NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

FACULTY OF SCIENCE
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

POSTGRADUATE PROGRAM

MASTER THESIS

Real-time high-resolution delay estimation in audio communication using inaudible pilot signals

Vaggelis G. Alexiou

Supervisor: **Alexandros Eleftheriadis,** Associate Professor

ATHENS

MAY 2014

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Εκτίμηση καθυστέρησης με υψηλή ακρίβεια κατά την μετάδοση ήχου σε συστήματα πραγματικού χρόνου χρησιμοποιώντας μη ακουστά πιλοτικά σήματα

Βαγγέλης Γ. Αλεξίου

Επιβλέπων: **Αλέξανδρος Ελευθεριάδης,** Αναπληρωτής Καθηγητής

ΑΘΗΝΑ

ΜΑΙΟΣ 2014

MASTER THESIS

Real-time high-resolution delay estimation in audio communication using inaudible pilot signals

Vaggelis G. Alexiou

Reg. Nu.: M1222

SUPERVISOR: **Alexandros Eleftheriadis,** Associate Professor

EXAMINATION COMMITTEE: **Sergios Theodoridis,** Professor

May 2014

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Εκτίμηση καθυστέρησης με υψηλή ακρίβεια κατά την μετάδοση ήχου σε συστήματα πραγματικού χρόνου χρησιμοποιώντας μη ακουστά πιλοτικά σήματα

Βαγγέλης Γ. Αλεξίου
Α.Μ.: M1222

ΕΠΙΒΛΕΠΩΝ: **Αλέξανδρος Ελευθεριάδης,** Αναπληρωτής Καθηγητής

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: **Σέργιος Θεοδωρίδης,** Καθηγητής

Μαίος 2014

Abstract

We describe the construction of inaudible pilot signals and associated receiver processing techniques that can be used in audio communication, so that an accurate estimation of the signal's end-to-end delay can be obtained. Considering the properties of the human auditory system, we structure the pilot signal so that it is acoustically untraceable and maximizes the time accuracy of the delay estimation. Analysis and mathematical modeling of our system is given, including using modulation techniques, pulse designs that could be adopted, receiver's processing and calculation of Power Spectral Density of the pilot signals that are finally proposed. We provide simulation results verifying the effectiveness of the techniques, and also demonstrate their superiority compared to systems with widely used pulses in similar schemes. The probability of error is derived assuming accuracy in the order of one sample. We also examine their performance under compression of the complex audio signal, and finally, we explore the capabilities of the system requiring fractional delay estimation.

SUBJECT AREA: Audio Communication

KEYWORDS: delay estimation, inaudible pilot signal, real-time systems, multimedia - audio, network music performance

Περίληψη

Μελετήθηκε η κατασκευή σημάτων- πιλοτικών καθώς επίσης και τεχνικές επεξεργασίας στον δέκτη που μπορούν να υιοθετηθούν κατά τη διάρκεια επικοινωνίας με ήχο, προκειμένου να εκτιμήσουμε με μεγάλη ακρίβεια την καθυστέρηση που υπόκειται ένα ηχητικό σήμα από το ένα τερματικό εως το άλλο. Λαμβάνοντας υπόψη τις ιδιότητες του ανθρώπινου μοντέλου ακουστότητας, κατασκευάζουμε τον πιλότο ώστε να μην γίνεται αντιληπτός από το ανθρώπινο αυτί και επηρεάζει την εμπειρία του χρήστη, καθώς ο πιλότος εισάγεται στο ηχητικό σήμα κατά την μετάδοση του τελευταίου. Ταυτόχρονα σκοπός είναι να μεγιστοποιήσουμε την ακρίβεια των εκτιμήσεων της καθυστέρησης που λαμβάνουμε στον δέκτη. Παρατίθεται μαθηματική μοντελοποίηση του συστήματος, ανάλυση τεχνικών διαμόρφωσης του σήματος και πιθανών παλμών προς χρήση, καθώς επίσης και υπολογισμός της φασματικής πυκνότητας ισχύος των προτεινόμενων πιλότων. Παρέχονται αποτελέσματα προσομοιώσεων που επαληθεύουν την αποδοτικότητα των τεχνικών που επιλέγονται και αναδεικνύουν την υπεροχή τους, σε σύγκριση με παραπλήσια συστήματα που υιοθετούν ευρέως χρησιμοποιούμενους παλμούς. Η πιθανότητα σφάλματος με ακρίβεια ενός δείγματος θεωρείται η κατάλληλη μετρική για τον έλεγχο της επίδοσης του συστήματος που προτείνουμε, καθώς απαιτείται υψηλή ακρίβεια. Επιπλέον, εξετάζουμε την απόδοση του συστήματος αφότου περάσει το σύνθετο ηχητικό σήμα από συμπίεση, και τέλος υπολογίζεται η πιθανότητα σφάλματος με ακρίβεια κλασματική του ενός δείγματος.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Επικοινωνία με ήχο

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: εκτίμηση καθυστέρησης, μη ακουστός σήμα-πιλότος, συστήματα πραγματικού χρόνου, πολυμέσα - ήχος, δικτυακη μουσική εκτέλεση

Συνοπτική Παρουσίαση

Το αντικείμενο της εργασίας αυτής είναι η μελέτη και κατασκευή ψηφιακού σήματος πιλότου για τον καθορσιμό της καθυστέρησης που υπόκειται ένα ηχητικό σήμα, κατά την διάδοσή του είτε μέσω διαδικτύου, είτε μέσω ενός αναλογικού καναλιού που εισάγει καθυστέρηση σε αυτό, υπό την παρουσία μουσικής εκτέλεσης.

Ο πιλότος εισάγεται κατά την λήψη του ηχητικού σήματος από έναν Η/Υ και μεταφέρονται μαζί στον δέκτη, όπου και αναπαράγονται. Προφανώς, ο πιλότος θα πρέπει να μην είναι αντιληπτός. Το πρωτότυπο κομμάτι της μελέτης, αφορά την επιλογή και την κατασκευή ενός σήματος, που βασίζεται στην απαλλαγή της όποιας δομής εισάγει φασματικές γραμμές, καθώς επίσης και στην αξιοποίηση των ορίων ακουστότητας, σύμφωνα με τις καμπύλες Fletcher-Munson. Με αυτόν τον τρόπο επιτυγχάνεται η κατασκευή σήματος με τη χαμηλότερη δυνατή ακουστότητα και δεδομένη ενέργεια. Εφόσον προηγείται μια γενική προσέγγιση της φασματικής πυκνότητας ισχύος της μουσικής λαμβάνοντας υπόψιν αντιπροσωπευτικά μουσικά δείγματα, καταλήξαμε στην κατασκευή του πιλότου με δυο πιθανούς τρόπους. Στον πρώτο, γίνεται χρήση σήματος ευρέως φάσματος, στα πλαίσια που ορίζει το εύρος ζώνης μιας κάρτας ήχου, ώστε να επιτύχουμε την μέγιστη δυνατή ακρίβεια στην εκτίμησή μας. Για την διεύρυνση του φάσματος γίνεται χρήση τεχνικών μεταπήδησης χρόνου (time-hopping) και άμεσης ακολουθίας (direct sequence). Στην δεύτερη προσέγγιση, διαμορφώνουμε το σήμα μας σε μια ζώνη συχνοτήτων που η αντίληψη του ανθρώπου, όσον αφορά τους ήχους, είναι σημαντικά υποβαθμισμένη.

Σε κάθε σενάριο κατασκευής του σήματος, υποθέτουμε ότι ο δέκτης έχει διαθέσιμη μια χρονική αναφορά της στιγμής που προστέθηκε ο πιλότος στο μουσικό σήμα, γνωρίζει το σήμα που αποστέλλεται και εκτελεί μια ετεροσυσχέτιση (cross-correlation) μεταξύ αυτού και της ενθόρυβης καθυστερημένης εκδοχής που λαμβάνει. Για την κατασκευή του σήματος αποστολής γίνεται μελέτη σχετικά με την χρήση κατάλληλων παλμών με την αντιστοίχως απαραίτητη ενέργεια ώστε να μην είναι ακουστοί. Περιλαμβάνεται θεωρητική ανάλυση και προσομοίωση της απόδοσης του συστήματος θεωρώντας Gaussian θόρυβο, όπως επίσης και έλεγχος της απόδοσης υπο την παρουσία αντιπροσωπευτικών δειγμάτων μουσικής. Η θεωρητική ανάλυση με την παρουσία μόνο μουσικής δεν είναι εφικτή, καθώς δεν υπάρχει κάποιο ισχυρό μαθηματικό μοντέλο για την περιγραφή της. Γίνεται επίσης σύγκριση των επιλεγμένων συστημάτων με παραπλήσια “δημοφιλή” σχήματα, όπου και αναδεικνύεται η υπεροχή τους. Τέλος, συγκρίνουμε τις τεχνικές αυτές ως προς την επίδοση στις εκτιμήσεις που πετυχαίνουν με ζητούμενη ακρίβεια ενός δείγματος, κλασματικής ακρίβειας του ενός δείγματος, και την ευρωστία που παρουσιάζουν εφαρμόζοντας κωδικοποίηση και αποκωδικοποίηση.

Contents

List of Figures

Preface

At this point I would like to thank my supervisor, Prof. Alexandros Eleftheriadis, for his guidance, support, and patience. I have been really lucky to have a supervisor who encouraged and advised me so much throughout all this time. During the last year of our pleasant cooperation, beyond the satisfaction of dealing with very interesting and modern fields of study, I experienced how high-level research should be and I am sure that these experiences will be proved critical in my near future. Furthermore, a special thanks to my family that supports me in each academic step in any possible way. Last but not least, I would like to thank all the new friends I made and with whom we walked through the new, for me, academic environment of the Informatics and Telecommunications Department of the National and Kapodistrian University of Athens.

1. The Problem of the Real-Time Estimation in Audio Systems and its Background

The problem we want to address in this Master thesis is the estimation of the delay that is inserted to audio signals, during their processing under real-time systems, or their transmission through net, using pilot waveforms. The pilot signal is injected and transmitted simultaneously with music, and if it is not designed appropriately, it will be probably hearable which is obviously unwanted. To have a closer view on parameters that affect the pilot transmission and design, a generalized Power Spectral Density (PSD) which is observed over various music samples and the Fletcher-Munson's curve for the Absolute Threshold of Hearing (ATH), are considered. Music is treated as noise in our case and the sophisticated part of this study is to construct a simple in generation, inaudible, and simultaneously powerful pilot trace that can be distinguished from the music signal. High performance for the delay estimation, is also required. Hence, we estimate the probability of correct signal's acquisition in the order of one sample, after processing the noisy and shifted signal, assuming that the pilot waveform is known at the receiver. Finally, we examine the pilot signal's behavior under compression, and its performance considering fractional delay estimation.

1.1 Introduction

In many cases of network communication, it is desirable to measure the end-to-end delay between two remote computer systems. In the case of real-time audiovisual communication, it is desirable to be able to measure the delay through the entire system, including digital-to-analog conversion and playback at the receiver. Such measurement can obtain the real delay experienced by an end-user. Although performance measurement of traditional videoconferencing systems can benefit from such a tool, there are applications where high-resolution delay measurement is critical. One example is Network Music Performance (NMP) systems, where the end-to-end delay tolerance is reported to be in the order of 20 msec. With such stringent requirements, it is important to be able to measure with high accuracy the delay across the entire path from capture, to encoding and transmission, decoding, as well as playback. In addition to performance analysis, the exact value of the delay that each participant is subject to can be used by the system in order to improve his/her experience. One example is tempo adaptation [2]. In various experiments that have been conducted, a general inverse relationship between tempo and delay was reported. For instance, slowing down the tempo of a particular musician who experiences a bigger delay than the rest of his group, can be beneficial and can enhance an ensemble's ability to play synchronously and with comfort.

We describe a novel system [1] that can be used in applications where audio data is exchanged between two or more end points, such as in an NMP. The proposed system is a high-resolution delay estimator that involves inaudible pilot signals. The objective is to measure the end-to-end delay in transmitting audio between two systems. The structure should provide "orthogonality" in that multiple independent measurements between a number of users across a network could be performed. A pilot signal is added to the audio during capture (e.g., at the computer used as a capture and transmitting device). At the receiver, the received "complex" signal is processed in order to obtain a high-resolution (at least in the order of one sample) estimate of the position of the pilot signal.

Assuming that a second time reference is provided, this can be used to obtain a highly accurate delay estimate. For example, in a laboratory setting, the audio signal together

with the pilot may be captured without being transmitted, so that it can be used as a reference. For example, monophonic outputs of the audio signal plus the pilot from the transmitting device and the receiving device can be recorded as analog signals on a stereo device. The two channels can then be processed, independently, to measure the exact delay. The proposed technique is robust to a number of signal processing operations.

This system can also be used in order to observe dynamically varying time-related parameters such as network jitter. This can be accomplished using periodic sequential transmission of the pilot signal. A benefit of the proposed technique is that it can measure computer hardware and operating system latency (play-out buffer management delay, etc.). The proposed system is designed as a performance analysis tool in an NMP system under development ("MusiNet").

1.2 Related Work

The system under consideration is strongly related to studies on the field of audio watermarking. Audio watermarking techniques are classified on time, frequency, wavelet and code domain method. The first techniques were directly inspired from previous research on image watermarking (Bender et al. [4]). The main objective consists in adding a signal, the watermark, to the original audio signal. The resulting watermarked signal must be perceived by the listener as identical to the original one. The watermark carries data that can be retrieved by a detector. The idea of spreading the watermark in frequency in order to maximize its power while keeping it inaudible and increasing its resistance to attacks, was introduced by Boney in [5] and Garcia in [6]. Boney's scope is to prevent piracy and his analysis assumes that the original audio file is known at the detection process. Garcia makes an extensive analysis on the same field, proposing a new algorithm that follows the same principles to our work. However, his algorithm for embedding and recovering of the watermark is more complex, as his purpose is to recover the encrypted data, even after numerous off-line processing operations.

The majority of the studies in this area, are focused on hiding and recovering data, mainly for piracy prevention purposes. In this work, the case of interest is the delay estimation and, hence, a simpler model is assumed. Embedding an inaudible pilot signal is similar to audio watermarking [8, 9]; the objective of identifying the exact position of the watermark, rather than its content, is, however, different.

1.3 Outline

In brief, we present the structure of this study. Audio signals, are expected to be present during pilot's transmission. In Chapter 2, the general spectral properties of the music are investigated based on some representative tracks, in order to design a pilot signal that could exploit the characteristics of this kind of noise. In addition, the ATH is presented to provide some general information about the human auditory system, which varies as a function of frequency. Based on these observations, we define two general techniques for the signal design.

In Chapter 3, widely used pulses are presented and evaluated in order to select the appropriate one that best fits, according to the system requirements. In our case, the pilot signal must be inaudible and should be capable of achieving high-performance delay estimation. For each candidate pulse, the shape as well as its auto-correlation function are given in both time and frequency domains. The width and the appropriate energy

of each used pulse, are defined, with respect to the system restrictions. In addition, the modulation techniques that are of interested are analyzed, and the PSD of the transmitting signal is calculate. Note that we ended up on two possible implementations for the pilot design and, hence, both of them will be investigated and evaluated through this study.

In Chapter 4, the system model is provided. The exact transmitted signal, the expected pilot at the receiver, the selected approach for the channel estimation, and the spot where the Signal to Noise Ratio (SNR) should be set, are defined. Furthermore, the products of signal and noise after filtering are analyzed, and the gain of filtering is calculated.

In Chapter 5, the theoretical performance of our system is analyzed, assuming that AWGN is dominating. Note that, different approaches are followed for the pilot designs. In fact, there is no consistent stochastic model for the exact music effects and, in general, we can't predict the system performance under its presence. Thus, we are confined in practical implementations for the latter case.

The simulations under AWGN, Pink noise which can be a useful model for rich musical sounds, and some representative music samples, are given in Chapter 6, with respect to each signal's construction. Furthermore, the validity of our theoretical analysis is evaluated, and a comparison between the selected techniques for the pilot design is performed.

Consequently, in Chapter 7, we propose a technique to improve system performance, in the sense of achieving a more accurate delay estimation. Specifically, upsampling at the receiver's side is performed and fractional delay estimation can be obtained. In addition, the robustness of the pilot under the OPUS codec is examined. According to [13], an Internet audio codec for applications of "Live Distributed Music Performances" should comply to specific limitations. Encoding and decoding with respect to these restrictions are performed, and preliminary results are presented. We explain, also, how to modify the proposed pilot signals, in order to comply to systems which adopt higher audio sampling frequencies. Finally, some noteworthy applications are presented, where this system could be used.

1.4 Contributions

In this study the newly introduced idea lies to the fact of injecting an inaudible pilot signal to an audio signal transmission, in order to estimate the delay of a channel (i.e., a real-time audio system, cables, or net). The sophisticated part of this study is the signal design, which exploits human perception on sounds and the general spectral characteristics of audio signals. It could be considered as an Audio Watermarking technique, however the goal of identifying the exact position of the watermark rather than its content, is essentially different. In addition, a simple signal generation and a reasonable performance under encoding/decoding process were considered more critical, in relevance to a potential complex signal design that could survive through multiple cases of processing without altering its content, or, a signal design that demands configuring of codecs. In the latter cases, it is denoted that enhanced complexity would be added to the system, which is undesirable. Thus, the signal design and generation, as well as the system setup are quite simple, which makes attractive and beneficial the use of such a system in relevant applications. Robustness and high-performance delay estimation, are also achieved. Indeed, under certain circumstances, delay estimation in the order of 3.78 μsec can be succeeded.

2. Music as Noise

Music, potentially, co-exists in the channel during the transmission of the pilot and, hence, it is treated as interferer - noise. In order to understand the effects of this kind of "noise", initially, a general modeling of the PSD is required to have a clear view on its spectral characteristics. It is obvious that for different kind of music, different PSDs occur. That's why there is no concise mathematical model that describes globally the effects of music, which could be used for theoretical analysis. However, certain properties are present in the majority of music tracks (if not to all) which are highlighted.

Representative music samples are considered and their PSDs are presented. The PSD is calculated as the square of the Fourier Transform's absolute value, divided by the corresponding time duration of each sample. In addition, the ATH based on Fletcher-Munson curves, is given. The goal is to take advantage of these properties and to devise general strategies for the structure of the pilot signal in order to be kept acoustically untraceable, and to be capable of maximizing the accuracy of the delay estimation. Certain considerations regarding the shape of the spectrum that our signal has to adopt, are explained. Widely used pulses along with their properties are examined, in order to select the optimal for our case. Finally, technical restrictions are considered for the construction of the pilot, such as the limits of the sound card's bandwidth and the global specifications of real time systems that process audio signals (sampling frequency, number of bits per sample).

2.1 Representative music samples

All considering samples were chosen from a free library [11], which has been developed by a professional musician and teacher with great experience in the field. These are music tracks from typical instruments that are used in orchestras and, thus, they are considered to be representative for our demands.Obviously, the estimated PSDs are related to the corresponding music tracks (i.e., different music track for each instrument will result to a different PSD). The title of each figure indicates the corresponding instrument. The PSD was derived, by converting each stereo track to a monophonic, and then by calculating the square of the absolute value of the Fourier Transform of each mono signal which is divided by the respective duration of the sample.

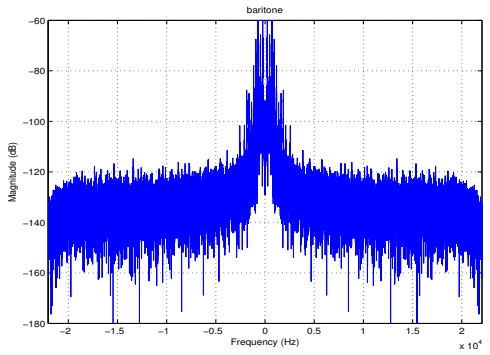

Figure 2.1.1: PSD of sample "baritone.wav".

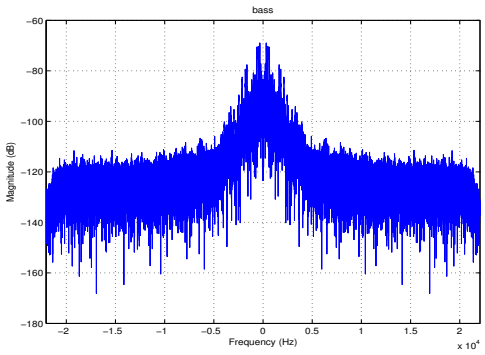

Figure 2.1.2: PSD of sample "bass.wav".

Considering the related figures, a general tendency of exhibiting lower power at higher frequencies was expected and observed in all tested music samples. Hence, we can assume that the PSD of music in general, could be modeled by the approximate waveform presented in Fig. 2.1.13.

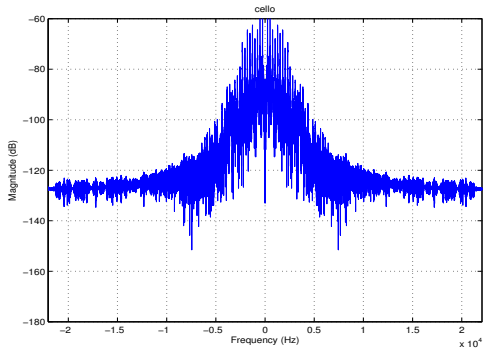

Figure 2.1.3: PSD of sample “cello.wav”.

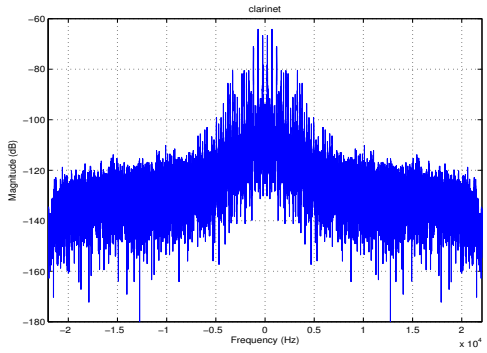

Figure 2.1.4: PSD of sample “clarinet.wav”.

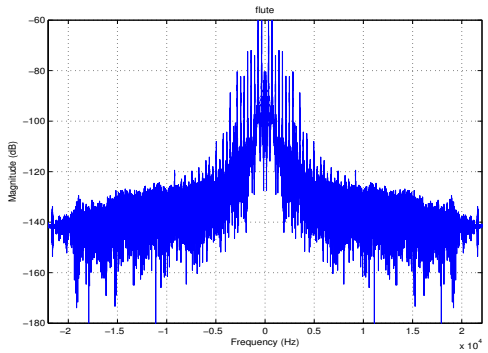

Figure 2.1.5: PSD of sample “flute.wav”.

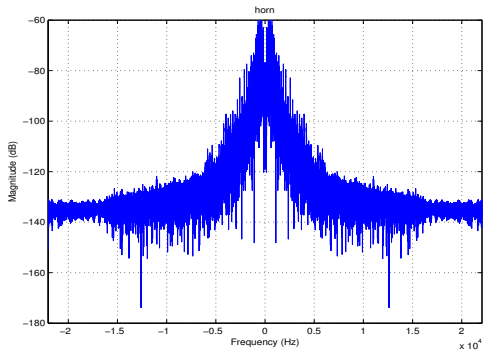

Figure 2.1.6: PSD of sample “horn.wav”.

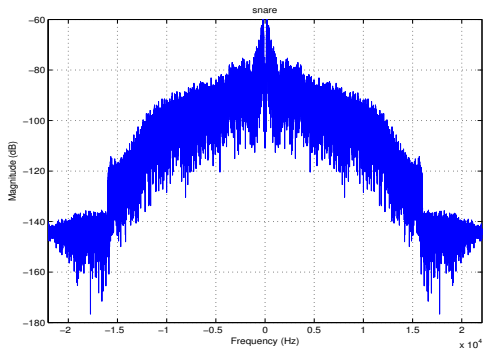

Figure 2.1.7: PSD of sample “snare.wav”.

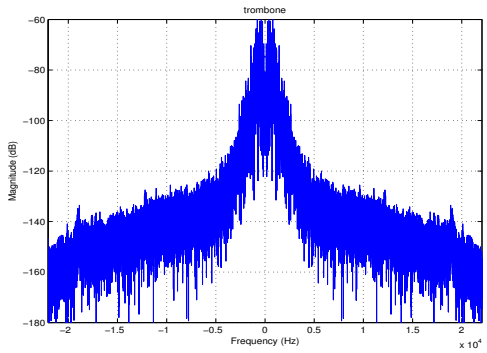

Figure 2.1.8: PSD of sample “trombone.wav”.

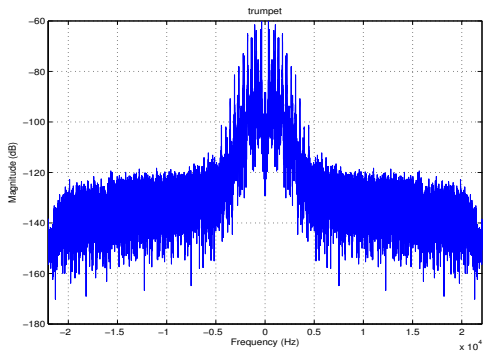

Figure 2.1.9: PSD of sample “trumpet.wav”.

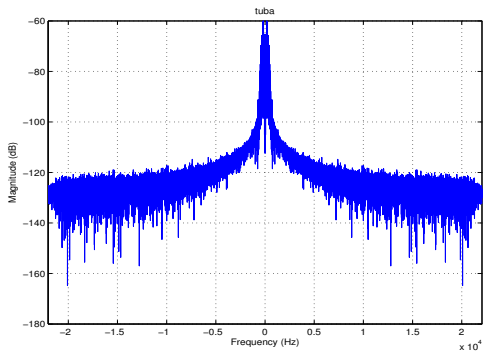

Figure 2.1.10: PSD of sample “tuba.wav”.

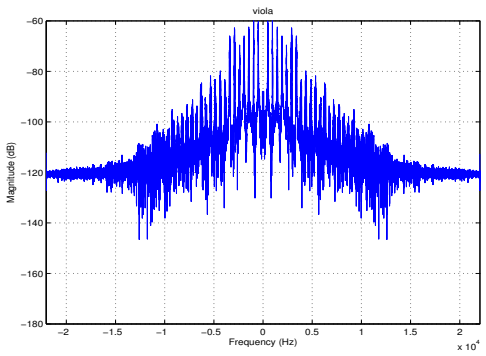

Figure 2.1.11: PSD of sample “viola.wav”.

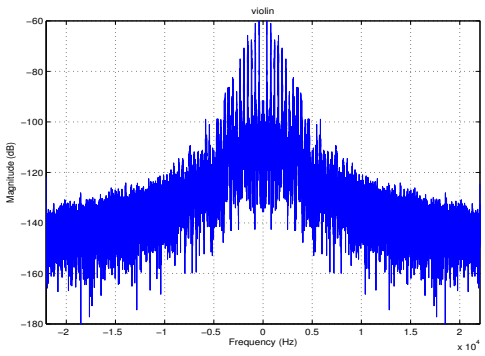

Figure 2.1.12: PSD of sample “violin.wav”.

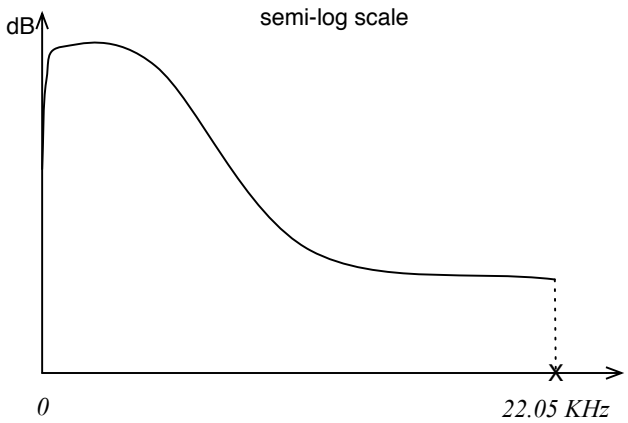

Figure 2.1.13: PSD of music in general.

2.2 Absolute Threshold of Hearing

The ATH based on Fletcher-Munson's curve is considered, in order to have a closer view of human perception on sounds. In Fig. 2.2.1, the curve that is a function of frequency, describes the absolute threshold of hearing.

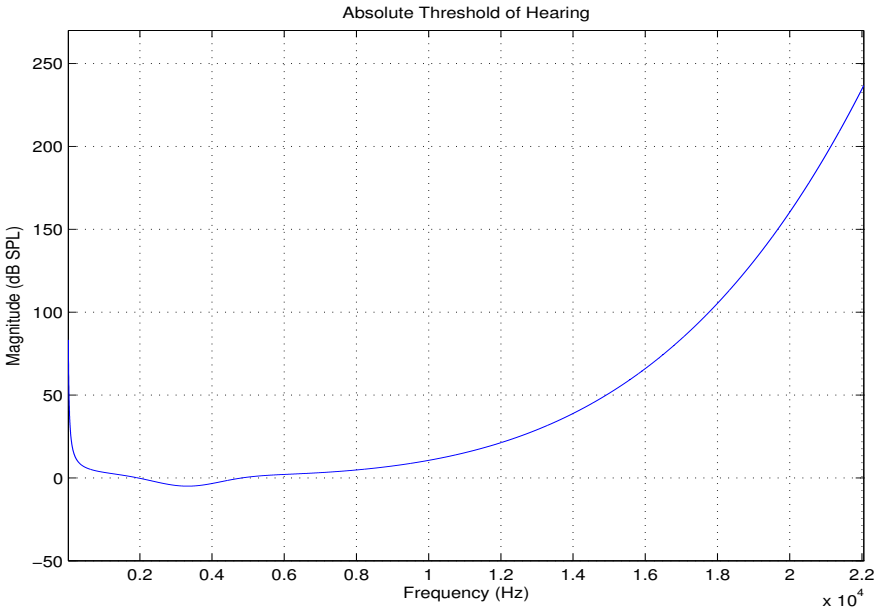

Figure 2.2.1: Fletcher-Munson curve.

The term SPL is a logarithmic measure of the effective sound pressure of a sound relative to a reference value. It is measured in decibels above a standard reference level. The standard reference sound pressure in air is $20\ \mu\mathrm{Pa}$, which is considered the threshold of human hearing (at 1 kHz), as depicted in Fig. 2.2.1, too.

$$\mathrm{SPL} = 20\ \log_{10} \frac{P_{\mathrm{rms}}}{P_{\mathrm{ref}}}\ \mathrm{dB} \tag{2.1}$$

where $P_{\mathrm{ref}} = 20\ \mu\mathrm{Pa}$ and P_{rms} is the RMS sound pressure being measured. The related figures that present the PSDs of music samples, are measured in dB and we can't compare them directly with the ATH curve. It is reasonable to assume that the magnitude of the PSD of an acoustic signal, is the sound power at the corresponding frequency and that metric is desirable for comparison. Sound power level (SWL) of a source is expressed in decibels relative to a reference sound power. In air this is normally taken to be equal to 10^{-12} watt, that is 0 dB SWL.

$$\mathrm{SWL} = 10\ \log_{10} \frac{P_1}{P_0}\ \mathrm{dB} \tag{2.2}$$

where $P_0,\ P_1$ are the sound powers and P_0 is the reference one. Unlike sound pressure, sound power is neither room dependent nor distance dependent. Sound power belongs strictly to the sound source. Sound pressure is a measurement at a point in space near the source, while sound power is the total power produced by the source in all directions.

It is known, that sound power level is equal to sound pressure level and intensity level at a distance of $r = 0.2821$ meter from the source at full sphere propagation, when $Q = 1$. Furthermore, 1 watt of sound power equals to 120 dB. Those two metrics are related by

Eq. 2.3, where it is used that sound intensity level and the sound pressure level at a distance of 1 meter from the source in a free field is 11 dB less than the sound power level of the source.

$$\mathrm{SPL} = \mathrm{SWL} \, - \, 20 \, \log_{10} r \, - \, 11 \tag{2.3}$$

Finally, the value of sound pressure level SPL at 0.2821 m is

$$\mathrm{SPL} = \mathrm{SWL} \, - \, 20 \, \log_{10} r \, - \, 11 \, = \, 120 \, - \, 20 \, \log_{10} 0.2821 \, - \, 11 \, = \, 120 \, \mathrm{dB}.$$

Now, we can calculate the SWL values of music samples at the corresponding frequencies, with respect to the PSDs figures. Assuming a distance equal to 0.2821 m from the source, those values would be equal to the SPL ones. For example, in Fig. 2.1.6, at 1 kHz the power is -80 dB. That is translated to 10^{-8} watts. Considering Eq. 2.2, SWL = $10 \, \log_{10} \frac{10^{-8}}{10^{-12}} \, = \, 10 \, \log_{10} 10^4 \, = \, 40$ dB. Using Eq. 2.3, the SPL value at a distance equal to 0.2821 m is 40 dB, which is above the ATH at the corresponding frequency.

2.3 Taking advantage of the music spectral properties

We try, now, to devise some general strategies for the structure of the pilot signal, assuming music as the interferer. The ATH is considered, in order to take advantage of the general principles that human perception follows, within the limits of the audio interface bandwidth. The main requirements of our system are: (a) to be inaudible and (b) to achieve high performance delay estimation. According to Fig. 2.2.1, it is obvious that human hearing is insensitive at very low (< 200 Hz) and high frequencies (> 15 kHz).

2.3.1 Case I: spectrum in high frequencies

In order to minimize the audible effects of the transmitted signal, we limit the pilot signal's spectrum at a high-frequency band. This can be achieved from the choice of the used pulse, by including a high-pass filter in the system architecture, or by modulation of band-limited pulses. In this work, the first and the latter choices are followed, since the second one is approached by the first. Adopting this design, we expect that the pilot signal with a given power will be less audible, in relevance to common signals with the same power and a spectrum located at lower frequency bands. In addition, according to Fig. 2.1.13 the proposed pilot signal is expected to be interfered by decreased values of the music power, resulting in better performance and enhanced robustness.

The high resolution requirement translates to the need for short pulses which, in turn, implies the use of pulses that have as wide bandwidth as possible (within the limits of the underlying digital representation). Here, we must do a distinction between the choices that are considered for the signal construction. Regarding the case of using pulses that place the signal spectrum in high-frequencies, they should be as small as possible, in order to achieve the maximum possible bandwidth. In the case of using a carrier, the signal spectrum will be transferred to high-frequencies and, thus, our band-limited pulse has to be selected appropriately, in order not to exceed the available bandwidth. Obviously, the limits that have to be fulfilled by pulse's spectrum, are related to the center frequency that is selected for the carrier wave. Finally, for the latter case, the smallest possible pulses with the minimum possible bandwidth, are considered, in order to meet the system requirements.

