

**DAS** Departamento de Automação e Sistemas  
**CTC** **Centro Tecnológico**  
**UFSC** Universidade Federal de Santa Catarina

# **Equalização Adaptativa Aplicada a um Sistema de Marcação de Áudio Digital**

*Monografia submetida à Universidade Federal de Santa Catarina  
como requisito para a aprovação da disciplina:*

***DAS 5511: Projeto de Fim de Curso***

***Samir Pauli Fiates***

*Florianópolis, Março de 2013*

# **Equalização Adaptativa Aplicada a um Sistema de Marcação de Áudio Digital**

***Samir Pauli Fiates***

Esta monografia foi julgada no contexto da disciplina  
**DAS 5511: Projeto de Fim de Curso**  
e aprovada na sua forma final pelo  
**Curso de Engenharia de Controle e Automação Industrial**

***Prof. Joceli Mayer, Ph.D.***

---

## Resumo

A transmissão de áudios marcados digitalmente pelo meio acústico ainda é um tema limitado na literatura. Poucos trabalhos nesta área mostram bons resultados frente a este tipo de transmissão. Mesmo as técnicas de estado da arte, apresentam limitações, na maioria das vezes relacionadas à transparência da marca.

Neste trabalho, são pontuados os problemas causados pela reverberação acústica no áudio transmitido, como a ISI e a distorção do espectro de frequência. Uma técnica de equalização adaptativa é proposta para reduzir os efeitos danosos do canal acústico sobre o áudio marcado e para melhorar a taxa de decodificação da informação inserida. Pretende-se através do aumento da robustez da marcação, conseguir reduzir energia da marca d'água e torná-la mais transparente, sem abrir mão de boas taxas de decodificação.

A equalização é feita baseada na proposta de [1], através de uma aproximação do canal de transmissão convolutivo e sua posterior inversão, caracterizando uma equalização *zero forcing*. A estimação do canal é realizada em pequenos intervalos de tempo por filtragem adaptativa, utilizando-se da sequência de sincronia da própria marca d'água para treinar o filtro adaptativo. Três algoritmos de filtragem adaptativa são testados: LMS, NLMS e AFA. Testes em ambiente simulado e real são apresentados comparando o desempenho da decodificação do sinal não equalizado com os sinais equalizados utilizando-se dos algoritmos de filtragem adaptativa citados.

**Palavras Chave:** marca d'água digital, processamento de áudio digital, comunicação digital, canal acústico, equalização, filtragem adaptativa.

## Abstract

The transmission of watermarked audio by acoustic channel is still a limited theme in literature. Few studies in this area have shown good results against this type of transmission. Even the state of the art techniques have limitations mostly related to the transparency of the mark.

This work presents problems caused by acoustic reverberation in the audio broadcast as the ISI and distortion of the frequency spectrum. An adaptive equalization technique is proposed to reduce the harmful effects of channel on watermarked audios and to improve the decoding rate. It is intended by increasing the watermarking robustness to reduce the watermark's energy and to boost its transparency without compromising good decoding rates.

The equalization is done based on the method proposed in [1] using an approximation of the transmission convolutional channel and its subsequent reversal, characterizing an zero forcing equalizer. The channel estimation is performed in small intervals of time by adaptive filtering using the watermark's synchronization sequence itself to train the adaptive filter. Three adaptive filtering algorithms are tested: LMS, NLMS and AFA. Tests on simulated and real environment are presented comparing the decoding rates of a non equalized signal with signals that were equalized using adaptive filtering algorithms cited.

**Keywords:** digital watermarking, digital audio processing, digital communication, acoustic channel, equalization, adaptive filtering.

# Sumário

<b>Abreviaturas</b>	<b>viii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Problema . . . . .	3
1.2 Estrutura do relatório . . . . .	4
<b>2 A Certimarca e o contexto do trabalho</b>	<b>6</b>
<b>3 Áudio digital</b>	<b>8</b>
3.1 Marcação em áudio digital . . . . .	9
3.1.1 Áreas de pesquisa . . . . .	9
3.1.1.1 Transparência . . . . .	10
3.1.1.2 Segurança . . . . .	10
3.1.1.3 Capacidade . . . . .	10
3.1.1.4 Robustez . . . . .	10
3.1.1.5 Algoritmos cego ou não-cego . . . . .	10
3.1.1.6 Complexidade e custo computacional . . . . .	11
3.1.1.7 Eficácia da inserção . . . . .	11
3.1.2 Técnicas de marcação em áudio digital . . . . .	12
3.1.3 LSB – <i>Least significant bit</i> . . . . .	12
3.1.4 Ocultamento de “eco” . . . . .	13
3.1.5 Codificação de fase . . . . .	14
3.1.6 Espalhamento espectral . . . . .	15
3.2 Ataques e Processamentos Comuns . . . . .	16

<b>4</b>	<b>O canal acústico e a equalização adaptativa</b>	<b>18</b>
4.1	Canal Acústico . . . . .	18
4.1.1	Estado da arte - Marca d'água digital no canal acústico . . . . .	21
4.2	Equalização . . . . .	21
4.2.1	Filtragem Adaptativa . . . . .	23
4.2.1.1	Filtro de Wiener . . . . .	26
4.2.1.2	LMS - <i>Least Mean Square</i> . . . . .	26
4.2.1.3	NLMS - <i>Normalized Least Mean Square</i> . . . . .	28
4.2.1.4	AFA - <i>Adaptive Filter with Averaging</i> . . . . .	28
<b>5</b>	<b>Marca d'água digital utilizando equalização adaptativa</b>	<b>29</b>
5.1	A marcação digital de áudio . . . . .	30
5.1.1	Incorporador . . . . .	31
5.1.2	Extrator . . . . .	32
5.2	Equalização com filtro <i>zero forcing</i> . . . . .	33
5.2.1	Estimação do canal . . . . .	33
5.2.2	Equalização do canal . . . . .	36
<b>6</b>	<b>Implementação</b>	<b>37</b>
6.1	Incorporador da marca . . . . .	37
6.2	Canal de transmissão . . . . .	38
6.3	Equalização <i>zero forcing</i> . . . . .	40
6.4	Extrator da marca . . . . .	43
<b>7</b>	<b>Resultados</b>	<b>45</b>
7.1	Variações do canal acústico . . . . .	45
7.2	Robustez e transparência . . . . .	47
7.3	Teste em condições reais . . . . .	49

<b>8 Conclusões e Perspectivas</b>	<b>52</b>
8.1 Perspectivas de trabalhos futuros . . . . .	54
<b>Referências Bibliográficas</b>	<b>55</b>

## Abreviaturas

AD	Analógico-Digital
AFA	Filtragem Adaptativa com Média ( <i>Adaptive Filtering with Averaging</i> )
dB	Decibel
BER	Taxa de Erro de <i>Bit</i> ( <i>Bit Error Rate</i> )
bps	<i>Bits</i> por segundo
DA	Digital-Analógico
DCT	Transformada Discreta de Cosseno ( <i>Discrete Cosine Transform</i> )
DFE	Equalização com Decisão Realimentada ( <i>Decision Feedback Equalizer</i> )
DFT	Transformada Discreta de Fourier ( <i>Discrete Fourier Transform</i> )
DWT	Transformada Discreta de <i>Wavelet</i> ( <i>Discrete Wavelet Transform</i> )
FFT	Transformada Rápida de Fourier ( <i>Fast Fourier Transform</i> )
FIR	Resposta ao Impulso com Duração Finita ( <i>Finite Impulse Response</i> )
Hz	Hertz
IFFT	Transformada Rápida de Fourier Inversa ( <i>Inverse Fast Fourier Transform</i> )
IIR	Resposta ao Impulso com Duração Infinita ( <i>Infinite Impulse Response</i> )
ISI	Interferência Intersimbólica
LCM	Mapeamento em Coordenadas Logarítmicas ( <i>Log Coordinate Mapping</i> )
LMS	Mínimos Quadrados Médios ( <i>Least Mean Squares</i> )
LSB	<i>Bit</i> menos significativa ( <i>Least Significant Bit</i> )
MMSE	Erro Quadrático Médio Mínimo ( <i>Minimum Mean Square Error</i> )
MSE	Erro Quadrático Médio ( <i>Mean Square Error</i> )
NLMS	Mínimos Quadrados Médios Normalizado ( <i>Normalized Least Mean Squares</i> )
PSD	Densidade de Potência Espectral ( <i>Power Spectral Density</i> )
RICE	Estimação de Canal com Redução da Interferência ( <i>Reduced Interference Channel Estimation</i> )
SNR	Relação de Sinal-Ruído ( <i>Signal-Noise Rate</i> )
SWR	Relação de Sinal-Marca d'água ( <i>Signal-Watermark Rate</i> )
ZFE	Equalizador Forçador a Zero ( <i>Zero Forcing Equalizer</i> )

# Capítulo 1: Introdução

O modo como vemos os acontecimentos do mundo mudou radicalmente nas últimas décadas. A evolução da tecnologia digital proporciona uma agilidade, comodidade e rapidez na troca de dados entre indivíduos ao redor de todo globo terrestre de uma forma nunca antes imaginada. Somos bombardeados a cada minuto com milhares de novas informações que, graças aos avanços tecnológicos, são retratadas com uma enorme riqueza de detalhes. Isto fica ainda mais evidenciado quando considera-se a miniaturização dos computadores, que há 30 anos ocupavam andares inteiros de prédios e recentemente, com a popularização dos *smartphones*, cabem nos bolsos. Ou seja, hoje, pode-se interagir com pessoas e eventos que acontecem a milhares de quilômetros de distância.

As mídias digitais tem papel fundamental neste novo universo cibernético em que vivemos. Grande parte dos usuários da rede mundial de computadores utiliza a *Internet* para visualizar ou compartilhar algum tipo de mídia digital, seja ela escrita ou através de vídeos, áudio e imagens. Notícias de diversos jornais distribuídos nos cinco continentes podem ser lidos ao alcance de um *click*, inclusive com conteúdos exclusivamente digitais como vídeos e áudios *online*. Músicas e filmes podem ter um alcance muito maior devido a sua exposição na grande rede. Além disso, o formato digital facilita a criação, edição e compartilhamento destas mídias sem perdas de qualidade e com um baixo custo. Estas vantagens determinaram a morte de tecnologias analógicas, como a fita VHS e o disco de vinil.

Porém, sob a ótica da proteção de direitos autorais, adulteração indevida de documentos e compartilhamento de informações sigilosas, as mídias digitais ainda apresentam graves falhas, que se tornaram fruto de intenso estudo nos últimos tempos. Dentro da questão de segurança destacam-se três tópicos que são: a integridade dos arquivos, visto que estes podem ser editados digitalmente; confidencialidade, o qual se refere à falta de sigilo da informação, pois os conteúdos disponíveis na grande rede são facilmente acessíveis; e a disponibilidade, já que há possibilidade de cópias e distribuição de mídias de forma não autorizada. A indústria da música e cinematográfica, por exemplo, perde uma grande quantidade de dinheiro todos os anos devido a distribuição e cópia ilegal de músicas e filmes.

Os dispositivos tradicionais de proteção de direitos autorais e de conteúdo já

não são suficientes para conter o avanço dos *hackers*. O combate à pirataria digital se dá através de mandados judiciais para o fechamento dos sites que compartilham estes arquivos. Todavia, esta prática se mostra muito onerosa e com pouca eficácia, dado que qualquer computador pessoal conectado à *Internet* e equipado com um gravador/leitor de mídia é capaz de copiar e distribuir conteúdos sem que os autores recebam qualquer *royalty* pela obra.

Esquemas de proteção antigos, como a inserção de informações no cabeçalho dos arquivos podem ser contornados com a simples edição e transformação do arquivo. O uso de criptografia para restringir o acesso do conteúdo a usuários que possuem uma chave criptográfica correta também é utilizado com frequência. Porém, após decifrado, o conteúdo fica disponível para a cópia e distribuição a critério do usuário portador da chave [2].

Assim, há a necessidade de se adotar um método de proteção que seja intrínseco ao próprio sinal de mídia. Mesmo com a edição ou cópia do conteúdo a informação deve persistir no arquivo. Neste cenário surge a marcação digital de conteúdos como mais uma forma de proteção, a qual, apesar de ainda não solucionar todos os problemas de distribuição, integridade e confidencialidade, permite inserir uma assinatura digital no conteúdo que pode ser utilizado em diversas aplicações como gerenciamento de direitos digitais, monitoramento de difusão, prova de adulteração, entre outras [2].

A informação inserida num sinal digital de mídia é chamada de marca d'água digital. O processo de marcação se dá através da incorporação de uma mensagem que contém as informações que se deseja inserir em um sinal hospedeiro, que pode ser um áudio, vídeo ou imagem. A saída do processo é um sinal que contém a mídia marcada com uma informação. A ideia é que o sinal marcado possa ser copiado, reproduzido, editado e/ou transformado e a marca seja sempre detectável e decodificável por um processo decodificador computacional. O processo decodificador é capaz de receber o sinal com a marca d'água e detectar e extrair a mensagem que foi inserida no sinal hospedeiro. A Figura 1.1 mostra o esquema de inserção e extração da marca.

Comumente técnicas de criptografia e esteganografia são utilizadas em conjunto na inserção da marca d'água digital. A criptografia tem o objetivo de ocultar o significado da mensagem, enquanto a esteganografia se preocupa em ocultar a existência da mensagem. Assim, a marca se torna menos perceptível e mais segura, pois fica escondida no sinal hospedeiro e só o portador da chave de criptografia é capaz de fazer a decodificação.

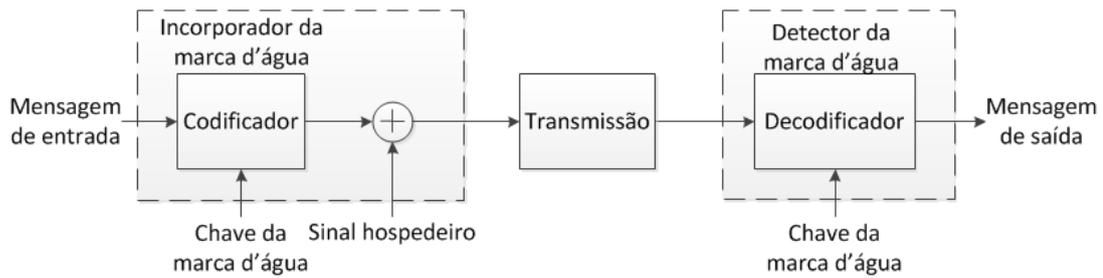


Figura 1.1: Esquema de inserção e extração da marca

Estas características tornam a marcação digital de mídia uma tecnologia extremamente interessante. Além das aplicações na área digital, existe um potencial da marcação de áudio ser utilizado em transmissões via canal acústico. Já que os processos de incorporação, geralmente, levam em consideração as propriedades esteganográficas da marca no sinal hospedeiro. Essa possibilidade abre um grande leque de novas aplicações para marcação de áudios. Portanto, basicamente há duas categorias de aplicações [1]:

- Aquelas orientadas à *copyright* e proteção de propriedade intelectual, que prezam pela robustez e segurança da marca à ataques maliciosos.
- Aquelas relacionadas com a transmissão de dados, que miram na inserção com uma grande capacidade de informação e alta transparência, adicionando valor à conteúdos digitais. Porém não há necessidade de alta robustez à ataque maliciosos, somente a manipulações clássicas de áudio (como filtragem passa-baixa e compressão com perdas).

## 1.1: Problema

A marcação digital em sua essência pode ser visto como uma forma de comunicação. É desejado comunicar uma mensagem do incorporador da marca para o receptor da marcação, passando por um canal de transmissão. É natural, portanto, que o processo de marcação seja modelado semelhante à um sistema deste tipo [3]. A partir do momento que o sinal é processado pelo incorporador ele está sujeito a ataques, que podem ser feitos de forma consciente, por um terceiro com intenções maléficas, ou pelo processamento natural do canal de transmissão do áudio. Estas distorções, muitas vezes, interferem na marcação tornando-as indecifráveis ou removendo a informação que tinha sido previamente inserida.

Há uma gama de estratégias bem desenvolvidas para o uso de marcações em ambientes digitais em que o áudio não sofre modificações ou que é considerado somente a reamostragem e compressão do sinal, que são processamentos comuns neste tipo de ambiente. Por outro lado, são poucos os estudos que consideram a conversão DA/AD e transmissão do áudio pelo meio acústico [4]. O canal acústico apresenta um obstáculo, pois cada ambiente apresenta uma determinada característica e resposta em frequência em cada instante de tempo devido a reverberação das ondas sonoras, ademais há o problema dos ruídos do ambiente que são imprevisíveis e podem assumir diversas formas.

Já existem técnicas de marcação que conseguem sobreviver à transmissão pelo canal acústico [1, 5, 6, 7, 8, 9]. Contudo, a área da marcação digital apresenta um paradigma que relaciona a transparência da marcação, a capacidade de armazenar informação e a robustez da marca. O uso marca d'água robusta, necessária para garantir a integridade da informação ao passar pelo canal de transmissão, em geral, introduz uma considerável degradação na qualidade do som, devido à força da marca em relação ao sinal hospedeiro [10].

Um grande desafio desta área de estudo está em tornar a marcação resistente ao canal acústico de forma a facilitar a extração da informação incorporada ao sinal hospedeiro, porém de forma a minimizar os problemas relacionados a transparência da marca frente ao áudio utilizado como hospedeiro. A equalização do canal acústico, apresentada em [1], é uma alternativa interessante e original para este problema. Propõe-se utilizar a filtragem adaptativa para seguir a estratégia proposta em [1] e assim tentar obter uma marcação que tenha robustez à transmissão pelo canal acústico e apresente um nível de transparência maior que a marcação sem a equalização.

## **1.2: Estrutura do relatório**

No Capítulo 2, é apresentada a empresa na qual o trabalho foi desenvolvido e o contexto no qual esta pesquisa é útil para o desenvolvimento dos produtos e serviços oferecidos pela organização.

No Capítulo 3, é feita uma rápida revisão sobre a marcação digital de áudios, lista-se algumas técnicas de incorporação e extração da marca e seus respectivos pontos fortes e fracos. Além de citar os principais tipos de ataques, maliciosos ou não, que estes áudios marcados podem sofrer.

Em seguida, no Capítulo 4, é feito um aprofundamento sobre o problema e a solução proposta. A transmissão pelo meio acústico de áudios marcados digitalmente é abordada sob o ponto de vista de sistemas de comunicação. Também é feito um embasamento teórico sobre a equalização e as ferramentas utilizadas para se chegar à solução pretendida.

