mathbf{U}\left[n\right]\hat{\mathbf{f}}\left[n\right]\right]^T\left[\mathbf{y}\left[n\right]-\mathbf{U}\left[n\right]\hat{\mathbf{f}}\left[n\right]\right]\right\}, \quad (2.31)$$

where the vector $\mathbf{e}\left[n\right]=\mathbf{y}\left[n\right]-\mathbf{U}\left[n\right]\hat{\mathbf{f}}\left[n\right]$ represents the error signal in the discrete time range $[1,n]$. Consequently, the Least Squares (LS) estimated coefficients of the RIR are given by [4]:

$$\hat{\mathbf{f}}_{LS}\left[n\right] = \left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{y}. \quad (2.32)$$

[2] Not to be confused with the notation used in (2.6) for MIMO scenario.

Combining the Equations (2.32)and (2.26), the LS estimator of the RIR can be even expressed as:

$$\hat{\mathbf{f}}_{LS}\left[n\right] = \left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{x} + \left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{v}. \tag{2.33}$$

Expressing the RIR estimation in the form of the Equation (2.33), it is evident that the factor $\left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{v}$ introduces an error into the estimation of the RIR. Due to the closed loop nature of the system, the source and loudspeaker signals will be correlated. Hence the bias of the estimated RIR coefficients is generally non zero:

$$bias\left\{\hat{\mathbf{f}}_{LS}\left[n\right]\right\} = E\left[\left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{v}\right] \neq \mathbf{0}. \tag{2.34}$$

The result is that the standard adaptive filter will not only predict and cancel the feedback component, but also cancel a part of the the source signal. Consequently the feedback compensated signal $e\left[n\right]$ will be a distorted version of the source signal. Another common problem in room acoustic applications is that the matrix $\mathbf{U}^T\mathbf{U}$, which is inverted in (2.33), is ill-conditioned or even singular due to poor excitation [51]. Indeed, the identifiability of the RIR $\mathbf{f}\left[n\right]$ is only guaranteed if the loudspeaker signal $u\left[n\right]$ is persistently exciting of order n_F [24]. However, the dynamics of a typical RIR can often only be captured with several thousands of coefficients. On the other hand, the loudspeaker signal is usually a speech or audio signal, and may exhibit a nearly harmonic spectrum, such that its excitation order is far below n_F .If such an ill-conditioned situation occurs, there will typically be a large variance on the resulting RIR estimate when the source signal $v\left[n\right]$ is non-zero, as can be seen from the LS estimate covariance matrix:

$$cov\left\{\hat{\mathbf{f}}_{LS}\right\} = \left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{R}_{\mathbf{v}}\mathbf{U}\left(\mathbf{U}^T\mathbf{U}\right)^{-1}, \tag{2.35}$$

with the source signal covariance matrix defined as:

$$\mathbf{R}_{\mathbf{v}} = E\left[\mathbf{v}\mathbf{v}^T\right], \tag{2.36}$$

with $E\left[\cdot\right]$ and $cov\{\cdot\}$ the expectation and covariance operators, respectively. The interpretation of (2.35) can be related to the double talk problem, usually occurring even in AEC context. Indeed, in AEC, when the loudspeaker signal is active while the source signal is not, the covariance matrix of the acoustic echo path LS estimate is relatively small since $\mathbf{R}_{\mathbf{v}} \approx \mathbf{0}$. However, when both signals are active at the same time, the covariance matrix may become large, influencing negatively on the convergence speed of the adaptive filter, or even diverge. In AFC, the closed signal loop results always in a double talk situation, where this is made even worse than AEC scenario because the correlation of the source and loudspeaker signals. A standard technique to turn an ill posed problem into a well posed problem is to apply a regularization procedure [51]. Most regularized linear adaptive filtering algorithm are based on the Tikhonov regularized LS estimate, namely:

$$\hat{\mathbf{f}}\left[n\right] = \Big(\mathbf{U}^T\mathbf{U} + \alpha_r\mathbf{I})^{-1}\mathbf{U}^T\mathbf{y}, \tag{2.37}$$

where α_r denotes the only regularization parameter. This represents the minimized estimate of modified LS criterion in which the squared Euclidean norm of the RIR estimate is added to the sum of squared errors and weighted with the regularization parameter α_r:

$$\min_{\hat{\mathbf{f}}[n]} \left\{ \left[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}\left[n\right]\right]^T \left[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}\left[n\right]\right] + \alpha_r \parallel \hat{\mathbf{f}}\left[n\right] \parallel_2^2 \right\}. \tag{2.38}$$

The Tikhonov regularized LS estimate in (2.37) may be calculated recursively by initializing the adaptive filter input correlation matrix as $\mathbf{R}(0) = \alpha_r\mathbf{I}$, applying the standard Recursive Least Square (RLS) algorithm:

$$\begin{cases} \hat{\mathbf{f}}\left[n\right] = \hat{\mathbf{f}}\left[n-1\right] + \mathbf{R}^{-1}\left[n\right]\mathbf{u}\left[n\right]\epsilon\left[n\right] \\ \mathbf{R}\left[n\right] = \mathbf{R}\left[n-1\right] + \mathbf{u}\left[n\right]\mathbf{u}^T\left[n\right] \\ \epsilon\left[n\right] = y\left[n\right] - \mathbf{u}^T\left[n\right]\hat{\mathbf{f}}\left[n-1\right] \end{cases} \tag{2.39}$$

It is noted that the a priori residual $\epsilon\left[n\right]$ in (2.39) differs from the a posteriori error signal $e\left[n\right]$, which is sent back to the far end side, since it is a function of the estimated RIR of the previous time index $\hat{\mathbf{f}}\left[n-1\right]$.

A different regularization technique usually applied to RLS adaptive filtering algorithms, is known as Levenberg-Marquardt regularization [51]. This method is very similar to Tikhonov regularization, however no correction term is subtracted from the RIR weight update. The Levenberg-Marquardt regularized RLS algorithm is given by:

$$\begin{cases} \hat{\mathbf{f}}\left[n\right] = \hat{\mathbf{f}}\left[n-1\right] + \mathbf{R}^{-1}\left[n\right]\mathbf{u}\left[n\right]\epsilon\left[n\right] \\ \mathbf{R}\left[n\right] = \lambda\mathbf{R}\left[n-1\right] + \mathbf{u}\left[n\right]\mathbf{u}^T\left[n\right] + \left(1-\lambda\right)\alpha\mathbf{I} \\ \epsilon\left[n\right] = y\left[n\right] - \mathbf{u}^T\left[n\right]\hat{\mathbf{f}}\left[n-1\right] \end{cases} \tag{2.40}$$

To achieve an optimal trade-off between convergence speed and robustness, the Gauss-Newton method (with update term $\mathbf{u}\left[n\right])\,\mathbf{u}^{T}\left[n\right])$ was combined with the steepest descent method (with update term $(1-\lambda\mathbf{I})$ in the correlation matrix update, by means of a steering factor α [51].

In room acoustic application, the Under-determined Recursive Least Square (URLS) family [52], of which the Normalized Least Mean Squares (NLMS) algorithm and the Affine Projection Algorithm (APA) are the most well-known members, is much more appealing from a computational point of view. However, due to their under-determined nature, these algorithms are even more susceptible to convergence problems resulting from poor excitation. Therefore, regularization is also included in nearly every algorithm from the URLS family. The regularized APA is similar to the Levenberg-Marquardt regularized RLS algorithm, in that a scaled identity matrix is added to the adaptive filter input correlation matrix before inversion:

$$\begin{cases}\hat{\mathbf{f}}\left[n\right]=\hat{\mathbf{f}}\left[n-1\right]+\mu\mathbf{U}_{M}\left[n\right]\left[\mathbf{U}_{M}^{T}\left[n\right]\mathbf{U}_{M}\left[n\right]+\alpha\mathbf{I}\right]^{-1}\boldsymbol{\epsilon}_{M}\left[n\right]\\ \boldsymbol{\epsilon}_{M}\left[n\right]=\mathbf{y}_{M}\left[n\right]-\mathbf{U}_{M}^{T}\left[n\right]\hat{\mathbf{f}}\left[n-1\right]\end{cases}\tag{2.41}$$

where now the identity matrix is of dimension $M\times M$ with M the projection order, μ represents the step size, and

$$\mathbf{y}_{M}\left[n\right]=\left[y\left[n\right]\;y\left[n-1\right]\;\ldots\;y\left[n-M+1\right]\right]^{T},\tag{2.42}$$

$$\mathbf{U}_{M}\left[n\right]=\left[\mathbf{u}\left[n\right]\;\mathbf{u}\left[n-1\right]\;\ldots\;\mathbf{u}\left[n-M+1\right]\right].\tag{2.43}$$

The NLMS algorithm can then be obtained from the APA by setting $M=1$:

$$\begin{cases}\hat{\mathbf{f}}\left[n\right]=\hat{\mathbf{f}}\left[n-1\right]+\mu\frac{\mathbf{u}[n]\epsilon[n]}{\mathbf{u}^{T}[n]\mathbf{u}[n]+\alpha}\\ \epsilon\left[n\right]=y\left[n\right]-\mathbf{u}^{T}\left[n\right]\hat{\mathbf{f}}\left[n-1\right]\end{cases}\tag{2.44}$$

The regularized APA and NLMS algorithms described above can also be obtained by minimization of the Mean Square Error (MSE) between the estimated and true RIR. The Minimum MSE RIR estimate will depend on the statistical assumptions on the source signal $v\left[n\right]$, as well as on the true RIR $\mathbf{f}$, and can also be obtained as the minimizing estimate of a weighted and regularized LS criterion [4, 51], as shown in the next section.

2.5.2 Optimal Regularized and Weighted MSE estimator

A more general approach towards regularization is to replace the scaled Euclidean norm $\alpha_r \parallel \hat{\mathbf{f}}\left[n\right] \parallel_2^2$ into the criterion in Equation (2.38) by the weighted Euclidean norm of the deviation $\left[\hat{\mathbf{f}}\left[n\right] - \boldsymbol{\xi}\right]$ of the estimated RIR $\hat{\mathbf{f}}\left[n\right]$ from a reference value $\boldsymbol{\xi}$, with an $n_f \times n_f$ weighting matrix $\boldsymbol{\Phi}_r$, including also a $n \times n$ weighting matrix $\mathbf{W}$ in the LS term[51]:

$$\min_{\hat{\mathbf{f}}[n]} \left\{ \left[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}\left[n\right]\right]^T \mathbf{W} \left[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}\left[n\right]\right] + \left[\hat{\mathbf{f}}\left[n\right] - \boldsymbol{\xi}\right]^T \boldsymbol{\Phi}_r \left[\hat{\mathbf{f}}\left[n\right] - \boldsymbol{\xi}\right] \right\}. \tag{2.45}$$

Minimizing the criterion in equation (2.45) leads to a weighted LS estimate:

$$\hat{\mathbf{f}}_{WLS}\left[n\right] = \boldsymbol{\xi} + \left(\mathbf{U}^T\mathbf{W}\mathbf{U} + \boldsymbol{\Phi}_r\right)^{-1} \mathbf{U}^T\mathbf{W}\left(\mathbf{y} - \mathbf{U}\boldsymbol{\xi}\right). \tag{2.46}$$

Depending on the choice of the weighted matrices $\mathbf{W}$ and $\boldsymbol{\Phi}_r$ and on the reference value $\boldsymbol{\xi}$, the properties of the above estimate will change. By choosing $\mathbf{W} = \sigma\mathbf{I}$ and $\boldsymbol{\Phi}_r = \nu\mathbf{I}$, such that $\sigma^{-1}\nu = \alpha_r$ and $\boldsymbol{\xi} = 0$, the traditional Tikhonov regularized LS estimate given in (2.37) is obtained. A desirable property of a linear estimate is that it has minimum variance [53]. However, for biased estimates, such as the estimate in (2.46), both the bias and the variance should be minimized. A straightforward choice is to minimize the mean square error (MSE) between the estimated and true RIR:

$$\min_{\hat{\mathbf{f}}[n]} \left\{ E\left[\left(\hat{\mathbf{f}}\left[n\right] - \mathbf{f}\left[n\right]\right)^T \left(\hat{\mathbf{f}}\left[n\right] - \mathbf{f}\left[n\right]\right)\right] \right\}. \tag{2.47}$$

Since in a deterministic framework, the above criterion will lead to a dependency on the unknown true RIR $\mathbf{f}$, a Bayesian approach is suggested. It is more useful to minimize (2.47) considering both the source signal vector $\mathbf{v}$ and the true RIR $\mathbf{f}$ as one particular realization of a stochastic vector process. Assuming the first and second order moments of the source signal and true RIR vectors to be:

$$E\left[\mathbf{v}\right] = 0 \qquad cov\left\{\mathbf{v}\right\} = E\left[\mathbf{v}\mathbf{v}^T\right] = \mathbf{R}_{\mathbf{v}}, \tag{2.48}$$

$$E\left[\mathbf{f}\right] = \mathbf{f}_0 \qquad cov\left\{\mathbf{f}\right\} = E\left[(\mathbf{f} - \mathbf{f}_0)(\mathbf{f} - \mathbf{f}_0)^T\right] = \mathbf{R}_{\mathbf{f}}, \tag{2.49}$$

the minimized estimation of the criterion (2.47) is given by [53]:

$$\hat{\mathbf{f}}\left[n\right] = \mathbf{f}_0 + \left(\mathbf{U}^T{\mathbf{R}_{\mathbf{v}}}^{-1}\mathbf{U} + {\mathbf{R}_{\mathbf{f}}}^{-1}\right)^{-1} \mathbf{U}^T{\mathbf{R}_{\mathbf{v}}}^{-1}\left(\mathbf{y} - \mathbf{U}\mathbf{f}_0\right). \tag{2.50}$$

Comparing the Equation (2.50) with the weighted and regularized LS estimate in (2.46), the MSE optimal choice for the matrices $\mathbf{W}$ and $\mathbf{\Phi}_r$, and for the reference value $\boldsymbol{\xi}$ is:

$$\mathbf{W} = \mathbf{R_v}^{-1} \tag{2.51}$$

$$\mathbf{\Phi}_r = \mathbf{R_f}^{-1} \tag{2.52}$$

$$\boldsymbol{\xi} = \mathbf{f}_0 \tag{2.53}$$

Hence the criterion for deriving MSE optimally weighted and regularized LS based adaptive filtering algorithms can be formulated as:

$$\min_{\hat{\mathbf{f}}[n]} \left\{ \left[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}\,[n]\right]^T \mathbf{R_v}^{-1} \left[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}\,[n]\right] + \left[\hat{\mathbf{f}}\,[n] - \mathbf{f}_0\right]^T \mathbf{R_f}^{-1} \left[\hat{\mathbf{f}}\,[n] - \mathbf{f}_0\right] \right\}. \tag{2.54}$$

If no prior knowledge on the true RIR is available, then effectively no regularization is applicable. Also, when the source signal power is small, then the data will be more reliable and the impact of the regularization term will decrease. On the other hand, if the source signal power increases, the signal-to noise ratio (SNR) decreases, and then regularization starts playing a more important role.

2.5.3 Source Signal Covariance Matrix $\mathbf{R_v}$

As already mentioned, in order to reduce the bias in the estimation, a decorrelation technique is usually embedded in the AFC framework. Selecting the decorrelating prefilter technique will influence the estimation of $\hat{\mathbf{f}}\,[n]$ in (2.50), acting mainly on the covariance matrix $\mathbf{R_v}$. An unbiased feedback path estimate can be obtained with the so called direct method [54], when a model of the near end source signal is taken into account in the identification, corresponding to the noise model in system identification theory. Many audio signals can be closely approximated as a low-order Auto Regressive (AR) random process as:

$$v\,[n] = H\,[q, n]\,w\,[n]\,. \tag{2.55}$$

In this way the data model can be rewritten as:

$$y\,[n] = F\,[q, n]\,u\,[n] + H\,[q, n]\,w\,[n]\,, \tag{2.56}$$

with $w[n]$ an uncorrelated sequence such as Gaussian white noise or a Dirac impulse, and

$$H[q,n] = \frac{1}{A[q,n]} = \frac{1}{1 + a_1[n]\, q^{-1} + \ldots + a_{n_A}[n]\, q^{-n_A}}, \tag{2.57}$$

representing the Time Varying (TV) filter of order n_A of the source signal autoregressive (AR) model. Because of the non stationary nature of speech and music signals, the near end signal model $H[q,n]$ is time varying and shoud be estimated concurrently with the acoustic feedback path $F[q,n]$. This is possible by applying the Prediction Error Method (PEM) [13, 55–57]. In (2.57), the near end signal model is assumed to be an all pole model, which is relevant assumption for speech application. If the near end signal is a tonal audio signal, then the all pole model is usually not appropriate. In this case a cascade of two linear models is preferred. In this case the data model can be rewritten as:

$$y[n] = F[q,n]\, u[n] + H_1[q,n]\, H_2[q,n]\, w[n], \tag{2.58}$$

with $H_1[q,n]$ the model for the tonal component, while $H_2[q,n]$ the model for the noise component (all pole model). In [58] the tonal components are modelled with different linear prediction LP models: all pole LP, pole zero LP, pitch prediction LP, frequency warped all pole or selective all pole models.

Assuming the near end signal model in Equation (2.57), the covariance matrix $\mathbf{R_v}$ is factorized as:

$$\mathbf{R_v} = E\left\{\mathbf{v}\mathbf{v}^T\right\} = E\left\{\mathbf{H}\boldsymbol{\omega}\boldsymbol{\omega}^T\mathbf{H}^T\right\}. \tag{2.59}$$

Accordingly, the inverse covariance matrix can be factorised as:

$$\mathbf{R_v}^{-1} = \mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A}, \tag{2.60}$$

with $\mathbf{A}$ an unit upper triangular matrix

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & 1 & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{2.61}$$

and Σ a diagonal matrix. It is convenient from both physical and computational point of views to model the source signal as autoregressive process of order n_A, with time varying AR coefficients $a_i\,[n]$ and residual signal $r\,[n]$ [57]:

$$v\,[n] = -\sum_{i=1}^{n_A} a_i\,[n]\,v\,[n-i] + r\,[n]\,. \tag{2.62}$$

In this way the matrices $\mathbf{A}$ and $\mathbf{\Sigma}$ can be rewritten as:

$$\mathbf{A} = \begin{bmatrix} 1 & a_1\,[n] & a_2\,[n] & \cdots & a_{n_A}\,[n] & 0 & \cdots & 0 \\ 0 & 1 & a_1\,[n-1] & \cdots & a_{n_A}\,[n-1] & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \tag{2.63}$$

and

$$\mathbf{\Sigma} = diag\left\{\sigma_r^2\,[n]\ldots\sigma_r^2\,[1]\right\}. \tag{2.64}$$

In this way, it is possible to write the criterion in (2.54) as:

$$\min_{\hat{\mathbf{f}}[n]}\left\{\left[\tilde{\mathbf{y}} - \tilde{\mathbf{U}}\hat{\mathbf{f}}\,[n]\right]^T \mathbf{\Sigma}^{-1}\left[\tilde{\mathbf{y}} - \tilde{\mathbf{U}}\hat{\mathbf{f}}\,[n]\right] + \left[\hat{\mathbf{f}}\,[n] - \mathbf{f}_0\right]^T \mathbf{R_f}^{-1}\left[\hat{\mathbf{f}}\,[n] - \mathbf{f}_0\right]\right\}, \tag{2.65}$$

where the prefiltering matrix $\mathbf{A}$ has been shifted into the data term $\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}\,[n]$. Accordingly, the prefiltered microphone and loudspeaker signals are defined as:

$$\tilde{\mathbf{U}} = \left[\tilde{\mathbf{u}}\,[n]\,\tilde{\mathbf{u}}\,[n-1]\ldots\tilde{\mathbf{u}}\,[1]\right]^T = \mathbf{AU} \tag{2.66}$$

$$\tilde{\mathbf{y}} = \left[\tilde{y}\,[n]\,\tilde{y}\,[n-1]\ldots\tilde{y}\,[1]\right]^T = \mathbf{Ay}. \tag{2.67}$$

Interestingly, from a computational point of view this method offers the possibility of identifying the near-end signal AR coefficients non recursively. If the near-end signal exhibits short-term stationary behaviour, then the AR model will not have to be recalculated at each time instant [13]. This approach is particularly suited to near-end speech signals since speech exhibits short-term stationary behaviour. Consequently, the AR model can also be identified on a batch of loudspeaker and microphone data.

In conclusion, by minimizing the energy of the so called prediction error $\tilde{\epsilon}\,[n] = \tilde{y}\,[n] - \tilde{\mathbf{u}}^T\,[n]\,\hat{\mathbf{f}}\,[n-1]$, the prediction error method produces an estimate $\hat{F}\,[q,n]$ and $\hat{H}\,[q,n]$ of the feedback path $F\,[q,n]$ and the desired signal model $H\,[q,n]$, respectively.

2.5.4 RIR covariance matrix $\mathbf{R_f}$

Further improvements can be achieved if reliable information on the true RIR is available through the mean $\mathbf{f_0}$ and covariance matrix $\mathbf{R_f}$. Taking into account the covariance matrix of the true RIR $\mathbf{R_f}$, two methods for the estimation of $\mathbf{R_f}$ were suggested in [51]. In both the suggested methods, the aim is to restrict the number of regularization parameters. An intuitive solution is to have a diagonal covariance matrix, in order to have only $n_f + 1$ regularization parameters, compared to the general case of a non-diagonal estimate where $(n_f + 1)^2$ parameters are needed. The first method is to perform a batch RIR identification during the initialization of the signal enhancement system. Suppose that in this way an initial RIR estimate is obtained:

$$\bar{\mathbf{f}} = \left[\bar{f}_0 \bar{f}_1 \ldots \bar{f}_{n_f}\right]^T , \tag{2.68}$$

then a diagonal estimate of the true RIR covariance matrix may be constructed as follows:

$$\hat{\mathbf{R}}_{\mathbf{f},init} = diag\left\{\bar{f}_0^2 \bar{f}_1^2 \ldots \bar{f}_{n_F}^2\right\} . \tag{2.69}$$

The main disadvantage of this method is that it is non-robust to RIR changes, since the estimate of $\mathbf{R_f}$ is based on one particular measurement. Alternatively, several initial measurements under different conditions (e.g., different microphone or loudspeaker positions) could be averaged before constructing $\hat{\mathbf{R}}_{\mathbf{f},init}$, or, if permitted by the application, the initial measurement could be updated once in a while.

The second method is based on Sabine 3-parameter RIR model, proposed in [59]. In [60], the cited model is based on the observation that RIRs have a typical form, which may be characterized by three parameters, as shown in Figure 2.7:

- the initial delay d: the time needed for the loudspeaker sound wave to reach the microphone through a direct path;
- the direct path attenuation A which determines the peak response in the RIR
- the exponential decay time constant τ_{RIR}, which models the envelope of the reverberation, characterized by the tail of the RIR.

These three parameters may be estimated from the acoustic set-up (distance between loudspeaker and microphone, acoustic absorption of the walls, room volume, etc.), using Sabine's reverberation formulas [59]. Hence they can be

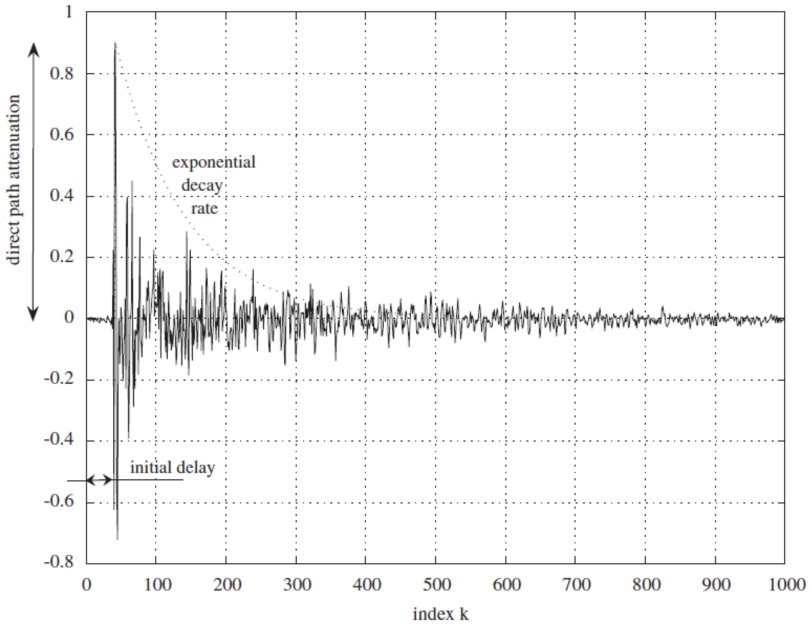


Fig. 2.7 Three parameters Sabine model of the Room Impulse Response (RIR) [24].

considered as prior knowledge. If these three parameters are taken into account, a diagonal estimate of the true RIR covariance matrix may be constructed as:

$$\hat{\mathbf{R}}_{\mathbf{f},3} = A \cdot diag \left\{ \underbrace{\beta_n \ldots \beta_n}_{d}, 1, e^{-\frac{2}{\tau_{RIR}}}, \cdots, e^{-2\frac{n_F - d}{\tau_{RIR}}} \right\}, \tag{2.70}$$

where β_n is a small number. The advantage of this method is that the parameter τ_{RIR} is invariant to RIR changes due to an arbitrary microphone or loudspeaker movement. Moreover, the other two parameters, d and A, are invariant to microphone or loudspeaker movements as long as the distance between the loudspeaker and the microphone remains constant. Hence the 3-parameter model is found to be more robust to RIR changes than the model based on an initial RIR measurement. Finally, two particular choices of $\mathbf{f_0}$ are of particular interest. Choosing $\mathbf{f_0} = \mathbf{0}$ leads to a Tikhonov type of regularization (TR), whereas choosing $\mathbf{f_0} = \hat{\mathbf{f}}\,[n-1]$ yields a Levenberg-Marquardt regularization (LMR) type. By choosing the latter approach, the Normalized Least Mean Square (NLMS) algorithm can be straightforward computed from the criteria in (2.65), leading to the PEM based

LMR-NLMS estimation of the RIR:

$$\hat{\mathbf{f}}\left[n\right] = \hat{\mathbf{f}}\left[n-1\right] + \mu \frac{\hat{\mathbf{R}}_{\mathbf{f},3}\tilde{\mathbf{u}}\left[n\right]\tilde{\epsilon}\left[n\right]}{\tilde{\mathbf{u}}^{T}\left[n\right]\hat{\mathbf{R}}_{\mathbf{f},3}\tilde{\mathbf{u}}\left[n\right] + \sigma_r^2}, \tag{2.71}$$

with μ the step size of NLMS algorithm and $\tilde{\epsilon}\left[n\right] = \tilde{y}\left[n\right] - \tilde{\mathbf{u}}^{T}\left[n\right]\hat{\mathbf{f}}\left[n-1\right]$ the pre-filtered version of the error signal.

Table 2.1 Summary table of different approaches used in acoustic feedback control.

Technique	**Pros**	**Cons**
Gain reduction (AGC,AEQ,NHS)	Reliability; robustness; extension of the approach to multi channel system straightforward (e.g. NHS)	Reactive method; limited Maximum Stable Gain $5-8d$B; high computational complexity.
Phase modulation (FS,PM,FM,DM)	Conceptually and computationally simple; robustness	Limited MSG up to $6d$B; signal distortion could appear; performance decreases as the number of channels increases.
Feedback cancellation (AFC,AIF)	Nearly complete elimination of acoustic coupling; large MSG	biased solution; high computational complexity.

2.6 Summary

In this chapter an overview on the acoustic feedback control for PA system was given. After a description in Section 2.2 of the acoustic feedback problem and the reason why a system becomes unstable, leading to ringing and howling effects, a review of existing acoustic feedback control techniques were discussed in Section 2.3. In Table 2.1, a summary of the different approaches used in acoustic feedback control was shown. In Section 2.4 feedforward suppression techniques were introduced. Although these techniques are robust and effective for feedback control, these methods have significant limitations. Not only is the MSG limited, but distortions are also introduced in the loudspeaker signals that will affect the sound quality. With the aim to preserve the quality of the sounds, in Section 2.5 feedback cancellation technique were considered.

In the next chapter, the basic concept of both radar micro-Doppler and electrodynamic transducer motion will be introduced, to better understand how radar sensors can provide benefits to acoustic transducers manufacturers on the upstream side of a loudspeaker production chain.

Chapter 3

Micro-Doppler effect in radar

3.1 Introduction

Radar systems have evolved tremendously since their early days when their functions were limited to target detection and target range determination. In fact, the word *RADAR* was originally an acronym that stood for RAdio Detection And Ranging. Recent breakthroughs in radar technology combined with the demand for compact, affordable and high precision radar for military and commercial applications, has led to a new season that can be defined as the “Modern Renaissance” in the methods and use of radar. This resurgence is a byproduct of escalating advancements in radar systems, driving the conventional radar solutions into obsolescence. Many of the upcoming sectors of technology growth, namely autonomous vehicles, unmanned aerial vehicle (UAV) and various commercial civilian applications, rely on new methods of fabrication and programming of radars. A contributing factor to enhanced capabilities and decreased costs is indeed the development of new antennas, such as phased array antennas allowing Radio Frequency (RF) sensors to raid in the automotive market.

Unlike optical sensors, the ability to penetrate sufficiently in adverse atmospheric conditions, such as fog, dust and rain made the radar a must-have sensor for the automotive industry. The reduced susceptibility of 79GHz millimeter wave frequency band radar to light condition, weather and clutter provides surveillance advantages over visual spectrum and IR camera technology, making radar sensors suitable for autonomous vehicles. The benefits of using this frequency band also enables the employment of radars in other applications, including detection and surveillance of Unmanned Aerial System (UAS) [61] and even medical monitoring. An example of it is the use of 80GHz band radar for remote heart rate monitoring, able to discriminate and characterize heartbeat accurately [62, 63].

The goal of this chapter is to provide the basic concept of micro-Doppler effect in radar. The working principle of a radar, namely the Doppler effect, is introduced in Section 3.2. To estimate and analyse correctly the Doppler shift, the coherent receiver is introduced in Section 3.2.1, leading to the formulation of the canonical form of a received radar signal. The effects of the micro-Doppler in radar are discussed in 3.3. An overview of the main time frequency analysis tools is given in 3.3.1. To measure the reflective strength of a target, the Radar Cross Section is considered in 3.3.2, where some models are also provided. The basic principles of rigid body motion in the context of radar signal processing are analysed in 3.4, with particular attention on the Euler angles. For a good interpretation of the received radar echoes from a vibrating surface such as a loudspeaker, two related canonical cases are considered: while micro-Doppler induced by a vibrating point is analysed in 3.5.1, in 3.5.2 the pendulum oscillation is considered to better understand the effect of non linear motion dynamic. Finally, with the aim to exploit the concept of radar micro-Doppler for condition monitoring of loudspeakers, in Section 3.6 some aspects of the electrodynamic transducer motion, and how to acoustically characterize the behaviour of a speaker are introduced.

3.2 The Doppler effect in radar

Observed for the first time in 1842 by the Austrian physicist Christian Doppler, the Doppler effect claims that the observed frequency of a light source depends on the velocity of the source relative to the observer. The apparent color of a star is changed by its motion: for a light source moving toward an observer, the color of the light source would appear bluer, while moving away from the observer the light would appear more red. In 1843, the Doppler phenomenon was experimentally proved by sound waves of a trumpeter of a train moving at different speed [64]. Valid for all kind of waves, the Doppler effect is also observed in radar.

A radar is an electrical system that transmits radio frequency (RF) electromagnetic (EM) waves toward a region of interest and receives and detects these EM waves when reflected from objects in that region. Although the details of a given radar system vary, the major subsystems must include a transmitter, antenna, receiver, and signal processor. Based on the antenna configurations, a radar system could be considered as monostatic or bistatic. In the monostatic configuration, one antenna serves both the transmitter and receiver. In the bistatic configuration,

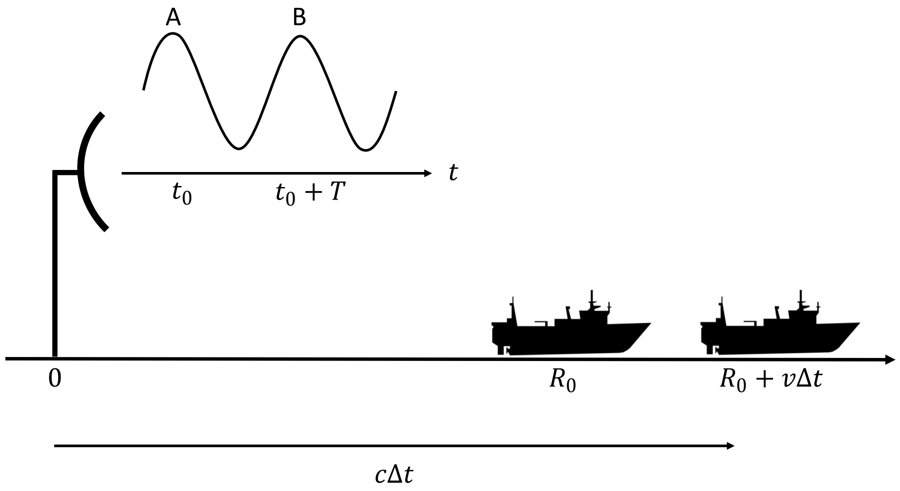


Fig. 3.1 The Doppler effect in radar: a radar transmits an EM signal and receives the return from a target.

there are separate antennas for the transmit and receive radar functions. Use of two antennas alone does not determine whether a system is monostatic or bistatic. If the two antennas are very close together, on the same structure for example, then the system is considered to be monostatic. The system is considered to be bistatic only if there is sufficient separation between the two antennas such that the angles or ranges to the target are sufficiently different [65].

Considering the monostatic configuration shown in Figure 3.1, an overview on radar working principle is provided. When an EM signal is transmitted with a target nearby, the received echo will be frequency shifted due to the target motion. The Doppler shift is determined by the target radial velocity, which represent the velocity component in the direction of the Line Of Sight (LOS). In $t = t_0$ the peak A is transmitted with the target in position R_0. The wave will reach the target after a time Δt, with propagation speed c ($c \approx 3 \cdot 10^8 m/s$). In that period, the target will travel the distance $v_t \Delta t$, with v_t its radial velocity. Thus, the travelling time of the signal can be extrapolated from the equivalence

$$c\Delta t = R_0 + v_t \Delta t \tag{3.1}$$

such that:

$$\Delta t = \frac{R_0}{c - v_t}, \tag{3.2}$$

with c the propagation speed of the EM wave. In $t = t_1$ the peak A returns to the radar, with t_1 equal to:

$$t_1 = t_0 + 2\Delta t = t_0 + \frac{2R_0}{c - v_t}. \tag{3.3}$$

Similar consideration can be done on the peak B, departed after a time T respect to the peak A, such that:

$$t_2 = t_0 + T + \frac{2R_1}{c - v_t}, \tag{3.4}$$

where R_1 is the target location when B leaves the radar, equal to:

$$R_1 = R_0 + v_t T. \tag{3.5}$$

The period of the received waveform T_R is equal to the difference between the arrival times of the two peaks:

$$T_R = t_2 - t_1 = t_0 + T + \frac{2\left(R_0 + v_t T\right)}{c - v_t} - \left(t_0 + \frac{2R_0}{c - v_t}\right) = T\frac{c + v_t}{c - v_t}. \tag{3.6}$$

Thus, the received frequency can be found from the ratio between the received and transmitted period as follow:

$$\frac{T_R}{T} = \frac{c + v_t}{c - v_t} \Rightarrow \frac{f_R}{f_0} = \frac{c - v_t}{c + v_t} = \frac{1 - v_t/c}{1 + v_t/c}. \tag{3.7}$$

As the velocity of the target v_t is usually much slower than the propagation speed c of the EM wave, such that $v_t \ll c$, then the following approximation can be used:

$$\frac{1}{1 + v_t/c} = 1 - \frac{v_t}{c} + \frac{v_t^2}{c^2} - \cdots. \tag{3.8}$$

Then, the received frequency can be approximated as:

$$f_R = f_0\left(1 - \frac{v_t}{c}\right)\left(1 - \frac{v_t}{c} + \frac{v_t^2}{c^2} - \cdots\right) = f_0\left(1 - \frac{2v_t}{c} + \cdots\right) \approx f_0\left(1 - \frac{2v_t}{c}\right). \tag{3.9}$$

Rewriting the equation (3.9) as

$$f_R \approx f_0 - \frac{2v_t}{c/f_0} = f_0 - \frac{2v_t}{\lambda} \tag{3.10}$$

with λ the wavelength of the transmitted signal, it is possible to extrapolate in easier way the Doppler frequency, defined as the difference between the received frequency and the transmitted one:

$$f_D = f_R - f_0 \approx -\frac{2v_t}{\lambda}. \tag{3.11}$$

If the radar is stationary, v_t will be the radial velocity of the target along the LOS of the radar. Velocity is defined positive when the object is moving away from the radar. As a consequence, the Doppler shift becomes negative.

3.2.1 Estimation and analysis of Doppler frequency: the quadrature detector

Radar systems can be classified as non-coherent or coherent configurations. A non-coherent system detects only the amplitude of the received signal, the coherent system detects the amplitude and phase. Noncoherent systems are often used to provide a two-dimensional display of target location in a ground map background. The amplitude of the signal at any instant in time will determine the brightness of the corresponding area of the display face. Noncoherent radars can be used in cases in which it is known that the desired target signal will exceed any competing clutter signal. All early radars were noncoherent; target detection depended on operator skill in discerning targets from the surrounding environment. For a coherent system, measurement of the phase of the received signal provides the ability to determine if the phase is changing, which can provide target motion characteristics.

The Doppler shift can be extracted using a coherent receiver, as the quadrature detector shown in Figure 3.2[64]. The output of the detector leads to the formulation of the canonical form of the received signal, composed by an in-phase component $I(t)$ and a quadrature phase component $Q(t)$ from the input signal. In the quadrature detector, the received signal is split into two mixers called synchronous detectors. In the synchronous detectors I, the received signal is mixed with a reference signal, the transmitted signal. In the other channel it is mixed with a 90° shift of the transmitted signal. Let $s_r(t)$ be the received signal, expressed as [64, 66]:

$$s_r(t) = a\cos\left[2\pi\left(f_0 + f_D\right)t\right] = a\cos\left[2\pi f_0 t + \Phi(t)\right] \tag{3.12}$$

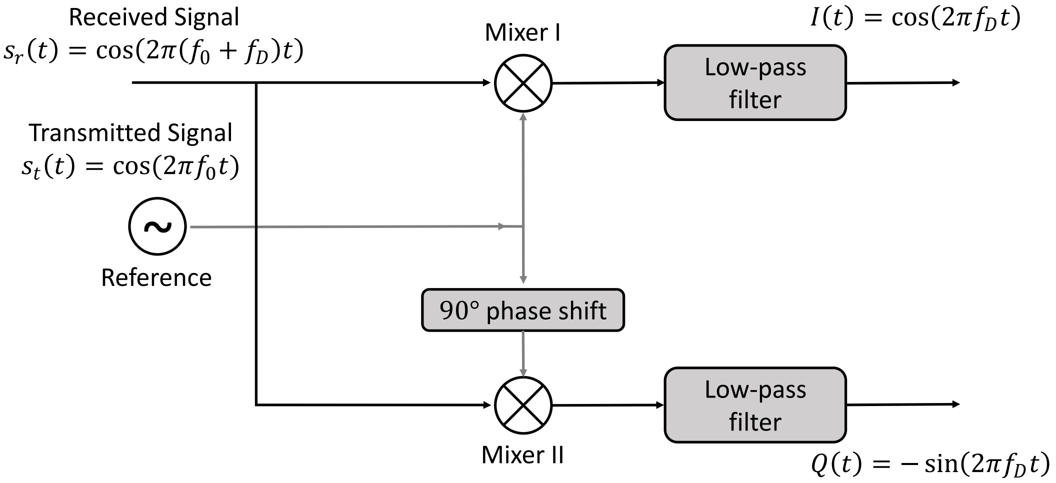

Received Signal
$s_r(t) = \cos(2\pi(f_0 + f_D)t)$
Transmitted Signal
$s_t(t) = \cos(2\pi f_0 t)$
Reference
Mixer I
Low-pass filter
$I(t) = \cos(2\pi f_D t)$
90° phase shift
Mixer II
Low-pass filter
$Q(t) = -\sin(2\pi f_D t)$

Fig. 3.2 Doppler shift extracted by a quadrature detector.

where a is the amplitude of the received signal, f_0 is the carrier frequency of the transmitter and $\Phi(t) = 2\pi f_D t$ is the phase shift on the received signal due to the target's motion. Mixing $s_r(t)$ with the transmitted signal $s_t(t)$

$$s_t(t) = \cos(2\pi f_0 t), \tag{3.13}$$

the output of the synchronous detector I can be obtained by using Werner formulas, and given by:

$$s_r(t)\, s_t(t) = \frac{a}{2} \cos[4\pi f_0 t + \Phi(t)] + \frac{a}{2} \cos \Phi(t). \tag{3.14}$$

After the low-pass filtering, the in-phase component $I(t)$ is found:

$$I(t) = \frac{a}{2} \cos \Phi(t). \tag{3.15}$$

The same reasoning is applied on the second channel by mixing $s_r(t)$ with the phase shifted transmitted signal $s_t^{90°}(t)$

$$s_t^{90°}(t) = \sin(2\pi f_0 t). \tag{3.16}$$

Similarly, the output of the synchronous detector II is computed:

$$s_r(t)\, s_t^{90°}(t) = \frac{a}{2} \sin[4\pi f_0 t + \Phi(t)] - \frac{a}{2} \sin \Phi(t). \tag{3.17}$$

After the low-pass filtering, the quadrature phase component $Q\left(t\right)$ is found:

$$Q\left(t\right)=-\frac{a}{2}\sin\Phi\left(t\right). \tag{3.18}$$

Combining the in-phase channel $I\left(t\right)$ and quadrature phase channel $Q\left(t\right)$, the complex Doppler signal can be formulated as:

$$s_D\left(t\right)=I\left(t\right)+jQ\left(t\right)=\frac{a}{2}\exp\left[-j\Phi\left(t\right)\right]=\frac{a}{2}\exp\left[-j2\pi f_D t\right]. \tag{3.19}$$

Thus, the Doppler frequency shift f_D can be estimated from the complex Doppler signal $s_D\left(t\right)$ by using frequency measurement tool. The simplest method is to use the Fast Fourier Transform (FFT), which is computationally efficient and easy to implement. However, its frequency resolution is limited to the reciprocal of the time interval of the signal and suffers from spectrum leakage associated with time windowing. The usual way to increase frequency resolution is to take FFT with a longer time duration of the analysed signal without zero padding. In the context of resolution, the type of the radar sensor is a key factor. Doppler radars include pure Continuous Wave (CW) radar, without modulation, Frequency Modulated Continuous Wave (FMCW) radar and coherent pulsed Doppler radar. Pure CW radar can only measure the velocity; FMCW and coherent pulsed Doppler radars can have wide frequency bandwidth to gain a high range resolution and measure both the range and Doppler information.

3.3 The Micro Doppler effect in radar

The micro Doppler effect was originally introduced in coherent laser [64, 67]. Due to the sensitivity of the phase of the returned signal to variation in range of the object, a half wavelength motion can cause a 360° phase change: for LADAR (LAser Detection And Ranging) system, with a wavelength of 2μm, a 1μm range variation can cause a 360° phase change. In case of a vibration, the maximum Doppler frequency variation in LOS is determined by:

$$\max\left\{f_D\right\}=\frac{2}{\lambda}D_v f_v, \tag{3.20}$$

where f_v is the vibration frequency and D_v is the amplitude of the vibration. As a consequence, in a high frequency system, even with a very low vibration frequency f_v, a very small displacement amplitude D_v can cause a large phase change, and thus Doppler shift can be easily detected.

The term “micro-motion” is referred to an object or any structural components of the object having oscillatory motions: in addition to the bulk motion of the object, any rotations or vibrations are called micro motions, introducing a frequency modulation on the radar received signal. For a pure periodic vibration or rotation, micro motions generate side band Doppler frequency shift about the center of the Doppler shifted carrier frequency. The modulation contains harmonic frequencies determined by the carrier frequency, the vibration or rotation rate, and the angle between the direction of vibration and the direction of incident wave. Thus, the kinematic properties of the object of interest can be determined by the frequency modulation, where the frequency shift coming from the micro-Doppler effect depends on the frequency band of the signal. In a radar system operating at microwave frequency bands, the micro Doppler may be observed if the product of the target’s oscillation rate and the displacement of the oscillation is high enough. For a X band radar, with 3cm of wavelength, a vibration rate of 15Hz and displacement of 0.3cm can induce a detectable maximum Doppler shift of 18.8Hz. An L band radar instead, to achieve the same micro Doppler shift, with same vibration rate of 15Hz, a displacement of 1cm is required.