2.3.2 Case II: spectrum in low frequencies

Another implementation of the pilot signal would be the transmission of broad pulses with the main energy of signal being in low frequencies (< 200 Hz if possible - the lower the better). The pulse has to be broad enough in order to minimize the spectrum of the pilot signal. The use of a low-pass filter is also a solution to delimit the bandwidth in the desirable band. The bigger the duration of the pulse, the smaller the bandwidth becomes. However, the operability of such a system is by definition limited. Particularly in IP networks where our system is expected to be used, the fact that a single packet could possibly contain a part of one single pulse, makes our pilot sensitive to network jitter and packet-loss. In addition, the accuracy of this system is expected to be degraded, as the time duration of each used pulse increases. Hence, this implementation is rejected and it is presented here, only as a possible solution to limited cases.

The general PSD of music plus the pilot for each case, are presented in Fig. 2.3.1 and 2.3.2, correspondingly.

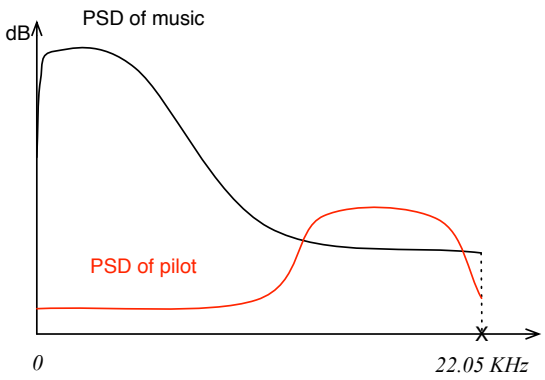

Figure 2.3.1: Case I.

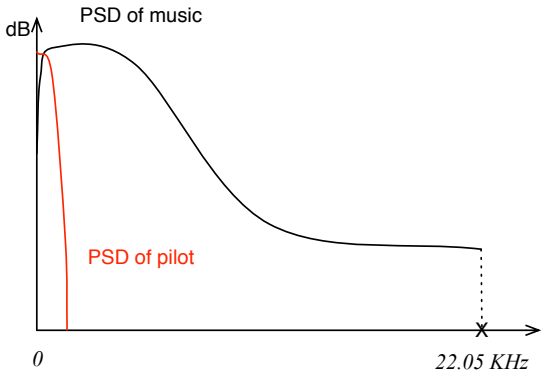

Figure 2.3.2: Case II.

3. Pilot Signal Design

The design of the pilot signal must be carefully considered, in order to ensure that it is inaudible and simultaneously it is capable of achieving high-resolution delay estimation. Furthermore, assuming that the pilot signal is added in the digital domain, its bandwidth is limited to that provided by the underlying digital representation. In this study, we assume a sampling frequency of 44.1 kHz and, thus, the pilot signal is limited to 22.05 kHz. We have defined two potential schemes for the construction of the pilot signal. We are going to describe both of them, along with their properties, in the time and frequency domain. The exact selection of the pulse, the modulation technique and the resulting PSD are going to be mathematically defined, for both schemes.

3.1 Pulse design

In order to meet the requirements defined in Section 2.3, for each case of the signal's construction, a different pulse could be used, in terms of shape, time duration and amplitude. In the case of adopting the pulse train as the pilot, the pulse spectrum must be placed at high frequencies and it should be as wide as possible, indicating the sharp shape of the pulse. In the case of using the carrier-wave as the pilot, the involved pulse should be band-limited with the minimum spectrum, given that it is narrow, too.

3.1.1 Evaluation of candidate pulses

A general review of widely used pulses is presented here. The scope is to demonstrate the properties of each candidate pulse in order to highlight the one that would best fit to our demands. For each pulse, a mathematical formula, its auto-correlation function and the corresponding Fourier Transform are given, to have a clear view on their properties and their adequacy. The auto-correlation function is required, because the delay estimation is performed under cross-correlation at the receiver's side, as it will be seen in the following chapters. In order to highlight the effectiveness of our techniques in dealing music, the pulses that do not meet the requirements that we set, will be used in the same system setup as baseline models, for comparison with the proposed one.

3.1.1.1 Rectangular pulse

Rectangular is chosen as a candidate transmitting pulse due to the simplicity of its generation. A general mathematical formula for this pulse, is given in Eq. 3.1

$$p(t) = \begin{cases} A, & \text{if} |t| < \frac{T_p}{2} \\ \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

where A denotes the amplitude and T_p the time duration of the pulse. We express Eq. 3.1 in a more convenient form, given below:

$$\Pi(wt) = \begin{cases} 1, & \text{if} |t| < \frac{1}{2w} \\ \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

where it is indicated that the amplitude of this pulse is 1 and its time duration equals $1/w$, with $w > 0$. To set a specific amplitude, a multiplication with the corresponding value is required. Comparing Eq, 3.1 and 3.2, $T_p = 1/w$ and the amplitude of $p(t)$ is A times bigger than $\Pi(w\,t)$.

It is well known that the Fourier Transform of a rectangular pulse is a sinc function, or a raised cosine pulse with roll-off factor $a = 0$. The Fourier Transform, which is denoted as $\mathcal{F}\{\}$, is calculated as follows.

$$\begin{aligned} \mathcal{F}\{p(t)\} &= \int_{-\infty}^{+\infty} p(t)\, \mathrm{e}^{-j2\pi ft}\, \mathrm{d}t \\ &= \int_{-\frac{T_p}{2}}^{\frac{T_p}{2}} \mathrm{e}^{-j2\pi ft}\, \mathrm{d}t \\ &= \frac{\mathrm{e}^{-j\pi fT_p} - \mathrm{e}^{j\pi fT_p}}{-2j\pi f} \\ &= \frac{\sin(\pi fT_p)}{\pi f} \\ &= T_p \operatorname{sinc}(\pi fT_p) \end{aligned}$$

The sinc function has the property to be decreasing, as the absolute value of the product $f\,T_p$ is increasing, based on the previous formula. In addition, its value for every multiple integer of T_p is zero. Theoretically its bandwidth is infinite, incurred by the indefinitely-sharp edges in the time domain definition, and expressed by the sinc function which constitutes its frequency domain representation.

In general cases, the form of the rectangular pulse of Eq. 3.2 is more useful, since a straight calculation of its Fourier Transform is obtained, by the following valid transformations.

$$\begin{aligned} \Pi(wt) &\overset{F.T.}{\longleftrightarrow} \frac{1}{w} \operatorname{sinc}\left(\frac{f}{w}\right) \\ \operatorname{sinc}(wt) &\overset{F.T.}{\longleftrightarrow} \frac{1}{w}\, \Pi\left(\frac{f}{w}\right) \end{aligned}$$

In Fig. 3.1.1, a typical implementation of the rectangular pulse along with its Fourier Transform are presented, by setting energy equal to 1 and time duration $T_p = 1$ sec.

In Eq. 3.3, the auto-correlation function of the rectangular pulse is given. To simplify calculations with integrals, we define the pulse duration in the range $[0,\, T_p]$.

$$R_{pp}(t+\tau, t) = \int_{-\infty}^{+\infty} p(t+\tau)\, p(t) dt = \begin{cases} \int_{-T_p}^{\tau} A^2 dt = A^2\,(T_p + \tau), & -T_p < \tau \leq 0 \\ \\ \int_{\tau}^{T_p} A^2 dt = A^2\,(T_p - \tau), & 0 < \tau \leq T_p \\ \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

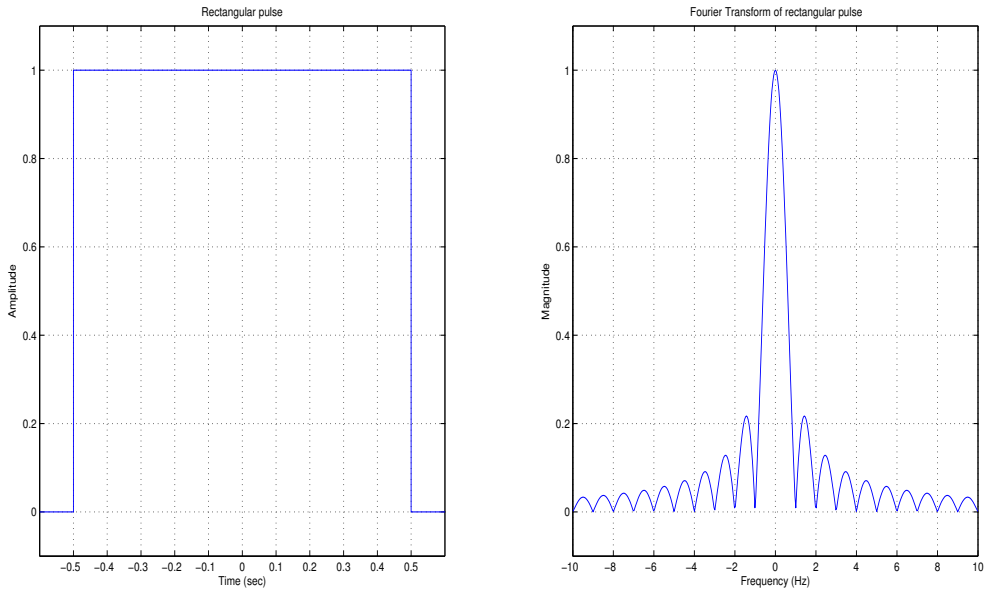

Figure 3.1.1: Rectangular pulse in time and frequency domain.

It is obvious that Eq. 3.3 depicts a triangle which is spread over the range $[-T_p,\, T_p]$ and its peak value, A^2, is obtained at the time instant $\tau = 0$.

In general, cross-correlation operation between the functions $f(t)$ and $g(t)$ is equivalent to the convolution of $f'(-t)$ and $g(t)$, where $x'(t)$ denotes the complex conjugate of $x(t)$. According to this observation, we come of to Eq. 3.4

$$f(t) \star g(t) = f'(-t) * g(t) \tag{3.4}$$

where "$\star$" denotes the cross-correlation and "$*$" denotes the convolution operation, respectively. Analogous to the convolution theorem, cross-correlation satisfies the following relation.

$$\mathcal{F}\{f(t) \star g(t)\} = \mathcal{F}\{f(t)\}' \, \mathcal{F}\{g(t)\} \tag{3.5}$$

Therefore the Fourier Transform of the rectangular's auto-correlation function is given in Eq. 3.6.

$$\mathcal{F}\{R_{pp}(t+\tau, t)\} = \mathcal{F}\{p(t) \star p(t)\} = \mathcal{F}\{p(t)\}' \, \mathcal{F}\{p(t)\} = P'(f)\, P(f) = |P(f)|^2 \tag{3.6}$$

In Fig. 3.1.2, the auto-correlation function of the rectangular pulse and its Fourier Transform are shown.

3.1.1.2 Square-root raised cosine pulse

Square-root raised cosine (SRRC) pulse is a possible transmitting pulse. It is frequently used as the transmission and reception filter in a digital communication system to perform

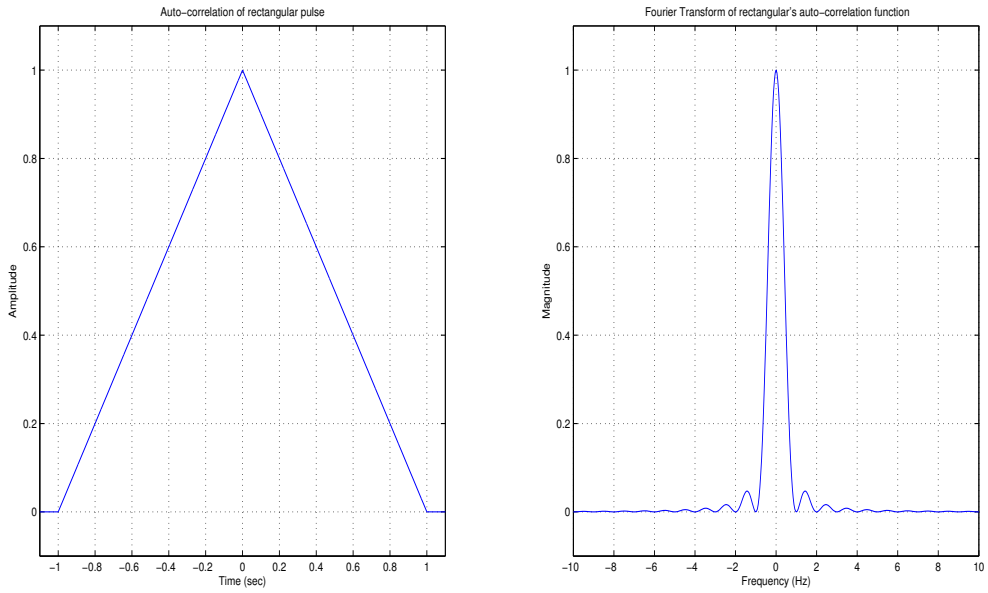

Figure 3.1.2: Auto-correlation of rectangular pulse in time and frequency domain.

matched filtering. This helps in minimizing inter-symbol interference (ISI). In our case though, a simpler channel which just shifts the input waveform, is considered. However, the extensive references and use in digital transmissions are appreciated.

The general formula of this pulse is given in Eq. 3.7

$$p(t) = \frac{4\,a}{\pi\sqrt{T}}\,\frac{\cos\left([1+a]\frac{\pi t}{T}\right) + T\,\frac{\sin\left([1-a]\frac{\pi t}{T}\right)}{4at}}{1-\left(\frac{4at}{T}\right)^2} \tag{3.7}$$

where T approximates the time duration of the half of the main lobe, and a is a parameter for pulse shaping (roll-off factor, which is a measure of the excess bandwidth). Both variables affect the bandwidth of the pulse which equals $(1+a)/(2T)$. Note that, the bandwidth definition expresses the spread of spectrum *only* in positive frequencies. The Fourier Transform of the SRRC is given in Eq. 3.8, according to [7].

$$\mathcal{F}\{p(t)\} = \begin{cases} \sqrt{T}, & 0 \le |f| < \frac{1-a}{2T} \\ \sqrt{\frac{T}{2}\left[1+\cos\left(\frac{\pi T}{a}\left[|f|-\frac{1-a}{2T}\right]\right)\right]}, & \frac{1-a}{2T} \le |f| < \frac{1+a}{2T} \\ 0, & |f| > \frac{1+a}{2T} \end{cases} \tag{3.8}$$

In Fig. 3.1.3, SRRC pulses are presented in time and frequency domain, assuming that energy equals 1, $T = 1$ and roll-off factor $a = \{0, 0.5, 1\}$. The definition of the SRRC pulse, involves a parameter, A, which is responsible for the duration of its two-sided tail. Each tail spans over $A\,T$ sec to each side. Typical values for A are 3, 6 and 12. In our case, $A = 6$. The time axis of the relative figure, is in the appropriate range $[-A,\ A]$ sec.

According to Eq. 3.4, the auto-correlation function of a SRRC pulse is straight equivalent to the convolution of two SRRC pulses, because they constitute even and real pulses. Furthermore, the convolution of the SRRC with itself is, by definition, equivalent to a raised cosine pulse. The general formula of the auto-correlation function of the SRRC pulse, or

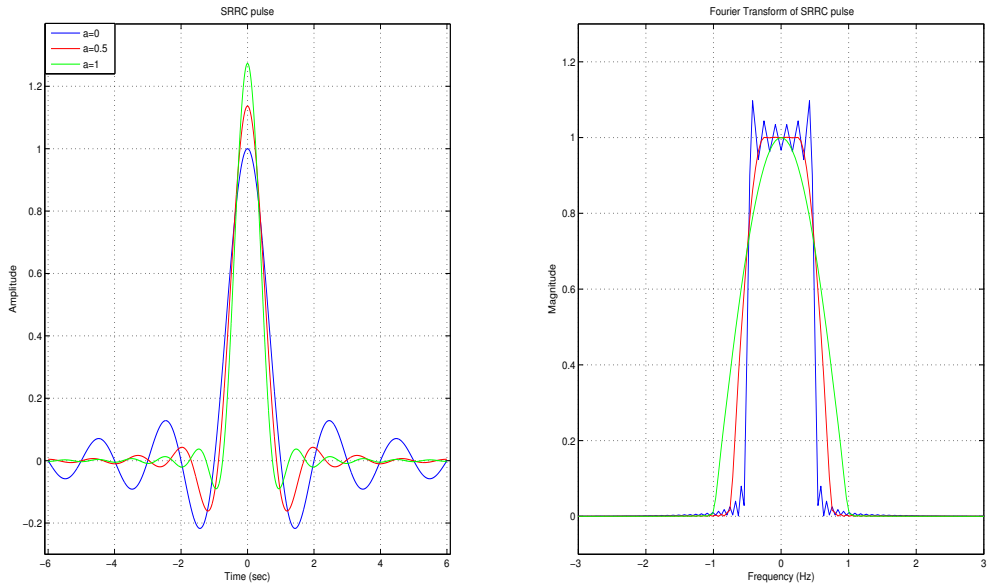

Figure 3.1.3: Srrc pulse in time and frequency domain.

in other terms, the raised cosine function is given by Eq. 3.9, according to [12]. The bandwidth of the latter is equal to $(1+a)/(2T)$, too.

$$R_{pp}(t+\tau,t) = \frac{\sin\left(\frac{\pi\tau}{T}\right)}{\frac{\pi\tau}{T}} \frac{\cos\left(\frac{a\pi\tau}{T}\right)}{1-\left(\frac{2a\tau}{T}\right)^2} \tag{3.9}$$

In addition, according to [12], the corresponding Fourier Transform of the raised cosine pulse is given as follows.

$$\mathcal{F}\{R_{pp}(t+\tau,t)\} = \begin{cases} T, & 0 \le |f| < \frac{1-a}{2T} \\ \frac{T}{2}\left[1+\cos\left(\frac{\pi T}{a}\left[|f|-\frac{1-a}{2T}\right]\right)\right], & \frac{1-a}{2T} \le |f| < \frac{1+a}{2T} \\ 0, & |f| > \frac{1+a}{2T} \end{cases} \tag{3.10}$$

The auto-correlation function of the SRRC pulse, in time and frequency domain, assuming $T=1$ and $a=0,\,0.5,\,1$, are given in Fig. 3.1.4.

3.1.1.3 Gaussian pulse & its derivatives

The Gaussian pulse, also, constitutes a candidate choice for transmission in our setup. The simple generation and the simplifications arising in mathematical calculations are convinient. A general mathematical formula for the Gaussian pulse is:

$$p(t) = A\,\mathrm{e}^{-\frac{t^2}{2\sigma^2}} \tag{3.11}$$

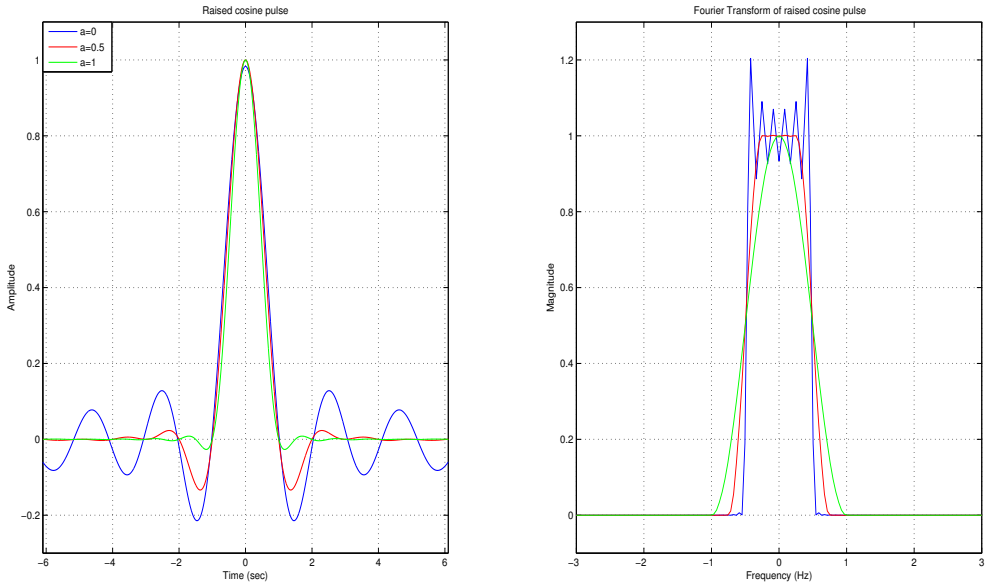

Figure 3.1.4: Auto-correlation of srrc pulse in time and frequency domain.

where A is the amplitude and σ^2 denotes the variance. Through observation, it was found that in order to define Gaussian pulse's time duration equal to T_p seconds, the value of its standard deviation has to be set as

$$\sigma = \frac{T_p}{4\sqrt{\pi}}. \tag{3.12}$$

In this way, the pulse's time duration T_p becomes a function of its standard deviation, and the Gaussian function's equation, takes the following form

$$p(t) = A\,\mathrm{e}^{-\frac{8\pi}{T_p^2}t^2} \tag{3.13}$$

It is well known[1] that

$$\mathcal{F}\{\mathrm{e}^{at^2}\} = \sqrt{\frac{\pi}{a}}\; e^{-\pi^2 f^2/a}.$$

Respectively, the Fourier Transform of the Gaussian pulse defined by Eq. 3.11, is derived as follows.

$$\mathcal{F}\{p(t)\} = A\sqrt{2\pi\sigma^2}\,\mathrm{e}^{-2\pi^2\sigma^2 f^2} \tag{3.14}$$

In the last expression, by setting standard deviation equal to $\sigma = \frac{B}{4\sqrt{\pi}}$ in order to handle the duration of the pulse in frequency domain, which by definition is the "total" bandwidth (twice of the bandwidth which is defined in positive frequencies, as it is assumed for the SRRC pulse earlier), we get

$$\mathcal{F}\{p(t)\} = A\sqrt{2\pi\left(\frac{B}{4\sqrt{\pi}}\right)^2}\,\mathrm{e}^{-2\pi^2\left(\frac{B}{4\sqrt{\pi}}\right)^2 f^2} = \frac{B}{4}A\sqrt{2}\,\mathrm{e}^{-\left(\frac{B^2\pi}{8}\right)f^2}.$$

[1] http://mathworld.wolfram.com/FourierTransformGaussian.html

Now, if we combine the exponentials included in the last formula and in Eq. 3.13, we get approximately, the relationship between the time duration T_p, and the bandwidth, B, of a Gaussian pulse which is defined according to Eq. 3.11. Then, we get

$$B \simeq \frac{8}{T_p}. \tag{3.15}$$

Alternatively, another widely used approximation to measure Gaussian pulse spread is to estimate the full width at the tenth of maximum, W_{10}[2], which is given as

$$W_{10} = 2\sqrt{2\ln 10}\sigma \approx 4.29193\,\sigma \tag{3.16}$$

assuming that Gaussian pulse is defined according to Eq. 3.11. This approximation could be used in both time and frequency domain. In the time domain, it can be used straight forward and, hence, $W_{10} = 4.29193\,\sigma$ sec. In the frequency domain, we have to express the product of Fourier Transform in the appropriate form, using Eq. 3.14. Specifically,

$$\mathcal{F}\{p(t)\} = A\sqrt{2\pi\sigma^2}\mathrm{e}^{-2\pi^2\sigma^2 f^2} = A\sqrt{2\pi\sigma^2}\mathrm{e}^{-\frac{f^2}{2\frac{1}{4\pi^2\sigma^2}}} = A\sqrt{2\pi\sigma^2}\mathrm{e}^{-\frac{f^2}{2\hat{\sigma}^2}}$$

where $\hat{\sigma}^2 = \frac{1}{4\pi^2\sigma^2} \iff \hat{\sigma} = \frac{1}{2\pi\sigma}$. We set the corresponding value of the standard deviation to Eq. 3.16 and we get that $W_{10} = 4.29193/(2\pi\sigma)$ Hz.

In order to evaluate the validity of the formulas stated above, they are considered for constructing a specified Gaussian pulse. Specifically in Fiq. 3.1.5, a Gaussian pulse with energy 1 and width $T_p = 1$ sec is given along with its Fourier Transform. To set the width, the relationship of Eq. 3.12 has been used with $T_p = 1$, and as it is observed, we can satisfactorily approximate the desirable value. The Fourier Transform was typically calculated and is shown, too. Taking into account Eq. 3.15, the bandwidth is estimated as 8 Hz, and as we observe, this approximation is also sufficient. Thus, both theoretical approaches (i.e., for pulse's duration and bandwidth) are compatible to the practical results obtained in the relevant figure.

In addition, using Eq. 3.16, we get that the W_{10} at the time domain is 0.6054 sec and at the frequency domain is 4.8429 Hz. In Fig. 3.1.5, a red line has been designed, which is equal to the one tenth of the max value of each "bell". Practically calculations show that, in the time domain, W_{10} value equals $0.8027 - 0.197 = 0.6057$ sec and in the frequency domain equals $2 \cdot 2.49 = 4.98$ Hz, indicating the effectiveness of this theoretical approach, as well.

Considering a Gaussian pulse given by Eq. 3.11, we calculate its auto-correlation function as follows.

[2]http://en.wikipedia.org/wiki/Gaussian_function

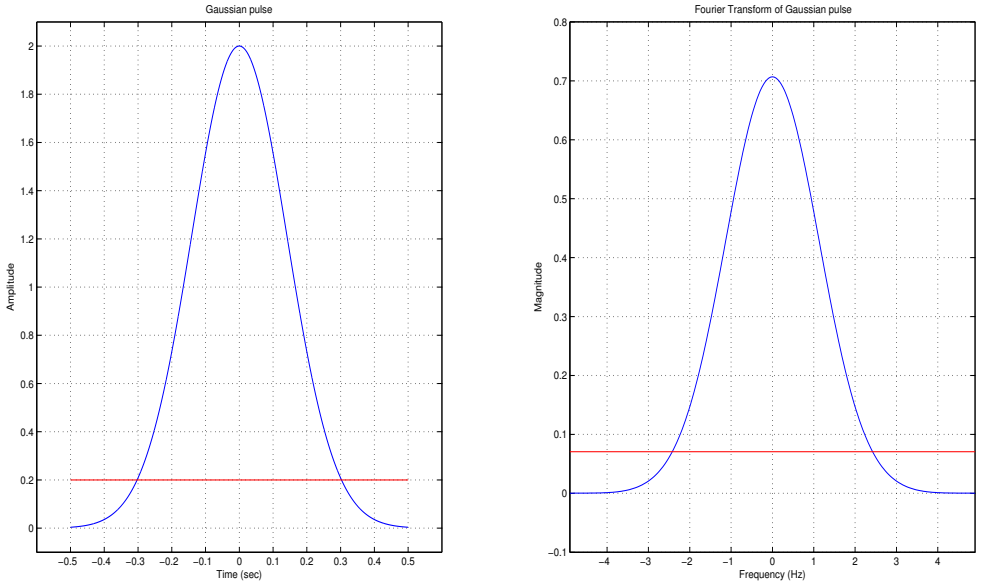

Figure 3.1.5: Gaussian pulse in time and frequency domain.

$$
\begin{aligned}
R_{pp}(t+\tau,t) &= \int_{-\infty}^{+\infty} p(t+\tau)\,p(t)\,\mathrm{d}t \\
&= A^2 \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{(t+\tau)^2}{2\sigma^2}}\,\mathrm{e}^{-\frac{t^2}{2\sigma^2}}\,\mathrm{d}t \\
&= A^2 \mathrm{e}^{-\frac{\tau^2}{2\sigma^2}} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{\sigma^2}t^2}\,\mathrm{e}^{-2\frac{\tau}{2\sigma^2}t}\,\mathrm{d}t \\
&= A^2\sqrt{\pi\sigma^2}\,\mathrm{e}^{-\frac{\tau^2}{2\sigma^2}}\,\mathrm{e}^{\frac{\left(\frac{\tau}{2\sigma^2}\right)^2}{\frac{1}{\sigma^2}}} \\
&= A^2\sigma\sqrt{\pi}\,\mathrm{e}^{-\frac{\tau^2}{4\sigma^2}}
\end{aligned}
$$

We define the scalar $c = A^2\sigma\sqrt{\pi}$ and, thus, we get

$$R_{pp}(t+\tau,t) = c\,\mathrm{e}^{-\frac{\tau^2}{4\sigma^2}}. \tag{3.17}$$

5

From the latter expression, it is easily observed that the auto-correlation function of the Gaussian pulse, is also a Gaussian pulse with bigger width. Specifically, its variance equals $\sigma_{ac}^2 = 2\sigma^2$ and equivalently its standard deviation is $\sigma_{ac} = \sqrt{2}\,\sigma$, in relevance to the Gaussian pulse of Eq. 3.11. In an analogous way and according to Eq. 3.12, we expect that the time duration of the auto-correlation function approximately equals $\sqrt{2}\,T_p$ sec. Its Fourier Transform (recalling link [1]), is given by Eq. 3.18.

$$\mathcal{F}\{R_{pp}(t+\tau,t)\} = 2\,c\,\sigma\sqrt{\pi}\,\mathrm{e}^{-4\pi^2\sigma^2 f^2} \tag{3.18}$$

Note that the bandwidth of the auto-correlation function is shrinked by a factor of $1/\sqrt{2}$, in relevance to the bandwidth of the Gaussian pulse. This result can be derived, if we set

the new value of width, $\sqrt{2}T_p$, to Eq. 3.15. These observations regarding the duration and the bandwidth are practically verified in Fig. 3.1.6, where the auto-correlation of the Gaussian pulse shown in Fig. 3.1.5 is given, both in time and in frequency domains.

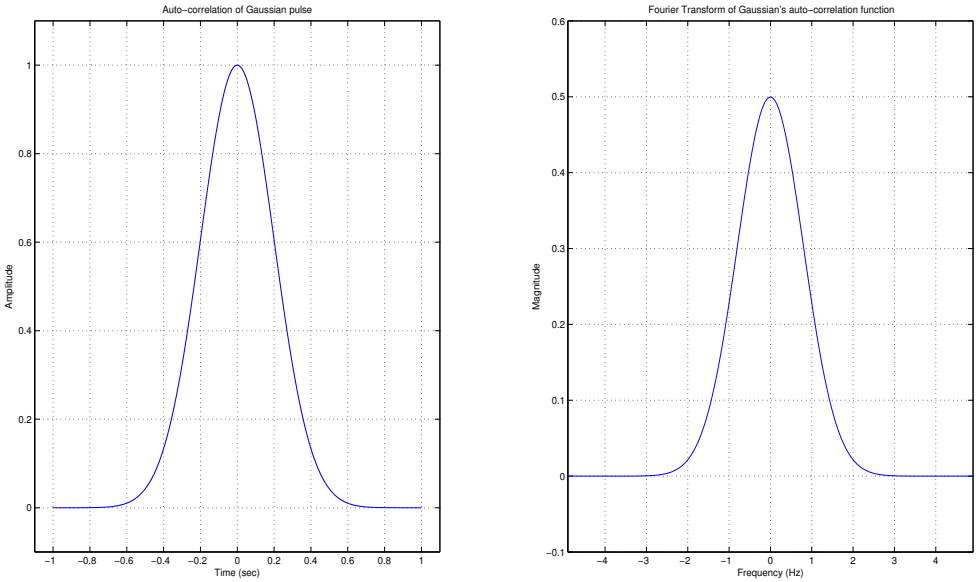

Figure 3.1.6: Auto-correlation of gaussian pulse in time and frequency domain.

We derive the mathematical formulas concerning the Gaussian derivatives. These pulses have properties that could be beneficial in our system setup. The advantage is that they behave like bandpass filters, with the main spectrum to be shifted towards higher frequencies, as the order of the derivative increases. This displacement of spectrum is even faster, as the time duration of the pulse is reduced.

Assuming the gerenal form of the Gaussian pulse, given by Eq. 3.11, its n^{th} derivative is expressed recursively by the following relation.

$$p^{(n)}(t) = -\frac{n-1}{\sigma^2}p^{(n-2)}(t) - \frac{t}{\sigma^2}p^{(n-1)}(t) \tag{3.19}$$

The corresponding Fourier Transform, is given in Eq. 3.20.

$$\mathcal{F}\{p^{(n)}(t)\} = A\sqrt{2\pi\sigma^2}\,(j2\pi f)^n\,\mathrm{e}^{-2\pi^2\sigma^2 f^2}. \tag{3.20}$$

In Fig. 3.1.7 and 3.1.8, the shape of the 15th derivative of the Gaussian pulse and its auto-correlation function are shown in both the time and frequency domains. The energy of the pulse is set 1 and the time duration $T_p = 1$ sec. Note that we have set standard deviation $\sigma = \frac{T_p}{4\sqrt{\pi}}$ and $T_p = 1$ sec and, as it is observed, the desirable time duration is approximated. For smaller derivatives, this approach is more accurate.

3.1.2 Pulse choice

The SRRC, the rectangular pulse, the Gaussian function and its derivatives are chosen as candidate transmitting pulses for the system. In any case, in order to achieve high

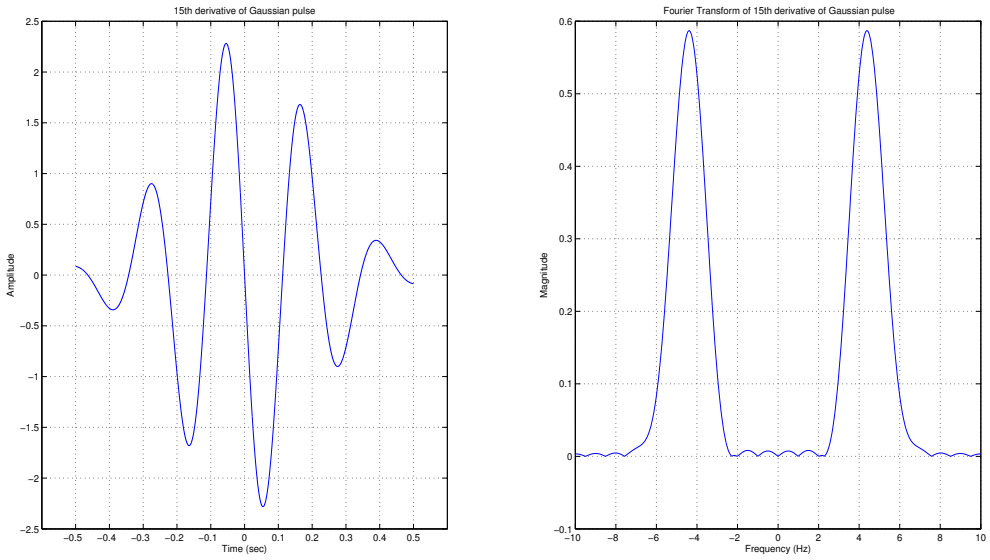

Figure 3.1.7: Gaussian pulse's derivative in time and frequency domain.

