Universidade Federal
do Rio de Janeiro
Escola Politécnica

AVALIAÇÃO DE ALGORITMO E MÉTRICA DE DESREVERBERAÇÃO DE SINAIS DE VOZ

Jéssica do Carmo Soares Veras

Projeto de Graduação apresentado ao Curso de Engenharia Eletrônica e de Computação da Escola Politécnica, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários da obtenção do título de Engenheira.

Orientadores: Sergio Lima Netto e Tadeu Nagashima Ferreira.

Rio de Janeiro

Abril de 2016

AVALIAÇÃO DE ALGORITMO E MÉTRICA DE DESREVERBERAÇÃO DE SINAIS DE VOZ

Jéssica do Carmo Soares Veras

PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO CURSO DE ENGENHARIA ELETRÔNICA E DE COMPUTAÇÃO DA ESCOLA POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE ENGENHEIRA ELETRÔNICA E DE COMPUTAÇÃO

Autora:

Jéssica do Carmo Soares Veras

Orientador:

Prof. Sergio Lima Netto, Ph. D.

Orientador:

Prof. Tadeu Nagashima Ferreira, D. Sc.

Examinador:

Prof. Eduardo Antônio Barros da Silva, Ph. D.

Examinador:

Prof. Thiago de Moura Prego, D. Sc.

Rio de Janeiro

Abril de 2016

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
Escola Politécnica - Departamento de Eletrônica e de Computação
Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária
Rio de Janeiro - RJ CEP 21949-900

DEDICATÓRIA

Dedico este trabalho a minha avó Nedir de Andrade Veras que sempre confiou no meu potencial e acreditou que eu alcançaria todos os meus sonhos, quaisquer que eles fossem.

AGRADECIMENTO

Agradeço aos meus pais Solange e Jefferson, e também ao meu irmão Rodrigo por todo o apoio e confiança durante o curso.

Agradeço a cada um dos meus familiares dentre avós, tios e primos por todas as palavras de conforto, momentos de descontração e sorrisos que tivemos e também aos que ainda estão por vir.

Peço obrigada também aos meus orientadores Sergio e Tadeu, que tiveram muita dedicação e paciência desde o inicio da pesquisa até hoje. Ao Thiago e Amaro que também fizeram parte da equipe de desenvolvimento desse projeto, e sempre foram muitos solícitos em me ajudar quando necessário. Não posso deixar de mencionar os demais professores da UFRJ e do CEFET; onde iniciei meus estudos técnicos e escolhi minha profissão, diga-se de passagem, fortemente influenciada pelo Diego.

Agradeço aos meus amigos de curso, em especial Felipe, Rafael, Michel e João Henrique que me acompanharam tanto nas noites de estudos quanto nas noites de festas. Sem esquecer também dos meus amigos mais antigos que estiveram presentes em muitas fases dessa jornada, principalmente a Jéssica Thiengo que me deu um apoio fundamental no início do curso e ainda o faz.

RESUMO

Este trabalho visa o aperfeiçoamento do sinal de voz, lidando principalmente com os efeitos negativos da reverberação em sinais de fala através de um algoritmo de subtração espectral. Além disso, é feita uma avaliação da qualidade percebida de sinais de voz submetidos ao algoritmo de desreverberação usando métricas como a QAreverb e outras objetivas de estimação de qualidade. Para a validação do processo, foram utilizados sinais providos pelo REVERB Challenge.

A técnica da desreverberação de sinais de voz é composta pelas seguintes etapas: janelamento, FFT, divisão em magnitude e fase, subtração, espectro de potência e IFFT. Pode-se dizer que a principal etapa do algoritmo é o bloco de subtração, que contém 4 parâmetros de ajuste representados por ϵ, a, ζ e ρ. Durante o treinamento do algoritmo, isto é, da escolha do valor dos parâmetros foi utilizada a base New Brazilian Portuguese (NBP) composta de 204 sinais, dentro deste total 4 são sinais anecóicos e 200 sinais reverberados.

A otimização do algoritmo é feita pela tentativa de maximizar ou minimizar, o que for mais conveniente, o valor de determinadas métricas de avaliação de qualidade. Neste trabalho são utilizadas até 8 métricas para julgar as características dos sinais, são elas: Q_{mos}, Relação de energia de modulação de voz para reverberação (SRMR), Distância Cepstral (CD), Razão do log da verossimilhança (LLR), SNR ponderadas em frequência (FWSS), Custo computacional (ATime e RTime) e Razão de palavras erradas (WER). Durante o treinamento do algoritmo foram utilizadas apenas as duas primeiras métricas, mas para a validação do programa todas as medidas foram empregadas.

A validação do processo de desreverberação foi feita durante o REVERB Challenge 2014; um evento internacional de grande prestígio na área de processamento de voz. Os organizadores do desafio ofereceram uma base de teste contendo 4211 sinais com diversas variações em relação à reverberação e a ruído de fundo. Os sinais utilizados no projeto foram criados tanto com simulações quanto com gravações de voz feitas diretamente no ambiente. Eles também variam na distância entre locutor e microfone, além do tamanho da sala em que foram gravados. Os resultados obtidos para cada métrica são detalhadamente apresentados em tabelas de acordo

com as classificações dos sinais.

Uma outra apresentação dos resultados é feita graficamente. A ideia é que o desempenho do algoritmo para uma dada métrica seja ilustrada para nossa equipe junto dos demais grupos participantes do REVERB Challenge 2014. Os projetos variavam principalmente pelo número de canais que o algoritmo utiliza e também pela forma como são agrupados os sinais durante o processo de desreverberação. Este projeto optou por usar sinais com 1 canal e processamento por lote completo de testes. Essa abordagem gráfica oferece uma visão mais ampla do desafio e permite comparar de forma efetiva o desempenho das equipes, de acordo com as ferramentas utilizadas por cada grupo. Os resultados mostram que no geral o sinal é aperfeiçoado, especialmente os sinais reais. Esse comportamento pode ser considerado positivo, pois descreve justamente as situações práticas e por isso de maior interesse.

Palavras-Chave: desreverberação, QAreverb, aperfeiçoamento da voz.

ABSTRACT

This work aims at the improvement of the speech signal, foccusing on the negatives reverberation effects in speech signal through a spectral subtraction algorithm. Also, an assessment of the perceived quality of speech signals subjected to the dereverberation algorithm was completed using metrics such as QAreverb and others. Signals provided by the REVERB Challenge were used to validate the process.

The technique of speech signal dereverberation consists of the following steps: windowing, FFT, magnitude and phase division, subtraction, power spectrum and IFFT. It can be said that the main step of the algorithm is the subtraction block, which contains four tuning parameters represented by ϵ, a, ζ e ρ. During the algorithm training , i.e., the choice of the parameters value, a base called New Brazilian Portuguese (NBP) was used. It consists of 204 signals, 4 of them are anechoic signals and 200 pf them reverberated signals.

The algorithm optimization is done by trying to maximize or minimize, whichever is more convenient, the value of certain quality evaluation metrics. This work used up to 8 metrics to rate the signal characteristics, they are: Q_{mos}, Speech-to-Reverberation Modulation energy Ratio (SRMR), Cepstral Distance (CD), Log-Likelihood Ratio (LLR), Frequency-Weighted Segmental SNR (FWSS), Computational cost (ATime and RTime) and Word Error Rate (WER). During the algorithm training only the first two metrics were used, however for program validation all measures were employed.

The evaluation of the dereverberation process was made during the REVERB Challenge 2014; an international event of great prestige in the voice processing area. The organizers of the challenge offered a test database containing 4211 signals with several variations from the reverberation and background noise. The signals used in the project were created either with simulations and voice recordings made directly in the environment. They also vary in distance between the speaker and microphone, in addition to the room size where they were recorded. The results obtained for each metric are presented in detailed tables according to the signals classification.

Another presentation of the results is done graphically. The idea is to illustrate the algorithm performance for a given metric either for our team or for other participating groups of the REVERB Challenge. The projects differed mainly by

the number of channels that the algorithm used and also by the way signals are grouped during the dereverberation process. This project chose to use signals with 1 channel and full batch processing. This graphical approach gives a broader view of the challenge and allows to compare effectively the performance of the teams, according to the tools used by each group. The results show that in general the signal is improved, especially real signals. This behavior can be considered positive, because it precisely describes the practical situations and therefore of interest.

Keywords: dereverberation, QAreverb, voice enhancement.

SIGLAS

CD - *Cepstral Distance*

EDC - *Energy Decay Curve*

FDR - *Free Decay Region*

FWSS - *Frequency-Weighted Segmental SNR*

LLR - *Log-Likelihood Ratio*

MOS - *Mean Opinion Score*

PESQ - *Perceptual Evaluation of Speech Quality*

REVERB Challenge - *REverberant Voice Enhancement and Recognition Benchmark* Challenge

SRMR - *Speech-to-Reverberation Modulation energy Ratio*

UFRJ - Universidade Federal do Rio de Janeiro

WER - *Word Error Rate*

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

O estudo de sinais de voz é uma área muito grande de pesquisa, pois tem aplicações em diversos segmentos como: telecomunicações, entretenimento, medicina e outras. Dentro da área de processamento de sinais, o tópico desreverberação de voz vem ganhando atenção nos últimos anos, e é justamente sobre esse assunto que iremos nos concentrar neste trabalho.

Este projeto está voltado para a desreverberação de um sinal de fala, ou seja, uma redução no efeito da reverberação em um sinal de voz. A reverberação nada mais é que uma alteração que o ambiente insere no sinal, associada às reflexões múltiplas que um sinal sofre no dado ambiente.