Interest in radar based micro-Doppler signature analysis has grown in the last decade, reaching a plethora of sectors and applications. Worthwhile to mention the use of micro-Doppler in defence, biomedical and automotive fields [66]. In defence application, the micro-Doppler signature has been used to detect and classify targets. In [68–70], it has been used for automatic helicopters classification. In [71, 72] the method is exploited to discriminate birds and small unmanned aerial vehicles (UAVs) with emphasis on micro-Doppler that can be extracted from time frequency distributions. In [73, 74] a micro-Doppler based algorithm was presented for ballistic missile classification. In the bio-medical field, the micro-Doppler signature is used to estimate the total human energy expenditure for walking and running activities [75]. In [76], radar micro-doppler is also used for human activuty recognition. In automotive applications, micro-Doppler analysis has been used to classify pedestrian activity for Automatic Drive Assistant Systems (ADAS) [77, 78]. The interest in micro-Doppler suggests that this technology is reliable, and that it is worthwhile investigating it in further applications domains. An overview of the main time frequency analysis tools and the basic principle of the motion of objects in the context of radar signal processing are introduced.

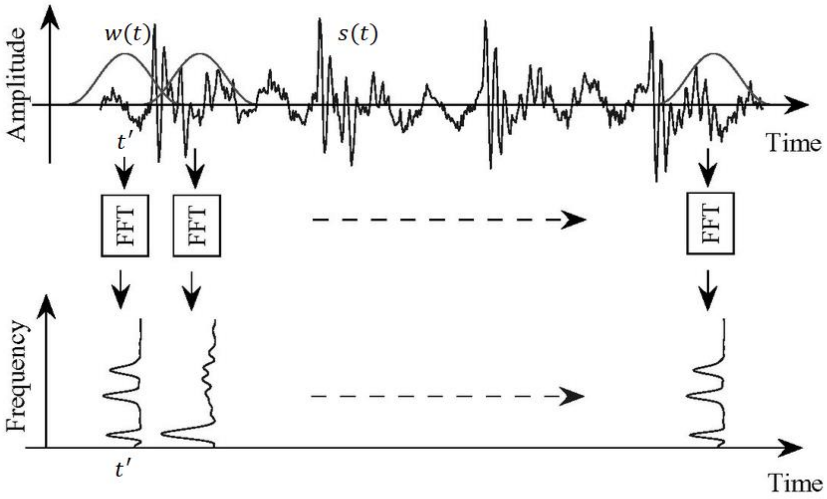


Fig. 3.3 Illustration of STFT processing on the signal $s(t)$.

3.3.1 Time Frequency Distributions

The micro Doppler shift is a time varying frequency shift that can be extracted from the complex output signal of a quadrature detector introduced in 3.2.1. For analysing time varying frequency features the traditional Fourier transform alone is not suitable because it can not provide time dependent spectral description. For this reason, the analysis of mD components into a received radar signal is generally conducted using more sophisticated tool, such as Time-Frequency Distributions (TFD). Specifically, TFD generates a 2-Dimensional (2D) representation in both time and frequency domains simultaneously, emphasizing the time-varying behaviour of the signal. A tool widely used to display time varying spectral density of a time varying signal $s(t)$ is the spectrogram. Defined as the squared magnitude of the Short-Time Fourier Transform (STFT), the spectrogram $\chi(\tau, f)$ is given by:

$$\chi(\tau, f) = |STFT\{s(t)\}|^2 = \left| \int_{-\infty}^{\infty} s(t)\, w(t-\tau) \exp^{-j2\pi ft} dt \right|^2, \tag{3.21}$$

where $w(t)$ is a time window function of choice. Differently from the conventional analysis in the Fourier domain obtained by applying the FT on the entire signal duration, the basic principle of STFT is the computation of FT onto shorter signal segments obtained by moving the window centre n along the signal time duration, as illustrated in Figure 3.3 [79]. In this way, the spectral analysis of the signal for different instants of time is provided. The time frequency resolution of the spectrogram is defined by the width of the windowing function, leading to the

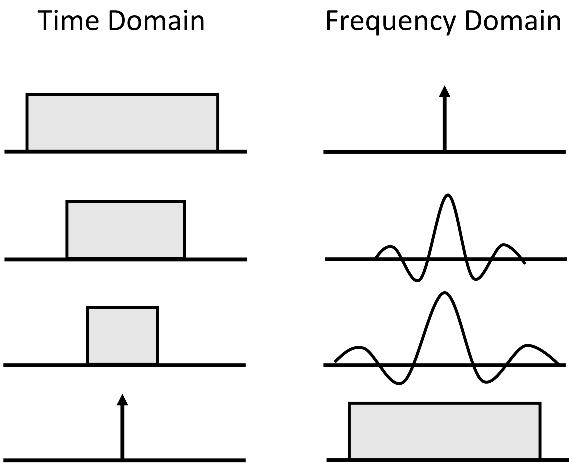


Fig. 3.4 Illustration of the Gabor uncertainly principle.

Gabor limit. Analogous to the Heisenberg uncertainty principle, it states that it is impossible to simultaneously identify a signal in both time and frequency domain. As illustrated in Figure 3.4, when a narrower window is used, a good time resolution is achieved, but with a poor frequency resolution. On the other hand, when a wide window is used, a good frequency resolution is achieved, but with a poor time resolution. This suggests that the error in measuring the frequency Δf is inversely related to the duration of the signal Δt, leading to:

$$\Delta f \Delta t \geqslant 1. \tag{3.22}$$

This is the Gabor uncertainty principle: the product of the uncertainties in frequency and time must exceed a fixed constant. The only case in which the time bandwidth product achieves the minimum value $\Delta f \Delta t = 1$ is when a Gaussian window is used:

$$w_G\left(t\right) = \frac{1}{\pi^{1/4}\sqrt{\sigma}} \exp^{-\frac{t^2}{2\sigma^2}} . \tag{3.23}$$

When the Gaussian window is used to compute the STFT, it is referred as Gabor Transform.

Whilst the windowing operation is useful to reduce the spectral leakage, some information can be lost due to the transition and the weighting. To reduce the information loss, overlap could be applied with higher computational cost. However, the overlap is limited since signal segments strongly correlated would not provide more information on the time variant spectral components [80]. To better analyse the time varying micro Doppler frequency characteristics and visualise the localised joint time and frequency information, the signal must be analysed by a high resolution TF transform, to characterise the spectral and temporal behaviour of the signal. Bilinear transforms, such as the Wigner-Ville Distribution (WVD), belongs to the high resolution TF transform. The WVD of a generic signal $s\left(t\right)$ is defined as:

$$WVD\left(t,f\right)=\int_{-\infty}^{\infty}s\left(t+\frac{\tau}{2}\right)s^{*}\left(t-\frac{\tau}{2}\right)\exp^{-j2\pi f\tau}d\tau. \tag{3.24}$$

Thus, the WVD can be interpreted as the FT of the time dependent autocorrelation function $s\left(t+\frac{\tau}{2}\right)s^{*}\left(t-\frac{\tau}{2}\right)$ of the signal $s\left(t\right)$. The main advantage of the WVD with respect to the more intuitive STFT is the absence of the trade-off among the time and frequency resolutions. However, due to the nature of bilinear functions, the WVD suffers of cross term interference. Considering two signals $s_1\left(t\right)$ and $s_2\left(t\right)$, the WVD of their sum is given by:

$$WVD_{s_1(t)+s_2(t)}\left(t,f\right)=WVD_{s_1(t)}\left(t,f\right)+WVD_{s_2(t)}\left(t,f\right)+2\mathbf{R}\left\{WVD_{s_1(t),s_2(t)}\left(t,f\right)\right\}, \tag{3.25}$$

where the term

$$WVD_{s_1(t),s_2(t)}\left(t,f\right)=\int_{-\infty}^{\infty}s_1\left(t+\frac{\tau}{2}\right)s_2^{*}\left(t-\frac{\tau}{2}\right)\exp^{-j2\pi f\tau}d\tau, \tag{3.26}$$

represents the cross term interference. Thus, if a signal contains more than one component in the joint time frequency domain, its WVD suffers of non-zero interference terms (cross-terms) that can affect the correct identification of the real signal components. To reduce the cross term interference, filtered WVDs have been used to preserve the useful property of the TFD with a slightly reduced time frequency resolution and a largely reduced cross term interference. The WVD with a linear low pass filter belongs to the Cohen class, generally defined as follow:

$$C\left(t,f\right)=\int\int s\left(u+\frac{\tau}{2}\right)s^{*}\left(u-\frac{\tau}{2}\right)\psi\left(t-u,\tau\right)\exp^{-j2\pi f\tau}dud\tau, \tag{3.27}$$

where the Fourier Transform of the low pass filter $\psi\left(t,\tau\right)$, denoted as $\Psi\left(\theta,\tau\right)$, is called Kernel function. With $\Psi\left(\theta,\tau\right)=1$, then $\psi\left(t,\tau\right)=\delta(t)$, the Cohen Class is reduced to the WVD. Over the years, different kernels have been developed, such as the Pseudo Wigner Ville Distribution (PWVD), defined by:

$$PWVD\left(t,f\right)=\int_{-\infty}^{\infty}w\left(\tau\right)s\left(t+\frac{\tau}{2}\right)s^{*}\left(t-\frac{\tau}{2}\right)\exp^{-j2\pi f\tau}d\tau. \tag{3.28}$$

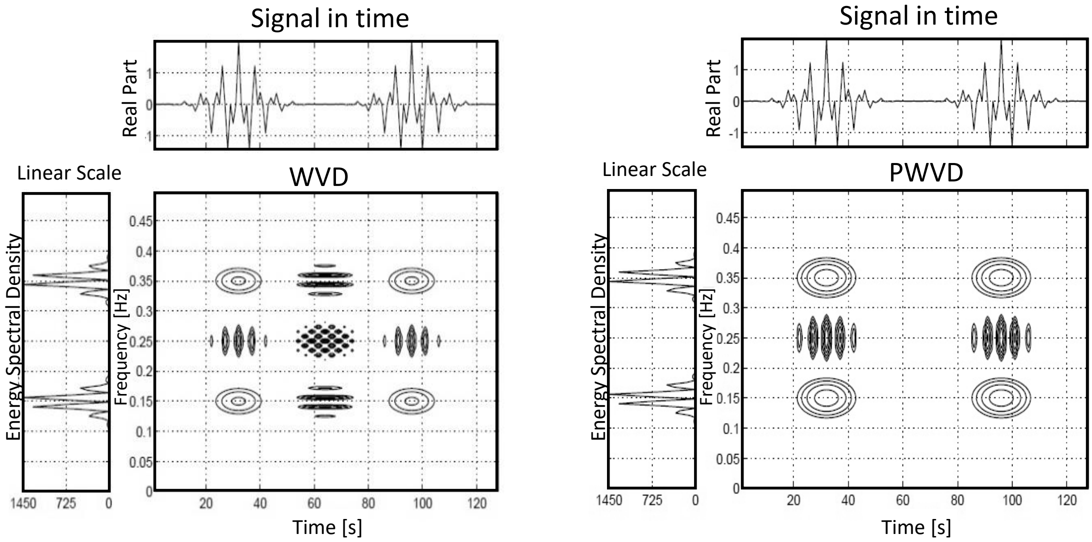


Fig. 3.5 Comparison between the Wigner Ville Distribution (WVD) and the Pseudo Wigner Ville distribution (PWVD).

In this case, a time window has been introduced. A comparison between WVD and PWVD is shown in Figure 3.5. Whilst the components caused by frequency oscillations are attenuated by the time window, the interference caused be the time oscillations are not. For this reason a second window can be introduced, in order to localise the signal in both time and frequency domain, still in accordance with Heisenberg principle. Thus, the Smoothed Pseudo Wigner Ville Distribution (SPWVD) is defined by:

$$SPWVD\left(t,f\right)=\int_{-\infty}^{\infty}w\left(\tau\right)\int_{-\infty}^{\infty}g\left(s-t\right)s\left(t+\frac{\tau}{2}\right)s^{*}\left(t-\frac{\tau}{2}\right)ds\exp^{-j2\pi f\tau}d\tau. \tag{3.29}$$

Thanks to the characteristic of its kernel with separated variables, such that $\Psi\left(\theta,\tau\right)=w\left(\tau\right)g\left(\theta\right)$, the SPWVD can be considered as the most versatile time frequency distribution, since it is possible to choose independently the filtering behaviour of the TFD over both time and frequency.

In conclusion, TFD are used to generate a 2D representation of signals, in both time and frequency domains simultaneously, emphasizing the time-varying behaviour of the signal. While the spectrogram is characterized by the trade-off between time and frequency resolution, the WVD is not. However, the WVD suffers of cross term interferences. These can be reduced using filtered WVDs, such as PWVD and SPWVD. Nevertheless, filtered WVDs has a limited use in real application because of the high computational complexity introduced by filtering process: in case of long observation time is required to collect a sufficient number of samples (e.g. targets with low oscillation rate or angular velocity), the computational burden increases massively making the spectrogram still the most used TFD.

3.3.2 Radar Cross Section of a Target

Electromagnetic scattering occurs when a target is illuminated by a radar transmitted electromagnetic (EM) waves. The incident wave induces electric and magnetic currents on the surface and within the volume of the target that will generate a scattered EM field which transmits waves in all the possible directions. If the target is at a distance far enough from the radar, the incident wavefront can be considered as a plane wave. The power of the scattered EM wave is measured by a bistatic scattering cross section of the target. If the direction is back to the radar, the bistatic scattering becomes backscattering and the cross section is a backscattering cross section, defined as Radar Cross section (RCS). According to the definition in [81], the RCS is described as a measure of the reflective strength of a target defined as 4π times the ratio of the power per unit solid angle scattered in a specified direction to the power per unit area in a plane wave incident on the scatterer from a specified direction. The RCS is formulated by:

$$\sigma = \lim_{r \to \infty} 4\pi r^2 \frac{E_s}{E_i}, \tag{3.30}$$

where E_s and E_i represent the intensity of the far field scattered and incident electric field, respectively, with r the distance between target and radar. It is normalised to the power density of the incident wave at the target in order to avoid any dependence on the distance. The RCS is dependent on the size, the geometry and material of the target, the frequency of the transmitter, the polarization of the transmitter and receiver and the aspect angle of the target with respect to the radar transmitter and receiver.

Three different scattering regions from the target are defined according to the radar wavelength and target dimensions:

- the Rayleigh region: it occurs when the wavelength is greater than the target dimension;
- the resonance region, even called Mie region: it occurs when the wavelength and target dimension are comparable;

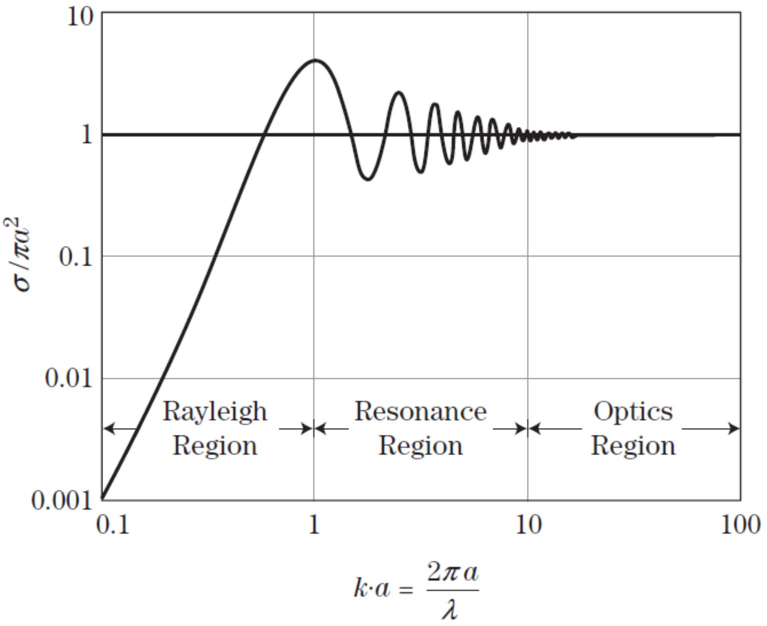


Fig. 3.6 Dependence of the RCS of sphere on wavelength [82].

- the optical region: it occurs when the wavelength is very small with respect to target sizes.

In Figure 3.6, the RCS of a sphere normalized with respect to the sphere area is shown, for different values of the ratio between the sphere radius and the EM wavelength. In the optical region, which corresponds to a large sphere compared to the wavelength, the RCS is a constant and can be simply expressed as:

$$\sigma_{sphere} = \pi a^2, \tag{3.31}$$

where a is the radius of the sphere greater than λ. In the Rayleigh region, the RCS is proportional to the waveform carrier frequency, which for a small sphere can be computed as:

$$\sigma_{sphere} = 9\pi a^2 \left(kr\right)^4, \tag{3.32}$$

where $k = 2\pi/\lambda$. Finally, in the resonance region, the RCS oscillates as a function of the carrier frequency, with a maximum value obtained for sphere radius equal to the wavelength.

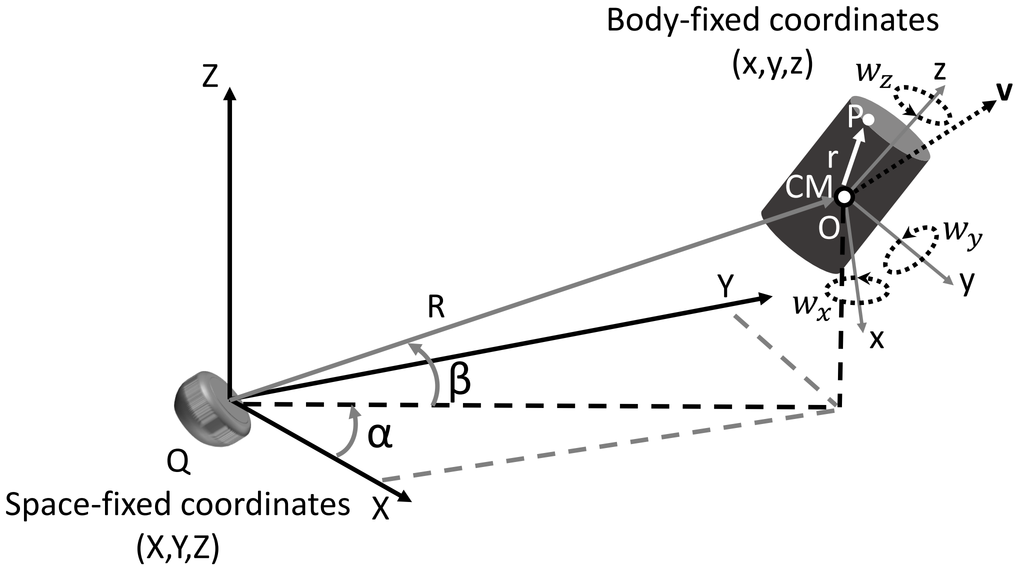


Fig. 3.7 Two coordinate systems are used to describe the motion of a rigid body: a space-fixed system (X, Y, Z) and a body-fixed system (x, y, z).

3.4 Rigid Body Motion

The micro Doppler effect induced by micro motion can be from a rigid or non rigid bodies. The former is defined as a solid body with a finite size, where the distance between any two particles does not vary with time. The mass of the rigid body is the sum of its particles and its general motion is a combination of translations and rotations of all the them. The latter is a deformable body. To compute the deformation, more complicated tools are needed, such as Finite Difference Method (FDM) and Finite Element Method (FEM). An alternative approach has been used in [64], where the non rigid body has been modelled by jointly connected rigid body segments. The motion of each segment is usually described by two coordinate systems, shown in Figure 3.7: the global or space-fixed system (X, Y, Z) and the local or body-fixed system (x, y, z). The range vector $\mathbf{R}$ is from the origin of the space fixed system to the origin of the body-fixed system, set in the center of mass (CM) of the body [83]. The orientation of the axes of the body-fixed system relative to the axes of the space-fixed system is given by three independent angles. Let r be the position of the body in the body-fixed system. Then its position in the space-fixed system is given by $\mathbf{r} + \mathbf{R}$, and its velocity is:

$$\mathbf{v} = \frac{d}{dt}(\mathbf{r} + \mathbf{R}) = \mathbf{V} + \boldsymbol{\Omega} \times \mathbf{r}, \tag{3.33}$$

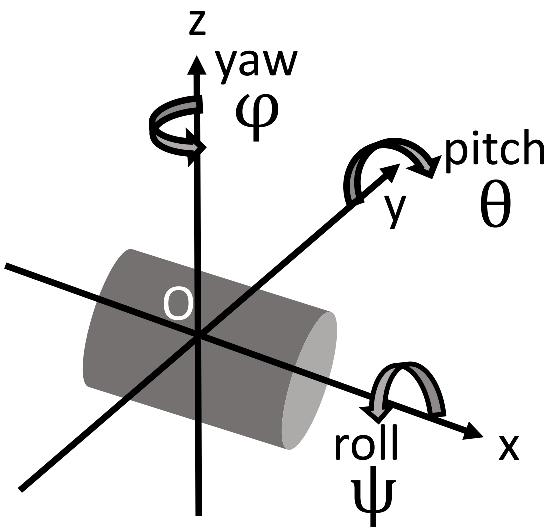


Fig. 3.8 The roll-pitch-yaw convention used to describe a flying aircraft.

where **V** is the translation velocity of the CM of the rigid body and $\mathbf{\Omega}$ is the angular velocity of the body rotation composed by the triad $(w_x, w_y, w_z)^T$.

3.4.1 Euler angles

To represent the rotation of an object, Euler angles and rotation matrices are commonly used. The rotation angles (φ, θ, ψ) are called Euler Angles, where φ is defined as the counter clockwise rotation around the z axis, θ is defined as the counter clockwise rotation around the y axis, and ψ is defined as the counter clockwise rotation around the x axis. Thus, the Euler angles are used to represent three successive rotation in a given rotation sequences. Different conventions exists, related to different successive rotation sequence. For example, the most common in aerospace engineering is the roll-pitch-yaw convention, or x-y-z sequence, show in Figure 3.8. To describe a flying aircraft three rotation angles about the x, y and z axis are used with the rotations in the roll-pitch-yaw (ψ, θ, φ) sequence. The pitch angle, or attitude, is defined by the rotation θ between $-\pi/2$ and $\pi/2$ about the y axis. The roll angle, or bank, is defined by the rotation ψ between $-\pi$ and π about the x axis. The yaw angle, or heading, is defined by the rotation φ between $-\pi$ and π about the z axis. Another commonly used rotation sequence is called x-convention. It follows the z-x-z sequence: first rotation by an angle about the z axis, the second rotation by an angle about the x axis and the third rotation by an angle around the z axis again. Thus, given a specific rotation sequence, the rotation matrix is an useful tool to compute rigid body rotations. According to Euler's rotation theorem, if the rotations are written in terms of elemental rotation matrices **D**, **C** and **B**, then a general rotation **A** can be written as:

$$\mathbf{A} = \mathbf{BCD}. \tag{3.34}$$

For the roll-pitch-yaw convention, the first step is rotating about the x axis $x = [1\,0\,0]^T$ by the angle ψ defined by the elemental rotation matrix $\mathcal{R}_X$:

$$\mathcal{R}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{bmatrix} \tag{3.35}$$

The second step is rotating about the new y axis $y_1 = [0 \cos\psi \; \sin\psi]^T$ by the angle θ defined by the elemental rotation matrix $\mathcal{R}_Y$:

$$\mathcal{R}_Y = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \tag{3.36}$$

The third step is rotating about the new z axis $z_2 = [\sin\theta \;\; -\cos\theta\sin\psi \;\; \cos\theta\cos\psi]^T$ by the angle φ defined by the elemental rotation matrix $\mathcal{R}_Z$:

$$\mathcal{R}_Z = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.37}$$

Thus, the general rotation matrix of the roll-pitch-yaw sequence $\mathbf{A} = \mathcal{R}_{X-Y-Z}$ can be written as:

$$\mathcal{R}_{X-Y-Z} = \mathcal{R}_Z \cdot (\mathcal{R}_Y \cdot \mathcal{R}_X) = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{3.38}$$

where the components of the rotation matrix $\mathcal{R}_{X-Y-Z}$ are:

$$\begin{cases} r_{11} = \cos\varphi\cos\theta \\ r_{12} = \sin\psi\sin\theta\cos\varphi + \cos\psi\sin\varphi \\ r_{13} = -\cos\psi\sin\theta\cos\varphi + \sin\psi\sin\varphi \\ r_{21} = -\cos\theta\sin\varphi \\ r_{22} = -sin\psi\sin\theta\sin\varphi + \cos\psi\cos\varphi \\ r_{23} = \cos\psi\sin\theta\sin\varphi + \sin\psi\cos\varphi \\ r_{31} = \sin\theta \\ r_{32} = -\sin\psi\cos\theta \\ r_{33} = \cos\psi\cos\theta \end{cases} \tag{3.39}$$

In the event the x-sequence is preferred, the general rotation matrix $\mathbf{A} = \mathcal{R}_{Z-X-Z}$ can be written as:

$$\mathcal{R}_{Z-X-Z} = \mathcal{R}_Z \cdot (\mathcal{R}_X \cdot \mathcal{R}_Z) = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{3.40}$$

where, this time, the components of the rotation matrix $\mathcal{R}_{Z-X-Z}$ are:

$$\begin{cases} r_{11} = -\sin\varphi\cos\theta\sin\psi + \cos\varphi\sin\psi \\ r_{12} = -\cos\varphi\cos\theta\sin\psi - \sin\varphi\sin\psi \\ r_{13} = \sin\theta\sin\psi \\ r_{21} = \sin\varphi\cos\theta\cos\psi + \cos\varphi\sin\psi \\ r_{22} = \cos\varphi\cos\theta\cos\psi - \sin\varphi\sin\psi \\ r_{23} = -\sin\theta\cos\psi \\ r_{31} = \sin\varphi\sin\theta \\ r_{32} = \cos\varphi\sin\theta \\ r_{33} = \cos\theta \end{cases} \tag{3.41}$$

In general, the rotation matrix $\mathcal{R}$ is a 3-by-3 matrix and must satisfy two conditions:

$$\begin{cases} \mathcal{R}^T\mathcal{R} = \mathbf{I} \\ \det\mathcal{R} = 1 \end{cases} \tag{3.42}$$

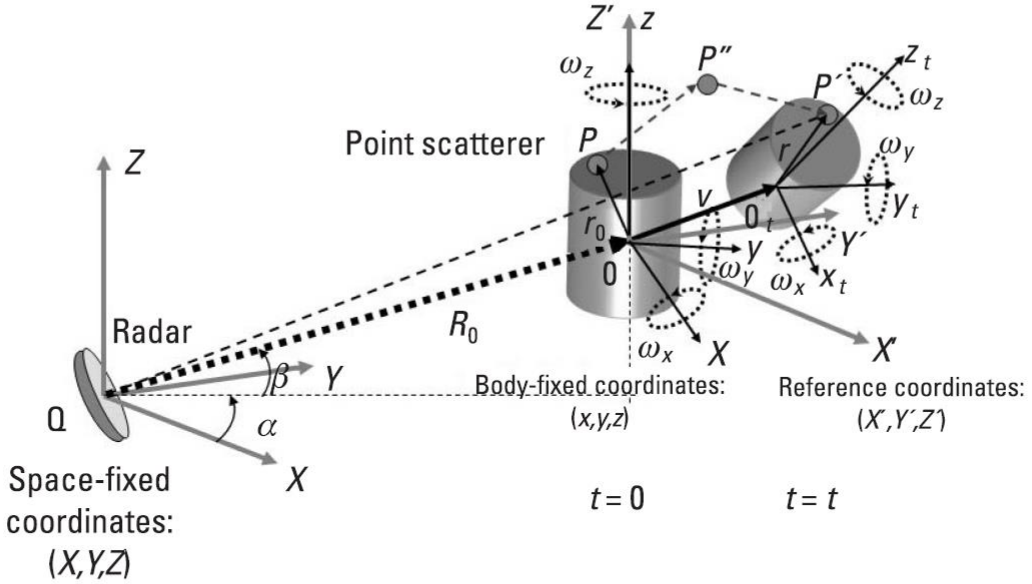


Fig. 3.9 Geometry of the radar and a target with translation and rotation [64].

Thus, if the product of the rotation matrix with its transposed matrix is a 3-by-3 unit matrix **I**, and the determinant of $\mathcal{R}$ is equal to 1, it means that the three column vectors of $\mathcal{R}$ must be orthonormal.

3.5 Micro-Doppler induced by target with micro motion

Introducing the micro motion of the target in the conventional Doppler analysis, it is possible to derive the mathematics of the micro Doppler [67]. In order to describe the effect of the translation and rotation of a target with respect to the radar, three coordinates systems are considered. Considering a stationary radar located at the origin of the space fixed coordinates system (X, Y, Z), as in Figure 3.9, the motion of a target can be described thanks to the local coordinates system (x, y, z), attached to the target such that the z axis corresponds with the symmetry axis of target [83]. To observe the target's rotations, a reference coordinates system (X', Y', Z') is introduced, which is parallel to the space fixed one and whose origin is shared with the target local coordinates system. Let P be a generic target's point scatterer located in $\mathbf{r}_0 = (X_0, Y_0, Z_0)$ at time $t = 0$. Considering a movement composed by two steps:

- a translation from P to P'', with translation velocity $\mathbf{v}$ with respect to the radar;
- a rotation from P'' to P' with angular velocity $\mathbf{w}$, which can be either represented in the body fixed coordinate system as $\mathbf{w} = (w_x, w_y, w_z)$ or represented in the space fixed coordinate system as $\mathbf{w} = (w_X, w_Y, w_Z)$.

It is possible to define the range dynamic with respect to the space fixed coordinates system of the particle P by using the time varying rotation matrix $\mathcal{R}_t$, such that:

$$r(t) = \| \mathbf{R}_0 + \mathbf{v}t + \mathcal{R}_t \mathbf{r}_0 \|, \tag{3.43}$$

where $\| \cdot \|$ represents the Euclidean norm, $\mathbf{R}_0$ the initial distance between the target and the radar at $t = 0$ and $r(t)$ the scalar range. If the radar transmits a sinusoidal waveform with carrier frequency f_0, the baseband received radar signal $s_r(t)$ from the point scatterer P is a function of $r(t)$:

$$s_r(t) = \sigma(x, y, z) \exp\left\{ j 2\pi f_0 \frac{2r(t)}{c} \right\} = \sigma(x, y, z) \exp\left\{ j\Phi\left[r(t)\right] \right\}, \tag{3.44}$$

where $\sigma(x, y, z)$ is the reflectivity function of the point scatterer P described in the body fixed coordinate system, c is the propagation speed of the electromagnetic wave and $\Phi[r(t)]$ the phase of the baseband signal, defined as:

$$\Phi[r(t)] = 2\pi f_0 \frac{2r(t)}{c}. \tag{3.45}$$

As in the standard Doppler analysis, by taking the derivative of the phase term, the Doppler frequency shift induced by the target's motion can be derived:

$$\begin{aligned} f_D &= \frac{1}{2\pi} \frac{d\Phi(t)}{dt} = \frac{2f_0}{c} \frac{d}{dt} r(t) = \\ &= \frac{2f_0}{c} \frac{1}{2r(t)} \frac{d}{dt} \left[(\mathbf{R}_0 + \mathbf{v}t + \mathcal{R}_t \mathbf{r}_0)^T (\mathbf{R} + \mathbf{v}t + \mathcal{R}_t \mathbf{r}) \right] = \\ &= \frac{2f_0}{c} \left[\mathbf{v} + \frac{d}{dt} (\mathcal{R}_t \mathbf{r}_0) \right]^T \mathbf{n}, \end{aligned} \tag{3.46}$$

where $\mathbf{n} = \frac{\mathbf{R}_0 + \mathbf{v}t + \mathcal{R}_t \mathbf{r}_0}{\|\mathbf{R}_0 + \mathbf{v}t + \mathcal{R}_t \mathbf{r}_0\|}$ is the unit vector from the origin of the space fixed coordinate system to the final position of the target. The Doppler shift introduced by rotation can be derived in easier way by introducing the relationship:

$$\mathbf{u} \times \mathbf{r} = \hat{\mathbf{u}} \mathbf{r}. \tag{3.47}$$

Given a vector $\mathbf{u} = (u_x, u_y, u_z)$ and a skew symmetric matrix $\hat{\mathbf{u}}$ defined as:

$$\hat{\mathbf{u}} = \begin{bmatrix} 0 & -u_z & u_y \\ u_z & 0 & -u_x \\ -u_y & u_x & 0 \end{bmatrix} \tag{3.48}$$

the cross product of the vector $\mathbf{u}$ and any vector $\mathbf{r}$ can be computed through the matrix computation:

$$\mathbf{u} \times \mathbf{r} = \begin{bmatrix} u_y r_z - u_z r_y \\ u_z r_x - u_x r_z \\ u_x r_y - u_y r_x \end{bmatrix} = \begin{bmatrix} 0 & -u_z & u_y \\ u_z & 0 & -u_x \\ -u_y & u_x & 0 \end{bmatrix} \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} = \hat{\mathbf{u}}\mathbf{r} \tag{3.49}$$

In the reference coordinate system, the angular rotation velocity can be described by $\mathbf{w} = (w_X, w_Y, w_Z)^T$, and the target will rotate along the unit vector $\mathbf{w}' = \mathbf{w} / \parallel \mathbf{w} \parallel$ with a scalar angular velocity $\Omega = \parallel \mathbf{w} \parallel$. By assuming a relatively low angular velocity compared to the radar sampling frequency, it is possible to write [64]:

$$\mathcal{R}_t = e^{\hat{\mathbf{w}}t}, \tag{3.50}$$

where $\hat{\mathbf{w}}$ is the skew symmetric matrix associated with $\mathbf{w}$, such that $\hat{\mathbf{w}}\mathbf{r} = \mathbf{w} \times \mathbf{r}$. Thus, the Doppler frequency shift in (3.46) becomes:

$$\begin{aligned} f_D &= \frac{2f_0}{c} \left[\mathbf{v} + \frac{d}{dt} (\mathcal{R}_t \mathbf{r}_0) \right]^T \mathbf{n} = \frac{2f_0}{c} \left[\mathbf{v} + \frac{d}{dt} \left(e^{\hat{\mathbf{w}}t} \mathbf{r}_0 \right) \right]^T \mathbf{n} = \\ &= \frac{2f_0}{c} \left[\mathbf{v} + \hat{\mathbf{w}} e^{\hat{\mathbf{w}}t} \mathbf{r}_0 \right]^T \mathbf{n} = \frac{2f_0}{c} \left[\mathbf{v} + \hat{\mathbf{w}}\mathbf{r} \right]^T \mathbf{n} = \frac{2f_0}{c} \left[\mathbf{v} + \mathbf{w} \times \mathbf{r} \right]^T \mathbf{n}. \end{aligned} \tag{3.51}$$

In the event $\mathbf{R}_0 \gg \parallel \mathbf{v}t + \mathcal{R}_t \mathbf{r}_0 \parallel$ the unit vector $\mathbf{n}$ can be approximated as $\mathbf{n} = \frac{\mathbf{R}_0}{\|\mathbf{R}_0\|}$, which is the direction of the radar LOS. This condition can be considered valid in most of the radar scenario, since the displacement is usually smaller than the range of the target.

Consequently, the Doppler frequency shift can be written as sum of two component:

$$f_D = \frac{2f_0}{c} \left[\mathbf{v} + \mathbf{w} \times \mathbf{r} \right] \cdot \mathbf{n}, \tag{3.52}$$

where the first term

$$f_{trans} = \frac{2f_0}{c} \mathbf{v} \cdot \mathbf{n}, \tag{3.53}$$

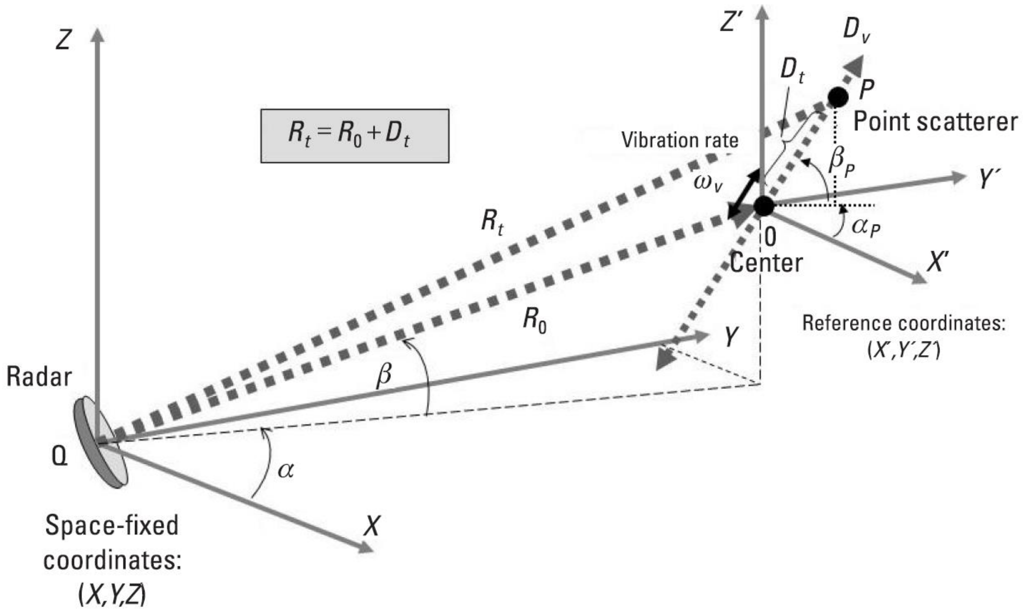


Fig. 3.10 Geometry of the radar and a vibrating point target [64].

is the main Doppler shift due to the translation and the second term

$$f_{mD} = \frac{2f_0}{c} [\mathbf{w} \times \mathbf{r}] \cdot \mathbf{n}. \tag{3.54}$$

is the micro-Doppler due to the rotation. In this way, the main Doppler and the micro-Doppler can be treated separately.

3.5.1 Micro Doppler of Vibrating point

In Figure 3.10, the geometry of a generic vibrating point target is shown, with the radar located at the origin of the space fixed coordinates system (X, Y, Z). The point scatterer P vibrates around the center point O, which is the origin of the reference coordinates system (X', Y', Z') placed at a distance R_0 from the radar. Assuming a stationary center point O, it is located in the radar coordinates system at the coordinates:

$$(R_0 \cos\beta \cos\alpha, R_0 \cos\beta \sin\alpha, R_0 \sin\beta), \tag{3.55}$$

with α and β the azimuth and elevation angle, respectively, of the point O respect to the radar. Then, the direction of the radar LOS $\mathbf{n}$ in (3.52) becomes:

$$\mathbf{n} = [\cos\beta \cos\alpha, \cos\beta \sin\alpha \sin\beta]^T. \tag{3.56}$$

At the instant the scatterer P starts vibrating at the frequency f_v with a displacement amplitude D_v, the range vector from the radar to the scatter point becomes:

$$\mathbf{R}_t = \mathbf{R}_0 + \mathbf{D}_t. \tag{3.57}$$

Thus, the scalar range can be expressed as:

$$\begin{aligned} R_t = \mid \mathbf{R}_t \mid = \Big[& \left(R_0 \cos\beta \cos\alpha + D_t \cos\beta_p \cos\alpha_p\right)^2 + \\ & + \left(R_0 \cos\beta \sin\alpha + D_t \cos\beta_p \sin\alpha_p\right)^2 + \left(R_0 \sin\beta + D_t \sin\beta_p\right)^2 \Big]^{\frac{1}{2}} , \end{aligned} \tag{3.58}$$

where α_p and β_p are the azimuth and elevation angles in the reference coordinates system. If the azimuth angle α and the elevation angle β_p of the scatterer point P are all zero and $R_0 \gg D_t$, the scalar range is approximated as:

$$R_t = \left(R_0^2 + D_t^2 + 2R_0 D_t \cos\beta \cos\alpha_p\right)^{\frac{1}{2}} \cong R_0 + D_t \cos\beta \cos\alpha_p. \tag{3.59}$$

Since the point scatterer vibrates with angular frequency ω_v with maximum displacement amplitude D_v, then $D_t = D_v \sin\omega_v t$. Thus, the scalar range can be expressed as:

$$R\left(t\right) = R_t = R_0 + D_v \sin\omega_v t \cos\beta \cos\alpha_p. \tag{3.60}$$

The baseband received radar signal $s_r\left(t\right)$ from the point scatterer becomes:

$$s_r\left(t\right) = \sigma \exp\left\{ j \left[2\pi f_0 t + 4\pi \frac{R\left(t\right)}{\lambda} \right] \right\} = \sigma \exp\left\{ j \left[2\pi f_0 t + \Phi\left(t\right) \right] \right\}, \tag{3.61}$$

where σ is the reflectivity of the the point scatterer, f_0 the carrier frequency of the transmitted signal, λ the wavelength and $\Phi\left(t\right) = 4\pi R\left(t\right)/\lambda$ the phase function. Defining $B = \left(4\pi/\lambda\right) D_v \cos\beta \cos\alpha_p$, the equation in (3.61) can be rewritten as:

$$s_r\left(t\right) = \sigma \exp\left\{ j \frac{4\pi R_0}{\lambda} \right\} \exp\left\{ j \left[2\pi f_0 t + B \sin\omega_v t \right] \right\}. \tag{3.62}$$

Defining the received radar signal as in (3.62), an analogy with the Bessel function is found. Let $J_k\left(B\right)$ be a Bessel function of the first kind of order k

$$J_k\left(B\right) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \exp\left\{ j \left(B \sin u - ku\right) \right\} du, \tag{3.63}$$

the baseband received radar signal can be expressed as a sum of Bessel function, such that:

$$\begin{aligned} s_r(t) = \sigma \exp\left\{j\frac{4\pi R_0}{\lambda}\right\} \sum_{k=-\infty}^{+\infty} J_k(B) \exp\left[j\left(2\pi f_0 + k\omega_v\right)t\right] = \\ = \sigma \exp\left\{j\frac{4\pi R_0}{\lambda}\right\} \Big\{J_0(B)\exp\left[j2\pi f_0 t\right] + J_1(B)\exp\left[j\left(2\pi f_0 + \omega_v\right)t\right] + \\ - J_1(B)\exp\left[j\left(2\pi f_0 - \omega_v\right)t\right] + J_2(B)\exp\left[j\left(2\pi f_0 + 2\omega_v\right)t\right] + \\ - J_2(B)\exp\left[j\left(2\pi f_0 - 2\omega_v\right)t\right] + \ldots\Big\}. \end{aligned} \tag{3.64}$$

Therefore, the micro-Doppler frequency spectrum consists of a pair of spectral lines around the center frequency f_0 and with spacing $\omega_v/(2\pi)$ between adjacent lines. As in (3.53), the micro-Doppler shift induced by the vibration is:

$$f_{mD} = \frac{2f_0}{c}\left(\mathbf{v}^T \cdot \mathbf{n}\right) = \frac{2f_0 f_v D_v}{c}\left[\cos\left(\alpha - \alpha_p\right)\cos\beta\cos\beta_p + \sin\beta\sin\beta_p\right]\cos\left(2\pi f_v t\right). \tag{3.65}$$

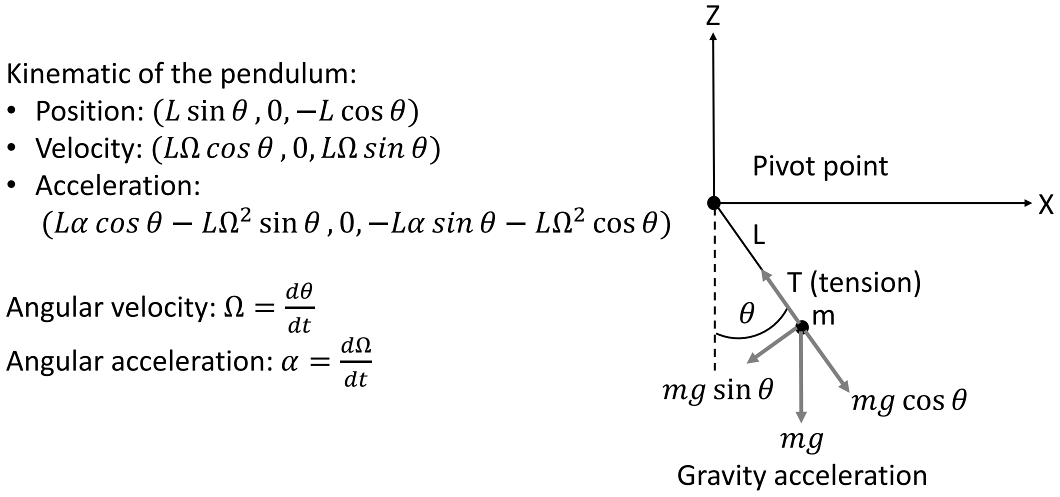
In the event the azimuth angle α and elevation angles β_p are both zero, then:

$$f_{mD} = \frac{2f_0 f_v D_v}{c}\cos\beta\cos\alpha_p\cos\left(2\pi f_v t\right). \tag{3.66}$$

When the direction of the vibration is along the projection of the direction of radar LOS, or $\alpha_p = 0$, and even the elevation angle β is null, the micro-Doppler achieves its maximum value of $2f_0 f_v D_v/c$.