O Capítulo 5, explica conceitualmente o tipo de marcação escolhida e a técnica utilizada para a equalização do sinal transmitido pelo canal acústico.

O Capítulo 6 aborda a implementação da solução e os problemas enfrentados durante este processo.

No Capítulo 7 é apresentado os resultados alcançados com a técnica proposta e a interpretação destes números.

## Capítulo 2: A Certimarca e o contexto do trabalho

A empresa na qual se desenvolveu este trabalho é a Certimarca Soluções Tecnológicas. A organização surgiu em 2009 como uma *spin-off* da Suntech Intelligence, empresa fornecedora de soluções de inteligência em comunicações, conhecida nacionalmente e internacionalmente pelas suas inovações e líder de mercado no Brasil para interceptação legal, retenção de dados e gerenciamento de rede. Em 2011, a Suntech Intelligence se tornou uma empresa do grupo Verint e a Certimarca passou a ser incubada no ParqTec Alfa, em Florianópolis.

A Certimarca oferece soluções para conexão de conteúdos digitais relevantes a produtos e bens de empresas. Com foco no uso de dispositivos móveis para disponibilizar dados e informações por meio de *call-to-actions*. Os produtos da empresa são apresentados em três linhas:

- Certibrand - Soluções voltadas para a checagem de características de produtos, tais como autenticidade, rastreabilidade e qualidade dos produtos, através de códigos 2D associados a cada item. Os produtos desta linha possuem um cliente *WEB* para gestão do conteúdo a ser disponibilizado e outro cliente móvel para a apresentação dos conteúdos relacionados aos produtos rastreados.
- Certidocs - Esta linha de serviços envolve soluções para a integridade, autenticidade e confidencialidade de documentos físicos em papel. Com códigos 2D, marcações “invisíveis” e tarjas criptografadas.
- Certicontent - As marcações de conteúdos digitais, tais como vídeos, áudios e imagens, são os produtos oferecidos nessa linha. Cada marcação é feita sob demanda para cada tipo de aplicação.

Nos últimos meses a equipe de pesquisa esteve envolvida com o desenvolvimento de marcações de áudios com alta robustez destinadas ao *fingerprinting* de áudios confidenciais armazenados em repositório na rede. A marca desenvolvida sobrevive à uma lista grande de ataques, inclusive à conversão DA/AD e transmissão via canal acústico.

A partir do desenvolvimento desta marcação, surgiu o interesse de integrar as marcações digitais de áudio com a linha Certibrand, utilizando marcas d'água trans-

mitidas via canal acústico. Porém, uma condição necessária para a viabilidade deste produto é a elevada transparência da marca para que a marcação não interfira na experiência do usuário final que deseja escutar o áudio com a mais alta qualidade. As marcações desenvolvidas até então, não contemplavam estas características.

Estudos foram feitos e novas técnicas implementadas, porém os trabalhos desenvolvidos nesta área com foco na transmissão da marcação pelo canal acústico são limitados, e até o momento nenhuma técnica testada conseguiu cumprir os requisitos de transparência e/ou robustez necessários para a aplicação desejada.

Em [1] propõe-se uma nova abordagem para o problema da transmissão via canal acústico com o uso de equalizadores. Todos os algoritmos testados e pesquisados apostam no método de inserção da marcação como meio de sobrevivência da marca frente a transmissão via acústica. A equalização do áudio transmitido é uma nova forma de adicionar robustez ao sistema, independente do tipo de marcação que se tenha.

Portanto, este trabalho pretende seguir a mesma linha proposta em [1], porém, propondo mudanças na forma da estimação do canal. Também, há a intenção de testar a solução em ambiente real, para avaliar se há melhora nas taxas de decodificação dos sinais marcados transmitidos em relação a um sistema que não usa a equalização. E assim, verificar se, com o aumento da robustez, é possível diminuir a energia de marcação e obter uma marca d'água mais transparente.

## Capítulo 3: Áudio digital

O crescente uso da *Internet* aliado à conversão do áudio analógico para o formato digital possibilitou a distribuição em massa de áudios com um custo muito pequeno. A manipulação dos áudios também foi facilitada, de forma tal que qualquer indivíduo possa fazer o *download* de *softwares* de edição que permitem executar *mixagens*, cortes, equalização, alterar características do som como *pitch*, volume, enfim, criar em casa um estúdio de edição com processos que antes só podiam ser feitos através de uma aparelhagem muito complexa. É inegável que o formato digital de áudio tornou muito mais conveniente o armazenamento, transmissão e manipulação de sons.

No caso do antigo sistema de gravação analógico de som, as ondas sonoras eram capturadas através de um transdutor, geralmente representado por um microfone. Através deste transdutor, as ondas são traduzidas em um sinal elétrico contínuo e amplificadas por um dispositivo eletrônico chamado de amplificador. Em seguida, esse sinal elétrico é convertido, por outro transdutor, em um sinal magnético ou mecânico, dependendo do tipo de armazenamento (fita k7 ou disco de vinil, por exemplo). Para se reproduzir o som, é necessário um alto-falante que recebe uma onda como um sinal elétrico amplificado, o qual foi transformado por um transdutor que interpretou o sinal do seu formato magnético ou mecânico. O alto falante gera o som como conhecemos, através de ondas de pressão no ar [11]. Estes dois processos são ilustrados na Figura 3.1.

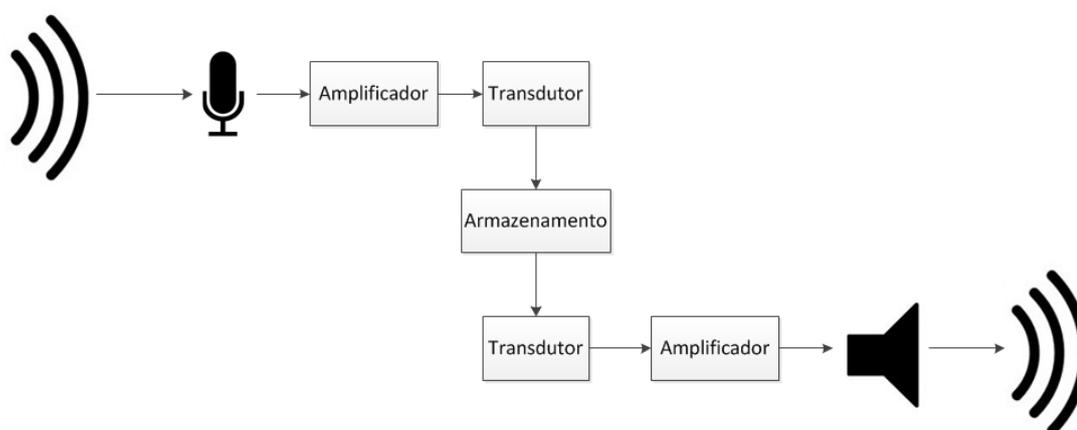


Figura 3.1: Sistema de gravação e reprodução analógico de áudio

O grande problema deste tipo de sistema é que cada vez que o áudio é trans-

formado por processos analógicos ele é distorcido e se distancia do sinal que foi capturado. Portanto, a cada cópia e edição destes áudios ocorre um desvio no sinal original. Mesmo quando armazenado o sinal sofre estas distorções. No caso da fita k7 distorções magnéticas e no vinil desgaste dos sulcos por onde passam as agulhas. Dessa forma, a vida útil e a capacidade de replicabilidade destas mídias analógicas é muito limitada se comparadas com as mídias digitais [11].

Já no sistema digital, é feita a mesma conversão de ondas sonoras para ondas elétricas, porém após este processo, o áudio é transformado em um vetor de *bits*, por uma placa AD instalada num computador. A onda elétrica é amostrada a uma dada taxa, chamada taxa de amostragem, normalmente dada em *Hertz*. A resolução da conversão é dada em *bits per sample*, ou seja, quantos *bits* compõe uma amostra do sinal. O sinal representado por um vetor de *bits* pode ser replicado infinitas vezes que não será distorcido. Essa forma de representação também facilita o processamento e edição do som. Porém, mesmo no formato digital, é inevitável a distorção causada pelos sistemas analógicos envolvidos no processo de gravação e reprodução, como os microfones, alto falantes, placas DA/AD e a própria transmissão pelo meio acústico.

### **3.1: Marcação em áudio digital**

A mídia digital, seja ela vídeo, áudio ou imagem, pode ser vista como um sinal. A marcação desta mídia consiste na incorporação de um sinal de informação dentro do sinal de mídia, de forma que o conteúdo da mídia continue inteligível. A replicabilidade das mídias digitais permite que a marcação persista na mídia independente de quantas vezes for copiada ou reproduzida. O campo da marcação digital de áudios, apresenta um grande desafio se comparada com a marcação digital de vídeos e imagens. O sistema auditivo humano tem uma sensibilidade muito apurada o que torna mais difícil ocultar uma informação no áudio do que numa imagem ou vídeo. Além disso, o sinal de áudio contém somente uma dimensão de dados, o que dificulta a inserção de *bits* sem a distorção na qualidade do som [12].

#### **3.1.1: Áreas de pesquisa**

As pesquisas na área de marca d'água em áudio seguem diversas abordagens. Geralmente, a aplicação define qual abordagem será privilegiada e qual técnica de marcação cumpre melhor os objetivos exigidos. Abaixo, são resumidas as proprieda-

des consideradas mais relevantes para algoritmos de marcação de áudio digital [2, 3].

#### 3.1.1.1: Transparência

Na maioria das aplicações, o algoritmo de inserção da marca deve inserir dados sem afetar a qualidade perceptual do sinal hospedeiro, ou seja do áudio original. A transparência da marca em um sinal é maior, quanto maior for a fidelidade do áudio marcado ao sinal original.

#### 3.1.1.2: Segurança

Dependendo da aplicação, pode ser desejável que somente pessoas autorizadas sejam capazes de extrair a informação inserida pela marcação. A segurança do processo de marcação é interpretada da mesma maneira que a segurança de criptografia é tratada. Nenhum indivíduo, mesmo que saiba que o sinal esteja carregando uma marca, deve ser capaz de extraí-la sem a chave de criptografia correta.

#### 3.1.1.3: Capacidade

A capacidade, ou *payload*, de uma marcação é o número de *bits* inseridos por uma unidade de tempo, geralmente dada em *bits* por segundo. O *payload* da marca d'água é definido pelo tipo de aplicação e algoritmo utilizados na marcação.

#### 3.1.1.4: Robustez

A robustez de um algoritmo é definido como a habilidade do detector da marca d'água de extrair a marca depois de manipulações, ou ataques ao sinal marcado. Aplicações geralmente requerem robustez frente a um universo pré-definido de ataques. Por outro lado, em alguns algoritmos a robustez é completamente indesejada, são os chamados algoritmos de marcação frágeis, que procuram identificar se o sinal passou por algum tipo de processamento em seu caminho até o destinatário.

#### 3.1.1.5: Algoritmos cego ou não-cego

Em algumas aplicações, os algoritmos de detecção podem precisar do áudio hospedeiro original para extrair a marca, esta é a chamada detecção não-cega. Porém,

se a aplicação não tem acesso ao sinal original, a extração deve ser feita somente com o sinal transmitido, essa é denominada detecção cega.

#### 3.1.1.6: Complexidade e custo computacional

O principal problema técnico que envolve a complexidade dos algoritmos de inserção e detecção de marcas é o requisito temporal. Enquanto em algumas aplicações a inserção e detecção deve ser feita em tempo real, em outras, o tempo não é um fator crucial. Outro fator importante é o tipo de *software* e *hardware* nos quais a implementação deve ser feita. É esperado que com o crescimento do uso dos dispositivos móveis, a demanda por aplicativos para estas plataformas aumente. Contudo, há de se considerar o poder de processamento e memória destes aparelhos, que muitas vezes, limita o tipo de aplicação ou algoritmo utilizado.

#### 3.1.1.7: Eficácia da inserção

A eficácia de um sistema de marcação é a probabilidade de decodificação da marca logo após a inserção, ou seja, sem passar por ataques ou processamentos. Essa definição implica que sistemas de marca d'água podem apresentar eficácia abaixo de 100%. A eficácia de 100%, apesar de ser almejada, só é alcançada a um custo muito alto em relação às outras propriedades. Dependendo da aplicação, pode ser aceitável sacrificar um pouco da eficácia para ter uma performance melhor em outras características, como transparência e capacidade.

É importante ressaltar que qualquer que seja a técnica implementada há sempre um *trade-off* entre robustez, capacidade de transmissão e transparência. Estas três características podem ser colocados num triângulo como o da Figura 3.2. Quando se quer uma alta robustez deve-se abrir mão de uma alta transparência da marca e capacidade ou vice-versa. Isto acontece, porquê uma alta robustez, geralmente, requer uma marca com uma energia muito forte, que compromete a transparência, e/ou a redundância da informação, que compromete a capacidade. Já marcas com alta capacidade de transmissão, apresentam sinais de marcação muito densos, que tornam mais difícil a tarefa de escondê-los no sinais hospedeiros sem comprometer a qualidade do áudio.

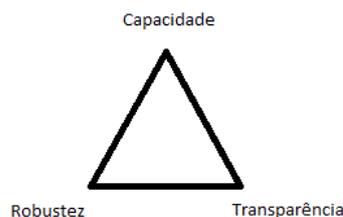


Figura 3.2: Triângulo de *trade-off*

### 3.1.2: Técnicas de marcação em áudio digital

A marcação digital de áudio é algo relativamente recente, os primeiros algoritmos que tratavam do assunto surgiram perto do ano de 1996. Algumas técnicas que foram desenvolvidas para a marcação de vídeos e imagens foram adaptadas para o uso em sinais de áudio [2]. Aos poucos novos métodos vêm sendo testados e outros melhorados através da utilização das características do áudio e do sistema auditivo humano. A seguir, revisa-se as ideias básicas de algumas das técnicas mais conhecidas na área da marcação digital de áudio.

### 3.1.3: LSB – *Least significant bit*

A técnica LSB foi uma das primeiras a ser estudada no mundo das marcas d'água digitais, dada a simplicidade de sua teoria e implementação [2]. Como o próprio nome já diz, esta abordagem se utiliza dos *bits* menos significativos das amostras do som para esconder a informação. Para a maior segurança da marcação, muitas vezes é gerada uma chave de criptografia que determina quais amostras serão selecionadas para a modificação dos *bits* menos significativos.

A decodificação é feita de forma similar. Através da chave secreta são selecionadas as amostras que foram modificadas e são lidos os *bits* menos significativos. Juntando todos os *bits* lidos é formada a informação que foi escondida no áudio. Para se aumentar a robustez é desejável repetir a informação várias vezes em um mesmo sinal.

A modificação dos *bits* menos significativos introduz no sinal um ruído branco gaussiano aditivo de baixa intensidade. A sensibilidade do sistema auditivo humano limita o número de *bits* que podem ser modificados imperceptivelmente [2]. A maior vantagem deste método é a alta capacidade de armazenagem de dados. Em um

áudio amostrado a 44 kHz é possível armazenar 44 kbps usando somente um *bit* por amostra. Por outro lado, a desvantagem evidente é a baixíssima robustez da marca. A simples adição de ruído, conversão DA/AD ou compressão podem comprometer a marcação.

### 3.1.4: Ocultamento de “eco”

A técnica de ocultamento por “eco” é bastante explorada na literatura, principalmente pelas suas características que garantem uma alta qualidade do sinal marcado. A transparência da marca é um ponto forte deste tipo de abordagem. Pode-se dizer que a diferença básica entre se ouvir um sinal de áudio sem marcação e um marcado com ocultamento por “eco”, é a mesma de se ouvir um som com fones de ouvido e caixas de som. Com os fones só se ouve o som original, já com as caixas se ouve o som original somado com “ecos”, causados pela acústica de um ambiente [13].

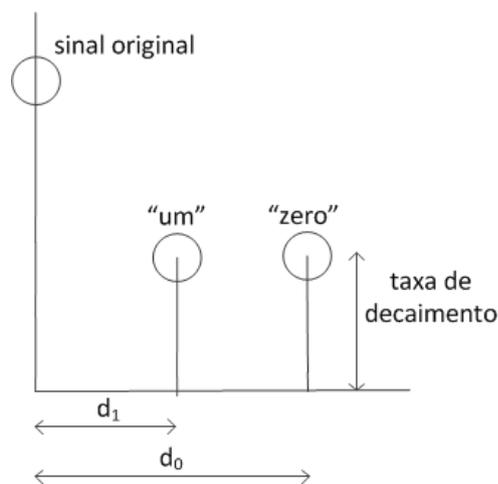


Figura 3.3: Parâmetros de inserção do “eco”

A abordagem original descrita em [13], utiliza somente um pulso de “eco” deslocado por um atraso  $\delta$ ,  $d_0$  ou  $d_1$  dependendo do *bit* que se quer inserir, e multiplicado por um fator  $\alpha$  de decaimento. Portanto, o áudio é dividido em várias partes e em cada parte é adicionado um “eco” com um atraso que representa o *bit* inserido. Na detecção o sinal também é dividido e a identificação dos “ecos” se dá pela análise do *cepstrum* de cada parte do sinal [13]. É verificado um grande pico no gráfico do *cepstrum* exatamente sobre o atraso que foi inserido durante a marcação, se o pico está em  $d_0$  o *bit* codificado é 0, caso esteja em  $d_1$  o *bit* é 1. Assim, após verificar todos os pedaços do áudio, se tem a informação que foi inserida pelo processo de marcação de áudio digital.

Apesar da vasta literatura trazendo melhorias para esta técnica, a robustez da marcação ainda é uma desvantagem se comparada a abordagens que utilizam o domínio da frequência para esconder informação. Por estar inserida no domínio do tempo a informação é facilmente corrompida pela compressão, adição de ruído, entre outros tipos de processamentos muito comuns no mundo do áudio digital.

### 3.1.5: Codificação de fase

Os métodos de codificação por fase se utilizam da falta de sensibilidade do sistema de audição humano a mudanças absolutas na fase de um sinal. A técnica funciona substituindo a fase do início de um segmento por uma de duas fases de referência. Estas fases de referência estão relacionadas com os *bits* 0 e 1. Para manter a distorção de fase no menor nível possível, são feitos ajustes nas fases dos segmentos adjacentes para que a diferença relativa entre eles seja mantida. A inserção é feita da seguinte forma:

1. O sinal original é dividido em pequenos segmentos cujo o tamanho é igual ao da mensagem que se quer inserir;
2. A DFT é aplicada em cada segmento e é obtido um sinal complexo, com magnitude e fase;
3. A diferença entre fases adjacentes são calculadas;
4. A mudança de fase entre segmentos consecutivos é facilmente detectável. Em outras palavras, o valor absoluto da fase dos segmentos pode ser mudada, porém a diferença relativa entre fases de segmentos adjacentes deve ser preservada. Portanto, a informação só é inserida no vetor de fase do primeiro segmento do sinal. Coloca-se uma fase de  $90^\circ$  para um *bit* 1 e  $-90^\circ$  para um *bit* 0, por exemplo;
5. Um novo sinal complexo é criado usando a nova fase do primeiro segmento e as diferenças originais entre as fases;
6. Usando a fase do novo sinal complexo e a magnitude do sinal complexo original, aplica-se a IDFT e concatena-se os segmentos.