A desreverberação de sinais pode ser aplicada em diversas situações: teleconferências, reconhecimento de voz em geral operando em ambiente fechado, ou até mesmo locais com características acústicas especiais como auditórios e teatros. Desta forma, se faz necessário um bom sistema de desreverberação para que a inteligibilidade da informação não seja comprometida.

A proposta deste projeto é validar a eficiência do algoritmo de desreverberação baseado em subtração espectral, assim como testar o avaliador de qualidade QAreverb.

Os testes foram feitos durante um desafio internacional de grande prestígio na área de processamento de voz, o que permite comparar o desempenho desse e algoritmo com outros propostos por várias equipes do mundo. Nesse cenário foram utilizados sinais com diversas características diferentes, como por exemplo a distância entre locutor e microfone, o tamanho das salas e a origem do sinal.

Os resultados do trabalho mostram que o desempenho do algoritmo foi intermediário, e podem ser considerados ainda melhores quando nos restringimos aos sinais reais. Essa situação é a mais interessante na prática, já que nas situações cotidianas só temos disponiveis os sinais degenerados, e não os anecóicos.

1.1 Descrição do trabalho

Os tópicos abordados nesse projeto serão apresentados na seguinte ordem:

No capítulo 2 será discutido em mais detalhes o que é o fenômeno da reverberação e suas principais variáveis como tempo de reverberação, variância espectral do ambiente e razão de energia direta sobre reverberante.

O capítulo 3 descreve como é obtida a métrica de avaliação da qualidade QAreverb. A seção descreve os cinco estágios do algoritmo do QAreverb, que são: pré - processamento, desconvolução, cálculo dos parâmetros, cálculo da métrica e mapeamento. Nesta seção também são ressaltadas as principais semelhanças e diferenças entre a versão padrão e a versão cega do QAreverb.

O capítulo 4 aborda o processo utilizado para combater a reverberação, ou seja, para realizar a desreverberação do sinal. O algoritmo possui várias fases que são: janelamento, FFT, divisão em módulo e fase, subtração espectral, cálculo do espectro e por fim IFFT. Ainda nessa seção são apresentados os quatro parâmetros de ajuste do algortimo: ϵ, a, ζ e ρ e também é explicado como funciona o treinamento para obtenção desses valores.

O capítulo 5 discute o trabalho realizado no âmbito do desafio internacional The REVERB (REverberant Voice Enhancement and Recognition Benchmark) Challenge 2014. Nessa seção é apresentada em detalhes a base de teste e a classificação dos sinais que a compõe. Os sinais variam principalmente quanto a origem, a distância locutor-microfone e também em relação ao tamanho da sala em que foram gravados. No final do capítulo são apresentadas todas as métricas empregadas no desafio: Q_{mos}, Relação de energia de modulação de voz para reverberação (SRMR), Distância Cepstral (CD), Razão do log da verossimilhança (LLR), SNR ponderadas em frequência (FWSS), Custo computacional (ATime e RTime) e Razão de palavras erradas (WER).

O capítulo 6 mostra os resultados obtidos durante o REVERB Challenge 2014

Em seguida é feita uma análise desses valores, e são indicadas possiveis razões que levaram a esses resultados. Para que se possa fazer uma comparação, são apresentados alguns gráficos para cada métrica com a performance de todos os algoritmos participantes do desafio.

Concluindo o trabalho, no Capítulo 7, discutimos o desempenho do avaliador automático no experimento e os principais problemas que ocorreram. Nesta seção também são apresentadas possíveis tópicos a se desenvolver visando a continuidade do trabalho.

Capítulo 2

Reverberação

2.1 Introdução

A reverberação de um sinal de áudio é um fenômeno inerente a quase todos os ambientes. É importante frisar que o sinal reverberado é escutado como um único sinal pelo ouvinte, o contrário do que ocorre com o eco em que é possível distinguir o sinal original de suas cópias atrasadas.

A intensidade da reverberação de uma sala varia de acordo com certas características. Uma delas é o volume do ambiente, pois quanto maiores as salas, maior o efeito da reverberação no sinal emitido neste local. A variação no nível de reverberação pode ser também devido à geometria do ambiente. Outro fator relevante é o material utilizado na construção que possui coeficiente de absorção dependente de sua natureza e que varia conforme a faixa de frequência do sinal. A madeira por exemplo possui um coeficiente de absorção dentre os mais altos ao contrário do vidro e do mármore que são mais reflexivos.

Existem três parâmetros de um ambiente que são essenciais para estudar as propriedades acústicas de um determinado local, são eles: tempo de reverberação, variância espectral do ambiente e razão de energia direta sobre reverberante que serão detalhados adiante.

2.2 Conceito de reverberação

A reverberação de um sinal de voz pode ser entendida como o efeito gerado pela interação do sinal original com os vários caminhos possíveis na sala entre a fonte e o ouvinte.

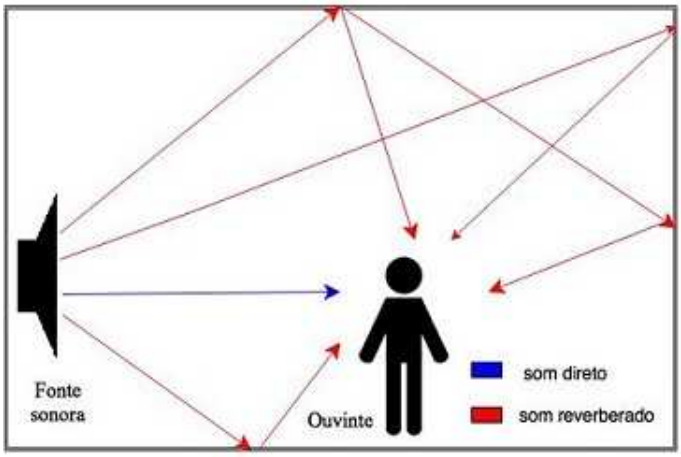

Figura 2.1: Imagem ilustrando os caminhos refletidos e direto entre a fonte sonora e o ouvinte.

A Figura 2.1 representa um ambiente fechado que contém uma fonte sonora e um ouvinte. Nesse cenário o som emitido pela fonte pode tanto alcançar o ouvinte por um caminho direto (linha azul) como através de caminhos alternativos (linhas vermelhas). Estes últimos são os percursos feitos pelas reflexões sofridas no teto, no chão e nas paredes e que são os principais responsáveis pelo efeito da reverberação no sinal.

Essas alterações feitas no sinal de áudio pelo ambiente podem ser caracterizadas pela resposta ao impulso da sala (RIR, do inglês *room impulse response*), como sugerem Neely e Allen [1]. Esse fenômeno é descrito pela seguinte expressão matemática:

$$s_r(t) = \int_0^\infty h(\tau)s(t-\tau)\,\mathrm{d}\tau, \tag{2.1}$$

onde $s(t)$ é o sinal original de aúdio, $s_r(t)$ é o sinal reverberante e $h(t)$ é a resposta ao impulso da sala.

2.3 Tempo de reverberação

O tempo de reverberação refere-se ao tempo necessário para um sinal deixar de ser percebido em um ambiente após sua emissão ter sido cessada. Uma definição mais difundida é a do tempo transcorrido até que sua potência seja reduzida em 60 dB e por isso um símbolo comumente utilizado para representar essa grandeza é T_{60}.

O método adotado neste trabalho para calcular essa medida foi desenvolvido por Schroeder [6]. O primeiro passo é estimar a resposta do ambiente quando se tem um pulso breve como entrada. Em seguida, traça-se uma curva de decaimento de energia (EDC, do inglês *Energy Decay Curve*) normalizada que é dada por:

$$EDC(t) = 10\log_{10}\left(\frac{\int_t^\infty h^2(\tau)\,\mathrm{d}\tau}{\int_0^\infty h^2(\tau)\,\mathrm{d}\tau}\right)[dB]. \tag{2.2}$$

Com uma aproximação desta curva é possível obter uma função de primeiro grau $r(t)$ que passa pelos pontos de -5 dB e o ponto de limiar de ruído [6] [7] [8]. E por fim deslocamos a reta $r(t)$ de forma a passar pela origem e gerar a reta $s(t)$ onde $s(T_{60}) = -60$ dB é o ponto desejado.

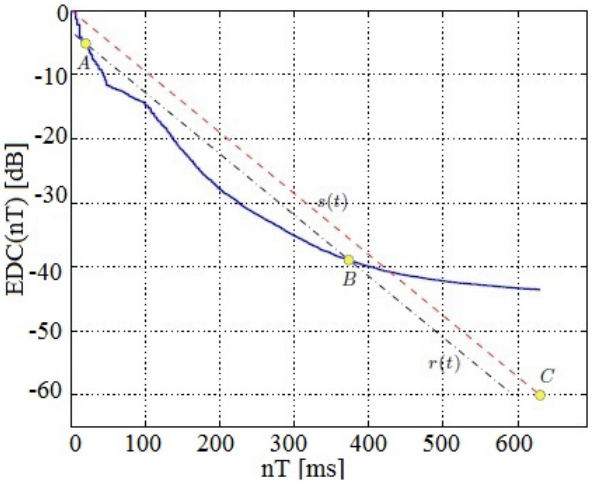

Figura 2.2: Gráfico com a função EDC e as retas $r(t)$ e $s(t)$ utilizadas para obter o T_{60} a partir do algoritmo de Schroeder [6]. Fonte [2].