3.5.2 Micro Doppler of a Pendulum

In this subsection, the classic example of an oscillation of a pendulum is considered in order to provide a better understanding of non linear motion dynamic effect on radar. A simple pendulum is modelled by a weighted small bob, attached to a pivot point through a weighted string, as shown in Figure 3.11. Under the influence of the gravity, the small bob swings back and forth along the $y - axis$. In the stable equilibrium position, the centre of mass of the pendulum is located at the coordinates $(x = 0, y = 0, z = L)$, where L is the length of the string. As Newton's law states, the total force acting on the pendulum is equal to the product between its mass and acceleration. If the position of the mass deviates from its equilibrium position of a swinging angle θ, two forces will act of the pendulum: the downward gravitational force, mg, and the tension T in the string. However, the



Kinematic of the pendulum:

- Position: $(L\sin\theta, 0, -L\cos\theta)$
- Velocity: $(L\Omega\, cos\,\theta, 0, L\Omega\, sin\,\theta)$
- Acceleration: $(L\alpha\, cos\,\theta - L\Omega^2\sin\theta, 0, -L\alpha\, sin\,\theta - L\Omega^2\cos\theta)$

Angular velocity: $\Omega = \frac{d\theta}{dt}$

Angular acceleration: $\alpha = \frac{d\Omega}{dt}$

Fig. 3.11 Kinematic of a simple pendulum.

tension has no contribution to the torque since the line of action goes through the pivot point. According to Newton's second law of motion, the angular equation of motion can be found by the applied torque τ on the mass m:

$$\tau = I\frac{d^2\theta}{dt^2}, \tag{3.67}$$

where I is moment of inertia defined as

$$I = mL^2, \tag{3.68}$$

and $(d^2\theta/dt^2)$ is the angular acceleration. The vector torque $\boldsymbol{\tau}$ is the cross product between the position vector $\mathbf{L}$ and the gravitational force vector $m\mathbf{g}$, such that $\boldsymbol{\tau} = \mathbf{L}\times m\mathbf{g}$. The magnitude of the torque is then

$$\tau = Lmg\sin\theta. \tag{3.69}$$

Thus, substituting (3.68) and (3.69) in (3.67), the equation of the pendulum is found:

$$mL\frac{d^2\theta}{dt^2} = -mg\sin\theta. \tag{3.70}$$

In (3.70), the relationship between the swing angle θ and its second time derivative is defined. Denoting the angular velocity $\Omega = d\theta/dt$, the equation of the pendulum can be rewritten as a set of two first order differential equations:

$$\begin{cases} \frac{d\theta}{dt} = \Omega \\ \frac{d\Omega}{dt} = -\frac{g}{L}\sin\theta \end{cases} \tag{3.71}$$

For a small angle, $\sin\theta$ can be substituted by θ and the pendulum becomes a linear oscillator; thus, the differential equation can be approximated as:

$$\frac{d^2\theta}{dt^2} + \omega_0^2\theta \cong 0, \tag{3.72}$$

where $\omega_0 = (g/L)^{1/2}$ is the angular frequency of the oscillating pendulum. Then the swinging angle, solution of the harmonic equation, is:

$$\theta\left(t\right) = \theta_0 \sin\omega_0 t, \tag{3.73}$$

with angular velocity:

$$\Omega\left(t\right) = \frac{d\theta\left(t\right)}{dt} = \theta_0\omega_0\cos\omega_0 t, \tag{3.74}$$

where θ_0 is the initial swinging angle of the pendulum, or initial amplitude.
For a given small initial amplitude θ_0, the period of the oscillating pendulum is determined by:

$$T_0 = 2\pi\frac{1}{\omega_0} = 2\pi\sqrt{\frac{L}{g}}, \tag{3.75}$$

with the inverse of the period T_0 be the frequency of the oscillation $f_0 = 1/T_0$. In a real scenario, however, a damping factor has to be considered also. Due to the linear friction, the extra term $-2\gamma d\theta/dt$ must be added on the right side in (3.70), leading to the new pendulum equation as:

$$\frac{d^2\theta}{dt^2} + 2\gamma\frac{d\theta}{dt} + \omega_0^2\sin\theta = 0, \tag{3.76}$$

where $\omega_0 = (g/L)^{1/2}$ is still the angular frequency of the free oscillations, and γ is the damping constant. Thus, the equation of pendulum can be rewritten as a set of two first order differential equations:

$$\begin{cases} \frac{d\theta}{dt} = \Omega \\ \frac{d\Omega}{dt} + 2\gamma\Omega = -\frac{g}{L}\sin\theta. \end{cases} \tag{3.77}$$

For a small angle, $\sin\theta \approx \theta$ and the pendulum equation is approximated as:

$$\frac{d^2\theta}{dt^2} + 2\gamma\frac{d\theta}{dt} + \omega_0^2\theta \cong 0. \tag{3.78}$$

If the friction is weak, such that $\gamma < \omega_0$, the solution of (3.78) is:

$$\theta\left(t\right) = \theta_0 e^{-\gamma t}\cos\left(\omega t + \Phi_0\right), \tag{3.79}$$

where θ_0 is the initial amplitude, Φ_0 the initial phase depending on the initial excitation and the exponential $\exp\left(-\gamma t\right)$ is a decreasing factor depending by the damping constant. The angular frequency of the oscillation ω is given by $\omega = \sqrt{\omega^2 + \gamma^2} = \omega_0\sqrt{1 - \left(\gamma/\omega_0\right)^2}$. When $\gamma < \omega_0$, the angular frequency of the oscillation and period can be approximated as:

$$\begin{aligned} \omega \approx& \omega_0 - \frac{\gamma^2}{2\omega_0}, \\ T \approx& T_0\left[1 + \frac{\gamma^2}{2\omega_0^2}\right], \end{aligned} \tag{3.80}$$

which both are close to the free oscillation frequency ω_0 and period T_0.

If a driving force is added in the pendulum oscillation, a further term must be added in the (3.78) becoming:

$$\frac{d^2\theta}{dt^2} + 2\gamma\frac{d\theta}{dt} + \frac{g}{L}\sin\theta = \frac{A_{Dr}}{mL}\cos\left(2\pi f_{Dr}t\right), \tag{3.81}$$

where γ is the damping constant, A_{Dr} is the amplitude of the driving force, and f_{Dr} is the driving frequency. Thus, in case of linear friction and driving force, the set of the two first order differential equations can be rewritten as:

$$\begin{cases} \frac{d\theta}{dt} = \Omega \\ \frac{d\Omega}{dt} + 2\gamma\Omega = -\frac{g}{L}\sin\theta + \frac{A_{Dr}}{mL}\cos\left(2\pi f_{Dr}t\right). \end{cases} \tag{3.82}$$

To calculate radar backscattering from an oscillating pendulum, the ordinary differential equations reported above are used for solving the swinging angle and the angular velocity: at each time instant during the radar observation time the position of the pendulum can be determined. Based on the location and orientation of the pendulum, the RCS of the pendulum and the radar received signal can be calculated.

If the radar transmits a sequence of rectangular pulses at the carrier frequency f_c,

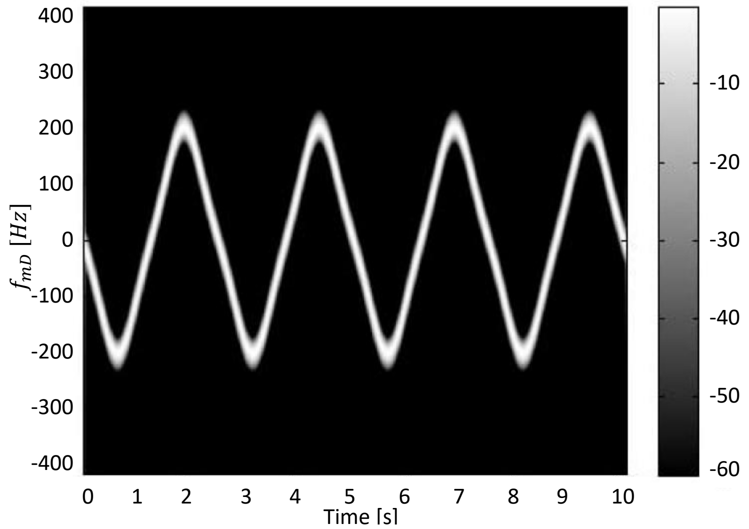


Fig. 3.12 Micro-Doppler signature of the simple free oscillating pendulum [64].

with pulse width Δ and pulse repetition interval Δt, the received baseband signal is:

$$s_r(t) = \sum_{k=1}^{n_p} \sqrt{\sigma_P(t)}\Pi\left\{t - k\Delta T - \frac{2R_P(t)}{c}\right\}\exp\left\{-j2\pi f_c \frac{2R_P(t)}{c}\right\}, \tag{3.83}$$

where n_p is the total number of received pulses, $R_P(t)$ is the distance from the radar to the bob at the time t and Π is the rectangular function defined as:

$$\Pi(t) = \begin{cases} 1 & 0 \leq t \leq \Delta \\ 0 & \text{otherwise} \end{cases} \tag{3.84}$$

The RCS $\sigma_P(t)$ of the small bob at time t is simulated by the point scatterer model since it can be seen as a point scatterer. Depending on its shape and geometry, the RCS $\sigma_P(t)$ can be defined. In case of a spherical shape, for example, the RCS can be expressed by mathematical formulation in (3.31) and (3.32), depending on the transmitted wavelength. Thus, equations (3.71), (3.77) and (3.82) are used to calculate the oscillating angle and angular velocity of simple pendulum, damping pendulum and damping pendulum with driving force, respectively. The joint time-frequency distribution are used to extract the micro Doppler signatures. The STFT is used in Figure 3.12 to show the micro-Doppler signature of a simple pendulum, where an oscillating frequency of 0.4Hz can be measured. Compared with the micro-Doppler signature of a simple pendulum, Figure 3.13 shows the micro-Doppler signature of a damping pendulum with $L = 1.5$m, $m = 20$g,

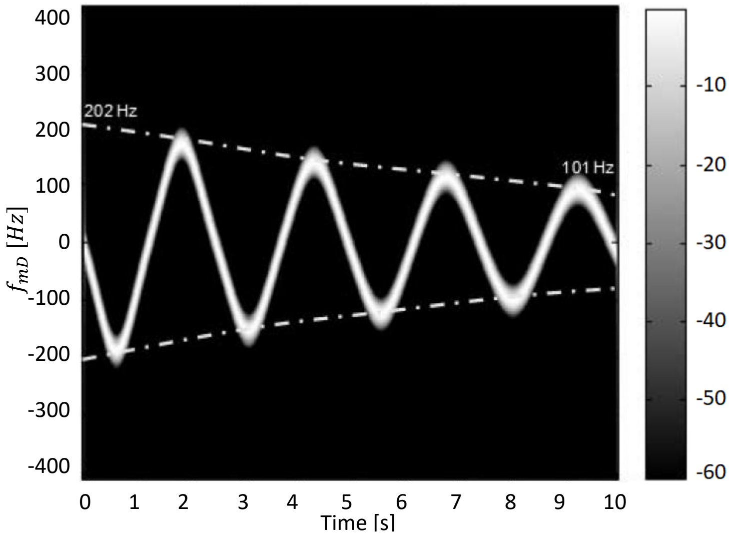


Fig. 3.13 Micro-Doppler signature of a damping oscillating pendulum [64].

$\gamma = 0.07$, $A_{Dr} = 15$ and $f_{Dr} = 0.2$Hz. In case a driving force is applied on the damping pendulum, the resultant effect are visible in the Figure 3.14. From the micro-Doppler signature in Figures 3.13 and 3.14, an oscillating frequency of 0.4Hz can be measured and the damping constant γ is measured from the change of the amplitude of the Doppler modulation during the observation time. During a 10 seconds time interval, the amplitude of the Doppler modulation changes from 202Hz to 101Hz, with the damping constant estimated as:

$$\gamma = -\ln\left(101/202\right)/10 = 0.069, \tag{3.85}$$

which is consistent with the damping constant of 0.07 used for the simulation.

3.6 Loudspeaker kinematic

With the aim to exploit the concept of radar micro-Doppler for condition monitoring of loudspeakers, it is important to understand the kinematic of an acoustic transducer. A magnetic type transducer is a device able to convert an electrical signal into sounds. Belonging to this class of devices are the electro-dynamic or moving coil loudspeaker. Among different type of transducers [84], each one relying on different working principles, in this work only direct radiator loudspeaker type is considered, where a cross sectional view is shown in Figure 3.15. The cone or diaphragm is made from a suitably light and stiff material; most of its stiffness comes from its profile. The profile can be designed as a straight line (a real cone)

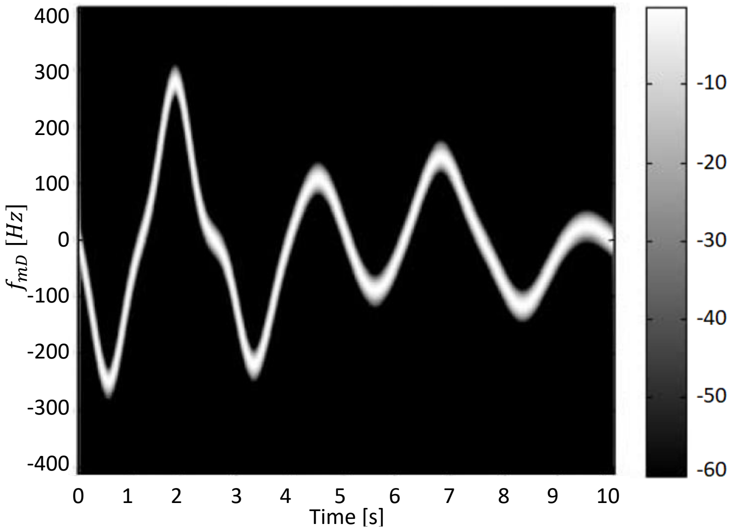


Fig. 3.14 Micro-Doppler signature of a damping and driving oscillating pendulum [64].

or curved. In order to prevent metallic dust falling into the magnetic gap a dust cap is placed in the centre; the dust cap is also useful to prevent sound from the back of the diaphragm to leak through the outer world. The coil is located in the gap of a magnetic path, comprising a pole piece and top plate, where the magnetic flux is produced by a permanent magnet, which is held in place by a basket structure. A surround and a spider are used to support the diaphragm at the rim and near the voice coil, respectively, so that it is free to move only in an axial direction. In general, sound from the back of the cone exits through holes in the basket, while sound from the back of the dust cap leaks through the magnetic gap and spider, which often presents holes or porosity, before exiting through the basket. When an audio signal is applied to the voice coil, the resulting current creates a magnetomotive force which interacts with the air-gap flux of the permanent magnet and causes a translatory movement of the voice coil and, hence, of the cone to which it is attached.

Sound waves are produced by the motion of the cone that displaces the air molecules at its surface. The loudness of the sound is, therefore, dependant on the acoustic pressure radiated by the membrane, proportional to the velocity, by which the cone moves and pushes the surrounding air [84]. The most widely used models for loudspeakers dynamics, assume that below 1kHz the drivers operate in what is referred to as the "piston mode", meaning that in the considered range of frequencies, the driver behaves as a rigid body. This assumption is not always verified, as measurements show that real drivers are never rigid. It is practically

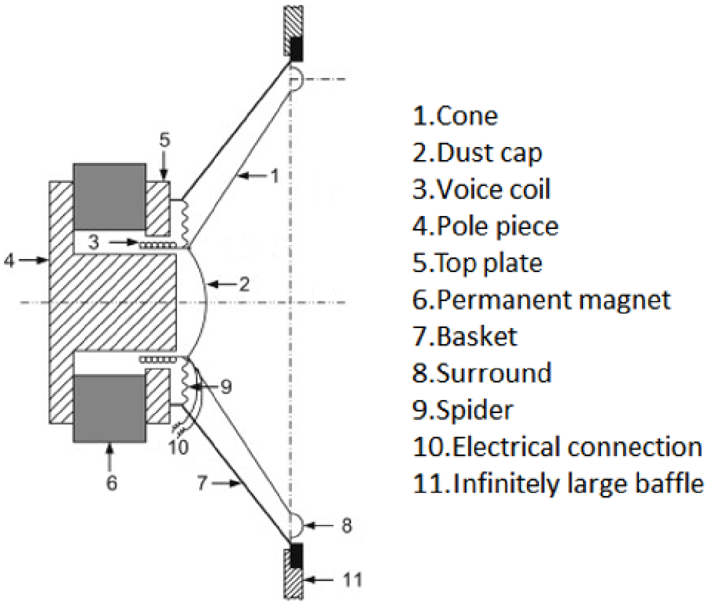


Fig. 3.15 Cross sectional sketch of a direct-radiator loudspeaker [84].

impossible to realise a perfect piston except for a small range of frequencies, which is related to the physical dimension of the diaphragm [85–87]. In this frequency band, force factor, stiffness and inductance introduce non linearities, generating spectral components that are not present in the input signal. An indication of the non linearities inherent in the system is given by the amplitude of the displacement. Due to the dynamical analogies, the differential equations describing the mechanical and acoustical behaviour can be solved by electrical circuit theory. In Figure 3.16, the mechanical and the electro-mechanical analogue circuits are shown. Thus, with the assumption of rigid body motion, the displacement of a loudspeaker can be computed as function of the frequency of the stimulus by considering the electro-mechanical components responsible of the dynamic response of the transducer, known as Thiele and Small (T&S) parameters [84]. In this way the voice coil displacement $\tilde{\eta}_c$, function of the acoustic frequency f_v, may be written as:

$$\tilde{\eta}_c\left(f_v\right) = \frac{\tilde{e}_g}{2\pi f_r Bl Q_{es}} \mid \gamma_c\left(f_v\right) \mid, \tag{3.86}$$

where $\tilde{e}_g$ is the voltage at the speaker's terminals, Bl is the force factor (magnetic flux density B multiplied by the length of the wire l), f_r is the resonance frequency of the speaker and $\gamma_c\left(f_v\right)$ is a dimensionless frequency response function given by:

$$\gamma_c\left(f_v\right) = \frac{1}{1 - \frac{f_v^2}{f_r^2} + j\frac{f_v}{f_r Q_{ts}}}, \tag{3.87}$$

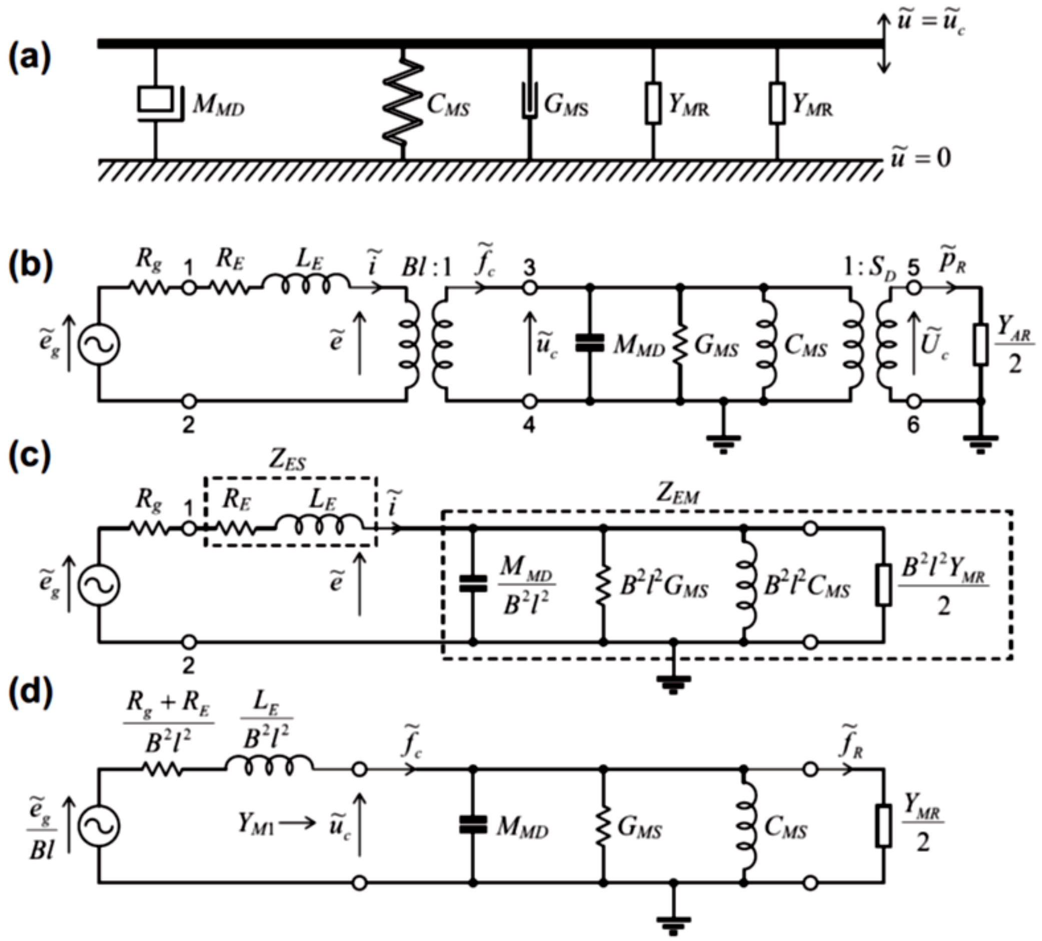


Fig. 3.16 a) Mechanical circuit of direct radiator loudspeaker; b) electro-mechano-acoustical analogous circuit of admittance type; c) electrical circuit showing static electrical impedance Z_{es} and motional electrical impedance Z_{em}; d) Analogous circuit of the admittance type with electrical quantities referred to the mechanical side.

where Q_{ts} represents the total damping effect, composed by the electrical damping Q_{es} and the mechanical damping Q_{ms}, and j is the imaginary unit. Equation (3.86) describes the frequency dependent behaviour of the loudspeaker displacement. The normalized voice coil displacement for different values of Q_{ts} is shown in Figure 3.17. It can be noted how the total damping affects the displacement behaviour around $f_v/f_r = 1$, while the displacement is virtually constant at $f_v/f_r \leq 1/3$ and proportional to $1/f_v^2$ at $f_v/f_r \geq 3$. Depending on the amplitude of the displacement, the transducer will generate distorted signals that can be classified as linear (low displacement amplitude) and non linear (high displacement amplitude) distortion. Both of them are regarded as regular distortions because they are accepted within the design process and are results of optimization process

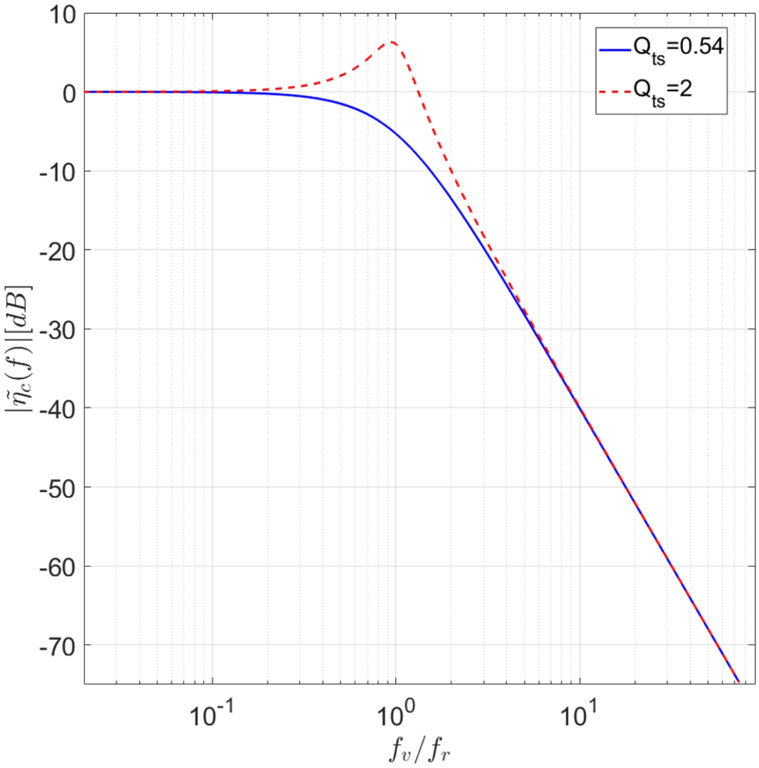


Fig. 3.17 Normalized voice coil displacement, with $Q_{ts} = 0.54$ (blue line) and $Q_{ts} = 2$ (red line).

giving the best compromise with other constraints (weight, cost, size). On the other hand, irregular distortions are non acceptable defects in a loudspeaker passing the End Of Line (EOL) tests. They are generated by defects caused during the manufacturing process, ageing and other external factor such as overload and temperature. A rubbing voice coil, buzzing parts, loose particles and air leaks are typical loudspeaker defects which produce irregular distortion, quite audible and not acceptable, generally defined as “rub & buzz”. In a woofer, for example, at higher frequencies (e.g. above 1kHz), the cone itself is not rigid and should be modelled as a flexible system. The vibrations travel transversally along the cone surface in what is generally referred to as cone break up. This behaviour generally becomes a dominant factor when the wavelength of the sound in air is comparable to or less than twice cone diameter [85, 86]. Above this frequency, radial and rocking modes are natural vibration patterns of the membrane, producing non linear or undesired output. Furthermore, the presence of any irregularities (e.g. mass distribution) produced in the manufacturing/assembling process, and/or diaphragms subjected to asymmetric acoustic loads enhance this phenomena. This becomes even more critical in small drivers such as headphones or micro-speakers, where small irregularities in the stiffness, mass and magnetic field distributions can affect dramatically the dynamic behaviour of these tiny structures [86–88]. Thus, all types of speakers (from the simple USB speakers to home theatre speakers up to the large professional speakers used in large concert halls) may be affected by physical defect during the assembly on the production line. For this reason, loudspeakers condition monitoring is an important topic in audio manufacturing in order to both fulfil the customer expectations and reduce the manufacturer costs to replace the damaged driver. The quality of these speakers is ensured by conducting inspection and quality control protocols during pre-production, production, and pre-shipment stage. In this domain, laser based analysis tools have been shown to yield significantly better results compared to traditional acoustic ones [89]. The former approach is more frequently used in advanced markets like automotive audio components and systems, while the latter is widely used in RD and manufacturing of acoustic transducers and consumer products (e.g. loudspeakers or audio products). Linear, non linear and irregular distortions depend highly on the amplitude and type of the stimulus.

A well known technique commonly used in audio environment to completely characterise the system with a single, fast and easy measurement was introduced in [90, 91]. It is based on exponentially swept sine signal defined as [92]:

$$x\left(t\right) = \sin\left[\frac{2\pi f_1 T}{ln\left(\frac{f_2}{f_1}\right)}\left(\left(\frac{f_2}{f_1}\right)^{\frac{t}{T}} - 1\right)\right], \tag{3.88}$$

where T is the length of the sine sweep in seconds, and f_1 and f_2 the starting and ending frequencies, respectively. This technique has the ability to separate the non-linear (distortion) responses from the linear response of the system. When the measured signal $y(t)$ is convolved with an inverse filter $g(t)$, namely the time reversal version of the test signal $x\left(t\right)$, the linear response compresses to an almost perfect impulse, with a delay equal to the length of the test signal. Simultaneously, the harmonic distortion responses compress to other smaller impulses, located at precise time delays occurring earlier than the impulse response. Applying a suitable time window it is possible to extract just the portion required, containing only the linear response or the distortion products. Thus, a Fourier transform can be applied, and both linear and non linear (harmonics) frequency responses can be displayed.

3.7 Summary

In this chapter, the basic concepts of micro-Doppler effect in radar and loudspeaker theory were introduced. Before to review the uses and models of micro-Doppler, the working principle of a radar, namely the Doppler effect, and the coherent receiver were introduced in Sections 3.2 and 3.2.1, respectively, leading to the formulation of the canonical form of a received radar signal. Having introduced the key concepts, an overview of the effect of the micro-Doppler in radar was provided in Section 3.3. In order to understand how extract the micro-Doppler signature, the concept of time-frequency analysis was introduced as well, with a detailed description of the commonly used time-frequency analysis tools in section 3.3.1. The basic principle of rigid body motion in the context of radar signal processing have been also analysed in Section 3.4, with a particular attention at the Euler angle in Section 3.4.1, to better understand the effect of translation and rotation of a target, introduced in section Section 3.5. To measure the reflective strength of a target, the Radar Cross Section were introduced in Section3.3.2, where some models were provided. For a good interpretation of the received radar echoes from a vibrating surface such as a loudspeaker and a better understanding of the effect of non linear motion dynamic, the micro-Doppler induced by both vibrating point and pendulum oscillation were also analysed in Sections 3.5.1 and 3.5.2, respectively. Finally, some of the aspects of the electrodynamic transducer motion, and how to acoustically characterize the behaviour of a speaker were introduced in Section 3.6. The concepts introduced in this chapter will be exploited to detect, confirm and characterize loudspeaker behaviour through radar micro-Doppler signature, focusing mainly on the rigid body motion of the acoustic driver.

Chapter 4

Deep learning in radar

4.1 Introduction

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains of modern society.

Conventional machine learning techniques were limited in their ability to process natural data in their raw form. For ages, constructing a machine learning system required careful engineering and considerable domain expertise to design a feature extractor that transformed raw data into a suitable internal representation of feature vector from which the learning subsystem, often a classifier, could detect or classify patterns in the input.

Representation learning is a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification. Deep learning methods are representation learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level into a representation of higher, slightly more abstract level [93].

For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations. An image, for example, comes in the form of array of pixel values and the learned feature in the first layer typically represents the presence or the absence of edges at particular orientations and locations in the image. The second layer detect motifs by spotting particular arrangements of edges, regardless of small variation in the

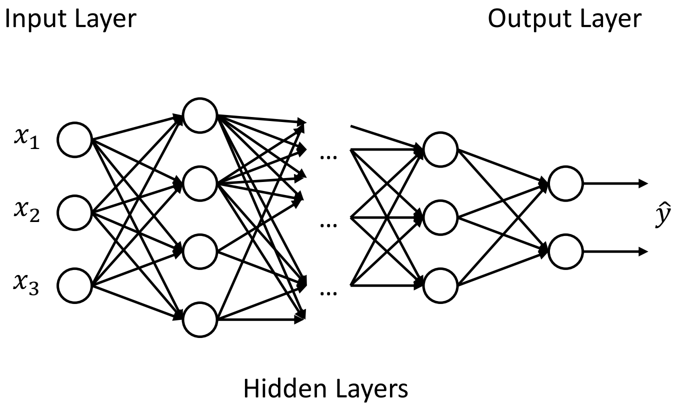


Fig. 4.1 Example of deep feed-forward neural network.

edge position. The key aspect of deep learning is that these layers of features are not designed by humans: they are learned from data using a general purpose learning procedure.

Motivated by the recent advances in deep learning, the goal of this chapter is to give an overview of different deep learning techniques. The most general architecture for deep learning is introduced in section 4.2, together with Convolutional Neural Network (CNN) and how it can be used in the radar domain. A deeper look on Recurrent Neural Network (RNN) is given in section 4.3. Finally, to cope the vanishing gradient problem that affect the performance of RNNs, the Long Short-Time Memory (LSTM) is introduced in section 4.4.

4.2 Deep Learning in Radar

Deep feed-forward neural networks are the most general architecture for deep learning. Deep networks differ from standard Artificial Neural Network (ANN) in terms of their depth, introducing new strength and capability to the network. An example of feed forward neural network is show in Figure 4.1, also known as Multi-Layer Perceptron (MLP), where the interconnected nodes are called "neurons" [94]. At each neuron, the inner products of the data with a weight matrix plus bias is fed as the input. Hence, the output of the $k-th$ layer and subsequently the input of the $(k+1)-th$ layer of the neural network can be characterised as:

$$\mathbf{a}_{k+1} = g\left(\mathbf{W}_k \mathbf{a}_k + \mathbf{b}_k\right), \tag{4.1}$$

where $\mathbf{a}_k \in \mathcal{R}^n$ is the input vector of the $k-th$ layer of the network, $g(\cdot)$ is an element wise non linear activation function of the neuron and $\mathbf{W}_k \in \mathcal{R}^{m\times n}$ and $\mathbf{b}_k$

are respectively the weight matrix and bias vector of the $k-th$ layer. Augmenting the bias vector $\mathbf{b}_k$ to the weight matrix $\mathbf{W}_k$ to form $\boldsymbol{\theta}$, and $\mathbf{a}_1$ equal to the input vector $\mathbf{x}$, the final output of a N layer network is given as:

$$\hat{\mathbf{y}} = g\left(\boldsymbol{\theta}_N \cdot g\left(\boldsymbol{\theta}_{N-1} \ldots \cdot g\left(\boldsymbol{\theta}_1 \cdot \mathbf{x}\right)\right)\right), \tag{4.2}$$

which defines the forward propagation. The network matrices $\mathbf{W}_k$ and bias $\mathbf{b}_k$ are learned by optimizing same criteria defined in terms of a mismatch between the true and the reconstructed signal, evaluated through a loss function. The optimization is carried out using the backpropagation algorithm [95, 96]. It is an efficient algorithm that is used to calculate the derivative of the loss function with respect to each parameter of the neural network, where weights and bias are updated through Stochastic Gradient Descent (SGD) optimizer or similar [95–97].
In radar domain, the use of deep learning methods may lead to some benefits, solving a broad range of problems. The most straightforward application of deep learning to radar is within the area of Synthetic Aperture Radar (SAR) for Automatic Target Recognition (ATR). In [98–100] deep learning algorithm is used to classify military target from SAR images. In the context of cognitive radar, in [101, 102] deep learning algorithms were used in order to detect the best sub-arrays of a phased array radar antenna, in order to increase the Direction of Arrival (DoA) estimation. Deep learning algorithms are also used for human detection and activity classification based on Doppler radar [103–105] as for hand gesture recognition using micro-Doppler signatures [106, 107].

In all these applications, Convolutional Neural Network (CNN) has shown to achieve great results. Ideal for image and video processing, CNN are deep learning algorithms widely used for computer vision tasks, mainly consisting of convolutional layers and pooling layers, followed by one or more fully connected layers as in a standard multilayer neural network introduced previously [93, 94, 108, 109]. The input to the CNN architecture shown in Figure 4.2 is a $m \times m \times r$ image where m is the height and width of the image and r is the number of channel, e.g. ab RGB image has $r = 3$. The convolutional layer will have k filters (or kernels) of size $n \times n \times q$ where n is smaller than the dimension of the image and q can either be the same as the number of channels r or smaller and may vary for each kernel. The size of the filters gives rise to the locally connected structure which are each convolved with the image to produce k feature maps of size $mn + 1$. Example of feature maps at the output of different convolutional layers are shown in Figure 4.3. After obtaining features using convolution, the

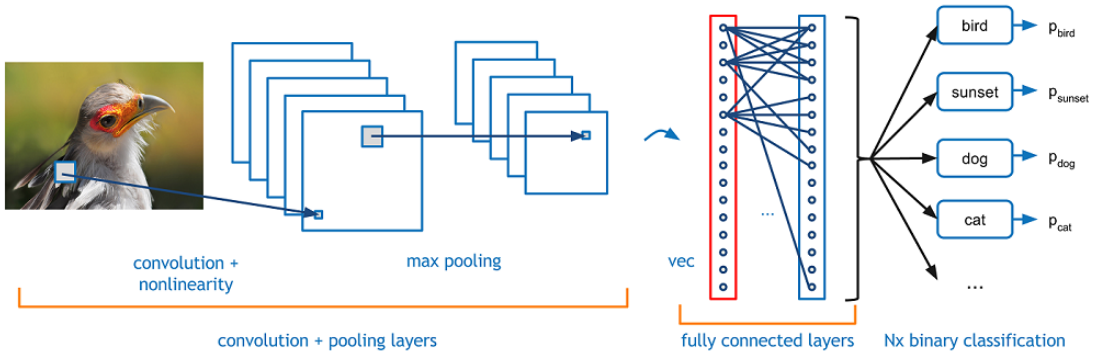


Fig. 4.2 Example of Convolutional Neural Network (CNN) [110].

aim is to use them for classification. In theory, all the extracted features could be used with a classifier such as a softmax classifier, but this can be computationally challenging. Considering for instance an images of size 96×96 pixels, and suppose 400 features have been learned using a 8×8 kernel. Each convolution results in an output of size $(968+1) \times (968+1) = 7921$, and since 400 features are available, this results in a vector of $89^2 \times 400 = 3168400$ features per example. Training a classifier with inputs having more than 3 millions of features can be unwieldy, and can also be prone to over-fitting. This phenomenon happens when the classifier learns the details and the noise in the training data: this affects negatively the performance on new data. To address this problem, a natural approach is to aggregate statistics of these features at various locations. For example, the mean or max value of a particular feature over a region of the image could be computed. These summary statistics are much lower in dimension (compared to using all of the extracted features) and can also improve results (less over-fitting). The aggregation operation is called "pooling", or sometimes "mean pooling" or "max pooling", depending on the pooling operation applied. Thus, each feature map is subsampled over $p \times p$ contiguous regions where p ranges between 2 for small images and is usually not more than 5 for larger inputs. Either before or after the pool layer an additive bias and sigmoidal non linearity is applied to each feature map. The mechanism by which CNN learns to recognise components of an image (e.g. lines, curves) and then learn to combine these components to recognise larger structures (e.g. faces, objects) is based on the research of same patterns in all the different subfields of the image (different regions of the space).

A different class of NN is preferred in case sequential data are processed: while a CNN learns to recognize patterns across space, Recurrent Neural Network (RNN) is trained to recognise patterns across time [112]. The RNN uses an architecture

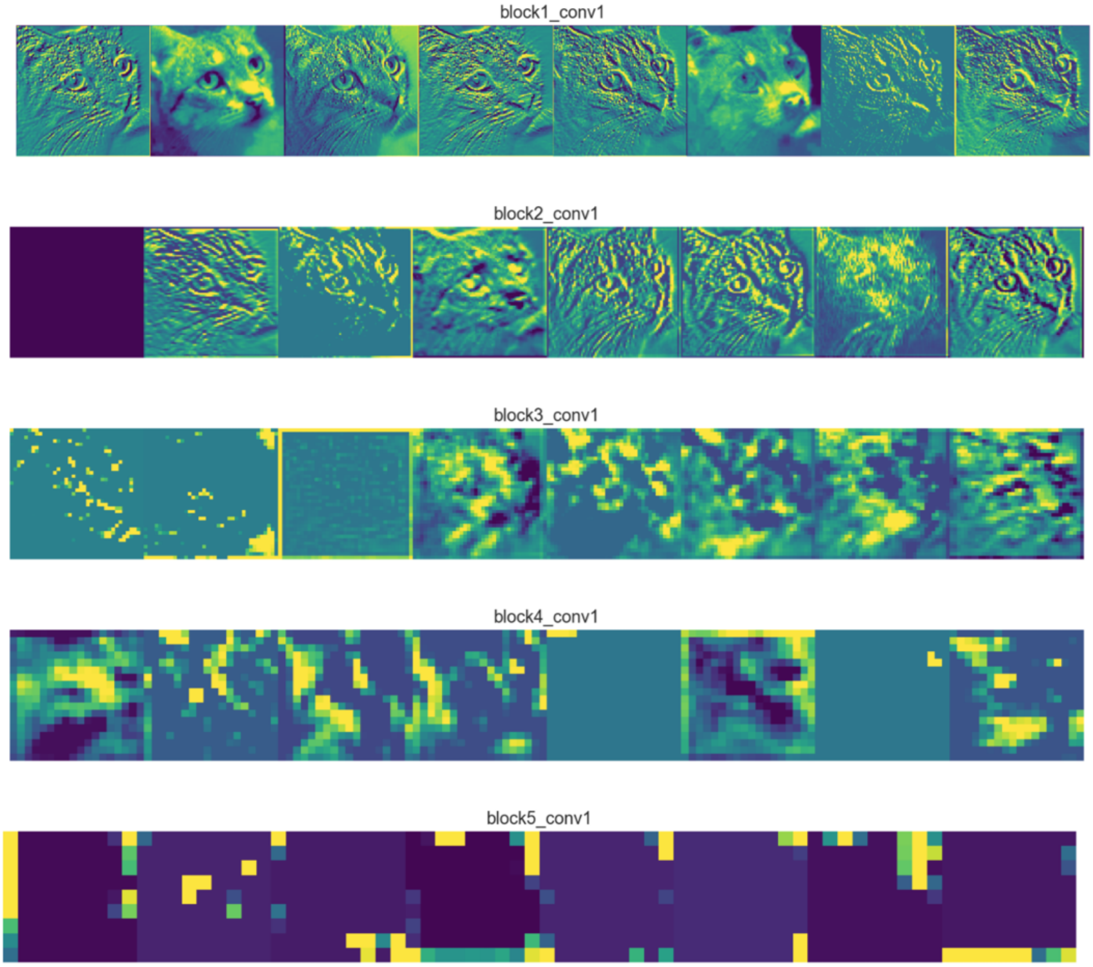


Fig. 4.3 Example of feature maps at the output of different convolutional layer [111].

that is not dissimilar to the traditional NN. The difference is that the RNN introduces the concept of memory, and it exists in the form of a different type of link. Unlike a feed forward NN, the outputs of some layers are fed back into the inputs of a previous layer. This addition allows for the analysis of sequential data, which is something that the traditional NN is incapable of, making RNNs ideal for applications with time component (audio, time-series data) and natural language processing (NLP).

4.3 Recurrent Neural Network

One of the most emerging Deep Neural Network (DNN) algorithm is Recurrent Neural Network (RNN), especially thanks to the success in NLP tasks and speech recognition. The idea behind RNN is to make use of sequential information. In a

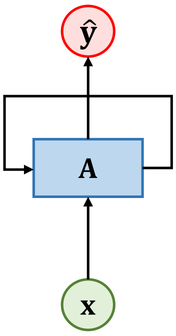


Fig. 4.4 Basic Recurrent Neural Network (RNN) with loops.

traditional NN, it is assumed that all the inputs (and outputs) are independent from each other. Making and analogy with human brain elaboration process, humans do not start their thinking from scratch every second. A reader, for example, understand each word based on his/her understanding of previous words. The reader does not throw everything away and start thinking from scratch again. His/her thoughts have persistence. Traditional NN cannot do this, and it seems like a major shortcoming. If a series of events happening in a movie wants to be classified at every point, it is unclear how a traditional NN could use its reasoning about previous events in the film to inform later ones [113]. Recurrent neural networks address this issue, by having loops in them, allowing information to persist. In Figure 4.4, a section of neural network $\mathbf{A}$, known as hidden layer, maps an input sequence of $\mathbf{x}$ values to a corresponding sequence of output $\hat{\mathbf{y}}$ values. A loop allows information to be passed from one step of the network to the next. These loops make recurrent neural networks to appear unpredictable. However, looking at RNN in a different way, it turns out that they are not all that different than a normal neural network. Showing the RNN in its unrolled version, as in Figure 4.5, reveals that recurrent neural networks are intimately related to kind of data like sequences and lists. They are the natural architecture of neural network to use for such data. A RNN can be thought of as multiple copies of the same network, which allow previous outputs to be used as inputs. By expressing the single hidden layer unit $\mathbf{A}_t$ as the graph in Figure 4.6 [114], it is possible to define the forward propagation equations. For each time step from $t = 1$ to $t = T$, the

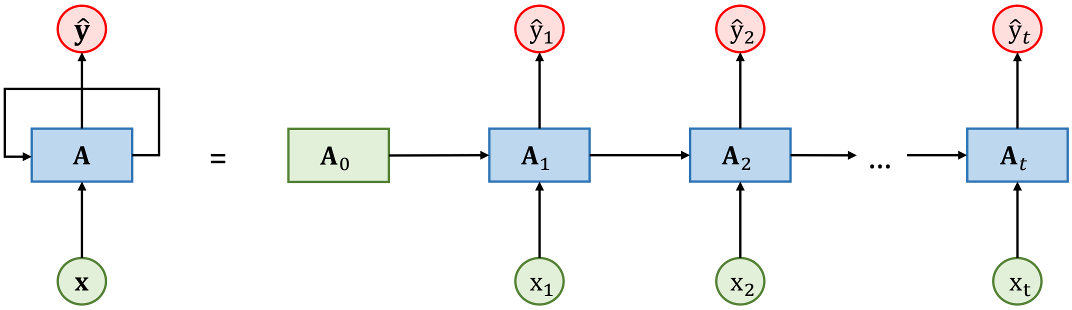


Fig. 4.5 Basic Recurrent Neural Network (RNN) in its unrolled version.