Para fazer a extração o decodificador deve saber exatamente o tamanho do segmento. Então é feita a DFT para se conseguir a fase do sinal e assim retirar a informação.

Uma grande desvantagem deste método é a baixa capacidade do algoritmo, visto que a mensagem é inserida somente no primeiro segmento do sinal [14].

### 3.1.6: Espalhamento espectral

O método de espalhamento em frequência é um exemplo de método de correlação que insere uma sequência pseudoaleatória e detecta a marca d'água calculando a correlação entre a sequência de ruído pseudoaleatório e o áudio marcado. Este método apresenta grande robustez frente a diversos ataques.

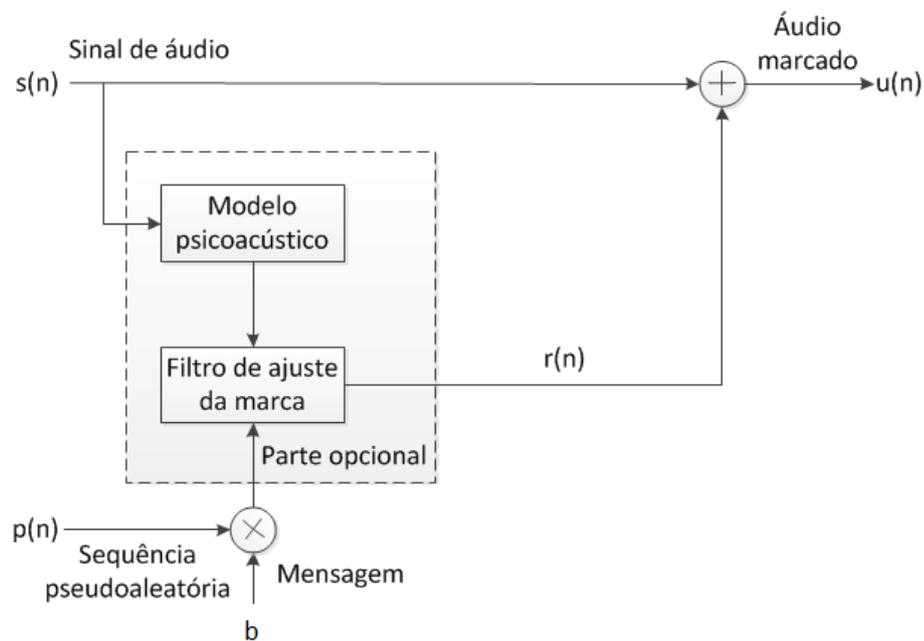


Figura 3.4: Um esquema típico de inserção do método de espalhamento em frequência

O princípio desta técnica é espalhar uma sequência pseudoaleatória por todo o sinal de áudio. O ruído de banda larga pode ser espalhado tanto no domínio do tempo quanto no transformado. Transformadas usadas frequentemente são a DCT, DFT e DWT. A mensagem binária da marca d'água bipolar  $b = \{-1, +1\}$  é modulada por uma sequência pseudoaleatória  $p(n)$  gerada por uma chave secreta. Então, a marcação modulada  $r(n) = bp(n)$  é escalada de acordo com a energia requerida do sinal hospedeiro  $s(n)$ . O fator de escala  $\alpha$  controla o compromisso entre robustez e transparência da marca. Segundo mostrado na Figura 3.4, o sinal modulado é somado ao áudio original para produzir o áudio marcado  $u(n)$  dessa forma:

$$u(n) = s(n) + \alpha r(n) \quad (3.1)$$

Dado que a sequência pseudoaleatória  $p(n)$  é conhecida e pode ser gerada de novo

através da chave secreta, as marcas são detectadas usando a correlação linear entre  $u(n)$  e  $p(n)$ :

$$\begin{aligned} c &= \frac{1}{N} \sum_{i=0}^{N-1} u(i)p(i) \\ &= \frac{1}{N} \sum_{i=0}^{N-1} s(i)p(i) + \frac{1}{N} \sum_{i=0}^{N-1} \alpha b p^2(i) \end{aligned} \quad (3.2)$$

Sendo  $N$  o comprimento do sinal. Assume-se que o primeiro termo da Equação (3.2) é muito pequeno, uma vez que  $s(n)$  e  $p(n)$  são descorrelacionados. Entretanto, não é sempre o caso. Se o primeiro termo tiver magnitude similar ou maior que o segundo termo, a detecção poderá ser incorreta. Usando o valor da correlação  $c$  e um limite pré-definido  $\tau$ , as saídas detectadas são:

$$m = \begin{cases} 1 & , \text{ se } c > \tau \\ 0 & , \text{ se } c \leq \tau \end{cases}$$

### 3.2: Ataques e Processamentos Comuns

A digitalização dos sinais sonoros traz muitas facilidades e comodidades, como citado anteriormente. O armazenamento, distribuição e manipulação de áudios é feita de forma mais conveniente e fácil. Contudo, o processamento de áudio digital, no caso de arquivos marcados, pode ser feito de forma maléfica a fim de destruir a marcação e impossibilitar decodificador de extrair a marca. Ou, mesmo que não tenham a intenção, certos tipos de processamento de sinal podem interferir na marcação feita no áudio, tornando a marca ilegível.

Dependendo da abordagem utilizada para inserir a marca, alguns tipos de processamento são mais danosos que outros. Abaixo segue uma lista com categorias de ataques baseadas no modo que o processamento afeta o áudio [15].

- Dinâmicos – Mudam o perfil de intensidades do arquivo de áudio. Aumentar ou diminuir as intensidades são os ataques mais básicos. Limitar, expandir e comprimir são ataques mais complexos, pois introduzem mudanças não lineares;
- Filtros – Filtros aumentam ou cortam uma parte do espectro de frequência. Os filtros mais básicos são os passa-baixa e passa-alta. Equalizadores também podem ser vistos como filtros, geralmente são usados para aumentar ou diminuir partes do espectro;

- Ambiente – Este grupo consiste de efeitos de áudio que simulam o efeito de uma sala. Os mais comuns são os efeitos de reverberação e atraso, que oferecem uma vasta gama de parâmetros dependendo da qualidade do efeito;
- Conversão – Arquivos de áudio estão sujeitos a mudanças no formato. Faixas de áudio mono devem ser dobradas para uso em ambientes estéreo. Frequências de amostragem podem variar de 8 kHz para 96 kHz, assim como a resolução que pode mudar de 16 para 24 *bits* por amostra, por exemplo;
- Compressão com perdas – Algoritmos de compressão de áudio baseados em efeitos psicoacústico são usados para reduzir a quantidade de informações de áudio por um determinado fator;
- Ruído – Ruído pode ser o resultado da maioria dos ataques aqui descritos. Muitos dos componentes de *hardware* na cadeia de áudio introduzem ruídos no sinal. Entretanto, pode-se inserir ruído num sinal com o objetivo de destruir a marca d'água do áudio;
- Modulação – Efeitos de modulação como vibrato (consiste na oscilação de uma corda de um instrumento musical), efeito de coro, modulação de amplitude ou efeito de *flanging* (quando a distância entre a fonte sonora, a área refletora e o ouvinte varia), presentes em *softwares* de edição de áudio podem afetar a marcação do sinal;
- Estreitamento do tempo, deslocamento de *pitch* e *tempo* – Estas manipulações mudam o tamanho de um áudio sem mudar o *pitch* ou mudam o *pitch* sem alterar o tamanho do áudio. Estes ataques também podem ser classificados como ataques de distorção geométrica. São utilizados para sintonia fina ou ajuste de partes de áudio em janelas de tempo. *Pitch* é uma sensação auditiva na qual um ouvinte atribui tons musicais para posições relativas a uma escala musical baseada primariamente na frequência de vibração [16]. Por outro lado, o *tempo* é a velocidade ou o ritmo de um determinado som. O *tempo* é medido em *beats per minute*, que indica quantas batidas devem ser tocadas por minuto;
- Ataques nas amostras – Este grupo consiste de algoritmos que não são usados para manipulação de áudio em ambientes normais. São ataques que, em sua maioria, têm objetivos nefastos em relação a marcação. Exemplos são permutação de amostras e retirada de amostras de forma periódica ou não.

## Capítulo 4: O canal acústico e a equalização adaptativa

A maioria das marcações de áudio são desenhadas para sobreviver somente à ataques no mundo digital. Processos de conversão de formato de arquivo, compressão e re quantização são comuns neste tipo de ambiente. Por outro lado, poucas são as marcações que consideram a transmissão via canal analógico [4]. Também são poucos trabalhos que se dedicam a entender os efeitos da conversão DA/AD e da transmissão pelo meio acústico [10, 17]. Em geral, pode-se dizer que o sinal transmitido por um canal acústico sofre com a adição de ruído, tanto nas placas DA/AD e amplificadores, como do próprio ambiente acústico; modificações na energia, devido aos amplificadores; modificações no espectro de frequência, pelas respostas em frequência do microfone e caixas acústicas; e por fim distorções devido a reverberação do som no ambiente [10].

Este trabalho pretende focar na redução das distorções causadas pelas reverberações do som no ambiente e verificar se através dessa redução é possível aumentar a robustez de uma marcação à conversão DA/AD e transmissão pelo meio acústico. Pretende-se com isso conseguir diminuir a energia da marcação, tornando-a imperceptível, porém mantendo um bom nível de robustez a este tipo de manipulação do áudio.



Figura 4.1: Conversão DA/AD com transmissão pelo canal acústico

### 4.1: Canal Acústico

O som é uma onda mecânica de pressão que se propaga de forma longitudinal. Diferente das ondas de luz, o som necessita de um meio para existir. Uma onda sonora pode viajar por diversos meios e materiais, entre os mais comuns estão o ar (meio gasoso), água (meio líquido) e metal (meio sólido). Os sinais sonoros apresentam características como comprimento de onda, amplitude e frequência. A velocidade de propagação das ondas depende do meio em que o som se desloca. Fatores como a umidade, temperatura e tipo de interação entre as moléculas do meio modificam a

velocidade em que estas ondas viajam. As ondas sonoras também obedecem aos mesmos fenômenos gerais da propagação ondulatória, sofrendo reflexão, refração, difusão e difração.

Devido a estes fenômenos, o sinal de som emitido por uma fonte sonora não é idêntico ao que chega num receptor. Quando uma onda sonora incide sobre um material, parte da energia é absorvida e outra refletida. Dependendo das características da superfície e da onda incidente, o som refletido é espalhado em direções aleatórias [18]. As ondas sonoras no interior de uma sala refletem em paredes e objetos, perdem energia, mudam de direção e sofrem atrasos com a propagação no ar, construindo interferências aditivas e destrutivas entre os sinais sonoros [19]. Um receptor, portanto, recebe o sinal sonoro que percorre o caminho direto sem influencia de obstáculos, e uma parcela indireta, constituída de todas estas outras ondas que foram refletidas pelo ambiente, e que chegam atrasadas e atenuadas. O som que viaja diretamente do emissor até o ouvinte contribui com a sensação de sonoridade, clareza e localização, já as reflexões que chegam atenuadas e atrasadas são responsáveis pela percepção de prolongamento do som [20].

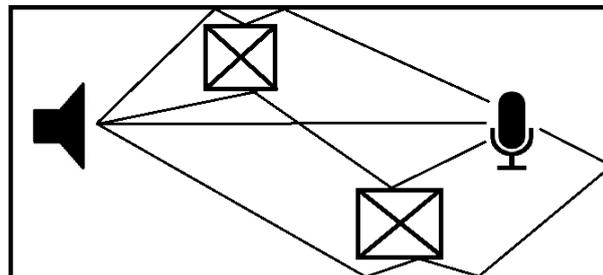
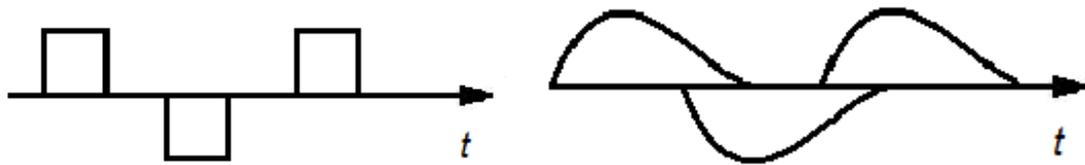


Figura 4.2: Ondas refletem no ambiente e chegam atrasadas ao receptor

Basicamente, a reverberação em um ambiente pode ser caracterizada pela resposta ao impulso do ambiente [21]. Esta resposta representa um modelo de como o som se comporta em relação àquela exata posição no ambiente e aquele instante de tempo. Variações no ambiente como mudança de lugar de objetos ou do receptor/emissor, obstáculos transitando e até mudanças na umidade e temperatura, podem levar a diferentes respostas ao impulso para um mesmo ambiente [22].

A característica de multipercurso dos sinais sonoros causa um efeito bem conhecido dos sistemas de telecomunicações, a interferência intersimbólica. Esta distorção é caracterizada pela interferência de um símbolo transmitido em símbolos seguintes. Acontece em sistemas de propagação múltipla e canais limitados em banda. Na Figura 4.3 podemos ver um exemplo do que esta distorção faz ao sinal. O símbolo enviado é

espalhado no tempo de uma maneira que os próximos pulsos que chegam ao receptor são afetados pela contribuição do símbolo anterior. Este efeito pode causar erros na interpretação dos pulsos transmitidos.



(a) Pulsos enviados

(b) Efeito do canal de transmissão sobre os pulsos

Figura 4.3: ISI na transmissão de pulsos

Existem algumas maneiras de reduzir os efeitos destas interferências. Uma maneira muito rudimentar, é aumentar o atraso entre os pulsos enviados até que toda a contribuição de um pulso tenha se extinguido. Porém, essa solução dificilmente é adotada nos dias de hoje, pois os sistemas atuais requerem uma taxa de transmissão cada vez maior. A opção adotada neste trabalho é a equalização, que pretende inverter o efeito do canal convolutivo através de um filtro, tornando a resposta em frequência do sistema plana.

Além da reverberação, o receptor está sujeito a interferência do ruído ambiente. Ruído pode ser considerado todo som que não é desejado e degrada o áudio que se deseja escutar. Estamos sujeitos ao ruído em qualquer ambiente que estivermos, seja o ruído da ventoinha do computador, o barulho do ar-condicionado, pessoas falando, etc.

Portanto, o canal acústico, num dado instante de tempo, pode ser modelado como um filtro linear do tipo FIR seguido de um ruído aditivo [23].

$$\hat{u}(n) = s(n) * c(n) + \eta(n) \quad (4.1)$$

Em que  $c(n)$  é a resposta ao impulso do canal acústico num instante de tempo,  $s(n)$  é o áudio emitido,  $\eta(n)$  representa o ruído ambiente e  $\hat{u}(n)$  é o áudio resultante após passar pelo canal de transmissão acústico.

### 4.1.1: Estado da arte - Marca d'água digital no canal acústico

A sobrevivência da marcação digital em ambientes acústicos é um tema ainda limitado na literatura. Poucos são os artigos que clamam a robustez da marcação de áudio e mostram os resultados de testes frente ao canal acústico, alguns exemplos são: [1, 5, 6, 7, 8, 9] .

Em [6] é utilizado a LCM, para incorporar e extrair a marca mesmo em situações de ataques extremos, como os geométricos de estreitamento de tempo e mudança de *pitch*. A marca é incorporada na frequência com a DFT e mapeada para o domínio logarítmico através da ferramenta de LCM. Já em [5] um sistema de transmissão de informações via rádio para passageiros de automóveis é proposto. A ideia é incorporar ao áudio do rádio uma marcação no domínio DCT. Outro que se utiliza do domínio DCT para adicionar a marca ao sinal hospedeiro é [9].

A transformada discreta de Fourier é usada por [7], porém um esquema envolvendo o re-ordenamento dos *bits* de informação é feito para melhorar a robustez e a transparência da marca. O autor em [8], desenvolve um sistema com dois alto falantes para misturar a marca com o áudio hospedeiro no ar em tempo real. O objetivo é viabilizar a marcação em concertos de música para que a marca seja recuperada por dispositivos móveis da plateia. Sua marcação é desenvolvida no domínio da frequência por uma matriz bi-dimensional pseudoaleatória.

Todos os autores citados se utilizam de técnicas baseadas no espalhamento de frequências abordado na Seção 3.1.6. A principal justificativa para o uso do espalhamento na frequência é sua robustez, se comparada com os outros métodos de marcação. O trabalho de [1] se diferencia pela elaboração de uma solução ativa além do projeto da marcação. É proposto um equalizador para inverter os efeitos danosos da reverberação em ambientes acústicos e assim tornar a marca mais robusta este tipo de ataque.

## 4.2: Equalização

A equalização por filtragem visa reverter efeitos indesejados em um sinal que passa por um canal não ideal de transmissão. Na prática, poucos canais podem ser considerados ideais com um espectro de frequência passa-tudo [24]. A maioria, possui atenuações e amplificações em diferentes frequências que modificam o sinal trans-

mitido, muitas vezes descaracterizando-o e causando problemas como a ISI. Abaixo segue um modelo básico de transmissão de dados através de um canal com equalizador.

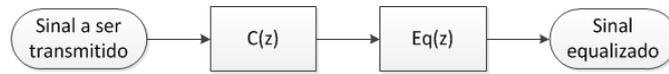


Figura 4.4: Transmissão com canal e equalizador

A intenção do equalizador  $Eq(z)$  é deixar a resposta em frequência do sistema plana. Idealmente, o sinal que for transmitido chegará no destino sem distorções, independente da sua banda de frequência. Isso significa que a resposta ao impulso do conjunto  $C(z) * Eq(z)$  será um impulso deslocado no tempo [25].

$$\begin{aligned}
 C(z) * Eq(z) &= \delta(z - D) \\
 Eq(z) &= \frac{\delta(z-D)}{C(z)}
 \end{aligned} \tag{4.2}$$

Esta solução é conhecida como *zero forcing equalizer*, pois visa compensar completamente os efeitos do canal de transmissão forçando o erro introduzido pela ISI a zero. O ZFE não considera o ruído que pode ser adicionado ao sinal durante a transmissão. Portanto, se o canal apresentar nulos espectrais, o equalizador apresentará um ganho enorme nessa faixa de frequência, na tentativa de inverter a atenuação causada pelo nulo espectral. Este ganho, por sua vez, pode amplificar o ruído, se este tiver componentes na faixa de frequência do nulo espectral [25, 23].