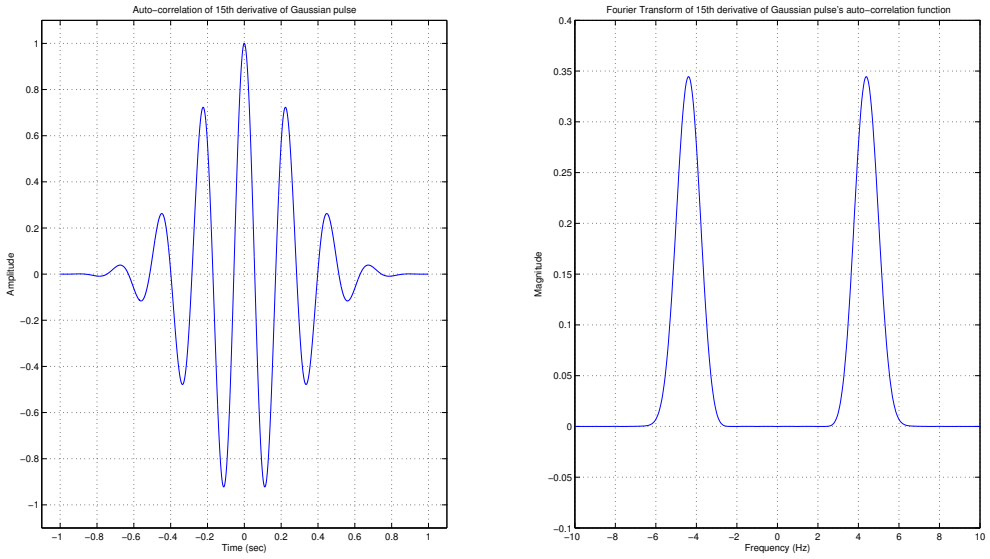

Figure 3.1.8: Auto-correlation of gaussian pulse's derivative in time and frequency domain.

performance, narrow pulses should be used. The minimum value of the time duration that does not alter the most of the candidate pulse shapes, found to be 0.3 msec. It is assumed that $F_s = 44100$ and, thus, 0.3 msec is translated to $\frac{0.0003}{1/44100} \approx 13$ samples. Considering the pulses analyzed in Section 3.1.1 to decide which one best fits, the normalized[3] magnitude of the Fourier Transform is estimated and shown in Fig. 3.1.9, given the aforementioned value for each pulse's duration.

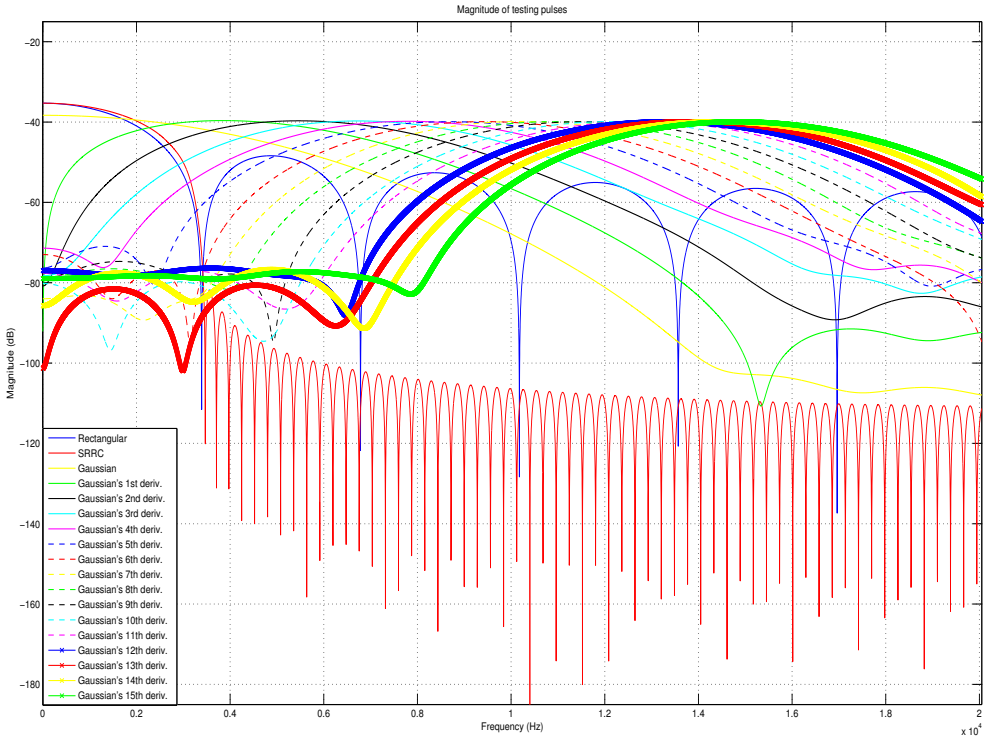

Figure 3.1.9: Magnitude of testing pulses, assuming $T_p = 0.3$ msec and energy equals 1.

3.1.2.1 15th derivative of Gaussian pulse, for pilot as pulse train

The pulse that conforms to our regulations and could be included to the construction of the pilot as a pulse train, is a high order derivative of the Gaussian pulse. We observe that, as the order of the derivative increases, the spectrum is shifted towards higher frequency bands. Hence, the order of the derivative has been augmented until the desirable pulse spectrum was achieved, given the specified value for pulse duration. Note that for broader pulses the spectrum becomes narrower and an even higher order derivative is necessary, in order to locate the spectrum to appropriate frequencies.

Based on Fig. 3.1.9 and according to our assumptions, the 15th derivative of the Gaussian pulse (plotted with bold green color) was selected. The other main requirement of

[3]The energy of each pulse equals 1, because we are interested in the shape of the corresponding spectrum.

the system is that the pilot signal must be acoustically untraceable. For that reason it was found, through experimentation, that each suggested pulse must have energy $\leq 4 \cdot 10^{-13}$, given that its duration equals 0.3 msec. Finally, in Fig. 3.1.10 the proposed pulse in both time and frequency domains, is shown. Note that, the frequency domain graph presents the square of the absolute value of the pulse's Fourier Transform, divided by its time duration.

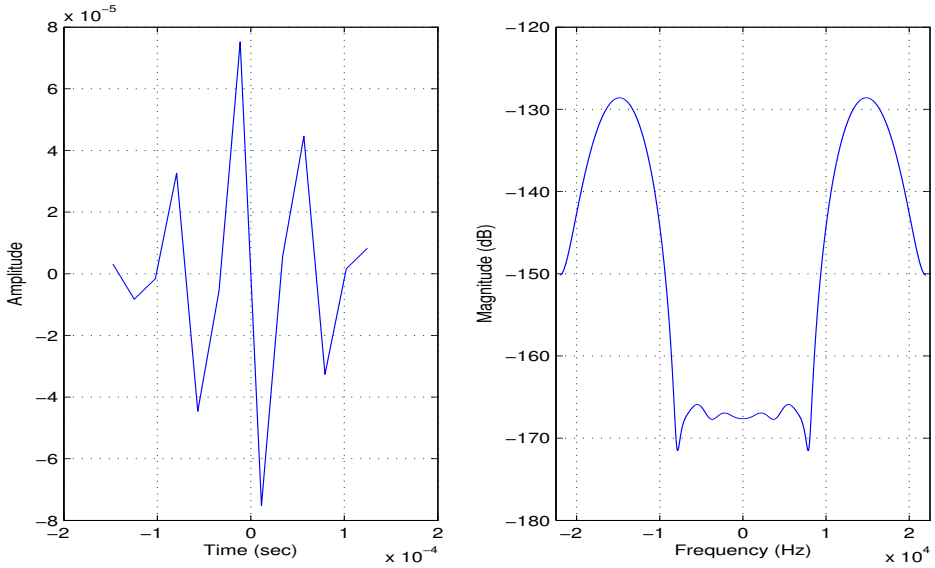

Figure 3.1.10: 15th derivative of the Gaussian pulse, with $T_p = 0.3$ msec and energy $4 \cdot 10^{-13}$.

The following process explains how we concluded to the use of the 15th derivative of the Gaussian pulse.

1. Firstly, we set a desirable power per pulse ($\sim 10^{-9}$). This value, obviously, has to be as big as possible, to ensure enhanced robustness. Simultaneously, the pulse has to be inaudible.

2. Secondly, we observed that the Gaussian derivatives behave as passband filters, and as the order of the derivative increases, the spectrum is transferred to higher frequencies.

3. In addition, as the width of these derivatives is decreasing, the same behavior is observed.

4. We exploit the Eq. 3.12, concerning the standard deviation that has to be defined, to achieve a specified duration. The time duration of each pulse, is a multiplication between the number of samples and the sampling period. Assuming that sampling frequency is constant, the pulse shape becomes a function of samples, given the order of the derivative.

5. We started from the minimum number of samples and, for each case, an evaluation for each derivative of the Gaussian pulse until the 15th order was performed. Gradually we reached at 13 samples for which the latter pulse was silent under the desirable power, and the shape was not essentially altered. Less pulse samples is desirable, because the lower derivatives ensure the bandwidth allocation in higher

frequencies, and thus less computational resources are needed. In addition, better performance for the delay estimation is expected.

3.1.2.2 Square-root raised cosine, for pilot with carrier

In this case, the selected pulse is the SRRC. Beside the fact that it is used widely in similar systems, its properties are appreciated for our goals. First of all, its spectrum is quite smooth and its bandwidth can be used to adjust the exact desirable limits by the roll-off factor. It is obvious that the exact band limits are related to the chosen center frequency. The spectrum of the modulated signal is equal to $(1 + a)/T$, located around the center frequency F_c, and the available bandwidth is limited to 22.05 kHz. Furthermore, the highest modulation frequency is desirable to be selected, in order to transfer signal's spectrum to a band where human is insensitive to sounds, according to Fletcher-Munson curve and, simultaneously, music power has decreased values based on Fig. 2.1.13.

In this study, the modulation of the baseband signal at the center frequency 18.997 kHz is selected, without limiting the generality of choosing another close value. To meet the aforementioned requirements, we set pulse duration equal to 0.35 msec and the roll-off factor $a = 0$, indicating that the bandwidth of modulated signal approximates 2.85 kHz and it is centered around 18.997 kHz, respecting the appropriate bandwidth limits. Finally, in order to be inaudible, it was found that each suggested pulse must have energy $\leq 5 \cdot 10^{-9}$, given that its duration is equal to 0.35 msec. It is noted that the restriction concerning the loudness of the pilot signal is more loose in this case, which is verified by ATH of Fig. 2.2.1. Fig. 3.1.11 shows the SRRC in time domain and its spectrum in both baseband and passband.

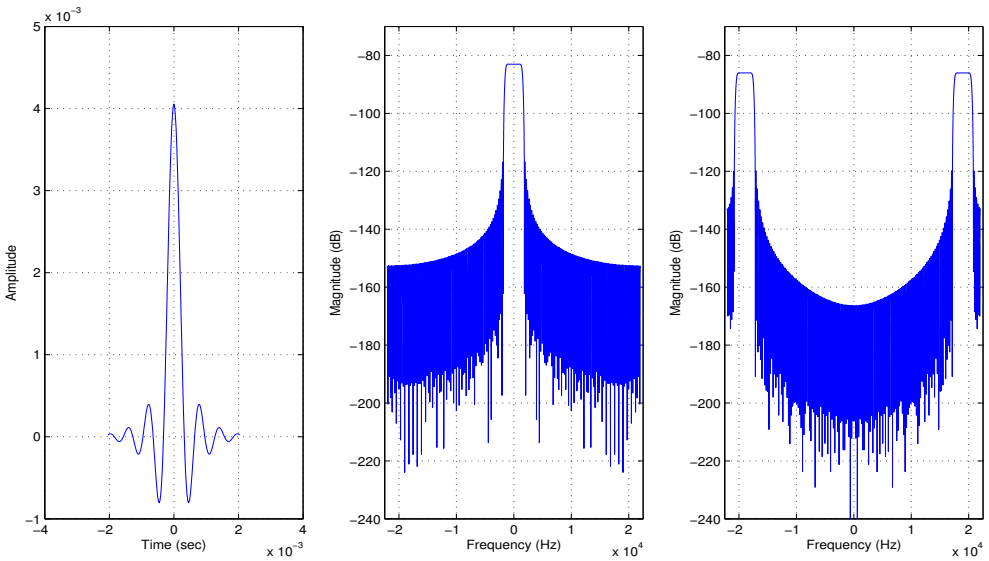

Figure 3.1.11: SRRC pulse, with $T_p = 0.35$ msec and energy $5 \cdot 10^{-9}$.

3.1.2.3 Remarks

In all cases we concluded to these values for energy per pulse, after experimentation. In general, a maximum energy per pulse has to be chosen, in order to make the transmission more robust and to obtain better performance under the same power of noise. The loudness of the signal is directly a function of power. Specifically, the loudness of the signal is derived from the loudness of the corresponding pulses that are included in the signal. Hence, to avoid misconceptions relative to the signal's structure, the loudness of the pulse and by extension of the signal, is considered a function of pulse's power, which is, by default, the ratio of the energy per pulse divided by its width and equals the signal power. The time duration of the signal is proportional to the number of the pulses that are included. By fixing the time duration of each pulse, in order to comply to bandwidth restrictions that we have set, the loudness of the signal is amplified when the energy per pulse is increased. The experimentations were performed under these observations and the maximum energy per pulse has been chosen, for which the signal was remaining silent under the evaluation of implementations where (a) only the corresponding pilot signal was included, and (b) a clear sound was also present. The values mentioned above concerning the width and energy of each pulse for each scenario, give us the corresponding power values.

- Power of 15th derivative of Gaussian pulse $\approx \frac{4 \cdot 10^{-13}}{0.3 \cdot 10^{-3}} = 1.33 \cdot 10^{-9}$.
- Power of SRRC pulse $\approx \frac{5 \cdot 10^{-9}}{0.35 \cdot 10^{-3}} = 1.47 \cdot 10^{-5}$.

We observe that in different bands, a different maximum value of power is obtained, in order for a single pulse to be untraceable. In general, that observation is expected and verified by the Fletcher-Munson curve, which is not flat over frequencies.

3.2 Modulation technique

Different modulation techniques will be used, for the various cases of the signal construction. Among them, spread spectrum techniques and modulation using a single-carrier, are considered. On one hand, the spread spectrum techniques ensure high performance in the sense of estimating the delay, on the other hand, locating the spectrum of the pilot signal in a band where the human is insensitive in hearing, ensures that stronger signals will be untraceable. Both approaches, will be analyzed and under simulations their effectiveness will be evaluated.

3.2.1 Spread Spectrum modulation

The proposed technique, here, is to construct a pilot as a pulse train and to spread its bandwidth as wide as possible, over the band that the audio interface defines, i.e., $[0,\ 22.05]$ KHz. The benefits of spread spectrum techniques, are derived from pseudo-random nature of the signal and the wide bandwidth. For that reason, the methods that are primarily used, is either Time Hopping (TH) or Direct Sequence (DS). The advantages are that they offer resistance to interferers, hide the signal from reception and improve multiple access capability. The two last properties are mainly convinient to our case. Before its transmission, the signal has to pass through 3 stages at the receiver, in order to spread and smooth its spectrum. The spectrum smoothness denotes that powerful

spectral lines at the frequency domain are avoided, which is obviously desirable. Firstly, a generator of randomized code is applied, in order to locate in a pseudo-random place each pulse, or to handle pseudo-randomly each pulse's amplitude, that corresponds to a single bit representation (TH and DS technique, respectively). Secondly, the modulation of the transmitting data takes place, i.e., the desirable modulation scheme is applied given the selected bit-code and, finally, the using pulse is generated.

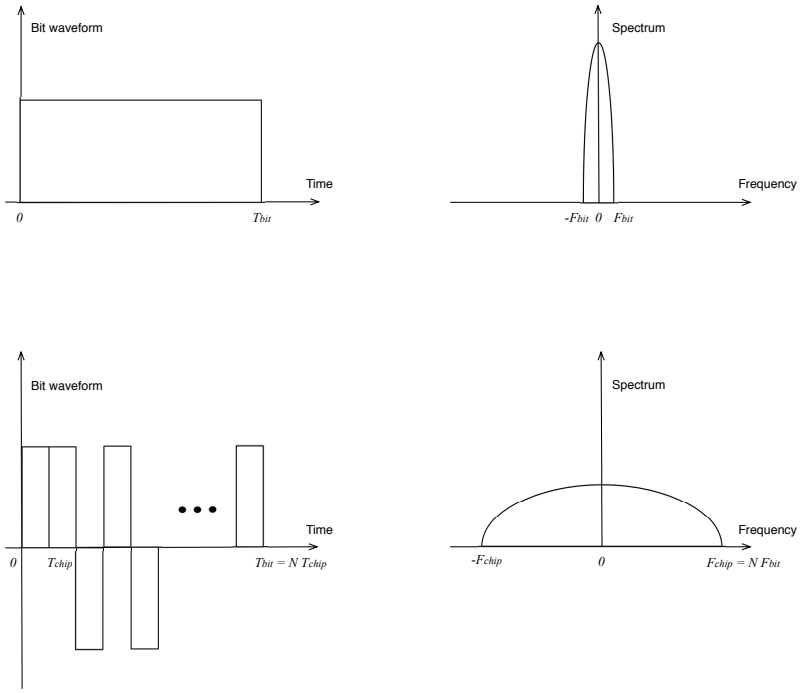

Figure 3.2.1: Spread Spectrum technique, in general.

The properties of the spread spectrum techniques are highlighted in Fig. 3.2.1, where the general shape of a typical bit-waveform and the one that is obtained by applying the DS technique are shown.

3.2.1.1 Time-Hopping

TH technique is based on multiple pulses that are placed pseudo-randomly in time, and consist an information bit. The waveform for a set of transmitted information bits b_i is:

$$b(t) = \sum_{i=0}^{N_b-1} b_i \, p_b(t - iT_b) \tag{3.21}$$

where $b_i \in \{-1, +1\}$ corresponding to bit 0 and 1, respectively, N_b denotes the number of transmitted bits, T_b is the bit duration, and $p_b(t)$ is the pulse used to represent each bit. To apply a TH spreading code to each bit, the following assumptions are considered. Time axis initially is divided in N_f non-overlapping time intervals, with duration equal to T_f sec that are called frames. An information bit is spread over $T_b = N_f \, T_f$ sec. In each frame a single pulse is placed to a pseudo-randomly position that is selected according to the user

code. To be able to determine the exact position, each frame is subdivided in N_c smaller non-overlapping intervals of duration T_c sec ($T_f = N_c T_c$ sec) that are called chips. In each chip a single pulse could be placed and in general, the pulse width is $\leq T_c$. In our case, we assume that the equality holds. The chip that is selected to contain the unique pulse over each frame, depends on the TH code c_j that was generated pseudo-randomly. Each value of $c_j, \forall j \in [0,\ N_f - 1]$ is selected by a uniform distribution over the range $[0,\ N_c - 1]$. The spreading waveform for a single bit will be given by

$$p_b^{(m)}(t) = \sum_{j=0}^{N_f - 1} p(t - jT_f - c_j^{(m)} T_c). \tag{3.22}$$

This technique offers multiple access capability, by defining a different TH code that corresponds to each user. The expansion factor of the bandwidth, which is called the "processing gain", is equal to $P_G = (N_f T_f)/T_c$, or in general $P_G = T_{\text{bit}}/T_{\text{chip}}$. With the above analysis, the transmitted pilot signal can be written as follows

$$s^{(m)}(t) = \sum_{i=0}^{N_b - 1} b_i^{(m)} \sum_{j=0}^{N_f - 1} p(t - iT_b - jT_f - c_j^{(m)} T_c) \tag{3.23}$$

where $p(t)$ is the using pulse, $c_j^{(m)}$ is the TH code of user m and $\{b_i^{(m)}\}$ sequence that corresponds to information bits of the m^{th} user.

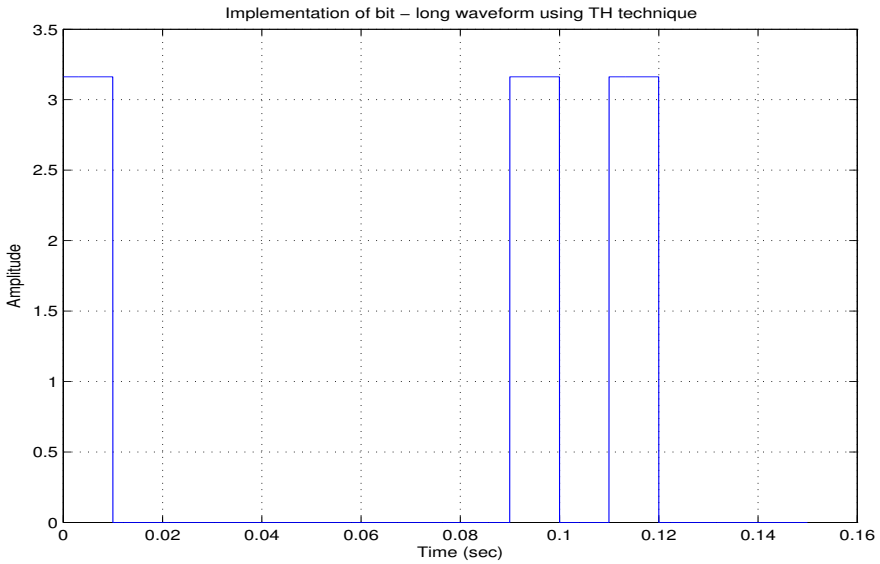

Figure 3.2.2: Typical implementation of the TH technique.

In Fig. 3.2.2, the bit duration is defined as $N_f T_f = 3 \cdot 0.05 = 0.15$ sec. We assume that the bit value is 1. Each pulse is a rectangular pulse with width equal to 0.01 sec and is normalized so that pulse's energy equals 0.1. We have defined $T_f = 0.05$ sec and $T_c = 0.01$ sec, which equals the pulse width. That indicates that $N_c = 5$ and a single pulse in each frame could be placed in any of the 5 chips that correspond to it. In this implementation we randomly selected $c = [0, 4, 1]$, i.e., for the first frame we set the pulse in chip 0, for the second frame the pulse is placed in the chip 4 and in the third frame in chip 1. To explain how the c_j code is applied, a calculation that returns the time intervals

where each pulse i is placed in every frame, is the following. Pulse $i, \forall i \in [0,\, N_f - 1]$, is placed in $[i\,T_f + c_i\,T_c,\ i\,T_f + (c_i + 1)\,T_c]$ time interval.

3.2.1.2 Direct-Sequence

The general structure of the DS modulation technique is based on the application of a spreading code to each transmitted bit pulse. The waveform for a set of transmitted information bits b_i is

$$b(t) = \sum_{i=0}^{N_b-1} b_i\, p_b(t - iT_b) \tag{3.24}$$

where $b_i \in \{-1, +1\}$ corresponding to bit 0 and 1, respectively, N_b denotes the number of transmitted bits, T_b is the bit duration, and $p_b(t)$ is the pulse used to represent each bit. To apply a spreading code to each bit, split the bit duration to N_c non-overlapping intervals called chips. Each chip has a duration equal to $T_c = T_b/N_c$. The processing gain, N_c, is assumed to be much larger than one. The spreading waveform for a single bit will be given by:

$$p_b(t) = \sum_{j=0}^{N_c-1} c_j\, p_c(t - jT_c) \tag{3.25}$$

where c_j is the spreading code and its value is selected equiprobably between ± 1. $p_c(t)$ is the chip pulse shape. The bandwidth of the spreading waveform is much greater than the bandwidth of the information signal and the spectral characteristics of the transmit signal are dominated by the spreading signal. Specifically, the chip rate and chip pulse shape along with the autocorrelation properties of the spreading sequence will determine the transmit signal's spectral properties.

The processing gain is defined identically, and is the same in relevance to TH technique. With the above analysis, the transmitted pilot signal can be written:

$$s^{(m)}(t) = \sum_{i=0}^{N_b-1} b_i^{(m)} \sum_{j=0}^{N_c-1} c_j^{(m)}\, p_c(t - iT_b - jT_c) \tag{3.26}$$

where $p_c(t)$ is the using pulse, $c_j^{(m)}$ is the DS code of user m and $\{b_i^{(m)}\}$ sequence that corresponds to information bits of the m^{th} user.

Summarizing, the total time duration of the pilot signal is defined as $T = N_b\,T_b$ and in each bit duration, T_b, the chosen pulses are placed sequentially over $N_c = T_b/T_c$ non-overlapping chips. The sign of each pulse placed in the pilot is defined by the DS code c_j in combination with the corresponding bit value b_i.

In Fig. 3.2.3, bit duration is defined as $T_b = N_c\,T_c = 15 \cdot 0.01 = 0.15$ sec, respectively. We assume that bit value is 1. Each pulse is a rectangular pulse with width equal to 0.01 sec and is normalized so that pulse's energy equals 0.1. Note that, T_c equals the pulse width. In Fig. 3.2.3, we used that $c = [-1, 1, 1, 1, -1, 1, -1, -1, -1, -1, 1, -1, 1, -1, 1]$.

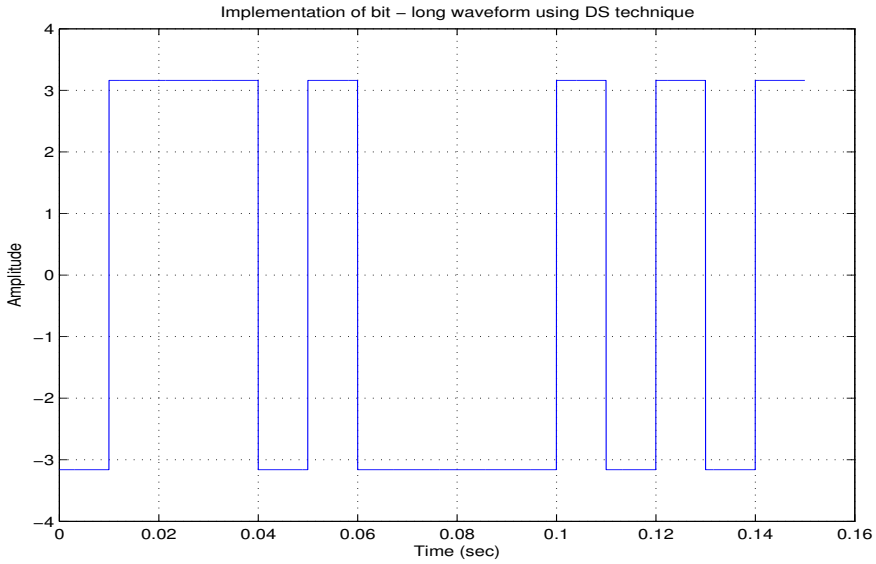

Figure 3.2.3: Typical implementation of the DS technique.

3.2.1.3 Selected spread spectrum technique

We choose DS modulation over TH for the following reasons. Firstly, the minimum duration of the pilot signal is desired, in order to minimize the probability of (a) being perceived by human, and (b) receiving a part of the transmitting signal due to packet loss or jitter. A cross-correlation technique is applied at the receiver and the delay estimation is based on detecting the peak at the output. Thus, the maximum possible signal energy is desirable. In addition, as previously mentioned, we have a practical constraint for energy per pulse, to ensure that it remains inaudible. We therefore need the maximum possible number of pulses, given a fixed value of the signal duration. This requirement is met with the DS method.

A noteworthy observation is that in our system, there is no need to transmit information bits. What is of interest is the exact temporal position of the information, and not its content. In fact, the information is already known at the receiver. This is obviously very different from a typical communication system, where the goal is to exchange information between the two ends. In Eq. 3.26, bits are selected in order to eliminate spectral lines that would be present in frequency domain, spaced by $1/T_b$. An equivalent system can be obtained if we assume that our total pilot duration corresponds to one bit duration (i.e., $T = T_b$, $N_b = 1$, and $b_0 = 1$). In this case, our system is a corner case of equivalent to a 2-PAM system, without using any carrier.

3.2.2 BPSK

The other approach to transfer the spectrum of our pilot to desirable frequencies, where human perception on sounds is quite low, is achieved by modulating the band-limited signal under the use of a carrier. The simplest and one of the most used techniques, is the Binary Phase-Shift Keying (BPSK). In Fig. 3.2.4 we demonstrate in frequency domain the basic property of modulating the signal, with F_c.

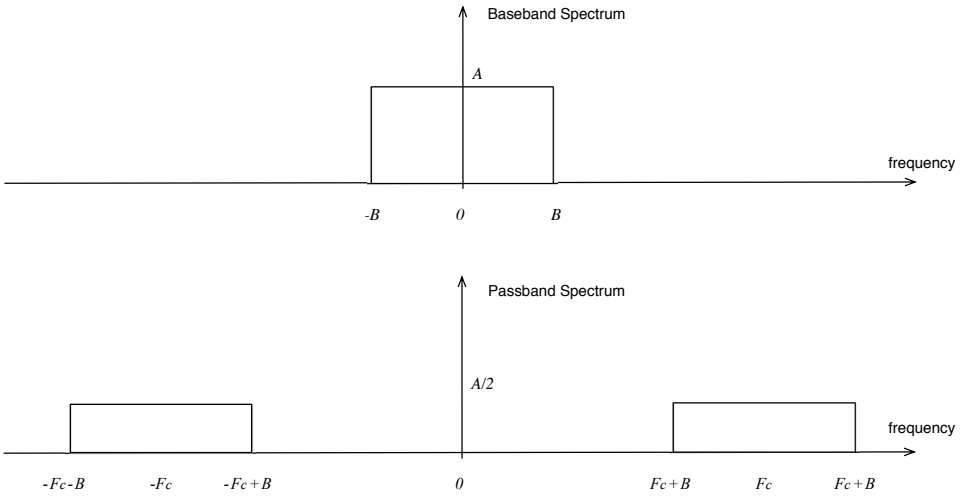

Figure 3.2.4: BPSK technique, in general.

BPSK (which is equivalent to 2-PAM) is the simplest form of phase shift keying (PSK). It uses two phases which are separated by 180°. This modulation is the most robust of all the PSKs since it takes the highest level of noise or distortion to make the demodulator reach an incorrect decision. It is able to modulate at 1 bit/symbol and, hence, it is unsuitable for high data-rate applications. We are not interested in data-rate nor the transmission of information symbols, as stated above. However, note that, if we consider the transmission of only one information bit (i.e. only bit 0, or 1), due to the resulting signal shape, the performance for the delay estimation will be decreased. In addition, powerful spectral lines that are spaced by $1/T_b$, will be appeared at the spectrum. In contrast, the constellation of BPSK in relevance to the appropriate mapping, is suitable in order to avoid those spectral lines.

Hence, numerous pulses are transmitted, with the same absolute value and different sign (defining BPSK), and the exact number of those pulses is derived as the ratio of the desirable duration of the transmitted signal divided by the pulse width. The main advantage that we are interested in using a single-carrier technique, is the fact that we can modulate typical pulses in specific frequencies, that could make the signal inaudible. Note that the baseband signal with the same power will be probably audible. Under these circumstances, the most simple single-carrier method was chosen.

A general mathematical representation of BPSK, is given in Eq. 3.27

$$s^{(m)}(t) = \sqrt{2} \sum_{i=0}^{N_b-1} b_i^{(m)} \, p(t - i\,T_b) \, \cos(2\,\pi\,F_c^{(m)}\,t) \tag{3.27}$$

where $p(t)$ is the using pulse, $\{b_i^{(m)}\}$ sequence that corresponds to information bits and $F_c^{(m)}$ the selected center frequency of the m^{th} user. In the previous formula, for each user, we assume the transmission of N_b information bits. For each pulse we assume a different sign, that is chosen equiprobably in the range $\{\pm 1\}$ and is given by $b_i, \forall i \in [0,\, N_b - 1]$.

In Fig. 3.2.5, a pilot signal with duration equal to $T = T_b\,(N_b - 1) = 0.01 \cdot 14 = 0.14$ sec, is defined and given both in baseband and in passband, assuming $F_c = 1.5$ kHz. Each transmitting pulse is a SRRC with width $T_b = 1$ sec, energy 0.1 per pulse, and its sign is selected equiprobably. We set signal's energy to be equal for both the baseband

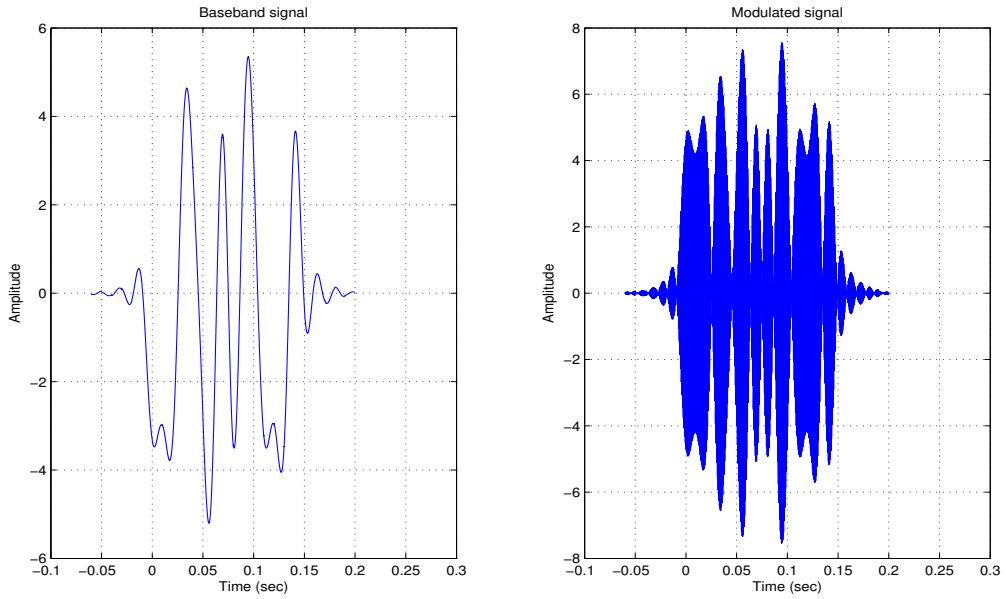

Figure 3.2.5: Typical implementation of BPSK technique - baseband and modulated signals.

and the modulated waveforms. The time duration of the useful signal is obtained as $(N_b - 1)\,T_b$, because the first bit value is placed at time instance $t = 0$. Actually, SRRC pulse involves a two-sided tail which is spread over $A\,T_b$ sec, for each side and in this case it is defined as $A = 6$. Thus, the signal is spread over $(N_b - 1)\,T_b + 2\,A\,T_b = 0.26$ sec. The modulation frequency is $F_c = 1.5$ kHz. For this specific implementation, the bit values are: $[-1, -1, -1, 1, 1, -1, -1, 1, -1, 1, 1, -1, -1, -1, 1]$. The time interval where each pulse $i, \forall i \in [0, N_b - 1]$ is placed, is $[(i-1)\,T_b,\ i\,T_b]$.