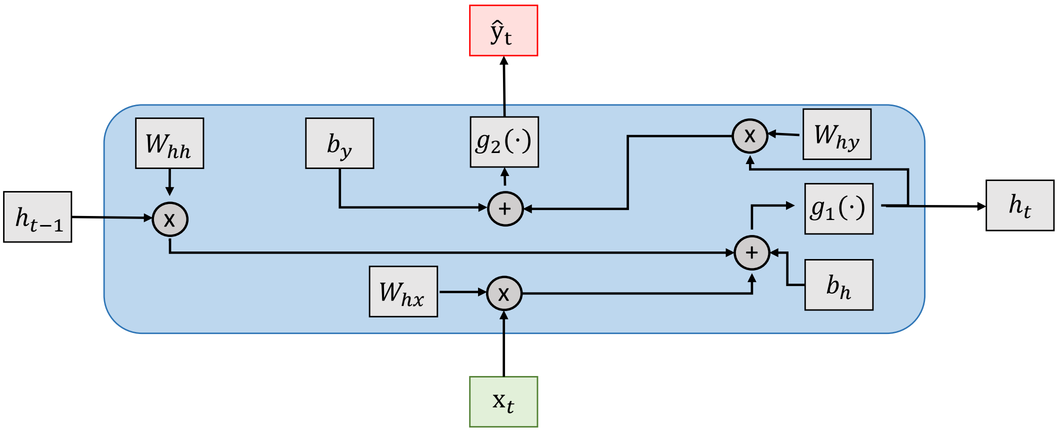


Fig. 4.6 Single hidden layer unit.

update equations for the RNN depicted in Figure 4.5 are [94]:

$$\begin{cases} \mathbf{a}_t = \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{W}_{hx}\mathbf{x}_t + \mathbf{b}_h \\ \mathbf{h}_t = g_1\left(\mathbf{a}_t\right) \\ \mathbf{o}_t = \mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y \\ \hat{\mathbf{y}}_t = g_2\left(\mathbf{o}_t\right) \end{cases} \tag{4.3}$$

where $\mathbf{h}_t$ is hidden layer activation, $\mathbf{b}_h$ and $\mathbf{b}_y$ are the bias vectors and $\mathbf{W}_{hx}$, $\mathbf{W}_{hh}$ and $\mathbf{W}_{hy}$ are the weight matrices respectively for the input to hidden, hidden to hidden and hidden to output connections. The hidden layer activation $\mathbf{h}_t$ and the output $\hat{\mathbf{y}}_t$ are obtained, respectively, by the activation functions $g_1(\cdot)$ and $g_2(\cdot)$ [115]. The activation function of the hidden layer in RNN is traditionally implemented by a logistic function or sigmoid function, defined as:

$$g_1(x) = \frac{1}{1+e^{-x}}. \tag{4.4}$$

Due to the S-shaped function in (4.4), the activation function reduces the value of the hidden layer into the range $[0,1]$. It is worth noting that another common choice for the activation function $g_1(\cdot)$ is the hyperbolic tangent, or tanh, defined as:

$$g_1(x) = \tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}. \tag{4.5}$$

It can be seen as the rescaled version of the sigmoid function, where the output range of the hidden layer is $[-1,1]$. Although it is not very much used in RNN, recent research has found a different activation function, namely the Rectified Linear Unit (ReLU) function, that often works better in practice for deep neural networks. This activation function is different from sigmoid and tanh because it is not bounded or continuously differentiable. The rectified linear activation function is given by:

$$g_1(x) = ReLu(x) = \max(0,x). \tag{4.6}$$

On the other hand, the activation function $g_2(\cdot)$ can be considered as the readout of the RNN. Usually it is represented by the softmax function, used to convert a vector of raw scores into class probabilities at the output layer of a Neural Network, for classification purpose [94]. As the sigmoid function, even the softmax function squashes the output of each unit to be into the range $[0,1]$. Given K different classes, the softmax function can be mathematically expressed as:

$$g_2(z_i) = softmax(z_i) = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}}, \tag{4.7}$$

where z is a vector of the inputs to the output layer, and i is the index of the output units, such that $i = 1,2,\ldots,K$ with $\sum_{k=1}^{K} g_2(z_k) = 1$.

4.3.1 RNN architectures

Due to a wide variety of architectures, recurrent neural network can be divided in three different classes [114].

The first class is composed by RNN that produce an output at each time step and have recurrent connections between hidden units. An example of such RNN is

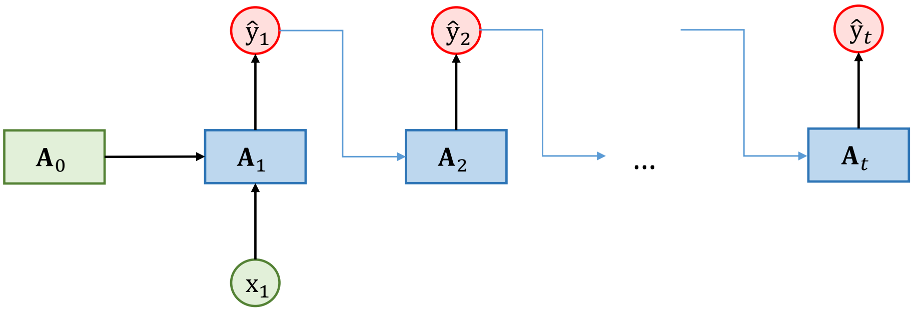


Fig. 4.7 RNN with one-to-many architecture with $T_x = 1$ and $T_y > 1$.

the architecture described so far, where the number of output values T_y are equal to the time steps of the input sequence T_x. In case $T_x = T_y = 1$, the traditional neural network is found, with architecture one-to-one. In case $T_x = T_y \neq 1$, the architecture is referred as many-to-many. An example of application for this kind of RNN is the Name Entity Recognition (NER) [116].

The second class is composed by RNN that produces an output at each time step and have recurrent connections only from the output at one time step to the hidden units at the next time step. An example of this class of RNN is given in Figure 4.7 and referred as one-to-many architecture: the RNN generates a series of output values based on a single input value. Compared to the first class of RNN, the architecture shown in Figure 4.7 is less powerful of the previous one. The RNN in Figure 4.5 can choose to put any information about the past into its hidden layer and transmit it to the future. The RNN in Figure 4.7, instead, is trained to put a specific output value into $\hat{y}$, and $\hat{y}$ is the only information it is allowed to send to the future. There are no direct connections between hidden units. Unless $\hat{y}$ is very high dimensional and rich, it will usually lack important information from the past. This makes the RNN in Figure 4.7 less powerful, but it may be easier to train because each time step can be trained in isolation from the others. A prime example for using such an architecture is for music generation task, where an input is the first note.

Last but not least, is the class of RNN with recurrent connections between hidden units - that reads an entire sequence and then produce either sequence (many-to-many) or single (many-to-one) output. In the event of output sequence, with $T_x \neq T_y$, an example of many-to-many architecture is shown in Figure 4.8. This

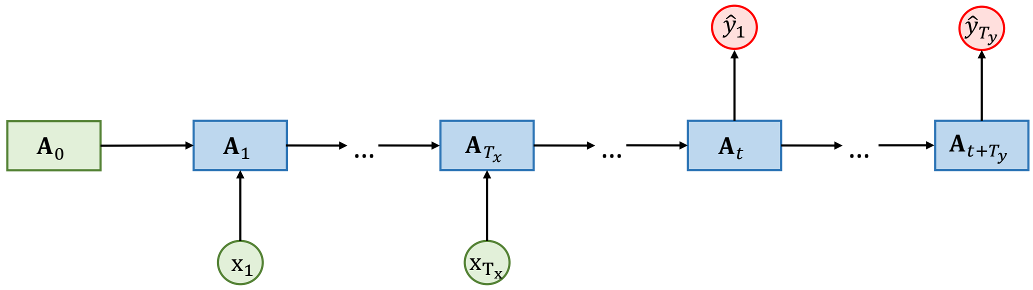


Fig. 4.8 RNN with many-to-many architecture with $T_x \neq T_y$.

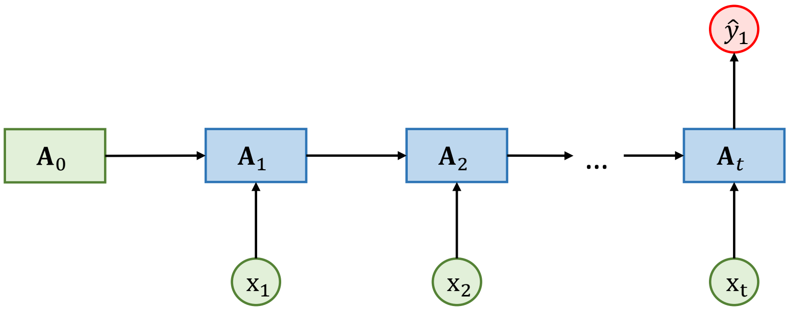


Fig. 4.9 RNN with many-to-one architecture with $T_x > 1$ and $T_y = 1$.

architecture refers to where many inputs are read to produce many outputs, where the length of inputs is not equal to the length of outputs. A prime example for using such an architecture is machine translation tasks: the part of the network which reads the sentence to be translated is referred as encoder, while the part of the network which translates the sentence into desired language is the decoder. In the event of single output, an example of many-to-one RNN architecture is shown in Figure 4.9: a suitable example for using such an architecture is for classification tasks. This is similar to the many-to-many RNN already discussed, but it only uses the final hidden state to produce the single output. The output will be a vector containing scores relatives to each classes: after the softmax layer, those scores are turned into probabilities and ultimately decide among the classes.

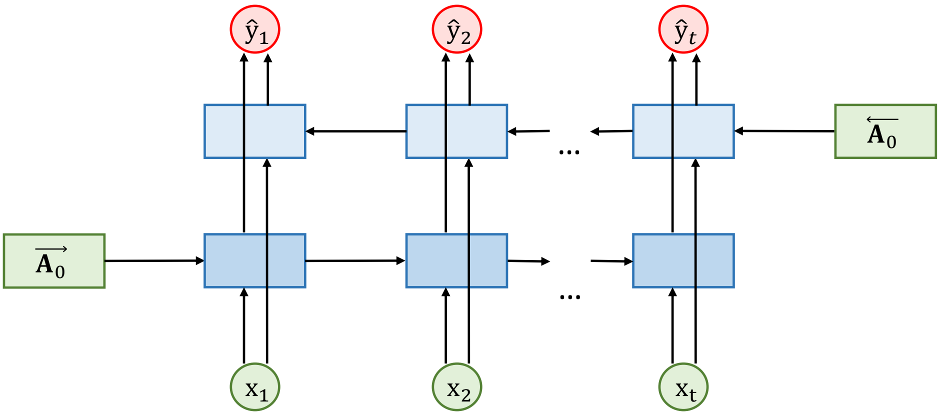


Fig. 4.10 Bidirectional recurrent neural network architecture.

4.3.2 Bidirectional RNN

Over the years researchers have developed more sophisticated types of RNNs in order to increase the learning capacity. In Figure 4.10, the architecture of a Bidirectional RNN (BRNN) is show [94, 114, 117]. BRNNs are based on the idea that the output at time t may not only depend on the previous elements in the sequence, but also future elements. For example, to predict a missing word in a sequence, it would be preferable to look at both the left and the right context. Bidirectional RNNs are quite simple: they are just two RNNs stacked on top of each other. The output is then computed based on the hidden state of both RNNs. Similar to BRNN are Deep (Bidirectional) RNNs, with the only difference of multiple layers per time step. These architectures ensure an increased learning capacity in case an increased amount of training data are used.

4.3.3 Training & Backpropagation Through Time

In supervised learning, at a given training set is associated a corresponding set of target values. Aim of the network is to minimize the error between the estimated outputs and the target values. This is done during the training stage thanks to the use of a loss function that calculates the error between the network output and the target solution with the current set of parameters, composed by weights and bias.

There is a natural choice of both output unit activation function and matching error function, according to the type of problem being solved:

- in regression problem, linear outputs and a sum of squares error are preferred;

- in multiple independent binary classifications problem, logistic sigmoid outputs and a cross-entropy error function are preferred;
- in multi class classification, softmax outputs with the corresponding multi-class cross-entropy error function are preferred.

The network in Figure 4.5, with its forward propagation equations in (4.3), is an example of a RNN that maps an input sequence to an output sequence of the same length. The total loss $L_{tot}\left(\mathbf{W},\mathbf{b}\right)$ for a given sequence of $\mathbf{x}$ values paired with a sequence of $\mathbf{y}$ values would then be just the sum of the losses $L_t\left(y_t,\hat{y}_t\right)$ over all the time steps. In case cross entropy is used as loss function, the total loss can be written as follow [94, 117]

$$L_{tot}\left(\mathbf{W},\mathbf{b}\right)=\sum_{t=1}^{T}L_t\left(y_t,\hat{y}_t\right)=-\sum_{t=1}^{T}y_t\ln\hat{y}_t. \tag{4.8}$$

In multi class classification problem, a training set $\mathbf{X}$ is considered. The training set is composed by a set of input vectors such that $\mathbf{X}=\{\mathbf{x}_1\ldots\mathbf{x}_N\}$, with N the number of independent and identically distributed observations. Each input is assigned to one of K mutually exclusive classes. The binary target variables $y_k\in\{0,1\}$ have a 1-of-K coding scheme indicating the class, and the network outputs are interpreted as conditional probability $\hat{y}_k=p\left(y_k=1|\mathbf{x},\mathbf{W}\right)$, leading to the following error function:

$$L_{tot}\left(\mathbf{W},\mathbf{b}\right)=-\sum_{n=1}^{N}\sum_{k=1}^{K}y_{nk}\ln\hat{y}_k, \tag{4.9}$$

where $\hat{y}_k$ is the output of the network, such that:

$$\hat{y}_k\left(\mathbf{x},\mathbf{W}\right)=\frac{e^{o_i}}{\sum_{k=1}^{K}e^{o_k}}, \tag{4.10}$$

and y_{nk} element of the matrix $\mathbf{Y}$ of dimension $N\times K$, indicating that the $n-th$ sample belongs to the $k-th$ class.

The goal is to minimize $L_{tot}\left(\mathbf{W},\mathbf{b}\right)$ as a function of $\mathbf{W}$ and $\mathbf{b}$. After having randomly initialized the weights and biases, a gradient-based optimization algorithm is applied. One of the most common gradient-based optimization algorithm used to learn the network parameters during the training stage is the Stochastic Gradient Descent (SGD) algorithm. The standard SGD algorithm updates the network parameters (weights and biases) to minimize the loss function by taking small steps at each iteration in the direction of the negative gradient of the loss, such that:

$$\boldsymbol{\theta}_l = \boldsymbol{\theta}_{l-1} - \alpha \nabla L_{tot} \left(\boldsymbol{\theta}_{l-1} \right), \tag{4.11}$$

where l is the iteration number, $\alpha > 0$ is the learning rate, $\boldsymbol{\theta}$ is the parameter vector, and $L_{tot} \left(\boldsymbol{\theta} \right)$ is the loss function. Usually the gradient of the loss function, $\nabla L_{tot} \left(\boldsymbol{\theta} \right)$, is evaluated using the entire training set at once. By contrast, a mini batch version of the SGD algorithm is often used, evaluating the gradient and updating the parameters using a subset of the training data, called mini-batch. The full pass of the training algorithm over the entire training set using mini-batches is one epoch. Many extensions of the SGD algorithm are available, as SGD with Momentum (SGDM). The standard SGD algorithm can oscillate along the path of steepest descent towards the optimum. Adding a momentum term to the parameter update is one way to reduce this oscillation [95]. The SGDM update equation is:

$$\boldsymbol{\theta}_l = \boldsymbol{\theta}_{l-1} - \alpha \nabla L_{tot} \left(\boldsymbol{\theta}_{l-1} \right) + \gamma \left(\boldsymbol{\theta}_{l-1} - \boldsymbol{\theta}_{l-2} \right), \tag{4.12}$$

where γ determines the contribution of the previous gradient step to the current iteration. During this process, the key step is the computation of the gradient and its partial derivatives. Once the total loss is obtained, the partial derivatives are computed using the BackPropagation Through Time (BPTT) algorithm in order to minimize the error [118]: the errors at one time step must be backpropagated "through time" to all previous time steps, hence the name BackPropagation Through Time (BPTT). As usual in backpropagation algorithm, there is the need to make a forward pass and a backward pass. In the forward pass, all the elements in the neural network are calculated, so all the hidden elements are computed, the prediction $\hat{y}_t$ is obtained and the losses L_t and L_{tot} can be computed. In the backward pass, the gradient of the loss has to be computed with respect to the biases $\mathbf{b}_h$ and $\mathbf{b}_y$, and the weights matrices $\mathbf{W}_{hy}$, $\mathbf{W}_{hx}$ and $\mathbf{W}_{hh}$. In usual feed forward neural network, the backpropagation is done only through layer, in vertical direction. In RNN instead sequences are involved, reason why backpropagation has to be applied not only through layer but even through time, in horizontal direction. As an example, to compute the gradient of the loss function with respect to the weight matrix $\mathbf{W}_{hy}$ all the gradients coming from all the time steps in the sequence need to be summed, such as:

$$\frac{\partial L}{\partial W_{hy}} = \sum_{t=0}^{T} \frac{\partial L_t}{\partial W_{hy}}. \tag{4.13}$$

It is necessary then calculate the loss in a specific time step t. For this reason, the chain rule is used, such as [94, 119]

$$\frac{\partial L_t}{\partial W_{hy}} = \frac{\partial L_t}{\partial \hat{y}_t}\frac{\partial \hat{y}_t}{\partial W_{hy}}. \tag{4.14}$$

As result, the partial derivative is the product of two component. The first element can be calculated, depending on the employed loss function. The second element can be easily computed also, since the prediction $\hat{y}_t$ depends on the matrix $\mathbf{W}_{hy}$ only in one point:

$$\hat{\mathbf{y}}_t = g_2\left(\mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y\right). \tag{4.15}$$

A more difficult case is the computation of the gradient of the loss with respect to the weight matrix $\mathbf{W}_{hh}$ [94, 119]. As before, the gradient of the loss function is given by

$$\frac{\partial L}{\partial W_{hh}} = \sum_{t=0}^{T}\frac{\partial L_t}{\partial W_{hh}}. \tag{4.16}$$

Proceeding as in the previous case, the gradient of the single time step t would be:

$$\frac{\partial L_t}{\partial W_{hh}} = \frac{\partial L_t}{\partial \hat{y}_t}\frac{\partial \hat{y}_t}{\partial h_t}\frac{\partial h_t}{\partial W_{hh}}. \tag{4.17}$$

Looking at the hidden unit formula, $\mathbf{h}_t$ does not depend on the weight matrix $\mathbf{W}_{hh}$ only at time step t, but also at time step $t-1$. For this reason, the chain rule is substituted by the formula of the total derivatives, such that:

$$\frac{\partial L_t}{\partial W_{hh}} = \frac{\partial L_t}{\partial \hat{y}_t}\frac{\partial \hat{y}_t}{\partial h_t}\left(\frac{\partial h_t}{\partial W_{hh}} + \frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial W_{hh}} + \cdots\right). \tag{4.18}$$

As result, the last part of the equation is the sum of the contributions from the all previous time steps to the gradient at time step t. To calculate the contribution from time step k to the gradient at time step t, there is the needs to go from the hidden units at time step t to hidden units at time step k. Then, (4.18) can be written as:

$$\frac{\partial L_t}{\partial W_{hh}} = \frac{\partial L_t}{\partial \hat{y}_t}\frac{\partial \hat{y}_t}{\partial h_t}\sum_{k=0}^{t}\left(\prod_{i=k+1}^{t}\frac{\partial h_i}{\partial h_{i-1}}\right)\frac{\partial h_k}{\partial W_{hh}}. \tag{4.19}$$

In each element of the sum, the product of the Jacobian matrices of the gradients of hidden units at one time step with respect to the hidden units at the previous time step is found. Ultimately (4.16) can be written as:

$$\frac{\partial L}{\partial W_{hh}} = \sum_{t=0}^{T} \frac{\partial L_t}{\partial W_{hh}} = \sum_{t=0}^{T} \frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \sum_{k=0}^{t} \left(\prod_{i=k+1}^{t} \frac{\partial h_i}{\partial h_{i-1}} \right) \frac{\partial h_k}{\partial W_{hh}}. \tag{4.20}$$

The same approach is applied also for the computation of the gradient of the loss with respect to the weight matrix $\mathbf{W}_{hx}$. As before, the dependence between hidden units and the weight matrix $\mathbf{W}_{hx}$ is not only in one place. Since the hidden unit at time step t depends by the hidden unit at time step $t-1$, hidden units of all the previous time steps also depend by the weight matrix $\mathbf{W}_{hx}$, leading to the gradient:

$$\frac{\partial L}{\partial W_{hx}} = \sum_{t=0}^{T} \frac{\partial L_t}{\partial W_{hx}} = \sum_{t=0}^{T} \frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \sum_{k=0}^{t} \left(\prod_{i=k+1}^{t} \frac{\partial h_i}{\partial h_{i-1}} \right) \frac{\partial h_k}{\partial W_{hh}}. \tag{4.21}$$

Similarly, the gradient of the loss with respect to the biases $\mathbf{b}_h$ and $\mathbf{b}_y$ is computed as:

$$\frac{\partial L}{\partial b_h} = \sum_{t=0}^{T} \frac{\partial h_t}{\partial b_h} \frac{\partial L_t}{\partial h_t} \qquad \frac{\partial L}{\partial b_y} = \sum_{t=0}^{T} \frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial b_y}. \tag{4.22}$$

4.3.4 Vanishing and exploding gradient problems

The basic principle about how RNN works and how BPTT algorithm can be used to train the network have been shown so far. As result of the gradient backpropagated not only through layers but also through time, two well known problems may arise: vanishing and exploding gradients [120]. In (4.20), the gradient of the loss with respect to the weight matrix $\mathbf{W}_{hh}$ is derived, where a product of Jacobian matrices is present in each term of the sum [95, 115]. Generally, the Jacobian matrices are subjected to two complementary hypothesis, namely:

$$\left\| \frac{\partial \mathbf{h}_i}{\partial \mathbf{h}_{i-1}} \right\|_2 < 1 \qquad \left\| \frac{\partial \mathbf{h}_i}{\partial \mathbf{h}_{i-1}} \right\|_2 > 1, \tag{4.23}$$

where $\| \cdot \|_2$ is the spectral matrix norm which is equal to the largest singular value of the matrix. In case all the Jacobian matrices in the product have the norms which are less than one, then their product goes to zero exponentially fast when the number of elements in this product tends to infinity. This problem is usually called the vanishing gradient problem, because a lot of elements in the gradient simply vanish and do not affect the training. As a result, in the first case, the contributions from the faraway steps go to zero and the gradient contains only the information about nearby steps. Thus it is difficult to learn long range dependencies with a simple recurrent neutral network. On the contrary, if all the Jacobian matrices have the norms which are higher than one, then their product goes to infinity exponentially fast, leading the gradient itself to increase as well. If an input sequence is long enough the gradient may even become a not-a-number in practice. This problem is called the exploding gradient problem and it makes the training very unstable. Both vanishing and exploding gradients are an issue in neural networks, but mostly RNNs suffer with vanishing gradients due to their large and complex structures. Fortunately, there are a few ways to cope with the vanishing gradient problem. A proper initialization of the weight matrices can reduce the effect of vanishing gradients, as the regularization approach, consisting to the addition of a regularization term into the error function, which not only reduces the effect of vanishing gradients but also avoids the overfitting problem. Overfitting refers to a model that models the training data too well: it happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model. The problem is that these concepts do not apply to new data and negatively impact the models ability to generalize. A traditional solution is to add a regularization term to the loss function [95], such that:

$$L_R\left(\boldsymbol{\theta}_l\right) = L\left(\boldsymbol{\theta}_l\right) + \lambda\Omega\left(\boldsymbol{\theta_l}\right), \tag{4.24}$$

where θ is the parameters vector, λ is the regularization coefficient and the regularization function $\Omega\left(\boldsymbol{\theta_l}\right)$ is:

$$\Omega\left(\boldsymbol{\theta}\right) = \frac{1}{2}\boldsymbol{\theta}^T\boldsymbol{\theta}. \tag{4.25}$$

A more preferred solution is to use Long Short-Term Memory (LSTM) [121], firstly proposed in 1997 explicitly designed to deal with vanishing gradients and efficiently learn long-range dependencies.

4.4 Long Short-Term Memory

LSTM recurrent neural networks are an improvement over the general recurrent neural networks, which possess a vanishing gradient problem, making it suitable to learn long-term dependencies between time steps of sequence data. As stated

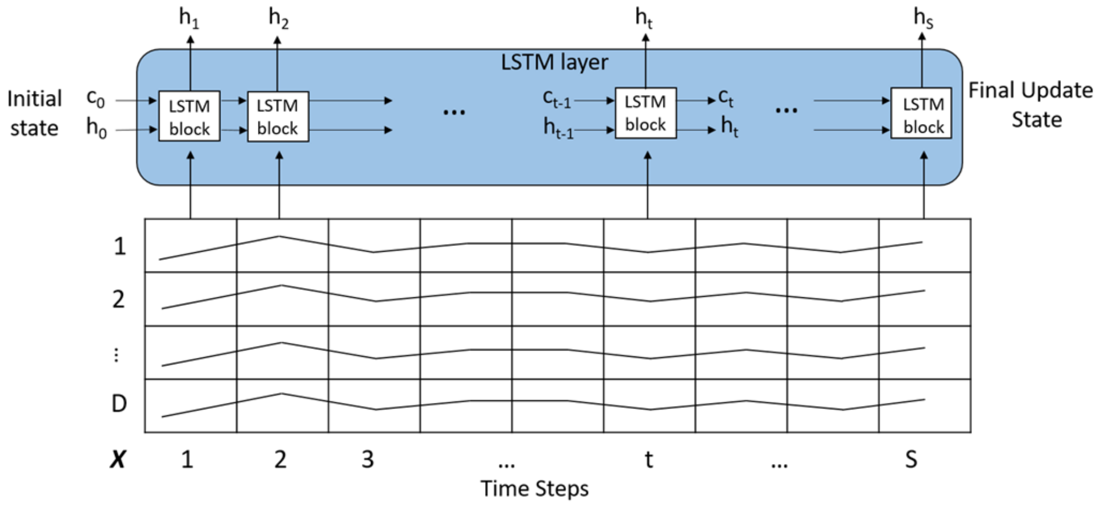


Fig. 4.11 Long Short Term Memory (LSTM) network architecture.

in [121], LSTM RNNs address the vanishing gradient problem commonly found in ordinary recurrent neural networks by incorporating gating functions into their state dynamics. The Figure 4.11 illustrates the flow of a time series $\mathbf{X}$ with D features of length S through an LSTM layer. In this diagram, h_t denotes the output (also known as the hidden state) and c_t denotes the cell state. The first LSTM block takes the initial state of the network and the first time step of the sequence $\mathbf{X}_1$, and then computes the first output h_1 and the updated cell state c_1. At time step t, the block takes the current state of the network (c_{t-1}, h_{t-1}) and the next time step of the sequence $\mathbf{X}_t$, and then computes the output h_t and the updated cell state c_t. The hidden state at time step t contains the output of the LSTM layer for this time step, while the cell state contains information learned from the previous time steps. At each time step, the layer adds information to or removes information from the cell state, controlling updates and outputs state using gates. The diagram in Figure 4.12 shows the flow of data at time step t and how the gates forget, update and output control the cell and hidden state. This is possible through the learnable weights. For an LSTM layer, they can be divided in input weights, recurrent weights and bias, which are concatenated as follow:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_i \\ \mathbf{W}_g \\ \mathbf{W}_f \\ \mathbf{W}_o \end{bmatrix} \qquad \mathbf{R} = \begin{bmatrix} \mathbf{R}_i \\ \mathbf{R}_g \\ \mathbf{R}_f \\ \mathbf{R}_o \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} \mathbf{b}_i \\ \mathbf{b}_g \\ \mathbf{b}_f \\ \mathbf{b}_o \end{bmatrix} \tag{4.26}$$

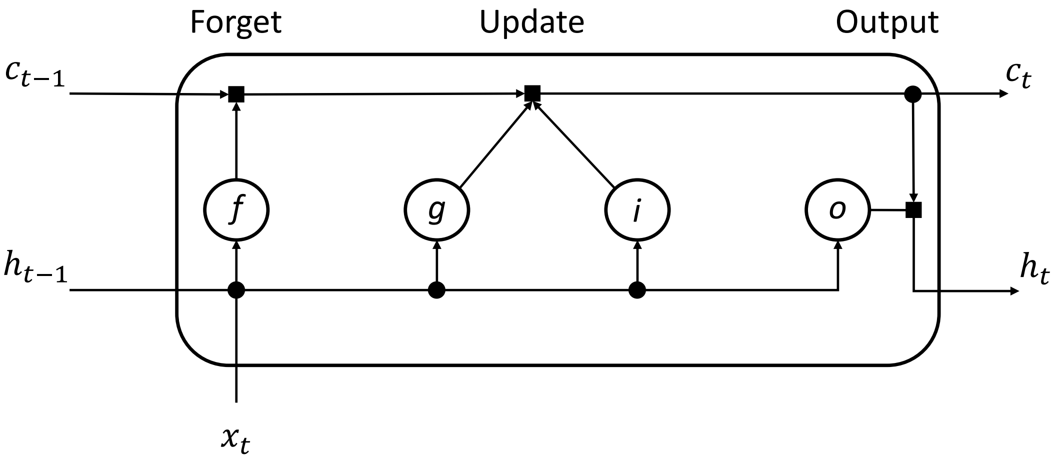


Fig. 4.12 Diagram of a single LSTM block: flow of the data at the time step t.

where the subscript i, f, g, and o denote the input gate, forget gate, cell candidate, and output gate, respectively. More concretely, the computation at time step t of each gates is defined as follow:

$$\begin{cases} \mathbf{i}_t = \sigma_g\left(\mathbf{W}_i\mathbf{X}_t + \mathbf{R}_i\mathbf{h}_{t-1} + \mathbf{b}_i\right) \\ \mathbf{f}_t = \sigma_g\left(\mathbf{W}_f\mathbf{X}_t + \mathbf{R}_f\mathbf{h}_{t-1} + \mathbf{b}_f\right) \\ \mathbf{g}_t = \sigma_c\left(\mathbf{W}_g\mathbf{X}_t + \mathbf{R}_g\mathbf{h}_{t-1} + \mathbf{b}_g\right) \\ \mathbf{o}_t = \sigma_g\left(\mathbf{W}_o\mathbf{X}_t + \mathbf{R}_o\mathbf{h}_{t-1} + \mathbf{b}_o\right) \\ \mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \\ \mathbf{h}_t = \mathbf{o}_t \odot \sigma_c\left(\mathbf{c}_t\right) \end{cases} \tag{4.27}$$

where $\odot$ denotes the element wise multiplication of vectors (Hadamard product). In this calculation, the activation function used for the input, forget and output gates is the sigmoid function defined as:

$$\sigma_g\left(x\right) = \frac{1}{\left(1 + e^{-x}\right)}, \tag{4.28}$$

while for the state activation function σ_c is used, representing the hyperbolic tangent function (tanh).

While LSTM is a well-known architecture among the RNNs able to deal with vanishing gradient problem, it could still be affected by exploding gradient. If the gradients increase in magnitude exponentially, the training loss may assume an indeterministic form. As a consequence then, the training becomes unstable and can diverge within a few iterations. In order to avoid this problem, gradient clipping can be used to prevent gradient explosion by stabilizing the training at higher learning rates and in the presence of outliers [122]. Gradient clipping enables networks to be trained faster, and does not usually impact the accuracy of the learned task. There are two types of gradient clipping:

- Norm-based gradient clipping rescales the gradient based on a threshold, and does not change the direction of the gradient. The *l2norm* and *global-l2norm* are norm-based gradient clipping methods.
- Value-based gradient clipping clips any partial derivative greater than the threshold, which can result in the gradient arbitrarily changing direction. Value-based gradient clipping can have unpredictable behaviour, but sufficiently small changes do not cause the network to diverge. The *absolute-value* is a value-based gradient clipping method.

While LSTMs possess the ability to learn temporal dependencies in sequences, they have difficulty with long term dependencies in long sequences. The solution proposed in [123] can help the LSTM RNN to learn these dependences. Combining a Bidirectional RNN (BRNN) with LSTM, it is possible to increase capacity of BRNNs by stacking hidden layers of LSTM cells in space, called deep bidirectional LSTM (BLSTM). In this way the output layer can get information from past (backwards) and future (forward) states simultaneously, making the BLSTM networks more powerful than unidirectional LSTM networks, by involving all information of input sequences in the computation.

4.5 Summary

In this chapter an overview of different deep learning techniques was provided. The most general architecture of deep learning, together with Convolutional Neural Network (CNN) and how it can be used in the radar domain were introduced in Section 4.2. A different architecture, namely Recurrent Neural Network (RNNs) was introduced in Section 4.3. At this purpose, different classes of RNNs and how train a RNN were also introduced in Section 4.3.3. Furthermore, in section 4.3.4 vanishing and exploding gradient problem were also tackled: since the performance of RNNs are affected by these problems, the Long Short-Time Memory (LSTM) has been finally introduced in section 4.4, with the aim to improve the deep learning capability for automatic classification purpose.

The concepts introduced in this chapter will be used to design a deep learning system able to detect and classify faulty speakers from its mechanical frequency response.

Chapter 5

Adaptive feedback cancellation algorithm for PA system

5.1 Introduction

The research presented in this chapter deals with signal processing algorithms in order to develop a robust downstream solutions of a loudspeaker production chain, providing an increased performance and sound quality for advanced acoustic systems in realistic conditions. In particular, in this chapter the problem of intelligibility of sound affected by the characteristics of typical acoustic environments is analysed, and a solution is presented. Audio intelligibility is often deteriorated due to the acoustic feedback problem. It occurs when the output signal of an audio device returns to its microphone and thereby forms an acoustic feedback loop. The typical consequences of acoustic feedback are sound quality degradation and, in the worst-case, howling. As stated in Chapter 2, acoustic feedback control in PA applications is often limited to the howling suppression, without modelling the room acoustics due to the complexity required to estimate the acoustic of a large venue.

Effective and robust methods to accomplish howling suppression are feedforward suppression techniques. However, these methods have significant limitations: not only is the achievable MSG limited, but also distortions are introduced in the loudspeaker signals affecting the sound quality. For this reason, room modelling methods are investigated. Typically, in the context of PA systems long impulse responses are involved (with reverberation time R60 from 1 to 5 seconds long [124]), and high sample frequencies are required in order to keep high the sounds quality. This leads to a high computational complexity in time domain algorithms.

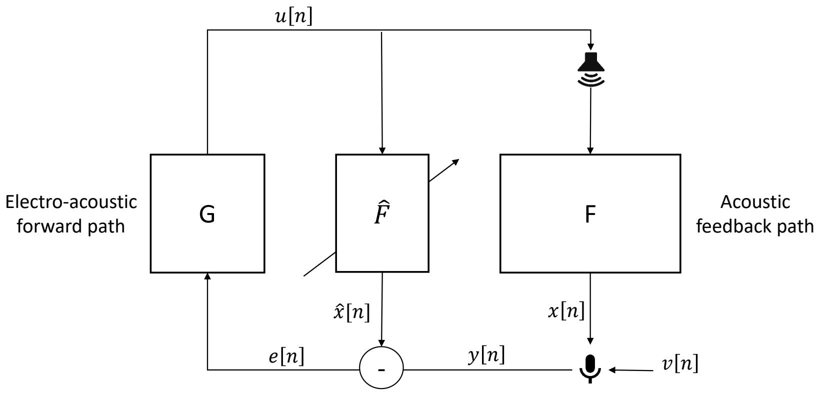


Fig. 5.1 Adaptive Feedback Cancellation (AFC) scheme in SISO scenario: the estimated feedback signal component $\hat{x}\,[n]$, obtained from $\hat{F}\,[q,n]$, is subtracted from the microphone signal $y\,[n]$.

Hence an approach able to provide an unbiased solution and a fast convergence rate is required.

The main scenario which acoustic feedback problem occurs is in hearing aids domain. Depending on the hearing aid style, the device is prone to the feedback and howling problems due to the close position of microphones and loudspeaker, typically only a few millimetres to a few centimetres apart. In this field, acoustic feedback cancellation using adaptive filter techniques (AFC) in a system identification configuration has become the state of the art method for reducing the effect of acoustic feedback [4, 125].

In Figure 5.1 a simple hearing aid system with an AFC system is illustrated, where an adaptive filter $\hat{F}\,[q,n]$ models and cancels the acoustic feedback path $F\,[q,n]$ from the hearing aid loudspeaker to the microphone. One of the challenges in using this basic AFC system is that whenever the correlation time of the incoming signal $v[n]$ is longer than the system latency (from the microphones to the loudspeaker) of the hearing aid, the signals $v[n]$ and $u[n]$ become correlated, and the adaptive filter estimate $\hat{F}\,[q,n]$ is biased. One of the solutions for this problem is therefore to reduce the correlation between the near end signal and the loudspeaker signal. There are different techniques to compensate for this biased estimation problem, introduced already in 2.5.1. A commonly used technique is based on Prediction Error Method (PEM), a prefilter used to whiten the component of the incoming signal $v[n]$ entering the adaptive filter estimation and thereby decorrelate it from $u[n]$. Practical PEM based AFC implementations often rely on computationally

simple time-domain stochastic gradient algorithms, such as the normalized least mean squares (NLMS) algorithm. Although this approach works most of the times, it may not be enough in other application domains, as PA systems.
In Chapter 2 the problem statement was introduced in detail, with its standard technique for the IR estimation. The aim of this chapter is to show a new framework for an unbiased estimation of long acoustic feedback paths, improving the sound intelligibility in scenarios such as public address systems. In Section 5.2, the Partitioned Block (PB) version of the traditional PEM based LMR-NLMS algorithm is introduced. The Partitioned Block approach consists of slicing the feedback path (e.g the impulse response of the system) to improve the algorithm performance. It can be applied either in the time domain or in the frequency domain, where the latter, called Partitioned Block Frequency Domain, shows faster convergence, lower computational cost and higher estimation accuracy. Finally, in Section 5.3 the results of the proposed framework are compared with the state of the art using real acoustic data showing superior performance in terms of Misalignment (MSL), Maximum Stable Gain (MSG) and less convergence time.

5.2 Proposed Method - PBTD and PBFD

In reverberant environments, PA systems face much longer impulse responses than the maximum $20ms$ experienced in hearing aids, usually longer than one second. Estimating a longer RIR means higher computational costs need to be accounted and a more efficient solution needs to be evaluated. In such scenarios, in order to achieve both fast convergence and low misalignment, a Partitioned Block Time & Frequency Domain (PBTD - PBFD) can be used, in combination with the Predicion Error Method (PEM) to decorrelate the loudspeaker signal. In case a priori knowledge of the IR is available, it could be used to increase the adaptive filter convergence speed [24]. Considering the scheme in Figure 5.2, the PEM based LMR-NLMS estimation of the RIR can be expressed as:

$$\hat{\mathbf{f}}\left[n\right] = \hat{\mathbf{f}}\left[n-1\right] + \mu \frac{\hat{\mathbf{R}}_{\mathbf{f},3}\tilde{\mathbf{u}}\left[n\right]\tilde{\epsilon}\left[n\right]}{\tilde{\mathbf{u}}^{T}\left[n\right]\hat{\mathbf{R}}_{\mathbf{f},3}\tilde{\mathbf{u}}\left[n\right]+\sigma_r^2}. \tag{5.1}$$

Let $\hat{\mathbf{f}}\left[n\right]$ be the estimated RIR of order $\hat{n}_F$ equal to n_F, order of the real RIR. With the partitioned block approach, the $\hat{n}_F$ taps feedback canceller $\hat{\mathbf{f}}\left[n\right]$ are

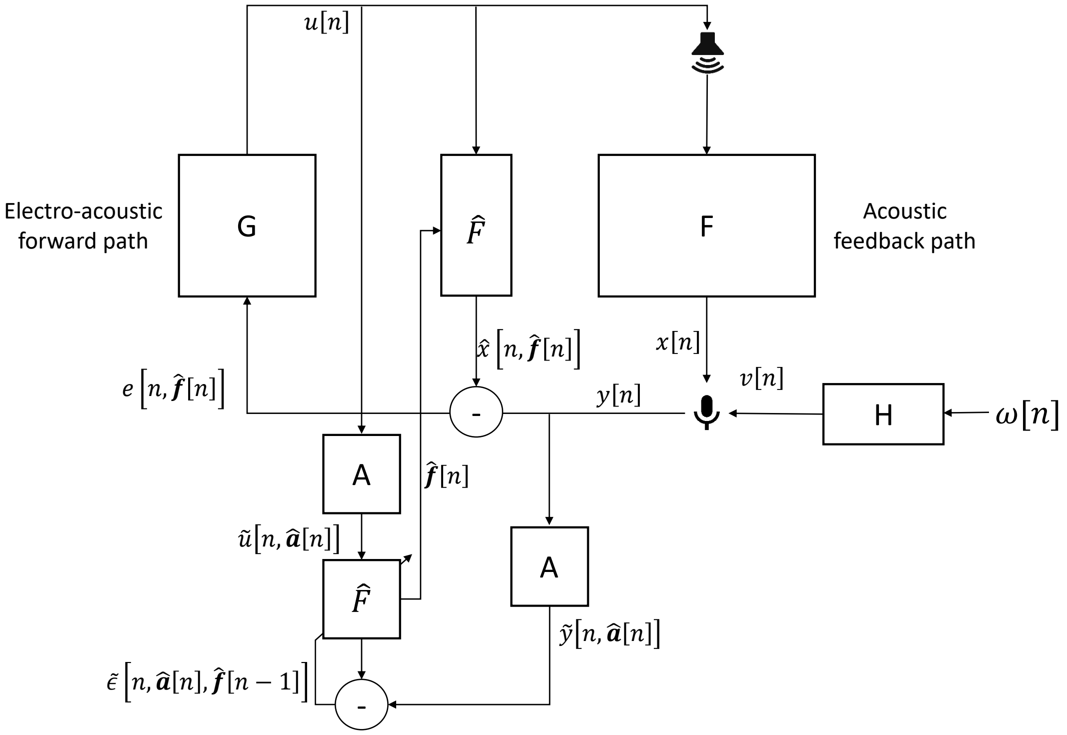


Fig. 5.2 Architecture of the Prediction Error Method (PEM) based Acoustic Feedback Cancellation (AFC), in a typical scenario.

partitioned into $\hat{n}_F/P^1$ segments $\hat{\mathbf{f}}_p[n]$ of length P each, such that:

$$\hat{\mathbf{f}}_p[n] = \left[\hat{f}_{pP}[n], \hat{f}_{pP+1}[n], \ldots, \hat{f}_{(p+1)P-1}[n]\right], \tag{5.2}$$

with $p = 0, \ldots, \frac{\hat{n_F}}{P} - 1$. In this case, the PBTD applied to PEM based LMR-NLMS algorithm in (2.71) will become:

$$\hat{\mathbf{f}}_p[n] = \hat{\mathbf{f}}_p[n-1] + \mu_p \frac{\mathbf{R}_{\mathbf{f}p}\tilde{\mathbf{u}}[n]\,\tilde{\epsilon}[n]}{\tilde{\mathbf{u}}^T[n]\,\mathbf{R}_{\mathbf{f}p}\tilde{\mathbf{u}}[n] + \sigma_r^2} \tag{5.3}$$

where $\mathbf{R}_{\mathbf{f}p}$ is the covariance matrix of the p-*th* IR block. An example of impulse response divided in p segments of length P each is shown in Figure 5.3. Usually, while moving on to the IR tail, the loop gain $|G(q,n)\,F(q,n)\,|$ will show a lower energy, thus producing a degraded estimation. To compensate this, a slower adaptation speed is preferable, leading to a choice of a Variable Step Size (VSS) μ_p, instead of a fixed one [126]. In order to get a faster convergence and a reduced complexity, a Partitioned Block Frequency Domain (PBFD) version of the algorithm has also been designed [45, 127, 126]. Applying the PB approach

[1] With $\hat{n}_F/P \in \mathbb{N}$, otherwise zero padding procedure is required.