A figura 2.2 acima ilustra um caso em que a EDC (linha cheia azul) para uma dada $h(t)$ possui os pontos A(0; -5) e B(373; -39). Este último ponto é obtido de acordo com a teoria proposta por Lundeby [7], que busca definir a inclinação da reta

$r(t)$ (tracejada e pontilhada preta) cuja aproximação de primeira ordem escolhida gere o menor valor de erro quadrático médio entre a EDC e $r(t)$. A partir de $r(t)$ é gerada a reta $s(t)$ que passa pela origem (0,0). Essa nova reta é necessária para se manter a função coerente com o caso real, ou seja, quando o som ainda não foi emitido, em $t = 0$, a energia EDC é nula. Nessa nova reta $s(t)$, já podemos buscar o ponto de interesse, representado pelo ponto com nível de energia correspondente a -60 dB. A coordenada encontrada é C (630; -60), com isso concluímos que $T_{60} =$ 630 ms.

2.4 Variância espectral da sala

Enquanto o T_{60} é uma medida de caracterização da reverberação no domínio do tempo, a variância espectral faz algo análogo mas no domínio da frequência. Jetz [10] desenvolveu uma forma de aferir a variância espectral que será descrita em mais detalhes a seguir.

Primeiramente devemos calcular a intensidade relativa $I(f)$. Dado que $H(f)$ é a transformada de Fourier da resposta ao impulso do ambiente, o cálculo e feito usando a seguinte fórmula:

$$I(f) = 10 \log_{10} \left(\frac{|H(f)|^2}{\int_{-\infty}^{\infty} |H(f)|^2 \, \mathrm{d}f} \right) [dB]. \tag{2.3}$$

Definindo $\overline{I(F)}$ como :

$$\overline{I(f)} = \int_{-\infty}^{\infty} I(f) \, \mathrm{d}f. \tag{2.4}$$

Podemos então calcular a variância espectral da sala que é dada por:

$$\sigma_r^2 = \int_{-\infty}^{\infty} (I(f) - \overline{I(f)})^2 \, \mathrm{d}f. \tag{2.5}$$

2.5 Razão de Energia Direta sobre Reverberante

Para o cálculo da energia direta sobre a reverberante precisamos definir um tempo t_d que é associado ao instante de maior valor da função de resposta ao impulso da sala $h(t)$.

A razão E_{dr} é dita como a razão entre a energia direta E_d (em torno de t_d) e a energia reverberante E_r (todo o restante) de $h(t)$, ou seja:

$$E_{dr} = \frac{E_d}{E_r} = \frac{\int_{t_d-t_1}^{t_d+t_2} h^2(\tau)\,\mathrm{d}\tau}{\int_{t_d+t_2}^{\infty} h^2(\tau)\,\mathrm{d}\tau}, \tag{2.6}$$

em que t_1 e t_2 delimitam um intervalo em torno de t_d associado a componente direta do sinal. Valores típicos para t_1 e t_2 são de 1 e 1,5 ms respectivamente [2].

As figuras abaixo mostram a resposta ao impulso de uma sala $h(t)$ produzida de duas formas:

- Artificialmente

  Nesse caso a entrada $s(t)$ é um impulso e a saída é igual à função de transferência $h(t)$, obtida através da resposta ao impulso do ambiente:

$$h(t) = \int_0^{\infty} h(\tau)\delta(t-\tau)\,\mathrm{d}\tau, \tag{2.7}$$

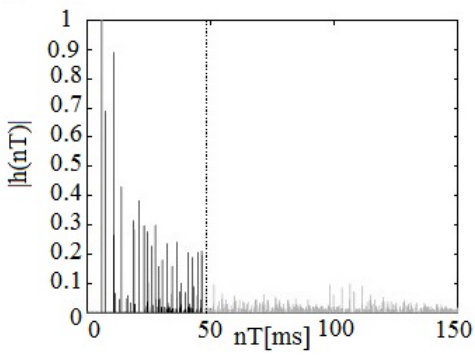

Figura 2.3: Exemplo de RIR artificial com primeiras reflexões em destaque e as reflexões tardias sombreadas. Fonte [2].

No exemplo deste sistema, utilizando a função de transferência artificial encontramos que t_d vale 8 ms;

- De maneira real

  Nesse processo $h(t)$ pode ser obtida através da transformada inversa de Fourier da razão das transformadas de Fourier entre sinais reverberado e não reverberado:

$$h(t) = IFFT\left[\frac{FFT[s_r(t)]}{FFT[s(t)]}\right]. \tag{2.8}$$

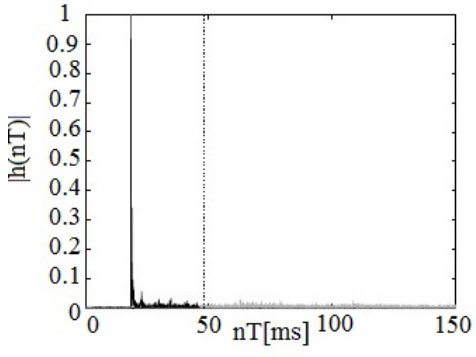

Figura 2.4: Exemplo de RIR real com primeiras reflexões em destaque e as reflexões tardias sombreadas. Fonte [2].

No exemplo deste sistema, utilizando a função de transferência real encontramos que t_d vale 20 ms.

Nas figuras 2.3 e 2.4, as amostras realçadas representam as primeiras reflexões e as demais amostras representam a reverberação tardia.

Kuster [11] diz que para reduzir o ruído é recomendável utilizar componentes de sinal 20 dB acima do ruído. Também é sugerido que o acúmulo de energia seja suspenso no mesmo ponto de parada definido pelo algoritmo do T_{60}.

2.6 Conclusão

Neste capítulo foi visto o que se entende por reverberação, como ela é originada e quais são seus principais efeitos em um sinal de áudio que no geral comprometem a inteligibilidade e por isso são indesejados.

Além disso, foram mostradas quais as variáveis que interferem na quantidade de reverberação de um sinal de voz e como calculá-las, destacando-se: tempo de reverberação (T_{60}), variância espectral da sala (σ_r^2) e razão de energia direta sobre reverberante (E_{dr}). A seguir será detalhado como fazer uso dessas grandezas para medir a qualidade do sinal de interesse.

Capítulo 3

QAreverb

3.1 Introdução

Para que se possa mensurar o quão melhor ou, apesar de indesejado, o quanto pior o sinal tratado pelo algoritmo de desreverberação está em relação à sua versão inicial são utilizadas diferentes métricas.

Em particular neste trabalho, usamos a métrica Q_{mos} derivada do sistema QAreverb proposto por Prego [2]. Neste capítulo são mostrados os princípios básicos do sistema QAreverb e sua variante cega, que utiliza apenas o sinal reverberante.

3.2 QAreverb

O sistema QAreverb é uma ferramenta para o estudo da reverberação de sinais. Esse sistema possui 5 principais estágios: pré-processamento, desconvolução, cálculo dos parâmetros, cálculo da métrica e mapeamento.

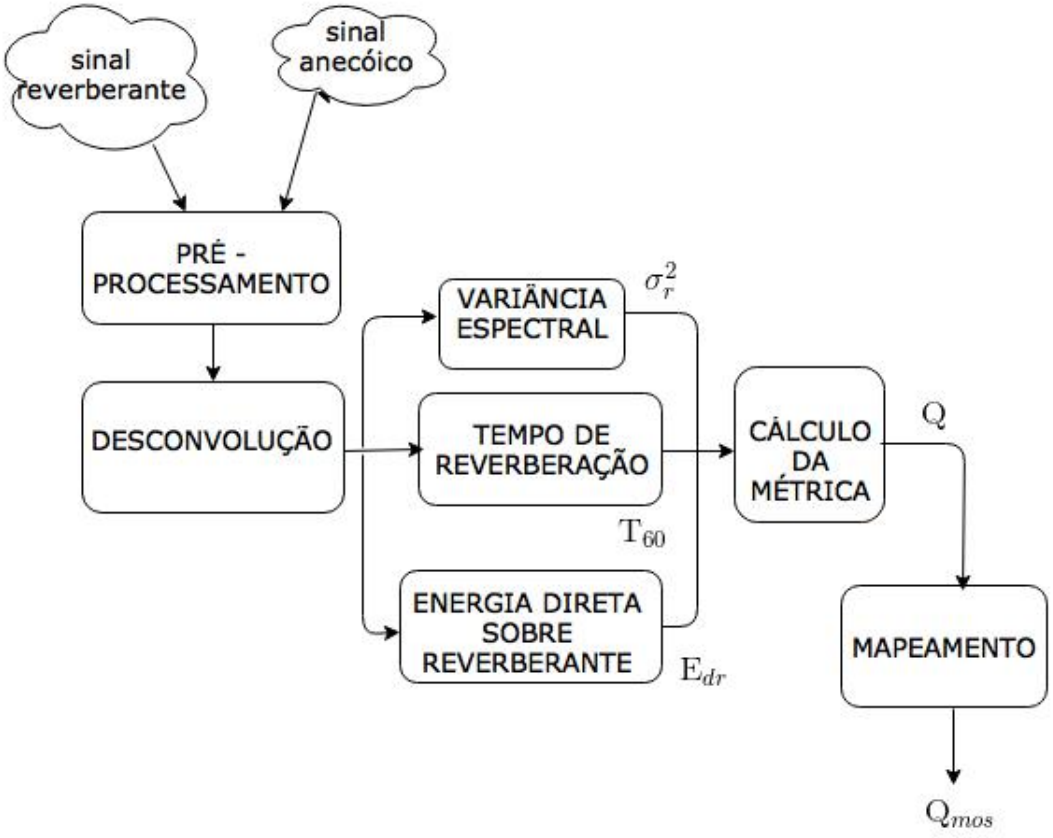

Figura 3.1: Diagrama de blocos ilustrando o processo de cálculo da métrica Q_{mos}.