Uma alternativa que considera a introdução do ruído no sistema é a equalização pela minimização do erro quadrático médio entre um sinal de treinamento e o sinal filtrado pelo equalizador. Esta solução é considerada um bom compromisso entre minimização da ISI e redução dos efeitos maléficis do ruído. Pois, o equalizador desvia da solução ótima da Equação 4.2, na tentativa de compensar o ruído. Por isso, num ambiente com ruído a função de transferência do equalizador não é exatamente a inversa do canal convolutivo [25].

Os dois tipos de equalizadores citados apresentam problemas quando as distorções devidas a ISI e ao ruído são muito severas [23]. Para estas situações um equalizador não-linear chamado *decision-feedback equalizer* pode ser utilizado. Este equalizador combina a solução do MMSE com a realimentação através de um segundo filtro adaptativo, que pretende utilizar a decisão de símbolos anteriores para anular a interferência destes símbolos sobre a decisão dos símbolos presentes [25].

Os sistemas de equalização podem ser classificados quanto ao seu modo de operação, em adaptativos ou não adaptativos. Os equalizadores não adaptativos foram os primeiros a serem desenvolvidos e eram usados em sistemas em que o canal era invariante. O equalizador não tinha o poder de modificar seus coeficientes ao longo do tempo e, portanto, tinha seu comportamento pré-estabelecido mediante prévio estudo do canal que iria ser equalizado. Já os equalizadores adaptativos são mais recentes e consideram que o canal apresenta características que variam ao longo do tempo, e assim não podem ser equalizados por um filtro estático. É necessário que os coeficientes do equalizador se adaptem às mudanças que ocorrem no canal ao longo do tempo.

Assim como os sistemas de marca d'água, os equalizadores adaptativos podem ser divididos em cegos (autodidatas), e não cegos (supervisionados). Os equalizadores não cegos se utilizam de sinais de treinamento transmitidos periodicamente pelo canal que se quer equalizar. Ao receber estes sinais modificados pelo canal, pode-se estimar a influência do canal convolutivo no sinal de treinamento e, então, é possível adaptar os coeficientes do equalizador. Os equalizadores cegos não dispõem de sinais de treinamento, por este motivo devem fazer a equalização às cegas, justificando seu nome. Nesta implementação, através das estatísticas do sinal emitido e do sinal recebido é possível estimar os efeitos do canal de comunicação e assim lidar melhor com o sinal distorcido.

#### **4.2.1: Filtragem Adaptativa**

Dentro do processamento de sinais, um recurso muito utilizado é a filtragem. Um filtro é um dispositivo que mapeia um sinal de entrada em um sinal de saída para facilitar a análise e extração de informações contidas no sinal de entrada. Se o filtro é linear, o sinal de saída será uma função linear do sinal de entrada. Os filtros podem ser classificados em dois grandes grupos: os filtros invariantes no tempo e os variantes [25].

Os invariantes, mantêm sua estrutura e parâmetros fixos ao longo do tempo e independentes do ambiente e sinais de entrada. Estes filtros são usados quando se tem uma ideia clara e definida das especificações requeridas pelo sistema. Uma vez adquiridas estas especificações são necessários três passos para se obter um filtro, são elas: a aproximação da especificação por uma função de transferência, a escolha de uma estrutura apropriada e definição da forma de implementação do algoritmo.

Problemas que não tem especificações bem definidas ou que variam ao longo do tempo são, geralmente, tratados com filtros adaptativos. Estes filtros são considerados variantes, pois há uma atualização dos seus parâmetros ao longo do tempo para atingir determinado requerimento. A variação dos parâmetros depende do sinal de entrada, de um sinal de referência e do algoritmo de minimização de erro adotado. Esta dependência para atualização dos parâmetros leva a outra consequência que é a não linearidade deste tipo de filtro. Porém se analisarmos a transformação desempenhada por ele em um determinado instante congelado de tempo, pode-se verificar a linearidade da operação, pois os sinais de saída são função linear dos sinais de entrada.

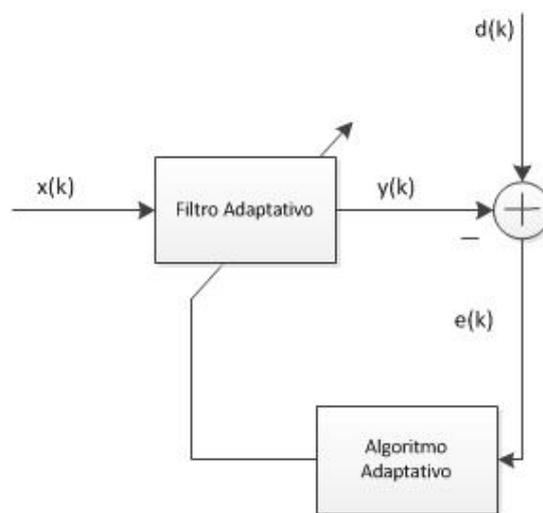


Figura 4.5: Configuração básica de um filtro adaptativo.

A Figura 4.5 mostra um exemplo básico da estrutura que um filtro adaptativo pode assumir. Percebe-se o sinal de entrada  $x(k)$  e o sinal desejado  $d(k)$ , sendo  $k$  é número da iteração. O sinal de erro  $e(k)$  é calculado como a diferença entre a saída do filtro e o sinal de referência. Este erro  $e(k)$  é entrada de um algoritmo que descreve certo desempenho esperado ou função objetivo, que determina a atualização necessária nos coeficientes do filtro.

Segundo [25], a especificação de um sistema adaptativo descrito na Figura 4.5, consiste de três itens:

1. Aplicação - O tipo de aplicação depende dos sinais medidos do ambiente e escolhidos pelo projetista para serem os sinais de entrada e referência. Há muitas aplicações às quais este tipo de solução é viável, entre elas está o cancelamento de eco, identificação de sistemas, *beamforming* adaptativo, equalização e cancelamento de ruído.

2. Estrutura do filtro adaptativo - A estrutura do filtro influencia na complexidade computacional do processo e no número necessário de iterações para chegar a um determinado nível de desempenho esperado. Basicamente há duas maneiras de se implementar um filtro digital adaptativo, e o que as distingue é a forma da resposta ao impulso. Há a resposta finita ao impulso e a resposta infinita ao impulso. A saída de um filtro FIR é simplesmente a soma ponderada das últimas amostras de entrada. A ordem do filtro FIR indica sua memória. Enquanto o filtro IIR faz o mesmo procedimento, porém adiciona a soma ponderada de amostras de saída passadas à soma descrita acima. Portanto, o filtro FIR é considerado não recursivo e o IIR recursivo. A recursividade do IIR pode, muitas vezes, levar à instabilidade devido ao crescimento das amostras de saída que por sua vez fazem os pesos do filtro crescerem exponencialmente, levando ao *overflow*.
3. Algoritmo - É o processo que determina o desempenho que se quer obter na utilização do filtro. O algoritmo é determinado pela definição do método de busca (ou algoritmo de minimização), função objetivo e natureza do sinal de erro.

O objetivo básico de um filtro adaptativo é achar um conjunto de coeficientes  $\theta(k)$  que minimize uma função  $F$ , que depende do sinal de entrada, do sinal de saída e da referência, portanto  $F(x(k), y(k), d(k))$ . Os coeficientes  $\theta(k)$  devem convergir para um  $\theta_o(k)$  que faça  $y(k)$  se aproximar ao máximo de  $d(k)$ , idealmente com erro igual a zero. Os métodos mais conhecidos e utilizados para esta minimização são: o método de Newton, métodos *Quasi-Newton* e o método do gradiente [25]. O primeiro é um método que procura o mínimo de uma aproximação de segunda ordem da função objetivo, usando uma fórmula de atualização de parâmetros iterativa. Os métodos *Quasi-Newton* são uma simplificação do método de Newton que se utiliza de recursividade para realizar a minimização. E o método do gradiente toma a direção oposta ao vetor gradiente da função objetivo para achar o ponto de mínimo desta função.

O método escolhido para implementação é o do gradiente, principalmente pela baixa complexidade. Os métodos de Newton e *Quasi-Newton* geralmente apresentam uma maior taxa de convergência, porém são mais custosos computacionalmente, além de que os métodos *Quasi-Newton* podem apresentar instabilidade devido a recursividade utilizada nos cálculos [25].

#### 4.2.1.1: Filtro de Wiener

Uma das funções de custo mais utilizadas é a do erro médio quadrático. Seguindo o exemplo da Figure 4.5, podemos defini-la assim:

$$F[e(k)] = \varepsilon(k) = E[e^2(k)] = E[d^2(k) - 2d(k)y(k) + y^2(k)] \quad (4.3)$$

Considerando o filtro com resposta ao impulso finita, temos que:

$$y(k) = \sum_{i=0}^{N-1} w_i(k)x(k-i) = \mathbf{w}^t(k)\mathbf{x}(k) \quad (4.4)$$

Sendo  $\mathbf{x}(k) = [x(k), x(k-1), \dots, x(k-N+1)]^t$  o vetor de entrada e  $\mathbf{w}(k) = [w_0(k), w_1(k), \dots, w_{N-1}(k)]^t$  os coeficientes do filtro, em que  $N$  é o tamanho do filtro. Em um caso linear com um filtro de coeficientes fixos e em um ambiente estacionário podemos reescrever a Equação 4.3 desta forma:

$$\begin{aligned} \varepsilon &= E[d^2(k)] - 2\mathbf{w}^t E[d(k)\mathbf{x}(k)] + \mathbf{w}^t E[\mathbf{x}(k)\mathbf{x}^t(k)]\mathbf{w} \\ &= E[d^2(k)] - 2\mathbf{w}^t \mathbf{p} + \mathbf{w}^t \mathbf{R} \mathbf{w} \end{aligned} \quad (4.5)$$

Em que  $\mathbf{R}$  é a auto correlação do vetor de entrada e  $\mathbf{p}$  é a correlação cruzada entre o sinal de referência e o sinal de entrada. O vetor gradiente do MSE relacionado aos coeficientes do filtro é:

$$\mathbf{g}_{\mathbf{w}} = \frac{\nabla \varepsilon}{\nabla \mathbf{w}} = -2\mathbf{p} + 2\mathbf{R}\mathbf{w} \quad (4.6)$$

Se igualarmos o vetor gradiente a zero e considerarmos  $\mathbf{R}$  não singular, os valores ótimos dos coeficientes do filtro de Wiener são dados por:

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} \quad (4.7)$$

Infelizmente, na prática não conhecemos com precisão  $\mathbf{R}$  e  $\mathbf{p}$ . Uma estimação destas matrizes é feita pelos algoritmos adaptativos e assim uma aproximação da solução de Wiener é alcançada.

#### 4.2.1.2: LMS - *Least Mean Square*

O algoritmo de busca LMS tenta achar os coeficientes de um filtro adaptativo capaz de minimizar o erro médio quadrático. Para isso, utiliza-se de simplificações do método do gradiente. O LMS é um dos métodos mais utilizados devido a sua baixa complexidade computacional, sua convergência em ambiente estacionário e

convergência, na média, para a solução de Wiener.

Toma-se o resultado da Seção 4.2.1.1 , a solução de Wiener:

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} \quad (4.8)$$

Sendo  $\mathbf{R} = E[\mathbf{x}(k)\mathbf{x}^t(k)]$ ,  $\mathbf{p} = E[d(k)\mathbf{x}(k)]$  e assumindo que  $d(k)$  e  $\mathbf{x}(k)$  são conjuntamente estacionários no sentido amplo. Se considerarmos o método do gradiente, que atualiza os coeficientes de acordo com o sentido contrário do gradiente multiplicado por um passo de convergência, e estimações das matrizes  $\mathbf{R}$  e  $\mathbf{p}$ , denominadas  $\hat{\mathbf{R}}$  e  $\hat{\mathbf{p}}$ , temos:

$$\begin{aligned} \mathbf{w}(k+1) &= \mathbf{w}(k) - \mu \hat{\mathbf{g}}_w(k) \\ &= \mathbf{w}(k) + 2\mu(\hat{\mathbf{p}}(k) - \hat{\mathbf{R}}(k)\mathbf{w}(k)) \end{aligned} \quad (4.9)$$

Em que  $\hat{\mathbf{g}}_w(k)$  representa uma estimacão do gradiente da função objetivo em relação aos coeficientes do filtro. Uma estimativa possível para  $\hat{\mathbf{R}}$  e  $\hat{\mathbf{p}}$  é:

$$\hat{\mathbf{R}} = \mathbf{x}(k)\mathbf{x}^t(k) \quad (4.10)$$

$$\hat{\mathbf{p}} = d(k)\mathbf{x}(k) \quad (4.11)$$

Pode-se assim definir a estimativa  $\hat{\mathbf{g}}_w(k)$  como:

$$\begin{aligned} \hat{\mathbf{g}}_w(k) &= -2d(k)\mathbf{x}(k) + 2\mathbf{x}(k)\mathbf{x}^t(k)\mathbf{w}(k) \\ &= 2\mathbf{x}(k)(-d(k) + \mathbf{x}^t(k)\mathbf{w}(k)) \\ &= -2e(k)\mathbf{x}(k) \end{aligned} \quad (4.12)$$

Substitui-se  $\hat{\mathbf{g}}_w(k)$  na Equação 4.9 e obtém-se :

$$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mu e(k)\mathbf{x}(k) \quad (4.13)$$

Que representa a equação de atualização dos coeficientes do filtro segundo o algoritmo LMS. O fator  $\mu$  é um passo de atualização que deve ser escolhido cuidadosamente, pois um valor muito pequeno pode fazer o algoritmo demorar a convergir, enquanto um valor grande pode fazer os coeficientes convergirem de forma rápida, porém com um erro em excesso grande, ou instabilizar o algoritmo.

O algoritmo nunca irá chegar à solução ótima de Wiener, devido às simplificações e estimacões assumidas para a realização do método. Todavia, os coeficientes do filtro tendem à solução de Wiener com um erro mínimo.

#### 4.2.1.3: NLMS - *Normalized Least Mean Square*

Muitos algoritmos surgiram a partir de derivações do LMS, o NLMS ou LMS normalizado é um destes métodos. Variações de potência em  $x(k)$  levam a variações significativas no ajuste dos coeficientes de  $w(k)$  no LMS. Na tentativa de melhorar o comportamento de convergência do algoritmo, buscou-se modificar o fator de atualização  $\mu$  de forma que este varie de acordo com o sinal de entrada  $x(k)$ , minimizando o erro instantâneo [25].

Faz-se a normalização em relação à potência de  $x(k)$ , deixando a função de atualização dos vetores do filtro desta forma:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\mu}{\varepsilon + \mathbf{x}^t(k)\mathbf{x}(k)} e(k)\mathbf{x}(k) \quad (4.14)$$

O parâmetro  $\varepsilon$  é um fator de segurança para não haver divisão por zero. E  $\mu$  é introduzido para controlar desajustes, pois todas as derivações para a construção deste algoritmo estão baseadas em valores instantâneos do erros quadráticos e não no erro quadrático médio [25].

#### 4.2.1.4: AFA - *Adaptive Filter with Averaging*

Outro método que apresenta um custo computacional baixo e resultados satisfatórios é o AFA. Este algoritmo leva em consideração os resultados de iterações passadas para atualizar os coeficientes do filtro. Toma-se a média dos coeficientes passados para se determinar o coeficiente presente, assim como a média da multiplicação entre o sinal de entrada e o erro. Abaixo segue a função de atualização dos coeficientes do filtro:

$$\mathbf{w}(k+1) = \bar{\mathbf{w}}(k) - \frac{1}{k^\gamma} \sum_{i=1}^k e(i)\mathbf{x}(i) \quad (4.15)$$

Em que:

$$\bar{\mathbf{w}}(k) = \frac{1}{k} \sum_{i=1}^k \mathbf{w}(i)$$
$$0.5 < \gamma < 1$$

O desempenho deste algoritmo é comprovado por [26] e [27]. Suas principais vantagens são a alta convergência e a simplicidade, tornando-o uma boa escolha para a implementação em problemas que necessitam de baixa complexidade computacional e rápida convergência.

## Capítulo 5: Marca d'água digital utilizando equalização adaptativa

Um sistema de marca d'água digital é em sua essência um sistema de comunicação [3]. Uma mensagem é modulada dentro de um sinal hospedeiro e transmitida por um canal até um receptor que faz a demodulação e decodificação desta informação. Para este sistema, o áudio hospedeiro é considerado um ruído e a marca é a informação relevante que deve ser recuperada. O principal desafio dos sistemas de marcação de áudio digital é incorporar a marcação no som hospedeiro de forma mais transparente possível, robusta à manipulações e com uma boa capacidade de transmissão de informação. Como foi citado na Seção 3.1, no projeto da marca d'água, sempre há um maior compromisso com uma destas características em detrimento das outras.

Poucos trabalhos na área de marcação de áudio digital apresentam bons resultados a transmissão em ambiente acústico e os que conseguem provar sua robustez, pecam na transparência da marca. Dado que, em geral, é difícil para os sistemas de marcação conciliar a transparência com a robustez, é testado um dispositivo independente à marca para tentar anular a influência da ISI do canal acústico e com isso alcançar uma melhor relação entre transparência e robustez. A equalização se mostra uma solução para este problema, pois seu objetivo é tornar o espectro de frequência do sistema de transmissão plano, reduzindo a ISI e a distorção do espectro. O canal acústico apresenta diferentes respostas impulsivas para diferentes posições no espaço e no tempo, por isso, o uso da filtragem adaptativa é indicada como meio de estimar os efeitos do canal.

As características dos sistemas de marcação apresentam boa compatibilidade com os sistemas de equalização adaptativa supervisionada, aquele que é necessário transmitir uma sequência de treinamento para estimar a influência do canal sobre o sinal em questão. É comum que marcações robustas apresentem dois tipos de informação incorporadas no sinal hospedeiro, uma informação referente à sincronização, devido à ataques de dessincronização, e outra à mensagem que se quer transmitir. A sequência de sincronização é conhecida do processo extrator e é transmitida periodicamente, para identificação do início e fim de cada marcação, e portanto, tem potencial para ser utilizada como sinal de treinamento para a equalização supervisi-

onada. Assim, não é necessário interromper o áudio hospedeiro para a transmissão de um sinal piloto de treinamento e pode-se atualizar a função de transferência do equalizador periodicamente.