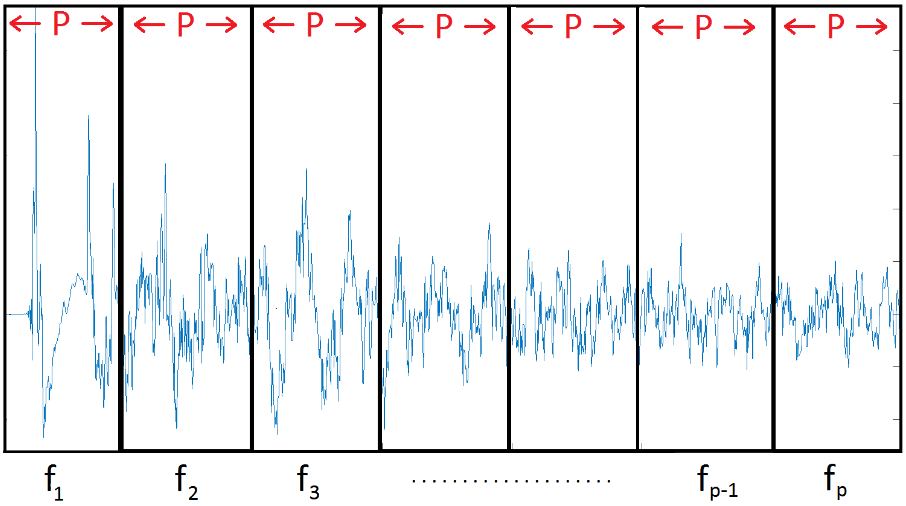


Fig. 5.3 Example of impulse response divided in p segments of length P each.

in frequency domain, an estimated frequency response can be obtained by the following equation:

$$\hat{\mathbf{F}}_{n,p} = \mathcal{F}_M \begin{bmatrix} \hat{\mathbf{f}}_p\left[n\right] \\ \mathbf{0} \end{bmatrix} \quad \text{with} \quad p = 0, \dots, \frac{\hat{n_F}}{P} - 1 \tag{5.4}$$

where $\mathcal{F}_M$ represents the $M \times M$ Discrete Fourier Transform (DFT) matrix. Defined the L-dimensional block of the loudspeaker signal $\mathbf{u}_n$ as:

$$\mathbf{u}_n = \Big[u\left[nL+1\right], \quad \dots \quad , u\left[\left(n+1\right)L\right]\Big]^T \tag{5.5}$$

with n the block time index. For each block $\mathbf{u}_n$ of input samples, the prefiltered version in frequency domain is obtained as:

$$\tilde{\mathbf{U}}_n = diag\left\{\mathcal{F}_M\left[\mathbf{A}\mathbf{u}_n\right]\right\} \tag{5.6}$$

For each time frame, M output samples of the prefiltered prediction signal are obtained:

$$\tilde{\mathbf{z}}_n = \Big[\mathbf{0} \quad \mathbf{I}_L\Big] \mathcal{F}_M^{-1} \sum_{p=0}^{\hat{n}_F/P} \tilde{\mathbf{U}}_n \hat{\mathbf{F}}_{n-1,p}, \tag{5.7}$$

The AR coefficients $\hat{\mathbf{a}}\left[n\right] = \left[1\, \hat{a}_1\left[n\right] \dots \hat{a}_{n_H}\left[n\right]\right]^T$ estimate the denominator coefficients of $H\left[q,n\right]$ and, equivalently, define the numerator coefficients of $A\left[q,n\right]$.

These coefficients are estimated using the Levinson-Durbin algorithm starting with the autocorrelation function of the prefiltered error signal [125], and used to compute the prefiltered version of both loudspeaker and microphone signals. To ensure correct operation, a constraint on the microphone samples R is applied in order to implement the PBFD with "overlap-and-save" method, a well-known technique for performing a linear convolution using FFT algorithms. By overlapping elements of the data sequences and retaining only a subset of the final DFT product, a linear convolution between a finite length sequence and an infinite-length sequence is obtained. In this case, the frequency-domain weight vector corresponds to the finite-length sequence, and the input signal corresponds to the infinite-length sequence. In order to generate P correct output samples, it will be necessary to use FFT of length $M \geq 2P - 1$. It turns out that DFT with $2P$ points ($M = 2P$) are suitable for this purposes and that the optimal block size is $R = L = P$ [128]. Thus, the adaptive filter coefficients are updated using an "overlap-and-save" method, producing (in the frequency domain) the prefiltered error signal:

$$\tilde{\mathbf{E}}\left[n\right] = \mathcal{F}\begin{bmatrix}\mathbf{0}\\ \tilde{\boldsymbol{\epsilon}}_n\end{bmatrix}, \tag{5.8}$$

where

$$\tilde{\boldsymbol{\epsilon}}_n = \tilde{\mathbf{y}}_n - \tilde{\mathbf{z}}_n \tag{5.9}$$

is the prefiltered error signal, with

$$\tilde{\mathbf{y}}_n = \left[\tilde{y}\left[nR+1\right], \ldots, \tilde{y}\left[(n+1)\,R\right]\right]^T. \tag{5.10}$$

the prefiltered microphone signal of R samples. Thus, for each time frame m, the algorithm produces $\hat{n}_F/P$ segments $\hat{\mathbf{f}}_p\left[n\right]$ of the estimated feedback path, generating $L = R$ output samples of the estimated feedback signal. Consequently, the error signal frame can be defined as the difference between the microphone signal frame and the estimated feedback signal frame:

$$\mathbf{e}\left[n\right] = \mathbf{y}_n - \hat{\mathbf{x}}\left[n\right] \tag{5.11}$$

In order to take full advantages of the a-priori knowledge of the scenario, such as public transportation facilities, live and recorded music venues, or any other venues where a reference impulse response is available, a constraint adaptation can be used, leading to the PEM based PBFD with VSS updating rule of the IR coefficients (in the frequency domain) [126]:

$$\begin{aligned}\hat{\mathbf{F}}_{n,p} =& \hat{\mathbf{F}}_{n-1,p} + \boldsymbol{\Delta}\left[n-1\right]\left[\mathcal{F}\mathbf{g}\mathcal{F}^{-1}\tilde{\mathbf{U}}_{n}^{H}\left[n\right]\tilde{\mathbf{E}}\left[n\right]\right] + \\ & - \boldsymbol{\Delta}\left[n-1\right]\left[\boldsymbol{\eta}\left(\hat{\mathbf{F}}_{n,p} - \mathbf{F}_{p_{ref}}\right)\right],\end{aligned} \tag{5.12}$$

where $\boldsymbol{\eta}$ is a diagonal matrix containing the trade off parameter η_k, with $k = 0, \ldots, \hat{n}_F/P$, $\mathbf{g}$ is an adaptation matrix build as

$$\mathbf{g} = \begin{bmatrix} \mathbf{I}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{M-P} \end{bmatrix}, \tag{5.13}$$

$\mathbf{F}_{p_{ref}}$ represents the FFT of reference measure of the $p-th$ block of the IR. In the frequency domain approach we express the regularization function as:

$$\boldsymbol{\Delta}\left[n\right] = diag\{\mu_0\left[n\right], \ldots, \mu_{M-1}\left[n\right]\}, \tag{5.14}$$

which is the diagonal matrix containing the VSS $\mu_k\left[n\right]$:

$$\mu_k\left[n\right] = \frac{\overline{\mu_k}}{|\tilde{E}_k\left[n\right]|^2 + |\tilde{U}_k\left[n\right]|^2 + \delta}. \tag{5.15}$$

The normalization is used to reduce the excess error in presence of signals with large power fluctuation. The denominator is the sum of the input power with the error power plus a small positive number δ to avoid division by zero.
Involving both the PEM and the a-priori knowledge into the PBFD with VSS, the bias into the estimation of the IR of large acoustic space has been drastically reduced. In Section 5.3 the relative performance of the new algorithms will be compared.

5.3 Performance Analysis

To assess the performance of the proposed algorithm, the *Misalignment* factor (MSL) and *Maximum Stable Gain* (MSG) are considered. The former is used to track the discrepancy between the true and the estimated feedback path, and it is defined as:

$$MSL\left(n\right) = \frac{\|\hat{\mathbf{f}}\left(n\right) - \mathbf{f}\left(n\right)\|}{\|\mathbf{f}\left(n\right)\|}. \tag{5.16}$$

The latter is the maximum allowable gain, assuming a flat frequency response of $G\left(q, n\right)$, as defined already in Chapter 2. From Equation 2.25, it immediately follows that a better estimation of the IR yields a larger MSG.

Table 5.1 Input parameters of the algorithm.

Parameters	**Value**
IRs length	1.34s
f_s	16kHz
n_F	21447
Source signal duration	10s
μ_{fix}	0.2
VSS μ	[0.2 0.08 0.02]
IR block length P	160
Source signal block length V	320
DFT coefficients M	320
PEM filter order n_A	30

In order to simulate the acoustic feedback, a pre-recorded female voice speech sampled at $f_s = 16$kHz of 10 seconds is considered as input source signal. For all the simulations the prediction error method is used, with filter order $n_A = 30$. The scenario in analysis is an auditorium of Tannoy Ltd, where IRs have been measured using a sine sweep method [90], with a microphone placed 3 meters far away from the loudspeaker, varying both microphone and loudspeaker positions into the room. The IRs length were approximately $1.34s$ long each, equal to $n_F = 21447$ samples. An overlap-and-save method has been used to implement the PBFD; the IR block size has been set with $P = 160$, together with a $M \times M$ DFT matrix with $M = 2P$. Since speech is considered to be stationary during $20ms$ frames, the block length of the source signal per each frame has been setted with $V = 320$ samples. A summary of the input parameters used for the algorithms are reported in Table 5.1.

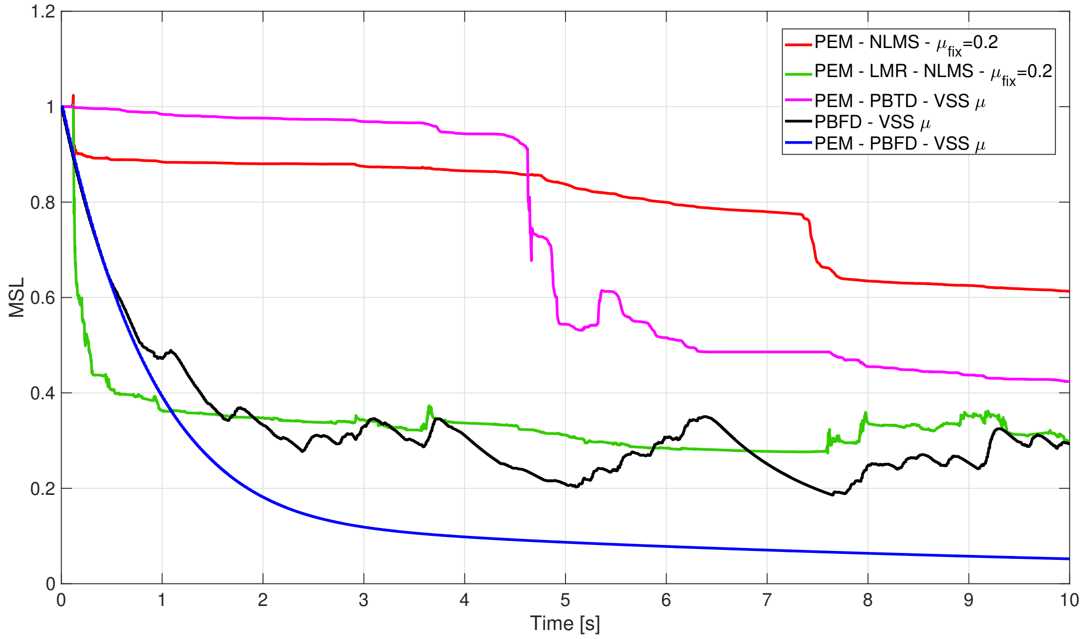


Fig. 5.4 Misalignment factor (MSL) of the AFC with a female speech as input signal of 10s long with $fs = 16kHz$, $n_F = 21447$, and $n_A = 30$: comparison among the NLMS and LMR-NLMS algorithms with fixed $\mu = 0.2$, and PBTD, PFDF and PEM-PBFD algorithms with VVS $\mu = [0.2\,0.08\,0.02]$.

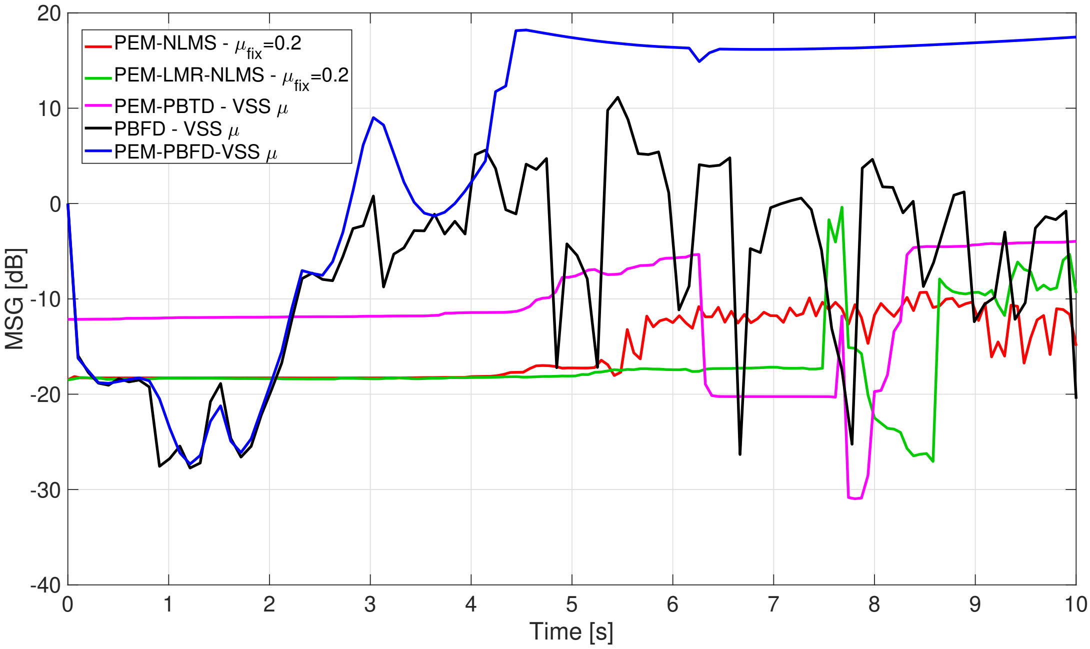


Fig. 5.5 Maximum Stable Gain (MSG) of the AFC with a female speech as input signal of 10s long with $fs = 16kHz$, $n_F = 21447$, and $n_A = 30$: comparison among the NLMS and LMR-NLMS algorithms with fixed $\mu = 0.2$, and PBTD, PFDF and PEM-PBFD algorithms with VVS $\mu = [0.2\,0.08\,0.02]$.

In Figures 5.4 and 5.5, both MLS factor and MSG of the different approaches are respectively compared. The basic PEM based NLMS algorithm is compared with the PEM based LMR-NLMS algorithm, both with a fixed step size $\mu_{fix} = 0.2$. The PEM based NLMS algorithm shows a high bias against the estimated IR, leading to a MSL equal to 0.61 and MSG equal to $-14.9dB$. Taking in account the a-priori knowledge through the covariance matrix $\mathbf{R_f}$ given by Sabine model 2, the bias is reduced with a MSL score equal to 0.3 and MSG equal to $-9.4dB$. In order to improve the performance, the partitioned block version of the LMR-NLMS algorithm with VSS in the time domain has been considered and compared with the previous one. In this case, using the VSS as control parameter, the performance does not show an improvement of estimated RIR since it shows a slower convergence rate. It requires a longer time to achieve a smaller MSL value. After 10s, an MSL equal to 0.42 and MSG equal to $-3dB$ are achieved. In the Figures 5.4 and 5.5, the PBFD with and without PEM are also compared. Clearly, the proposed PEM based PBFD with VSS version outperforms all other algorithms. It is important to stress that despite the PEM not being very relevant for the estimation of the first IR slices, it then becomes essential on the estimation of the IR tail, where the energy is lower: comparing the last blocks of the estimated and real RIR, highly biased estimation is achieved. While the PBFD with VSS achieves a MSL value of 0.29 and MSG of $-20dB$, the PEM based PBFD with VSS algorithm outperforms all the other algorithm with a MSL equal to 0.05 and MSG of $18dB$.

In order to show the re-adaptation capability of the proposed PEM based PBFD with VSS algorithm, a non-stationary feedback path scenario is simulated with the results reported in Figure 5.6 and 5.7. In this case the impulse response has been changed after 5 seconds of input source signal in order to simulate a change of position between the microphone and the loudspeaker (e.g. a speaker moving on the stage) and consequently a change of the impulse response. In particular, the results show the ability of the algorithm to re-adapt the coefficient of the estimated RIR, leading to a misalignment equal to 0.21 with a maximum stable gain of $10dB$, after 2 seconds from the change of the IR.

Finally the execution time of the AFC algorithms have been evaluated and reported in the Table 5.2. It can be concluded from Figures 5.4 and 5.5 and from the Table 5.2, that the PBTD shows a slower convergence rate and a high computational cost. On the contrary, the frequency domain method avoids both problems wherein the PBFD substantially decreases the computational burden, and shows

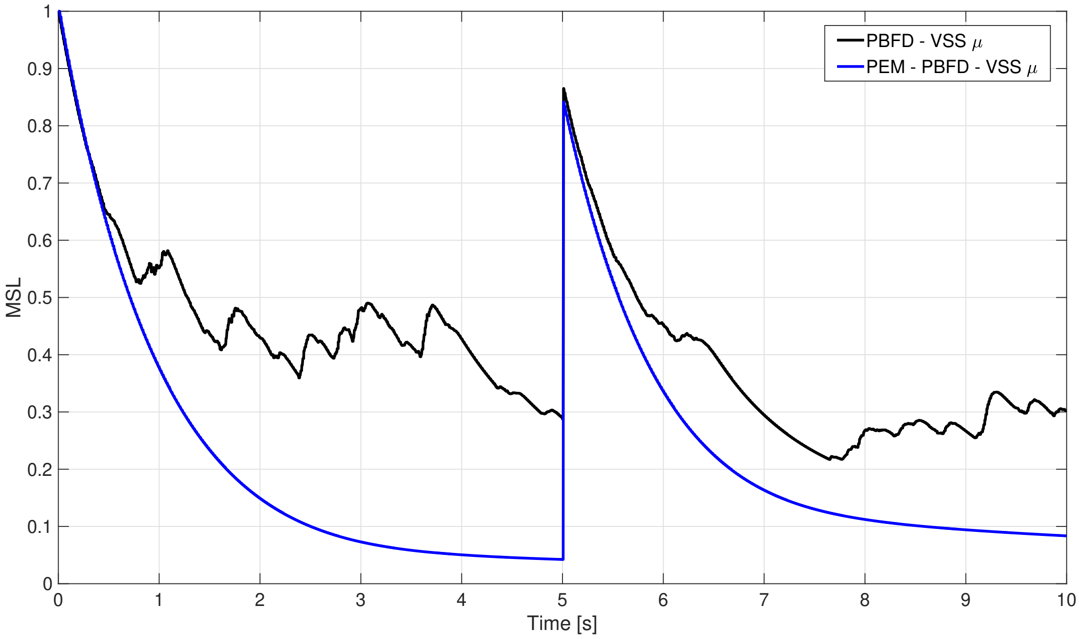


Fig. 5.6 Misalignment factor (MSL) of the AFC with a female speech as input signal of $10s$ long with $fs = 16kHz$, $n_F = 21447$, $n_A = 30$ and variable step size VSS $\mu = [0.2\,0.08\,0.02]$: comparison between PBFD with and without PEM in a non stationary feedback path scenario.

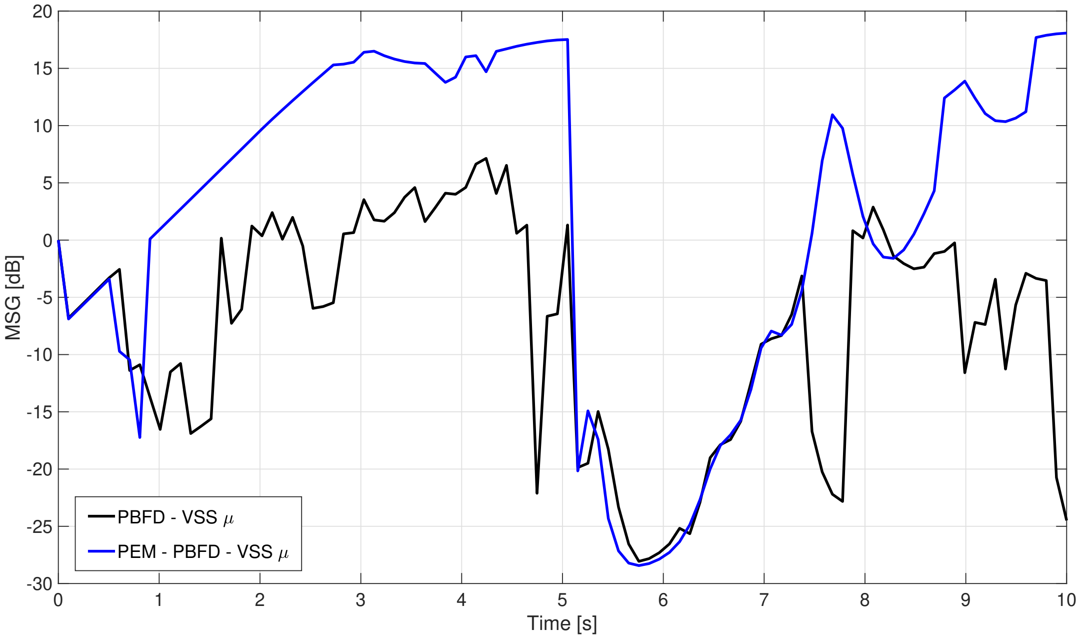


Fig. 5.7 Maximum Stable Gain (MSG) of the AFC with a female speech as input signal of $10s$ long with $fs = 16kHz$, $n_F = 21447$, $n_A = 30$ and variable step size VSS $\mu = [0.2\,0.08\,0.02]$: comparison between PBFD with and without PEM in a non stationary feedback path scenario.

Table 5.2 Execution time of the AFC algorithms with a female speech signal of $10s$ long as input signal, with $fs = 16kHz$, $n_F = 21447$, $n_A = 30$, with Matlab R2018a on pc with 8GB Ram and a CPU Intel i5 Quad Core 4^{th} generation.

Algorithm	**Elaboration Time** [s]
PEM based NLMS	230
PEM based LMR-NLMS	249.4
PEM based PBTD with VSS	16417
PBFD with VSS	153
PEM based PBFD with VSS	205

a convergence rate faster than the all other algorithms, thus becoming a suitable choice for real time implementations.

5.4 Summary

In this chapter, a new framework to tackle the acoustic feedback problem in large acoustic spaces was presented. It is based on the Frequency Domain Adaptive Filtering (FDAF) implementation of the Normalized Least Mean Square (NLMS) algorithm. Since the traditional LS-based adaptive filtering algorithm converge to a biased solution of the acoustic feedback path due to a considerable correlation between loudspeaker and microphone signals, a signal decorrelation method has been used. Inspired by hearing aids device, the Prediction Error Method (PEM) was introduced. In order to further decrease the bias into the estimated feedback path the a-priori knowledge was considered. Based on acoustic set-up information (distance between loudspeaker and microphone, acoustic absorption of the walls, room volume, etc.), a robust estimator of the RIR covariance matrix can be obtained from Sabine 3-parameter RIR model. Applying the Levenberg-Marquardt Regularization (LMR), the PEM based LMR NLMS was obtained. Dealing with long impulse response, meaning higher computational costs, a more efficient solution needed to be evaluated. For this reason, the Partitioned Block approach has been considered. Moving towards the IR tail, the loop gain $|G(q,n)\,F(q,n)\,|$ showed a lower energy, thus producing a degraded estimation. To compensate this, a slower adaptation speed was used, leading to the PBTD version of the algorithm with Variable Step Size. Performance analysis have shown that PBTD version with Variable Step Size has a slower convergence rate and a higher computational cost than PEM based LMR-NLMS algorithm. A faster convergence was achieved by designing the algorithm in the frequency domain. Finally, in order to take the full advantages of the a-priori knowledge of the scenario, such as public transportation facilities, live and recorded music venues, or any other venues where a reference impulse response is available, a constraint adaptation could be used, leading to the proposed PEM based PBFD with VSS. The results showed that this technique outperform previous approaches, with superior performance in terms of Misalignment (MSL), Maximum Stable Gain (MSG), achieving up to 18dB of MSG and 30 seconds less convergence time. However, the proposed algorithm has significant limitations. Due to the simplicity of source signal model, the algorithm showed good results on pre-recorded speech signal. In case of sound or music signals are considered, the algorithm would not be able to achieve the same performance: a more complex source signal model should be used in this case. Furthermore, a low misalignment and high gain is achieved by considering the a priori knowledge of the RIR. In case a priori knowledge is not available, the regularization methods could no be applied leading to a degraded estimation of the feedback path.

Moving on the upstream side of loudspeaker production chain, the novel approach for loudspeaker EOL test based on radar micro-Doppler analysis is introduced in the next chapter.

Chapter 6

Micro-Doppler analysis of loudspeakers

6.1 Introduction

In Chapter 5 a new downstream solution to acoustic feedback was presented. In this chapter a novel solution is provided from the upstream point of view. To successfully accomplish this task, the concepts of micro-Doppler in radar introduced in Chapter 3 are exploited and applied to loudspeaker condition monitoring.

Audio intelligibility is not only affected by the characteristics of typical acoustic environments, but also from physical defect of a speaker. For this reason, the novel use of radar micro-Doppler for loudspeaker analysis is proposed for the first time. This approach offers the potential benefits to characterize the mechanical motion of a loudspeaker in order to identify defects and design issues. One of the important topics in audio manufacturing is the loudspeaker condition monitoring. Increasing quality checks at various stages of production (with limited costs) can provide substantial benefits to loudspeaker manufacturers. As reported in Section 3.6, laser based analysis tools have been shown to yield significantly better results compared to traditional acoustic ones [89]. However, both acoustic and laser analysis have technical and practical limitations that do not apply to the use of our radar based method, as shown along the chapter.

The effectiveness of acoustic End-Of-Line tests (EOL) or acoustic measurements is limited by the surrounding environment; as it normally requires specifically designed insulated booths or silent areas for the signal-to-noise ratio of audio data to be meaningful. There are two main limitations when laser-based scanner vibrometer systems (Scanning Vibrometer System (SCN) [89]) are used in place of the traditional acoustic approach. The first is the requirement of a very large sets of measurements (up to almost 3000 points) to fully characterize a loudspeaker and its non linearities, thus being a serious time consuming activity. The second is the limitation due to the presence of any physical obstacle in the line of sight between the laser source and the membrane (or acoustic source) under test [85, 86].

The interest in micro-Doppler suggests that this technology is reliable, and that it is worth investigating it in further applications domains, such as the acoustic monitoring. In [129], the authors introduce for the first time a novel approach based on radar micro-Doppler to analyse and measure the return from loudspeaker. This approach was motivated by the potential cost effectiveness and operational advantages that a radar based approach could introduce over acoustic and laser based ones. With respect to the traditional acoustic measurement, a radar based approach is not affected by the acoustic environmental factors, allowing its use for End of Line (EoL) test. Unlike the SCN system, the micro-Doppler has the ability to cope with visual occlusion due to plastic parts and the capability of separation metallic components of a loudspeaker from non metallic ones through the use of the back-scattering intensity.

In this chapter a novel approach to measure loudspeaker characteristic will be proposed, exploiting low cost radar sensors and the micro-Doppler signatures. The novelty introduced in this chapter can be summarised as follow:

- A model for the radar return from a loudspeaker based on the Thiele and Small parameters.
- A methodology to measure mechanical frequency response of loudspeakers in order to characterise the speaker with a single radar measurement.

The remainder of the chapter is organised as follows. Starting from the concept of micro-Doppler introduced in Chapter 3, together with loudspeaker kinematic, the micro-Doppler signature of a speaker playing a single tone is modelled and analysed in Section 6.2.1, while in Section 6.2.2 the micro-Doppler signature of speaker playing a chirp signal is investigated. By using a chirp signal, the concept of mechanical frequency response is introduced in order to characterize the speaker with a single measurement, in Section 6.2.3. Results of experimental acquisitions are compared with the simulated data in Section 6.3.1 and 6.3.2, for both single tone and sine sweep analysis, in order to validate the expected micro-Doppler modulations. In Section 6.3.3 the matched filter approach is also applied to real data. Finally in Section 6.4, conclusions and future developments are proposed.

6.2 Micro-Doppler Analysis

An operating loudspeaker presents a complex scenario of moving parts which can generate a multifaceted pattern of vibrations. A radar sensor can be used to identify the vibration pattern of the transducer. From the basic concept of radar micro-Doppler and loudspeaker kinematic, introduced respectively in Section 3.3 and 3.6, the micro-Doppler signature of a speaker playing single tone signal is analysed in Section 6.2.1. To fully understand and interpret correctly the radar micro-Doppler phenomena in complex and realistic scenario, the exponential sine sweep in Equation (3.88) is considered, with the relative micro-Doppler signature analysed in Section 6.2.2. Finally, in Section 6.2.3 the radar based mechanical characterization of the speaker will be introduced. For all these analyses, a sampling frequency $f_s = 22$kHz is considered, in order to compare the simulated signals with real signals acquired with the available hardware.

6.2.1 Single tone Analysis

Radar micro-Doppler effect can be understood by considering target's micro-motions. As stated in Section 3.2, in coherent radars, the range variations cause a phase change in the returned signal from a target. Thus, the Doppler frequency shift, representing the change of phase function over time, can be used to detect vibrations or rotations of structures in a target. In Figure 6.1 the geometry used to analyse the micro-Doppler induced by a loudspeaker is shown [66, 129]. The

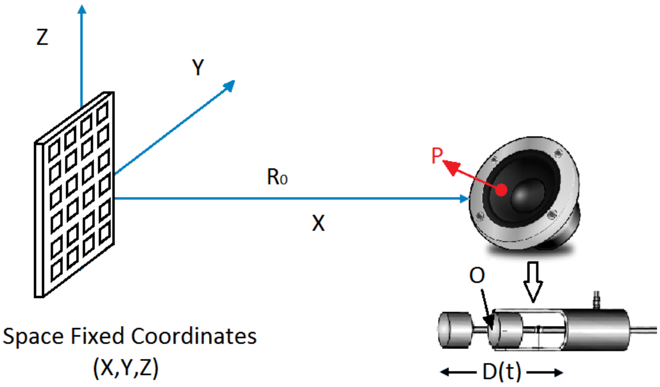


Fig. 6.1 Geometry for the radar and generic vibrating point: the motion of a speaker can be described as rigid body motion having a piston mode when the input to the loudspeaker is a signal with frequency range up to 1kHz.

signal from a target as a function of time is modelled as follows:

$$s_r\left(t\right) = \rho \exp\left\{j\left[2\pi f_0 t + \Phi\left(t\right)\right]\right\} \tag{6.1}$$

where ρ is the reflectivity of the vibrating point scatterer, f_0 is the carrier frequency of the transmitted signal and $\Phi\left(t\right)$ is the time varying phase change of the vibrating scatterer. Letting R_0 be the distance between the radar and the speaker's initial position O, then the range function varies with time due to the speaker micro-motion:

$$R\left(t\right) = R_0 + D\left(t\right) \tag{6.2}$$

Assuming an arbitrary point of the cone located in P vibrates with sinusoidal frequency f_v and maximum cone displacement $\tilde{\eta}_c\left(f_v\right)$ defined by Equation 3.86, the displacement function will be of the kind:

$$D\left(t\right) = \tilde{\eta}_c\left(f_v\right)\sin\left(2\pi f_v t\right) \tag{6.3}$$

while assuming the radar being in the line of the sight with the speaker [64, 66]. Then, the time varying phase can be written as:

$$\Phi\left(t\right) = \beta_{wn} R\left(t\right) = \beta\left[R_0 + \tilde{\eta}_c\left(f_v\right)\sin\left(2\pi f_v t\right)\right] \tag{6.4}$$

where the angular wavenumber is

$$\beta_{wn} = \frac{4\pi}{\lambda} \tag{6.5}$$

with λ the wavelength of the transmitted signal. Substituting (6.4) in (6.1) the received signal can be expressed as [129]:

$$\begin{aligned} s_r\left(t\right) = &\rho \exp\left\{j\frac{4\pi R_0}{\lambda}\right\} \\ &\exp\left\{j2\pi f_0 t + j\frac{4\pi\tilde{\eta}_c\left(f_v\right)}{\lambda}\sin\left(\omega_v t\right)\right\} \end{aligned} \tag{6.6}$$

where $\omega_v = 2\pi f_v$. In order to simulate a received radar signal, the backscattering coefficient ρ [64, 66] relative to the only vibrating metallic component, namely the voice coil, is modelled as flat circular plate and calculated as:

$$\rho = \frac{4\pi^3 r^4}{\lambda^2} \tag{6.7}$$

with λ the radar signal wavelength and r is the radius of voice coil. From (6.6), the derivative of the second phase term leads to the expression of the micro-Doppler shift:

$$f_{mD}(t) = \frac{1}{2\pi}\frac{d\Phi}{dt} = \frac{4\pi}{\lambda}\tilde{\eta}_c(f_v)\, f_v \cos(2\pi f_v t) \tag{6.8}$$

In Figure 6.2 the theoretical micro Doppler of a speaker moving at its resonance frequency of 67Hz, with output voltage of 5V and 10V, is shown. From (3.86), the theoretical displacement $\tilde{\eta}_c$ at 5V and 10V of output voltage is computed. With a $\tilde{\eta}_{c,5V} = 1.1$mm and $\tilde{\eta}_{c,10V} = 2.2$mm, the maximum Doppler shift achievable is 73.90Hz and 147.80Hz, respectively. The spectrum of a typical simulated received radar signal is shown in the Figure 6.3, for 5V and 10V of applied voltage. By Fourier analysis the vibration frequency of the coil can be detected, where the number of visible harmonics depends on the displacement amplitude, directly related to the micro Doppler. This result is in complete agreement with loudspeaker modelling theory.

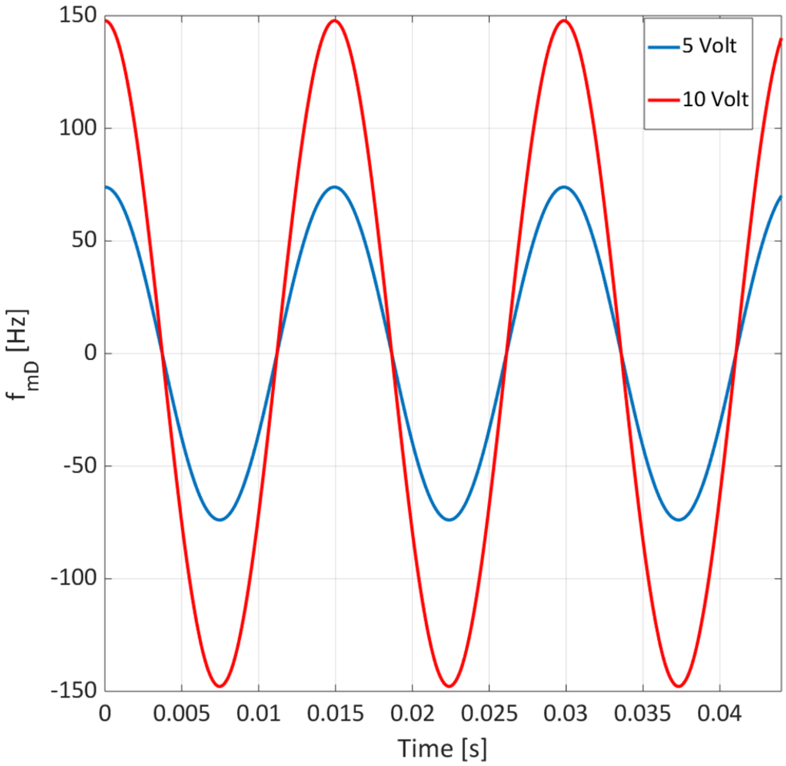


Fig. 6.2 Theoretical micro Doppler of a speaker moving at its resonance frequency of $f_v = f_r = 67$Hz with output voltage of 5V and 10V, modelled as a flat circular plate having maximum displacement of $\tilde{\eta}_{c,5V} = 1.1$mm and $\tilde{\eta}_{c,10V} = 2.2$mm.

As stated in Section 3.6, loudspeakers and other kinds of actuators which produce sounds or vibrations behave differently at small and high displacement amplitudes.

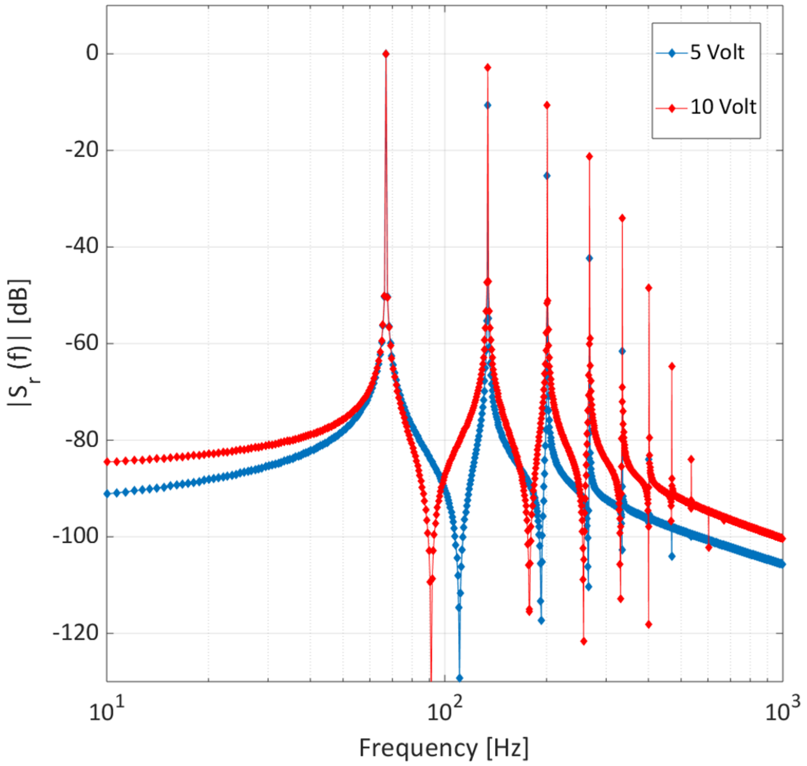


Fig. 6.3 Normalised spectrum of the simulated received signal of a speaker moving at its resonance frequency of $f_v = 67$Hz with output voltage of 5V and 10V, modelled as a flat circular plate having maximum displacement of $\tilde{\eta}_{c,5V} = 1.1$mm and $\tilde{\eta}_{c,10V} = 2.2$mm.

The dependency of the displacement amplitude is an indication of non linearities inherent in the system. As the displacement amplitude increases, particularly at low frequencies, the most dominant non linearities effects are introduced by stiffness $K_{ms}(\tilde{\eta}_c)$ (reciprocal of the compliance $C_{ms}(\tilde{\eta}_c)$), force factor $Bl(\tilde{\eta}_c)$ and inductance $L_e(\tilde{\eta}_c)$, function of the displacement $\tilde{\eta}_c$ [84]. A second non linear effect is the generation of additional spectral components which are not in the exciting stimulus; these components are generally multiple of the fundamental frequency and thus labelled as harmonic and intermodulations distortion[86].

From Equations (6.4),(6.6) and (6.8) that describe the micro-Doppler signature we can deduce the capability to retrieve information about the behaviour, anomalies and failures of a loudspeaker from the radar returned. While the spectral composition of a signal varies as function of the time, the conventional Fourier transform cannot provide a time dependent spectral description. Thus, a joint time-frequency distribution, introduced in Section 3.3.1, provides more insight into the time-varying behaviour of the signal. In the Figures 6.4 and 6.5, spectrograms of

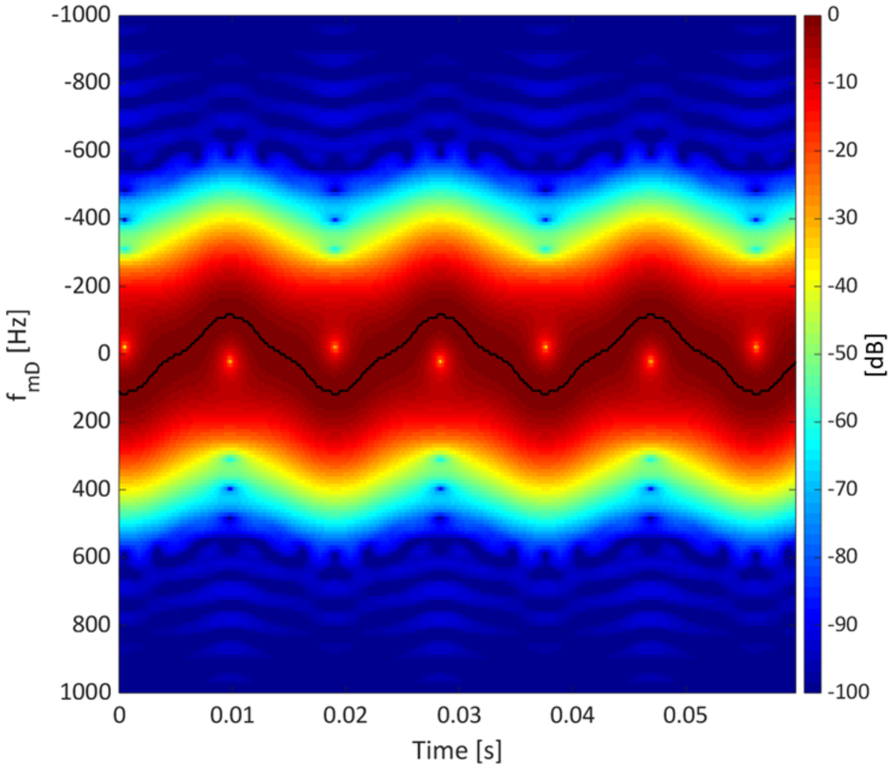


Fig. 6.4 Spectrogram with Blackman-Harris window of 11.6ms of the simulated received radar signal from a speaker moving at its resonance frequency $f_v = f_r =$ 67Hz with output voltage of 5V, modelled as a flat circular plate having maximum displacement $\tilde{\eta}_{c,5V} = 1.1$mm. The maximum Doppler shift is highlighted with a black line.

the simulated received radar signal of a speaker moving at its resonance frequency $f_v = f_r = 67$Hz with output voltage of 5V are shown. For a better understanding, the maximum Doppler shift is highlighted with a black line. In Figure 6.4, the spectrogram with a Blackman-Harris window of 11.6ms, confirming the sinusoidal-like motion. Due to the trade off between time-frequency resolution, the Doppler shift is bigger than the theoretical one (approximately 10Hz). A better frequency resolution can be achieved by increasing the window length to 23.2ms, as shown in Figure 6.5 where a Doppler shift of 75Hz is found, in agreement with the theoretical one. With an output voltage of 10V, the spectrogram in Figure 6.6 reveals a maximum Doppler shift of 150Hz, in agreement with what is expected from theory in (6.8).