No primeiro estágio o sistema remove o nível médio dos sinais reverberante e anecoico gerando respectivamente $s'_r(n)$ e $s'(n)$.

Em seguida, durante a desconvolução, estima-se a resposta ao impulso $\hat{h}(n)$ usando a mesma proposta da equação (2.8), porém reescrita no domínio do tempo discreto dada por:

$$\hat{h}(n) = IDFT\left[\frac{DFT[s'_r(n)]}{DFT[s'(n)]}\right]. \tag{3.1}$$

Com essa nova equação surge a necessidade de criar um limiar ξ para que caso o denominador $S'(k) = DFT[s'(n)]$ tenha um valor muito pequeno, alteremos para $S'(k) = \xi$, isto é :

$$|S'(k)| < \xi \Rightarrow S'(k) = \xi, \text{ válido } \forall k. \tag{3.2}$$

Deve-se ressaltar que ϵ é ajustado para cada base de treinamento.

As três últimas fases do sistema QAreverb podem ser mais facilmente entendidas quando descritas em conjunto, são elas: cálculo dos parâmetros, cálculo da

métrica e mapeamento.

O desenvolvimento de uma medida para avaliação da qualidade da desreverberação é um dos principais objetivos do sistema, e para isso é necessário o cálculo de certas grandezas.

Na área de reverberação de sinal existem 3 parâmetros que se destacam na literatura, são eles: tempo de reverberação T_{60} por Karjalainen [9], a variância espectral σ_r^2 por Jetz [10] e a energia direta sobre reverberante E_{dr} por Kuster [11]. Por isso mesmo o QAreverb faz uma combinação dessas variáveis utilizando os algoritmos dos pesquisadores citados acima para obter o avaliador Q definido como:

$$Q = \frac{-T_{60}\sigma_r^2}{E_{dr}^{\gamma}}, \tag{3.3}$$

sendo $\gamma = 0{,}3$ um valor de ajuste encontrado empiricamente através de testes por Prego [2].

Em seguida, com o intuito de facilitar a sua interpretação, o valor de Q é mapeado e definido como Q_{mos} (mos do inglês, *mean opinion score*) numa escala que varia entre 1 (muito reverberado) e 5 (idealmente sem reverberação).

3.3 QAreverb Cego

Geralmente em uma situação real não se têm disponíveis sinais anecóicos, e sim apenas o reverberado. Além disso vale ressaltar que estes sinais são considerados discretos no tempo por isso a notação adotada neste trabalho é $s(n)$.

Desta necessidade de medir a qualidade de reverberação em um sinal sem a sua versão limpa surge o QAreverb cego. Para determinar os valores dos parâmetros T_{60}, σ_r^2 e E_{dr} o sistema utiliza técnicas um pouco diferentes das descritas anteriormente e que serão mais detalhadas a seguir.

3.3.1 Tempo de reverberação sem referência

Dentre os parâmetros da nota Q, um dos mais explorados pela comunidade científica é o cálculo do T_{60}. Várias técnicas já foram apresentadas, mas todas partem do princípio de modelar uma função exponencial decrescente e sua constante de decaimento através do sinal $s_r(n)$ conforme ilustrado no Capítulo 2.

O que varia entre os algoritmos é se a estimativa do T_{60} será a partir do sinal completo como sugere Ratnam [13] e [14] ou de apenas um trecho dele, conhecido como região de decaimento livre (FDR, do inglês *free decay region*) apresentado por Vieira [15]. As FDRs podem ser entendidas como trechos do sinal com energia sonora decresce em diversas amostras consecutivas.

Uma alternativa é o algoritmo utilizado neste trabalho proposto em [2]. Esse processo também adota as FDRs mas faz isso dentro de cada sub-banda do sinal. Essas regiões são obtidas pela decomposição em frequência do sinal que fornecem estimativas parciais de T_{60}.

A figura 3.2 ilustra o processo para uma gravação real numa sala com $T_{60} = 0,7$ e distância entre locutor e microfone de 100 cm.

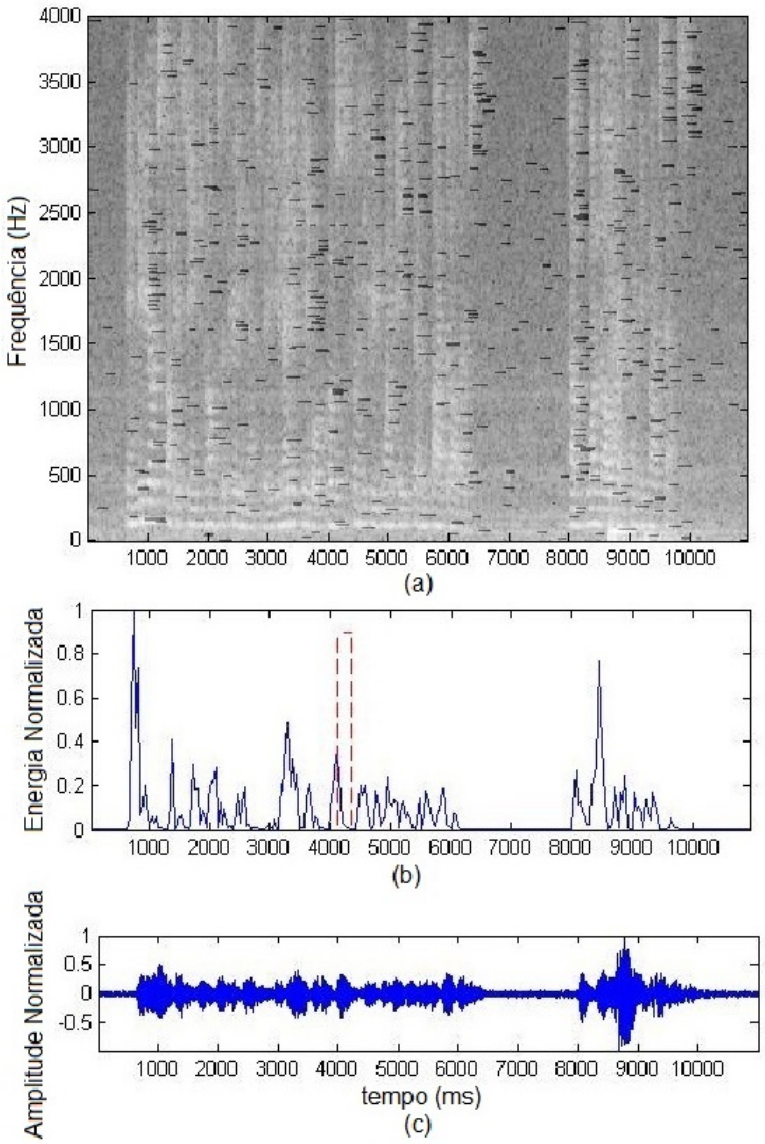

Figura 3.2: Distribuição das FDRs em sub-bandas: (a) Sinal no domínio da frequência mostrando cada sub-banda e suas correspondentes FDRs representadas pelas linhas escuras; (b) Energia normalizada para a sub-banda com frequência central em 132 Hz e em destaque a FDR com linhas tracejadas; (c) Amplitude normalizada do sinal de fala no domínio do tempo.

É interessante notar que a distribuição das FDRs (linhas horizontais pretas) para cada sub-banda tem uma forma particular, mas sempre predominam nos instantes iniciais do silêncio. Isso acontece pois é nesses intervalos que o efeito da reverberação se destaca.

Em seguida, é feita uma análise estatística a partir dos valores parciais de

T_{60} de cada sub-banda para gerar a estimativa final do parâmetro.

Supondo que foram obtidas R_k FDRs na k-ésima sub-banda, cada estimativa parcial do tempo de reverberação pode ser denotada por $T_{60}^s(r;k)$, para r = 0, 1, ..., $(R-1)$. A partir destes valores pode-se calcular a mediana $\hat{T}_{60}^s(r)$ para a dada banda.

O processo é repetido para todas as sub-bandas gerando k estimativas $\hat{T}_{60}^s(k)$ que após um novo cálculo de mediana produzem o valor $\bar{T}_{60}^s$ de banda completa. Um mapeamento de $\bar{T}_{60}^s$ se faz necessário para que os valores encontrados variem no mesmo intervalo dinâmico da base de referência. No sistema proposto, usamos um mapeamento do tipo:

$$\tilde{T}_{60}^s = \alpha_{nr}\bar{T}_{60}^s + \beta_{nr}, \tag{3.4}$$

em que α_{nr} e β_{nr} são dois coeficientes obtidos durante o treinamento da base e que não afetam a correlação entre as estimativas e os valores de referência.

3.3.2 Variância espectral sem referência

O modelo utilizado para calcular σ_r^2 de forma cega foi proposto por Habets [16]. O artigo apresenta a mesma ideia de usar a transformada de Fourier discreta de $s_r(n)$ e $h(n)$ para buscar FDRs e em seguida calcular a variância desse sinal em pequenos intervalos de frequência.

Assume-se que $S_r(k,l)$ e $H(k,l)$ são as STFTs (do inglês, *Short-Time Fourier Transform*) do sinal reverberado e da RIR janelados com uma função de Hamming de tamanho M, sobreposição de V amostras e frequência de amostragem F_s. Definem-se também l como $0 \leq l \leq L$ em que L é o total de segmentos no tempo e $0 \leq k \leq K$ em que K é o total de bins da DFT.