3.3 Power spectral density

PSD shows, “on average”, the allocation of power in the frequency domain and, hence, it is a meaningful metric for the transmitted signal, in order to understand its spectral properties. For a large range of signals, it is valid to assume that the PSD of a signal is equal to the “average” value of its periodogram, when its time duration T tends to infinity. Specifically, the periodogram of a signal $x(t)$ is defined as $\frac{|X(f)|^2}{T}$, where $X(f)$ is the Fourier Transform of the signal and T is the time duration over which it is spread. According to the previous statement, the PSD is defined as

$$S_X(f) = \lim_{T \to \infty} \frac{E\{|X(f)|^2\}}{T}. \tag{3.28}$$

For a stationary in wide-sense (WSS) random process $x(t)$, the PSD is equal to the Fourier Transform of its auto-correlation function.

$$S_X(f) = F.T.\{R_x(\tau)\} \tag{3.29}$$

Furthermore, if $x(t)$ is a cyclo-stationary process, in the wide-sense, then

$$S_X(f) = F.T.\{\overline{R}_x(\tau)\},\ \text{where}\ \overline{R}_x(\tau) = \frac{1}{T}\int_T R_{xx}(t+\tau, t)\,\mathrm{d}t. \tag{3.30}$$

Note that, in our analysis below, we calculate the average PSD over numerous implementations of each considering signal construction.

3.3.1 Direct Sequence

For the DS modulation technique, we defined the general form of the transmitting signal in Eq. 3.26. To calculate the PSD, we assume $m = 1$ user and we define the time axis in a convinient way, as follows

$$s(t) = \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} b_i \sum_{j=-\tilde{N}_c}^{\tilde{N}_c} c_j \, p_c(t - iT_b - jT_c) \tag{3.31}$$

where $p_c(t)$ denotes the using pulse. Note that, the range of the sum regarding the number of information bits, is defined as $[-\tilde{N}_b,\, \tilde{N}_b]$ which corresponds to $[-(N_b-1)/2,\, (N_b-1)/2]$, when N_b is an odd number, and $[-N_b/2+1,\, N_b/2]$, when N_b is even. In other terms, we assume that the equality $2\,\tilde{N}_b + 1 = N_b$, always holds. Identical assumptions are also considered for the number of chips N_c. The random variables b_i and c_j are independent and uniformly distributed. Specifically, we assume that $b_i \in \{\pm 1\},\ \forall i \in [-\tilde{N}_b,\, \tilde{N}_b]$ and $c_j \in \{\pm 1\},\ \forall j \in [-\tilde{N}_c,\, \tilde{N}_c]$. The statistics of b, are given below.

$$\begin{aligned} &\mu_b = E\{b\} = 0 \\ &\mathrm{var}\{b\} = E\{(b - E\{b\})^2\} = \sigma_b^2 = 1 \end{aligned}$$

It is obvious that the statistics of c, are identical. Then, we calculate the mean value and the auto-correlation function of the transmitting signal, in order to derive its PSD.

$$E\{s(t)\} = \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} E\{b_i\} \sum_{j=-\tilde{N}_c}^{\tilde{N}_c} E\{c_j\}\, p_c(t - iT_b - jT_c) = 0$$

The auto-correlation function, $R_{ss}(t+\tau, t)$, is calculated as follows

$$\begin{aligned} R_{ss}(t+\tau, t) &= E\{s(t+\tau)\, s(t)\} \\ &= E\left\{ \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} b_i \sum_{j=-\tilde{N}_c}^{\tilde{N}_c} c_j p_c(t+\tau - iT_b - jT_c) \sum_{k=-\tilde{N}_b}^{\tilde{N}_b} b_k \sum_{n=-\tilde{N}_c}^{\tilde{N}_c} c_n p_c(t - kT_b - nT_c) \right\} \\ &= E\left\{ \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} \sum_{k=-\tilde{N}_b}^{\tilde{N}_b} b_i b_k \sum_{j=-\tilde{N}_c}^{\tilde{N}_c} \sum_{n=-\tilde{N}_c}^{\tilde{N}_c} c_j c_n p_c(t+\tau - iT_b - jT_c) p_c(t - kT_b - nT_c) \right\} \\ &\overset{*}{=} \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} E\left\{b_i^2\right\} \sum_{j=-\tilde{N}_c}^{\tilde{N}_c} E\left\{c_j^2\right\}\, p_c(t+\tau - iT_b - jT_c)\, p_c(t - iT_b - jT_c) \\ &= \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} \sum_{j=-\tilde{N}_c}^{\tilde{N}_c} p_c(t+\tau - iT_b - jT_c)\, p_c(t - iT_b - jT_c) \end{aligned}$$

$$\overset{**}{=} \sum_{m=-N}^{N} p_c(t+\tau-mT_c)\, p_c(t-mT_c)$$

where at the forth equality ("$*$") we used that the double sums are reduced to single ones, because only the products for $k=i$ and $n=j$, respectively, do survive. At the last equality ("$**$"), an equivalent expression for the double sum was considered, by taking advantage of the relationship between the chip and the bit duration ($T_b = N_c\, T_c$), and by modifying the range of the double sum appropriately, using that $2\,N+1 = N_b\, N_c$. It is obvious that $N \gg 1$.

It is easy to be proved that $R_{ss}(t+\tau,t) = R_{ss}(t+T_c+\tau, t+T_c)$, when N tends to infinity, which means that $R_{ss}(t+\tau,t)$ is periodic to the variable T_c. Based on the derivation of $s(t)$'s mean value, its auto-correlation function and the last observation concerning periodicity, we conclude that $s(t)$ is a cyclo-stationary process in the wide-sense, with period T_c. The PSD of $s(t)$ is then given, according to Eq. 3.30, as $S_s(f) = \mathcal{F}\{\overline{R}_s(\tau)\}$, with

$$\begin{aligned}
\overline{R}_s(\tau) &= \frac{1}{T_c}\int_{T_c} R_{ss}(t+\tau,t)\,\mathrm{d}t \\
&= \frac{1}{T_c}\int_{-T_c/2}^{T_c/2} \sum_{m=-N}^{N} p_c(t+\tau-mT_c)\, p_c(t-mT_c)\,\mathrm{d}t \\
&= \frac{1}{T_c}\sum_{m=-N}^{N} \int_{-T_c/2}^{T_c/2} p_c(t+\tau-mT_c)\, p_c(t-mT_c)\,\mathrm{d}t \\
&\overset{!}{=} \frac{1}{T_c}\int_{-N\cdot T_c-T_c/2}^{N\cdot T_c+T_c/2} p_c(t+\tau)\, p_c(t)\,\mathrm{d}t \\
&= \frac{1}{T_c} R_{p_c p_c}(t+\tau,t) \\
&= \frac{1}{T_c}\, p_c(\tau) * p_c(-\tau)\,\mathrm{d}t.
\end{aligned}$$

The following calculations derive the expression at the spot "!".

$$\sum_{m=-N}^{N} \int_{-T_c/2}^{T_c/2} p_c(t+\tau-mT_c)\, p_c(t-mT_c)\,\mathrm{d}t \overset{\grave{t}=t-mT_c}{=} \sum_{m=-N}^{N} \int_{-mT_c-T_c/2}^{-mT_c+T_c/2} p_c(\grave{t}+\tau)\, p_c(\grave{t})\,\mathrm{d}\grave{t}$$

The last expression denotes the sum of $2\,N+1$ integrals, over sequential and non-overlapping intervals of length equal to T_c. Hence, it is obvious, that the sum of integrals over the subintervals, is equivalent to the integral over the total interval, which is given at the right part of equality at the spot "!". Finally, note that $N\,T_c + T_c/2 \gg T_c$ in general, where T_c denotes the range over which the product $p_c(t+\tau)\, p_c(t)$ has non-zero values, on the positive axis.

Based on our preceding analysis, the PSD of $s(t)$ is given in Eq. 3.32.

$$S_s(f) = F.T.\left\{\overline{R}_s(\tau)\right\} = \frac{1}{T_c}\, P_c(f)\, P_c'(f)\, dt = \frac{1}{T_c}\,|P_c(f)|^2 \tag{3.32}$$

where $P_c(f) = \mathcal{F}\{p_c(t)\}$ and $X'(f)$ denotes the complex conjugate of $X(f)$. Eq. 3.32 indicates that the shape of the transmitting signal PSD is fully determined by the using pulse $p_c(t)$ and its Fourier Transform.

3.3.2 BPSK

For the BPSK modulation technique, we set the general form of the transmitting signal in Eq. 3.27. To calculate the PSD, we define the time axis in a convinient way, as follows

$$s(t) = \sqrt{2} \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} b_i \, p(t - iT_b) \, \cos(2\pi F_c t) \tag{3.33}$$

where $p(t)$ denotes the using pulse, $2\,\tilde{N}_b + 1 = N_b$ which is the number of information bits and $b_i \in \{\pm 1\},\ \forall i \in [-\tilde{N}_b,\ \tilde{N}_b]$. The statistics of b, are:

$$\begin{aligned} \mu_b &= E\{b\} = 0 \\ \mathrm{var}\{b\} &= E\{(b - E\{b\})^2\} = \sigma_b^2 = 1. \end{aligned}$$

An equivalent expression of the modulated transmitting signal is the following

$$s(t) = \sqrt{2} s_b(t) \, \cos(2\pi F_c t) \tag{3.34}$$

where the baseband signal is defined as

$$s_b(t) = \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} b_i \, p(t - iT_b). \tag{3.35}$$

Firstly, we are focused on the calculation of the PSD of the baseband signal. Its mean value is:

$$E\{s_b(t)\} = E\left\{ \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} b_i \, p(t - iT_b) \right\} = \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} E\{b_i\} \, p(t - iT_b) = 0.$$

The auto-correlation function $R_{s_b s_b}(t + \tau, t)$, is:

$$\begin{aligned} R_{s_b s_b}(t + \tau, t) &= E\{s_b(t + \tau) \, s_b(t)\} \\ &= E\left\{ \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} b_i \, p(t + \tau - iT_b) \sum_{k=-\tilde{N}_b}^{\tilde{N}_b} b_k \, p(t - kT_b) \right\} \\ &= \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} \sum_{k=-\tilde{N}_b}^{\tilde{N}_b} E\left\{b_i \, b_k\right\} p(t + \tau - iT_b) \, p(t - kT_b) \\ &\overset{\star}{=} \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} E\left\{b_i^2\right\} p(t + \tau - iT_b) \, p(t - iT_b) \\ &= \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} p(t + \tau - iT_b) \, p(t - iT_b) \end{aligned}$$

where at the forth equality ("$\star$") we used that the double sum is eliminated, because only the products for $k = i$ survive. In addition, the last expression of the auto-correlation

function is familiar to us. Recalling the last form of the DS auto-correlation function in Section 3.3.1, noted by "$**$", we observe that both expressions are identical by replacing T_c with T_b, N with N_b and $p_c(t)$ with $p(t)$. Hence, the baseband's PSD is given in an analog way, in Eq. 3.36.

$$S_{s_b}(f) = F.T.\left\{\overline{R}_{s_b}(\tau)\right\} = \frac{1}{T_b}\,P(f)\,P^{'}(f)\,dt = \frac{1}{T_b}\,|P(f)|^2 \tag{3.36}$$

The relationship between the spectrum of the baseband and the modulated signal is quite simple. The auto-correlation function of the modulated signal, is calculated as follows.

$$\begin{aligned}
R_{ss}(t+\tau,t) &= E\{s(t+\tau)\,s(t)\} \\
&= E\{\sqrt{2}\,s_b(t+\tau)\,\cos(2\pi F_c[t+\tau])\,\sqrt{2}\,s_b(t)\,\cos(2\pi F_c t)\,\} \\
&= 2\,E\{s_b(t+\tau)\,s_b(t)\}\cos(2\pi F_c[t+\tau])\,\cos(2\pi F_c t) \\
&= 2\,R_{s_b s_b}(t+\tau,t)\,\cos(2\pi F_c[t+\tau])\,\cos(2\pi F_c t) \\
&= 2\,R_{s_b s_b}(t+\tau,t)\,\frac{1}{2}\,[\cos(2\pi F_c[t+\tau]-t)+\cos(2\pi F_c[t+\tau]+t)] \\
&= R_{s_b s_b}(t+\tau,t)\,[\cos(2\pi F_c\tau)+\cos(2\pi F_c[2\,t+\tau])]
\end{aligned}$$

At this point, we use equation 8.2.22 of [12] and we get that

$$\overline{R}_s(\tau) = \overline{R}_{s_b}(\tau)\cos(2\pi F_c\tau).$$

The PSD of the passband signal is, finally, given by the Fourier Transform of $\overline{R}_s(\tau)$, as

$$S_s(f) = \frac{1}{2}\,[S_{s_b}(f-F_c)+S_{s_b}(f+F_c)] = \frac{1}{2\,T_b}\left[|P(f-F_c)|^2+|P(f+F_c)|^2\right] \tag{3.37}$$

using Eq. 3.36.

4. System Model

Fig. 4.0.1 shows the different processing steps for the pilot injection and analysis. At the transmitter, the captured audio signal m is mixed with the pilot signal s. The exact time for the "injection" is not relevant. The mixed signal is potentially encoded and transmitted to the receiver. We make no assumptions about the delivery channel. While a typical path would be packet-based digital distribution (e.g., over the Internet), the proposed scheme works for both analog as well as non-packetized communication. At the receiver, the mixed signal is optionally decoded. If we do not want to measure delays associated with receiver processing such as play-out buffer management and operating system delays, the delay analysis can be performed at this stage ("channel delay estimation"). If, however, we want to include such delays in the measurement, then the endpoint will have to play the received audio, and we will need another device to capture the played back, mixed audio signal. The analysis of this signal will indicate the precise position where the pilot was injected ("total delay estimation"). Note that this position does not necessarily relate to any particular clock, except that of the device where the analysis is performed.

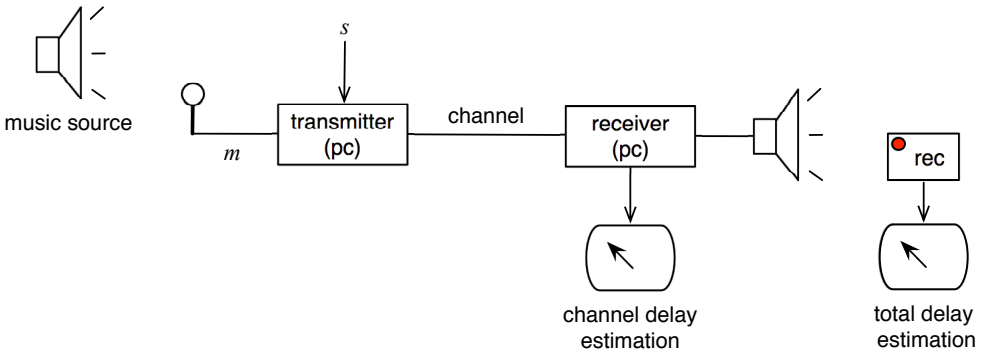

Figure 4.0.1: System diagram.

In order to measure the delay, we need to obtain a time reference. Fig. 4.0.2 depicts a diagram of an example delay measurement configuration such as the one that would be used in laboratory-based measurements. The pilot signal is transmitted mixed with the audio/music signal. Note that for the user at the receiver side, the pilot is the noise that may deteriorate his musical experience. In contrast, from the system aspect, the music signal is treated as "noise". The mixed signal is transmitted over a delay-prone path to a receiver. In order to create a reference, we can assume that the mixed signal without delay is also made available to the receiver. This can be accomplished in a laboratory setup by having a direct, analog or digital audio path between the sender and the receiver. The two signals - with, and without delay - can be captured as a stereo signal at the receiver. Note that the capture may be performed in the analog domain if we wish to include in the delay measurement the delays associated with receiver processing (operating system, play-out buffer management, D/A conversion).

In the Sections below, operations that are performed on the transmitter and the receiver, along with our system assumptions, are going to be described. The exact transmitted waveform, the received signal, the operation to estimate channel's delay and the definition of the SNR of the system, will be analyzed.

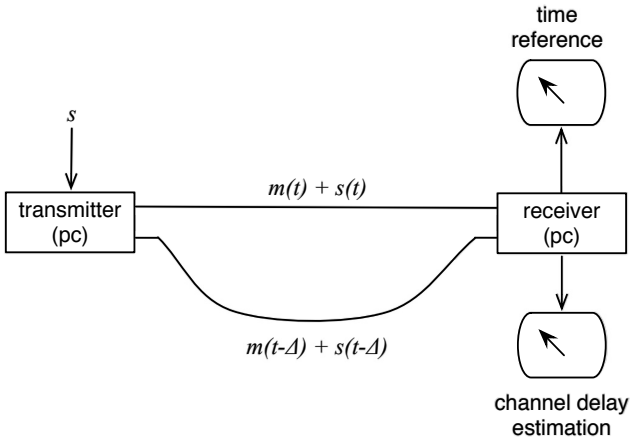

Figure 4.0.2: Example configuration for laboratory-based delay measurement.

4.1 The trasmitted signal

The analysis of Chapter 3 is based on two potential schemes for the construction of the pilot signal. In this Section, we will present, in brief, a typical implementation of the transmitting signal and the corresponding PSD, for each considering system setup.

Pulse train implementation:

The using pulse in this case, is the 15th derivative of the Gaussian pulse and its shape is given in Fig. 3.1.10. We set the time duration of the pilot equal to 0.5 sec and, hence, approximately 1667 pulses are placed sequentially. Energy per pulse equals $4 \cdot 10^{-13}$. The pilot construction is based on the DS technique and, hence, multiple-access capability is achieved by setting to each user a specified spreading code. The length of the code is a variable that can be determined at will, for each implementation. As a result, the number of transmitted information bits is not constant and depends on the exact pilot duration and the length of the spreading code. The formula that describes the transmitting waveform is given in Eq. 3.26 and the corresponding expression for its PSD is given in Eq. 3.32.

In Fig. 4.1.1, the pilot signal in specified time axis and its PSD are shown. The spreading code consists of 16 chips and the starting 7 are given in this figure. Specifically, the sequence $[1, -1, -1, -1, 1, -1, 1]$ is shown. From the shape of the PSD, it is obvious that the desirable effect of locating the signal spectrum in high frequencies, is achieved.

Single-carrier implementation:

In the second considering scheme we use a carrier wave to transfer the spectrum of our band-limited pulse to the desirable frequencies. The SRRC pulse was chosen for this purpose and its shape is given in Fig. 3.1.11. We set the time duration of the pilot equal to 0.5 sec, too, and approximately 1428 pulses are transmitted. Energy per pulse equals $5 \cdot 10^{-9}$. The BPSK modulation technique is selected, since we are not interested on the data rate. Hence, the number of information bits equals the number of pulses. We assume that different sequences of bits are generated for the various users that potentially use the system simultaneously. In this way, orthogonality is achieved, since the inner

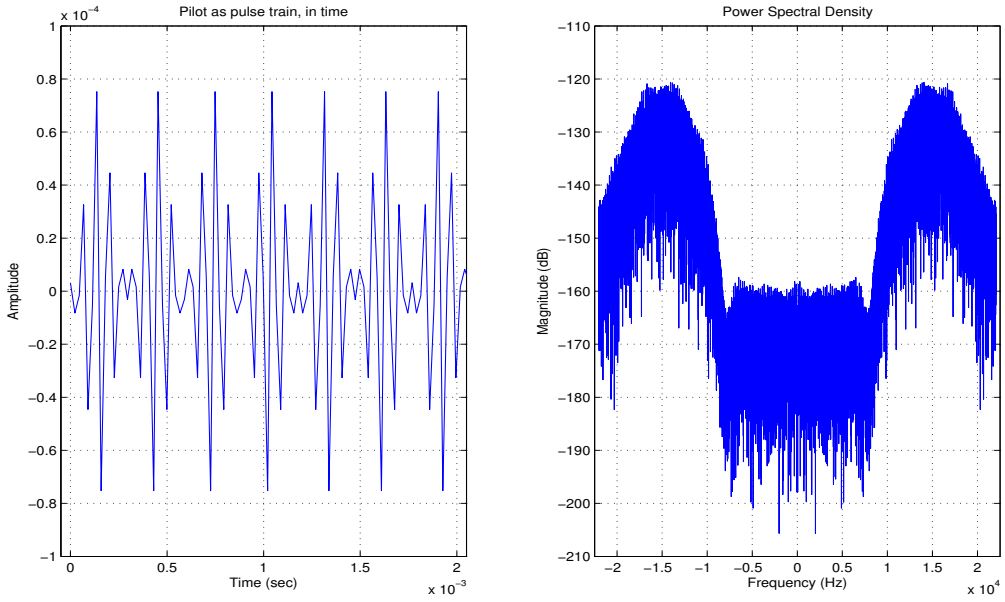

Figure 4.1.1: Pilot signal, in time and frequency domain.

product between each pair of users bit sequences is zero. The formula that describes the transmitting waveform is given in Eq. 3.27 and the corresponding expressions for the PSD of the baseband and the passband signal are given in Eq. 3.36 and 3.37, respectively.

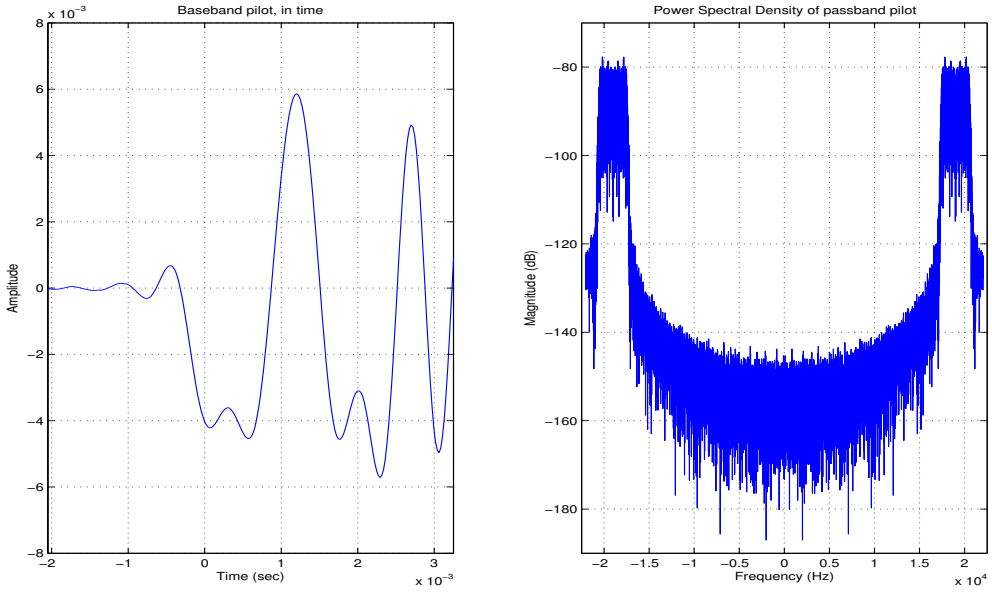

Figure 4.1.2: Pilot signal, in time and frequency domain.

In Fig. 4.1.2, the baseband signal in specified time axis and its passband PSD, are shown. Center frequency $F_c = 18.997$ kHz is selected and the transfer of signal's spectrum at the corresponding band, is obvious. The bandwidth of the modulated pilot equals 2.85 kHz, according to our definitions. The information bits sequence, for this figure, is $[-1, -1,$

$-1, 1, 1, -1, -1, -1, 1, -1]$. Note that the baseband signal will be probably audible, in contrast to the passband. For more details regarding the specified parameters, for both systems, please refer to Section 3.1.2.

4.2 The signal at the receiver

In realistic implementations of our system, the received signal is a delayed version of the transmitted one, caused by processing of real-time systems and/or net. Hence, we assume that the channel of our system inserts just a delay and, in general, the transmitted waveform is multiplied by a scalar factor $a \neq 0$. Furthermore, the pilot is added during music transmission along with the thermal noise which is modeled as AWGN (through cables, digital equipment etc). Under these circumstances, we can assume that a general diagram of our system is given in Fig. 4.2.1, and the formula that describes the received signal is given in Eq. 4.1

$$r(t) = a\, s(t - \Delta) + a\, m(t - \Delta) + n(t) \tag{4.1}$$

where a is a scalar, $n(t)$ and $m(t)$ denote AWGN and music, correspondingly.

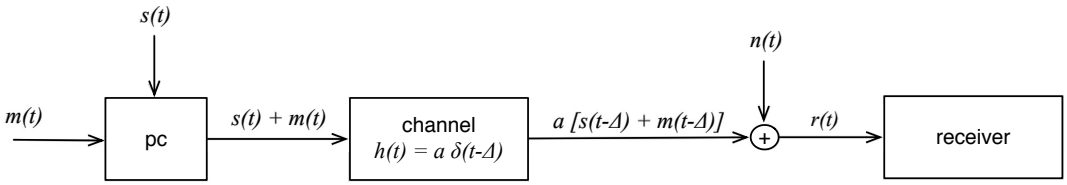

Figure 4.2.1: General system diagram

4.3 Channel estimation

The pilot signal at the receiver is a scaled and delayed version of the transmitted one. In such cases, cross-correlation can be performed, and it is considered as one of the simplest approaches for the delay estimation of the channel. In addition, for both cases of signal construction, a correlator receiver can be applied. As a result, cross-correlation technique is chosen for our purpose. Mathematical calculations in order to estimate the signal and the noise products after the filter at the receiver, will be performed.

Let the transmitted signal be $s(t)$ and the received signal be $r(t) = s(t - \Delta) + m(t - \Delta) + n(t)$, assuming $a = 1$, $n(t)$ AWGN that follows $N(0, \sigma_n^2)$ and $m(t)$ denotes music. The channel, shifts the transmitted waveforms by Δ sec. At the receiver, we assume that the pilot signal is known and, hence, cross-correlation with a determined waveform is performed in order to calculate the delay of the channel.

In Fig. 4.3.1, a general diagram that depicts receiver's various stages, is shown. A simplified cross-correlation product of $R_{rs}(t + \tau, t)$ is included, assuming that the signal dominates against noise.

$$r(\tau) \star s(\tau) = R_{rs}(t + \tau, t) = \int_{-\infty}^{+\infty} r'(t + \tau)\, s(t)\, \mathrm{d}t = \int_{-\infty}^{+\infty} r(t + \tau)\, s(t)\, \mathrm{d}t$$

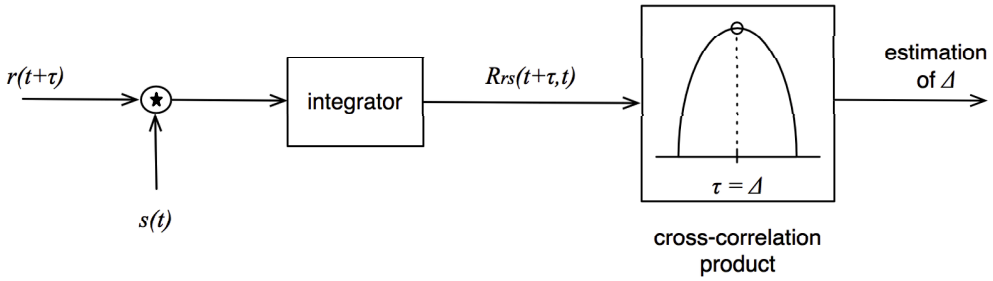

Figure 4.3.1: Receiver diagram

where $\star$ denotes the operation of cross-correlation and $x'(t)$ the complex conjugate of $x(t)$. Note that, in our case, $r(t)$ is a real signal and, thus, the last equality holds.

For both modulation techniques, Fig. 4.3.1 is precise with the operations that are performed at the receiver. However, assuming BPSK, the stage of transferring the spectrum from the passband back to the baseband, is not included. Specifically, before the cross-correlation process, we assume that demodulation has been performed and the notation $r(t)$ in Fig. 4.3.1 is replaced by the following term $r_{\text{dem}}(t) = r(t)\sqrt{2}\cos(2\pi F_c t + \phi)$. Afterwards, the exact process will be described. In this study, we assume coherent demodulation, i.e., the phase is known at the receiver. In general, the carrier recovery could be achieved by a phase-lock loop (PLL). Analyzing $R_{rs}(t+\tau, t)$, we get

$$\begin{aligned} R_{rs}(t+\tau,t) &= \int_{-\infty}^{+\infty} [s(t+\tau-\Delta) + m(t+\tau-\Delta) + n(t+\tau)]\, s(t)\,\mathrm{d}t \\ &= \int_{-\infty}^{+\infty} s(t+\tau-\Delta)s(t)\,\mathrm{d}t + \int_{-\infty}^{+\infty} m(t+\tau-\Delta)s(t)\,\mathrm{d}t + \int_{-\infty}^{+\infty} n(t+\tau)s(t)\,\mathrm{d}t. \end{aligned}$$

Last expression, indicates the products of signal and both type of noises, occurring after cross-correlation. Equivalently, we define

$$R_{rs}(t+\tau,t) = R_{ss}(t+\tau-\Delta,t) + R_{ms}(t+\tau-\Delta,t) + R_{ns}(t+\tau,t) = R_s(\tau-\Delta) + y(\tau-\Delta) + w(\tau). \tag{4.2}$$

4.3.1 Signal

We are focused on the signal product of Eq. 4.2, $R_{ss}(t+\tau-\Delta,t)$, which denotes the auto-correlation function of the pilot. The transmitted signal, for each user, is known at the receiver and, hence, it is assumed as a deterministic waveform.

Pulse train implementation:

Assuming the DS technique, we use Eq. 3.31 to get the formula of the transmitted signal. Then, the signal product at the output of cross-correlation is analyzed as follows. Note that after some calculations, $R_{ss}(t+\tau-\Delta,t)$ becomes a function of $\tau-\Delta$ and, hence, for space saving reasons the valid abbreviation $R_s(\tau-\Delta)$ is used.

$$
\begin{aligned}
R_s(\tau-\Delta) &= \int_{-\infty}^{+\infty} s(t+\tau-\Delta)\, s(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{+\infty} \sum_{i=-\tilde{N}_b}^{\tilde{N}_b}\sum_{k=-\tilde{N}_b}^{\tilde{N}_b} b_i d b_k \sum_{j=-\tilde{N}_c}^{\tilde{N}_c}\sum_{n=-\tilde{N}_c}^{\tilde{N}_c} c_j c_n p_c(t+\tau-\Delta-iT_b-jT_c)p_c(t-kT_b-nT_c)\mathrm{d}t \\
&= \sum_{i=-\tilde{N}_b}^{\tilde{N}_b}\sum_{k=-\tilde{N}_b}^{\tilde{N}_b} b_i b_k \sum_{j=-\tilde{N}_c}^{\tilde{N}_c}\sum_{n=-\tilde{N}_c}^{\tilde{N}_c} c_j c_n \int_{-\infty}^{+\infty} p_c(t+\tau-\Delta-iT_b-jT_c)p_c(t-kT_b-nT_c)\mathrm{d}t \\
&\overset{*}{=} \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} b_i^2 \sum_{j=-\tilde{N}_c}^{\tilde{N}_c} c_j^2 \int_{-\infty}^{+\infty} p_c(t+\tau-\Delta-iT_b-jT_c)\, p_c(t-iT_b-jT_c)\,\mathrm{d}t \\
&= \sum_{i=-\tilde{N}_b}^{\tilde{N}_b}\sum_{j=-\tilde{N}_c}^{\tilde{N}_c} \int_{-\infty}^{+\infty} p_c(t+\tau-\Delta-iT_b-jT_c)\, p_c(t-iT_b-jT_c)\,\mathrm{d}t \\
&= (2\,\tilde{N}_c+1)\sum_{i=-\tilde{N}_b}^{\tilde{N}_b} \int_{-\infty}^{+\infty} p_c(t+\tau-\Delta-iT_b)\, p_c(t-iT_b)\,\mathrm{d}t \\
&= (2\,\tilde{N}_b+1)\,(2\,\tilde{N}_c+1)\int_{-\infty}^{+\infty} p_c(t+\tau-\Delta)\, p_c(t)\,\mathrm{d}t \\
&= N_b\,N_c \int_{-\infty}^{+\infty} p_c(t+\tau-\Delta)\, p_c(t)\,\mathrm{d}t \\
&= N_b\,N_c\,R_{p_c}(\tau-\Delta)
\end{aligned}
$$

where we used that $\int_{-\infty}^{+\infty} p_c(t+\tau-\Delta-iT_b-jT_c)\, p_c(t-kT_b-nT_c)\,\mathrm{d}t = 0$ in general, when $k \neq i$ and $n \neq j$. In other words, only the terms for $k=i$ and $n=j$ do survive, at the corresponding double sums. Finally, we recall that $(2\,\tilde{N}_b+1) \equiv N_b$ and $(2\,\tilde{N}_c+1) \equiv N_c$.