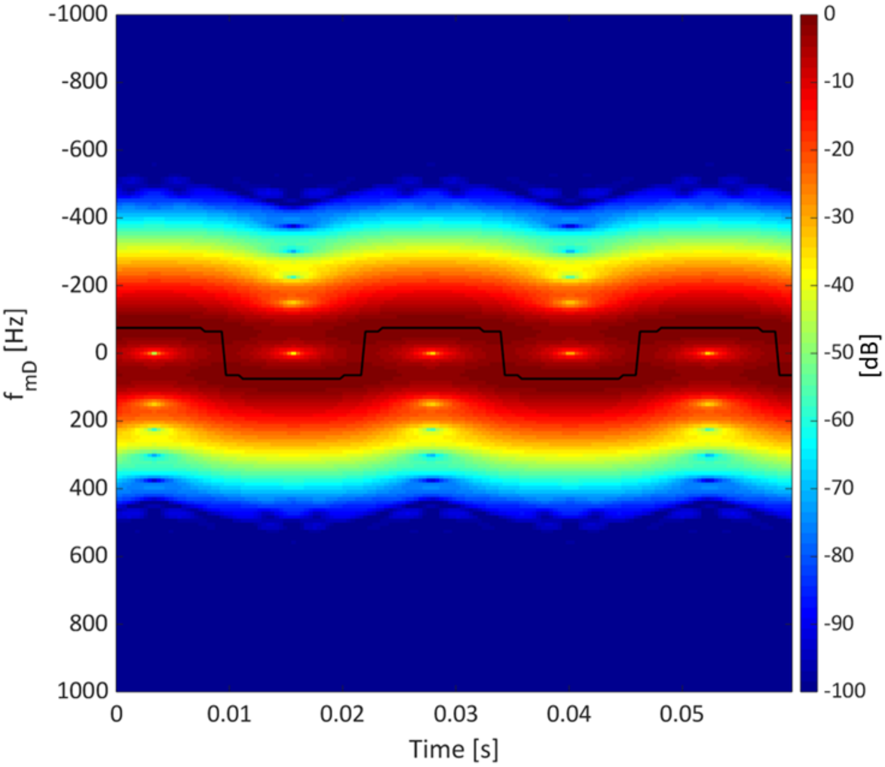


Fig. 6.5 Spectrogram with Blackman-Harris window of 23.2ms of the simulated received radar signal from a speaker moving at its resonance frequency $f_v = f_r = $ 67Hz with output voltage of 5V, modelled as a flat circular plate having maximum displacement $\tilde{\eta}_{c,5V} = 1.1$mm. The maximum Doppler shift is highlighted with a black line.

6.2.2 Sine Sweep Analysis

Using the chirp signal $x\left(t\right)$ in Equation (3.88), the ideal received radar signal $s_r\left(t\right)$ in baseband is modelled as:

$$s_r\left(t\right) = \rho \exp\left\{j\frac{4\pi\tilde{\eta}_c\left(f_v\left(t\right)\right)}{\lambda}x\left(t\right)\right\} \tag{6.9}$$

with the displacement $\tilde{\eta}_c$ be a time varying function of $f_v\left(t\right)$, as described in Equation (3.86). For an exponential sine sweep, the instantaneous vibration frequency $f_v\left(t\right)$ is defined as:

$$f_v\left(t\right) = f_1 k^t = f_1\left(\frac{f_2}{f_1}\right)^{\frac{t}{T}} \tag{6.10}$$

with k the exponential chirp rate. Depending on the behaviour of the displacement, the micro Doppler will show a different envelope, strictly related to voice coil

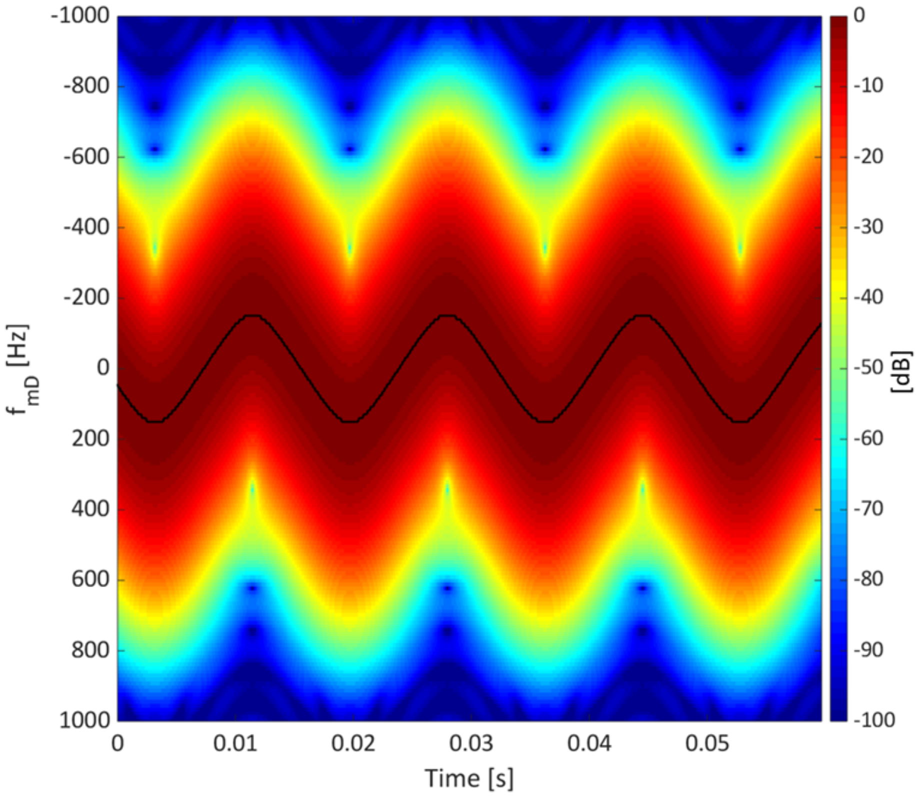


Fig. 6.6 Spectrogram of the simulated received radar signal from a speaker moving at its resonance frequency $f_v = f_r = 67$Hz with output voltage of 10V, modelled as a flat circular plate having maximum displacement $\tilde{\eta}_{c,10V} = 2.2$mm, with Blackman-Harris window of 5.8ms. The maximum Doppler shift is highlighted with a black line.

motion. In case of constant displacement $\tilde{\eta}_c$ during the sweep, the maximum micro Doppler increases linearly with the frequency of the stimulus, with fixed modulation index $\Upsilon = \beta\tilde{\eta}_c$, as in (6.8). In Figure 6.7 the maximum micro Doppler for different fixed displacement at different vibration frequencies are shown. Considering a fixed displacement $\tilde{\eta}_c = 10$mm with an exponential chirp of $T = 60$ seconds long in the frequency band $f_v \in [20, 1000]$Hz, the theoretical maximum micro-Doppler shift achievable is 10kHz. The spectrogram of the simulated received radar signal is shown in Figure 6.8. The sinusoidal like motion of the micro Doppler is visible at low vibration frequency. Due to high velocity of the target at high vibration frequency, only the maximum Doppler shift is visible. Furthermore, at high vibration frequency, it is easier to distinguish different harmonics components of the stimulus. In a more realistic scenario the voice coil, modelled as in Equation (3.86), can be considered constant before the resonance frequency, while after it decreases as the square of vibration frequency f_v. Then, it is necessary consider

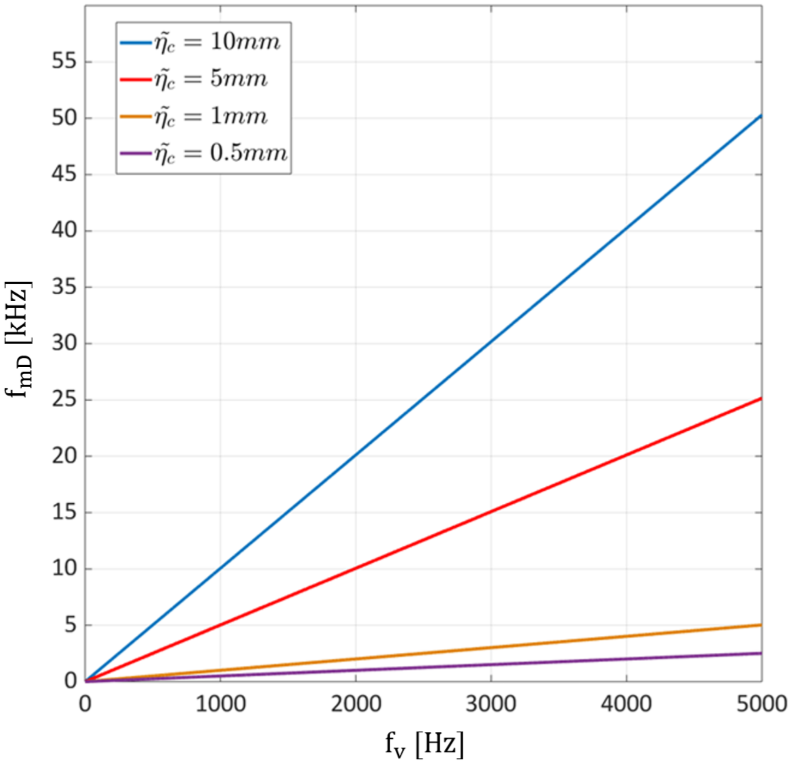


Fig. 6.7 Theoretical micro Doppler for different displacement $\tilde{\eta}_c$ at different vibration frequencies f_v, with a fixed wavelength $\lambda = 1.25$cm.

both displacement and vibration frequency as function of the time. With this assumption the theoretical micro Doppler equation will become the sum of two components, namely:

$$f_{mD}\left(t\right) = \frac{2}{\lambda}\frac{d\tilde{\eta}_c\left(t\right)}{dt}x\left(t\right) + \frac{4\pi f_v\left(t\right)}{\lambda}\tilde{\eta}_c\left(t\right)\frac{dx\left(t\right)}{dt} \tag{6.11}$$

Let's consider a loudspeaker with resonance frequency $f_r = 67.50$Hz, playing an exponential sine sweep of length $T = 60$ seconds, with instantaneous vibration frequency $f_v \in [20, 5000]$Hz. In the hypotheses of initial displacement $\tilde{\eta}_c(t = 0) =$ 2.26cm, related to initial vibration frequency $f_v(t = 0) = 20$Hz, the theoretical micro Doppler frequency is computed by equation (6.11) and shown in Figure 6.9. Unlike the constant displacement scenario, the micro Doppler frequency achieves its maximum value $f_{mD} = 887$Hz at the time instant $t_{max} = 13.2305$s, namely the instant which the vibration frequency f_v matches the resonance frequency f_r of the speaker itself. As expected, this suggests that the highest micro Doppler shift is achieved at the highest velocity of the speaker, namely at the resonance frequency. Notice that at high vibration frequency, the micro Doppler tends

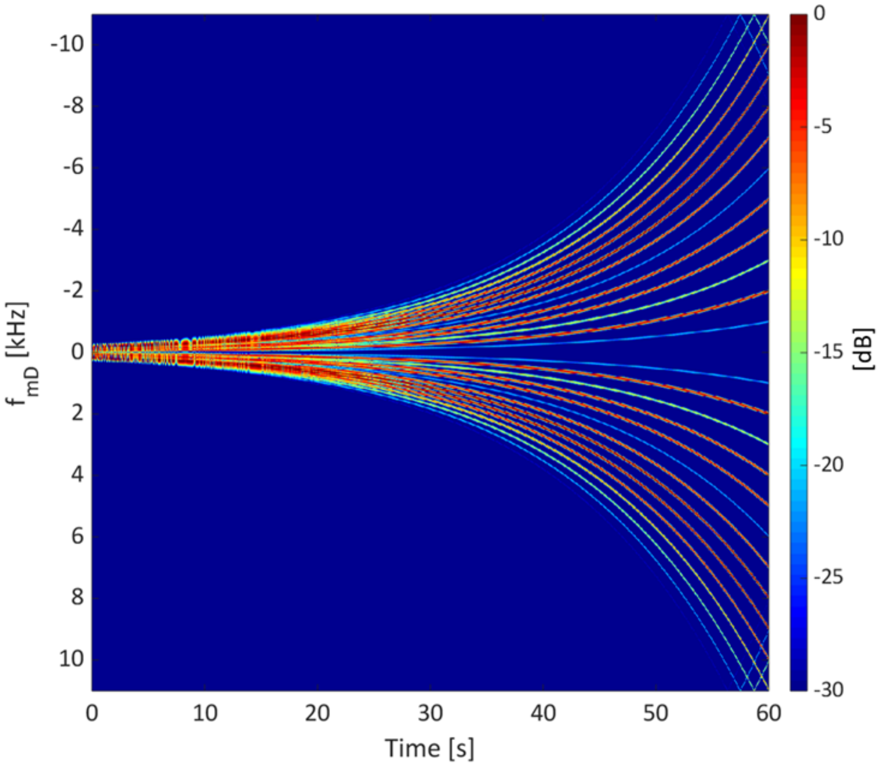


Fig. 6.8 Spectrogram of the simulated received radar signal from a speaker playing a $T = 60$ seconds exponential sine sweep with $f_v \in [20, 1000]$Hz, with fixed displacement equal $\tilde{\eta}_c = 10$mm.

towards zero due to the displacement function. The spectrogram of the simulated received radar signal has shown in Figure 6.10, where a Blackman-Harris window of 46.5ms is used. From Figure 6.10 the behaviour of the micro Doppler frequency is confirmed. While the sinusoidal like motion of the micro Doppler is still visible at low vibration frequency achieving the maximum value at the resonance frequency, at high vibration frequency it is clear the strong component at the zero frequency. The spectrogram of a time window of 0.05s of the simulated received radar signal around t_{max} is shown in Figure 6.11, confirming the sinusoidal like motion of the micro Doppler. Due to the fast vibration of the speaker on the Line of Sight (LOS), a phenomenon known as coupled echoes will appear [66]. The result of this effect will be the presence of “ghost returns” in the Doppler direction on both sides of the original target. So the speaker vibration can then introduce an infinite series of paired echoes m because, when considering (6.11), the received signal s_r in (6.9) may be expressed as a series of expansion of Bessel functions of the first

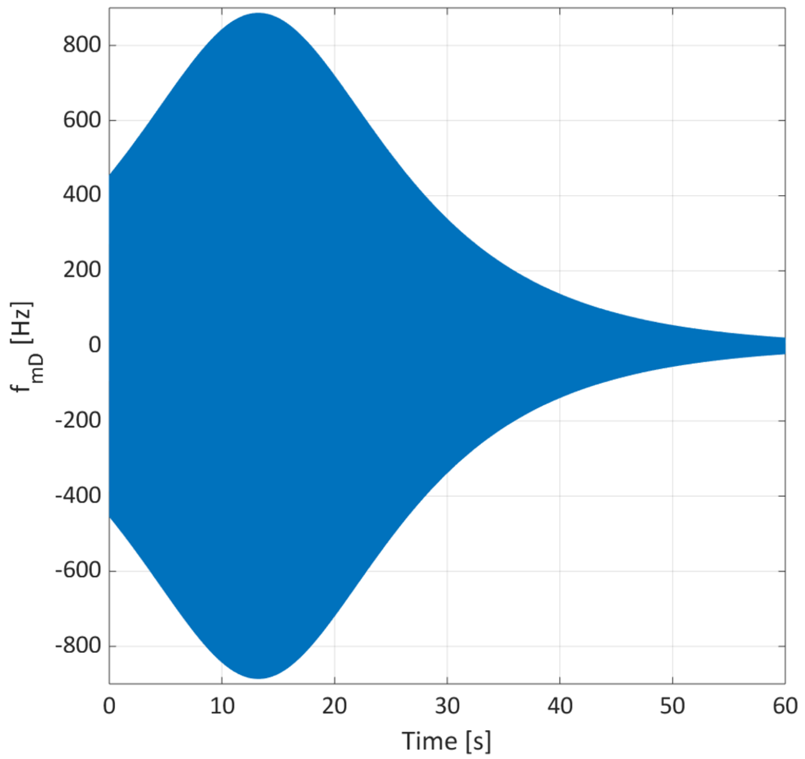


Fig. 6.9 Theoretical micro-Doppler frequency shift from of a speaker playing an exponential sine sweep of $T = 60$s, with $f_v \in [20, 5000]$Hz, and initial displacement $\tilde{\eta}_c(t = 0) = 2.26$cm and $f_v(t = 0) = 20$Hz.

kind of order m:

$$J_m\left(\frac{4\pi\tilde{\eta}_c}{\lambda}\right) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \exp\left\{j\left(\frac{4\pi\tilde{\eta}_c}{\lambda}\sin(u) - mu\right)\right\} du \tag{6.12}$$

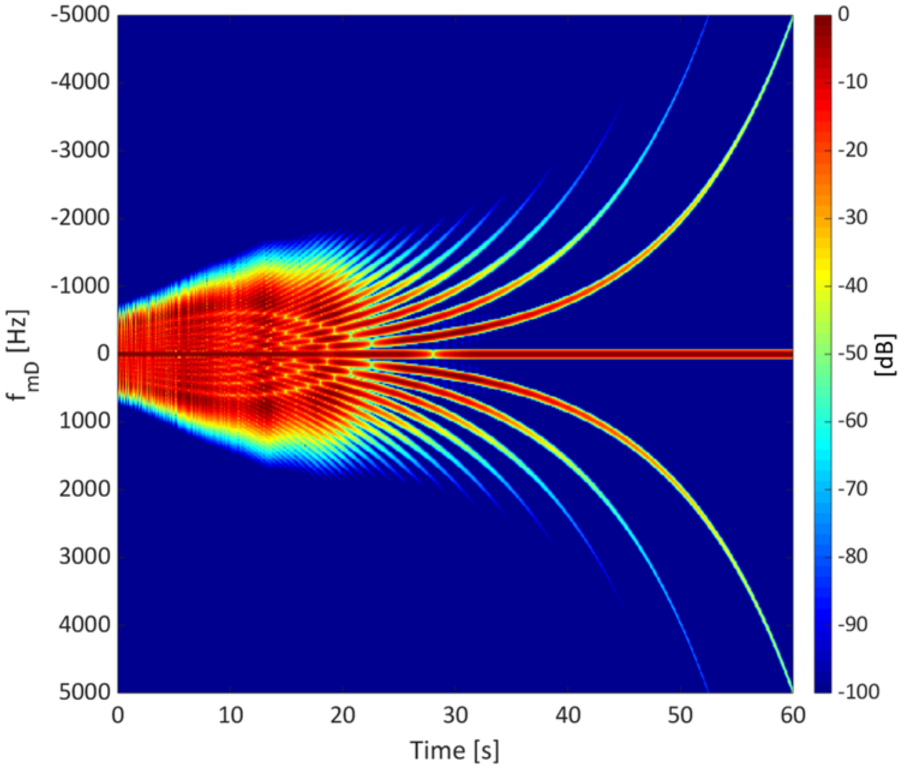


Fig. 6.10 Spectrogram of the simulated received radar signal from a speaker playing an exponential sine sweep of $T = 60\text{s}$, with $f_v \in [20, 5000]\text{Hz}$, and initial displacement $\tilde{\eta}_c(t = 0) = 2.26\text{cm}$ and $f_v(t = 0) = 20\text{Hz}$.

such that the received radar signal s_r in baseband can be expressed as:

$$\begin{aligned} s_r(t) =& \rho \exp\left\{j\frac{4\pi R_0}{\lambda}\right\} \exp\left\{j2\pi f_0 t\right\} \\ & \sum_{m=-\infty}^{+\infty} J_m\left(\frac{4\pi\tilde{\eta}_c}{\lambda}\right) \exp\left\{mj\frac{2\pi f_1\left(k^t-1\right)}{\log k}\right\} = \\ & \rho \exp\left\{j\frac{4\pi R_0}{\lambda}\right\} \exp\left\{j2\pi f_0 t\right\} \left\{J_0\left(\frac{4\pi\tilde{\eta}_c}{\lambda}\right) + \right. \\ & + J_1\left(\frac{4\pi\tilde{\eta}_c}{\lambda}\right) \exp\left\{j\frac{2\pi f_1\left(k^t-1\right)}{\log k}\right\} + \\ & - J_1\left(\frac{4\pi\tilde{\eta}_c}{\lambda}\right) \exp\left\{-j\frac{2\pi f_1\left(k^t-1\right)}{\log k}\right\} + \\ & + J_2\left(\frac{4\pi\tilde{\eta}_c}{\lambda}\right) \exp\left\{j\frac{4\pi f_1\left(k^t-1\right)}{\log k}\right\} + \\ & \left. -J_2\left(\frac{4\pi\tilde{\eta}_c}{\lambda}\right) \exp\left\{-j\frac{4\pi f_1\left(k^t-1\right)}{\log k}\right\} + \ldots \right\} \end{aligned} \tag{6.13}$$

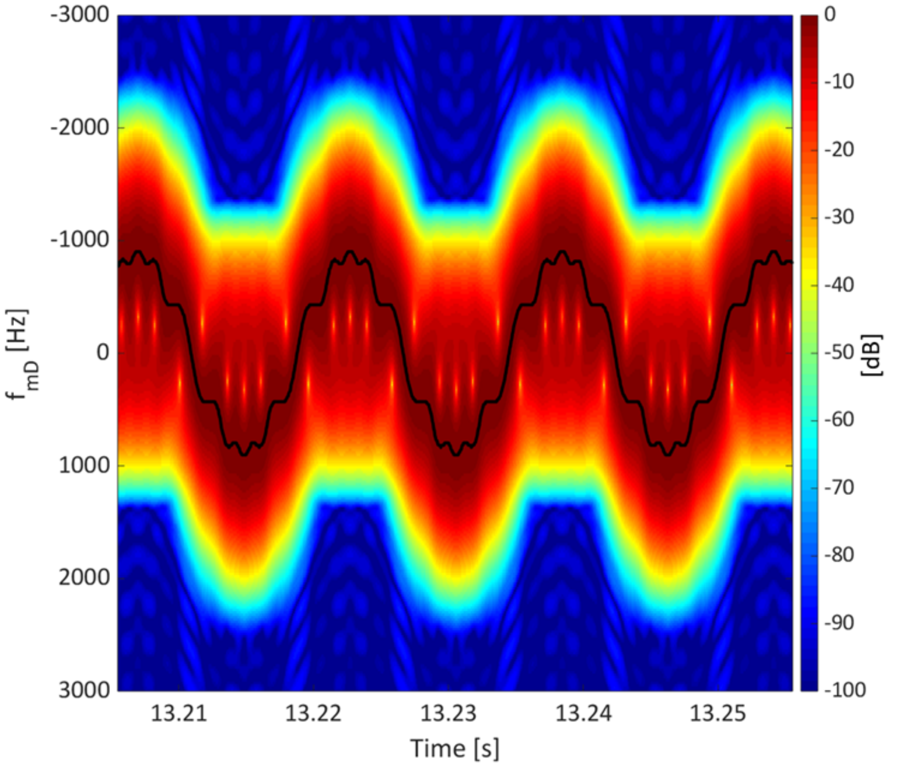


Fig. 6.11 Spectrogram of time window of 0.05s of the simulated received radar signal around the time instant t_{max}, where Doppler shift achieves its maximum, namely in proximity of the resonance frequency f_r of the speaker. The maximum Doppler shift is highlighted with a black line.

with k the exponential chirp rate. Therefore, the micro-Doppler frequency spectrum consists of pairs of spectral lines around the center frequency f_0 with spacing f_v between adjacent lines. The intensity of paired echoes visible depends on the modulation index $\Upsilon = \beta\tilde{\eta}_c$. In case of wideband modulation ($\Upsilon > 1$) more spectral lines appear. This is visible in the spectrograms in Figure 6.8 where, a fixed displacement of 10mm makes the signal wideband modulated with fixed modulation index: multiple and very densely spaced paired echoes are introduced. Due to a non constant displacement, a different behaviour is shown through the spectrogram in Figure 6.10 where the received signal is wideband modulated at low vibrational frequency and narrowband modulated at high vibrational frequency. Thus, more harmonics are detected at low vibration frequencies. This results is in agreement with loudspeaker modelling theory, where the harmonics components at low vibration frequencies are defined as regular non linear distortions components, generated by the non linear behaviour of stiffness, force factor and inductance of the driver.

6.2.3 Mechanical Characterization of a Speaker

The ability of the radar technology to detect the motion of a speaker has been described above. When a chirp is used in the simulated scenario, the voltage as the force factor is supposed to be constant with the frequency of the stimulus. In a real scenario this hypothesis is not justified due to the non linearities, introduced for example by stiffness ($K_{ms}\left(\tilde{\eta}_c\right)$), force factor ($Bl\left(\tilde{\eta}_c\right)$) and inductance ($L_e\left(\tilde{\eta}_c\right)$). To understand the influence of the non linearities on the speaker behaviour, in terms of deviation from the ideal piston mode behaviour, an alternative approach is considered. Commonly used in radar is the matched filter technique, obtained by correlating a known signal with an unknown signal to detect the presence of the template in the unknown signal. This is the radar equivalent of the acoustic measurement technique introduced in the Section 3.6, where the unknown signal is convolved with a conjugated time-reversed version of the template [90–92]. With this technique the speaker can be mechanically characterized. The matched filter is the optimal linear filter for maximizing the signal-to-noise ratio (SNR) in the presence of additive white Gaussian noise. If the model of the ideal received radar signal is found, matched filter technique could be applied to RF sensors in order to characterize the mechanical behaviour of the speaker. Using (6.9) the received radar signal in baseband of an ideal loudspeaker, behaving as piston mode in the full frequency band, is described. It can be seen as the product between the magnitude and phase components. While the phase component depends on both T&S parameters of the speaker and the stimulus waveform, the magnitude component introduces uncertainty since it is an estimation of the target reflectivity, which is usually difficult to estimate. For this reason, to reduce the amount of uncertainty, the system's impulse response can be computed by simply correlating the phase of the measured radar signal $y(t)$ with the phase of the simulated signal $s_r\left(t\right)$, such that:

$$h\left(t\right) = \angle y\left(t\right) \star \angle s_r\left(t\right). \tag{6.14}$$

where $\star$ is the correlation operator. With an exponential sine sweep of $T = 60$ seconds long as a test signal, and with the hypothesis of a linear system, the results would be a perfect peak centred in T, defined as linear impulse response, as shown in the Figure 6.12. In a real scenario instead, where the Device Under Test (DUT) is never linear, along with the linear impulse responses, non linear impulse responses are also obtained, corresponding to the various harmonics of the input signal. With the exponential sine sweep, these non linear products, do not contaminate the linear impulse response, as they are occurring at very precise

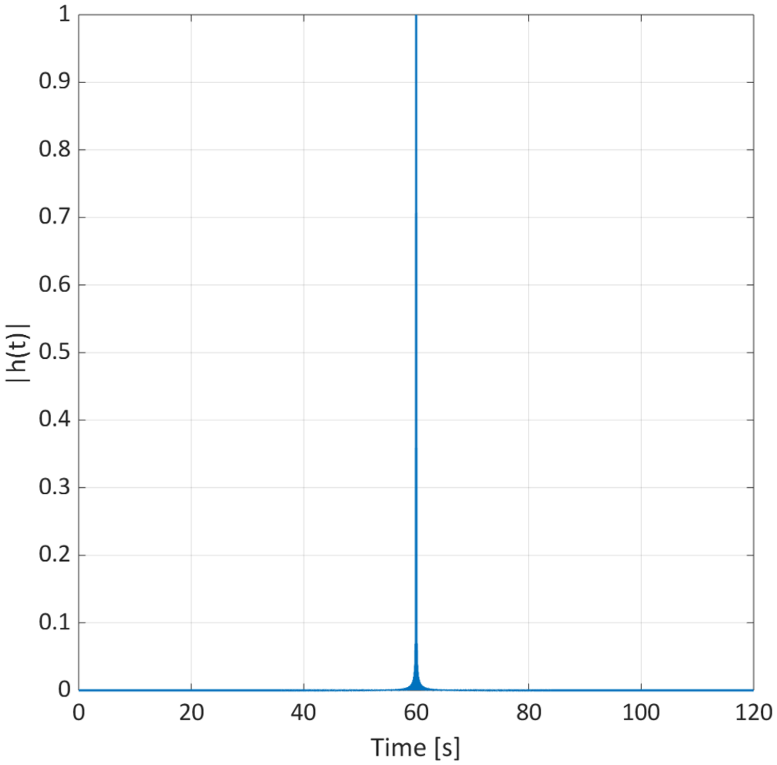


Fig. 6.12 Matched filter output of the simulated radar signal from an ideal speaker, playing an exponential sine sweep of $T = 60$s with $f_v \in [20, 5000]$Hz.

anticipatory times Δt before the linear response, namely:

$$\Delta t = T \frac{\ln (N)}{\ln \left(\frac{f_2}{f_1}\right)} \tag{6.15}$$

where N is the Nth distortion component. Thus, the signal component of the time waveform at the output of the matched filter is actually the autocorrelation function r_{s_r,s_r} of the ideal signal. The matched filter peak $h(T)$ is then $r_{s_r,s_r}(0) = E_{s_r}$, where E_{s_r} is the total energy in the signal $s_r(t)$ [130]. Applying a window around the peak $h(T)$, it is possible to compute the linear frequency response through Fourier Transform. In the event where no window is applied, the Power Spectral Density (PSD) of the signal is computed, where the harmonic components and noise are incorporated into the frequency response. In Figure 6.13 the PSD of the simulated radar signal from an ideal speaker is shown. Later in this chapter, the harmonic distortion responses will not be discarded but analysed. The system's response is affected in varying ways by different irregular defects, making the non-linear behaviour of the loudspeaker vibrations a powerful indicator of possible manufacturing problems. Thus, in the real scenario it is possible to define the

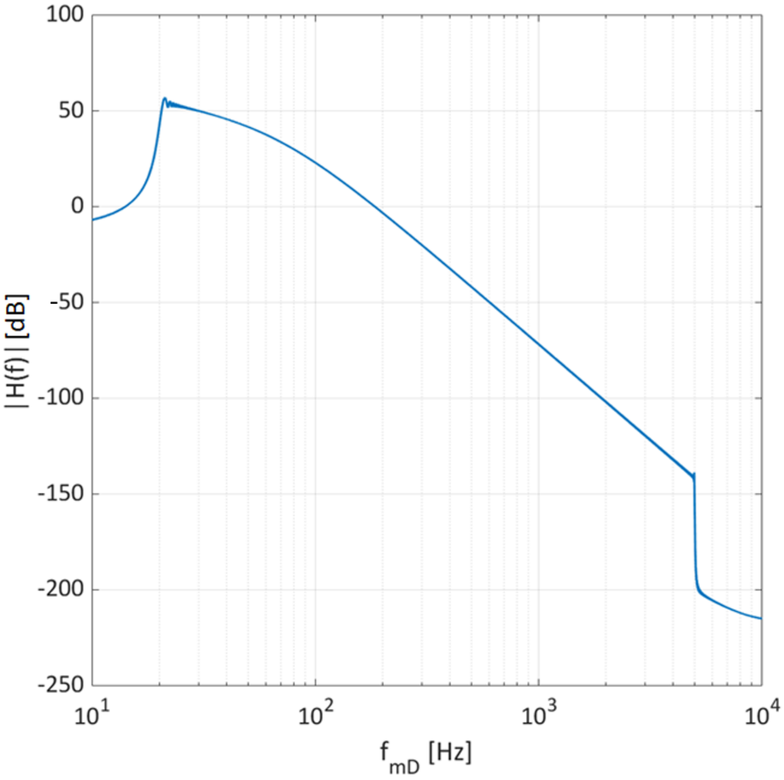


Fig. 6.13 Normalised frequency responses of an ideal speaker, playing a exponential sine sweep of $T = 60$s with $f_v \in [20, 5000]$Hz.

time waveform at the output of the matched filter as the cross correlation function r_{y,s_r} between the measured signal $y(t)$ and the ideal one $s_r(t)$, where its Fourier Transform is referred as Cross Power Spectrum Density (CPSD).

6.3 Real measurement analysis

In this section, real data acquisitions are analysed and compared with simulation results. In Section 6.3.1, the micro-Doppler signature is analysed considering a single tone acoustic signal. The micro-Doppler signature of a speaker playing a sine sweep is analysed in Section 6.3.2. Finally, in Section 6.3.3 the mechanical characterization of a real speaker is shown. For all of the analyses, the signal amplitude was set to $-6dB$ for the standard "Loudness units relative to Full Scale" (LUFS) to prevent any digital or analog clipping in the measurement chain. In order to simulate a received radar signal, a diameter of 25cm (which is a typical dimension for a loudspeaker operating in this frequency range) has been considered. The backscattering coefficient ρ [66] coming from the only vibrating metallic component, namely the voice coil, is calculated as in (6.7), with the

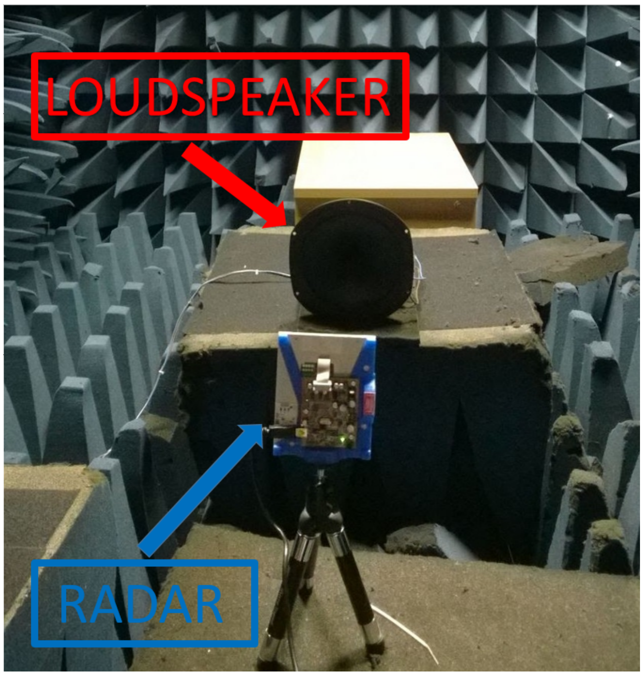


Fig. 6.14 Experiment Setup.

Table 6.1 Measured Thiele&Small Parameters of *B&C 10CL51* LF driver.

Parameter	**Value**
f_r	67.57 Hz
Bl	$9.67 N/A$
Q_{ts}	0.5425
Q_{es}	0.6075

radius of voice coil $r = 2.55$cm. The measurements acquisition was conducted through a bespoke 24GHz CW radar made by *WhiteHorse Radar LTD*. It has been used to measure the returns from a 25cm low frequency driver placed 1m away from the radar on the Line Of Sight (LOS). For both simulated and real data, a sampling frequency $f_s = 22$kHz is considered, due to the hardware limitation. Through a *Clio Pocket* board [131], the electromechanical parameters needed to feed the model of the ideal received radar signal are computed. The measured T&S parameters of *B&C 10CL51* LF driver [132] are reported in Table 6.1. In Figure 6.14, the system set up is shown. The input signal to the loudspeaker has been generated by *Adobe Audition 3.0*, while the received signal is acquired by the radar through *Matlab R2018a*, also used to process the data.

6.3.1 Micro Doppler Signature: Single tone analysis

In this section, the single tone analysis is performed, where real data are analysed and compared to simulation results. A single tone has been chosen as acoustic input to the loudspeaker with frequency $f_v = 67\text{Hz}$ to drive an ideally flat and rigid disk behaving in piston mode, at its resonance frequency. To understand the ability of the radar to detect the motion of the speaker, two different output voltage were taken into consideration. Setting the voltage at the loudspeaker terminals to be 5V, the normalized spectrum of the received signal is shown in Figure 6.15. Having the signal in baseband, in the spectrum in Figure 6.15 the positive frequency band is referred as positive direction while the negative as negative displacement. By Fourier analysis the vibration frequency of the coil is detected correctly. Despite the presence of the noise floor, the fundamental component and its harmonics are visible and in agreement with the spectrum of the simulated signal. The discrepancy between the ideal and the real spectrum is related to the non linear effect of the DUT, previously defined. Moreover, the discrepancy between the positive (blue curve) and negative direction (red curve) may be related to the effect of non linear stiffness. In Figure 6.16, a

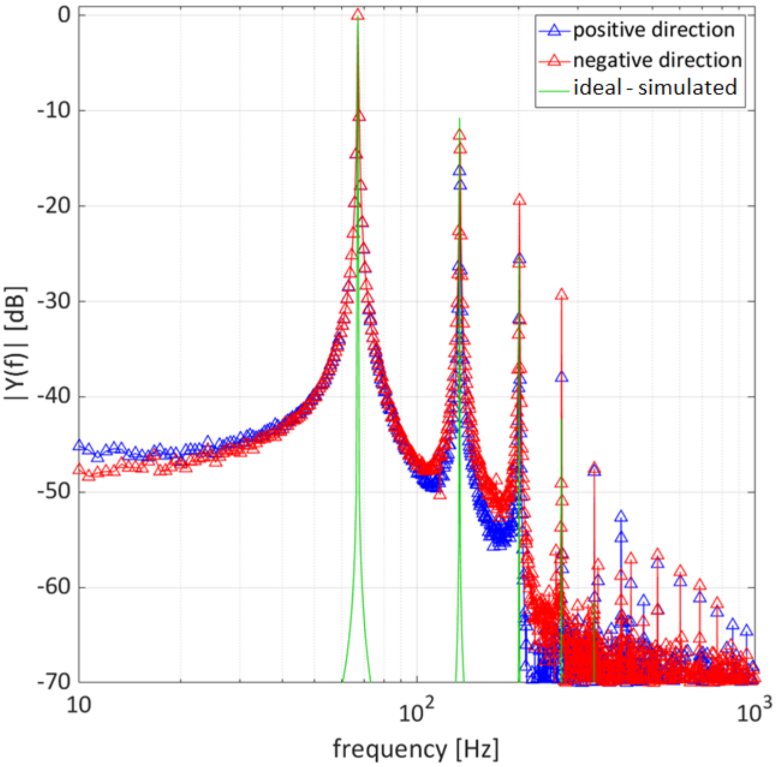


Fig. 6.15 Spectrum of real radar measurement from a 25cm loudspeaker playing a single tone $f_v = 67\text{Hz}$ at its resonance frequency at 5V output voltage.

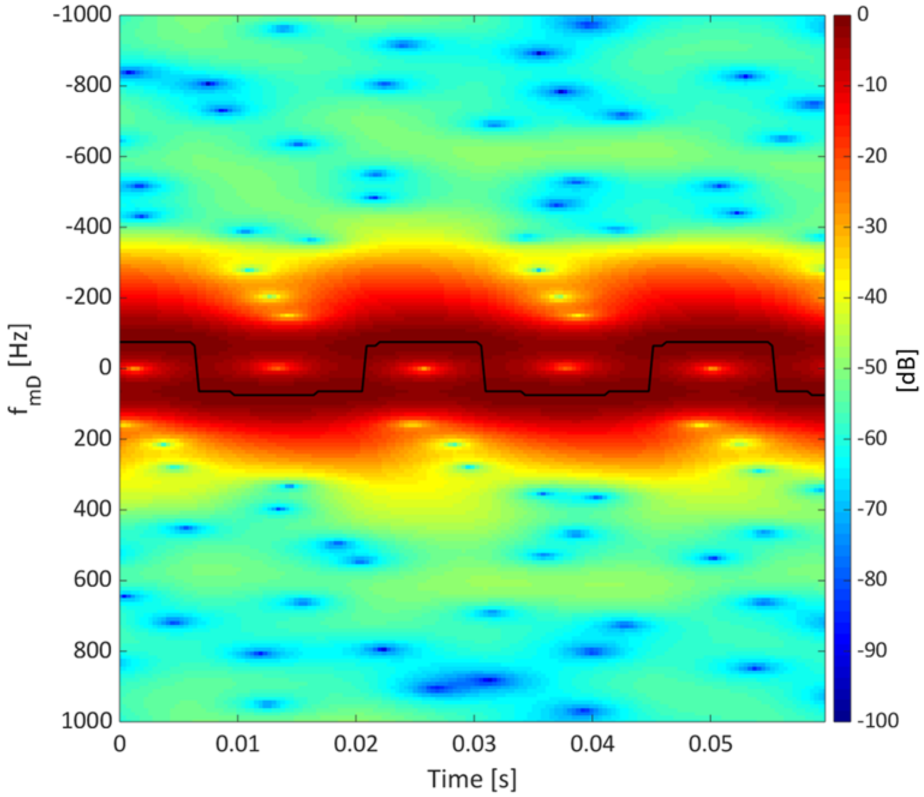


Fig. 6.16 Spectrogram of real radar measurement from a 25cm loudspeaker playing f_v = 67Hz single tone at 5V output voltage, with Blackman-Harris window of 23.2ms. The maximum Doppler shift is highlighted with a black line.

Blackman-Harris window of 23.2ms long is used to generate the spectrogram of the radar signal. From the spectrogram the maximum frequency Doppler shift can be evaluated and from this the displacement. Inverting (6.8), the maximum value of $\tilde{\eta}_{c,5V}$ can be obtained. As in the simulated scenario (in Figure 6.5), the micro Doppler has a maximum value equal to 75Hz, in both positive and negative direction, leading to the estimation of the displacement equal to 1.1mm. In the case of an output voltage of 10V, the speaker should be more prone to distortion. This is visible from both the spectrum in Figure 6.17 and the spectrogram in Figure 6.18, where harmonics with higher magnitude appear due to a larger displacement. Although the behaviour is still in agreement with the model in Figure 6.6, some discrepancies appear. The differences with the ideal micro Doppler are illustrated in Figure 6.19.

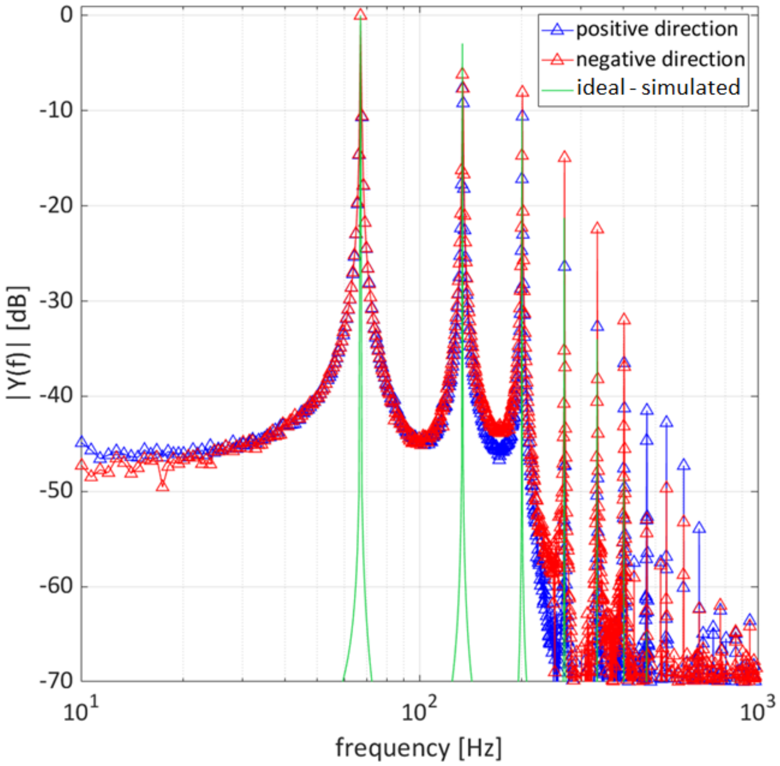


Fig. 6.17 Spectrum of real radar measurement from a 25cm loudspeaker playing a single tone $f_v = 67$Hz at its resonance frequency at 10V output voltage.

While in the simulated scenario the micro Doppler profile has a maximum and minimum value equal to 150Hz, the measured one resulted to be 150Hz in the positive direction and 172Hz in the negative direction. This suggests that, due to non linear effects, the voice coil is susceptible to acceleration in order to reach the farthest point from the radar, visible through the different rising and falling front from the simulated one. This can be confirmed by the phase of the signal.