Sendo $B_d(k)$ e $B_r(k;l)$ variáveis aleatórias gaussianas centradas em zero independentes e identicamente distribuídas, $R = M - V$ a distância entre dois segmentos consecutivos e $\tau(k)$ a taxa de decaimento definida como:

$$\tau(k) = \frac{3\ln 10}{T_{60}(k)F_s}, \tag{3.5}$$

Empregando-se o conceito de primeiras reflexões e reverberação tardia $H(k,l)$ pode ser apresentada na seguinte forma:

$$H(k;l) = \begin{cases} B_d(k), & l = 0, \\ B_r(k,l)e^{-\tau(k)lR} & l > 0. \end{cases} \tag{3.6}$$

A função $B_d(k)$ possui as informações do caminho direto e primeiras reflexões, já $B_r(k,l)$ refere-se as reflexões tardias. Com isso, podemos calcular a $E_{dr}(k)$ dada por:

$$E_{dr} = 10\log_{10}\left(\frac{1-e^{-2\tau(k)R}}{e^{-2\tau(k)R}}\frac{1}{\kappa(k)}\right), \tag{3.7}$$

sendo

$$\kappa(k) = \frac{E[B_d(k)^2]}{E[B_r(k,l)^2]}. \tag{3.8}$$

Com o valor de $\kappa(k)$ determinado, podemos encontrar a variância da região de reverberação $\sigma_{pt}^2(k,l)$ que é dada por:

$$\sigma_{pt}^2(k,l) = (1-\kappa(k))\eta\sigma_{pt}^2(k,l-1) + \kappa(k)\eta\sigma_{sr}^2(k,l-1), \tag{3.9}$$

em que $\eta = e^{-2\tau(k)R}$ e $\sigma_{sr}^2(k,l) = E[|S_r(k,l)|^2]$. A partir desse ponto calcula-se σ_t^2 referente apenas a parcela da reverberação tardia. Supondo que existam N_e amostras referentes as primeiras reflexões, σ_t^2 é dada por:

$$\sigma_t^2(k,l) = e^{2\tau(k)R(N_e-1)}\sigma_{pt}^2(k,l-Ne+1). \tag{3.10}$$

Nessa etapa já se pode fazer um tratamento estatístico desses valores de forma semelhante ao que ocorreu na determinação do T_{60}. Começamos calculando a estimativa de variância em cada banda com a fórmula:

$$\sigma_t{}^2(k) = \sum_{l=0}^{L-1}\sigma_t{}^2(k,l). \tag{3.11}$$

Em seguida uma estimativa considerando todas as sub-bandas é dada por:

$$\bar{\sigma_t}{}^2 = \sum_{k=0}^{K-1}\sigma_t{}^2(k). \tag{3.12}$$

Por fim se faz o mapeamento semelhante ao utilizado no T_{60} para que se possa obter um σ_r^2 total:

$$\hat{\sigma}_r^2 = \alpha_\sigma\bar{\sigma_t}{}^2 + \beta_\sigma, \tag{3.13}$$

na qual α_σ e β_σ são constantes determinadas durante o treinamento do algoritmo.

É importante ressaltar que, como visto, para se calcular a variância espectral, se faz necessário o uso do T_{60} e da E_{dr} entre os passos intermediários. Essa abordagem faz com que a medida fique mais sensível à propagação de erros, mas ainda é a que produz melhores resultados atualmente comparada aos outros estimadores da mesma classe e por isso foi escolhido para compor o sistema QAreverb cego no trabalho de [2].

3.3.3 Energia direta sobre reverberante sem referência

Para o cálculo da E_{dr} o procedimento adotado também foi elaborado por [2] e possui algumas semelhanças ao anterior (determinação do T_{60}).

Inicialmente faz-se uma busca por FDRs no sinal reverberante no domínio do tempo e logo após, outra procura no domínio da frequência utilizando-se os mesmos tamanhos de janela M, sobreposição V e número de segmentos L.

O processo consiste em encontrar t_{hr} segmentos consecutivos com energia decrescente. Supondo uma frequência de amostragem F_s o limiar t_{hr} tem inicialmente o valor de $t_{hr} = \frac{0{,}5F_s}{M}$. Caso não seja encontrada nenhuma FDR, t_{hr} é decrementado e faz-se uma nova busca, restringindo-se t_{hr} a ser no mínimo 3.

Supondo que foram encontradas R_1 FDRs no domínio do tempo, já se pode calcular as $\hat{E}_{dr}(r,k)$ parciais com a r-ésima FDR temporal e a k-ésima FDR espectral através da equação (2.6) que será repetida aqui para maior comodidade do leitor:

$$E_{dr} = \frac{E_d}{E_r} = \frac{\int_{t_d-t_1}^{t_d+t_2} h^2(\tau)\,\mathrm{d}\tau}{\int_{t_d+t_2}^{\infty} h^2(\tau)\,\mathrm{d}\tau}. \tag{3.14}$$

Outro conjunto de FDRs é procurado no espectro com uma abordagem semelhante à feita na busca destas regiões para o T_{60}. Assim, são geradas $R_2(k)$ FDRs para o k-ésimo bin da DFT e mais estimativas.

Com as novas FDRs têm-se disponíveis $R_1 + R_2(k)$ estimativas para cada bin, que são combinadas através da seguinte fórmula:

$$\hat{E}_{dr}(k) = \frac{\sum_{r=1}^{R1+R_2(k)} \hat{E}_{dr}(r,k)}{R1 + R_2(k)} \tag{3.15}$$

Em seguida, para encontrar a estimativa parcial $\bar{E}_{dr}$ faz-se a média das $\hat{E}_{dr}(k)$, para $\frac{k}{8} + 1 \leq k \leq \frac{3k}{8}$ (o que é equivalente a utilizar somente os bins da DFT relativos ao intervalo contínuo entre 500 Hz e 1500 Hz.)

Por último, através de um mapeamento obtém-se a $\tilde{E}_{dr}$ total dada por:

$$\tilde{E}_{dr} = \alpha_p \bar{E}_{dr} + \beta_p. \tag{3.16}$$

em que α_p e β_p são constantes calculadas a partir da base de treinamento.

3.4 Conclusão

Neste capítulo foi mostrado o que é o sistema QAreverb e também a métrica Q_{mos} baseada nos valores de T_{60}, σ_r^2 e E_{dr}, que surge como uma alternativa aos avaliadores de qualidade mais comuns.

Além disso foi vista uma versão sem referência conhecida como QAreverb cego, que como o próprio nome sugere utiliza apenas o sinal reverberante para o cálculo de Q_{mos} e consequentemente utiliza técnicas diferentes das utilizadas para obter o Q_{mos} tradicional quando se buscam os parâmetros T_{60}, σ_r^2 e E_{dr}.

Já em posse dessas informações, podemos prosseguir para o processo de desreverberação propriamente dito. A métrica Q_{mos} será utilizada após a técnica de desreverberação como um medidor da qualidade da voz, ou seja, um medidor da eficiência do método aplicado.

Capítulo 4

Desreverberação

4.1 Introdução

Neste capítulo será abordado o processo de desreverberação do sinal de voz, ou seja, o processo que tem por objetivo fazer uma compensação do efeito da reverberação no sinal causado pelo ambiente.

A técnica utilizada para realizar essa tarefa foi proposta em [2]. Nessa estratégia utiliza-se um método conhecido como algoritmo de desreverberação baseado em subtração espectral que será mais detalhado a seguir.

4.2 Algoritmo de desreverberação - subtração espectral

O algoritmo de subtração espectral tem por finalidade reduzir o efeito da reverberação tardia no sinal discreto de entrada que aqui será representado por:

$$s_r(n) = \sum_{l=0}^{N} h(l)s(n-l) \ , \tag{4.1}$$

para $0 \leq n \leq N-1$, em que N é número de amostras. O sinal reverberante é representado por $z(n)$, já $s(n)$ é o sinal original e por fim $h(n)$ é a resposta ao impulso da sala. Nesse algoritmo considera-se como entrada o sinal gerado por um único microfone.

A figura 4.1 apresenta os blocos que compõem o algoritmo.

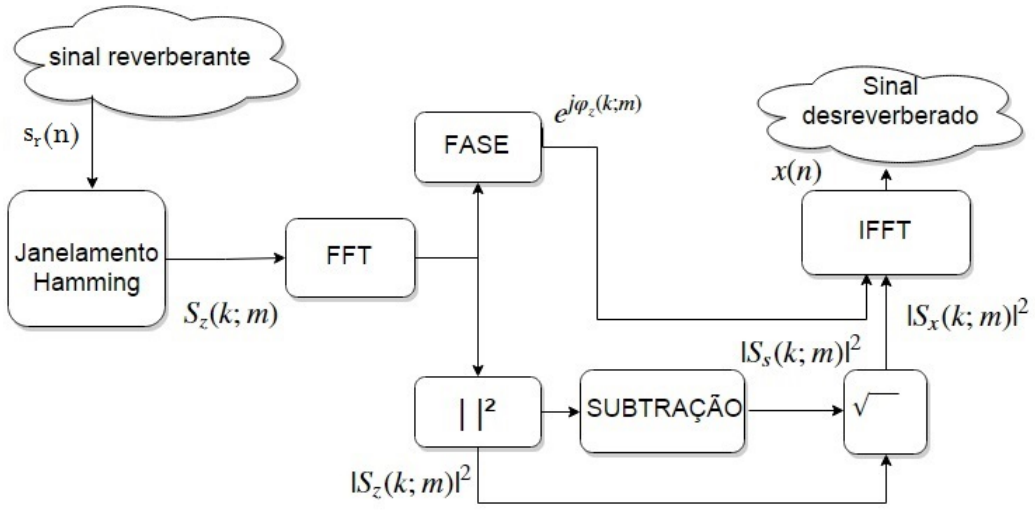

Figura 4.1: Diagrama do algoritmo de subtração espectral.