Em [1] é proposto um sistema inovador para aumentar a robustez da marcação à transmissão no ambiente acústico incluindo, justamente, a equalização do canal para reduzir os efeitos da ISI. A solução é formulada em dois passos, a estimação do canal acústico, através de uma técnica denominada de RICE, e sua posterior inversão. Utiliza-se um filtro *zero forcing*, não causal, com resposta finita ao impulso para equalizar o canal. O autor mostra através de testes no canal acústico a redução da BER de 0.2 para  $9 \cdot 10^{-4}$  quando o *bitrate* é de 100 bps, o que significa um excelente resultado. Porém não há clareza das condições de realização dos testes e não há indicação sobre a força da marcação utilizada. O tempo necessário para que cada estimação de canal chegue a uma boa aproximação do canal real é de aproximadamente 2,5 segundos, que é um tempo muito longo na qual nenhuma informação útil é transmitida, apenas é feito o treinamento do equalizador. Além disso, o autor considera a marcação sempre sincronizada, o que inviabiliza a aplicação prática da solução, uma vez que a conversão AD/DA e a transmissão do som pelo canal acústico é uma das formas mais comuns de dessincronização [17].

Com a substituição do método de estimação de [1] por uma técnica de filtragem adaptativa, procura-se fazer uma estimação mais rápida do canal e em intervalos de tempo menores, para que a equalização acompanhe as modificações que possam acontecer no canal acústico e não se desperdice tanta banda com o treinamento do equalizador. Apesar da equalização *zero forcing* não ser a mais indicada para o caso do canal de transmissão acústico, devido ao problema com a intensificação do ruído, averigua-se a existência ou não de ganhos para o sistema com o uso desta solução em situações reais. O objetivo é facilitar o processo de decodificação, adicionando uma maior robustez a transmissão acústica do sinal e assim conseguir diminuir a energia da marca até um ponto que a torne inaudível.

## 5.1: A marcação digital de áudio

O esquema de inserção escolhido para ser implementado é o mesmo usado em [1] e melhor descrita em [28]. É um método aditivo, ou seja, que incorpora a marca através da soma do sinal hospedeiro com o sinal de marcação. O autor propõe uma marca focada na transmissão de uma informação oculta em um sinal de áudio e não na

proteção de conteúdo, portanto a inaudibilidade da marca frente ao áudio hospedeiro é privilegiada em relação à robustez do sistema à ataques e manipulações. A detecção é feita às cegas, pois o sinal hospedeiro não está disponível em sua forma original para o detector, somente o sinal distorcido pela marcação.

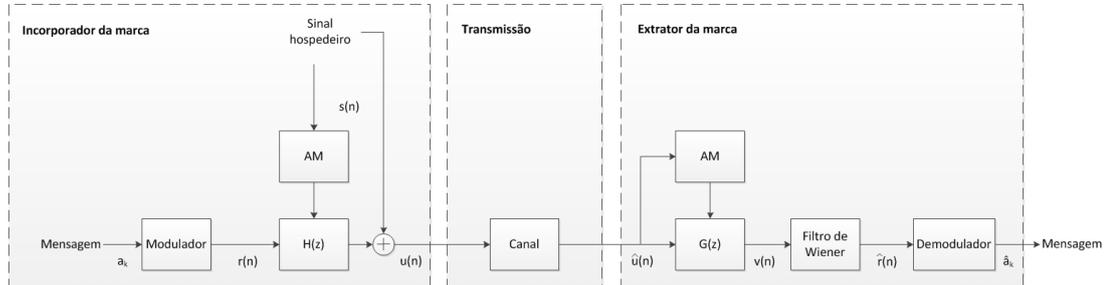


Figura 5.1: Esquema de marcação

Um esquema da marcação pode ser visualizado na Figura 5.1, para a análise do sistema de marcação o canal é considerado ideal. O sinal  $s(n)$  representa um sinal portador, como uma música ou qualquer outro áudio, que deve carregar a informação que se deseja esconder, representada por  $a_k$ . Este sinal por sua vez, passa por uma modulação e pelo mascaramento psicoacústico que esconde a informação nas frequências de  $s(n)$ .

Para a extração da informação, o mesmo processamento psicoacústico é feito, porém desta vez sobre o sinal  $\hat{u}(n)$ , já que não dispomos de  $s(n)$ . As características psicoacústicas de  $s(n)$  e  $\hat{u}(n)$  podem ser consideradas muito similares já que a contribuição de energia da marcação sobre o sinal portador é muito pequena. Assim, consegue-se inverter o efeito do mascaramento deixando as frequências em que a informação foi inserida mais aparentes. Após isso, o sinal  $v(n)$  passa por filtros de Wiener, que realçam as frequências em que o sinal modulado se encontra. E por fim, há a demodulação e a decodificação que recupera a informação inserida no início do processo.

### 5.1.1: Incorporador

O incorporador da marca é feito de forma muito simples com o auxílio do mascaramento psicoacústico. Em [28], a informação que se deseja esconder é traduzida em símbolos  $a_k$  que são selecionado de um alfabeto  $A = \{a_0, \dots, a_{K-1}\}$  de símbolos. O modulador contém  $K$  vetores  $p_k(n)$ , ortogonais e espalhados na frequência, e cada vetor corresponde a um símbolo  $a_k$ . Portanto, o sinal de marcação  $r(n)$  é obtido através

deste dicionário de acordo com a informação que se deseja inserir no áudio. É importante salientar que é utilizada uma sequência de sincronia, modulada da mesma forma, antes de cada bloco de informação. A sincronização é importante para o extrator tomar conhecimento de onde inicia e termina uma marcação.

Para a construção do filtro  $H(z)$  é utilizada a modelagem psicoacústica do áudio  $s(n)$ . Esta modelagem tira proveito das propriedades psicoacústicas do ouvido humano. Um limiar de mascaramento  $M_s(f)$  é criado a partir do modelamento psicoacústico em janelas de  $N$  amostras do sinal  $s(n)$ . O módulo do filtro  $H(z)$  corresponde ao limiar de mascaramento em frequência  $M_s(f)$ , e é atualizado a cada  $N$  amostras. Finalmente, o sinal  $r(n)$  é filtrado por  $H(z)$  e adicionado ao áudio portador  $s(n)$ , como mostrado abaixo:

$$u(n) = r(n) * h(n) + s(n) \quad (5.1)$$

### 5.1.2: Extrator

O extrator da marca utiliza filtros de Wiener IIR não causais, com intuito de minimizar o MSE entre as sequências pseudoaleatórias do alfabeto e as sequências pseudoaleatórias corrompidas com o áudio hospedeiro, e assim, facilitar a extração da sequência  $\hat{p}_k(n)$  que corresponde ao símbolo  $\hat{a}_k$ . Antes, o sinal  $\hat{u}(n)$  é recebido e passa pelo processamento psicoacústico para gerar o filtro  $G(z)$ . O limiar de mascaramento do sinal  $s(n)$  e  $\hat{u}(n)$ ,  $M_s(f)$  e  $M_{\hat{u}}(f)$ , respectivamente, podem ser considerados muito parecidos devido à influência que  $s(n)$  tem sobre  $\hat{u}(n)$ . Portanto, considera-se  $H^{-1}(z) \cong G(z)$ , e assim,  $G(z)$  inverte o efeito de mascaramento deixando o sinal  $\hat{r}(n)$  mais aparente.

Para a realização da filtragem de Wiener IIR não causal, tem-se que [29]:

$$v(n) = g(n) * \hat{u}(n) \approx r(n) + \eta(n) \quad (5.2)$$

$$\eta(n) = s(n) * g(n) \quad (5.3)$$

Em que  $g(n)$  é a resposta ao impulso do filtro  $G(z)$  e  $\hat{u}(n)$  é  $u(n)$  distorcido pelo canal. Assumindo que  $\eta(n)$  e  $r(n)$  são processos de média nula, decorrelacionados e aleatórios, o filtro de Wiener IIR não causal para realce é dado por [29]:

$$FW(e^{j\omega}) = \frac{P_r(e^{j\omega})}{P_r(e^{j\omega}) + P_\eta(e^{j\omega})} \quad (5.4)$$

Em que  $P_r(e^{j\omega})$  é o espectro de potência de  $r(n)$  e  $P_v(e^{j\omega}) = P_r(e^{j\omega}) + P_\eta(e^{j\omega})$  é o espectro de potência de  $v(n)$ . Percebe-se que nas bandas de frequências em que  $P_r \gg P_\eta$  a magnitude da resposta é aproximadamente um, enquanto nas outras bandas de frequências há uma atenuação. Deste modo, há um realce dos segmentos do sinal em que cada sequência  $p_k(n)$  foi inserida. E através da correlação pode-se encontrar qual símbolo  $a_k$  do dicionário  $A$  foi adicionado naquela porção de sinal.

## 5.2: Equalização com filtro *zero forcing*

Em [1] é apresentado uma equalização do canal acústico através de um filtro *zero forcing*. No artigo, o filtro é estruturado em duas fases. Primeiro uma estimação do canal acústico através de um filtro FIR, para conhecer suas características e seu efeito no áudio transmitido. E depois, é feita a inversão do filtro de estimação com o algoritmo de *least squares*.

A solução desenvolvida em [1] não apresenta compromisso com as variações do canal acústico. O canal é considerado estático durante o período entre duas estimações e isto pode levar até 5 segundos, que significa um longo período. A estimação é feita através de um sinal de treinamento de tamanho considerável, aproximadamente 2,5 segundos. Durante todo este tempo, a banda é utilizada para passar informações irrelevantes para o extrator da marca, caracterizando um enorme desperdício. A proposta é fazer esta estimação em períodos menores e, portanto, a premissa de que entre estes períodos o canal é estático pode ser aceita, além de liberar a banda de transmissão para transmitir informações realmente relevantes.

### 5.2.1: Estimação do canal

Para a estimação do canal, usa-se a filtragem adaptativa em uma estrutura propícia para a aproximação do canal. Esta estrutura pode ser vista pela Figura 5.2 que utiliza a mesma notação da Seção 4.2.1, com  $x(k)$ ,  $y(k)$ ,  $d(k)$  e  $e(k)$ , sendo o sinal de entrada do filtro adaptativo, sinal filtrado por  $w(k)$ , sinal desejado e sinal de erro, respectivamente.

A estrutura adaptada ao sistema de marcação digital de áudio fica como demonstrado na Figura 5.3, em que  $u(n)$ , é o sinal marcado,  $r(n)$  é o sinal de marcação, que nesta Seção para fins de simplificação é considerado contendo somente a sequência de sincronia, e  $e(n)$  é o sinal de erro. O ideal é que o mesmo sinal que passe pelo

canal desconhecido seja a entrada do filtro adaptativo, porém isso não é possível devido à indisponibilidade do sinal hospedeiro no extrator, uma vez que o algoritmo de marca d'água utilizado é do tipo cego. A solução é usar a sequência de sincronia, que está disponível no extrator, como sinal de treinamento do filtro adaptativo. Assim, o sinal hospedeiro é considerado um ruído no sinal de sincronia.

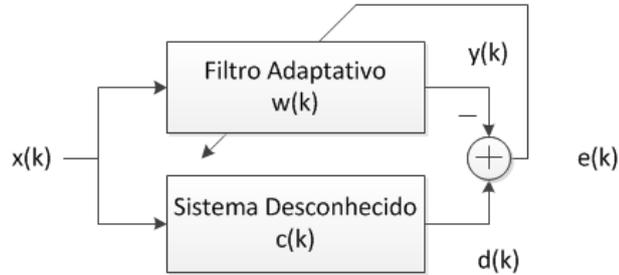


Figura 5.2: Sistema de identificação

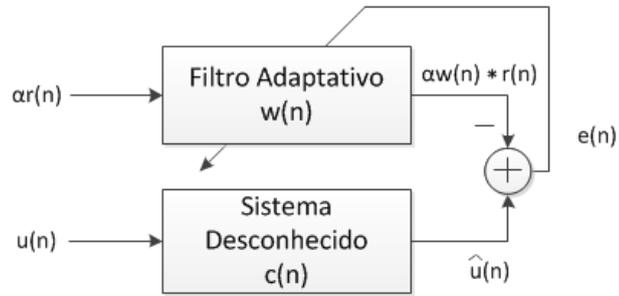


Figura 5.3: Sistema de identificação baseado na marca d'água

O sinal marcado após passar pelo canal é:

$$\begin{aligned}\hat{u}(n) &= \hat{s}(n) + \alpha\hat{r}(n) + \eta(n) \\ &= c(n) * s(n) + \alpha c(n) * r(n) + \eta(n)\end{aligned}\quad (5.5)$$

Em que  $\hat{s}(n)$  e  $\alpha\hat{r}(n)$  são versões distorcidas pelo canal  $c(n)$  dos sinais  $s(n)$  e  $\alpha r(n)$  e  $\eta(n)$  é o ruído ambiente. Para desenvolver esta estimação, ignora-se os efeitos dos mascaramentos psicoacústicos e considera-se  $s(n)$ ,  $r(n)$  e  $\eta(n)$  independentes entre si e com média nula. Outra premissa adotada é que o canal é estacionário durante o período de estimação. Conforme a Figura 5.3 o sinal desejado é o sinal marcado distorcido pelo canal acústico:

$$d(n) = \hat{u}(n) = \hat{s}(n) + \alpha\hat{r}(n) + \eta(n)\quad (5.6)$$

O sinal de entrada do filtro adaptativo é:

$$x(n) = \alpha r(n)\quad (5.7)$$

E, portanto, o sinal de erro é:

$$e(n) = \hat{u}(n) - \alpha w(n) * r(n) \quad (5.8)$$

Em seguida, define-se  $\mathbf{s}(n) = [s(n), s(n-1), \dots, s(n-L+1)]^t$  e  $\mathbf{r}(n) = [r(n), r(n-1), \dots, r(n-L+1)]^t$  como notações matriciais de  $s(n)$  e  $r(n)$  e os vetores  $\mathbf{c} = [c_0, c_1, \dots, c_{L-1}]^t$  e  $\mathbf{w} = [w_0, w_1, \dots, w_{L-1}]^t$ , como os coeficientes do canal e filtro adaptativo, respectivamente. A intenção é que os coeficientes do filtro adaptativo  $\mathbf{w}$  se aproximem ao máximo dos coeficientes do canal  $\mathbf{c}$ , e desta forma:

$$\alpha \mathbf{c}^t \mathbf{r}(n) \approx \alpha \mathbf{w}^t \mathbf{r}(n) \quad (5.9)$$

Utiliza-se o critério de minimização do erro quadrático:

$$\begin{aligned} J(\mathbf{w}) &= E[e^2(n)] \\ &= E[\hat{u}^2(n)] - \alpha^2 \mathbf{w}^t E[\mathbf{r}(n) \mathbf{c}^t \mathbf{r}(n)] - \alpha^2 E[\mathbf{r}(n) \mathbf{c}^t \mathbf{r}(n)]^t \mathbf{w} + \alpha^2 \mathbf{w}^t E[\mathbf{r}(n) \mathbf{r}^t(n)] \mathbf{w} \end{aligned} \quad (5.10)$$

Pode-se reescrever os termos acima como:

$$E[\hat{u}^2(n)] = \sigma_{\hat{u}}^2 \quad (5.11)$$

$$E[\mathbf{r}(n) \mathbf{c}^t \mathbf{r}(n)] = \mathbf{p}_r \quad (5.12)$$

$$E[\mathbf{r}(n) \mathbf{r}^t(n)] = \mathbf{R}_r \quad (5.13)$$

Em que  $\mathbf{R}_r$  e  $\mathbf{p}_r$  são, respectivamente, a aproximação da autocorrelação e da correlação cruzada do sinal da marca d'água com sua versão distorcida pelo canal. Com isso a superfície de desempenho é:

$$J(\mathbf{w}) = \sigma_{\hat{u}}^2 - 2\alpha^2 \mathbf{p}_r^t \mathbf{w} + \alpha^2 \mathbf{w}^t \mathbf{R}_r \mathbf{w} \quad (5.14)$$

Derivando em relação aos coeficientes do filtro adaptativo e igualando a zero para se obter o ponto de mínimo e o valor de mínimo da função de custo, chega-se à:

$$\mathbf{w}_o = \mathbf{R}_r^{-1} \mathbf{p}_r \quad (5.15)$$

$$J_{min}(\mathbf{w}) = \sigma_{\hat{u}}^2 - \mathbf{w}_o^t \mathbf{p}_r \quad (5.16)$$

Percebe-se então que os coeficientes do filtro adaptativo  $\mathbf{w}_o$  minimizam a superfície de desempenho e convergem para os coeficientes do filtro do canal de transmissão  $\mathbf{c}$ . O filtro  $\mathbf{w}_o$  não depende das características do sinal hospedeiro. Isso comprova que é possível fazer a estimação do canal, mesmo transmitindo a marca junto de um sinal

hospedeiro.

### 5.2.2: Equalização do canal

Na sequência é feito o mesmo procedimento descrito em [1] para a inversão do canal estimado. Considera-se que  $w(n)$  é uma boa estimativa de  $c(n)$ . O filtro FIR  $w^{-1}(n)$  é a estimativa ótima, pelo algoritmo de mínimos quadrados, do filtro inverso que cancela os efeitos da ISI devido ao canal acústico. Para isso:

$$w(n) * w^{-1}(n) = \delta(n) \quad (5.17)$$

A solução final na forma matricial da solução do mínimos quadrados é dada por:

$$\begin{bmatrix} w_0^{-1} \\ \vdots \\ w_{N'_w-1}^{-1} \end{bmatrix} = (\mathbf{W}^t \mathbf{W})^{-1} \mathbf{W}^t \mathbf{d} \quad (5.18)$$

Em que  $N'_w$  e  $N_w$  são os tamanhos dos vetores de estimativa  $w^{-1}$  e  $w$  respectivamente,  $\mathbf{W}$  é a matriz Toeplitz de dimensões  $(N'_w + N_w) \times N'_w$  construída a partir da estimativa  $w$  e  $\mathbf{d} = [0, \dots, 0, 1, 0, \dots, 0]^t$  é a representação vetorial do impulso unitário atrasado de  $N'_w/2 + 1$  amostras [1, 25]. Em [30], há um estudo mais aprofundado sobre a formulação de equalizadores *zero forcing* com o algoritmo de *least squares*.

## Capítulo 6: Implementação

A solução desenvolvida foi implementada no ambiente do MatLab<sup>®</sup> devido a familiaridade e facilidade com a programação nesta plataforma. Além disso, o projeto foi feito em caráter de protótipo, portanto, não houve a exigência de cumprir requisitos rígidos de tempo de processamento e alocação de memória.