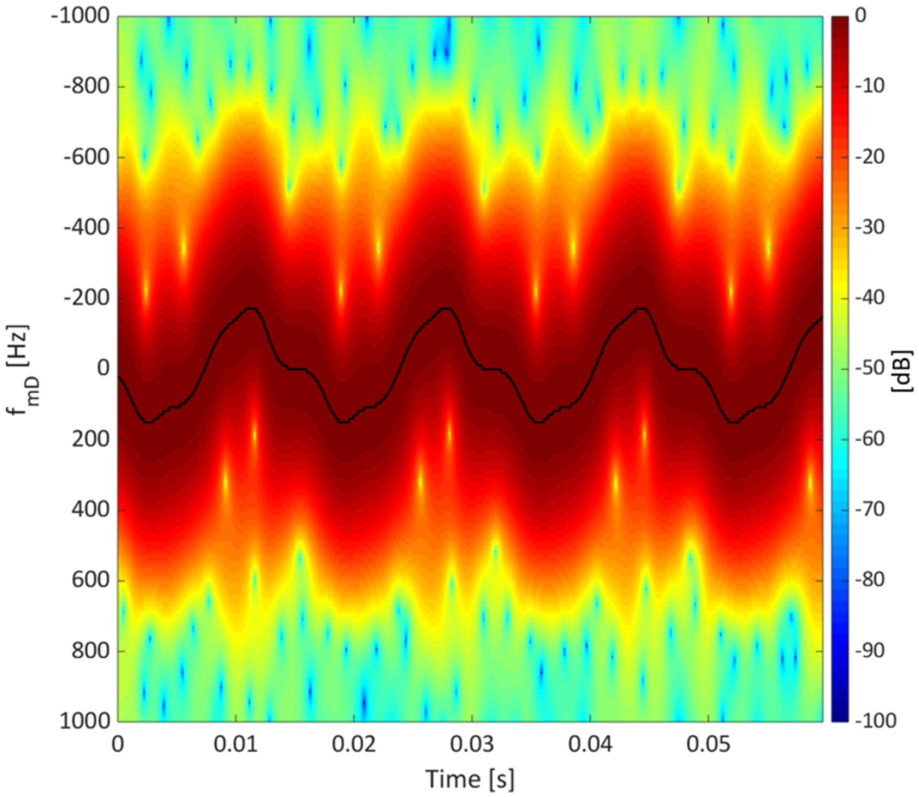


Fig. 6.18 Spectrogram of real radar measurement from a 25cm loudspeaker playing $f_v = 67$Hz single tone at 10V output voltage, with Blackman-Harris window of 5.8ms. The maximum Doppler shift is highlighted with a black line.

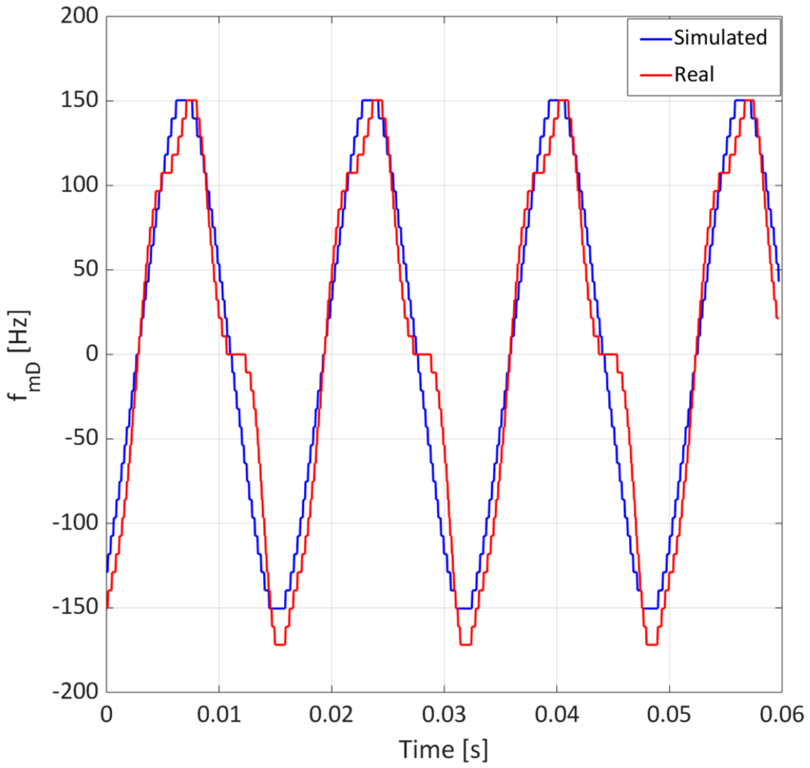


Fig. 6.19 Micro Doppler comparison between the simulated signal and the real radar measurement from a 25cm loudspeaker playing 67Hz single tone at 10V output voltage, with Blackman-Harris window of 5.8ms.

In Figure 6.20 and 6.21 the phase of the real and simulated radar signals are compared when an output voltage is set to 5V and 10V, respectively. It can be seen that the phase of the real data matches the simulated one in terms of sinusoidal-like motion, especially when the applied voltage is 5V. However discrepancies between the simulated and real data appear at 10V of applied voltage. Comparing the rising and falling front of the measured phase in Figure 6.21 with the corresponding simulated one, it is possible to observe that actual motion does not match completely the ideal one. From the plot it is possible to infer that the voice coil spends more time in the position further away from the radar than the piston model suggests. A possible explanation of this phenomenon could be the superposition of two components. The first is the non linearity introduced by the stiffness. Loudspeakers use a suspension system to center the coil in gap and to generate a restoring force which moves the coil back to the rest position. Thus, spider and surround behave like a normal spring: at low displacement there is an almost linear relationship, while at high displacement the suspension responds with more force than the predicted one. The second component could derive from the non linearity introduced by the force generated by the magnetic field times the length of the voice coil immersed in the gap: if the coil windings leave the gap, the force factor decreases [86]. Consequently, the non linearities appear and the coil is pushed back by magnetic field, from the nearest to the furthest position from the radar, earlier than its ideal one.

6.3.2 Micro Doppler Signature: Sine Sweep Analysis

In this section the micro Doppler signature of a speaker playing a sine sweep will be analysed and compared to the simulation results. The acoustic tone varies from a starting frequency $f_1 = 20$Hz at time $t = 0$, ending at the time $T = 60$s with frequency $f_2 = 5$kHz. The speaker was connected to an amplifier, whose output was set to $\tilde{e}_g = 3$V at 1kHz. With the speaker parameters in Table 6.1, the theoretical displacement is modelled through (3.86): at time $t = 0$ the maximum value of $\tilde{\eta}_c = 1.1$mm is found. The spectrograms of both simulated and real radar signal are compared and shown in Figures 6.22 and 6.23, respectively. Both the spectrograms are produced using a Blackman-Harris window of 46.5ms, with an overlap of 99%, with column based normalization. The first difference that can be immediately noted, it is the presence of the noise floor. While the simulated signal has been generated in absence of noise, the real one shows a background noise increasing with time. This result is in agreement with radar sensitivity,

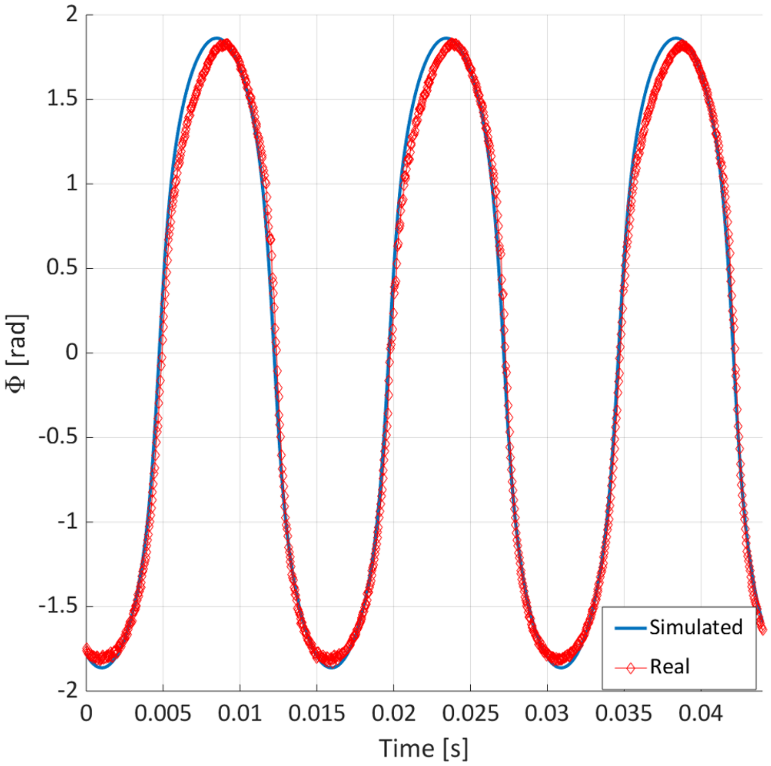


Fig. 6.20 Comparison between the phase of the simulated and real radar signal of the speaker playing tone $f_v = 67\text{Hz}$ equal at its resonance frequency f_r, with output voltage set to 5V.

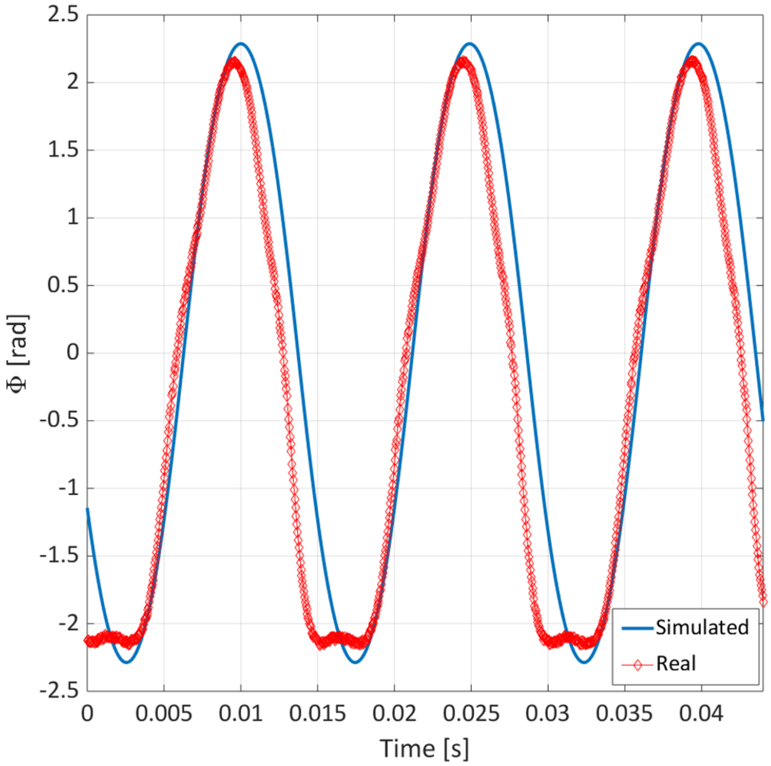


Fig. 6.21 Comparison between the phase of the simulated and real radar signal of the speaker playing tone $f_v = 67\text{Hz}$ equal at its resonance frequency f_r, with output voltage set to 10V.

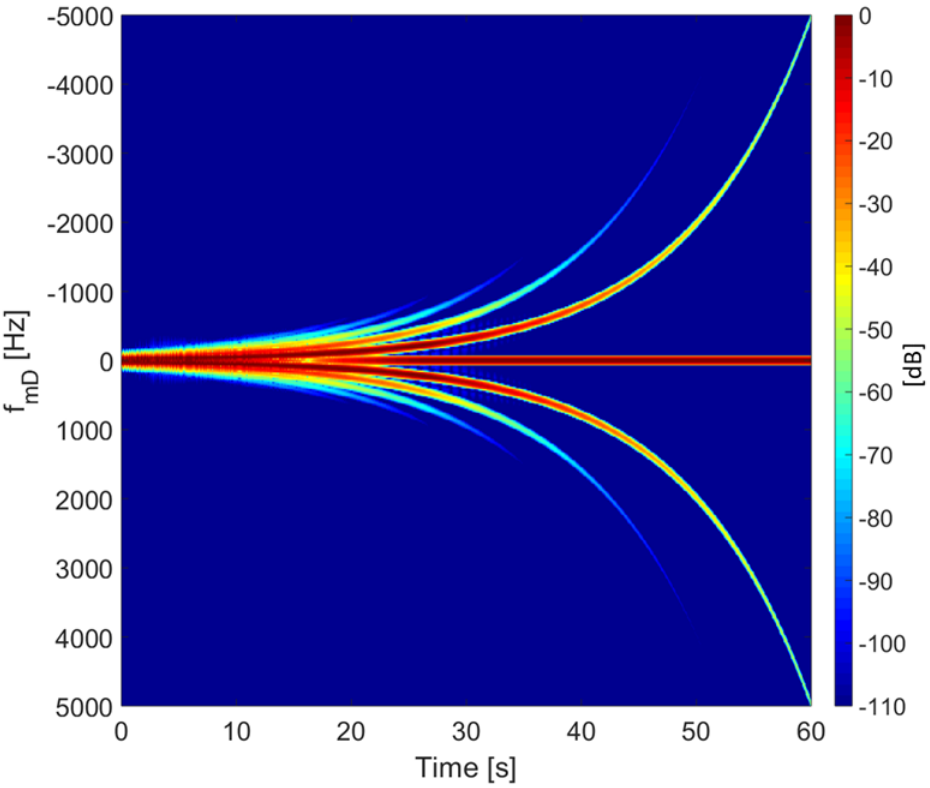


Fig. 6.22 Spectrogram of the simulated received radar signal from a speaker playing an exponential chirp, with T&S parameter of Table 6.1, with Blackman-Harris window of 46.5ms.

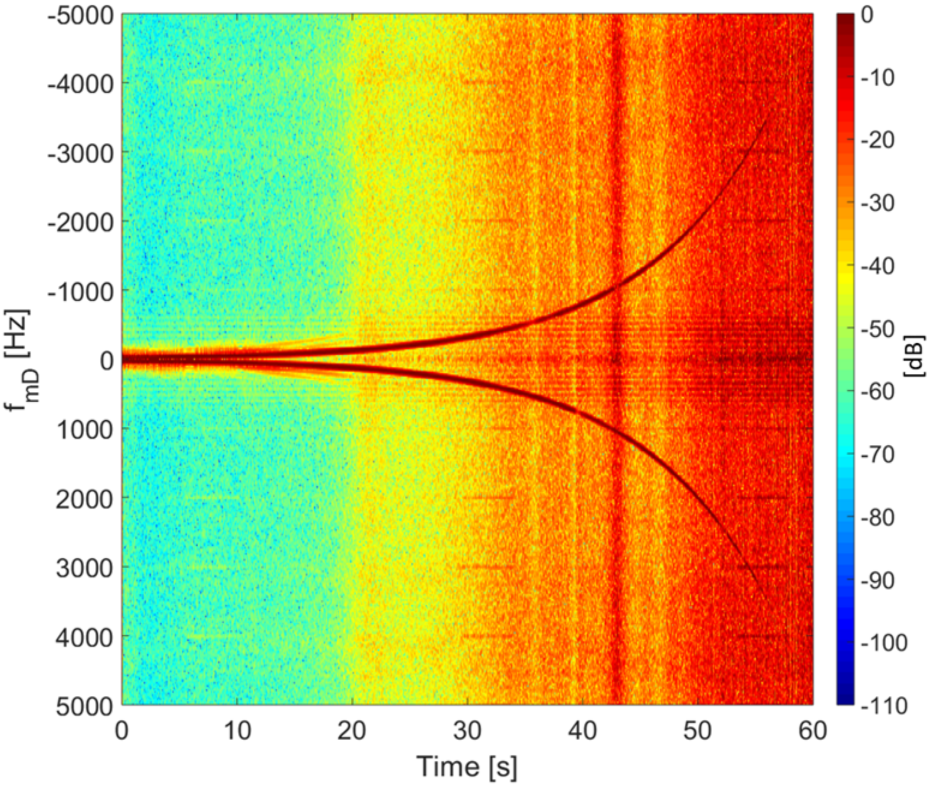


Fig. 6.23 Spectrogram of the real received radar signal from a speaker playing an exponential chirp, with T&S parameter of Table 6.1, with Blackman-Harris window of 46.5ms.

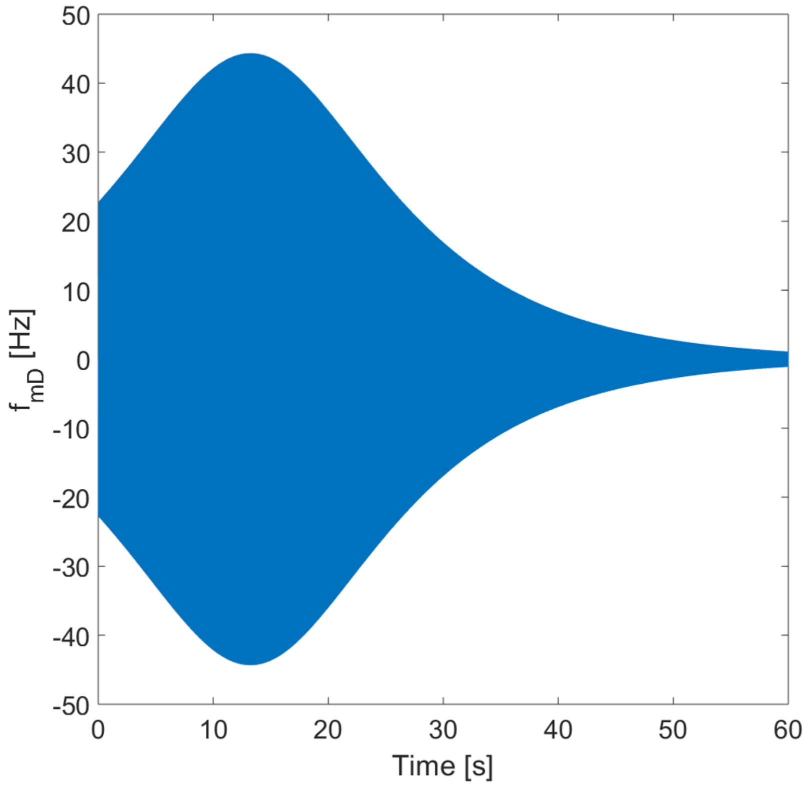


Fig. 6.24 Theoretical micro Doppler frequency shift of a speaker playing a 60 seconds exponential sine sweep with $f_v \in [20, 5000]$Hz, with T&S parameter of Table 6.1.

shown in Table 6.2: as a vibration amplitude of a micron results in a phase shift of only 0.06 deg, it is almost undetectable. It can be seen from Figure 6.23 that a high intensity distortion is visible at the time instant $t = 42$s, with vibration frequency f_v approximately 1KHz. Due to rocking modes, DC displacement and motor instability, the speaker deviates from the "piston mode", making voice coil rubbing and hard bottoming typical defects.

From the radar point of view this effect can be explained through the concept of disruptive interference. Whenever waves originating from two or more sources interact with each other, there will be phasing effects leading to an increase or

Table 6.2 Sensitivity of a 24GHz CW radar to different vibration amplitude.

Vibration Amplitude	**Maximum Phase Shift**
1cm	576 deg
1mm	58 deg
1μm	0.06 deg

decrease in wave energy at the point of combination. When elastic waves of the same frequency meet in such a way that their displacements are precisely synchronized (in phase, or 0 degree phase angle), the wave energies will add together to create a larger amplitude wave. If they meet in such a way that their displacements are exactly opposite (180 degrees out of phase), then the wave energies will cancel each other. At phase angles between 0 degrees and 180 degrees, there will be a range of intermediate stages between full addition and full cancellation. Using the mathematical formulation in (6.11), the theoretical micro Doppler related to the theoretical displacement is shown in Figure 6.24. Also in this case the maximum Doppler shift happens at the resonance frequency of the speaker, with maximum value of $f_{mD_{MAX}} = 44.3$Hz at the time $t_{max} = 13.23$s. The spectrograms of both simulated and real radar signal around t_{max} are shown in Figure 6.25 and 6.26, respectively. From the spectrograms in Figures 6.25 and 6.26 is possible to appreciate how the simulated signal matches with the real measurement. In both the spectrograms, a maximum Doppler shift of 107Hz is detected with a Blackman-Harris window of 11.6ms. Due to both the spectrogram time-frequency dilemma and coupled echoes phenomenon, the maximum Doppler shift detected differs from the theoretical one making the echoes stronger than the main component. Thus, for a correct characterization of the speaker, an alternative approach is needed and introduced in the next section.

6.3.3 Mechanical Characterization of a speaker

With the displacement model introduced in the section 3.6, the performance of the loudspeaker at low frequency can be estimated. This estimation is computed considering small input signal levels for which the mechanical behaviour of the driver is effectively linear. In order to understand the effects introduced by the non linear components, the matched filter approach is used. In case of perfect linear system, the matched filter output would consist in a perfect peak centred at the instant T equal to the length of the test signal, defined as linear impulse response. Real devices unfortunately are never linear; thus, not only a linear impulse response appears, but also non linear impulse responses are obtained, corresponding to the various harmonics of the input signal. This is visible in the Figure 6.27, where the matched filter is applied to a real measurement. In agreement with (6.15), the non linear products occur at very precise anticipatory time before the linear response, namely at $\Delta t_{2nd} = 7.50$s, $\Delta t_{3rd} = 11.93$s and $\Delta t_{4th} = 15.04$s. Applying a Fourier Transform to the matched filter output, the

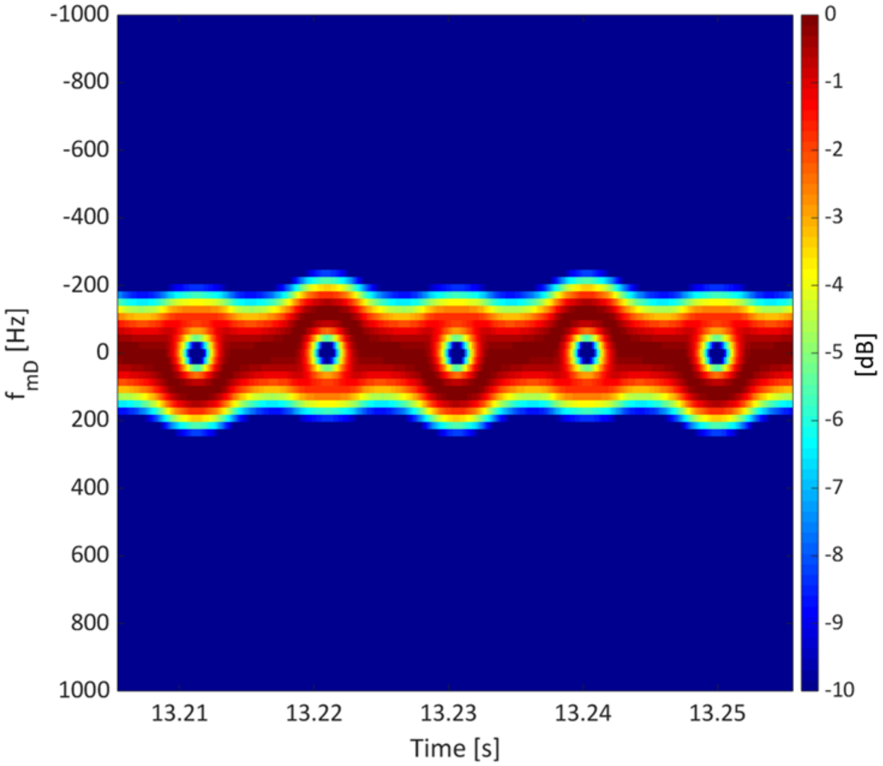


Fig. 6.25 Spectrogram of the simulated received radar signal from a speaker playing an exponential chirp, with T&S parameter of Table 6.1, at the $t = t_{max}$ and $f_v = f_r$, with Blackman-Harris window of 11.6ms.

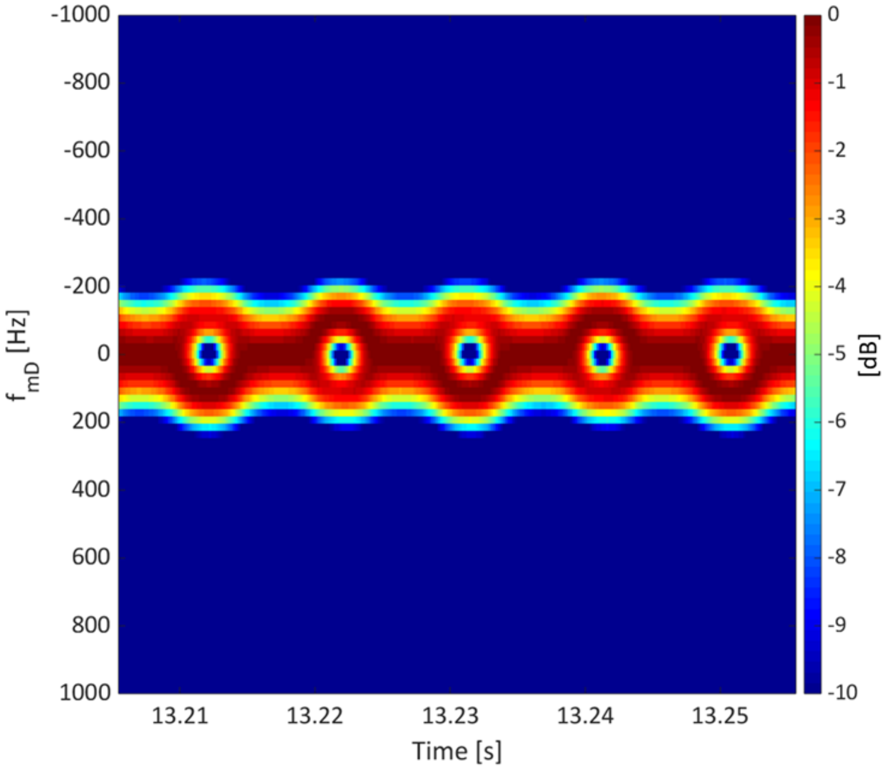


Fig. 6.26 Spectrogram of the real received radar signal from a speaker playing an exponential chirp, with T&S parameter of Table 6.1, at the $t = t_{max}$ and $f_v = f_r$, with Blackman-Harris window of 11.6ms.

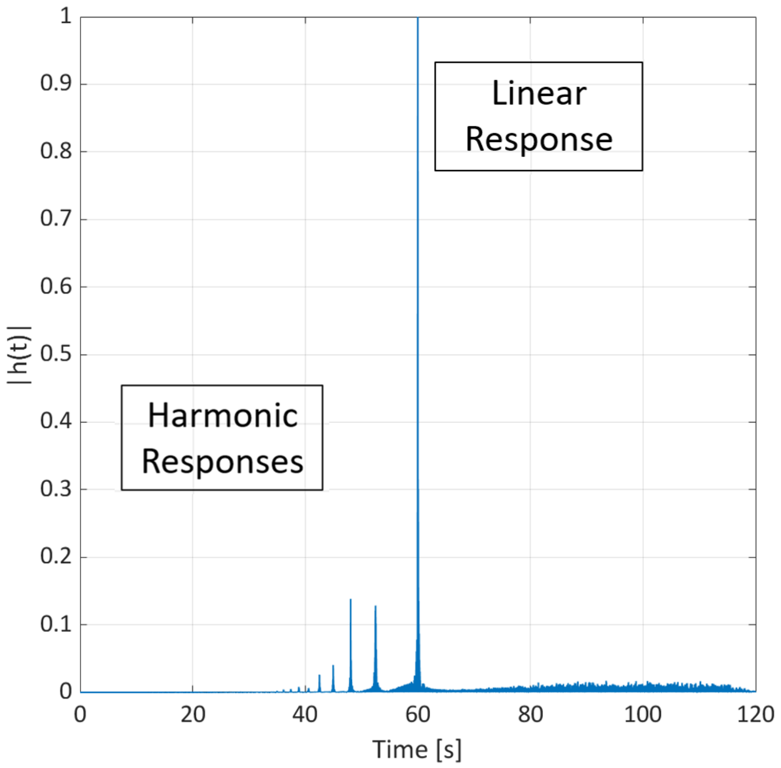


Fig. 6.27 Matched filter output: linear and non linear impulse responses of the DUT, with measured T&S parameters of table 6.1.

frequency response of the DUT can be evaluated. In Figure 6.28, CPSD, linear and harmonic frequency responses are shown, considering 2048 frequency bins. In case of no windowing, the CPSD of the signal is computed in Figure 6.28 (blue line) where harmonics products and noise are incorporated into the frequency response. Applying a window of 4096 samples around the peak in $h(T)$, the linear frequency response can be assessed. Since the non linear products are a powerful indicator of possible manufacturing problems, they are analysed too. For this reason, windows are applied to harmonic responses as well. As it can be noted from the Figure 6.28, the harmonic products affect the behaviour of the speaker mainly at low frequency, where the device is more susceptible to the non linear effects, in agreement with loudspeaker model theory.

6.4 Summary

In this chapter a novel approach for condition monitoring of loudspeakers EOL test based on radar micro Doppler was proposed. With the assumption of rigid body motion at low frequency, the displacement of a loudspeaker has been modelled as

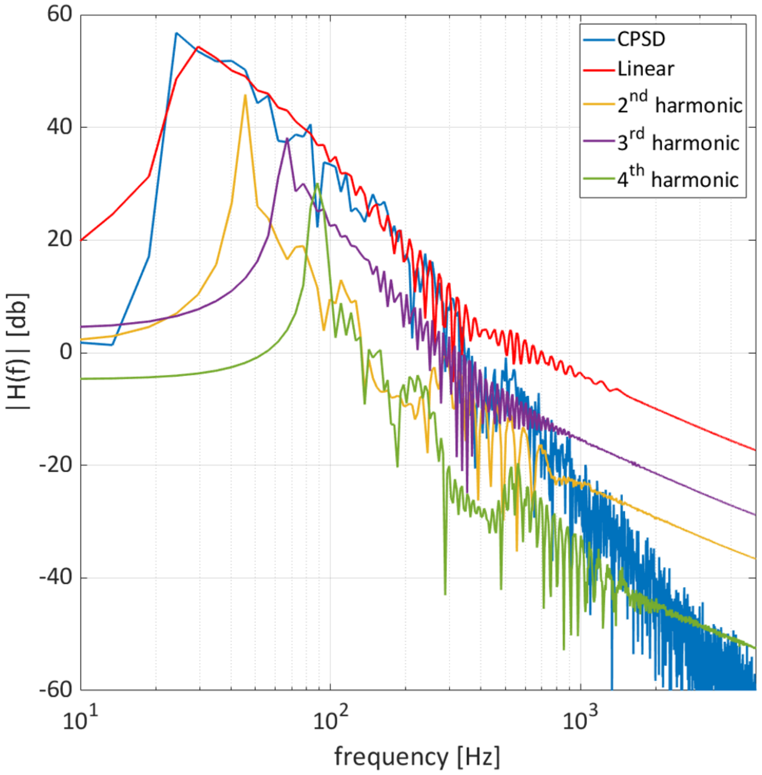


Fig. 6.28 CPSD, linear frequency response and harmonic frequency responses of the DUT, with measured T&S parameters of table 6.1.

function of the frequency of the stimulus by considering the electro-mechanical components responsible of the dynamic response of the transducer. Modelling the displacement with T&S parameters, the micro-Doppler analysis of loudspeaker could be generalized for any frequency of the stimulus.

In this work, both single tone and sine sweep analysis were performed. In the first case, the phase, the spectrum and the spectrogram of the received radar signal were compared to those from the model confirming that loudspeaker behaviour can be detected from radar. In particular, taking in account both micro-Doppler shift and the phase component of the received signal, the information of the displacement motion can be extracted. By increasing the voltage applied to the terminals of the driver, a resulting discrepancy between real and simulated signal appeared due to the non linear effects of the speaker. When the sine sweep test signal was used, some discrepancy between real and simulated signal appeared, as the rocking modes effect have not been taken in consideration in the displacement model. Nevertheless, the spectral analysis results demonstrate good ability in detecting irregular defects affecting the motion of the voice coil. Finally, a matched filter based approach was proposed to mechanically characterise the speaker. Cross power spectrum density, linear frequency response and harmonic frequency responses were analysed. These powerful indicator of possible manufacturing problems can be used as features for an automatic anomalies detection of loudspeaker defects, as it will be shown in Chapter 7.

Finally, the micro-Doppler analysis of loudspeaker presented in this chapter was restricted only to woofer-type of transducers. Due to the limited sampling frequency $f_s = 22kHz$ of the 24GHz CW radar, the maximum vibration frequency detectable would be 11kHz. This was sufficient to analyse correctly woofer, since it is designed to reproduce low frequency sounds (up to 5kHz usually). Contrary, it would fail in case of tweeter, since it is used to reproduce high frequency sounds (up to 20kHz). In this case, if the vibration amplitude would big enough to generate a meaningful phase shift, aliasing phenomenon appears. Thus, to extend this approach to any kind of speakers, a reasonable solution would be the use of mm-wave radar. Characterized by higher carrier frequency (77-80GHz), mm-wave radars would be more sensitive to smaller displacement leading to higher Doppler shifts and fine range resolution, if used in FMCW mode. These would allow to perform different kind of signal processing, such as the range-Doppler map: it could be possible to slice the radar return in range, in order to get localized responses of the driver. Thus, tweeters and micro speakers could be also analysed, providing surely benefits to loudspeakers manufacturers with limited cost.

Chapter 7

Loudspeaker faults detection and classification

7.1 Introduction

In order to show the ability of the radar technology to automatically classify faulty speakers, a framework based on the mechanical impulse response computation is proposed in this chapter. Following the recent trends of the manufacturing industry, a deep learning architecture will be introduced as classifier. Although Convolutional Neural Networks (CNN) are mostly preferred in radar domain, here Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) is investigated. The reason behind the choice of such architecture lies on the nature of the data. Handling the mechanical frequency response of the Device Under Test (DUT) as time series or sequence data, it enables the use of LSTM-RNN architecture introduced in the chapter 4 for classification purpose. The proposed classification framework is presented in Section 7.2, with all the information needed to generate the training, validation and test dataset. In Section 7.3, the architecture of the proposed deep learning based classifier is introduced. In order to avoid the traditional problem in deep learning, namely overfitting, vanishing and exploding gradient problems, some solutions are embedded in the proposed network. Finally, in order to show the accuracy of the classification, in Section 7.4 the performance of the deep learning based classifier are shown and compared with the traditional *k*-NN.

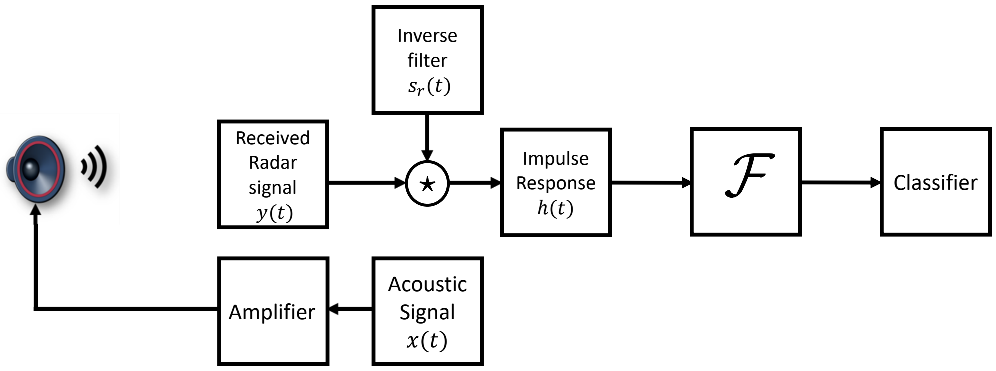


Fig. 7.1 Block diagram of the proposed framework.

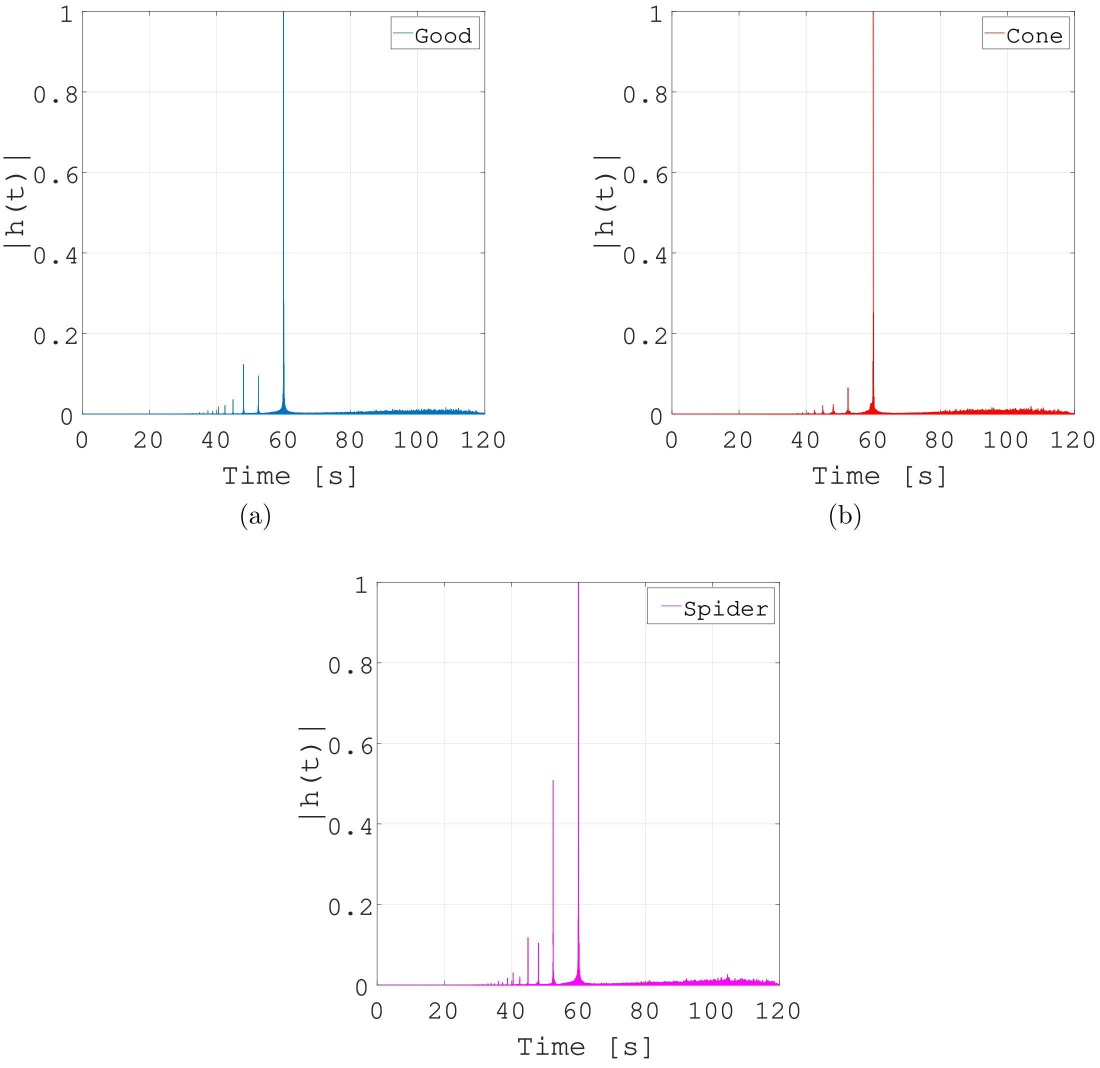
7.2 Proposed classification framework and dataset generation

To assess the ability of the radar to automatically detect and classify loudspeaker defects, match filtering approach is used. The block diagram of the proposed classification framework is shown in Figure 7.1. An acoustic sine sweep of 60 seconds long with amplitude of $-6d$B for the standard "Loudness units relative to Full Scale" (LUFS) is sent to an amplifier, connected to the loudspeaker. A 24GHz Continuous Wave (CW) radar is placed 1m away, in line of sight with the driver. Thus, the received radar signal $y\left(t\right)$ is acquired with sampling frequency $f_s = 22$kHz. The collected signal is then correlated with the inverse filter $s_r\left(t\right)$. The matched filter output corresponds to the mechanical impulse response of the DUT: applying the FFT, the mechanical frequency responses are obtained and sent as input to the LSTM classifier introduced in chapter 4.
Due to the difficulty of acquiring samples with labelled defects, two different manipulations have been artificially applied to the speaker, namely:

- cone defect manipulation: in order to simulate a scenario which the mass of the cone deviates from the designed one, an extra mass of $10gr$ is attached on the diaphragm;
- spider defect: in order to simulate a defect during the loudspeaker bonding process, vinyl glue is applied on the ring of the spider.

Depending on the status of the speaker, the matched filter output will show a different mechanical impulse response. An example of the matched filter output of the same speaker, before and after the manipulations, are shown in the Figures

7.2a, 7.2b and 7.2c. As expected, different defects affect the impulse response of



(a)

(b)

(c)

Fig. 7.2 Linear and non linear responses of the DUT, with measured T& S parameters of table 6.1. (a) Good speaker. (b) Cone manipulation. (c) Spider manipulation.

the system in different ways. Using FFT, the CPSD, linear frequency response and harmonic frequency responses can be computed. Since all of them are considered as powerful indicator of possible manufacturing problems, they can be used as input sequence to the network. Increasing the number of channel of the input sequence can help to achieve a better classification accuracy. In this case, the highest accuracy is achieved by considering an input sequence with 5 channel: the CPSD, linear frequency response and the first three harmonic frequency responses of the DUT are assigned to each channel. Thus, with the number of fft bins set equal to $S = 2048$, the dimension of the single sequence data is found to be equal to 5×2048 and used as input to the network.

In order to have better classification performance, large amounts of balanced and diversified data are needed. For this purpose, a total amount of 620 measurements have been obtained from 24GHz CW radar pointing at four different speakers of the same brand (B&C 10CL51 LF driver of $10in$). Three of these speaker have been considered as golden unit, and the related measurement labelled as good. Subsequently, a cone manipulation is applied on each of them and the same amounts of measurements have been obtained. Finally, due to the irreversible damage caused by the application of the glue on the spider, the fourth speaker has been used exclusively to collect signals of speaker affected by spider defect. Thus, after selecting a balanced number of measurements for each class and speaker, approximately 70% of the total measurements are used to populate the training dataset, and 15% for both the validation and test datasets, with overall dimensions of $5 \times 2048 \times 450$, $5 \times 2048 \times 85$ and $5 \times 2048 \times 85$ respectively.

In a real scenario, two practical problems need to be also tackled, namely:

- the heating effect: depending on the usage of the driver, the magnitude of the frequency responses will be affected;
- features with different ranges: the input sequence, composed by five channel, will contains the CPSD, linear frequency response and the harmonic frequency responses of the DUT, all with different magnitudes.

To solve both the problems, pre-processing of the data is necessary. The average and the standard deviation of each channel are first calculated from the training dataset, then used to normalize each channel of training, validation and test dataset.

7.3 Proposed Deep Learning Architecture

After normalization procedure, the input sequence are ready to be sent to the network, which architecture is shown in Figure 7.3. The choice of this architecture resides in the nature of the input sequence. Handling the mechanical frequency responses as time series, enables us to use RNN network for classification purposes. In particular, combining together the BRNN and LSTM architecture, respectively introduced in Section 4.3 and 4.4, bidirectional LSTM (BiLSTM) architecture can be used, since the full sequence is accessible at prediction time. Namely, a

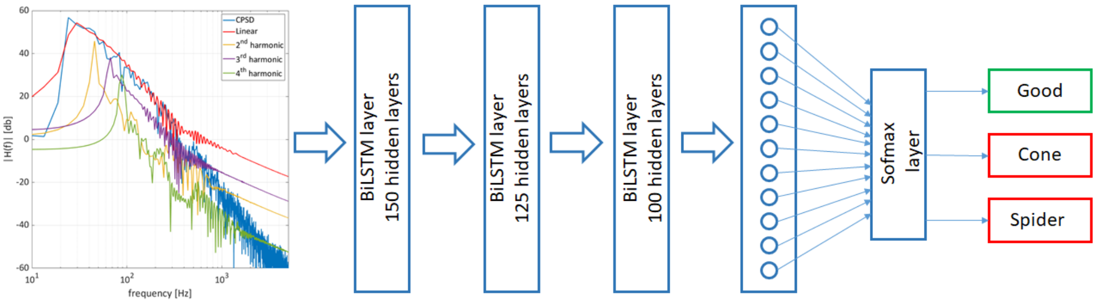


Fig. 7.3 Architecture of the implemented BiLSTM network: input layer, three BiLSTM layers, full connected layer, sofmax layer and classification layer.

BiLSTM layer learns bidirectional long-term dependencies between time steps of time series or sequence data. Since all the samples of the input time series are available from the beginning, the Bidirectional LSTM layers can be used in order to get information from past (backwards) and future (forward) states, simultaneously. Here, three BiLSTM layers are used, with decreasing number of hidden units: 150, 125 and 100 for the first, second and third layer respectively. In particular, the selection of the number of layers as well as the size of the hidden units is based on empirical analysis. These parameters might depend on several aspects of the problem, such as the complexity of the dataset, the number of features and the number of data points. For instance, in case only CPSDs and linear frequency responses of the DUTs were considered, the input sequence of the classifier would have only two channels. In this case, the highest probability of correct classification would be achieved with only two BiLSTM layer, with an accuracy lower than 90%. Higher accuracy could be reached by adding another layer. In this case, due to the low number of training samples, adding a third layer would lead to over-fitting problem. To enable the use of three BiLSTM layers without the over-fitting problem, five channels of the input sequence are considered, taking into account the CPSD, linear frequency response and the first three harmonic frequency responses.