A primeira fase do processo consiste em convoluir o sinal $s_r(n)$ e uma janela de Hamming assimétrica com duração de 32 ms e 24 ms de sobreposição.

No segundo passo é feita a FFT do sinal de entrada, gerando $S_z(k;m)$ de cada uma das m janelas. Em seguida, já se pode separar o sinal em suas componentes de módulo $|S_z(k;m)|$ e fase $e^{j\varphi_z(k;m)}$.

No estágio de subtração espectral apenas o módulo do sinal é necessário. Nesse bloco são utilizados quatro parâmetros que servem para ajustar o algoritmo a uma determinada base de sinais, são eles: ϵ, a, ζ e ρ.

- Parâmetro a

  Esta variável é responsável pelo tamanho da janela de atenuação que será usada no bloco de subtração. A função que descreve essa janela segue a distribuição de Rayleigh e é dada por:

$$w(m) = \begin{cases} (\frac{m+a}{a^2})e^{-\frac{(m+a)^2}{2a^2}}, & m > -a \\ 0, & m \leq -a \end{cases} \tag{4.2}$$

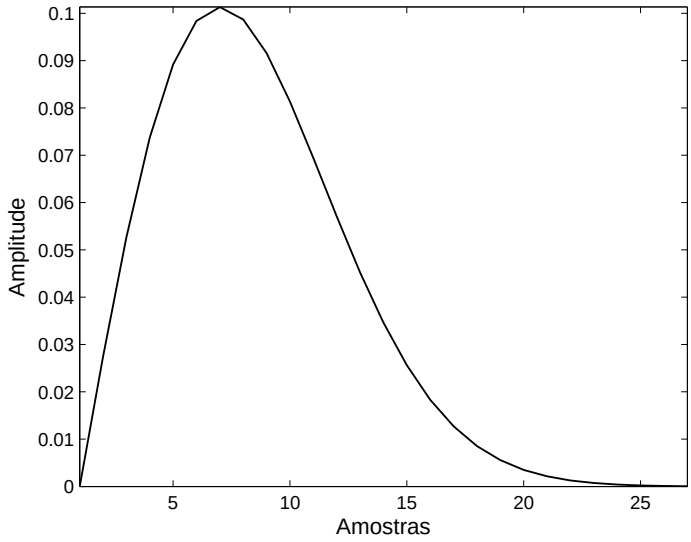

Figura 4.2: Janela de Rayleigh.

A figura 4.2 mostra o formato da janela de Rayleigh para $a = 6$. Podemos notar que a função tem um aspecto que varia lentamente, isso permite que o sinal seja janelado sem cortes abruptos, evitando danos no sinal processado.

- Parâmetros ζ e ρ

Para se obter o sinal desreverberado devemos primeiro calcular a potência espectral das reflexões tardias. Sabendo que o sinal pode ser dividido em primeiras reflexões e reflexões tardias, e que estas são descorrelacionadas entre si, a fórmula que descreve o processo é dada por:

$$|S_l(k;m)|^2 = \sum_{m=-\infty}^{\infty} \zeta w(m-\rho)|S_z(k;m)|^2, \tag{4.3}$$

onde k é o índice do bin de frequência, m é o índice do bloco no tempo, $w(m)$ é a janela de Rayleigh já mencionada com ρ deslocamentos no tempo e ζ é um fator de escala.

O parâmetro ζ pode ser entendido como uma variável que define a influência das componentes tardias. Por outro lado, ρ é o número de blocos que contém as primeiras reflexões. O parâmetro ρ pode ser considerado como um atraso

na janela utilizada para segmentar o sinal. ρ possui uma relação direta com o parâmetro a, em que $a < \rho$. Essa regra é utilizada para que haja uma correspondência razoável com o formato da resposta ao impulso gerada em relação ao modelo esperado.

- Parâmetro ϵ

  Para se obter o percentual de potência correspondente à potência espectral das primeiras reflexões devemos remover a parcela referente à reverberação tardia através da fórmula normalizada:

$$|S_s(k;m)|^2 = \max\left[1 - \frac{|S_l(k;m)|^2}{|S_z(k;m)|^2}, \epsilon\right], \tag{4.4}$$

  em que o parâmetro ϵ é um limite de atenuação, ou seja, um limiar para que S_s nunca fique nulo.

No penúltimo bloco do diagrama da figura 4.1 calcula-se a potência referente à parte desreverberada utilizando o peso encontrado acima. Para isso, basta aplicar a seguinte fórmula:

$$|S_x(k;m)|^2 = \sqrt{|S_s(k;m)|^2 \cdot |S_z(k;m)|^2}, \tag{4.5}$$

para então incluir a fase $\varphi_z(k;m)$ do sinal de entrada e finalmente usar a IFFT com o intuito de obter o sinal já desreverberado representado por $x(n)$. A figura 4.3 ilustra melhor o resultado do algoritmo.

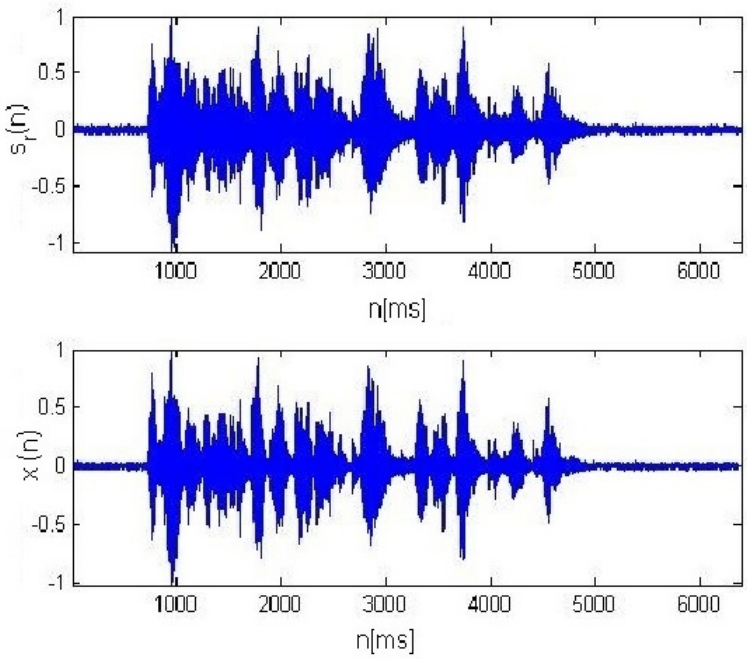

Figura 4.3: Exemplo de sinal antes do processo de desreverberação com curvas mais suaves e depois com curvas mais profundas.

Neste exemplo o sinal utilizado foi gravado em uma sala com $T_{60} = 0,7$ e distância entre fonte e microfone de 200 cm. Como a figura 4.3 sugere, o gráfico do sinal reverberado possui as reflexões tardias sobrepostas ao sinal desejado, por isso o envelope do sinal tem uma variação menor. Já o sinal desreverberado possui uma envoltória mais bem definidas se comparado ao sinal reverberado. Uma consequência imediata desta característica é que se pode distinguir com mais facilidade cada fonema e por isso a inteligibilidade geral do sinal fica melhorada.

4.3 Treinamento do algoritmo

Como mencionado anteriormente o algoritmo de desreverberação possui quatro parâmetros ajustáveis. Os sinais utilizados nesse trabalho para o treinamento do algoritmo, isto é, escolha dos valores desses parâmetros, provém de uma base desenvolvida no trabalho de [2] conhecida como base New Brazilian Portuguese (NBP).

Essa base é formada originalmente por 4 sinais sem reverberação, dos quais 2 são com voz masculina e os outros 2 com voz feminina. Para gerar outros 200 sinais

reverberados os sinais anecóicos foram expostos a 3 diferentes tipos de reverberação: artificial, natural e real que serão mais bem explicadas a seguir.

- Reverberação artificial

  Nesse conjunto são gerados 24 sinais reverberados a partir da convolução dos 4 sinais anecóicos com as 6 diferentes RIRs geradas artificialmente. As funções de resposta ao impulso são oriundas de uma modelagem virtual de salas com dimensões físicas fixas e distância entre locutor e microfone de 180 cm. A única diferença entre os ambientes nesse simulador era o valor escolhido para T_{60} que variavam entre 200 e 700 ms. O tempo médio de reverberação em cada uma das funções de resposta ao impulso foi de: 196, 292, 387, 469, 574 e 664 ms.

- reverberação natural

  Essa abordagem utiliza 17 RIRs obtidas de 4 salas reais. Os ambientes possuem diferentes tamanhos e distâncias entre locutor-microfone, que variam entre 50 e 1020 cm. Os sinais reverberados são gerados a partir da convolução dos 4 sinais anecóicos com cada uma das RIRs em questão. São gerados nesse grupo 68 sinais reverberados e o tempo médio de reverberação em cada sala é de: 120, 230, 430 e 780 ms.

- reverberação real

  Nesse caso os 108 sinais reverberados são gravados diretamente no ambiente, sem o auxílio da técnica de convolução entre os sinais anecóicos e as RIRs. A técnica consiste em emitir o sinal de voz através de uma caixa de som e gravá-lo com um microfone.

  Foram utilizadas 7 salas com diferentes tamanhos e pelo menos 3 diferentes distâncias entre fonte - microfone, que variam de 50 a 400 cm, resultando em 27 RIRs. O tempo médio de reverberação de cada sala é de: 140, 390, 570, 650, 700, 890, 920 ms.