Devido a problemas com a implementação do mascaramento psicoacústico, prevista em [1, 28], esta etapa da marcação foi retirada da solução. Porém, isto não compromete os objetivos do projeto. A implicação da retirada desta etapa é que a marca sem o mascaramento deve ter uma força menor para ser inaudível aos ouvidos humanos.

Este capítulo está dividido em quatro seções que contêm a explicação específica de cada processo envolvido e representado no esquema simplificado da Figura 6.1.



Figura 6.1: Esquema geral do sistema implementado

### 6.1: Incorporador da marca

O primeiro passo foi implementar o sistema de marcação com o espalhamento na frequência. O algoritmo de incorporação da marca é bastante simples e foi feito como descrito na Seção 5.1.1. O sinal de informação  $\mathbf{b} = [b(0), b(1), \dots, b(K-1)]^t$  é composta por uma sequência binária, que representa a mensagem que será inserida no sinal hospedeiro  $s(n)$ . A notação  $\mathbf{s}_i(n) = [s(n), s(n+1), \dots, s(n+i-1)]^t$  é utilizada quando queremos representar um *frame* de  $s(n)$  com tamanho  $i$ . O alfabeto foi simplificado e substituído por dois vetores de modulação  $\mathbf{p}_1 = [p(0), p(1), \dots, p(N_p-1)]^t$  e  $\mathbf{p}_0 = [-p(0), -p(1), \dots, -p(N_p-1)]^t$ , que representam as sequências pseudoaleatórias de distribuição gaussiana branca. A sequência  $\mathbf{p}_1$  é usada para representar o *bit* 1 e  $\mathbf{p}_0$  para representar o *bit* 0. O sinal de sincronia  $\mathbf{sync} = [sync(0), sync(1), \dots, sync(N_{sync}-1)]^t$  também é uma sequência pseudoaleatória de distribuição gaussiana branca e é incorporado antes de cada mensagem modulada, assim se sabe quando começa e

termina cada marca.  $N_{sync}$ ,  $N_p$  e  $K$  representam o tamanho do vetor de sincronia, do vetor de modulação de informação e da mensagem em *bits*, respectivamente. Portanto, o sinal de marcação completo tem tamanho  $N = N_{sync} + K * N_p$ . A Figura 6.2 mostra o esquema do incorporador implementado.

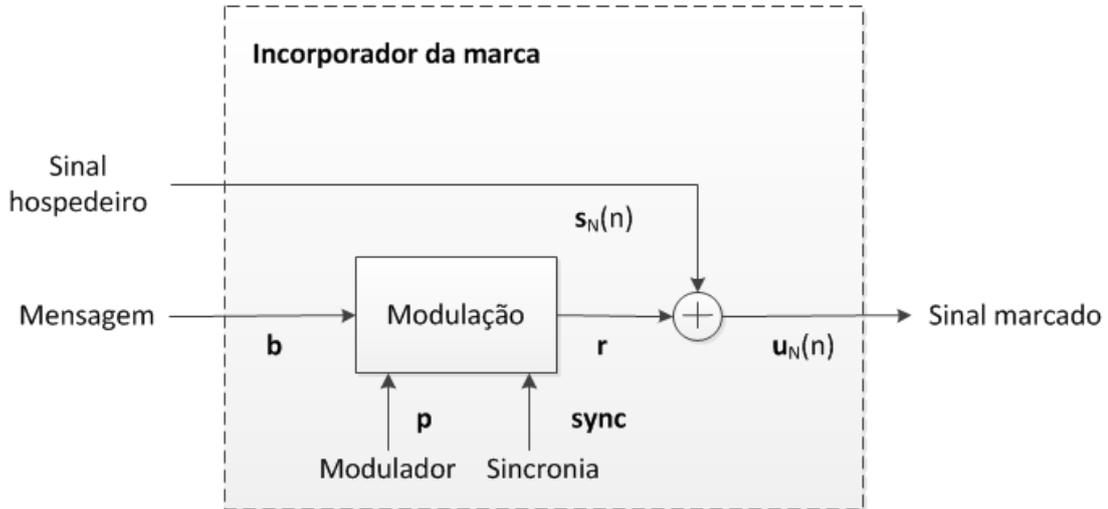


Figura 6.2: Esquema do incorporador da marca

Abaixo segue os passos utilizados para a incorporação:

1. A função de incorporação recebe como argumentos o áudio hospedeiro  $s(n)$  e a mensagem  $b$ .
2. É modulado o sinal da marcação com a sequência de sincronia *sync* seguida de  $K$  sequências  $\mathbf{p}_{b(k)}$ . O sinal de marcação completo é  $\mathbf{r} = [\mathbf{sync}^t, \mathbf{p}_{b(0)}^t, \mathbf{p}_{b(1)}^t, \dots, \mathbf{p}_{b(K-1)}^t]^t$ .
3. A marca é multiplicada pelo fator de escala  $\alpha$  e inserida no sinal hospedeiro  $\mathbf{u}_N(n) = \mathbf{s}_N(n) + \alpha \mathbf{r}$ .
4. Se for necessário, outras marcações podem ser inseridas repetindo o processo a partir do Item 2, com  $n = n + N$ .
5. A função retorna o sinal marcado e o número de marcações inseridas.

## 6.2: Canal de transmissão

O efeito do canal acústico na transmissão da marcação é simulado através da modelagem digital deste canal de transmissão, assim, pode-se isolar e controlar os

efeitos introduzidos no sinal marcado. O uso do modelo economiza tempo e torna os testes da solução implementada mais ágeis e fáceis de mensurar. O modelo é

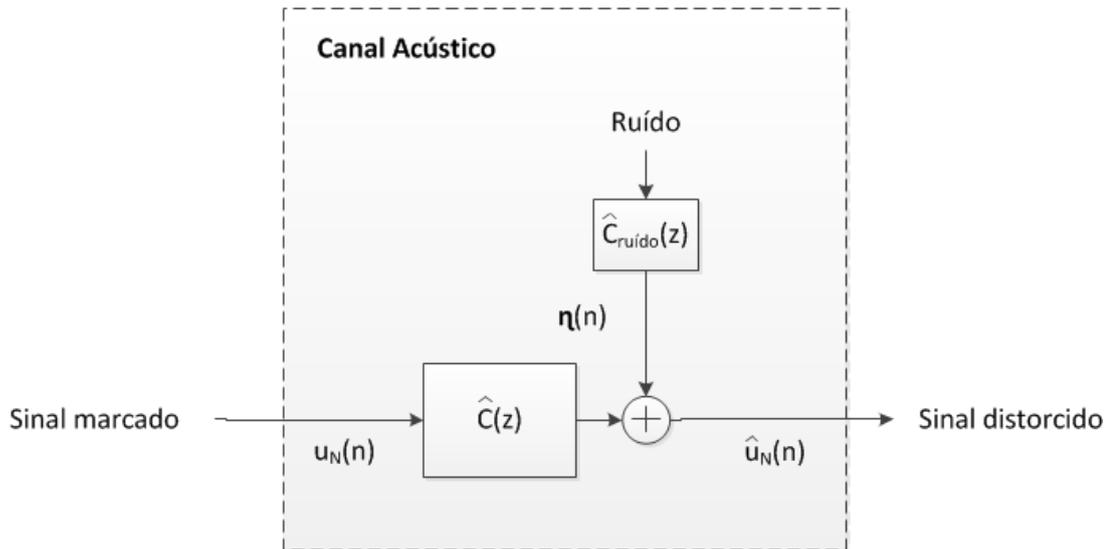


Figura 6.3: Modelo do canal acústico

feito de forma aproximada através da transmissão de um sinal já conhecido pelas caixas acústicas e sua posterior gravação por um microfone conectado ao computador. O sinal transmitido deve ter componentes em toda a banda de frequência para que os efeitos do canal sejam avaliados em todo espectro. Por isso, o sinal escolhido para a transmissão é um ruído branco. O canal convolutivo estimado é um filtro FIR representado matricialmente por  $\hat{c}(n) = [\hat{c}_0(n), \hat{c}_1(n), \dots, \hat{c}_{N_{\hat{c}}-1}(n)]^t$ . O ruído acústico é capturado do ambiente e usado para criar um filtro que dá forma ao ruído branco gaussiano que é adicionado ao modelo. O código para a modelagem foi implementado em MatLab<sup>®</sup> por outro membro da empresa para testes da mesma natureza.

Na Figura 6.4, pode-se observar o espectro de potência de um sinal conhecido passando pelo modelo e o mesmo sinal transmitido pelas caixas acústicas e gravado por um microfone. O modelo e a gravação foram feitas na mesma posição relativa entre o emissor do som e o receptor. A aproximação é considerada boa e válida para testar a solução, dada a proximidade das duas respostas.

Infelizmente, através deste método não se consegue isolar a influência acústica da reverberação. O sinal que é transmitido passa, além do canal acústico, pela conversão AD e DA, cuja influência cria ruídos de quantização, distorce a escala temporal, provoca mudanças no espectro de frequência do áudio, principalmente devido à resposta em frequência da caixa acústica e microfone [4, 10]. Porém, como o filtro adaptativo estima os efeitos aplicados sobre sinal de treinamento, a partir do momento que

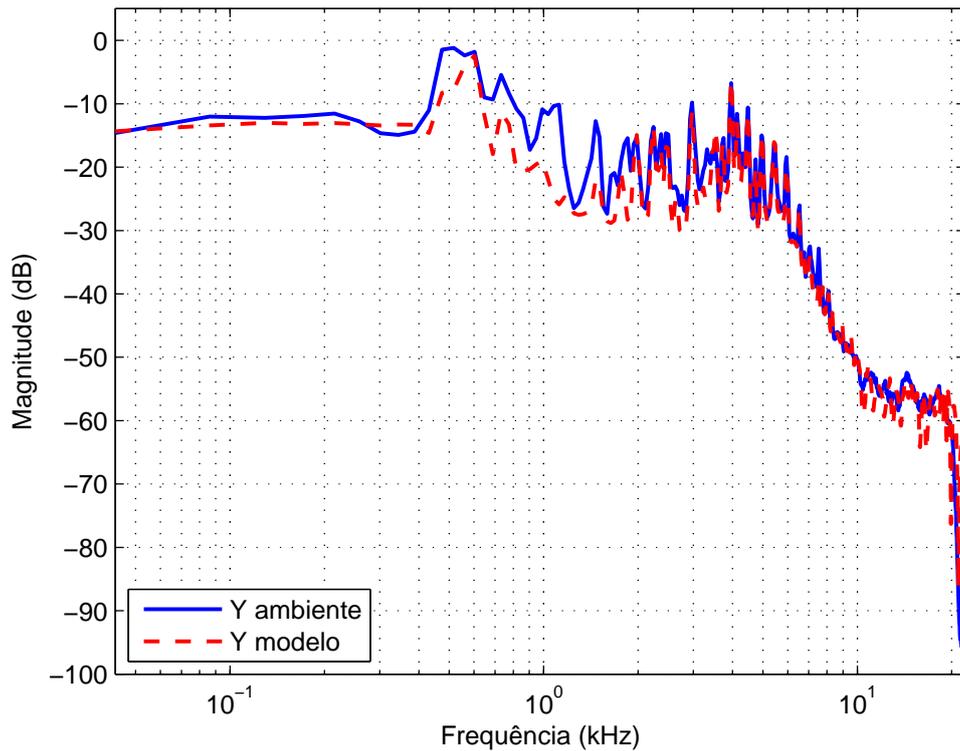


Figura 6.4: Espectro de potência do sinal gravado e sinal estimado pelo modelo

este é inserido no sinal hospedeiro até sua gravação para decodificação, a influência combinada da resposta em frequência das caixas acústicas, canal acústico e microfone é equalizada pelo filtro *zero forcing*.

### 6.3: Equalização *zero forcing*

O filtro *zero forcing* é desenvolvido em duas etapas. Primeiro, estima-se o canal de transmissão do áudio com o auxílio da filtragem adaptativa que em seguida, é utilizado para inverter os efeitos do canal. Foram selecionados os três algoritmos descritos na Seção 4.2.1 para a estimação, são eles: LMS, NLMS e AFA. Os principais motivos para a utilização destes algoritmos foram a simplicidade e baixa complexidade computacional.

Inicialmente, é feita a sincronia do sinal distorcido, tendo em vista que a conversão DA/AD desloca o sinal no tempo. O filtro de Wiener IIR não causal é construído com o PSD do sinal de sincronia, que é previamente calculado, dividido pelo PSD do *frame*  $\hat{u}_{N_{sync}}(n)$ , composto por  $N_{sync}$  amostras.

$$WR(e^{j\omega}) = \frac{PSD(\mathbf{sync})}{PSD(\hat{\mathbf{u}}_{N_{sync}}(n))} \quad (6.1)$$

O filtro é truncado e aplicado sobre o *frame*  $\hat{\mathbf{u}}_{N_{sync}}(n)$  do sinal distorcido para realce da marca de treinamento. Assim que é identificado o pico pela correlação é feita a sincronização das amostras. Se o pico não for identificado, repete-se o procedimento de confecção do filtro de Wiener e filtragem para o *frame*  $\hat{\mathbf{u}}_{N_{sync}}(n)$ , com  $n = n + \frac{N_{sync}}{2}$ , até que o pico seja encontrado.

Sobre o *frame*  $\hat{\mathbf{u}}_{N_{sync}}(n)$  sincronizado são aplicados os algoritmos de filtragem adaptativa para estimar o canal desconhecido. O sinal de sincronia é conhecido e utilizado para avaliar a influência do canal sobre o sinal de sincronia transmitido. O resultado da filtragem adaptativa é o filtro  $w(n)$  que é a aproximação do filtro  $\hat{c}(n)$ , empregado para simular os efeitos do ambiente acústico.

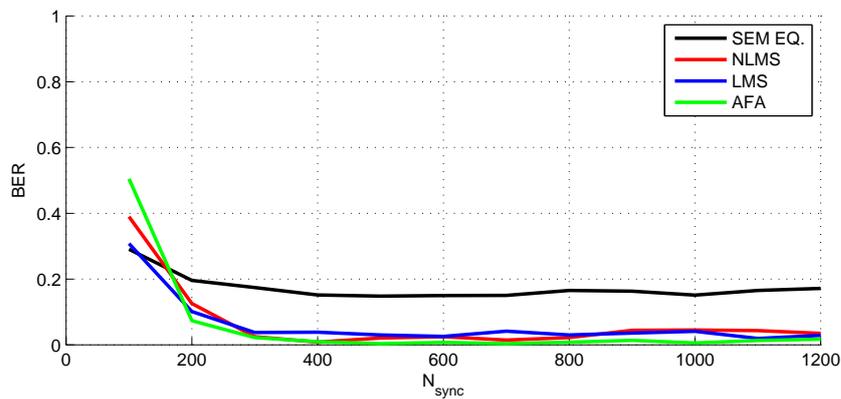


Figura 6.5: *Bit error rate* médio x  $N_s$  para 100 áudios

O número de iterações dos algoritmos é dado pelo tamanho do sinal de sincronia  $N_{sync}$ , que foi escolhido empiricamente através de simulações. Deseja-se que  $N_{sync}$  seja grande o suficiente para permitir a estimação do canal convolutivo e também possibilitar a sincronização do equalizador e do decodificador, e pequeno o suficiente para não desperdiçar a banda de transmissão, pois, do ponto de vista do sistema de comunicação, a sequência de treinamento é uma informação sem utilidade.

Para determinar  $N_{sync}$ , foi gerado o gráfico da Figura 6.5 que relaciona esta variável com o objetivo final da marcação que é a menor BER possível. Com trechos de músicas dos mais variados estilos foram incorporadas com 2000 *bits* de informação e diferentes tamanhos de sequências de treinamento. Assim, consegue-se avaliar o impacto de  $N_{sync}$  sobre a decodificação dos *bits*. Verifica-se que a partir de  $N_{sync} = 400$  há uma estabilização do erro, portanto é escolhido um valor acima deste.

Em seguida, é feita a inversão do filtro adaptativo, com o procedimento apresentado na Seção 5.2.2, resultando no filtro FIR de equalização  $w^{-1}(n)$ . Com isso, o equalizador está pronto para filtrar o *frame*  $\hat{u}_N(n)$ , de tamanho  $N$ , sendo iniciado um novo processo de sincronização, estimação e equalização a cada marcação inserida. Desta forma, a cada  $N$  amostras há uma nova estimação do canal e um novo filtro de equalização.

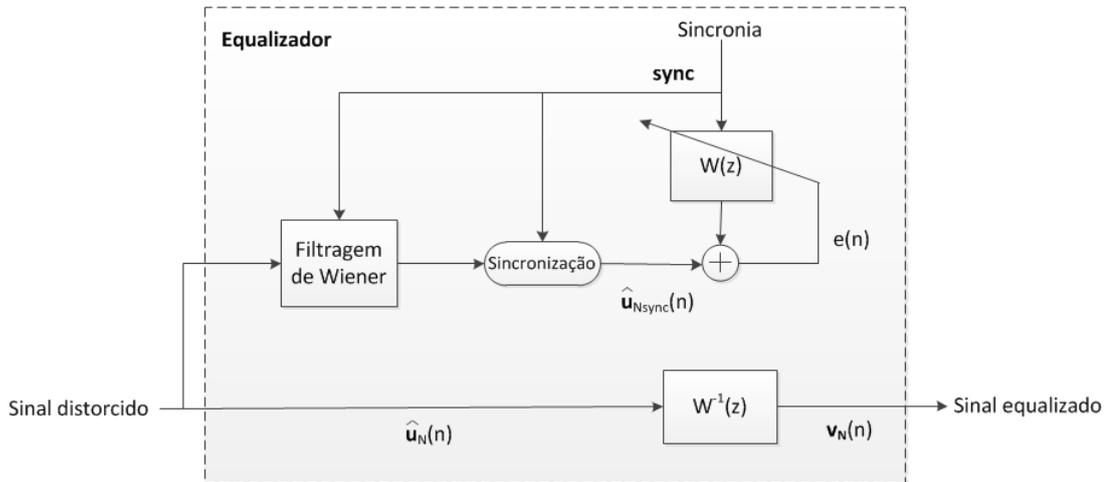


Figura 6.6: Esquema do equalizador

As etapas para a confecção do equalizador adaptativo são:

1. A função recebe o sinal transmitido  $\hat{u}(n)$ .
2. É calculado o PSD de *sync* e o PSD do *frame*  $\hat{u}_{N_{sync}}(n)$ , para criar o filtro de Wiener IIR não causal.
3. O filtro de Wiener é truncado e aplicado no *frame*  $\hat{u}_{N_{sync}}(n)$  para o realce do sinal *sync*.
4. Através do pico da correlação entre o *frame*  $\hat{u}_{N_{sync}}(n)$  filtrado e *sync* é identificado o local em  $\hat{u}_{N_{sync}}(n)$  que o sinal de treinamento se encontra. Se o pico não for encontrado, repete-se os procedimentos a partir do Item 2, com  $n = n + \frac{N_{sync}}{2}$ .
5. O *frame*  $\hat{u}_{N_{sync}}(n)$  sincronizado é conduzido até os algoritmos adaptativos como sinal desejado e *sync* como sinal de entrada.
6. O algoritmo adaptativo é aplicado com  $N_{sync}$  iterações para a estimação das distorções sofridas por  $\hat{u}_{N_{sync}}(n)$ . Com isso, é obtido o filtro  $w(n)$ , de tamanho  $N_w$ , que deve ser a aproximação do canal convolutivo  $c(n)$ .