To classify the class labels, the network ends with a fully connected layer, a softmax layer, and a classification output layer. All weights are initialized from a Gaussian distribution with zero mean and standard deviation of 0.01, while for the initial bias the values are set to be zero.

Unlike the traditional Stochastic Gradient Descent algorithm with Momentum introduced in Section 4.3.3, where only single learning rate was used to update all the network parameters, an alternative optimization algorithms that is capable of improving network training by using learning rates that differ by parameter and can automatically adapt to the loss function being optimized is preferred here. ADAptive Moment(Adam) estimation is one such algorithm [94]: it keeps an element-wise moving average of both the parameter gradients and their squared values, such that [133]:

$$\begin{cases} \mathbf{m}_l = \beta_1 \mathbf{m}_{l-1} + (1 - \beta_1) \nabla L(\boldsymbol{\theta}_l) \\ \mathbf{v}_l = \beta_2 \mathbf{v}_{l-1} + (1 - \beta_2) \left[\nabla L(\boldsymbol{\theta}_l)\right]^2 \end{cases} \tag{7.1}$$

where β_1 and β_2 are the gradient decay factor and the squared gradient decay factor of the moving averages $\mathbf{m}_l$ and $\mathbf{v}_l$, respectively. Thus, Adam algorithm is using the moving averages to update the network parameters as:

$$\boldsymbol{\theta}_l = \boldsymbol{\theta}_{l-1} - \frac{\alpha \mathbf{m}_l}{\sqrt{\mathbf{v}_l} + \epsilon} \tag{7.2}$$

If gradients over many iterations are similar, then using a moving average of the gradient enables the parameter updates to pick up momentum in a certain direction. If the gradients contain mostly noise, then the moving average of the gradient becomes smaller, and so the parameter updates become smaller too [133]. In addition, the learning rate α in (7.2) is not a fixed learning rate. An alternative to using a fixed learning rate is to indeed vary the learning rate over the training process. The way in which the learning rate changes over time (training epochs) is referred to as the learning rate schedule or learning rate decay. The simplest learning rate schedule is to decrease the learning rate linearly from a large initial value to a small value. This allows large weight changes in the beginning of the learning process and small changes or fine-tuning towards the end of the learning process [93]. In fact, using a learning rate schedule may be a best practice when training neural networks. Instead of choosing a fixed learning rate hyperparameter, the configuration challenge involves choosing the initial learning rate and a learning rate schedule. The learning rate can be decayed to a small value close to zero. Alternatively, the learning rate can be decayed over a fixed number of training epochs, then kept constant at a small value for the remaining training epochs to facilitate more time fine-tuning. In this case, an initial learning rate $\alpha = 0.01$ is choose, and decreased of a factor 10^-1 every 100 epochs. In order to avoid exploding gradient problem introduced in Section 4.3.4, l2norm-based gradient clipping has been used in this work. Finally, the problem of overfitting is also tackled by adding a regularization term to the loss function. The training

Table 7.1 Training option and parameters of the proposed BiLSTM network.

Parameter	**Value**
Epoch	500
Mini Batch	30
Solver	Adam
Initial Leaning Rate	$\alpha = 0.01$
Learning rate drop period	100 Epochs
Learning rate drop factor	0.1
Gradient Decay Factor	$\beta_1 = 0.9$
Squared Gradient Decay Factor	$\beta_2 = 0.99$
Gradient Threshold Method	l2norm
Gradient Threshold	5
Regularization coefficient	$\lambda = 0.0001$

option and parameters of the proposed BiLSTM network are summarised in Table 7.1.

7.4 Performance Analysis

In order to assess the performance of the proposed deep learning based classifier, the loss function and accuracy have been considered. In Figure 7.4, the loss function of the training dataset is shown, over 500 epochs. In Figure 7.5, the accuracy of the training dataset is compared with the accuracy of the validation dataset. After 500 epoch, a training dataset has been classified correctly with 100% of accuracy, compared to the 98.82% on the validation dataset. As observed from Figures 7.4 and 7.5, the loss decreases rapidly in the first 100 epochs, with learning rate $\alpha = 0.01$ and an accuracy already above the 90%. During this period, due to the fluctuation of the loss, the accuracy drops down in correspondence of the loss peaks. This is the effect mainly of a large learning rate: it speeds up the learning process but without finding a converging solution. By decreasing the learning rate this effect can be significantly reduced, until the loss reaches approximately zero: with a learning rate of $\alpha = 10^{-4}$, a better tuning of the network parameters is found, leading to an accuracy of 100%. The comparison with a validation dataset is also a valid proof to show that the proposed network is not prone to overfitting problem. To further validate the performance of the BiLSTM network, the confusion matrices of both validation dataset and test dataset are shown in the Tables 7.2 and 7.3, respectively. Although the accuracy of the test dataset achieves the maximum percentage, an accuracy mismatch between training and

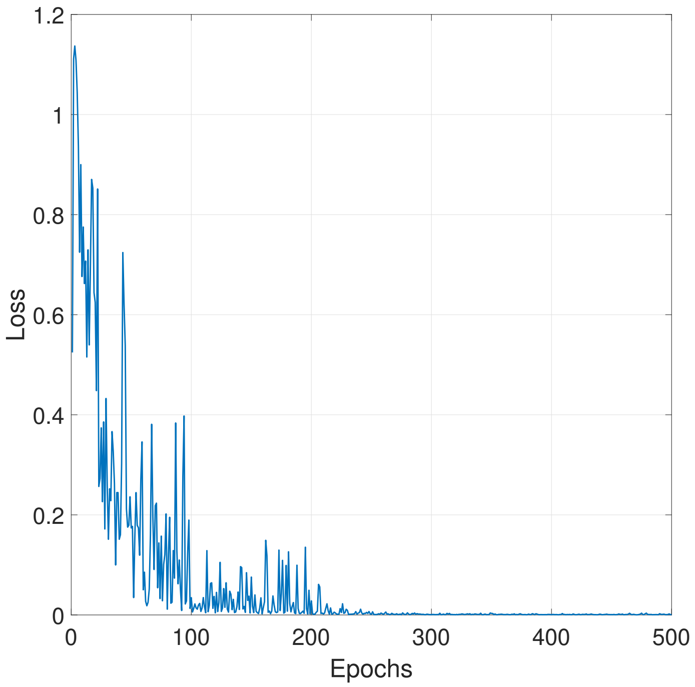


Fig. 7.4 Loss function of the proposed BiLSTM network, over 500 epochs on the training dataset.

		Target Class		
		Good	Cone	Spider
Output Class	Good	100%	0	0
	Cone	0	100%	0
	Spider	0	4%	96%

Table 7.2 Confusion matrix for the validation dataset: results of the classification for each class.

validation dataset is registered. It comes from a single wrong fault classification, namely a spider defect that it has been classified as cone defect. This accuracy mismatch may arise from a non perfect measurement acquisition: due to heating and overload usage, the speaker deviates from its behaviour, creating an higher ambiguity with the cone class.

As benchmark, the proposed BiLSTM based classifier has also been compared with the standard k-Nearest Neighbour classifier (k-NN). To empathise the ability

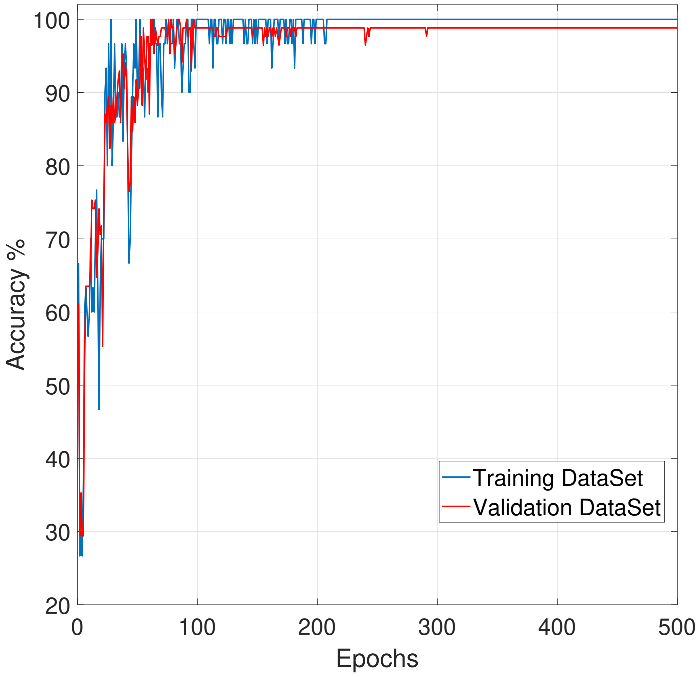


Fig. 7.5 Training and validation accuracy over 500 epochs proposed BiLSTM network.

		Target Class		
		Good	Cone	Spider
Output Class	Good	100%	0	0
	Cone	0	100%	0
	Spider	0	0	100%

Table 7.3 Confusion matrix for the test dataset: results of the classification for each class.

of deep learning models to learn features directly from the data without the need for manual feature extraction, all the 5×2048 input data are used for the k-NN. By setting the Euclidean distance as distance metric, and the number of nearest neighbours equal to the closest odd integer to the square root of the training dataset samples, namely $k = 21$, the overall accuracy is $91,76\%$ on the cross-validation dataset. In Table 7.4, the confusion matrix of the cross-validation dataset using the k-NN classifier is shown. Due to noise, heating effect and

		Target Class		
		Good	Cone	Spider
Output Class	Good	88.33%	11.67%	0
	Cone	10%	90%	0
	Spider	2%	0	98%

Table 7.4 Classification performance of the k-NN classifier: confusion matrix of the cross-validation dataset, with $k = 21$.

overload usage of the driver, the k-NN classifier seems to be less robust of the proposed BiLSTM classifier, since multidimensional spaces sufficiently separated can not be found.

7.5 Summary

In this Chapter, the capability of the joint use of radar and deep learning technology to automatically detect and classify faulty speakers has been evaluated. The classification framework in 7.2 has been used to classify real received radar signal from speakers, affected by different damages.

By using a 24GHz CW radar, a real dataset has been obtained by acquiring in laboratory the signals scattered by the DUTs: due to the difficulty of acquiring samples with labelled defects, two different manipulations have been artificially applied to the speaker, namely on cone and spider. The output of the matched filter shows discrepancies between mechanical impulse response of the speaker without defects and the ones with defects. To restrict the number of samples, Fourier Transform is applied on both linear and harmonic impulse responses. In this way, mechanical frequency responses are obtained and used as input vector of the network.

Handling the mechanical frequency responses as time series, allow the use of RNNs network for classification purposes. Since the full sequence, representing the mechanical frequency response, is accessible at prediction time, a combination of BRNN and LSTM architecture is preferred, leading to the BiLSTM architecture introduced in 7.3. The reliability of deep learning based classifier has been demonstrated by testing the network on training, validation and test dataset. The results have shown that, for all the real data, the proposed approaches ensure a probability of correct classification above the 98%.

Moreover, the proposed classifier as been compared with the traditional k-NN classifier, where the probability of correct classification achieves $91,76\%$ on the cross-validation dataset. In this comparison, to empathise the ability of deep learning models to learn features directly from the data without the need for manual feature extraction, the same input sequences used for BiLSTM architecture are used also for the k-NN. Higher classification accuracy can be achieved with k-NN in case manual feature extraction is applied on the input sequence. However, if feature extraction is important for the performance of the k-NN classifier, the selection of the number of layer as well as the size of the hidden units is crucial for correct classification with deep learning architectures. The performance of the network is highly correlated to several factors, such as the complexity of the dataset, the number of features and the number of data points. For this reason, many trials are required to tune properly all the network parameters.

In conclusion, from the analysis of the performance, one can deduce that the proposed framework outperform the traditional k-NN, and a deep learning based classifier is reliable for the classification of faulty speaker.

Chapter 8

Conclusions and future works

8.1 Conclusion

The research presented in this thesis investigated new signal processing solutions capable of providing increased performance and a better sound quality for advanced acoustic system in realistic conditions. From loudspeakers manufacturers perspective, the intelligibility of sound can be affected on both downstream and upstream side of a loudspeaker production chain, differently. On downstream side, the sound is partially reflected by the physical boundaries of the environment, leading to reverberation, echo and feedback problems. On the upstream side, the quality of the driver could be compromised during the production stage. In this regard, two solutions were developed respectively on downstream and upstream side of a loudspeaker production chain.

On the downstream side, the concept of the acoustic feedback problem was taken in consideration and presented in Chapter 2. Both feedforward suppression and feedback cancellation techniques were introduced, with a particular focus on Adaptive Feedback Cancellation (AFC) technique. Furthermore, different solutions mainly used in hearing aids domain were also introduced in details, taking in account both room impulse response and source signal covariance matrices, with the aim to consider into the optimum estimator the a-priori knowledge of the scenario and whitening the source signal components, respectively.

In order to present a solution on the upstream side as well, the basic concept of micro-Doppler effect in radar was introduced in Chapter 3. The Doppler effect and the canonical form of a received radar signal were discussed with the aim to review the uses and models of micro-Doppler. In order to understand how to extract the micro-Doppler signature, the most used Time-Frequency Distribution (TFD) for observing how the frequencies components of a signal varies on time were exhaustively introduced, describing the trade-off which each function poses in terms of time-frequency resolution, computational complexity and production of artefacts. Furthermore, the basic principle of rigid body motion in the context of radar signal processing were also analysed to understand the effect of translation and rotation of a target. For a good interpretation of the received radar echoes from a vibrating surface such as a loudspeaker and a better understanding of the effect of non linear motion dynamic, two related canonical cases were analysed, namely micro-Doppler induced by a vibrating point and pendulum oscillation. Finally, some of the aspects of the electrodynamic transducer motion, and how to acoustically characterize the behaviour of a speaker were also introduced. The concepts introduced in Chapter 3 were successively exploited to detect, confirm and characterize loudspeaker behaviour through radar micro-Doppler signature, focusing mainly on the rigid body motion of the acoustic driver.

Motivated by recent advances in deep learning application in different fields, a brief overview of different deep learning architectures was provided in the Chapter 4. After introducing the most general architecture for deep learning, together with Convolutional Neural Network (CNN) and how they can be used in the radar domain, a deeper look was given on Recurrent Neural Network (RNN). Finally, to cope with the vanishing gradient problem that affect the performance of RNNs, the Long Short-Time Memory (LSTM) was also introduced.

In Chapter 5, a new framework to tackle the acoustic feedback problem in large acoustic spaces was presented as downstream solution of a loudspeaker production chain. It is based on the Frequency Domain Adaptive Filtering (FDAF) implementation of the Normalized Least Mean Square (NLMS) algorithm. Since the traditional LS-based adaptive filtering algorithm converge to a biased solution of the acoustic feedback path due to a considerable correlation between loudspeaker and microphone signals, a signal decorrelation method was used. Inspired by hearing aids device, the Prediction Error Method (PEM) was introduced. In order to further decrease the bias into the estimated feedback path the a-priori knowledge was considered. Based on acoustic set-up information (distance between loudspeaker and microphone, acoustic absorption of the walls, room volume, etc.), a robust estimator of the RIR covariance matrix was obtained from Sabine

3-parameter RIR model. Applying the Levenberg-Marquardt Regularization (LMR), the PEM based LMR NLMS was obtained. Dealing with long impulse response, meaning higher computational costs, a more efficient solution needed to be evaluated. For this reason, the Partitioned Block approach was considered: it consists of slicing the feedback path in p segments of length P each. Moving towards the IR tail, the loop gain $|G\left(q,n\right)F\left(q,n\right)|$ showed a lower energy, thus producing a degraded estimation. To compensate this, a slower adaptation speed was used, leading to the PBTD version of the algorithm with Variable Step Size. Performance analysis shown that PBTD version with Variable Step Size had a slower convergence rate and a higher computational cost than PEM based LMR-NLMS algorithm. A faster convergence was achieved by designing the algorithm in the frequency domain. Finally, in order to take the full advantages of the a-priori knowledge of the scenario, such as public transportation facilities, live and recorded music venues, or any other venues where a reference impulse response is available, a constrained adaptation could be used, leading to the proposed PEM based PBFD with VSS. The results shown that this technique outperform previous approaches, achieving a lower estimation error and a faster convergence rate. The results of the proposed framework were compared with the state of the art using real acoustic data showing superior performance with up to 18dB Maximum Stable Gain (MSG) and 30 seconds less convergence time. However, the proposed algorithm showed significant limitations. Due to the simplicity of source signal model, the algorithm showed good results on pre-recorded speech signal. In case of sound or music signals are considered, the algorithm would not be able to achieve the same performance: a more complex source signal model should be used in this case. Furthermore, a low misalignment and high gain is achieved by considering the a priori knowledge of the RIR. In case a priori knowledge is not available, the regularization methods could no be applied leading to a degraded estimation of the feedback path.

A solution on the upstream side of the loudspeaker production chain was provided in Chapter 6, where a novel approach for condition monitoring of loudspeakers based on radar micro Doppler was proposed. Due to hardware limitation, the radar micro-Doppler based loudspeaker analysis was restricted to woofer-type of speaker. With the assumption of rigid body motion at low frequency, the displacement of a loudspeaker was modelled as function of the frequency of the stimulus by considering the electro-mechanical components responsible of the dynamic response of the transducer, in both single tone and sine sweep analysis.

In the first case, the phase, the spectrum and the spectrogram of the received radar signal were compared to those from the model confirming that loudspeaker behaviour can be detected from radar. In particular, taking in account both micro-Doppler shift and the phase component of the received signal, the information of the displacement motion were extracted. By increasing the voltage applied to the terminals of the driver, a resulting discrepancy between real and simulated signal appeared due to the non linear effects of the speaker. When the sine sweep test signal was used, some discrepancy between real and simulated signal appeared, as the rocking modes effect were not been taken in consideration in the displacement model. Nevertheless, the spectral analysis results demonstrated good ability in detecting irregular defects affecting the motion of the voice coil. Finally, a matched filter based approach was proposed to mechanically characterise the speaker. Cross power spectrum density, linear frequency response and harmonic frequency responses were analysed and considered as powerful indicator of possible manufacturing problems.

In order to detect and classify automatically faulty speakers, a framework based on radar and deep learning technologies was also designed in Chapter 7. By using a 24GHz CW radar, a real dataset has been obtained by acquiring in laboratory the signals scattered by the DUTs: due to the difficulty of acquiring samples with labelled defects, two different manipulations have been artificially applied to the speaker, namely on cone and spider. The output of the matched filter shows discrepancies between mechanical impulse response of the speaker without defects and the ones with defects. To restrict the number of samples, Fourier Transform is applied on both linear and harmonic impulse responses. In this way, mechanical frequency responses are obtained and used as input vector of the network.
Handling the mechanical frequency responses as time series, allow the use of RNNs network for classification purposes. Since the full sequence, representing the mechanical frequency response, is accessible at prediction time, a combination of BRNN and LSTM architecture is preferred, leading to the BiLSTM architecture introduced in 7.3. The reliability of deep learning based classifier has been demonstrated by testing the network on training, validation and test dataset. The results have shown that, for all the real data, the proposed approaches ensure a probability of correct classification above the 98%, outperforming the traditional k-NN classifier, used as benchmark.

8.2 Future works

On the downstream side of loudspeaker production chain, the PBFD approach could be used on board of loudspeakers array equipped with an integrated cutting edge DSP (e.g. Qflex series of Tannoy Ltd) for real time implementation. In this domain, the use of adaptive filtering could be limited. In case high sampling rate is required to obtain a good sound quality (especially for audio applications), the impulse response would be densely sampled hence requiring many coefficients, and moreover, a large number of adaptive filter iterations has to be performed per second. Another great challenge in acoustic feedback control, and in AFC in particular, is to generalize the methods proposed in a single-channel context to multichannel systems. Since the number of acoustic feedback paths in a multichannel system equals the number of loudspeakers times the number of microphones, the AFC computational complexity can be expected to increase very quickly in a multi-channel context. For this reason, high complexity puts a limit on the generalization of the AFC approach also to multi-channel systems. Since no results are available on how to exploit the fact that the different acoustic feedback path impulse responses of a multichannel system share some underlying room acoustic properties, the state of the art in multi-channel AFC consists in applying $S \cdot L$ single-channel AFC algorithms in a system having S microphones and L loudspeakers, hence the resulting complexity also increases with a factor $S{\cdot}L$.

On the upstream side of loudspeaker production chain, the use of the new generation of radar, namely mm-wave radars would be more sensitive to smaller displacement leading to higher Doppler shifts and fine range resolution, if used in FMCW mode. These would allow to perform different kind of signal processing, such as the range-Doppler map: it could be possible to slice the radar return in range, in order to get localized responses of the driver. Thus, tweeters and micro speakers could be also analysed. The use of mm-wave radar could be introduced in additional manufacturing applications. For example, in the same loudspeaker testing domain it could be integrated in Linear Suspension Testing aimed at assessing the quality control of moving parts. In other domains, the proposed technique could find application in testing of lightweight components (i.e.: made of carbon fiber) for aerospace use as well as in vibration analysis of machines such as the gearbox of a wind turbine.

In the context of deep learning applied to loudspeaker testing domain, a combination of CNN and BiLSTM based classifier on micro-Doppler signature and mechanical frequency response could be used to increase the performance capability, as well as the amount of classes to be detected. Finally, it would be nice to find the correlation between acoustic and radar measurements: in this domain deep learning approach could be used to estimate the acoustic frequency response of the driver starting from its mechanical frequency response.

References

[1] D. W. Stacks and M. B. Salwen. *An Integrated Approach to Communication Theory and Research.* Routledge, 2009.

[2] T. D. Rossing. *Handbook of acoustics.* Springer, 2007.

[3] G. Schmidt. Application of acoustic echo control - an overview. In *12th European Signal Processing Conference (EUSIPCO 2004)*, pages 9–16, Sept 2004.

[4] T. VanWaterschoot and M. Moonen. Fifty years of acoustic feedback control: State of the art and future challenges. *Proceedings of IEEE*, 99(2):288–327, February 2011.

[5] H. Schepker, S. E. Nordholm, L. T. T. Tran, and S. Doclo. Null-steering beamformer-based feedback cancellation for multi-microphone hearing aids with incoming signal preservation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(4):679–691, April 2019.

[6] M. G. Siqueira and A. Alwan. Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids. *IEEE Transactions on Speech and Audio Processing*, 8(4):443 – 453, July 2000.

[7] A. Spriet, M. Moonen, and I. Proudler. Feedback cancellation in hearing aids: an unbiased modelling approach. In *11th Europian Signal Processing Conference (EUSIPCO '02)*, September 2002.

[8] S. S. Yaxleys. *The First Outside Broadcast 1915.* 2002. History of PA Charity Trust. Retrieved 25 November 2011.

[9] R. L. Weaver and O. I. Lobkis. On the linewidth of the ultrasonic larsen effect in a reverberant body. *The Journal of the Acoustical Society of America*, 120(1):102–109, 2006.

[10] T. VanWaterschoot and M. Moonen. Assessing the acoustic feedback control performance of adaptive feedback cancellation in sound reinforcement systems. In *17th European Signal Processing Conference (EUSIPCO '09)*, August 2009.

[11] J. C. Willems. *The Analysis of Feedback Systems.* MIT Press, 1970.

[12] H. Nyquist. Regeneration theory. *Bell System Technical Journal*, pages 126–147, January 1932.

[13] G. Rombouts, T. VanWaterschoot, K. Struyve, and M. Moonen. Acoustic feedback cancellation for long acoustic paths using a nonstationary source model. *IEEE Transactions on Signal Processing*, 54(9):3426–3434, August 2006.

[14] M. R. Schroeder. Improvement of acoustic feedback stability by frequency shifting. *The Journal of Acoustic Society of America*, 36(9):1718–1724, September 1964.

[15] P. Mapp and C. Ellis. Improvements in acoustic feedback margin in sound reinforcement systems. In *106th Convention of Audio Engigneer Socity (AES)*, May 1999.

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

[17] J. Patronis. Acoustic feedback detector and automatic gain control. *United States Patent*, March 1978.

[18] S. Ando. Howling detection and prevention circuits and a loudspeaker system employing the same. *United States Patent*, June 2001.

[19] M. Hanajima. Howling eliminating appaaratus. *United States Patent*, September 2000.

[20] P. G. Cacho, T. VanWaterschoot, M. Moonen, and S. H. Jensen. Regularized adaptive notch filters for acoustic howling suppression. In *17th Europian Singanl Processing Conference (EUSIPCO'09)*, August 2009.

[21] S. Kuehl, C. Anemueller, C. Antweiler, P. Jax, F. Heese, and P. Vicinus. Acoustic howling detection and suppression for ip-based teleconference systems. In *Speech Communication; 13th ITG-Symposium*, pages 1–5, Oct 2018.

[22] T. Jithin, K.K. Mohamed Salih, and A.R. Jayan. Real time suppression of howling noise in public address system. *Procedia Technology*, 24:933 – 940, 2016. International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2016.

[23] Philipp Bulling, Klaus Linhard, Arthur Wolf, and Gerhard Schmidt. Automatic equalization for in-car communication systems. In André Berton, Udo Haiber, and Wolfgang Minker, editors, *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2018*, pages 7–14. TUDpress, Dresden, 2018.

[24] T. Van Waterschoot. *Design and evaluation of Digital Signal Processing algorithms for acoustic feedback and echo cancellation.* Katholieke Universiteit Leuven, 2009.

[25] S. Haykin and K. J. R. Liu. *Handbook on Array Processing and Sensor Networks.* Wiley-IEEE press, 2010.

[26] M. R. Schroeder. Improvement of acoustic feedback stability in public address system. *The Journal of Acoustic Society of America*, 31(6):851, 1959.

[27] M. R. Schroeder. Improvement of feedback stability of public address systems by frequency shifting. In *13rd Audio Engigneer Socity (AES) Convention*, October 1961.

[28] C. V. Deutschbein. Digital frequency shifting for electroacoustic feedback suppression. In *118th Audio Engigneer Socity (AES) Convention*, May 2005.

[29] Etienne Thuillier, Otso Lähdeoja, and Vesa Välimäki. Feedback control in an actuated acoustic guitar using frequency shifting. *J. Audio Eng. Soc*, 67(6):373–381, 2019.

[30] L.N. Mishin. A method for increasing the stability of sound amplification systems. *Soviet Physics Acoustic*, 4:64–71, 1958.

[31] G. Nishinomiya. Improvement of acoustic feedback stability of public address system by warbling. *6th International Congress of Acoustics (ICA) Convention*, E:93–96, 1968.

[32] M. Guo, S. H. Jensen, J. Jensen, and S. L. Grant. On the use of a phase modulation method for decorrelation in acoustic feedback cancellation. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 2000–2004, Aug 2012.

[33] M. A. Poletti. The stability if multichannel sound system with frequency shifting. *The Journal of Acoustic Society of America*, 116(2):853–871, August 2004.

[34] G. Schmidt and T. Haulick. Signal processing for in car communication systems. *Signal Processing, Elsevier*, 86(6):1307–1326, June 2006.

[35] Wancheng Zhang, A. W. H. Khong, and P. A. Naylor. Adaptive inverse filtering of room acoustics. In *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 788–792, Oct 2008.

[36] T. VanWaterschoot, G. Rombouts, and M. Moonen. On the performance of decorrelation by prefiltering for adaptive feedback cancellation in public address systems. In *IEEE Benelux Signal Processing Symposium (SPS 2004)*, April 2004.

[37] S. H. Ibaraki, H. N. Furukawa, and H. H. Naono. Howling canceller. *United State Patent*, May 1988.

[38] J. H. Ibaraki and N. D. Wells. Method and apparatus fro reduction of unwanted feedback. *United State Patent*, July 2001.

[39] A. Goertz. An adaptive subtraction filter for feedback cancellation in public address sound systems. In *15th International Congress on Acoustics (ICA '95)*, June 1995.

[40] C. P. Janse and H. J. W. Belt. Sound reinforcement system having an echo suppressor and loudspeaker beamformer. *World Intellectual Property Organization (WIPO)*, February 2003.

[41] S. Kamerling, K. Janse, and F. Van Der Meulen. Feedback cancellation in hearing aids: results from using frequency-domain adaptive filters. In *IEEE International Symposium on Circuits and Systems (ISCAS '94)*, June 1994.

[42] J. Hellgren and U. Forssell. Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification. *IEEE Transactions on Speech and Audio Processing*, 9(7):906 – 913, November 2001.

[43] L. T. T. Tran, S. E. Nordholm, H. Schepker, H. H. Dam, and S. Doclo. Two-microphone hearing aids using prediction error method for adaptive feedback control. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(5):909–923, May 2018.

[44] T. VanWaterschoot, K. Eneman, and M. Moonen. Instrumental variable methods for acoustic feedback cancellation. Technical Report ESAT-SISTA TR 05-14, Katholieke Universiteit Leuven, Belgium, October 2004. Available at https://ftp.esat.kuleuven.be/pub/sista/vanwaterschoot/abstracts/05-14.html.

[45] J. M. Gil-Cacho, T. VanWaterschoot, M. Moonen, and S. H. Jensen. A frequency-domain adaptive filter (fdaf) prediction error method (pem) framework for double-talk-robust acoustic echo cancellation. *IEEE Transaction on Audio, Speech and language processing*, 22(12):2074–2086, December 2014.

[46] G. Bernardi, T. van Waterschoot, J. Wouters, and M. Moonen. Adaptive feedback cancellation using a partitioned-block frequency-domain kalman filter approach with pem-based signal prewhitening. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(9):1784–1798, Sep. 2017.

[47] J. Franzen and T. Fingscheidt. Improved measurement noise covariance estimation for n-channel feedback cancellation based on the frequency domain adaptive kalman filter. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 965–969, May 2019.

[48] J. A. Maxwell and P. M. Zurek. Reducing acoustic feedback in hearing aids. *IEEE Transactions on Speech and Audio Processing*, 3(4):304–313, July 1995.

[49] J.M. Kates. Feedback cancellation in hearing aids: results from a computer simulation. *IEEE Transactions on Signal Processing*, 39(3):553 – 562, March 1991.

[50] S. Thipphayathetthana and C. Chinrungrueng. Variable step-size of the least-mean-square algorithm for reducing acoustic feedback in hearing aids. In *IEEE Asia-Pacific Conference on Circuits and Systems (APCCAS 2000), Electronic Communication Systems*, December 2000.

[51] T. VanWaterschoot, G. Rombouts, and M. Moonen. Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement. *Signal Processing, Elsevier*, 88(3):594–611, March 2008.

[52] B. Baykal and A. G. Constantinides. Underdetermined-order recursive least-squares adaptive filtering: the concept and algorithms. *IEEE Transactions on Signal Processing*, 45(2):346–362, July 1997.

[53] S. M. Key. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice-Hall, 1993.

[54] U. Forssell and L. Ljungs. Closed loop identification revisited. *Automatica, Elsevier*, 35(7):1215–1241, July 1999.

[55] G. Rombouts, T. VanWaterschoot, and M. Moonen. Robust and efficient implementation of the pem-afrow algorithm for acoustic feedback cancellation. *Journal of Audio Engeneer Socity*, 55(11):955–966, November 2007.

[56] A. Spriet, I. Proudler, M Moonen, and J. Wouters. Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal. *IEEE Transactions on Signal Processing*, 53(10):3749–3763, October 2005.

[57] T. VanWaterschoot, G. Rombouts, P. Verhoeve, and M. Moonen. Double talk robust prediction error identification algorithm for acoustic echo cancellation. *IEEE Transactions on Signal Processing*, 55(3):846–858, March 2007.

[58] T. VanWaterschoot and M. Moonen. Comparison of linear prediction models for audio signals. *EURASIP Journal on Audio, Speech, and Music Processing*, December 2009.

[59] W.C. Sabine. *Collected papers on acoustics.* Peninsula Publishing, Los Altos, CA, 1992.

[60] T. VanWaterschoot, G. Rombouts, and M. Moonen. Towards optimal regularization by incorporating prior knowledge in an acoustic echo canceller. In *2005 International Workshop on Acoustic Echo and Noise Control (IWAENC 2005)*, September 2005.

[61] M. Caris, W. Johannes, S. Sieger, V. Port, and S. Stanko. Detection of small uas with w-band radar. In *2017 18th International Radar Symposium (IRS)*, pages 1–6, June 2017.

[62] M. Alizadeh, G. Shaker, and S. Safavi-Naeini. Remote heart rate sensing with mm-wave radar. In *2018 18th International Symposium on Antenna Technology and Applied Electromagnetics (ANTEM)*, pages 1–2, Aug 2018.

[63] S. Wang. A novel ultra-wideband 80 ghz fmcw radar system for contactless monitoring of vital signs. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4978–4981, 2015.

[64] V. C. Chen. *The Micro-Doppler Effect in Radar.* Artech House, 2019.

[65] IEEE. Ieee standard for radar definitions. *IEEE Std 686-2017 (Revision of IEEE Std 686-2008)*, pages 1–54, Sep 2017.

[66] C. Clemente, A. Balleri, K. Woodbridge, and J. J. Soraghan. Developments in target micro-doppler signatures analysis: radar imaging, ultrasound and through-the-wall radar. *EURASIP Journal on Advances in Signal Processing*, 2013(1):47, Mar 2013. Available at https://doi.org/10.1186/1687-6180-2013-47.

[67] V. C. Chen, F. Li, S. Ho, and H. Wechsler. Micro-doppler effect in radar: phenomenon, model, and simulation study. *IEEE Transactions on Aerospace and Electronic Systems*, 42(1):2–21, Jan 2006.

[68] D. Gaglione, C. Clemente, F. K. Coutts, G. Li, and J. J. Soraghan. Model-based sparse recovery method for automatic classification of helicopters. In *2015 IEEE Radar Conference (RadarCon)*, pages 1161–1165, May 2015.

[69] C. Clemente, L. Pallotta, A. De Maio, J. J. Soraghan, and A. Farina. A novel algorithm for radar classification based on doppler characteristics exploiting orthogonal pseudo-zernike polynomials. *IEEE Transactions on Aerospace and Electronic Systems*, 51(1):417–430, January 2015.

[70] V. Agnihotri, M. Sabharwal, and V. Goyal. The extraction of key distinct features for identification and classification of helicopters using micro-doppler signatures. In *2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, pages 893–896, Aug 2019.

[71] R. I. A. Harmanny, J. J. M. de Wit, and G. P. Cabic. Radar micro-doppler feature extraction using the spectrogram and the cepstrogram. In *2014 11th European Radar Conference*, pages 165–168, Oct 2014.

[72] Y. Zhao and Y. Su. The extraction of micro-doppler signal with emd algorithm for radar-based small uavs' detection. *IEEE Transactions on Instrumentation and Measurement*, 69(3):929–940, March 2020.

[73] A. R. Persico, C. Clemente, D. Gaglione, C. V. Ilioudis, J. Cao, L. Pallotta, A. De Maio, I. Proudler, and J. J. Soraghan. On model, algorithms, and experiment for micro-doppler-based recognition of ballistic targets. *IEEE Transactions on Aerospace and Electronic Systems*, 53(3):1088–1108, June 2017.

[74] E. Alhadhrami, M. Al-Mufti, B. Taha, and N. Werghi. Learned micro-doppler representations for targets classification based on spectrogram images. *IEEE Access*, 7:139377–139387, 2019.

[75] Y. Kim, S. Choudhury, and H. Kong. Application of micro-doppler signatures for estimation of total energy expenditure in humans for walking/running activities. *IEEE Access*, 4:1560–1569, 2016.

[76] W. Ye and H. Q. Chen. Human activity classification based on micro-doppler signatures by multi-scale and multi-task fourier convolutional neural network. *IEEE Sensors Journal*, pages 1–1, 2020.

[77] A. Assmann, A. Izzo, and C. Clemente. Efficient micro-doppler based pedestrian activity classification for adas systems using krawtchouk moments. In *11th International Conference on Mathematics in Signal Processing (IMA)*, Dec 2016. Available at https://strathprints.strath.ac.uk/id/eprint/66617.

[78] X. Huang, J. Ding, D. Liang, and L. Wen. Multi-person recognition using separated micro-doppler signatures. *IEEE Sensors Journal*, pages 1–1, 2020.

[79] L. Cohen. *Time-frequency Analysis: Theory and Applications.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1995.

[80] G. Heinzel, A. Rüdiger, and S. Rüdiger. Spectrum and spectral density estimation by the discrete fourier transform (dft), including a comprehensive list of window functions and some new flat-top windows. *Max Plank Inst*, 12, 01 2002.

[81] E.F. Knott, J.F. Shaeffer, and M.T. Tuley. *Radar Cross Section.* Radar Library. Artech House, 1993. Available at https://books.google.co.uk/books?id=YhQntAEACAAJ.

[82] M.A. Richards, W.A. Holm, and J. Scheer. *Principles of Modern Radar: Basic Principles.* Electromagnetics and Radar. Institution of Engineering and Technology, 2010. Available at https://books.google.co.uk/books?id=nD7tGAAACAAJ.

[83] H. Gao, L. Xie, S. Wen, and Y. Kuang. Micro-doppler signature extraction from ballistic target with micro-motions. *IEEE Transactions on Aerospace and Electronic Systems*, 46(4):1969–1982, Oct 2010.

[84] Leo L. Beranek and Tim J. Mellow. Chapter 6 - electrodynamic loudspeakers. In Leo L. Beranek and Tim J. Mellow, editors, *Acoustics: Sound Fields and Transducers*, pages 241–288. Academic Press, 2012.

[85] W. Klippel and J. Schlechter. Measurement and visualization of loudspeaker cone vibration. In *Audio Engineering Society Convention 121*, Oct 2006.

[86] W. Klippel. Tutorial: Loudspeaker nonlinearities - causes, parameters, symptoms. *Journal of Audio Engineering Society*, 54(10):907–939, 2006.

[87] Guangjian Ni and S. J. Elliott. Wave interpretation of numerical results for the vibration in thin conical shells. *Journal of Sound and Vibration*, 333(10):2750–2758, May 2014.

[88] W . Klippel and J. Schlechter. Distributed mechanical parameters describing vibration and sound radiation of loudspeaker drive units. In *Audio Engineering Society Convention 125*, Oct 2008.

[89] W. Klippel. Scanning Vibrometer C5 Hardware and Software Module of the KLIPPEL R&D SYSTEM. Technical report, Klippel GmbH, Mendelssohnallee 30, 01309 Dresden, Germany, 2008.

[90] Angelo Farina. Advancements in impulse response measurements by sine sweeps. In *Audio Engineering Society Convention 122*, May 2007.

[91] A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio Engineering Society Convention 108*, Feb 2000.

[92] M. C. Bellini, L. Collini, A. Farina, D. Pinardi, and K. Riabova. Measurement of loudspeakers with a laser doppler vibrometer and the exponential sine sweep excitation technique. *Journal of Audio Engineering Society*, 65(7/8):600–612, 2017.

[93] Y. Lecun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[94] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning.* MIT Press, 2016. Available at https://www.deeplearningbook.org.

[95] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Springer-Verlag, Berlin, Heidelberg, 2006.

[96] S. Amari. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4):185 – 196, 1993.

[97] Michael A. Nielsen. *Neural Network and Deep Learning.* Determination Press, 2015. Available at https://neuralnetworksanddeeplearning.com.

[98] D.A.E. Morgan. Deep convolutional neural networks for atr from sar imagery. volume 9475, 2015.

[99] M. Wilmanski, C. Kreucher, and J. Lauer. Modern approaches in deep learning for sar atr. volume 9843, 2016.

[100] Sun J. Wang J. Hussain A. Yang E.) year=2019 month=Dec. title=A New Algorithm for SAR Image Target Recognition Based on an Improved Deep Convolutional Neural Network journal=Cognitive Computation pages=809-824 volume= 11 number=6 Gao F., Huang T.

[101] A.M. Elbir, K.V. Mishra, and Y.C. Eldar. Cognitive radar antenna selection via deep learning. *IET Radar, Sonar and Navigation*, 13(6):871–880, 2019.

[102] D. Brodeski, I. Bilik, and R. Giryes. Deep radar detector. In *2019 IEEE Radar Conference (RadarConf)*, pages 1–6, April 2019.

[103] Y. Kim and T. Moon. Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1):8–12, 2016.

[104] Y. Kim and H. Ling. Human activity classification based on micro-doppler signatures using an artificial neural network. 2008.

[105] Yang Yang, Chunping Hou, Yue Lang, Dai Guan, Danyang Huang, and Jinchen Xu. Open-set human activity recognition based on micro-doppler signatures. *Pattern Recognition*, 85:60 – 69, 2019.

[106] Y. Kim and B. Toomajian. Hand gesture recognition using micro-doppler signatures with convolutional neural network. *IEEE Access*, 4:7125–7130, 2016.

[107] M. G. Amin, Z. Zeng, and T. Shan. Hand gesture recognition based on radar micro-doppler signature envelopes. In *2019 IEEE Radar Conference (RadarConf)*, pages 1–6, April 2019.

[108] A. Ng, J. Ngiam, C. Y. Foo, Y. Mai, C. Suen, A. Coates, A. Maas, A. Hannun, B. Huval, T. Wang, and S. Tandon. *Deep Learning*. Available at https://deeplearning.stanford.edu/tutorial,StanfordUniversity, ComputerScienceDepartment.

[109] C. Olah. *Convolutional Neural Networks*. Available at https://colah.github.io/posts/tags/convolutional_neural_networks.html.

[110] Adit Deshpande. A beginner's guide to understanding convolutional neural networks, 2016.

[111] Arden Dertat. Applied deep learning: Convolutional neural network, 2017.

[112] S. Ghelani. *Text classification - RNN's or CNN's*. Available at https://towardsdatascience.com/text-classification-rnns-or-cnn-s-98c86a0dd361.

[113] C. Olah. *Understanding LSTM network*. Available at https://colah.github.io/posts/2015-08-Understanding-LSTMs.

[114] A. Amidi and S. Amidi. *Recurrent Neural Networks cheatsheet*. Available at https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks.

[115] F. M. Bianchi. *Recurrent Neural Networks - A quick overview*. Machine Learning Group, University of Tromso, available at https://www.sintef.no/contentassets/9689621c8cda44d9b2bc91c5fba21135/bianchi.pdf.

[116] Q. Wang and M. Iwaihara. Deep neural architectures for joint named entity recognition and disambiguation. In *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 1–4, 2019.

[117] D. Britz. *Artificial Intelligence, Deep Learning and NLP*. Wildml. Available at https://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns.

[118] P.J. Werbos. Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.

[119] E. Alese. *RNN Training - Deriving the gradients for Backward propagation in RNN*. Available at https://medium.com/learn-love-ai/step-by-step-walkthrough-of-rnn-training-part-ii-7141084d274b.

[120] E. Alese. *The curious case of the vanishing and exploding gradient.* Available at https://medium.com/learn-love-ai/the-curious-case-of-the-vanishing-exploding-gradient-bf58ec6822eb.

[121] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[122] T. Mikolov R. Pascanu and Y. Bengio. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28, page 1310–1318, 2013.

[123] A. Graves and J. Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, 18(5-6):602–610, 2005.

[124] R. Spagnolo. *Manuale di acustica applicata.* UTET Libreria, 2002.

[125] S. Haykin. *Adaptive Filter Theory, 5th edition.* Pearson, 2014.

[126] A. Spriet, G. Rombouts, M. Moonen, and J. Wouters. Adaptive feedback cancellation in hearing aids. *Elseiver, Journal of the Franklin Institute*, 343(3):545–573, August 2006.

[127] H. Schepker, L.T.T. Tran, S. Nordholm, and S. Doclo. Improving adaptive feedback cancellation in hearing aids using an affine combination of filters. *IEEE ICASIP 2016*, 88(3):231–235, March 2016.

[128] J.J. Shynk. Frequency-domain and multirate adaptive filtering. *IEEE Signal Processing Magazine*, 9(1):14–37, January 1992.

[129] A. Izzo, L. Ausiello, C. Clemente, and J. J. Soraghan. Radar micro-doppler for loudspeaker analysis: An industrial process application. In *International Conference on Radar Systems (Radar 2017)*, pages 1–6, Oct 2017.

[130] M. A. Richards, J. A. Scheer, and W. A. Holm. *Principles of Modern Radar.* Number v. 1 in Principles of Modern Radar. SciTech Publishing, Incorporated, 2010.

[131] Audiomatica S.r.l. Clio poket 2.0, 2001. http://www.audiomatica.com/wp/?page_id=2429.

[132] B & C Speaker. B & C 10CL51 LF driver, 1946. https://www.bcspeakers.com/en/products/lf-driver/10-0/8/10cl51-8.

[133] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR) 2014*, April 2014.