Single-carrier implementation:

In this case, firstly, we apply the demodulation process to $s(t)$ given by Eq. 3.34 and yielding $s_{\mathrm{dem}}(t)$. Secondly, we perform the cross-correlation operation between the latter product and the baseband signal $s_b(t)$.

$$
\begin{aligned}
s_{\mathrm{dem}}(t) &= \sqrt{2}\; s(t)\; \cos(2\pi F_c t+\phi) \\
&= 2\; s_b(t)\; \cos(2\pi F_c t+\phi)\cos(2\pi F_c t+\phi) \\
&= 2\; s_b(t)\; \frac{1}{2}[\cos(4\pi F_c t+2\phi)+\cos(0)]
\end{aligned}
$$

$$= s_b(t) + s_b(t)\cos(4\pi F_c t + 2\phi)$$

The spectrum of the last term in the previous expression, is located in a band around $2\,F_c$ and, hence, we expect to be eliminated after the cross-correlation with the band-limited signal that follows. Thus, it is not considered in our calculations below.

We use Eq. 3.35 to get the expression of the baseband signal. The cross-correlation between $s_{\mathrm{dem}}(t)$ and $s_b(t)$ follows, which are both baseband signals. Similarly, $R_{s_b}(\tau-\Delta)$ is used instead of $R_{s_b s_b}(t+\tau-\Delta, t)$.

$$\begin{aligned}
R_{s_b}(\tau-\Delta) &= \int_{-\infty}^{+\infty} s_{\mathrm{dem}}(t+\tau-\Delta)\, s_b(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{+\infty} s_b(t+\tau-\Delta)\, s_b(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{+\infty} \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} b_i\, p(t+\tau-\Delta-iT_b) \sum_{k=-\tilde{N}_b}^{\tilde{N}_b} b_k\, p(t-kT_b)\,\mathrm{d}t \\
&= \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} \sum_{k=-\tilde{N}_b}^{\tilde{N}_b} b_i\, b_k \int_{-\infty}^{+\infty} p(t+\tau-\Delta-iT_b)\, p(t-kT_b)\,\mathrm{d}t \\
&\overset{\star}{=} \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} b_i^2 \int_{-\infty}^{+\infty} p(t+\tau-\Delta-iT_b)\, p(t-iT_b)\,\mathrm{d}t \\
&= \sum_{i=-\tilde{N}_b}^{\tilde{N}_b} \int_{-\infty}^{+\infty} p(t+\tau-\Delta-iT_b)\, p(t-iT_b)\,\mathrm{d}t \\
&= N_b \int_{-\infty}^{+\infty} p(t+\tau-\Delta)\, p(t)\,\mathrm{d}t \\
&= N_b\, R_p(\tau-\Delta)
\end{aligned}$$

where we used that $\int_{-\infty}^{+\infty} p(t+\tau-\Delta-iT_b)\, p(t-kT_b)\,\mathrm{d}t = 0$, in general, when $k \neq i$.

In both cases, the terms $R_{p_c}(\tau-\Delta),\ R_p(\tau-\Delta)$ are the auto-correlation functions of the corresponding used pulses. It is well-known that an auto-correlation operation will always have a peak at a lag of zero unless the signal is a trivial zero signal[1]. Hence, at the correct time instant $\tau = \Delta$, we obtain the maximum value of the pulse and, by extension, of the signal's auto-correlation function. This peak value, using DS equals $N_b\, N_c\, E_p \equiv N_p\, E_p \equiv E_s$, where E_p is the energy per pulse, N_p the total number of transmitted pulses and E_s the energy of the pilot, respectively. Using BPSK technique, equals $N_b\, E_p \equiv E_s$, where N_b is the number of information bits, which equals the total number of transmitted pulses. Recalling the construction assumptions we made at Sections 3.1.2.1 and 3.1.2.2, each pulse has a specified energy E_p and, according to 4.1, each pilot contains a constant number of pulses, for each implementation. Hence, all these variables are known and the peak value equals the energy of the corresponding signal, in any case.

[1] Recalling the well-known property for the auto-correlation of a signal $x(t)$: $R_x(0) \geq |R_x(\tau)|,\ \forall\tau$

4.3.2 Noise

In this section, we are focused on the other parts of Eq. 4.2 concerning the products caused by music and noise. It is well-known that there is no mathematical model that can provide a concise and general stochastic description of music and, hence, the product $R_{ms}(t+\tau-\Delta,t) \equiv y(\tau-\Delta)$ can't be further analyzed. Concerning the product of AWGN, $R_{ns}(t+\tau,t) \equiv w(\tau)$, its statistics will be calculated.

At first, more convenient calculations arise if we perform convolution instead of cross-correlation, in an equivalent way. In general, the following relation is valid between the two operations:

$$x(t) \star y(t) = x(t) * y(-t) \tag{4.3}$$

where the signs "$\star$" and "$*$" denote cross-correlation and convolution, respectively. Thus, the noise product can be expressed as:

$$w(\tau) = \int_{-\infty}^{+\infty} n(t+\tau)\, s(t)\, \mathrm{d}t = n(\tau) \star s(\tau) = n(\tau) * s(-\tau).$$

Noise term $n(\tau)$ follows $N(0,\, \sigma_n^2)$ distribution and it is a WSS process with auto-correlation function $R_n(\tau) = \sigma_n^2\, \delta(\tau)$. Its PSD is flat and equal to N_o, spread over the whole available bandwidth. Note that according to the system assumptions, the transmitted audio waveforms are, in general, digital signals that resulted by an A/D conversion, with a sampling frequency F_s. Consequently, the available bandwidth is defined over the analog frequencies $[-F_s/2,\, F_s/2]$. The integral over the PSD of a signal equals its power and, thus, $\int_{-F_s/2}^{F_s/2} N_o\, \mathrm{d}f = N_o\, F_s = \sigma_n^2$. Hence, the AWGN variance is different from its PSD value. If we consider the frequency axis of discrete signals, i.e., $[0,\, 1]$, observe that $N_o = \sigma_n^2$. Thus, in our case the following equality holds.

$$N_o = \frac{\sigma_n^2}{F_s} \tag{4.4}$$

Coming back to the estimation of $w(\tau)$'s statistics, different calculations should be performed for each case of the pilot construction.

Pulse train implementation:

Firstly, we estimate the mean value of $w(\tau)$, assuming that our pilot adopts the DS technique.

$$\mu_w = E\left\{\int_{-\infty}^{+\infty} n(t+\tau)\, s(t)\, \mathrm{d}t\right\} = \int_{-\infty}^{+\infty} E\left\{n(t+\tau)\right\}\, s(t)\, \mathrm{d}t = 0$$

because $n(t)$ and $s(t)$ are independent.

Secondly, we calculate its auto-correlation function. For this purpose, we use equations 4.2.9 and 4.2.10 of [12], slightly altered[2], since in our case cross-correlation instead of convolution is performed. Then, we get

[2]The exact calculations to obtain the equations that we use, are performed in Section 8 of the Appendix.

$$R_{xy}(\tau) = R_x(\tau) * h(\tau) \tag{4.5}$$

$$R_y(\tau) = R_x(\tau) * h(\tau) * h(-\tau) \tag{4.6}$$

where $x(t)$ is the WSS input, $h(t)$ the Linear Time Invariant (LTI) filter and $y(t)$ is the output which is WSS, too. According to Eq. 4.6, we have

$$\begin{aligned} R_w(\tau) &= R_n(\tau) * s(\tau) * s(-\tau) = \sigma_n^2\, s(\tau) * s(-\tau) \overset{4.3}{=} \sigma_n^2\, s(\tau) \star s(\tau) = \sigma_n^2\, R_s(\tau) \\ &= \sigma_n^2\, N_b\, N_c\, R_{p_c}(\tau) \end{aligned}$$

where $s(t)$ is determined at the receiver and, hence, it is assumed as an LTI filter, and the last equality holds under the derivations of Section 4.3.1.

Single-carrier implementation:

In this case, firstly, the demodulation process is applied and, secondly, the cross - correlation operation between the latter product and the baseband signal $s_b(t)$. After demodulation the noise term is:

$$n_{\text{dem}}(t) = \sqrt{2}\, n(t)\, \cos(2\pi F_c t + \phi).$$

The statistics of $n_{\text{dem}}(t)$, are required. Its mean value is:

$$E\{n_{\text{dem}}\} = E\left\{\sqrt{2}\, n(t)\, \cos(2\pi F_c t + \phi)\right\} = \sqrt{2}\, E\{n(t)\}\; E\{\cos(2\pi F_c t + \phi)\} = 0$$

and its auto-correlation function is calculated as follows

$$\begin{aligned} R_{n_{\text{dem}} n_{\text{dem}}}(t+\tau, t) &= E\{n_{\text{dem}}(t+\tau)\, n_{\text{dem}}(t)\} \\ &= E\left\{\sqrt{2}\, n(t+\tau)\, \cos(2\pi F_c[t+\tau] + \phi)\, \sqrt{2}\, n(t)\, \cos(2\pi F_c t + \phi)\right\} \\ &= 2\, E\{n(t+\tau)\, n(t)\}\; E\{\cos(2\pi F_c[t+\tau] + \phi)\, \cos(2\pi F_c t + \phi)\} \\ &= 2\, R_{nn}(\tau)\, E\left\{\frac{1}{2}[\cos(2\pi F_c[2t+\tau] + 2\phi) + \cos(2\pi F_c \tau)]\right\} \\ &= R_{nn}(\tau)\, \cos(2\pi F_c \tau) \\ &= \sigma_n^2\, \delta(\tau)\, \cos(2\pi F_c \tau) \\ &= \sigma_n^2\, \delta(\tau) \end{aligned}$$

where ϕ is considered as a random variable which follows a uniform distribution over $[0,\, 2\pi]$ and, hence, $E\{\cos(2\pi F_c[2t+\tau] + 2\phi)\} = 0$. Through previous calculations, it is proven that $n_{\text{dem}}(t)$ is a WSS process and, thus, $R_{n_{\text{dem}} n_{\text{dem}}}(t+\tau, t) = R_{n_{\text{dem}}}(\tau)$.

Until now, we estimated the statistics of the demodulation product, because they will be used below. After the demodulation, the cross-correlation operation follows.

$$w(\tau) = \int_{-\infty}^{+\infty} n_{\text{dem}}(t+\tau)\, s_b(t)\, \mathrm{d}t$$

$$= \sqrt{2} \int_{-\infty}^{+\infty} n(t+\tau)\, \cos(2\pi F_c[t+\tau] + \phi)\, s_b(t)\, \mathrm{d}t$$

Firstly, we get the mean value of $w(\tau)$.

$$\mu_w = \sqrt{2} \int_{-\infty}^{+\infty} E\left\{n(t+\tau)\right\}\, E\left\{\cos(2\pi F_c[t+\tau] + \phi)\right\}\, s_b(t)\, \mathrm{d}t = 0$$

because $n(t)$, $s_b(t)$ and $\cos(2\pi F_c t + \phi)$ are independent. Secondly, we estimate its auto-correlation function, using Eq. 4.6, in coordinance to our system parameters.

$$\begin{aligned} R_w(\tau) &= R_{n_{\mathrm{dem}}}(\tau) * s_b(\tau) * s_b(-\tau) = \sigma_n^2\, s_b(\tau) * s_b(-\tau) = \sigma_n^2\, s(\tau) \star s(\tau) = \sigma_n^2\, R_{s_b}(\tau) \\ &= \sigma_n^2\, N_b\, R_p(\tau) \end{aligned}$$

For both cases of the signal construction, we have identical expressions for the auto-correlation functions of the noise product at the output of the filter. In brief, we get

$$R_w(\tau) = \begin{cases} \sigma_n^2\, N_b\, N_c\, R_{p_c}(\tau), & \text{for DS} \\ \\ \sigma_n^2\, N_b\, R_p(\tau), & \text{for BPSK} \end{cases}$$

which is a function of input noise power, total number of pulses and the auto-correlation of the used pulse, in each case. In addition, it is obvious that $w(\tau)$ is not AWGN. Specifically, it is a colored WSS process (since $R_w(\tau) \neq A\,\delta(\tau)$) and the "order of dependency" is relevant to the corresponding width of the using pulse. Furthermore, the output of an LTI filter assuming a WSS process as an input, is a WSS process too.

Finally, we want to estimate the power of $w(\tau)$. In general, the PSD of a WSS process at the output of an LTI filter, $h(t)$, is given as

$$S_O(f) = S_I(f)\, |\mathcal{F}\{h(t)\}|^2$$

where $S_O(f)$ and $S_I(f)$ is the PSD of the output and the input, respectively. In our case, the PSD of the input is constant and equals N_o and the filter $h(\tau)$ is known. Thus,

$$\sigma_w^2 = \int_{-\infty}^{+\infty} N_o\, |\mathcal{F}\{h(t)\}|^2\, \mathrm{d}f = N_o \int_{-\infty}^{+\infty} |\mathcal{F}\{h(t)\}|^2\, \mathrm{d}f = N_o \int_{-\infty}^{+\infty} |h(t)|^2\, \mathrm{d}t = N_o\, E_s$$

using Parseval's theorem. In the last equality E_s denotes the energy of the pilot and is known for both cases. For DS, $h(t) = s(t)$, given by Eq. 3.31. For BPSK, $h(t) = s_b(t)$, given by Eq. 3.35. Combining the latter expression of σ_w^2 with Eq. 4.4, we get

$$\sigma_w^2 = N_o\, E_s = \frac{\sigma_n^2\, E_s}{F_s}. \tag{4.7}$$

4.4 SNR of the system

In this section, the definition and noteworthy remarks concerning the SNR of our system are discussed. Based on bibliography, we set

$$\mathrm{SNR} = \frac{\sigma_s^2}{\sigma_n^2 + \sigma_m^2} = \frac{P_s}{P_n + P_m} = \frac{\int_{-\infty}^{+\infty} S_s(f)\,\mathrm{d}f}{\int_{-\infty}^{+\infty} S_n(f)\,\mathrm{d}f + \int_{-\infty}^{+\infty} S_m(f)\,\mathrm{d}f}$$

where σ_s^2, σ_n^2, σ_m^2 are the variances, P_s, P_n, P_m the powers and $S_s(f)$, $S_n(f)$, $S_m(f)$ the PSDs of the signal, noise and music correspondingly. In Fig. 4.4.1 the system diagram is depicted and the spot where the SNR is defined, is highlighted.

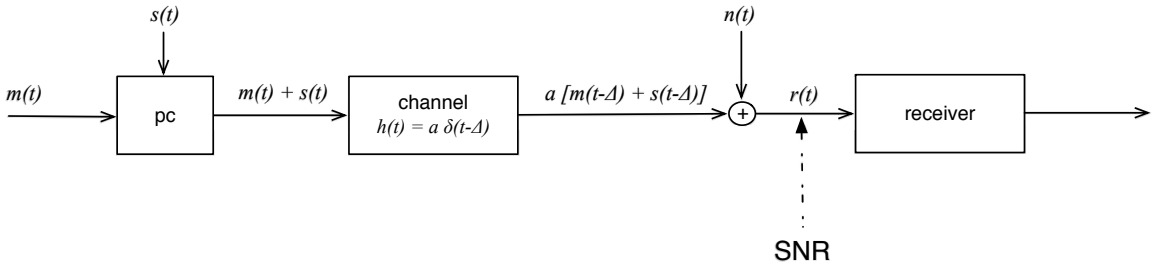

Figure 4.4.1: SNR definition

Note that, in our application, both the actual noise and the audio/music signal are considered noise, since the actual information we are interested in is the pilot signal. The power of each of these signals can be calculated as the integral over the corresponding PSD. Given that the power of the pilot signal is known, the evaluation of interfering music track's power can be calculated directly by setting the SNR's value.

However, the system performance can be theoretically approached assuming just the presence of AWGN, since there is no consistent mathematical modeling for music effects. Below, we define a special case, in which we don't take into account the presence of music and we assume the existence only of AWGN in our system, which appears in the form of thermal noise (i.e., through cables, digital equipment, etc.).

4.4.1 Input SNR

Assuming, simply, the transmission of our pilot in an AWGN channel, the simplified system diagram is shown in Fig. 4.4.2 and the SNR ratio becomes

$$\mathrm{SNR} = \frac{\sigma_s^2}{\sigma_n^2} = \frac{P_s}{P_n} = \frac{\int_{-\infty}^{+\infty} S_{ss}(f)\,\mathrm{d}f}{\int_{-\infty}^{+\infty} S_{nn}(f)\,\mathrm{d}f} \tag{4.8}$$

Note that, assuming just the presence of AWGN in the channel, the cross-correlation output of Eq. 4.2, is given as follows.

$$R_{rs}(t+\tau, t) = R_{ss}(t+\tau-\Delta, t) \; + \; R_{ns}(t+\tau, t) = R_s(\tau-\Delta) \; + \; w(\tau). \tag{4.9}$$

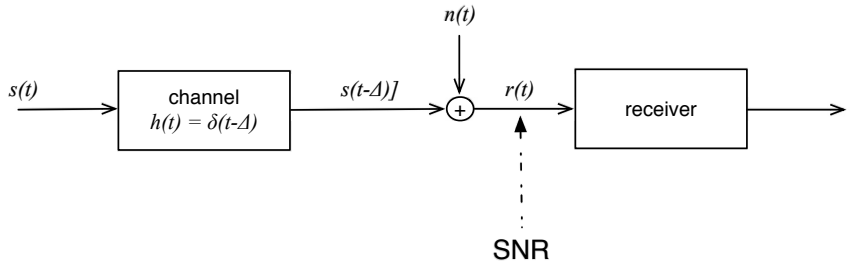

Figure 4.4.2: Simplified system diagram assuming SNR at the input.

4.4.2 Output SNR

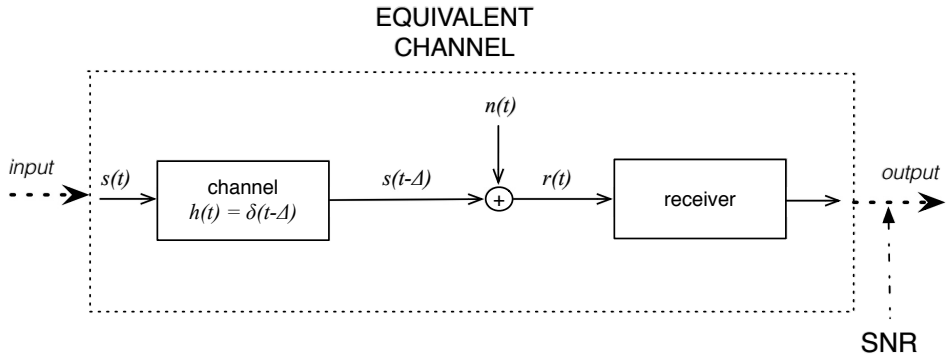

Figure 4.4.3: Simplified system diagram assuming SNR at the output.

In typical digital communication systems, it is a common method to define the SNR at the output of the system, replacing the whole processing at the receiver by an equivalent channel. In this case, the system diagram is given by Fig. 4.4.3.

The SNR at the output, is defined as the ratio of the square of the signal product at the appropriate time $\tau = \Delta$, divided by the noise product variance. According to Section 4.3.1, $R_s^2(\tau - \Delta)|_{\tau=\Delta} = R_s^2(0) = E_s^2$, and recalling Eq. 4.7, we get the following expression.

$$\mathrm{SNR}_{out} = \frac{E_s^2}{\sigma_n^2\, E_s/F_s} = \frac{F_s\, E_s}{\sigma_n^2} = \frac{E_s}{N_o} \tag{4.10}$$

The ratio of the two SNR definitions (given by Eq. 4.8 and 4.10), is obtained as

$$\frac{\mathrm{SNR}}{\mathrm{SNR}_{out}} = \frac{\sigma_s^2/\sigma_n^2}{F_s\, E_s/\sigma_n^2} = \frac{\sigma_s^2}{F_s\, E_s} = \frac{E_s/T}{F_s\, E_s} = \frac{1}{T\, F_s}$$

The difference in dB, is given in Eq. 4.11.

$$\mathrm{DSNR}_{\mathrm{dB}} \equiv \mathrm{SNR}_{\mathrm{dB}} - \mathrm{SNR}_{out,\,\mathrm{dB}} = -10\,\log_{10} T - 10\,\log_{10} F_s \tag{4.11}$$

In the last equation, we calculate essentially the SNR gain after filtering at the receiver.

5. Theoretical Analysis of Delay Accuracy

In this Chapter, the theoretical performance of our system will be analyzed. The definition of the SNR before the processing at the receiver is appropriate, considering realistic implementations of our system. As it is mentioned, we can't theoretically predict the effects of music, in general, since they are differentiated per music track and related to the music genre, instrumentation and tempo. The formulas below, could be used directly to estimate the performance when AWGN is dominating, which is considered as a corner case of our system. For the DS technique, a simplified approach is considered, in which dirac pulses are placed in the position of the Gaussian derivatives, to avoid too complex calculations. As it will be seen through simulations, the performance of both systems is quite similar. The scope, here, is to estimate the probability of error for the estimation of the true time delay. Hence, a binary decision is applied and if and only if the estimated delay equals the true delay of the channel, no error occurs. For the BPSK technique, a different approach is considered and confidence intervals are calculated.

5.1 Probability of peak detection using Dirac pulses

In this case, the transmitted signal is given by Eq. 3.31 and the dirac pulse is assumed as the using pulse. Recalling the calculations of Section 4.3.1, and specifically the "Pulse train implementation" subsection, we get that the cross-correlation at the output, regarding the signal product, is given as

$$R_s(\tau - \Delta) = N_b\, N_c\, R_{p_c}(\tau - \Delta) = N_b\, N_c\, \delta(t) \star \delta(t + \tau - \Delta) = N_b\, N_c\, \delta(\tau - \Delta)$$

which is a scaled dirac pulse. Note that, the amplitude of the used pulse should be normalized in order to achieve the specified energy per pulse E_p and, by extension, the signal energy to be $E_s \equiv R_s(0) = N_b\, N_c\, E_p$. Similarly, according to Section 4.3.2, the noise product at the output of the cross-correlation is, in general, colored noise and its variance is given by Eq. 4.7, as $\sigma_w^2 = \sigma_n^2\, E_s/F_s$, where σ_n^2 is the power of the AWGN at the input of the filter. However, in this specific case, the auto-correlation function of the noise at the output is given

$$R_w(\tau) = \sigma_n^2\, N_b\, N_c\, R_{p_c p_c}(\tau) = N_b\, N_c\, \delta(\tau)$$

and, hence, $w(\tau)$ is AWGN.

It is more convenient for our calculations to use samples instead of time intervals in an equivalent way. The noise product at the output, is expressed as a stochastic process $W = (W_{\tau_1}, W_{\tau_2}, \ldots, W_{\tau_N})$. This collection of random variables includes independent and identical distributed (i.i.d.) samples and, thus, $W_{\tau_i} \sim N(0, \sigma_w^2), \quad \forall i \in \{1, N\}$. Similarly, the signal output at the appropriate time instant $\tau = \Delta$ has a constant value $X_\Delta = N_b\, N_c\, \delta(\tau - \Delta)$ and $X_{\tau_i} = 0, \quad \forall i \in \{1, N\}$, with $\tau_i \neq \Delta$. The discrete equivalent modeling of the cross-correlation output, then, is given as

$$Y = X_\Delta + W$$

where Y is a stochastic process, with the following statistics:

$$\begin{aligned} Y_{\tau_i} &\sim N(0, \sigma_w^2), \quad \forall i \in \{1,\, N\}, \text{ with } \tau_i \neq \Delta \\ Y_{\tau_i} &\sim N(E_s, \sigma_w^2), \quad \tau_i = \Delta \end{aligned} \tag{5.1}$$

and all its samples are independent. The probability of the correct estimation of the time delay $\tau = \Delta$, equals the correct estimation of the dirac's peak, which is located at that particular time instant. Given the waveform Y, the equivalent problem we have to address here, is the following

$$P_{\text{corr}} = \bigcap_{\substack{\forall i \in \{1,\, N\} \\ i \neq \Delta}} P(Y_{\tau_i} < Y_\Delta) = P(\underbrace{Y_{\tau_1}, Y_{\tau_2}, \ldots, Y_{\Delta-1}, Y_{\Delta+1}, \ldots, Y_{\tau_N}}_{N-1 \text{ terms}} < Y_\Delta)$$

where analyzing the last expression, we get

$$\begin{aligned} P_{\text{corr}} &= \int_{-\infty}^{+\infty} \underbrace{\int_{-\infty}^{y_\Delta} \cdots \int_{-\infty}^{y_\Delta} \int_{-\infty}^{y_\Delta}}_{N-1} f_{Y_{\tau_1}, Y_{\tau_2}, \cdots, Y_{\tau_N}, Y_\Delta}(y_1, y_2, \cdots, y_N, y_\Delta)\, \mathrm{d}y_1\, \mathrm{d}y_2\, \cdots\, \mathrm{d}y_N\, \mathrm{d}y_\Delta \\ &= \int_{-\infty}^{+\infty} \underbrace{\left[\int_{-\infty}^{y_\Delta} f_{Y_{\tau_1}}(y_1)\, \mathrm{d}y_1 \int_{-\infty}^{y_\Delta} f_{Y_{\tau_2}}(y_2)\, \mathrm{d}y_2 \cdots \int_{-\infty}^{y_\Delta} f_{Y_{\tau_N}}(y_N)\, \mathrm{d}y_N\right]}_{N-1} f_{Y_\Delta}(y_\Delta)\, \mathrm{d}y_\Delta \\ &= \int_{-\infty}^{+\infty} \underbrace{\left[F_{Y_{\tau_1}}(y_\Delta) F_{Y_{\tau_2}}(y_\Delta) \cdots F_{Y_{\tau_N}}(y_\Delta)\right]}_{N-1} f_{Y_\Delta}(y_\Delta)\, \mathrm{d}y_\Delta \\ &= \int_{-\infty}^{+\infty} \left[F_{Y_\tau}(y_\Delta)\right]^{N-1} f_{Y_\Delta}(y_\Delta)\, \mathrm{d}y_\Delta \end{aligned}$$

where $f_Z(z)$ is the probability density function (PDF) and $F_Z(z)$ is the cumulative density function (CDF) of the random variable Z. In the previous equalities it was used that the random variables $Y_{\tau_1}, Y_{\tau_2}, \ldots, Y_{\tau_N}$ and Y_Δ are independent (i.e., $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$), which is valid by definition. According to Eq. 5.1, the quantities included at the probability of correct estimation, are given below:

$$f_{Y_\Delta}(y) = \frac{1}{\sqrt{2\pi\sigma_w^2}}\, \mathrm{e}^{-\frac{(y-E_s)^2}{2\sigma_w^2}}$$

$$F_{Y_\tau}(y) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{y}{\sqrt{2\sigma_w^2}}\right)\right].$$

Finally, the probability of error is given in Eq. 5.2

$$P_e = 1 - \int_{-\infty}^{+\infty} \left[F_{Y_\tau}(y_\Delta)\right]^{N-1} f_{Y_\Delta}(y_\Delta)\, \mathrm{d}y_\Delta \tag{5.2}$$

and it is obvious that the performance of our system is related to the signal energy E_s and the SNR of the system (i.e., combining Eq. 4.7 with 4.8 the variance σ_w^2 becomes straightly a function of the SNR, since the other variables are constant).

5.2 Confidence intervals applied in raised cosine pulses

In this section we calculate the 95% confidence interval for estimating the peak of a raised cosine pulse [3]. According to calculations of Section 4.3.1 and, specifically, the "Single-carrier implementation" subsection, we get that the cross-correlation at the output, regarding the signal product, is given as

$$R_{s_b s_b}(t+\tau, t) \equiv R_{s_b}(\tau) = N_b\, R_p(\tau)$$

by setting $\tau = \tau - \Delta$. Note that, the peak value is achieved for $\tau = 0$, after the change of variables. In addition, $R_p(\tau) = p(t) \star p(t) = p(t) * p(-t) = p(t) * p(t)$ is valid, because the used SRRC pulse, is an even and real function. By definition, the convolution of a SRRC with itself, results to a raised cosine pulse. Hence, $R_p(\tau) = E_p\, p_{rc}(\tau)$, where the $p_{rc}(\tau)$ denotes the raised cosine pulse. According to Eq. 3.9, its time domain representation is given below.

$$p_{rc}(\tau) = \frac{\sin\left(\frac{\pi\tau}{T_p}\right)}{\frac{\pi\tau}{T_p}} \frac{\cos\left(\frac{a\pi\tau}{T_p}\right)}{1 - \left(\frac{2a\tau}{T_p}\right)^2}$$

where T_p is the time duration of the raised cosine pulse and a is the roll-off factor. Assuming that $a = 0$, we get that

$$p_{rc}(\tau) = \frac{\sin\left(\frac{\pi\tau}{T_p}\right)}{\frac{\pi\tau}{T_p}}$$

which will be used for our calculations. We want to define an estimator of $R_{s_b}(\tau) \equiv N_b\, E_p\, p_{rc}(\tau) \equiv E_s\, p_{rc}(\tau)$ for $\tau = 0$. The expansion of $\sin\left(\frac{\pi\tau}{T_p}\right)$ near $\tau = 0$ yields

$$\sin\left(\frac{\pi\tau}{T_p}\right) \simeq \frac{\pi\tau}{T_p} - \frac{1}{6}\left(\frac{\pi\tau}{T_p}\right)^3.$$

Hence, near $\tau = 0$, the estimate $\hat{R}_{s_b}(\tau)$ is given as

$$\hat{R}_{s_b}(\tau) \simeq E_s \frac{T_p}{\pi\tau}\left[\frac{\pi\tau}{T_p} - \frac{1}{6}\left(\frac{\pi\tau}{T_p}\right)^3\right] = E_s\left[1 - \frac{1}{6}\left(\frac{\pi\tau}{T_p}\right)^2\right].$$

The statistics of the estimator, are calculated below. Its mean value is:

$$E\left\{\hat{R}_{s_b}(0)\right\} = E_s = R_{s_b}(0)$$

and, thus, the estimator is unbiased. The variance is estimated below. Note that, for an unbiased estimator, the variance equals to the mean square error.

$$\mathrm{var}\left\{\hat{R}_{s_b}(\tau)\right\} = E\left\{\left(\hat{R}_{s_b}(\tau) - R_{s_b}(\tau)\right)^2\right\}$$

$$
\begin{aligned}
&= E\left\{\left(E_s\left[1-\frac{1}{6}\left(\frac{\pi\tau}{T_p}\right)^2\right]-E_s\right)^2\right\}\\
&= E\left\{\left[\frac{E_s}{6}\left(\frac{\pi\tau}{T_p}\right)^2\right]^2\right\}\\
&= \frac{E_s^2\,\pi^4}{36\,T_p^4}\;E\left\{\tau^4\right\}
\end{aligned}
$$

Finally, the normalized mean square error, which is the variance divided by the true value for an unbiased estimator, is calculated as follows.

$$
\varepsilon^2\left\{\hat{R}_{s_b}(0)\right\}=\frac{\mathrm{var}\{\hat{R}_{s_b}(0)\}}{R_{s_b}^2(0)}=\frac{1}{E_s^2}\,\frac{E_s^2\,\pi^4}{36\,T_p^4}\;E\left\{\tau^4\right\}=\frac{\pi^4}{36\,T_p^4}\;E\left\{\tau^4\right\}
$$

Assuming that the time delay follows a Gaussian distribution with variance $\sigma_1^2(\tau)$ near $\tau=0$, $E\left\{\tau^4\right\}=3\,\sigma_1^4(\tau)$. Combining the latter expression with the normalized mean square error formula, we get

$$
\varepsilon^2\left\{\hat{R}_s(0)\right\}=\frac{\pi^4}{36\,T_p^4}\;3\,\sigma_1^4(\tau)=\frac{\pi^4}{12\,T_p^4}\,\sigma_1^4(\tau)\iff\;\sigma_1(\tau)=12^{\frac{1}{4}}\,\frac{T_p}{\pi}\,\varepsilon\left\{\hat{R}_{s_b}(0)\right\}^{\frac{1}{2}}\iff
$$

$$
\sigma_1(\tau)=1.8612\;\frac{T_p}{\pi}\sqrt{\varepsilon\left\{\hat{R}_{s_b}(0)\right\}}. \tag{5.3}
$$

Our goal now is to calculate the normalized mean square error value of $R_{s_b}(\tau)$, for $\tau=0$. According to Section 4.4.1, we define the following system: the transmitted signal $s(t)$ passes through an AWGN channel where it is delayed by Δ seconds. Let $r(t)=s(t-\Delta)+n(t)$ be the received waveform and after the stage of demodulation at the receiver, we get $r_{\mathrm{dem}}(t)=s_{\mathrm{dem}}(t-\Delta)+n_{\mathrm{dem}}(t)$. The operation of cross-correlation is followed and we define the estimator of $R_{s_b}(\tau)$, as the product $R_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta,t)$

$$
\begin{aligned}
\hat{R}_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta,t)&=\int_{-\infty}^{+\infty}r_{\mathrm{dem}}(t+\tau+\Delta)\,s_b(t)\,\mathrm{d}t\\
&=\int_{-\infty}^{+\infty}\left[s_{\mathrm{dem}}(t+\tau+\Delta-\Delta)+n_{\mathrm{dem}}(t+\tau+\Delta)\right]s_b(t)\,\mathrm{d}t\\
&=\int_{-\infty}^{+\infty}s_b(t+\tau)\,s_b(t)\,\mathrm{d}t+\int_{-\infty}^{+\infty}n_{\mathrm{dem}}(t+\tau+\Delta)\,s_b(t)\,\mathrm{d}t\\
&=R_{s_bs_b}(t+\tau,t)+\int_{-\infty}^{+\infty}n_{\mathrm{dem}}(t+\tau+\Delta)\,s_b(t)\,\mathrm{d}t\\
&=R_{s_b}(\tau)+\int_{-\infty}^{+\infty}n_{\mathrm{dem}}(t+\tau+\Delta)\,s_b(t)\,\mathrm{d}t.
\end{aligned}
$$

The mean value of $\hat{R}_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta,t)$ is:

$$
E\left\{\hat{R}_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta,t)\right\}=E\left\{R_{s_b}(\tau)\right\}+E\left\{\int_{-\infty}^{+\infty}n_{\mathrm{dem}}(t+\tau+\Delta)\,s_b(t)\,\mathrm{d}t\right\}
$$

$$= R_{s_b}(\tau) + \int_{-\infty}^{+\infty} E\left\{n_{\text{dem}}(t+\tau+\Delta)\right\}\, s_b(t)\,\mathrm{d}t$$
$$= R_{s_b}(\tau).$$

Hence, the estimator $\hat{R}_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)$ is unbiased. Then, the variance is calculated.