7. O filtro  $w(n)$  é invertido pelo método de mínimos quadrados e se obtém o filtro *zero forcing*,  $w^{-1}(n)$ , com tamanho  $N'_w$ .
8. O *frame*  $\hat{u}_N(n)$  que contém o marcação com o sinal de sincronia referente a estimação é equalizado pelo filtro *zero forcing*.
9. O processo se repete, a partir do Item 2, com  $n = n + N$ , até que todos os *frames* do sinal distorcido que tenham marcações sejam equalizados.
10. A função retorna o sinal equalizado.

## 6.4: Extrator da marca

O sinal equalizado é utilizado para a extração da mensagem que foi inserida no sinal hospedeiro. Se a equalização tiver sucesso, os efeitos do canal sobre o sinal marcado serão reduzidos e a decodificação será facilitada.

Sincroniza-se a marcação novamente, conforme descrito na Seção 6.3, a filtração de Wiener realça os *frames* em que o sinal de sincronia, *sync*, está presente e pelo pico da correlação se faz o ajuste necessário para a sincronização.

Em seguida, com o sinal sincronizado, parte-se para a extração dos *bits*, com  $n = n + N_{sync}$ . Utiliza-se o filtro de Wiener, construído com o PSD da sequência de modulação  $\mathbf{p}$  e o PSD do *frame*  $\mathbf{v}_{N_p}(n)$  de tamanho  $N_p$ , que contém a sequência de modulação  $\mathbf{p}_{b(k)}$ .

$$H(e^{j\omega}) = \frac{PSD(\mathbf{p})}{PSD(\mathbf{v}_{N_p}(n))} \quad (6.2)$$

Após o realce do sinal de modulação, é feita a correlação entre o sinal realçado e a sequência de modulação  $\mathbf{p}$ . A partir do pico da correlação é decidido se o *bit* ali presente é 1 ou 0. Se o pico for positivo, o *bit* é identificado como 1, se for negativo é 0. A cada  $N_p$  amostras um novo filtro de Wiener é calculado e o *frame*  $\mathbf{v}_{N_p}(n)$ , com  $n = n + N_p$ , é filtrado, assim outro *bit* é decodificado. O processo se repete por mais  $K - 1$  vezes, até que todos os *bits* tenham sido decodificados. Se houver outras marcas no sinal, inicia-se um novo procedimento de decodificação pela sincronia da marcação.

Para a extração as etapas são:

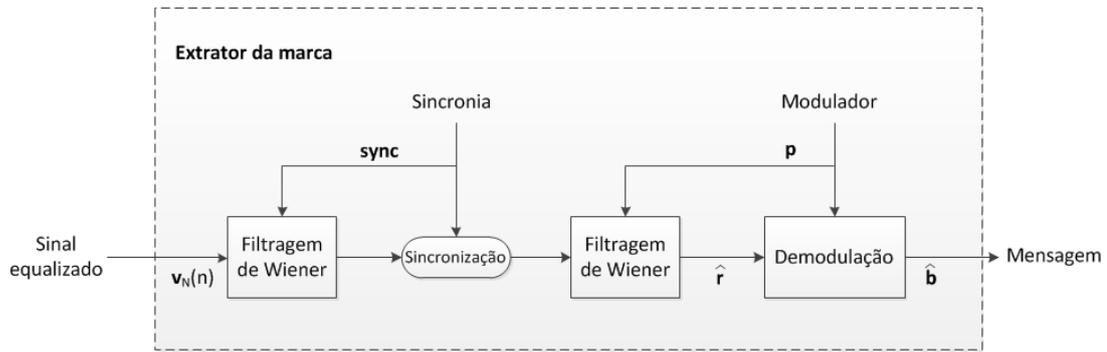


Figura 6.7: Esquema do extrator da marca

1. A função de extração recebe como argumento o sinal equalizado  $v(n)$ .
2. É calculado o PSD das sequências  $sync$  e  $p$ .
3. O PSD do *frame*  $v_{N_{sync}}(n)$ , com tamanho  $N_{sync}$ , é calculado e utilizado junto com o PSD de  $sync$ , para criar o filtro de Wiener IIR não causal.
4. O filtro de Wiener é truncado e aplicado no *frame*  $v_{N_{sync}}(n)$  para o realce do sinal  $sync$ .
5. Através do pico da correlação entre o *frame*  $v_{N_{sync}}(n)$  filtrado e  $sync$  é identificado o local em  $v_{N_{sync}}(n)$  que o sinal de sincronia se encontra. Se o pico não for encontrado, repete-se os procedimentos a partir do Item 3 com  $n = n + N_{sync}/2$ .
6. Parte-se para a extração da informação, com  $n = n + N_{sync}$  usando um novo filtro de Wiener composto pelo PSD da sequência de modulação  $p$ , previamente calculado, e o PSD do *frame*  $v_{N_p}(n)$  sincronizado, com tamanho  $N_p$ .
7. O filtro de Wiener IIR não causal é truncado e aplicado no *frame*  $v_{N_p}(n)$  para o realce da sequência de modulação.
8. É feita a correlação entre o sinal realçado e a sequência pseudoaleatória  $p$ .
9. Se o pico da correlação for positivo, se interpreta o *bit* como 1, caso contrário o *bit* é 0. O processo se repete por  $K - 1$  vezes, a partir do Item 6, com  $n = n + N_p$ , até que todos os *bits* sejam decodificados.
10. Se houver novas marcas no sinal  $v(n)$ , uma nova sincronia é buscada, a partir da posição final da última marcação encontrada. Reinicia-se os procedimentos pelo Item 3.
11. A função retorna os *bits* decodificados.

## Capítulo 7: Resultados

Neste capítulo serão apresentados os resultados obtidos pela implementação da solução descrita nos capítulos anteriores. Foi desenvolvido um equalizador para atuar em conjunto com a marca d'água digital e inverter os efeitos da reverberação do canal acústico, permitindo assim, sua transmissão por este meio. Os resultados obtidos com a implementação da solução serão apresentados em três seções, duas em ambiente simulado, destinadas a mostrar a robustez relacionada à transparência da marcação e a capacidade da solução se adaptar à diferentes canais, e outra com a intenção de testar o equalizador no ambiente acústico real.

O objetivo da solução é tornar a marca mais robusta à transmissão acústica. Robustez é a qualidade que se dá a algo que mantém o desempenho esperado independente das variações do ambiente. No caso da transmissão de uma marca d'água, é esperado que se consiga extrair a informação inserida no áudio hospedeiro sem erros, apesar da influência do canal de transmissão. Um indicador muito utilizado para medir o desempenho de um sistema de comunicação, como é a marcação digital, é o *bit error rate*. O BER aponta a taxa de *bits* decodificados de forma incorreta, quanto maior o seu valor, pior é o desempenho do sistema.

Nos testes simulados foram utilizados 100 trechos de áudios dos mais variados estilos musicais, todos em WAV no formato PCM, com 16 *bits* por amostra, taxa de amostragem de 44100 kHz e mono. Introduziu-se em cada trecho de áudio 2000 *bits* aleatórios de informação. Os resultados são apresentados comparando o desempenho do sistema sem equalização com os sistemas equalizados utilizando a inversão: do canal real, que representa uma estimativa ótima do canal de transmissão, dos canais aproximados pelos diferentes algoritmos de estimativa implementados.

### 7.1: Variações do canal acústico

O canal de transmissão acústico como exposto na Seção 4.1.1 apresenta variações em suas características ao longo do tempo e espaço. As mudanças no canal convolutivo são causadas, principalmente, pela alteração na composição do ambiente no qual as ondas transmitidas por um emissor reverberam até chegar no receptor. É desejável que um sistema que tenha como objetivo estimar e inverter este efeito

consiga aproximar canais de diferentes respostas impulsivas.

Verifica-se o potencial da solução em equalizar o sistema de transmissão em diferentes posições, que conseqüentemente apresentam respostas ao impulso distintas. Foram modeladas posições em 5 distâncias, com o receptor disposto com  $0^\circ$  e  $90^\circ$  em relação ao emissor. Quanto mais distante e mais obstáculos existirem entre as caixas de som e o gravador, mais o sinal de som marcado é atenuado em relação ao ruído ambiente capturado pelo microfone. Portanto, há um aumento da contaminação do áudio pelo ruído ambiente com o aumento da distância da gravação e mudança da direção do microfone de  $0^\circ$  para  $90^\circ$ . Os modelos foram obtidos com a transmissão dos sons através de caixas acústicas genéricas e a gravação pelo microfone de um celular *Sony Xperia Arc*<sup>®</sup>. A Tabela 7.1 mostra o desempenho do sistema na forma de BER médio para os diferentes sistemas.

Distâncias	15 cm		30 cm		60 cm		90 cm		150 cm	
	$0^\circ$	$90^\circ$								
SEM EQ.	0.266	0.325	0.433	0.483	0.326	0.445	0.362	0.490	0.477	0.492
REAL	0.003	0.007	0.001	0.049	0.008	0.039	0.050	0.095	0.006	0.008
LMS	0.110	0.238	0.251	0.343	0.200	0.326	0.275	0.435	0.391	0.465
NLMS	0.187	0.235	0.289	0.338	0.263	0.291	0.256	0.449	0.405	0.484
AFA	0.073	0.181	0.214	0.271	0.216	0.251	0.244	0.439	0.338	0.432

Tabela 7.1: *Bit error rate* médio em relação à distância e ângulo entre emissor e receptor

Percebe-se que há uma melhora nas taxas de decodificação dos sinais que passaram pela solução proposta, sobretudo nas menores distâncias entre o emissor e receptor. De acordo com o aumento da distância as taxas de erro vão subindo e a equalização tem a eficácia reduzida na decodificação correta dos *bits*. Este comportamento também é verificado quando comparamos o sinal que foi obtido com o microfone virado diretamente para a caixa acústica e o que foi obtido com o microfone virado em  $90^\circ$ . Neste caso, as taxas de erro aumentam, porém, em geral, ainda ficam abaixo das taxas de erro obtidas com o sistema sem equalização.

Uma condição para o funcionamento da estimação do canal através da filtragem adaptativa é que se consiga identificar a correlação entre o sinal transmitido pelo canal e o sinal desejado. O aumento do ruído torna difícil para o sistema de estimação encontrar a similaridade entre o sinal que chega ao microfone e o sinal de treinamento. Dessa forma, o estimador não consegue aproximar de forma correta os efeitos do canal acústico no sinal transmitido e por consequência a equalização não tem o desempenho esperado.

Esta condição também é necessária para a sincronização e decodificação, pois estes processos fazem uso da operação de correlação entre o sinal transmitido e as sequências pseudoaleatórias de sincronização e de informação. Isto explica o aumento do erro na extração dos *bits* pelo sistema não equalizado com o aumento da distância e ângulo entre o emissor e receptor.

O sistema equalizado com a inversão do canal real mantém as baixas taxas de erro nos diferentes canais modelados. A resposta em frequência do conjunto canal/equalizador, neste caso, deve ser plana e com ganho unitário, independente do modelo que se esteja testando, portanto, o sinal não sofre atenuação em relação ao ruído quando é transmitido pelo canal simulado e tem a decodificação facilitada.

O modelo feito à 15 cm com  $0^\circ$  entre emissor e receptor, que apresenta as melhores taxas de decodificação, tanto para o sistema equalizado, quanto para o sistema sem equalização, será utilizado nos testes seguintes, como meio de avaliar a influência de outras variáveis sobre a solução implementada.

## 7.2: Robustez e transparência

Do ponto de vista da transparência, a marcação pode ser considerada um ruído que é adicionado ao sinal hospedeiro. Quanto maior for a contribuição de energia do sinal de ruído sobre o sinal hospedeiro, mais evidente para os ouvidos humanos a marcação será, e portanto, menos transparente. O SNR, ou *signal to noise ratio*, é usado para medir a relação de energia entre os sinais. Para sistemas de marca d'água, a nomenclatura *signal to watermark ratio* substitui o termo SNR. Este será o indicador para a transparência da marcação em relação a robustez.

Um gráfico que relacione estes dois indicadores é utilizado para medir a robustez da marca para diversos níveis de transparência. O que se espera com a equalização é alcançar uma transparência maior para um mesmo nível de robustez. Não foi adicionado ruído ambiente para este teste, pois o objetivo é verificar a influência isolada da força da marcação sobre a decodificação dos *bits*. Com isso, a situação ótima de funcionamento da equalização *zero forcing* é posta à prova.

As curvas tomam uma forma crescente, quanto mais fraca e transparente for a marcação, maior o erro na decodificação. Percebe-se que, como desejado, as curvas que representam os sistemas com equalizador estão abaixo da curva do sistema não equalizado. Isso significa que para uma mesma BER, a solução proposta possibilita

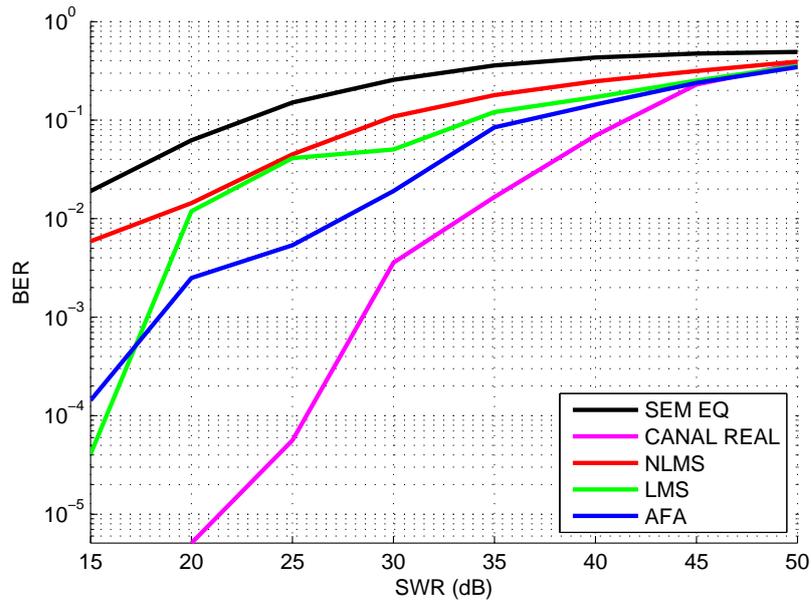


Figura 7.1: *Bit error rate* médio x SWR sem adição de ruído

uma maior transparência ao conjunto. Contudo, quando a marca se torna muito transparente, as taxas de erro dos sistemas que passam pela equalização convergem para a mesma BER do sistema sem equalização, pois a interferência do sinal hospedeiro sobre a marca d'água se torna tão forte que compromete o processo de estimação do canal e de decodificação da marca.

Outro teste proposto, devido às limitações do equalizador *zero forcing* em lidar com ambientes ruidosos, é relacionar a BER com elevação do ruído no canal. A elevação do ruído é dada na forma de SNR e para estes testes o SWR das marcações foi ajustado em torno de 25 dB.

Nota-se que com o ruído severo tanto o sistema que utiliza o equalizador quanto o sem equalização, não conseguem extrair com sucesso a marca d'água, a BER fica em torno de 50%. Mas, a partir de 10 dB de SNR há uma diminuição acentuada da BER com o uso de equalizadores, enquanto o sistema sem equalização reduz a BER de forma muito mais suave. Porém, através da curva do sistema que é equalizado pelo canal real, fica clara a limitação da filtragem *zero forcing* em ambientes ruidosos, mesmo que o canal seja estimado de forma ótima, por não possuir nenhum mecanismo para supressão do ruído ambiente.

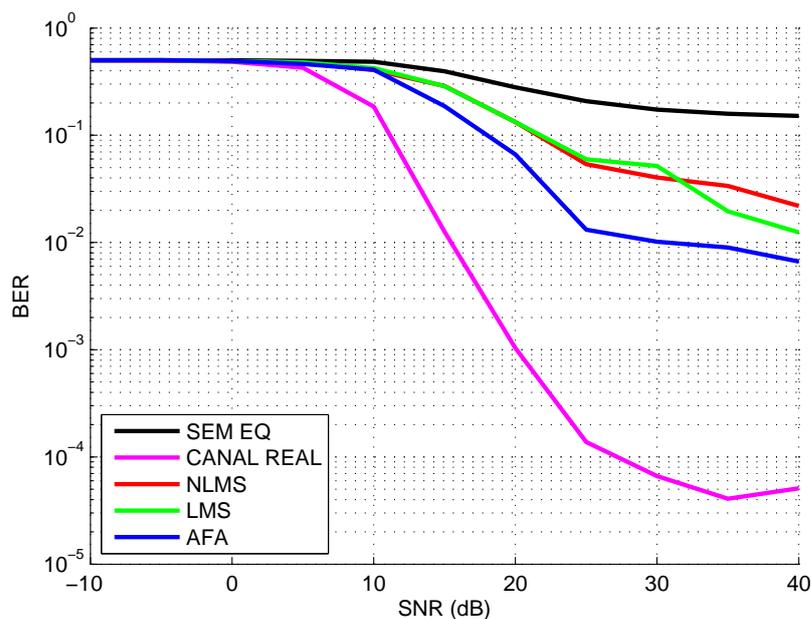


Figura 7.2: *Bit error rate* médio x SNR para SWR = 25 dB

### 7.3: Teste em condições reais

Como teste final para a solução é proposto verificar o desempenho do equalizador em condições reais de funcionamento. O sinal é marcado digitalmente através do código implementado em MatLab<sup>®</sup>, o resultado é um arquivo com extensão WAV contendo o áudio com a marca d'água. Em seguida, este áudio é transmitido via caixas acústicas conectadas ao computador, passando pelo ambiente acústico até ser capturado por um microfone que grava o som no formato WAV com a mesma taxa de amostragem e resolução do arquivo executado anteriormente pelo computador. O áudio contido neste arquivo passa pelo código implementado para a estimação do canal pelo qual foi transmitido, e em seguida pelo processo de extração e decodificação.