Para o treinamento do algoritmo foram selecionados 18 sinais, um para cada ambiente (1 anecóico, 6 RIRs artificias, 4 salas naturais, 7 salas reais). Após o treinamento os valores definidos para os parâmetros foram:

- $\zeta = 0,35$,
- $\rho = 7$,
- $\epsilon = 0,001$,
- $a = 6$.

Esses números aumentam o valor médio da métrica sem o treinamento $Q_{mos} = 3,46$ para $Q_{mos} = 3,78$ após o processo.

4.4 Conclusão

Este capítulo apresentou um método para reduzir o efeito da reverberação em sinais de voz através do algoritmo de subtração espectral para 1 canal. Foi explicado em detalhes cada um dos seus blocos passando desde o janelamento no tempo, conversão para o domínio da frequência, até o bloco de subtração em si e por fim a conversão para o domínio do tempo novamente.

É importante ressaltar que o valor dos parâmetros ϵ, a, ζ e ρ foram definidos através de busca exaustiva, utilizando-se a base de treinamento deste projeto para então serem efetivamente testados em outra base. O teste foi realizado num contexto de um evento internacional com diversas equipes competindo entre si pelo melhor resultado de desreverberação e será mais detalhado no próximo capitulo.

Capítulo 5

REVERB Challenge

5.1 Introdução

O REVERB (REverberant Voice Enhancement and Recognition Benchmark) Challenge é um desafio promovido por pesquisadores de diferentes organizações destacando-se: NTT, International Audio Labs Erlangen, Paderborn University, Beuth University of Applied Sciences - Berlin, University of Erlangen-Nuremberg, Bar-Ilan University e Mellon University.

A proposta do programa é convidar cientistas de diversos países para testar seus próprios algoritmos de desreverberação e/ou reconhecimento de voz em sinais de áudio e concluir o quão eficiente foi o processo através de algumas métricas.

Na etapa de desreverberação era necessário considerar que esta técnica poderia ser utilizada em diversas situações desde aprimoramento de aparelhos auditivos a reconhecimento automático de fala. Além disso, as métricas propostas pelo desafio abrangiam tanto a avaliação objetiva quanto a subjetiva. A ideia é revelar vantagens e desvantagens de diferentes abordagens. Já para o reconhecedor de voz automático pode-se escolher qualquer modelo acústico, critério de formação e estratégia de decodificação que gere o melhor resultado.

Os arquivos de áudio oferecidos pelo programa possuem diferentes características. Nesse projeto focamos nos algoritmos para sinais adquiridos com um único canal, mas é valido citar que no REVERB Challenge havia outras categorias para sinais

de multicanais com 2 ou 8 canais. Outras variações são quanto à distância entre microfone e locutor, origem e tamanho da sala que serão mais detalhadas adiante.

Após o período de avaliação dos dados, os grupos de pesquisas são orientados a escrever um artigo detalhando o processo e os resultados obtidos, além de uma apresentação durante a conferência propriamente dita. O artigo produzido pela minha equipe pode ser verificado em [3].

5.2 Base de dados

A base oferecida pelo grupo REVERB Challenge pode ser dividida em duas sub-bases: desenvolvimento e avaliação. O desafio sugere a utilização dos sinais da base de desenvolvimento para treino e otimização dos parâmetros do algoritmo. Já os sinais da base de avaliação deveriam ser desreverberados propriamente e medidos de acordo com as métricas propostas.

Entretanto, nossa equipe preferiu utilizar ambas as bases para a realização de testes e a base NBP para a realização do treinamento em si, como mencionado na seção 4.3.

Cada uma dessas sub-bases fornecidas pelo desafio são compostas por sinais que classificam-se em:

- Simulados - quando são obtidos através da convolução do sinal anecóico (sem reverberação) com a resposta ao impulso (RIR) do ambiente em estudo.
- Reais - quando são obtidos diretamente de um microfone de uma sala com ruído e reverberação.

O dispositivo utilizado para medir a resposta ao impulso foi um microfone de 8 canais e 20 cm que é exibido a seguir na Figura 5.1. Também foi adicionado à RIR um ruído de fundo previamente gravado, basicamente composto pelo sistema de refrigeração, com uma razão sinal - ruído (SNR) fixa de 20 dB.

Esse mesmo aparato foi utilizado para gravar os sinais Reais, que já continham um ruído ambiente estacionário.

Figura 5.1: Microfones utilizados para medir as RIRs no contexto do REVERB Challenge. Fonte [17].

O número de sinais de cada uma das bases são:

- 1484 sinais Simulados da base de desenvolvimento
- 179 sinais Reais da base de desenvolvimento
- 2176 sinais Simulados da base de avaliação
- 372 sinais Reais da base de avaliação

Outra possível classificação dos sinais se deve ao tamanho da sala onde foi adquirido o sinal de voz. As salas para os sinais Simulados podem variar entre: Pequena - Sala 1, Média - Sala 2 e Grande - Sala 3, com T_{60} de 0,25 s, 0,5 s, 0,7 s respectivamente. Isso nos permite avaliar a capacidade do algoritmo e da métrica em atuar em diferentes ambientes de reverberação.

Entretanto, para os sinais Reais somente um tipo de sala foi utilizada, correspondente à um T_{60} de 0,7 s. Nesse caso estamos interessados em observar a robustez das ferramentas de avaliação em situações que não podem ser reproduzidas com facilidade artificialmente.

Além destas categorias já citadas, mais uma divisão pode ser feita quanto a distância entre o microfone e o locutor. A distância pode ser dita como Perto (50 cm - sinais Simulados e 100 cm - sinais Reais) ou Longe (200 cm - sinais Simulados e 250 cm - sinais Reais.)

As tabelas 5.1 e 5.2 mostram respectivamente a quantidade de sinais em cada classe para as duas sub-bases: desenvolvimento e avaliação.

Tabela 5.1: Tabela com a distribuição dos sinais para base de desenvolvimento.

Desenvolvimento							
Simulado						Real	
Sala 1		Sala 2		Sala 3		Sala 1	
Perto	Longe	Perto	Longe	Perto	Longe	Perto	Longe
248	248	247	247	247	247	89	90

Tabela 5.2: Tabela com a distribuição dos sinais para base de avaliação.

Avaliação							
Simulado						Real	
Sala 1		Sala 2		Sala 3		Sala 1	
Perto	Longe	Perto	Longe	Perto	Longe	Perto	Longe
363	363	363	363	362	362	186	186

5.3 Algoritmo

O algoritmo para desreverberação aplicado inicialmente durante o desafio é o mesmo que foi mencionado no Capítulo 4.

A abordagem utilizada para o processamento dos sinais através do algoritmo poderia ser feita de três diferentes formas: lote completo de testes, lote dividido de testes ou ainda tempo real.

- Lote completo de testes - sugere que os sinais com características semelhantes em relação à origem na sala e/ ou distância locutor - microfone podem ser processados juntos. Esse método permite otimizar os parâmetros do algoritmo de acordo com as particularidades de cada grupo de sinais.

- Lote dividido de testes - esquema em que os sinais são analisados individualmente, independente de suas características comuns.

- Tempo real - método que utiliza trechos próximos do bloco atual em análise para processar de melhor forma, além disso alguns atrasos pré-fixados pelos participantes também podem ser empregados. O processamento também é feito individualmente para cada sinal nessa abordagem.

A minha equipe optou pelo método de processamento por lote completo.

Antes de utilizar o algoritmo propriamente para desreverberar os sinais foi feita uma otimização de parâmetros em que algumas configurações foram testadas. A ideia é variar os valores de ϵ, a, γ e ρ para obter resultados específicos e buscar os que melhor atendem ao objetivo do REVERB Challlenge. Os números encontrados e usados neste desafio foram os mesmos citados na seção 4.3 que são :

- $\epsilon = 0{,}001$;
- a = 6;
- $\zeta = 0{,}35$;
- $\rho = 7$.

5.4 Métricas

As métricas utilizadas para avaliar a qualidade do sinal de voz sugeridas pelo REVERB Challenge foram: Distância Cepstral (CD), Razão do log da verossimilhança (LLR), SNR ponderadas em frequência (FWSS), Relação de energia de modulação de voz para reverberação (SRMR), Razão de palavras erradas (WER), ATime e RTime. Adicionamente a estas, avaliamos o algoritmo também através da métrica Q_{mos}. Todas essas métricas serão descritas em mais detalhes a seguir.

- Q_{mos}: mede a qualidade do sinal através dos parâmetros T_{60}, variância espectral e energia direta sobre reverberante. Sua escala varia de 1 (muito reverberante) a 5 (idealmente sem reverberação), por isso quanto maior o valor de Q_{mos}, melhor o sinal de voz. Esta métrica pode ser aplicada tanto nos sinais reais quanto nos simulados.

- Distância Cepstral (CD, do inglês *Cepstral Distance*): mede a distância entre os cepestros dos sinais degradados e limpos. O cálculo é feito usando a raiz da média quadrática da diferença dos dois cepestros. Só pode ser avaliada nos dados simulados já que precisa do sinal sem reverberação. Quanto menor o valor da distância cepstral de um sinal, melhor.

- Razão do log da verossimilhança (LLR, do inglês *Log-Likelihood Ratio*): mede o grau da discrepância entre os espectros do sinal degradado e do sinal de referência. O valor é obtido utilizando-se Coeficientes de Predição Linear (LPC, do inglês *Linear Prediction Coefficients*). Só pode ser medida nos dados simulados já que precisa do sinal limpo. Quanto menor o valor de LLR, melhor o sinal.

- SNR ponderadas em frequência (FWSS, do inglês *Frequency-Weighted Segmental SNR*): mede a relação entre a potência do sinal de voz e do ruído no domínio da frequência. Quanto maior o valor da $FWSS$, melhor o sinal de voz.