$$\text{var}\left\{\hat{R}_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\} = E\left\{\left(\hat{R}_{r_{\text{dem}}s_b}(t+\tau+\Delta,t) - E\left\{\hat{R}_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\}\right)^2\right\}$$
$$= E\left\{\hat{R}^2_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\} - E\left\{\hat{R}_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\}^2 \iff$$

$$\text{var}\left\{\hat{R}_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\} = E\left\{\hat{R}^2_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\} - R^2_{s_b}(\tau) \tag{5.4}$$

Here, we have to calculate the quantity $E\left\{\hat{R}^2_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\}$ and we will come back to Eq. 5.4.

$$E\left\{\hat{R}^2_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\} = E\left\{\left[\int_{-\infty}^{+\infty} r_{\text{dem}}(t+\tau+\Delta)\, s_b(t)\,\mathrm{d}t\right]^2\right\}$$
$$= E\left\{\left[\int_{-\infty}^{+\infty} \left[s_{\text{dem}}(t+\tau+\Delta-\Delta) + n_{\text{dem}}(t+\tau+\Delta)\right] s_b(t)\,\mathrm{d}t\right]^2\right\}$$
$$= E\left\{\left[\int_{-\infty}^{+\infty} s_b(t+\tau)\, s_b(t)\,\mathrm{d}t + \int_{-\infty}^{+\infty} n_{\text{dem}}(t+\tau+\Delta)\, s_b(t)\,\mathrm{d}t\right]^2\right\}$$
$$= E\left\{\left[R_{s_b}(\tau) + \int_{-\infty}^{+\infty} n_{\text{dem}}(t+\tau+\Delta)\, s_b(t)\,\mathrm{d}t\right]^2\right\}$$
$$= R^2_{s_b}(\tau) + 2\, R_{s_b}(\tau) \int_{-\infty}^{+\infty} E\left\{n_{\text{dem}}(t+\tau+\Delta)\right\}\, s_b(t)\,\mathrm{d}t +$$
$$\iint_{-\infty}^{+\infty} E\left\{n_{\text{dem}}(t+\tau+\Delta)\, n_{\text{dem}}(\tilde{t}+\tau+\Delta)\right\}\, s_b(t)\, s_b(\tilde{t})\,\mathrm{d}t\,\mathrm{d}\tilde{t}$$

In the last equality the second term is zero and at the double integral of the last term, only the parameter for $t=\tilde{t}$ survives, because $R_{n_{\text{dem}}n_{\text{dem}}}(t_1,t_2) = \sigma_n^2\,\delta(t_1-t_2)$. Hence,

$$E\left\{\hat{R}^2_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\} = R^2_{s_b}(\tau) + \int_{-\infty}^{+\infty} E\left\{n^2_{\text{dem}}(t+\tau+\Delta)\right\}\, s_b^2(t)\,\mathrm{d}t$$
$$= R^2_{s_b}(\tau) + N_o \int_{-\infty}^{+\infty} s^2(t)\,dt$$
$$= R^2_{s_b}(\tau) + \frac{\sigma_n^2}{F_s}\, E_s.$$

As it is shown in Section 4.3.2, the variance of the noise product at the output of cross-correlation is $\sigma_w^2 = \frac{\sigma_n^2\, E_s}{F_s}$. Finally, as it is proven

$$E\left\{\hat{R}^2_{r_{\text{dem}}s_b}(t+\tau+\Delta,t)\right\} = R^2_{s_b}(\tau) + \frac{\sigma_n^2\, E_s}{F_s}. \tag{5.5}$$

Combining Eq. 5.5 with 5.4, we get

$$\mathrm{var}\left\{\hat{R}_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta,t)\right\} = R^2_{s_b}(\tau) + \frac{\sigma_n^2\, E_s}{F_s} - R^2_{s_b}(\tau) \iff$$

$$\mathrm{var}\left\{\hat{R}_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta,t)\right\} = \frac{\sigma_n^2\, E_s}{F_s}. \tag{5.6}$$

The normalized mean square error is given as the ratio of variance divided by the square of true value. Using Eq. 5.6, we get

$$\varepsilon^2\left\{\hat{R}_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta,t)\right\} = \frac{\mathrm{var}\left\{\hat{R}_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta,t)\right\}}{R^2_{s_b}(\tau)} = \frac{\frac{\sigma_n^2\, E_s}{F_s}}{R^2_{s_b}(\tau)}.$$

From Eq. 4.8, we have that $\sigma_n^2 = \frac{\sigma_s^2}{10^{\mathrm{SNR_{dB}}/10}} = \frac{E_s}{T\,10^{\mathrm{SNR_{dB}}/10}}$, where T is the time duration of signal. Combining last two equalities, normalized mean square error is given as

$$\varepsilon^2\left\{\hat{R}_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta,t)\right\} = \frac{\frac{E_s^2}{T\,F_s\,10^{\mathrm{SNR_{dB}}/10}}}{R^2_{s_b}(\tau)} = \frac{E_s^2}{R^2_{s_b}(\tau)\,T\,F_s\,10^{\mathrm{SNR_{dB}}/10}}.$$

Recalling our scope, we wanted to estimate the normalized mean square error of $R_{s_b}(\tau)$ at $\tau = 0$, which is given in the following form.

$$\varepsilon^2\left\{\hat{R}_{s_b}(0)\right\} \simeq \varepsilon^2\left\{\hat{R}_{r_{\mathrm{den}}s_b}(t+\Delta,t)\right\} = \frac{E_s^2}{E_s^2\,T\,F_s\,10^{\mathrm{SNR_{dB}}/10}} = \frac{1}{T\,F_s\,10^{\mathrm{SNR_{dB}}/10}}$$

Finally, replacing the last equality in Eq. 5.3, we get

$$\sigma_1(\tau) = 1.8612\,\frac{T_p}{\pi}\left(T\,F_s\,10^{\frac{\mathrm{SNR_{dB}}}{10}}\right)^{-\frac{1}{4}}. \tag{5.7}$$

According to Eq. 5.7, an approximate 95% confidence interval for the time delay Δ in seconds, is:

$$[\hat{\Delta} - 2\,\sigma_1(\tau) \le \Delta \le \hat{\Delta} + 2\,\sigma_1(\tau)]. \tag{5.8}$$

Note that, the range of the confidence interval is related to the pulse duration (which is directly associated to the bandwidth of the raised cosine pulse, BW$= (1+a)/(2\,T_p)$, in our case), the signal duration and the value of the SNR. Specifically, as the pulse duration is decreasing, the margin of the confidence interval is also reduced. Finally, as the value of the SNR and/or the signal duration are increasing, the range of the confidence interval is decreasing.

6. Simulations

In this Chapter, experimental results will be presented to verify the effectiveness of the chosen techniques and also demonstrate their superiority compared to typical systems in similar schemes.

6.1 Experimental results for the DS technique

In this case, the baseline models for comparison, consist of widely used pulses in the position of the proposed 15th derivative of the Gaussian pulse. The structure of the defined transmitted signal remains identical, including the number of using pulses, the selected DS code, and the specified energy per pulse. The scenarios to determine the valuation of our techniques, involve the presence of (a) AWGN, (b) Pink noise and (c) representative music samples. In Fig. 6.1.1 we compare the performance of our system for different selections of the pilot's pulse. We compare Dirac, rectangular, Gaussian and the 15th derivative of the Gaussian. The effectiveness of the system is measured using the probability of error. A binary decision for the delay estimation is considered: if the delay estimation is correct (the estimated equals the true time delay), no error occurs. The probability of error is calculated as the ratio of the total number of errors divided by the number of repetitions, for each value of SNR. The best performance under AWGN is expected to be achieved with Dirac pulses, which indeed is the case. The theoretical probability of error using Dirac pulses, is also considered. We observe that the 15th derivative of the Gaussian pulse achieves similar performance to the Dirac, which can be successfully predicted using Eq. 5.2.

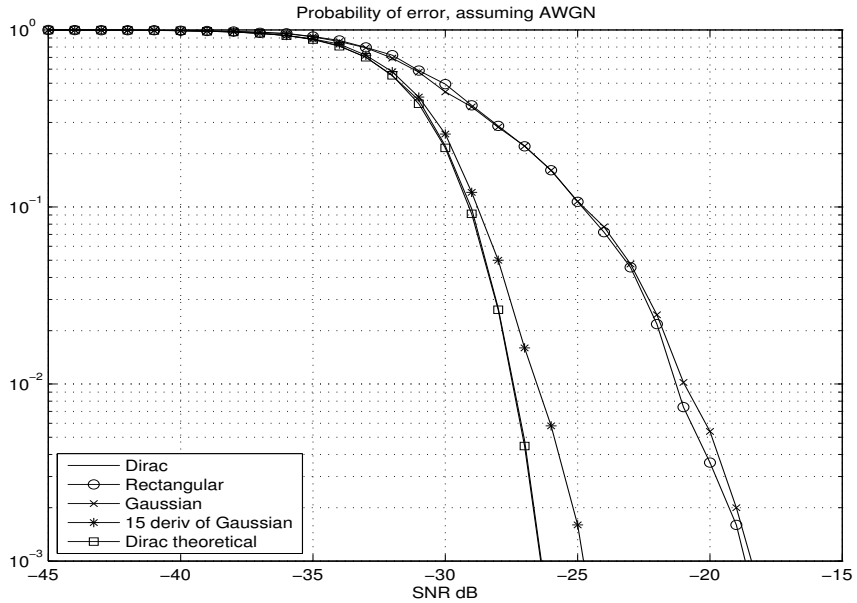

Figure 6.1.1: AWGN.

Pink noise, where each octave contains the same amount of energy, can be a useful model for rich musical sounds. In Fig. 6.1.2 we present the performance of different pulses under Pink noise. In this case our proposed pulse performs much better than the other candidate pulses. The reason for the substantial difference in performance is that, by design, the spectrum of our pulse is positioned in frequencies where the Pink noise's

power is decreased. Consequently, a significant increase in SNR at the output of the cross-correlation is observed, in contrast to the other candidate pulses. Specifically, it was calculated[1] that the proposed system under Pink noise gains approximately $7-8$ dB, in relevance to the case where AWGN is considered, which is verified by the experimental results.

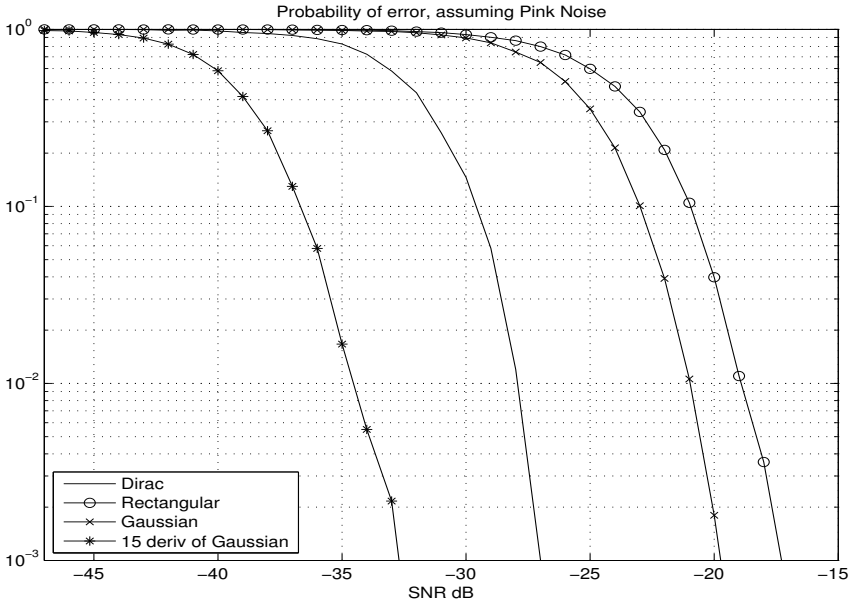

Figure 6.1.2: Pink noise.

Finally, the performance of the system assuming the co-existence of various music samples is shown in Fig. 6.1.3. The selected samples are based on typical orchestra instrumentation involving baritone, bass, cello, clarinet, flute, horn, snare, trombone, trumpet, tuba, viola and violin tracks, whose spectral characteristics are highlighted in Section 2.1. Given the pulse that is used, it is obvious that different behavior is obtained for different music samples, as the corresponding PSD differs. Hence, we calculated the average probability of error over all the selected music samples for each compared system, in Fig. 6.1.3. The exact performance of the proposed system for each music track, is given in Fig. 6.1.4. The results show that the proposed pulse outperforms all the rest by a significant margin. It is important to note, that by setting the specified value for energy per pulse, all the candidate pilot signals were audible except of the proposed.

6.2 Experimental results for the BPSK technique

In this case, the baseline models for comparison, consist of the same pilot signals, but, with different selected center frequency F_c. Typical values were chosen, $F_c = 2, 7, 10$ kHz and compared to the proposed 18.997 kHz. It is already mentioned, that the selected center frequency does not restrict the general case, where a center frequency that pertain to high band should be selected - in general: higher passband, better expected performance, with respect to the available bandwidth. The main idea remains the same, to transfer the signal spectrum to high frequencies and the scope is to demonstrate that in this simple case of signal construction, the selected center frequency is crucial for the performance of our system, under the presence of music. Similarly to the DS assump-

[1]The calculations to quantify the difference under Pink noise and AWGN, are in Appendix II.

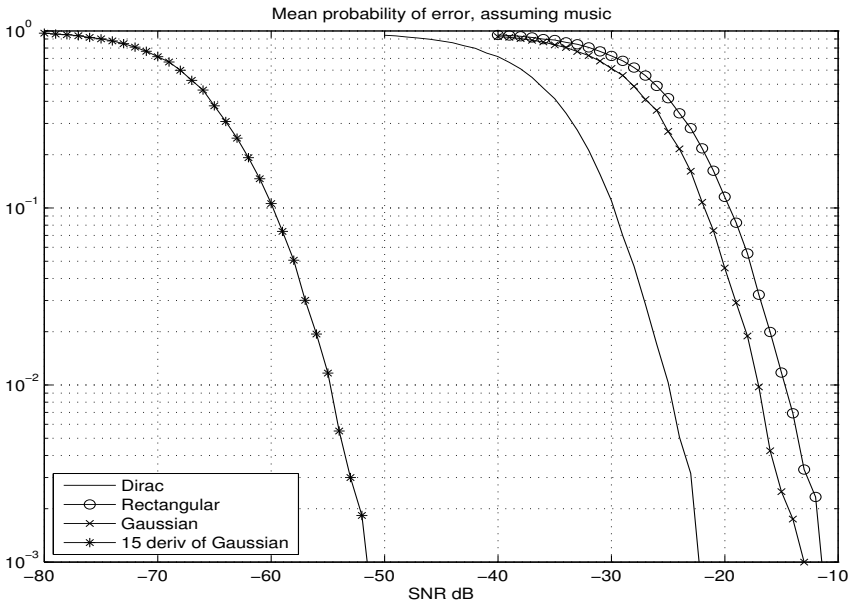

Figure 6.1.3: Music.

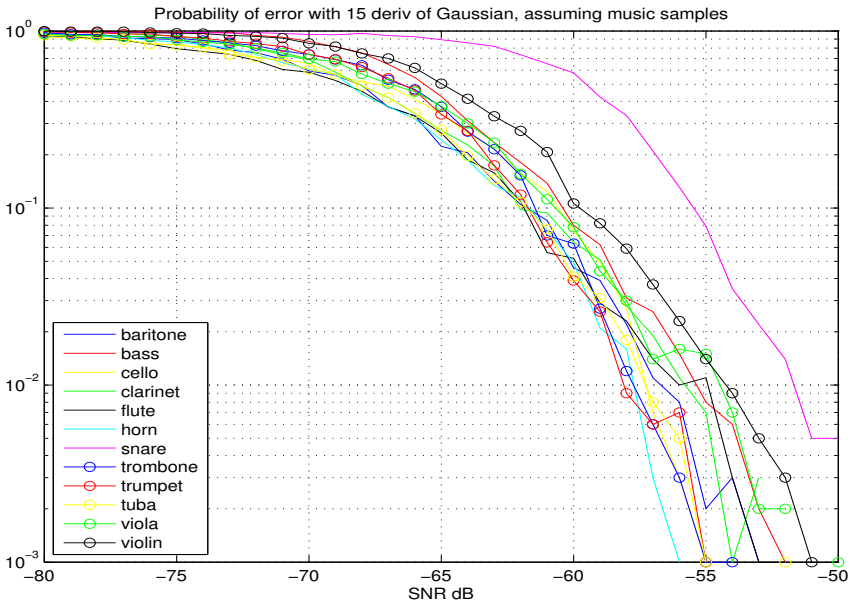

Figure 6.1.4: Representative music samples.

tions, the structure of the signal's construction remains identical under all baseline models and the presence of (a) AWGN, (b) Pink noise and (c) representative music samples, was considered. At first, a binary decision for the delay estimation was applied. In Fig. 6.2.1 the performance of all the systems is shown and we observe that, identical graphs are derived for all the signals, which was expected under AWGN, because its PSD is flat over the total bandwidth. It is, also, noted that their performance is reduced, in relevance to the corresponding DS case.

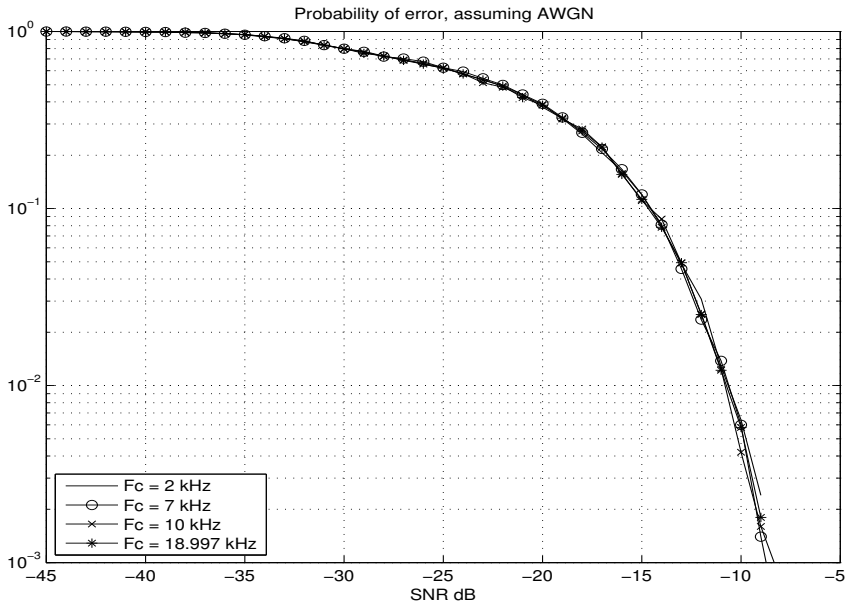

Figure 6.2.1: AWGN.

Concerning the degradation of the performance in relevance to the DS case, the signal product at the output of the cross-correlation is a raised cosine pulse and its amplitude nearby the peak, can lead to wrong decisions regarding the delay estimate. It is quite possible, then, to decide that the channel delay equals to a value that corresponds to a sample nearby its peak, under the presence of noise. Thus, in Fig. 6.2.2 we apply the confidence intervals that were calculated in Section 5.2, and simultaneously we include the probability of error assuming the Dirac pulse train that was analyzed in Section 5.1. We observe that, by applying the 95% confidence intervals, the performance of all the baseline models is almost identical to the performance achieved by the Dirac pulse train. However, small differences are noted for lower values of the SNR, or equivalently, for higher power of noise. This is observed, because through our calculations of Section 5.2 we defined the estimator of $\hat{R}_{s_b}(\tau)$ as the product $R_{r_{\mathrm{dem}}s_b}(t+\tau+\Delta, t)$ which includes a noise term. Specifically, we set $\varepsilon^2\left\{\hat{R}_{s_b}(0)\right\} = \varepsilon^2\left\{\hat{R}_{r_{\mathrm{den}}s_b}(t+\Delta, t)\right\}$, which is an approximation and the equality is achieved for low power of noise. These inaccuracies were expected and are eliminated, as the power of noise is decreasing. Fig. 6.2.2 is quite complex and is based on the following operations: (1) The confidence intervals were estimated using Eq. 5.8. Note that, the confidence interval is related to the signal's time duration T, the sampling frequency F_s, which are both considered constants and the SNR value. Hence, we get a different interval for a different value of SNR. (2) The estimation of the delay was performed. It is considered correct, if the estimated delay is lying within the range of the corresponding confidence interval. Otherwise, an error occurs.

The confidence intervals are defined over the time scale. Multiplying the correspond-

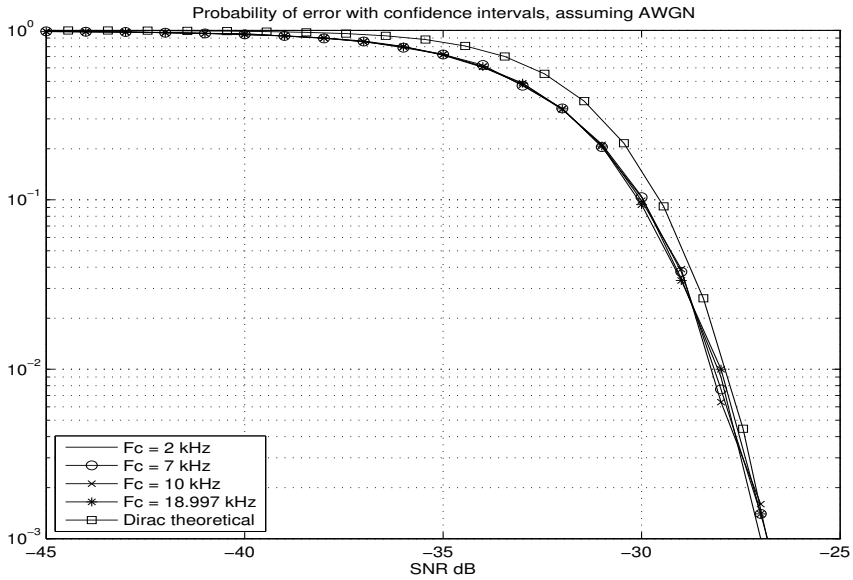

Figure 6.2.2: AWGN.

ing time intervals, $2\,\sigma_1(\tau)$, with the sampling frequency F_s, we get the same quantities in samples. In Fig. 6.2.3, we present the calculated confidence intervals, $2\,\sigma_1(\tau)$, in samples. Note that, the first value corresponds to SNR$= -45$ and the last to SNR$= -15$ dB.

```
Command Window
>> confidence_intervals

confidence_intervals =

  Columns 1 through 10

   19.4497   18.3617   17.3345   16.3649   15.4494   14.5852   13.7693   12.9991   12.2719   11.5854

  Columns 11 through 20

   10.9374   10.3255    9.7479    9.2026    8.6878    8.2019    7.7431    7.3099    6.9010    6.5150

  Columns 21 through 30

    6.1505    5.8065    5.4817    5.1750    4.8855    4.6122    4.3542    4.1107    3.8807    3.6636

  Column 31

    3.4587
```

Figure 6.2.3: Samples that correspond to confidence intervals.

Finally, these results help us to define the operation point of our system, as follows. When we know about, or we can estimate the conditions of the channel that is involved during our pilot transmission, i.e., the SNR value, we can calculate and define the 95% confidence interval over which the peak should be located with the corresponding probability of error. At that spot, where we have the interval and the peak is located in it with a specified probability, either more processing should be performed in order to detect the exact position of the peak, or we can assume that the accuracy of our estimation is decreasing as the margin of the confidence interval is increasing, respectively.

In Fig. 6.2.4 we present the performance of the considering signals. It is obvious that the signal with the proposed center frequency outperforms the other ones. This was expected, since the location of the signal spectrum at higher bands, result in a more robust communication due to the decrease of Pink noise power at higher frequencies.

The same reason leads to the variation of the performance for the various cases under Pink noise, in contrast to the corresponding cases under AWGN. Hence, for the proposed center frequency, the SNR at the output of the cross-correlation is increased in relevance to the SNR value at the input. However, the binary decision for the delay estimation, results to poor performance, compared to the corresponding case of the DS technique.

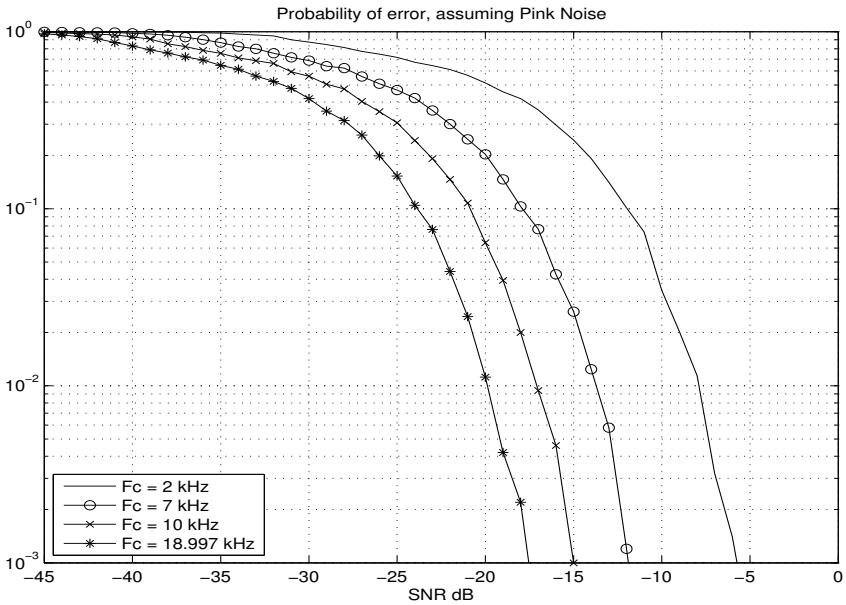

Figure 6.2.4: Pink noise.

The performance of the system with the representative music samples, is given in Fig. 6.2.5, which shows the average probability of error over all the selected music samples, for each compared system. The exact performance of the proposed system, for each music track, is given in Fig. 6.2.6. The results show that the proposed pulse outperforms all the rest by a significant margin. In addition, by setting the specified value for energy per pulse, all the pilot signals were audible except of the proposed, due to the selected center frequency.

6.3 Comparison between DS and BPSK techniques

Finally, we examine the performances of both the signal structures under the same power of music tracks. The reason is that the SNR is defined as the power of the signal divided by the power of the corresponding music track. However, as depicted in Section 3.1.2.3, different values were defined for the different cases of signal construction, concerning the pulse energy and duration. Note that the total signal duration remains constant. Specifically, according to our definitions, we get that the signal power under the DS and BPSK technique are given as $P_{\mathrm{DS}} = 1.33 \cdot 10^{-9}$ and $P_{\mathrm{BPSK}} = 1.47 \cdot 10^{-5}$, respectively. Hence, it is interesting to evaluate both techniques, assuming the existence of music with the same power. Then,

$$\mathrm{SNR}_{\mathrm{DS}} = \frac{P_{\mathrm{DS}}}{\sigma_n^2}$$

and

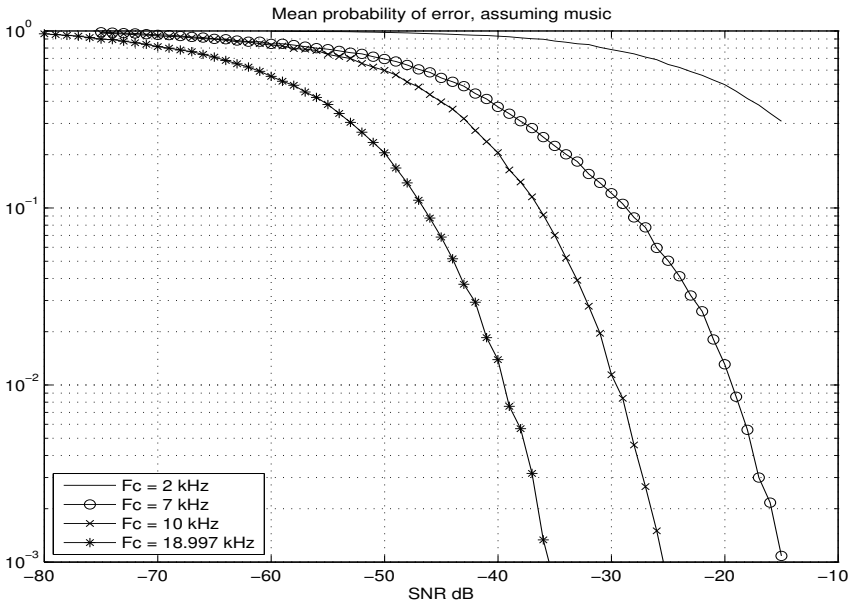

Figure 6.2.5: Music.

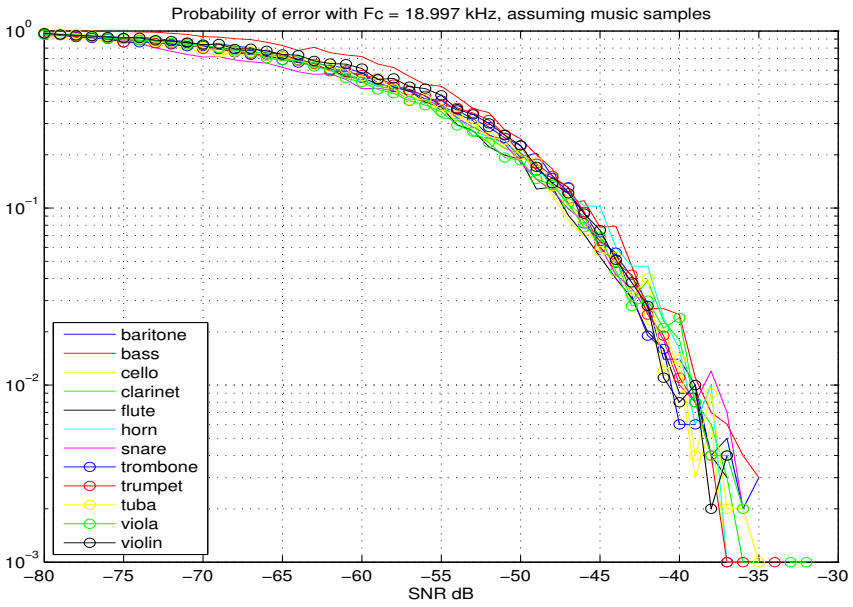

Figure 6.2.6: Representative music samples.

$$\mathrm{SNR}_{\mathrm{BPSK}} = \frac{P_{\mathrm{BPSK}}}{\sigma_n^2} = \frac{P_{\mathrm{BPSK}}}{P_{\mathrm{DS}}/\mathrm{SNR}_{\mathrm{DS}}} = \frac{P_{\mathrm{BPSK}}}{P_{\mathrm{DS}}} \cdot \mathrm{SNR}_{\mathrm{DS}}.$$

In a logarithmic scale, the latter relationship is given as

$$\mathrm{SNR}_{\mathrm{BPSK,dB}} = 10 \cdot \log_{10}\left(\frac{P_{\mathrm{BPSK}}}{P_{\mathrm{DS}}}\right) + \mathrm{SNR}_{\mathrm{DS,dB}} = -40.34 + \mathrm{SNR}_{\mathrm{DS,dB}}$$

according to the specified values. To explain the previous expression, in order to define the same power of noise for both techniques, the $\mathrm{SNR}_{\mathrm{BPSK,dB}}$ should be reduced by 40.34 dB, in relevance to the $\mathrm{SNR}_{\mathrm{DS,dB}}$ value. Thus, a simple transfer of the SNR axis is sufficient in order to compare both performances. However, note that if we set as a reference the $\mathrm{SNR}_{\mathrm{BPSK,dB}}$ values, right shift of 40.34 dB should be applied at the DS curve. If we set as a reference the $\mathrm{SNR}_{\mathrm{DS,dB}}$ values, left shift of -40.34 dB is appropriate for the BPSK graph. In Fig. 6.3.1, for both systems, the mean probability of error over all chosen music samples is shown, considering the latter choice.

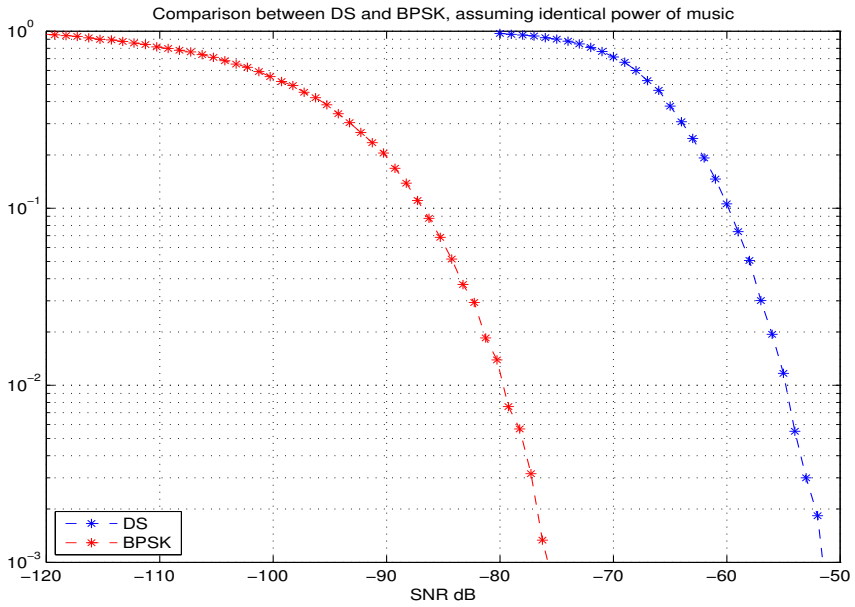

Figure 6.3.1: Comparison with Music.

We observe that, the performance of the DS technique is increasing with a higher rate, in relevance to the BPSK, which can be also detected in Fig. 6.1.3 and 6.2.5, correspondingly. However, in BPSK we use band-limited pulses that are transferred to higher bands and stronger signals can be inaudible. This difference of signal powers is critical and results to better performance of the BPSK technique, under the same power of music. Note that a value of SNR in the range $\{-65, -70\}$ dB according to Fig. 6.3.1, leads to amplitudes of tracks that do not usually exceed 1 (where the operation of clipping is performed to a .wav file).