Até agora, os efeitos convolutivos do canal de transmissão e o ruído adicionado no sistema, eram isolados e controlados. Na situação proposta, apenas a força da marcação pode ser controlada. A conversão DA/AD insere novas distorções no sinal, nas quais a robustez da solução ainda não foi testada. Espera-se, que apesar destas novas modificações no sinal transmitido, a solução desenvolvida ainda traga ganhos para o sistema na forma de uma menor BER.

Para simular o caso de uso real do sistema de marca d'água foi utilizado caixas acústicas genéricas para executar o áudio e um celular para fazer a gravação. O celu-

lar foi segurado com o microfone virado para as caixas acústicas à aproximadamente 15 cm de distância, onde o ruído ambiente é reduzido em relação ao sinal de interesse e o equalizador tem chances de um melhor desempenho. A informação inserida contém 5000 *bits* divididos em marcações de 100 *bits*. A inserção foi feita em quatro áudios de estilos diferentes. Uma música com diversos instrumentos tocando simultaneamente, uma música com instrumento de corda, outra com um de sopro, e um áudio com um discurso lido por uma voz masculina. Foram feitos 6 ensaios com cada áudio para cada valor de SWR e as gravações foram feitas no ambiente de escritório.

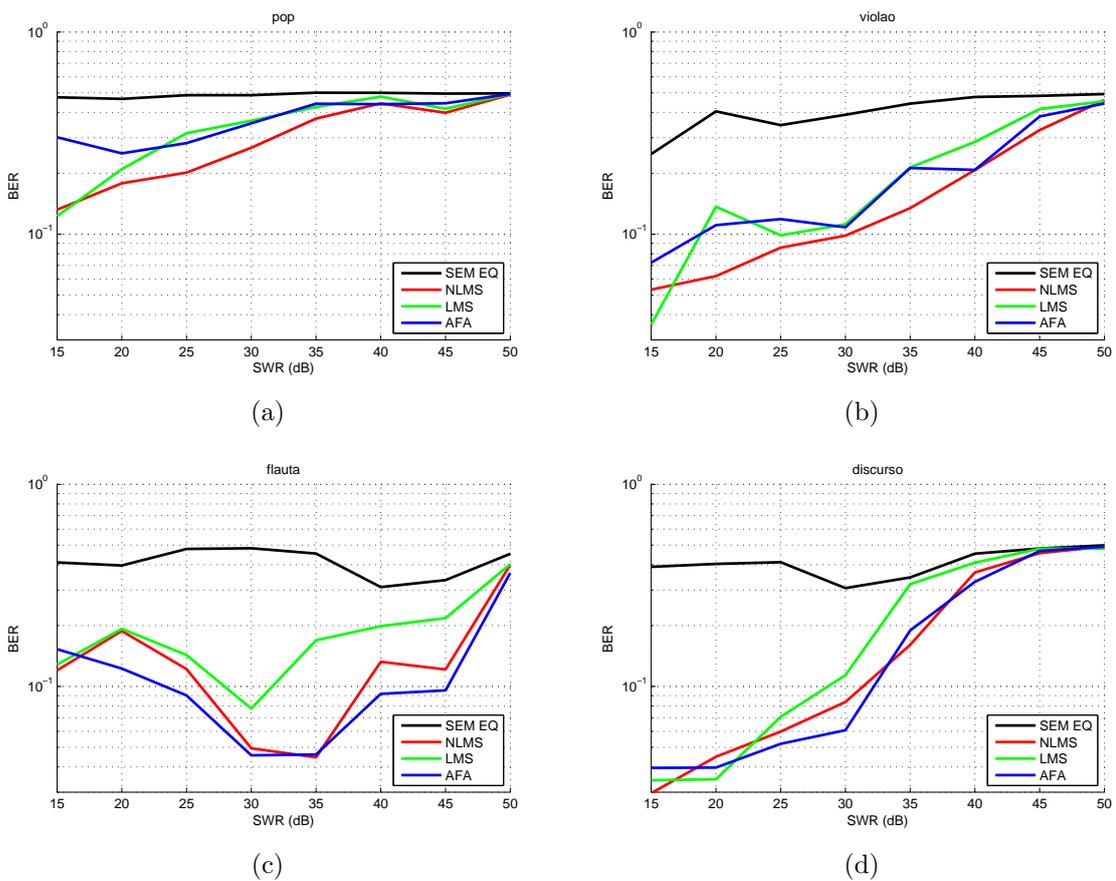


Figura 7.3: *Bit error rate* médio x SWR em ambiente acústico real

Verifica-se que há um ganho com o uso do equalizador. O sinal sem equalização apresenta taxas de decodificação de *bits* sempre próximas à 50%, enquanto os sinais que passaram pela solução proposta chegaram a apresentar taxas menores que 4% de erro. As curvas acompanham a tendência crescente das curvas apresentadas na Figura 7.1, com exceção da curva que representa o sistema sem equalização, que não apresentou taxas de erro menores que 30% e praticamente não obteve sucesso na extração da informação.

A música pop marcada foi a que apresentou as maiores taxas de erro, devido à

grande faixa de espectro ocupada. Em contrapartida, os sons de flauta, violão e o discurso, apresentam sinais mais comportados, limitados à faixas estreitas de frequência, facilitando a identificação da marca.

Testes com um número maior de músicas e canais são necessários para verificar o desempenho da solução em diferentes ambientes e em condições variadas. Porém, nos sistemas testados, tanto em situações simuladas, quanto em situação real de funcionamento, a equalização apresentou uma diminuição dos efeitos maléficos do canal sobre o sinal transmitido, traduzidos em uma menor taxa de erro na decodificação dos *bits* inseridos. Este ganho, porém é limitado, principalmente, pela força da marcação em relação ao sinal hospedeiro e pelo ruído ambiente introduzido no sistema.

## Capítulo 8: Conclusões e Perspectivas

As mídias digitais têm um papel fundamental no mundo moderno. Seja para trazer informações, lazer ou como meio de comunicação, as mídias eletrônicas foram incorporadas ao cotidiano da maioria das pessoas. Elas trouxeram mais agilidade na criação e divulgação de conteúdos e facilitaram processos que antes eram impossíveis ou muito complexos de se fazer no mundo analógico. Além disso, este tipo de mídia pode ser reutilizado inúmeras vezes sem perda de qualidade do conteúdo, diferente das mídias analógicas, que sofriam degradação a cada nova reprodução e cópia.

Neste contexto, a proteção de conteúdos digitais é algo crítico, devido à facilidade de cópia e manipulação destes conteúdos. Para ser mais um dispositivo contra a ação ilegal de terceiros, novos estudos surgiram envolvendo técnicas de marcação de mídias digitais. Neste tipo de proteção uma informação é inserida imperceptivelmente dentro de um sinal hospedeiro, que pode ser um vídeo, imagem ou áudio. Com isso, a marca deve persistir no sinal de mídia transmitido, mesmo que esse seja copiado. Estas características tornam a marcação digital de áudio uma ferramenta interessante para aplicações que se utilizam do meio acústico para transmissão.

A transmissão de sinais marcados digitalmente pelo meio acústico, porém, requer um nível muito alto de robustez da marca. A conversão DA/AD e os efeitos do canal ambiente distorcem o sinal marcado e trazem dificuldades para a decodificação correta da informação inserida. Neste trabalho, foi revisada a ação maléfica dos canais acústicos sobre sinais transmitidos. As ondas sonoras emitidas refletem nos obstáculos e percorrem diferentes caminhos no ambiente, e assim, chegam atrasadas e atenuadas no receptor. Por isso, o canal acústico pode ser considerado um canal dispersivo e há a ação da ISI sobre o sinal transmitido.

A solução proposta para reduzir estes efeitos indesejados do canal acústico sobre o sinal marcado é a equalização *zero forcing*. O canal de transmissão é estimado através de um filtro FIR, cuja a estimação é feita através da filtragem adaptativa com os algoritmos LMS, NLMS e AFA, pois o canal é variante no tempo. É de se destacar o fato de que a transmissão do sinal hospedeiro não precisa ser interrompida para a transmissão do sinal de treinamento do filtro adaptativo, para isso é utilizado a própria sequência de sincronia da marcação. Em seguida, é feita sua inversão e sua aplicação sobre o sinal transmitido, com intuito de deixar a resposta em frequência do sistema

plana e anular a ISI. Apesar de reduzir os problemas com a ISI a equalização *zero forcing* não dispõe de nenhum recurso para a redução do ruído presente na transmissão e pode enfrentar problemas se este estiver presente no sistema.

Contudo, nos testes e simulações feitas, a equalização tornou o sistema mais robusto, indicado pela diminuição da BER no processo de decodificação da informação inserida no áudio. Canais em diversas posições foram modelados e aplicados no domínio digital sobre o sinal marcado para verificar o desempenho do equalizador em condições diferentes. O equalizador trouxe ganhos se comparado com o sistema sem equalização, principalmente quando a distância entre emissor e o receptor é pequena. Também, foi testada a eficácia da solução quando diferentes intensidade de ruído incidem sobre o canal. Há melhoras com o uso de equalizadores, porém, como esperado, em sistemas muito ruidosos a BER se aproxima da decodificação feita sem o processo de equalização e a solução perde a eficácia. Outro fator limitante no uso desta estratégia para a equalização é a intensidade da marcação, tendo em vista que para fazer a estimação do canal considera-se que a sequência de treinamento pode ser identificada no sinal transmitido.

A solução implementada, ainda foi posta à prova em condições reais de funcionamento. Quatro áudios foram marcados, transmitidos pelas caixas acústicas conectadas ao computador e gravados por um celular no ambiente do escritório, após isso o áudio foi transferido para o computador e foi executado o algoritmo de equalização e decodificação do sinal. Nesta situação, o sinal está exposto às distorções causadas pelo microfone, caixas acústicas e placas eletrônicas DA/AD, além dos efeitos do canal acústico. Pode-se dizer que a solução proposta viabilizou a decodificação da informação do sinal transmitido, já que quando o sinal não é equalizado as taxas de erro de *bits* ficam em torno de 50%, porém, com a equalização, consegue-se taxas de erro menores que 10%. O aumento da robustez da marca permite a diminuição da energia da marcação e um ganho na transparência, como era desejado. Porém, a melhora no desempenho da BER ainda não é suficiente para baixar a energia a ponto de deixar a marcação inaudível.

Os resultados foram satisfatórios, tendo em vista que a técnica de marcação não foi desenvolvida para sobreviver a este tipo de ataque e a equalização *zero forcing* é o tipo mais simples de equalização e enfrenta dificuldades quando as distorções causadas pela ISI e por ruído são severas. Apesar de não atingir as taxas de erro apresentadas em [1], o tempo necessário para a estimação do canal, em que é transmitido um sinal de treinamento, é reduzido. Além disso, o mesmo sinal utilizado para o trei-

namento do filtro adaptativo é usado para realizar a sincronia da marcação, que em [1] é desconsiderada.

## 8.1: Perspectivas de trabalhos futuros

Ainda há espaço para o aperfeiçoamento do desempenho da equalização aplicada para melhoria no sistema de marca d'água digital transmitida por meios acústicos. Como a BER alcançada é maior que zero, seria necessário o uso de códigos corretores de erro, com o intuito de corrigir os *bits* decodificados de forma errônea e assim obter a mensagem completa e correta. Os códigos corretores BCH, são de fácil implementação e tem utilização em diversas aplicações, entre elas sistemas de comunicação e sistemas de marca d'água digital. Para a aplicação em questão, é razoável a correção de até 20% de *bits* inseridos. Acima disso, o *payload* da marcação fica comprometido, pois é necessário a transmissão de muitos *bits* corretores para corrigir poucos *bits* de informação útil. Ademais há de se solucionar o problema com a implementação do mascaramento de frequências que pode melhorar a transparência da marca. Também, pode-se aumentar a robustez do sistema com o uso da redundância na marcação, ou seja, inserir várias vezes a mesma informação no sinal hospedeiro, aumentando as chances de decodificação.

O sistema de marcação utilizado apresenta baixa robustez à conversão DA/AD e transmissão via canal acústico. O uso da equalização trouxe claros benefícios à decodificação da marca em condições para as quais não fora desenvolvida. Pode-se estudar a implementação desta técnica em sistemas de marcação mais robustas, desenhadas para enfrentar estes tipos de ataque.

Outra linha de trabalho seria a mudança do sistema de equalização, pois a solução *zero forcing* é considerada a mais simples entre as opções de equalização. Pretende-se adaptar outros equalizadores mais avançados para o sistema de marca d'água, como o equalizador linear pela minimização do erro quadrático médio e o equalizador não-linear com realimentação, o DFE, citados na Seção 4.2. Estes equalizadores prometem melhores resultados, pois consideram a influência do ruído na estimação do filtro de equalização, e portanto, conseguem lidar com distorções mais severas na transmissão do sinal.

## Referências Bibliográficas

- [1] WOLFF, E.; BARAS, C.; SICLET, C. Toward robustness of audio watermarking systems to acoustic channels. In: Proceedings of EUSIPCO. Aalborg, Denmark: [s.n.], 2010. p. 4.
- [2] CVEJIC, N. Algorithms for Audio Watermarking and Steganography. Dissertação (Mestrado) — University of Oulu, 2004.
- [3] COX, I. et al. Digital Watermarking and Steganography. 2. ed. [S.l.]: Morgan Kaufmann Publishers Inc., 2008.
- [4] XIANG, S. Audio watermarking robust against D/A and A/D conversions. EURASIP Journal on Advances in Signal Processing, Springer International Publishing, CH, p. 1–14, 2011.
- [5] HUANG, C.-H. et al. Information delivery systems for car passengers without networking capabilities. In: Proceedings of the multimedia 8th Pacific Rim conference on Advances in multimedia information processing. [S.l.: s.n.], 2007.
- [6] KANG, X.; YANG, R.; HUANG, J. Geometric Invariant Audio Watermarking Based on an LCM Feature. IEEE Transactions on Multimedia, v. 13, n. 2, p. 181–190, 2011.
- [7] QIAN, K. et al. Audio annotation watermarking with robustness against da/ad conversion. In: Media Forensics and Security. [S.l.]: SPIE, 2010.
- [8] TACHIBANA, R. Audio Watermarking for Live Performance. In: Proc. of SPIE Int. Conf. on Security and Watermarking of Multimedia Contents V. [S.l.: s.n.], 2003. p. 32–43.
- [9] ZHANG, X. et al. An audio digital watermarking algorithm transmitted via air channel in double dct domain. In: Multimedia Technology (ICMT), 2011 International Conference on. [S.l.: s.n.], 2011. p. 2926 –2930.
- [10] STEINEBACH, M. et al. Audio watermarking quality evaluation: Robustness to da/ad processes. In: Proceedings of the International Conference on Information Technology: Coding and Computing. [S.l.]: IEEE Computer Society, 2002. (ITCC '02), p. 100.
- [11] MIRANDA, S. G. Sistemas analógicos e digitais. Revista UNORP, v. 4, p. 73–87, 2003.
- [12] XIANG, Y. et al. Effective Pseudonoise Sequence and Decoding Function for Imperceptibility and Robustness Enhancement in Time-Spread Echo-Based Audio Watermarking. IEEE Transactions on Multimedia, v. 13, p. 2–13, 2011.
- [13] GRUHL, D.; BENDER, W.; LU, A. Echo hiding. In: Information Hiding. [S.l.: s.n.], 1996. v. 1174, p. 293–315.

- [14] DUTTA, P.; BHATTACHARYYA, D.; KIM, T.-H. Data Hiding in Audio Signal: A Review. International Journal of Database Theory and Application, v. 2, 2009.
- [15] STEINEBACH, M. et al. Stirmark benchmark: audio watermarking attacks. In: Proceedings Information Technology: Coding and Computing. [S.l.: s.n.], 2001. p. 49–54.
- [16] PLACK, C. J.; OXENHAM, A. J.; FAY, R. R. Pitch: neural coding and perception. [S.l.]: Springer, 2005.
- [17] XIANG, S.; HUANG, J. Analysis of Quantization-based Audio Watermarking to D/A and A/D Conversions. I. J. Network Security, v. 3, p. 230–238, 2006.
- [18] PASSERO, C. R. M. Avaliação e adequação acústica de escritórios panorâmicos através de medições e simulações. Tese (Doutorado) — Universidade Federal do Paraná, 2009.
- [19] NEELY, S. T.; ALEEN, J. B. Invertibility of a room impulse response. Journal of Acoustical Society of America, v. 66, p. 165–169, 1979.
- [20] CAVANAUGH, W.; WILKES, J. Architectural Acoustics: Principles and Practice. [S.l.]: John Wiley & Sons, 1998.
- [21] CARSON, M.; GIESBRECHT, H.; PERRY, T. Surround Sound Impulse Response. July 2009.
- [22] CECCHI, S. et al. An adaptive multiple position room response equalizer in warped domain. In: 20th European Signal Processing Conference. [S.l.: s.n.], 2012.
- [23] PROAKIS, J. Digital Communications. [S.l.]: McGraw-Hill, 2001. (McGraw-Hill Series in Electrical and Computer Engineering).
- [24] JUNIOR, A. N. L.; MIRANDA, M. D. Técnicas de equalização adaptativa usadas em sistemas de comunicação digital. In: MACKENZIE (Ed.). Caderno de Pós-Graduação em Engenharia Elétrica. [S.l.: s.n.], 2003.
- [25] DINIZ, P. Adaptive Filtering: Algorithms and Practical Implementation. [S.l.]: Kluwer Academic Pub, 2002.
- [26] ILIEV, G.; KASABOV, N. Adaptive Filtering with Averaging in Noise Cancellation for Voice and Speech Recognition.
- [27] ILIEV, G.; KASABOV, N. Channel Equalization Using Adaptive Filtering with Averaging.
- [28] LARBI, S.; JAIDANE, M.; MOREAU, N. A new wiener filtering based detection scheme for time domain perceptual audio watermarking. In: Acoustics, Speech, and Signal Processing, 2004. IEEE International Conference. [S.l.: s.n.], 2004. v. 5, p. 949–52.
- [29] TCHESLAVSKI, G. V. Optimum filters. 2008. Disponível em: <ee.lamar.edu/gleb/adsp/lecture 06 - optimum filters.pdf>.
- [30] RUPP, M. Robust design of adaptive equalizers. Signal Processing, IEEE Transactions on, v. 60, n. 4, p. 1612–1626, 2012.