Useful Note: Finally, assuming an increase of pilot duration, for each case of signal construction and under each form of noise, the performance will be increased by $10 \cdot$

$\log 10(k)$ dB, where k denotes the ratio of the new duration T_{signal}, divided by our specified value 0.5 sec. For example, if we double the time duration, the peak of the auto-correlation of the signal product will be also doubled, indicating an increase in performance, or a shift of the SNR axis by -3 dB. In this way, the desirable performance becomes a function of the pilot's duration.

7. Performance Evaluation under Compression and Fractional Delay Estimation

In this Chapter, further evaluation of the system capabilities is performed. Firstly, we propose a technique to estimate fractional delays of the channel, and practical simulations are provided. Secondly, we investigate the performance of our system under compression, and the preliminary results are presented. Consequently, we explain how to use this system under higher sampling frequencies. Finally, some noteworthy applications are presented, where the system could be adopted.

7.1 Fractional delay estimation

We examine to improve the performance of the system, requiring a fractional delay estimation. The proposed technique is based on upsampling the received signal by a factor L, just before applying cross-correlation. This method indicates that the cross-correlation waveform acquires higher resolution, and thus, more accurate delay estimation is achieved. Specifically, in Fig. 7.1.1 the original pulse, along with the upsampled by a factor of 4, are shown in the time domain. It is obvious, that the duration of the letter, is four times bigger. In addition, in Fig 7.1.2, the spectrum of these pulses is presented. It is noted that the upsampling leads to the shrink of spectrum by $1/4$ in the digital representation of frequencies, as it will be derived also through the following mathematical calculations.

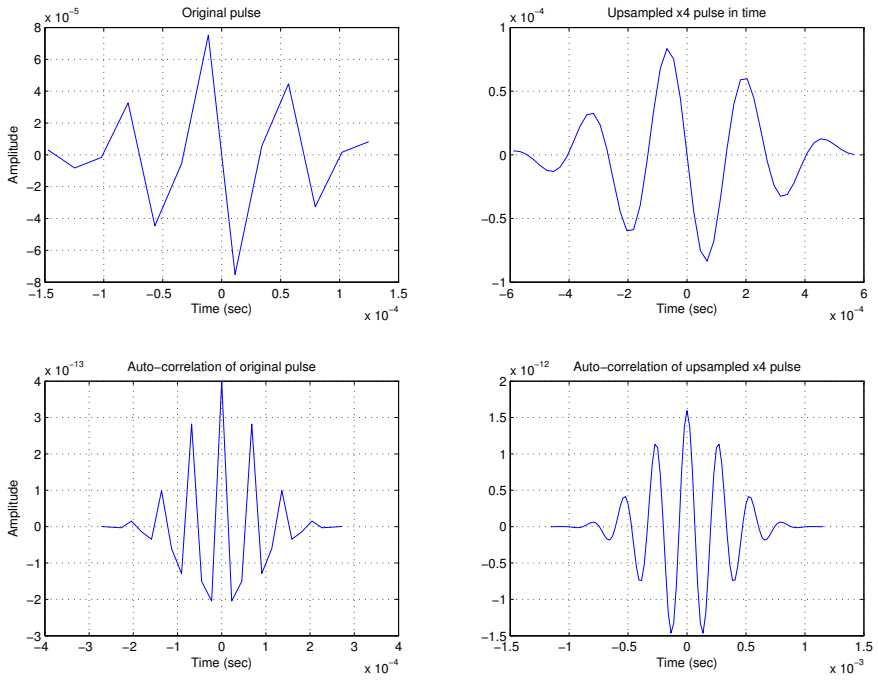

Figure 7.1.1: Time domain waveforms.

Typically, upsampling process involves two steps. $L-1$ zeros are inserted between each sample of the original signal, followed by filtering. The use of a filter is critical, in order to fill the zero samples that were inserted, in time domain. An other equivalent aspect that indicates the necessity for using a filter, can be shown in frequency domain. Zero-inserted signal is composed of L repetitions of the original signal's spectrum and,

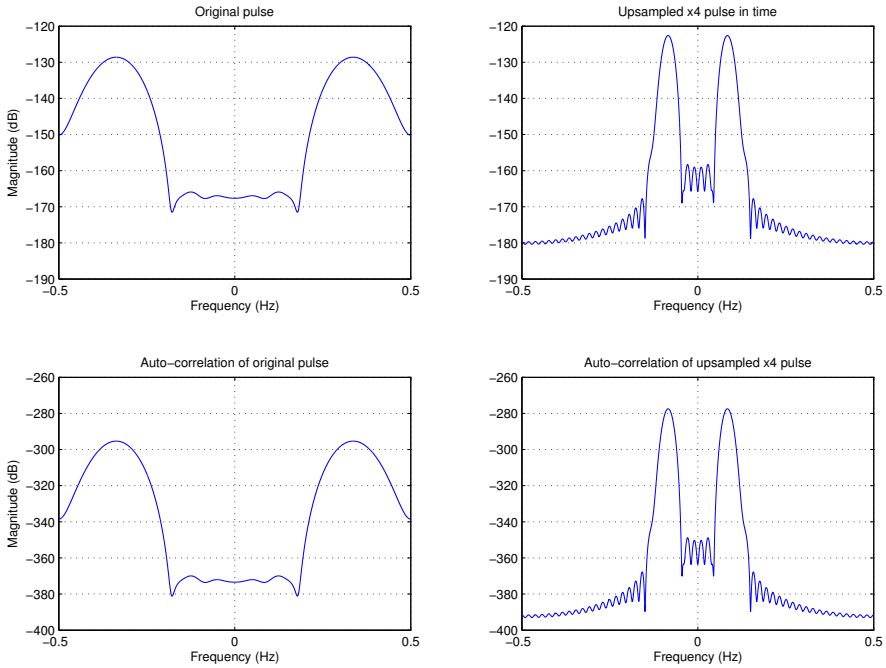

Figure 7.1.2: Frequency domain waveforms.

hence, in order to maintain a single scalable version of the original spectrum, filtering is needed. Specifically, zeros-inserted signal, which is obtained by inserting $L-1$ zeros between every two samples of $x(m)$, can be expressed by Eq. 7.1

$$x_z(m) = \begin{cases} x\left(\frac{m}{L}\right), & m = 0, \pm L, \pm 2L, \dots \\ 0, & \text{otherwise} \end{cases} \tag{7.1}$$

The spectrum of the zeros-inserted signal is related to the spectrum of the original discrete - time signal, as the following calculations indicate.

$$\begin{aligned} X_z(f) &= \sum_{m=-\infty}^{+\infty} x_z(m)\, \mathrm{e}^{-j2\pi f m} \\ &= \sum_{\substack{m=-\infty, \\ m\%L=0}}^{+\infty} x\left(\frac{m}{L}\right)\, \mathrm{e}^{-j2\pi f m} \\ &\overset{!}{=} \sum_{n=-\infty}^{+\infty} x(n)\, \mathrm{e}^{-j2\pi (fL) n} \\ &= X(f\,L) \end{aligned}$$

where at the spot "!", we set $n = m/L$ and an equivalent expression was derived as a function of n. In Fig. 7.1.3, the appropriate stages to perform upsampling on signal $x(n)$ are shown, while in Fig. 7.1.4, the signal spectrum in each step of upsampling process is analyzed and the significance of the filter is highlighted. Note that, digital frequencies are considered for the spectrum projection.

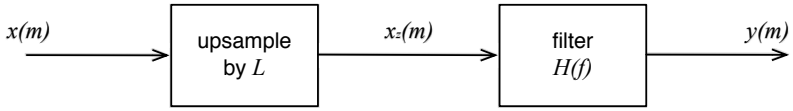

Figure 7.1.3: Upsampling steps.

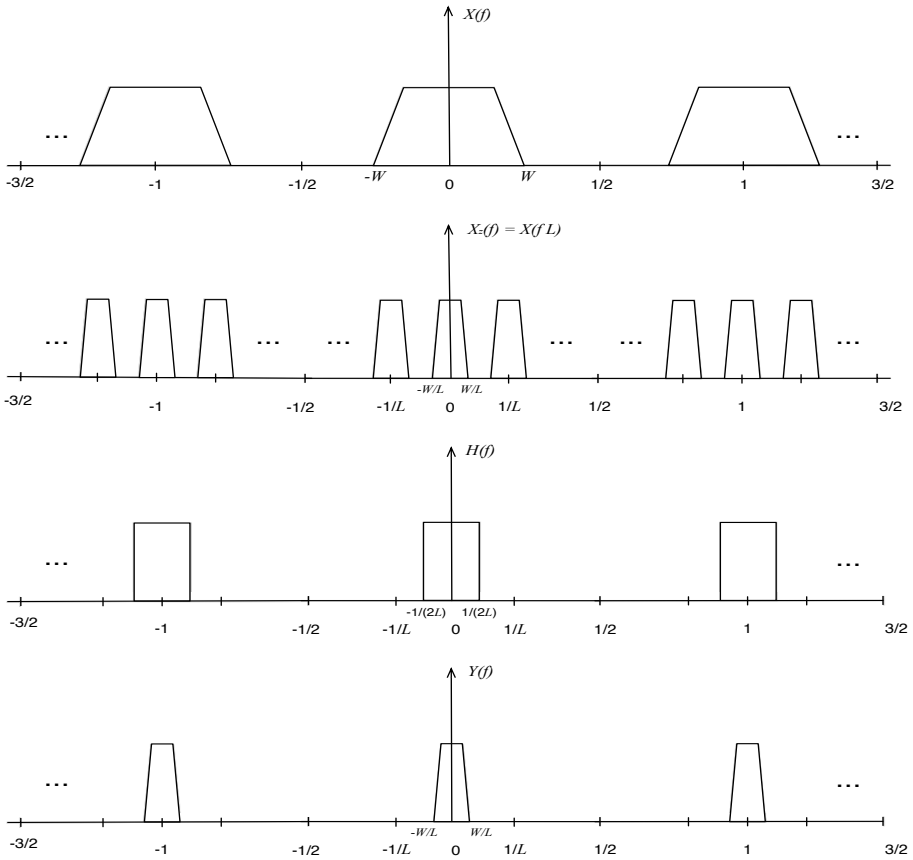

Figure 7.1.4: Spectrum changes during upsampling steps.

In order to evaluate our system, a channel delay in a sub-sample scale should be considered. To implement such a delay, upsampling by a factor of L, appropriate shifting of the resampled signal and downsampling by a factor of L took part, before the transmission of the pilot signal. In brief, the process of downsampling by a factor of K is based on two stages as Fig. 7.1.5 depicts. Let $x(n)$ be the signal at which downsampling will be performed. At first, a low-pass filter with a cutoff frequency $1/(2\,K)$ in digital representation of the frequencies, should be applied to $x(n)$. And secondly, the discarding of $K-1$ samples for every K samples. The use of a low-pass filter is critical, because after the discarding of samples, signal spectrum expands by a factor of K and, hence, it constitutes an anti-aliasing process. In our case, downsampling after upsampling, both by L took part.

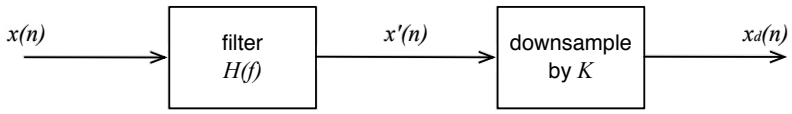

Figure 7.1.5: Downsampling steps.

DS pilot performance

In Fig. 7.1.6 and 7.1.7, the performance of our system under AWGN and Pink noise, respectively, is shown. The received signal was upsampled by a factor $L = 2, 3, 4, 5$ and 6. The rule under which a correct decision is obtained, is to estimate the corresponding fractional delay of the channel. Our results show, that for $L = 2$ we can succeed a half-sample accuracy with almost identical performance, under AWGN and Pink noise. However, for $L > 2$, system performance is slowly decreasing as the number of L is increasing.

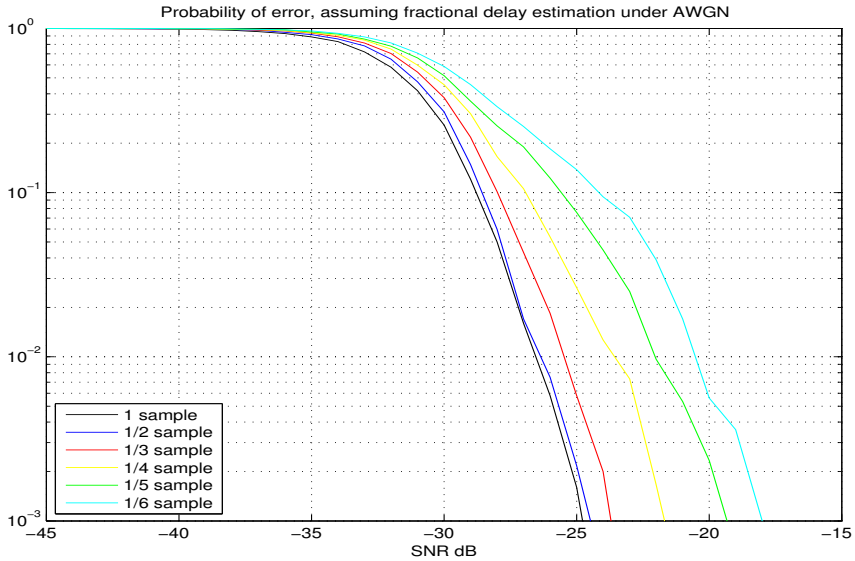

Figure 7.1.6: Fractional delay under AWGN.

This is observed, because after upsampling process, signal's spectrum shrinks by $1/L$

and the requirement for wide bandwidth is not met anymore, resulting in degradation of estimation's accuracy. This effect is enhanced under the BPSK signal construction, as it will be seen. In the DS case, the auto-correlation function of the used pulse leads to a shape where the peak value is differentiated enough by the second bigger amplitude. Indeed, it was calculated that the decrease in performance, assuming AWGN or Pink noise, under the upsampling process by any factor L, was related to errors where the decision of the peak value was located to a sample next to the sample that corresponds to the true peak, according to Fig. 7.1.1. Note that the next sample, corresponds to a value in seconds related to the upsampling process (i.e., for $L = 4$, a sample corresponds to $T_s/4$).

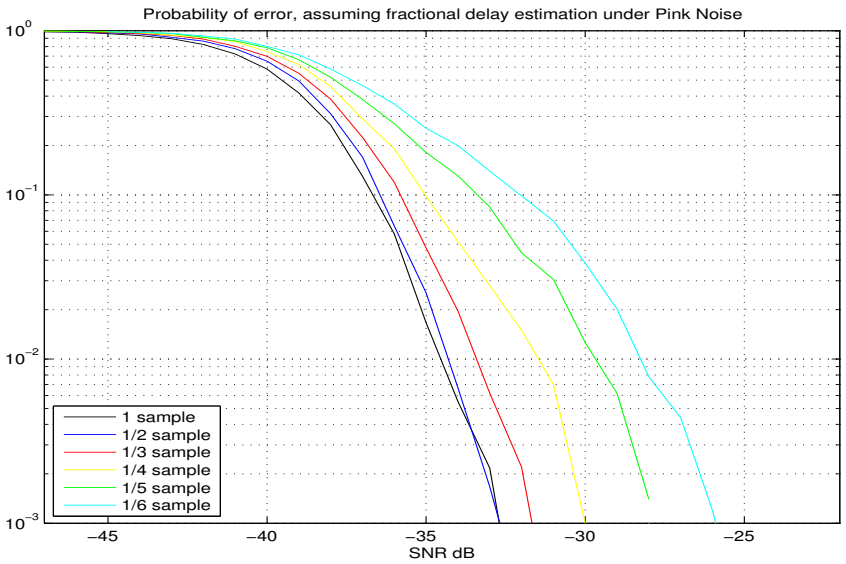

Figure 7.1.7: Fractional delay under Pink noise.

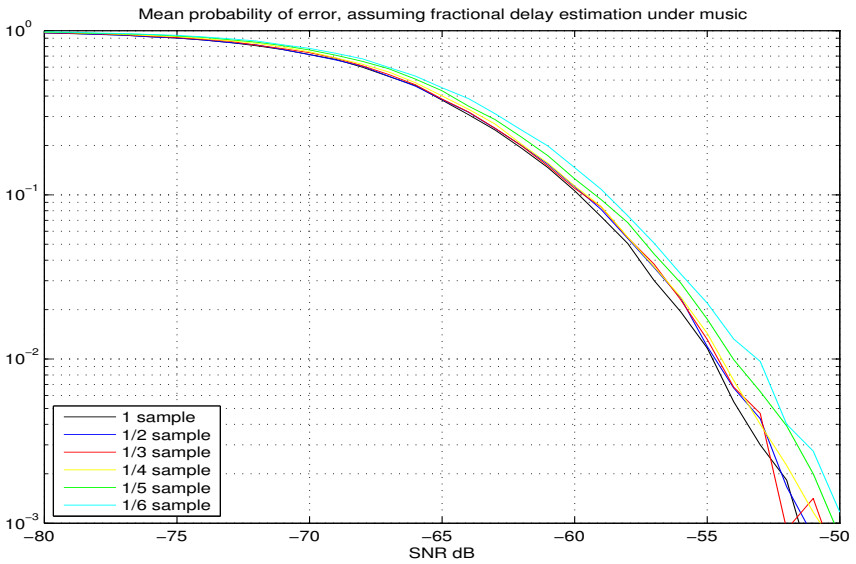

Figure 7.1.8: Fractional delay under music.

Furthermore, the performance of the proposed system, is evaluated under the pres-

ence of the representative music samples. In Fig. 7.1.8, the average probability of error over all the selected tracks is given, for each value of upsampling factor L. It is noted that the performances are almost identical, with a small decrease when $L = 6$. It is obvious that when music exists, the effects of upsampling process are eliminated. The properties of the auto-correlation of the used pulse in time domain, along with the music structure, are considered responsible for this behavior. Specifically, as it is derived through the general PSD model of music in Fig. 2.1.13, the expected "noise" after the filtering, assuming just the music signal as input, is a slow varying waveform which does not alter the efficiency of estimating the peak of the auto-correlation function given in Fig. 7.1.1. Finally, we should note that same performances we get if we assume randomly selected fractional delays.

BPSK pilot performance

The same technique was also considered assuming the pilot that adopts the BPSK structure. The steps and the assumptions were identical. Note that, the upsampling process at the receiver was applied after the coherent demodulation. In Fig. 7.1.9 and 7.1.10, the performance of our system under AWGN and Pink noise, respectively, is shown. The received signal was upsampled by a factor $L = 2, 3, 4$ and 5. The rule was to estimate the corresponding fractional delay of the channel, which was selected randomly.

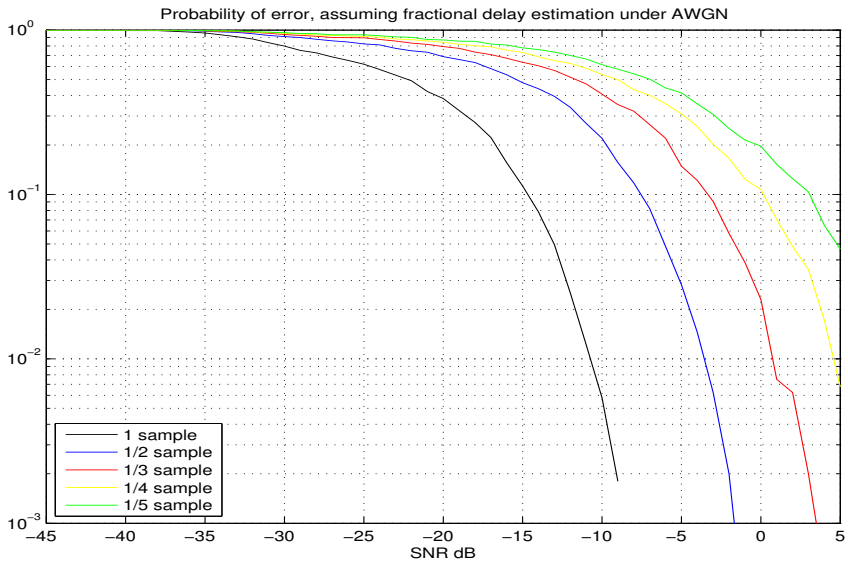

Figure 7.1.9: Fractional delay under AWGN.

As it is observed, the performance is decreasing with a higher rate, as the factor L is increasing, in contrast to the DS case. Specifically, to estimate the accurate value of the delay under half-sample accuracy, the performance is decreased by approximately 7 dB for both AWGN and Pink noise. Finally, the performance of our system including the proposed technique for the fractional delay estimation, is evaluated under the presence of the representative music samples. In Fig. 7.1.11, we present the average probability of error over all the selected tracks, for each value of the upsampling factor L.

It is obvious that even for half-sample accuracy on delay estimation, the performance is substantially decreased. The shrink of spectrum is crucial, since pulse's shape is ex-

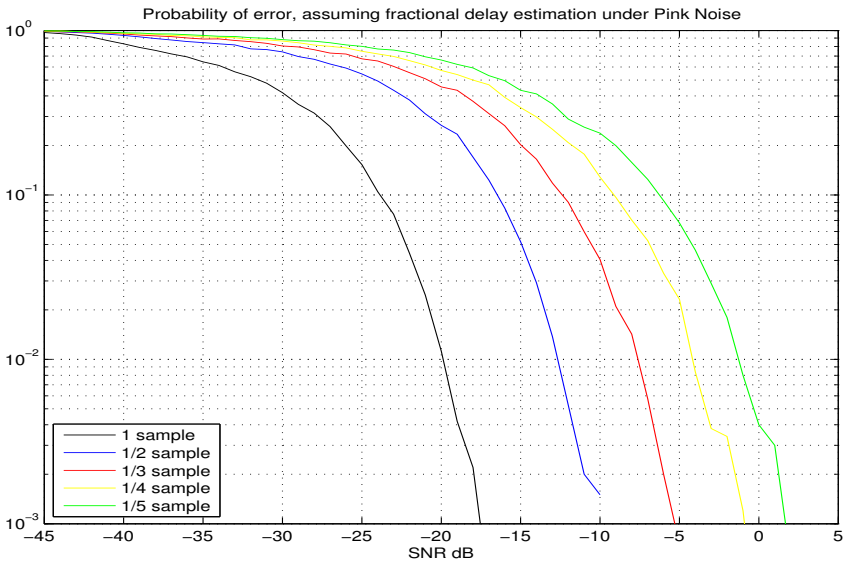

Figure 7.1.10: Fractional delay under Pink noise.

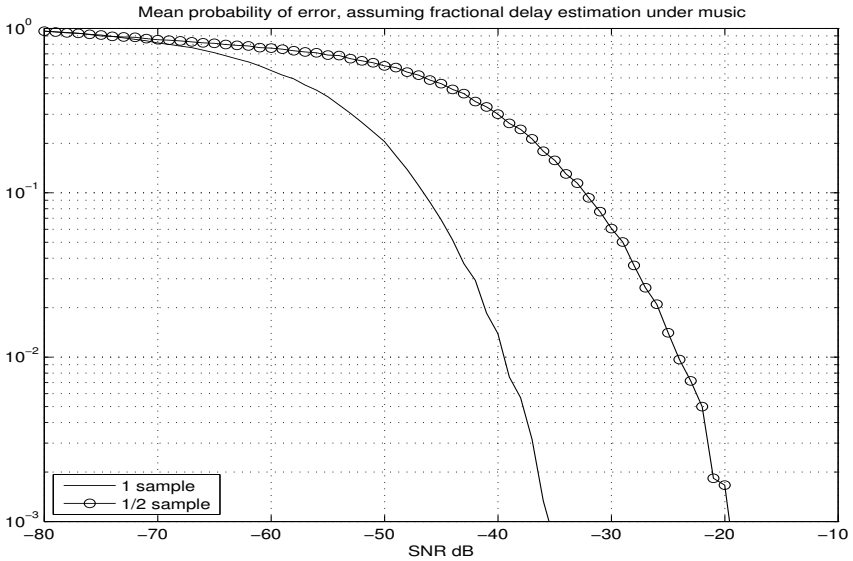

Figure 7.1.11: Fractional delay under music.

tended after upsampling and, hence, its auto-correlation results to a raised-cosine pulse with bigger width. Given the shape of the resulting pulse, it is obvious that lots of samples correspond to values around the peak and, thus, it is getting harder to estimate the exact peak value. Furthermore, after the demodulation process, the music spectrum is transferred to high frequencies, resulting in a quickly varying noisy waveform which is added to the auto-correlation function of the SRRC, and makes even harder the localization of the correct peak which corresponds to the true delay. That observation explains the results of the augmented decrease in performance under the presence of music, in relevance to the case where the AWGN is assumed. According to Fig. 7.1.11, in order to achieve half-sample accuracy, a degradation of approximately 15 dB is observed, in contrast to the DS case, where the performances are identical. Thus, it is obvious that the BPSK technique outperforms the DS, when accuracy of one sample is required, but, in order to achieve sub-sample accuracy the DS method is better.

7.2 Robustness under Compression

Preliminary results show that the considering pilot constructions are merely "**robust**" against typical encoding/decoding processing. Based on [13], an Internet audio codec for applications of "Live Distributed Music Performances", such as NMP, should meet the following requirements. The full audio bandwidth (44.1 or 48 kHz sampling rate) should be transmitted and audio codecs must be able to efficiently operate within the range of bit rates $64 - 128$ kbits/sec. The performance of our pilot signal was examined under the use of the OPUS codec[1]. The following process was applied: (1) At first, we constructed our pilot equivalently, defining $F_s = 48$ kHz. In order to be in synchronization with our previous results signal energy should be the same. Thus, identical number of samples per pulse was set and equal number of pulses per signal, resulting to the same total amount of samples. In this case, though, it is obvious that the signal duration in seconds is decreased by a factor of $44.1/48$, in relevance to the signal duration with $F_s = 44.1$ kHz. (2) Secondly, we quantized the transmitted pilot using Pulse-Coding Modulation (PCM) with 16 bit depth, sampling frequency of 48 kHz and we stored the corresponding file in a .wav form. Slight distortion, almost negligible, was caused by the quantization process. Specifically, Root-Mean-Square Error (RMSE) was approximately 10^{-7}. (3) Then, we used the free software XLD[2] in order to transcode the audio files from .wav to .pcm and vice versa. (4) The encoding and decoding process is applied in the latter form of the audio files, which is raw PCM data with little-endian ordering, according to the specified requirements of the executable file opus_demo. The bit rate of OPUS codec was set at the lowest and biggest value. (5) The decoded file was transcoded from .pcm to .wav form and the evaluation of the latter, with respect to the original pilot signal, was performed. The same 5-step process repeated multiple times assuming (a) just the presence of the pilot, simulating a silent channel, and (b) the co-existence of the pilot and a randomly selected track from the representative music samples, considering a typical implementation of our system.

DS pilot robustness

The file that was including just the pilot, firstly, was evaluated, in order to observe the effects of the encoding/decoding. After the decoding process, increased distortions are

[1] http://www.opus-codec.org
[2] http://tmkk.undo.jp/xld/index_e.html

presented in the pilot signal $\hat{s}(t)$, as bit rate decreases. Performing cross-correlation with the original signal to estimate the delay value by locating the peak of the resulting product, interesting results came up. Setting 128 kbits/sec bit rate, the peak value of cross-correlation, $R_{\hat{s}s}(\Delta)$, between the altered and the original signal, found to be almost identical to the peak value of the original signal's auto-correlation function, $R_{ss}(\Delta)$. By setting 64 kbits/sec bit rate, the peak value of $R_{\hat{s}s}(\Delta)$ was approximately equal to $0.5\,R_{ss}(\Delta)$. In addition, typical RMSE values were $1.8 \cdot 10^{-5}$ and $4.6 \cdot 10^{-5}$, correspondingly. Secondly, the file that was including the complex audio signal, was evaluated. The latter results, are verified through the figures below, where the cross-correlation product of a typical implementation assuming the co-existence of a music sample with the pilot signal, are presented. At the received signal, we apply (a) no coding, (b) encoding/decoding with bit rate 64 kbits/sec and (c) encoding/decoding with bit rate 128 kbits/sec, in Fig. 7.2.1, 7.2.2 and 7.2.3, respectively.

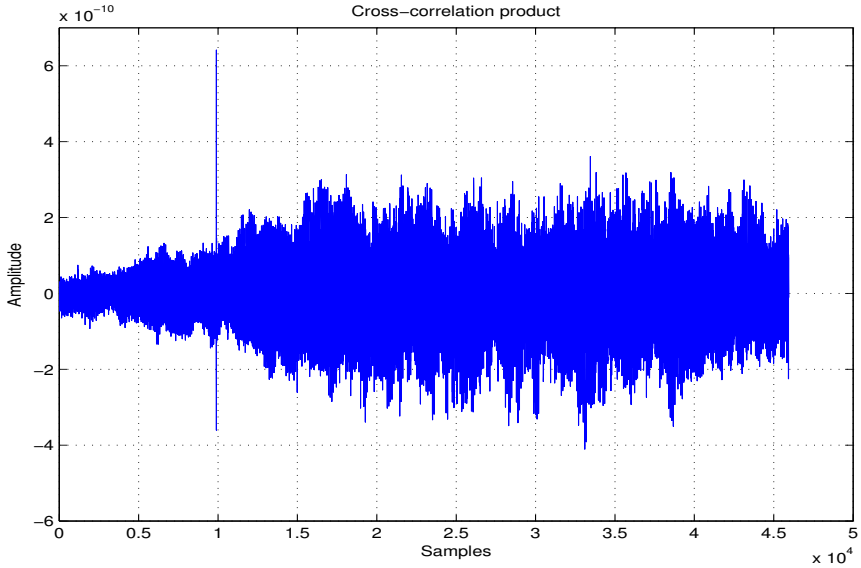

Figure 7.2.1: Cross-correlation product.

Our results lead to the following assumptions. At 128 kbits/sec bit rate, we expect similar performance to the one we get when codec is not present. At 64 kbits/sec, the system performance is expected to be degraded by a factor of approximately 3 dB. In either case, the pilot is not audible.

BPSK pilot robustness

The same process was also followed for the BPSK pilot. In this case, although the pilot is stronger, the encoding/decoding process results to increased distortions, in relevance to the DS case. It is obvious that the choice of the passband is critical, in order to minimize these effects. Hence, there is a trade-off between the audibility and the encoding/decoding effects. In this case, performing cross-correlation with the original signal to estimate the delay value by locating the peak of the resulting product, we found that the peak value $R_{\hat{s}s}(\Delta)$ was approximately equal to $0.8\,R_{ss}(\Delta)$, by setting 128 kbits/sec bit rate. When 64 kbits/sec bit rate is used, in each trial we observed that the peak is buried and the waveform $R_{\hat{s}s}(\tau)$ contains just noise, when only the pilot is transmitted. In addition, for the latter

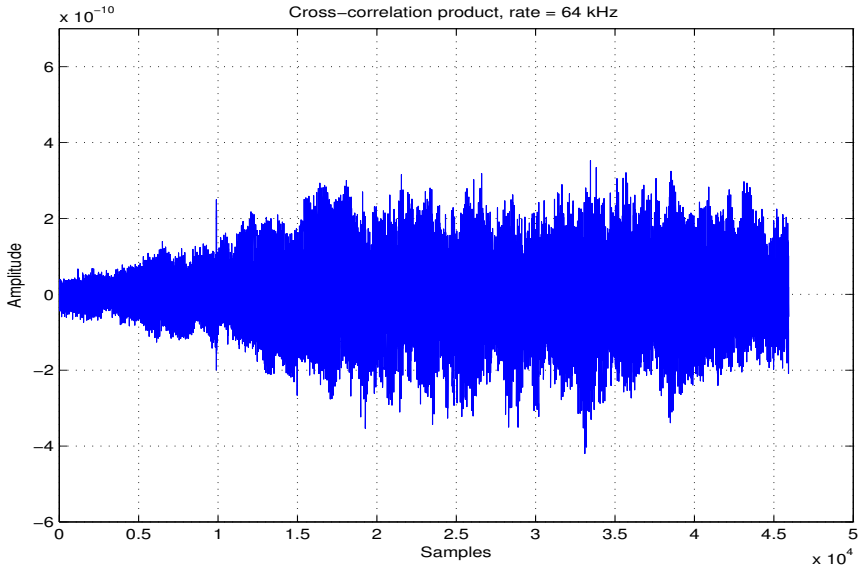

Figure 7.2.2: Cross-correlation product, after encoder/decoder with bit rate 64 kbits/sec.

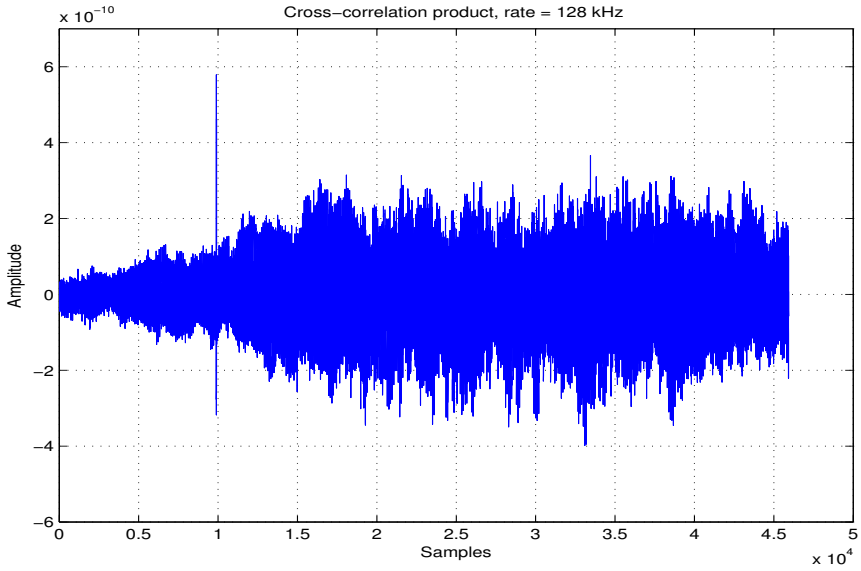

Figure 7.2.3: Cross-correlation product, after encoder/decoder with bit rate 128 kbits/sec.

case, the received pilot was audible. Typical RMSE values were 0.0055 and 0.0019 for 64 and 128 kbits/sec bit rates, respectively. Then, the file that was including the complex audio signal, was evaluated. Note that the track in both cases of interest, is the same. The music signal was used identically (assuming no normalization) and, hence, the SNR value for the DS case was $\text{SNR}_{\text{DS,dB}} = -57$ dB and in this case was $\text{SNR}_{\text{BPSK,dB}} = -16.66$ dB, which is straightly calculated according to Section 6.3 results. In the figures below, we present the cross-correlation product of a typical implementation assuming the co-existence of the music sample with the pilot. At the received signal, we apply (a) no coding, (b) encoding/decoding with bit rate 64 kbits/sec and (c) encoding/decoding with bit rate 128 kbits/sec, in Fig. 7.2.4, 7.2.5 and 7.2.6, respectively. When 64 kbit/sec bit rate was used, the spectrum of the received pilot when it is transmitted on silence, is distorted and extended, in contrast to the case where music signal is also present. Obviously, the OPUS codec processing is responsible for these observations and as depicted in Fig. 7.2.5, the peak can be found. In this case, though, the musical experience of user is deteriorated, since the pilot becomes audible.

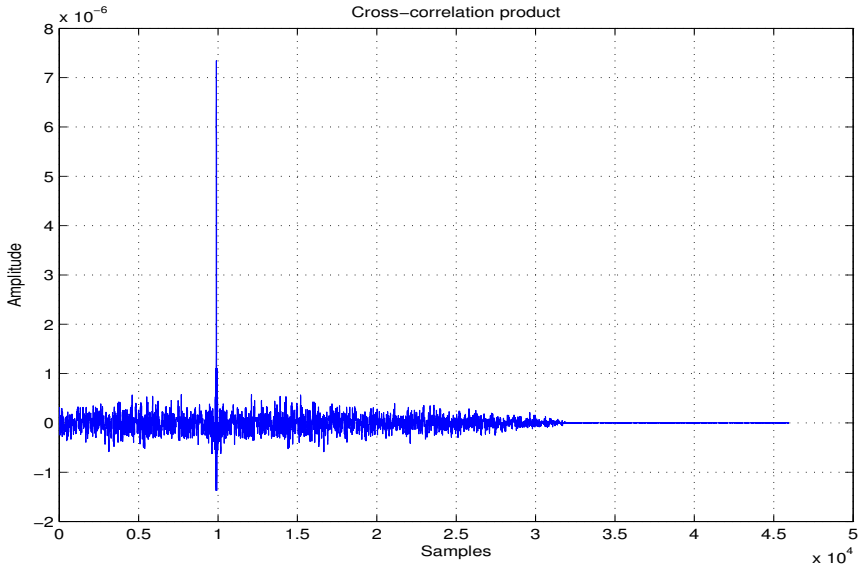

Figure 7.2.4: Cross-correlation product.

7.3 Generalizing at sampling frequencies for high definition audio signals

In this Section we provide details concerning the handling of the pilot signal, when higher sampling frequencies are used in professional audio devices (i.e. 48, 96 and 192 kHz). For both cases of the signal design, in order to comply to these parameters and achieve similar performance to the depicted through this study, the pulse duration should be reduced by a factor of $F_s/44.1$, where F_s denotes the new sampling frequency.

When the BPSK technique is used, the only parameter that has to be re-defined is the center frequency. The pilot spectrum has to be transferred at the highest possible band, with respect to the corresponding available bandwidth, given by $(1 + a)/T$. In Fig. 7.3.1, considering the preceding observations, the pulse in time domain and the spectrum in baseband and passband are shown. Note, that this graph is, basically, identical to the Fig. 3.1.11 on different axis.

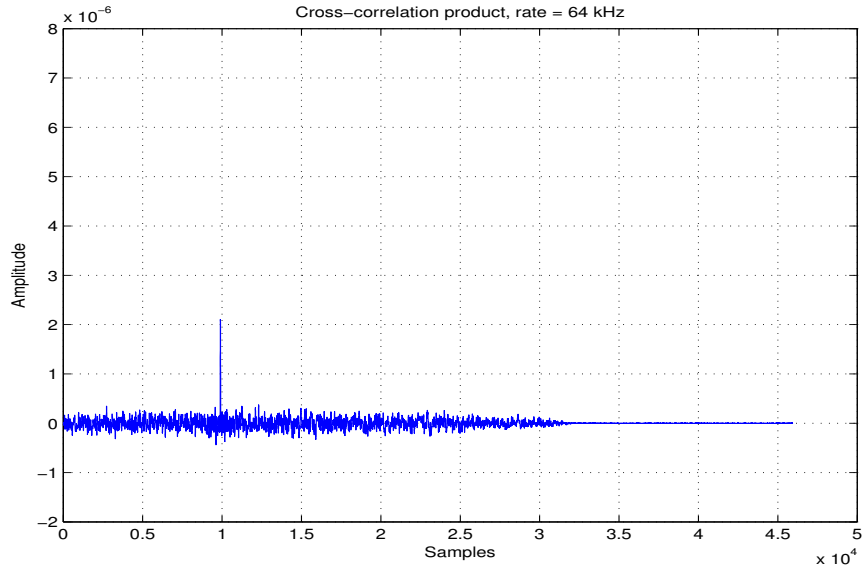

Figure 7.2.5: Cross-correlation product, after encoder/decoder with bit rate 64 kbits/sec.

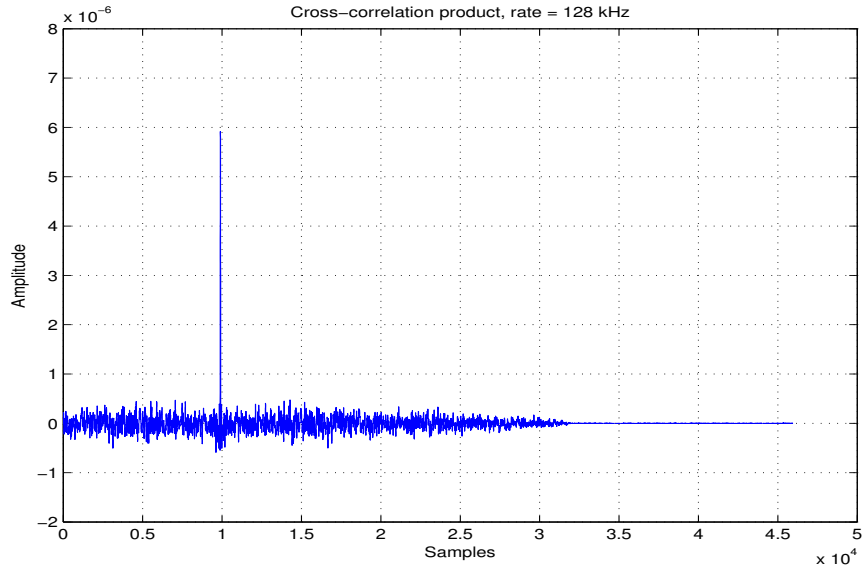

Figure 7.2.6: Cross-correlation product, after encoder/decoder with bit rate 128 kbits/sec.

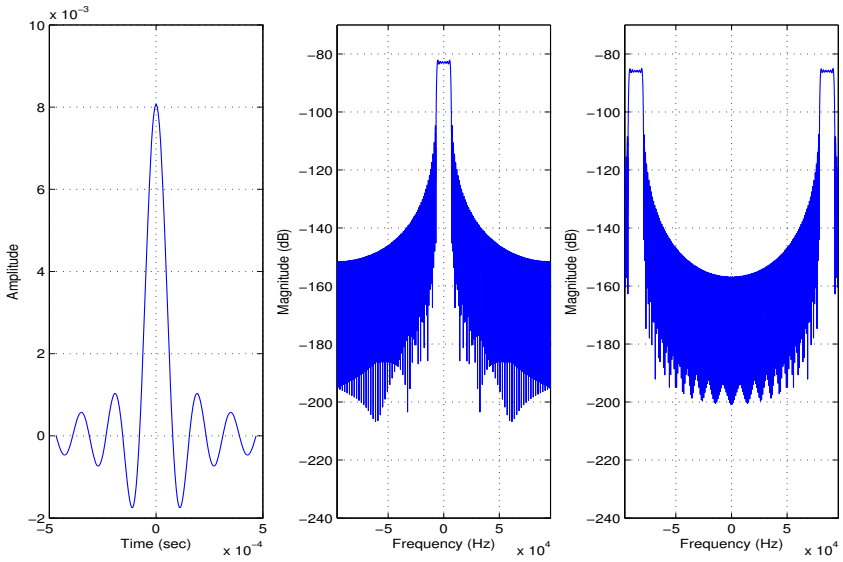

Figure 7.3.1: SRRC assuming $F_s = 192$ kHz and appropriate normalization.

When the DS technique is used, identical number of samples should be set for the pulse generation. When higher sampling frequencies are assumed, a straight calculation of the resulting pulse is obtained, considering the process described in Section 3.1.2.1. The number of samples remains constant and equal to 13, and the only term that is modified is the sampling frequency. Of course, the time duration of the pulse is reduced, in the sense of seconds, but, the spectrum is designed identically, with respect to the new frequency axis. In Fig 7.3.2, we present the new pulse, assuming $F_s = 192$ kHz in both the time and in frequency domain, whose shape is identical to the pulse of Fig. 3.1.10, on different axis.

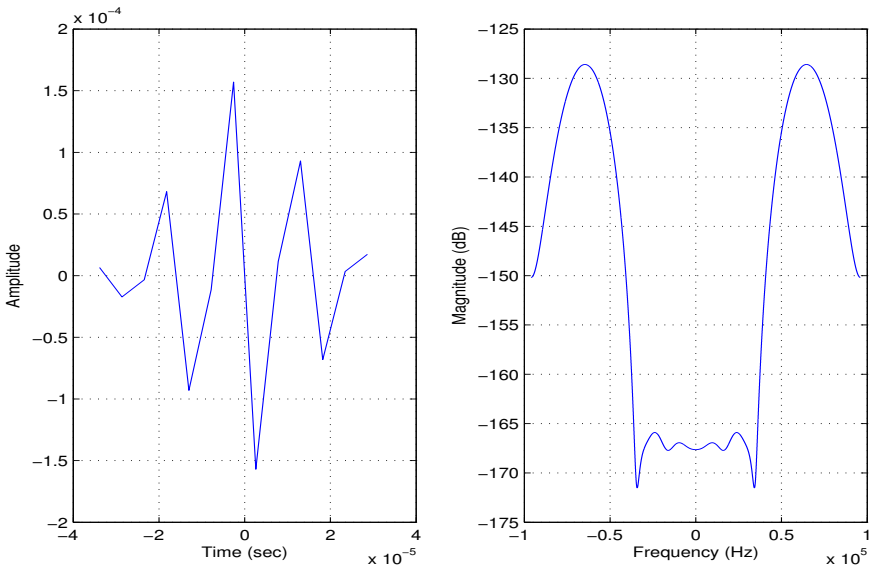

Figure 7.3.2: 15th derivative of Gaussian assuming $F_s = 192$ kHz and appropriate normalization.

7.4 Applications

In Fig. 4.0.1 a typical diagram is presented, where the proposed techniques and signals could be adopted. In general, the pilot can be injected at a specified initial stage of signal's transmission in order to measure the time delay that a following real-time system inserts during its processing, to audio signals. In fact, in specific cases the delay can be estimated with accuracy in the order of $3.78 \cdot 10^{-6}$ seconds. Focusing on internet applications, the Fig. 7.4.1 depicts a more detailed version.

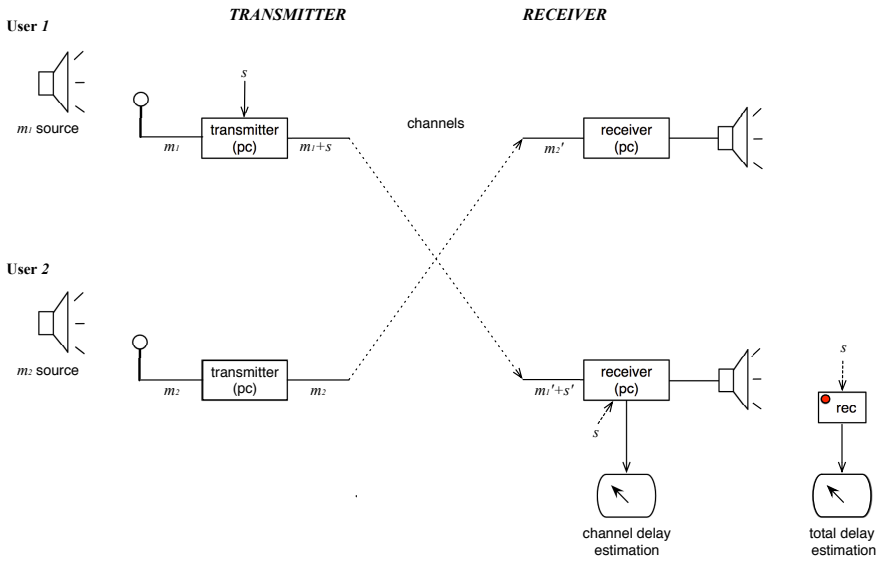

Figure 7.4.1

Once again, two delay measurements can be obtained at the receiver, given the time reference of the pilot injection at the transmitter side. The first one concerns the "channel delay estimation", potentially with encoding and decoding of the signal, and the second one regards to the "total delay estimation", where the receiver processing (play-out buffer management, operating system delays etc.) are also included.

In Fig. 7.4.2, a generalized version of the previous system is shown. The audio signals are transmitted from more than one locations. For each music source, a computer is placed that captures the corresponding acoustic signal. Each transmitter could inject a pilot signal that is unique for each user. By definition, the pilot signals are orthogonal in order to be distinguished each of them from the others, in the case of simultaneous transmissions. Furthermore, we can estimate delays of the same streams on different terminals, which enables the mix phase difference measurement.[3]

Mix phase difference

Beyond the typical case of use for the described system, which is the delay estimation of a channel, and/or, the delay that is inserted by a real-time system that processes audio signals, the defined system could be also used in order to evaluate a newly introduced parameter which could be proved critical in a Network Music Performance. The "maximum mix phase difference", *was defined by Eleftheriadis*, and it may introduce further

[3]The difference of delay of the same audio signals which are transmitted through different channels to different terminals.

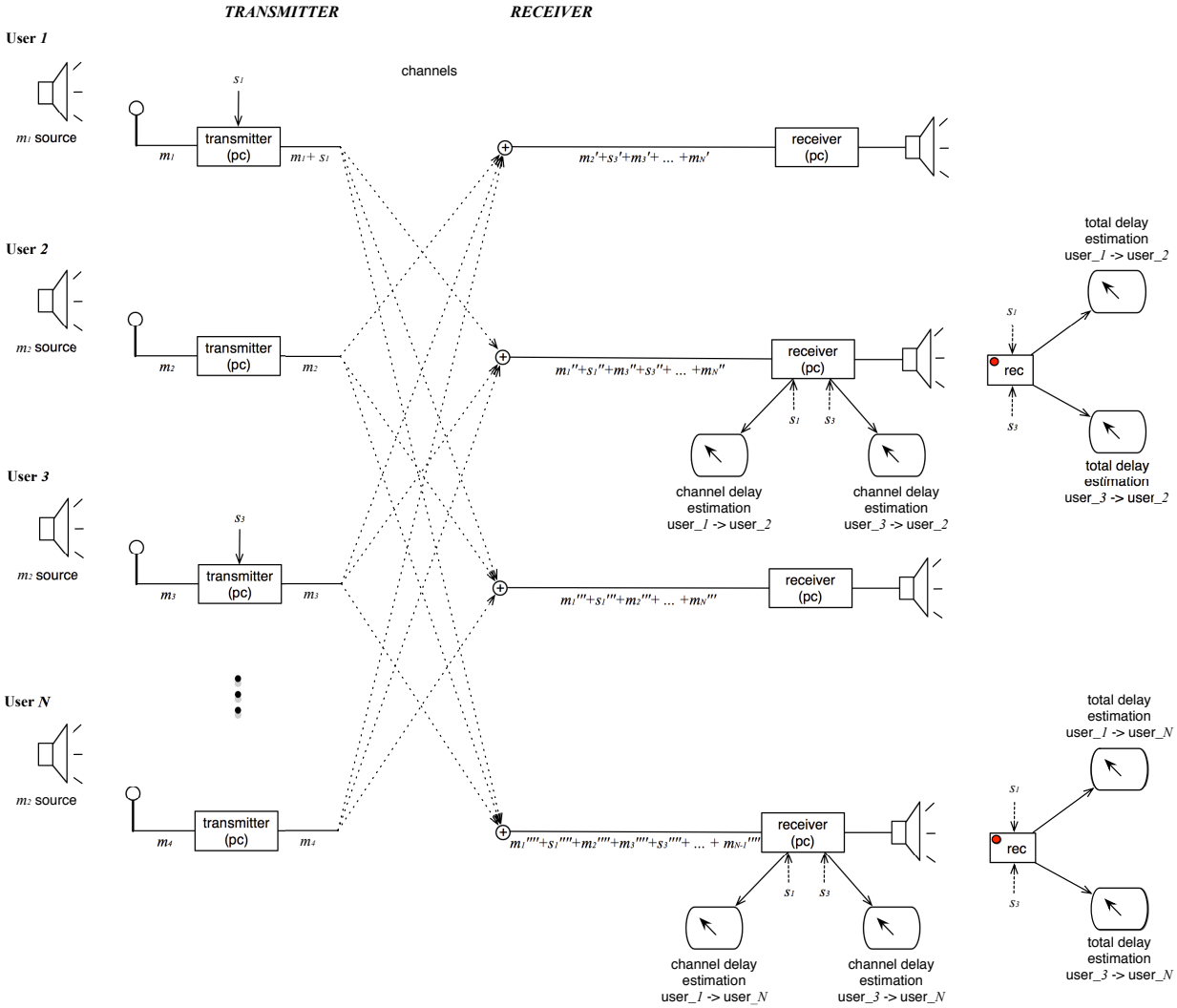
TRANSMITTER
RECEIVER
User 1
User 2
User 3
User N
channels
m1 source
m2 source
m2 source
m2 source
transmitter (pc)
transmitter (pc)
transmitter (pc)
transmitter (pc)
receiver (pc)
receiver (pc)
receiver (pc)
receiver (pc)
channel delay estimation user_1 -> user_2
channel delay estimation user_3 -> user_2
channel delay estimation user_1 -> user_N
channel delay estimation user_3 -> user_N
rec
rec
total delay estimation user_1 -> user_2
total delay estimation user_3 -> user_2
total delay estimation user_1 -> user_N
total delay estimation user_3 -> user_N

Figure 7.4.2

qualitative restrictions in order to maintain the comfort of musicians who collaborate in such a network. Specifically, according to Fig. 7.4.3, let four musicians be placed in four different places, which correspond to the four nodes of the graph.

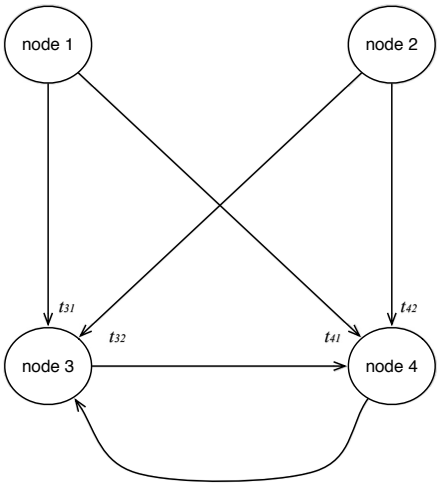

Figure 7.4.3

The audio signal from each node is received at different time instances, by the three other nodes. Note that only the streams of interest are included in the previous graph, for simplicity reasons. We are interested in the quantification of the time difference of the same audio signals that are received from two nodes, enabling the collaboration between the latter. For example, at node 3, the audio signals from node 1 and 2 are received at time instances of t_{31} and t_{32}, respectively, forming a phase mixing at node 3, which is given as $\Delta_3 = t_{31} - t_{32}$. Note that we could define, equivalently, $\Delta_3 = t_{32} - t_{31}$. In addition, the audio signals from the same transmitters, are received from node 4, at time instances t_{41} and t_{42}. The desirable parameter is the difference between those two mix phases, which is defined as follows.

$$D \equiv \Delta_3 - \Delta_4 = [t_{31} - t_{32}] - [t_{41} - t_{42}]$$

At this spot, an upper and a lower bound should be defined, $L \leq D \leq U$, which permit the synchronization of nodes 3 and 4. In fact, by re-arranging the previous quantity, we get

$$D \equiv \Delta_3 - \Delta_4 = [t_{31} - t_{41}] - [t_{32} - t_{42}]$$

which defines the inter-arrival difference, in the sense of the delay between successful arrivals of the same signal (node 1) to different nodes (3 and 4). In this case, we set as our benchmark the time when the signal reaches the node 3. The D is the difference between inter-arrivals arising from the corresponding scheme that is assumed. The quantity can be named as “network delay diffusion difference”, in the sense that each term of the difference indicates how the delay from the particular node varies as the information reaches to the various recipients. Hence, the parameter of interest has a dual substance. Under the musical aspect, it regards to the difference of the phase mixing, and in terms of network, the inter-arrival difference. Obviously in both cases the amount may be an indication of synchronization capabilities of nodes 3 and 4, in this case.

8. Conclusions and Future Work

We have presented two systems for accurately measuring the delay that is inserted to audio signals, during their processing under real-time systems, or their transmission through net, using inaudible pilot waveforms. Noteworthy parameters that affect the pilot transmission and design, such as the generalized Power Spectral Density of music and the Absolute Threshold of Hearing, were considered. We concluded in two different approaches for the signal construction, that ensure: (a) simplicity in generation, (b) inaudibility, and (c) high performance at the delay estimation. On one hand, we show how to select an appropriate pilot pulse and use spread-spectrum techniques to both make the pilot inaudible as well as increase its ability to be detected at the receiver, and on the other hand, a single-carrier technique involving typical pulses in cooperation with an appropriate selection of the center frequency to transfer the signal spectrum in a high band. Comparison with typical implementations that have identical structure was performed for both cases, as well as a comparison between them. Specifically, for the first approach other candidate pulses such as the Dirac, rectangular and Gaussian were tested and for the second one, randomly chosen center frequencies were selected as the baseline models. We concluded, that the use of the BPSK technique is a more consistent approach that leads to enhanced robustness, since the used pilot is much more stronger, and remains inaudible as well. However, the performance is essentially decreased when delay estimation with sub-sample accuracy is required. The use of the 15th derivative of the Gaussian pulse with the DS spread spectrum technique leads to high-performance, in a less noisy environment, though. In addition, the suggested technique for sub-sample accuracy seems to overcome the expected decrease under the presence of music samples and an accuracy of 3.78 μsec seems to be achieved, almost with identical performance. Finally, preliminary results show that the first implementation is more robust under the OPUS codec, and in any case of encoding/decoding, with respect to the "Live Distributed Music Performances" specifications, the waveform after processing remains silent. In contrast, the BPSK signal does not guarantee the same behavior, except for the case where the highest bit-rate 128 kbits/sec is set. Hence, the selection between the proposed techniques, depends on the application of use and the noise power level. In any case, the simplicity of injecting and distinguishing, the pilot, in and from the audio signal, is achieved.

Further work should be done in order to optimize the using pulses, under each case of the signal design. Concerning the *BPSK* technique, it is obvious that there is a trade-off between the accuracy and the audibility of the pilot. Hence, an optimized operation point should be found, by decreasing the pulse duration, which in turn corresponds to the extension of the signal bandwidth. Under these circumstances, the energy per pulse should be reduced, since the pilot will be probably audible. Note also, that this optimal operation point will be probably different under various cases of the co-existing audio signal. Thus, either a worst-case version of the pilot, that meet the practical constraints is required, or a bank of pilots should be used, where the selection of the appropriate signal is triggered after the evaluation of the music signal that is present and somehow the receiver gets the information for the expected receiving waveform (for example, initial test of all the pilots from the bank). Concerning the *DS*, after the issues we experienced in this study, we concluded that the optimized pulse for use should have the minimum spectrum located to the highest possible frequency band (quick varying pulse), its auto-correlation function should be a Dirac pulse (or a "similar" shape), and under upsampling operation it should not loose these properties! In other terms, a middle solution that could achieve enhanced robustness in relevance to the DS technique, and higher performance

for the delay estimation with sub-sample accuracy in contrast to the BPSK, will be the optimal. The aforementioned approaches are expected to succeed these goals.

Another idea that could enhance the use of the system is its extension to an on-line mode. In this way the delay of the channel can be estimated synchronously and lots of applications could, potentially, benefit by such a measure. Of course, it is indicated that the synchronization of clocks between two terminals, should be as accurate as possible. To overcome this problem, a potential coding technique that carries the information that is needed to synchronize transmitter and receiver, could also be adopted.

This system, finally, could be adapted under different level of existing noise. In terms, we could estimate the level of existing noise in channel, and based on that observation, the pilot to trigger specific actions in order to increase, or decrease its energy and/or duration. The decision of transmitting or not the pilot, should be determined under a specified threshold for the level of noise.

Abbreviations - Acronyms

ATH Absolute Threshold of Hearing

AWGN Additive White Gaussian Noise

BPSK Binary Phase Shift Keying

CDF Cumulative Distribution Function

dB Decibel

DS Direct-Sequence

Eq. Equation

Eq. Equation

Fiq. Figure

F.T. Fourier Transform

Hz Hertz

i.e. In Explain

i.i.d. Independent and Identical Distributed

ISI InterSymbol Interference

kHz Kilohertz

LTI Linear Time Invariant

msec Milliseconds

NMP Network Music Performance

PCM Pulse Coding Modulation

PDF Probability Distribution Function

PLL Phase-Lock Loop

PSD Power Spectral Density

RMSE Root-Mean Square Error

sec Seconds

SNR Signal-to-Noise Ratio

SPL Sound Pressure Level

SRRC Square-Root Raised Cosine

SWL Sound Power Level

TH Time-Hopping

WSS Wide-Sense Stationary

Appendices

Appendix I

Relationship between auto-correlation functions of input and output

We present the proof of Eq. 4.5 and 4.6, according to [10], in Section 4.3.2. Let, $x(t)$, be AWGN and $h(t)$ be a deterministic waveform under which cross-correlation is performed. Then, the output $y(t)$ is given as

$$y(t) = x(t) \star h(t) = x(t) * h(-t) = \int x(\tau)\, h(t+\tau)\, \mathrm{d}\tau = \int h(\tau)\, x(t+\tau)\, \mathrm{d}\tau$$

using Eq. 4.3. Then, the auto-correlation product between the input $x(t)$ and the output $y(t)$, is given below.

$$\begin{aligned} R_{yx}(t_1, t_2) &= E\{y(t_1)\, x(t_2)\} \\ &= E\left\{ \int h(\tau)\, x(t_1+\tau)\, \mathrm{d}\tau\, x(t_2) \right\} \\ &= \int h(\tau)\, E\{x(t_1+\tau)\, x(t_2)\}\, \mathrm{d}\tau \\ &= \int h(\tau)\, R_{xx}(t_1+\tau, t_2)\, \mathrm{d}\tau \\ &= \int h(\tau)\, R_x(t_1+\tau-t_2)\, \mathrm{d}\tau \\ &= \int h(\tau)\, R_x([t_1-t_2]+\tau)\, \mathrm{d}\tau \end{aligned}$$

where we used that $x(t)$ is a WSS process (by definition of AWGN) and $R_{xx}(k, m) = R_x(k-m)$. According to the last expression of $R_{yx}(t_1, t_2)$, it is obvious that the cross-correlation of $x(t)$ and $y(t)$ depends only on the difference $[t_1 - t_2]$. In addition, the mean value of the input $\mu_x = 0$ and $\mu_y = \mu_x \int h(t)\, \mathrm{d}t = 0$, indicating that $x(t)$ and $y(t)$ are jointly WSS. Thus, $R_{yx}(t_1, t_2) = R_{yx}(t_1 - t_2)$. Finally, by setting $u = t_1 - t_2$, we get

$$\begin{aligned} R_{yx}(u) &= \int h(\tau)\, R_x(u+\tau)\, \mathrm{d}\tau \\ &= R_x(u) * h(-u) \end{aligned}$$

The auto-correlation of the output $y(t)$ is given below.

$$\begin{aligned} R_{yy}(t_1, t_2) &= E\{y(t_1)\, y(t_2)\} \\ &= E\left\{ y(t_1) \int h(\tau)\, x(t_2 + \tau)\, \mathrm{d}\tau \right\} \\ &= \int E\left\{ y(t_1)\, x(t_2 + \tau) \right\} h(\tau)\, \mathrm{d}\tau \\ &= \int R_{yx}(t_1, t_2 + \tau)\, h(\tau)\, \mathrm{d}\tau \\ &= \int R_{yx}([t_1 - t_2] - \tau)\, h(\tau)\, \mathrm{d}\tau \end{aligned}$$

where we used that $y(t)$ and $x(t)$ are jointly WSS and, hence, $R_{yx}(k, m) = R_{yx}(k - m)$. At this spot, we observe that $R_{yy}(t_1, t_2)$ depends on the difference $[t_1 - t_2]$ and $\mu_y = 0$, indicating that $y(t)$ is a WSS process. Thus, $R_{yy}(t_1, t_2) = R_y(t_1 - t_2)$. By setting $u = t_1 - t_2$ in the last equality and after the change of variables, we get

$$\begin{aligned} R_y(u) &= \int h(\tau)\, R_{yx}(u - \tau)\, \mathrm{d}\tau \\ &= R_{yx}(u) * h(u) \end{aligned}$$

Finally, we get the desirable result:

$$R_y(u) = R_x(u) * h(-u) * h(u)$$

Appendix II

Relationship of system performance assuming different noise types: AWGN and Pink noise

We present a coarse proof of our findings in Section 6.1, concerning the system performance under AWGN and Pink noise. Let $n_1(t)$ be AWGN and $n_2(t)$ Pink noise with the same power, in our system. It is known that the PSD of $n_1(t)$ is flat and we assume that it is equal to a. Then, we get

$$P = \int_{-22.05\cdot 10^3}^{22.05\cdot 10^3} a \, \mathrm{d}f = 2 \cdot 22.05 \cdot 10^3 \cdot a = 44100 \cdot a$$

The power of $n_2(t)$ in the frequency domain is decreased by 3 dB, in each octave. Hence, the power as the integral over its PSD, can be calculated as follows.

$$\begin{aligned} P &= 2\int_{2^0}^{2^1} b \, \mathrm{d}f + 2\int_{2^1}^{2^2} b \cdot 2^{-1} \, \mathrm{d}f + \ \ldots \ + 2\int_{2^{13}}^{2^{14}} b \cdot 2^{-13} \, \mathrm{d}f + 2\int_{2^{14}}^{22.05\cdot 10^3} b \cdot 2^{-14} \, \mathrm{d}f \\ &= 2 \cdot b + 2 \cdot b \cdot 2^{-1} \, 2 + 2 \cdot b \cdot 2^{-2} \cdot 2^2 + \cdots + 2 \cdot b \cdot 2^{-13} \, 2^{13} + 2 \cdot b \cdot [22.05 \cdot 10^{-3} - 2^{14}] \cdot 2^{-14} \\ &= 2 \cdot b \cdot [14 + 5666 \cdot 2^{-14}] \\ &= 2 \cdot b \cdot 14.3458 \\ &= 28.69 \cdot b \end{aligned}$$

We have set equal the powers of both noises. Hence, we get the following equality between a and b variables, which are related to the PSDs of AWGN and Pink noise, respectively.

$$28.69 \cdot b = 44100 \cdot a \Longleftrightarrow b = \frac{44100}{28.69} \cdot a$$

Our pilot signal is located, approximately, to frequencies > 8 kHz. We assume that Pink noise's PSD value is constant and equal to $b \cdot 2^{-13}$ over the range $[8192, \, 2205]$, which is not totally accurate, since it is constant and equal to $b \cdot 2^{-13}$ over the range $[2^{13}, \, 2^{14}]$, and equals $b \cdot 2^{-14}$ within $[2^{14}, \, 22050]$. Thereby, we will get an upper limit of the increase that we expect under Pink noise, in relevance to AWGN. Finally, under our assumptions, at the frequency band of interest the Pink noise PSD will be $\simeq 2^{-13} \cdot b = 2^{-13} \cdot \frac{44100}{28.69} \cdot a = 0.154 \cdot a$, which corresponds to $\simeq -8$ dB. Hence, we expect an increase in system performance by 8 dB, at best, which is validated from our experimental results.

Bibliography

[1] V. Alexiou and A. Eleftheriadis. “real-time high-resolution delay estimation in audio communication using inaudible pilot signals”. In *Submitted to the 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, 16 December 2013.

[2] A. Barbosa. *“Displaced Soundscapes”*. PhD thesis, Pompeu Fabra University, Barcelona, Spain, 2006.

[3] J. S. Bendat and A. G. Piersol. *“Random Data Analysis and Measurement Procedures”*. Wiley-Interscience, 3rd edition, 2000.

[4] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. “Techniques for Data Hiding”. *IBM System Journal*, 35:313–336, 1996.

[5] L. Boney, A. Tewfik, and K. Hamdy. “Digital Watermarks for Audio Signals”. In *Proceedings of 3rd International Conference on Multimedia Computing and Systems*, pages 473 – 480, Hiroshima, June 1996.

[6] R. Garcia. “Digital Watermarking of Audio Signals using a Psychoacoustic Model and Spread Spectrum Theory”. In *Proceedings of 107th Convention of the Audio Engineering Society (AES)*, New York, NY, USA, 24-27 September 1999.

[7] K. Gentile. “Raised Cosine Filters”. *RF Design*, April 2002.

[8] K. Goenka and P. Patil. “Overview of Audio Watermarking Techniques”. *International Journal of Emerging Technology and Advanced Engineering*, 2(2):67–70, February 2012.

[9] G. B. Khatri and D. S. Chaudhari. “Digital Audio Watermarking Applications and Techniques”. *International Journal of Electronics and Communication Engineering and Technology (ICEJET)*, 4(2):109–115, March-April 2013.

[10] A. Liavas. *“Communication Systems II”*. Technical University of Crete, School of Electronic and Computer Engineering, 2009.

[11] M. Martinson and B. Martinson. http://www.beginband.com.

[12] J. G. Proakis and M. Salehi. *“Communication Systems Engineering”*. Prentice Halls, 2002.

[13] J. Valin and K. Vos. “Requirements for an Internet Audio Codec”. RFC 6366, IETF, August 2011.