- Relação de energia de modulação de voz para reverberação (SRMR, do inglês *Speech-to-Reverberation Modulation energy Ratio*): supostamente mede a qualidade de percepção de um sinal de fala degradado por ruído e reverberação. Pode ser usado tanto para os dados simulados quanto para os dados reais. Quanto maior o valor da $SRMR$, melhor.

- Custo computacional: mede em segundos quanto tempo (ATime) o algoritmo levou para processar um determinado conjunto de dados. Como esta medida é fortemente dependente da plataforma de configuração, o custo computacional (RTime) do código de referência dado também é calculado para cada conjunto de dados neste trabalho. Pela própria definição desta métrica, não exite medição de ATime para os sinais não processados, já que não foram utilizados em nenhum algoritmo. O algoritmo foi rodado em MATLAB Versão 7.12.0.635 (R2011a) de 64 bits em um ambiente de computação com sistema operacional Windows 7 de 64 bits, processador AMD Visão dupla E-350 1.60 GHz Core e 4 GB de RAM. Sendo assim, observa-se que para as métricas ATime e RTime, quanto menor o valor, melhor.

- Razão de palavras erradas (WER, do inglês *Word Error Rate*): métrica comum para medir desempenho de sistemas de reconhecimento de voz. O valor de WER é medido após o conjunto de dados ser processado pelo algoritmo de desreverberação e o algoritmo de referência para reconhecimento automático de fala dado pelo REVERB Challenge. No caso da WER quanto menor a

nota, melhor a qualidade do sinal avaliado. O algoritmo de reconhecimento de voz automático foi usado em um Linux ubuntu 12.04 máquina virtual em um MAC OS X 10.864-bits, com um processador de 2,3 GHz e i7 intel quadcore 8 GB de RAM.

Os valores obtidos por cada uma das métrica citadas foram testados tanto na base de desenvolvimento quanto na base de avaliação. Esses resultados serão mostrados em mais detalhes no próximo capítulo.

Capítulo 6

Resultados

6.1 Introdução

Nesse capítulo serão apresentados os números encontrados ao testar as bases do REVERB Challenge com as métricas: Distância Cepstral (CD), Razão de log-verossimilhança (LLR), SNR ponderadas em frequência (FWSS), Relação de energia de modulação de voz para reverberação (SRMR), Q_{mos}, Custo computacional (ATime e Rtime) e razão de palavras erradas (WER).

6.2 Valores obtidos

Os resultados apresentados a seguir mostram que pela média todas as bases têm uma melhora nas métricas FWSS, SRMR, Q_{mos} e WER, quando comparamos os sinais antes (originais) e após (processados) o tratamento com o algoritmo.

Alguns resultados parciais e inclusive as médias de CD e LLR podem dar a falsa impressão de que o tratamento reduziu a qualidade do áudio processado em comparação ao áudio original. Uma justificativa para essas variações é que elas estão dentro da margem de erro esperado.

Uma exceção do caso citado acima é a WER para a sala 1 dos arquivos simulados. Nesta situação houve uma redução da qualidade (aumento na taxa de erro) pois antes mesmo do processamento esses áudios já possuíam uma nota Q_{mos} alta que os classificariam como 'bom', dispensando assim o tratamento da reverberação.

Tabela 6.1: Resultados utilizando sinais simulados originais da base de desenvolvimento.

Métrica	Sala 1		Sala 2		Sala 3		Média
-	Perto	Longe	Perto	Longe	Perto	Longe	-
CD	1,96	2,65	4,58	5,08	4,2	4,82	3,88
LLR	0,34	0,38	0,51	0,77	0,65	0,85	0,58
FWSS	8,1	6,75	3,07	0,53	2,32	0,14	3,49
SRMR	4,37	4,63	3,67	2,94	3,66	2,76	3,67
QMOS	4,23	3,87	3,35	1,52	3,27	2,35	3,10
WER (%)	15,3	25,3	43,9	85,8	52,0	88,9	51,8

Tabela 6.2: Resultados utilizando sinais simulados processados da base de desenvolvimento.

Métrica	Sala 1		Sala 2		Sala 3		Média
-	Perto	Longe	Perto	Longe	Perto	Longe	-
CD	3,46	3,46	4,64	4,78	4,27	4,44	4,17
LLR	0,51	0,52	0,51	0,69	0,64	0,77	0,61
FWSS	8,07	7,56	5,39	2,55	4,19	1,96	4,96
SRMR	5,06	5,68	4,71	4,32	4,74	4,13	4,77
QMOS	4,21	3,96	3,81	2,42	3,69	2,85	3,49
WER (%)	36,5	46,0	34,6	63,2	45,3	64,5	48,3
ATime	1167	1200	1185	1667	1067	1206	1249
RTime	181	164	189	199	181	192	184

Tabela 6.3: Resultados utilizando sinais reais da base de desenvolvimento.

Métrica	Originais			Processados		
-	Perto	Longe	Média	Perto	Longe	Média
SRMR	4,06	3,52	3,79	6,51	5,74	6,13
QMOS	2,45	2,41	2,43	3,72	3,64	3,68
WER (%)	88,7	88,3	88,5	69,0	62,9	66,0
ATime	-	-	-	340	329	335
RTime	-	-	-	56	53	55

A respeito da base de desenvolvimento simulada contida nas tabelas 6.1 e 6.2, as métricas objetivas CD, LLR e FWSS apresentaram um aumento de 7%, 5%, 42%. Para estes mesmos sinais, as métricas de percepção SRMR, Q_{mos} e WER obtiveram um acréscimo de 30%, 13% e 3,5%.

Já para os áudios reais contido na tabela 6.3, as métricas SRMR, Q_{mos} aumentaram em 62% e 51% nessa ordem, mas a métrica WER reduziu 22,5% para os mesmos sinais.

Tabela 6.4: Resultados utilizando sinais simulados originais da base de avaliação.

Métrica	Sala 1		Sala 2		Sala 3		Média
-	Perto	Longe	Perto	Longe	Perto	Longe	-
CD	1,99	2,67	4,63	5,21	4,38	4,96	3,97
LLR	0,35	0,38	0,49	0,75	0,65	0,84	0,58
FWSS	8,12	6,68	3,35	1,04	2,27	0,24	3,62
SRMR	4,5	4,58	3,74	2,97	3,57	2,73	3,68
QMOS	4,24	3,96	3,61	2,37	3,2	2,4	3,30
WER (%)	18,1	25,4	43,0	82,2	53,5	88,0	51,7

Tabela 6.5: Resultados utilizando sinais simulados processados da base de avaliação.

Métrica	Sala 1		Sala 2		Sala 3		Média
-	Perto	Longe	Perto	Longe	Perto	Longe	-
CD	3,49	3,53	4,62	4,86	4,29	4,55	4,22
LLR	0,53	0,53	0,48	0,65	0,62	0,74	0,59
FWSS	7,97	7,65	5,85	3,14	4,3	2,03	5,16
SRMR	5,21	5,55	4,9	4,35	4,8	4,1	4,82
QMOS	4,22	4,02	3,99	2,87	3,73	3,88	3,79
WER (%)	47,5	52,5	38,4	57,1	43,4	66,2	50,8
ATime	1661	2028	1754	1834	1760	1709	1791
RTime	331	247	290	328	278	307	297

Tabela 6.6: Resultados utilizando sinais reais da base de avaliação.

Métrica	Originais			Processados		
-	Perto	Longe	Média	Perto	Longe	Média
SRMR	3,17	3,19	3,18	5,08	5,12	5,10
QMOS	2,51	2,57	2,54	3,79	3,8	3,80
WER (%)	89,7	87,3	88,5	76,3	71,5	73,9
ATime	-	-	-	736	622	679
RTime	-	-	-	138	126	132

Em relação à base de teste simulada contida nas tabelas 6.4 e 6.5 as métricas de CD, LLR , FWSS , SRMR , QMOS e WER mostraram um crescimento de 6%, 2%, 43%, 31%, 15%, 0,9%. Já para os sinais reais contidos na tabela 6.6 SRMR e Q_{mos} aumentaram em 60% e 50% e WER diminuiu em 14,6%.

Esses valores nos mostram que em geral o sinal é aperfeiçoado, principalmente quando se refere ao caso dos sinais Reais. Esse aspecto é exatamente o desejado já que na prática não temos os sinais que desejamos melhorar são os utilizados sinais provenientes das condições reais.

É valido ressaltar que o algoritmo QAreverb utilizado para o cálculo da métrica Q_{mos} para os sinais com referência é o algoritmo tradicional. Já nos sinais sem referência é utilizado o algoritmo QAreverb cego que calcula o Q_{mos} adaptado.

6.3 Outros algoritmos

Como já mencionado anteriormente o REVERB Challenge é um desafio internacional que contou com a participação de equipes de diversos países.

Nessa seção serão mostrados gráficos que comparam os resultados de alguns algoritmos para uma dada métrica.

Aqui neste trabalho será exibido um gráfico para cada métrica. Nessas figuras estarão representados os algoritmos que fizeram uso das mesmas ferramentas e dados que nós, no caso: 1 canal e processamento por lote completo de testes. Esse perfil será doravante denominado configuração restrita.

6.3.1 CD

Analisando o gráfico da Figura 6.1 é possível observar que para a distância cepstral (CD) o nosso algoritmo (linha marrom) possui uma performance melhor em ambientes com grandes dimensões; dado que quanto menor o valor de CD, melhor para o sinal. Como mostra a imagem, CD tem um valor menor no sinal processado nas salas 2 e 3 do que na sala 1 quando comparado ao sinal original.

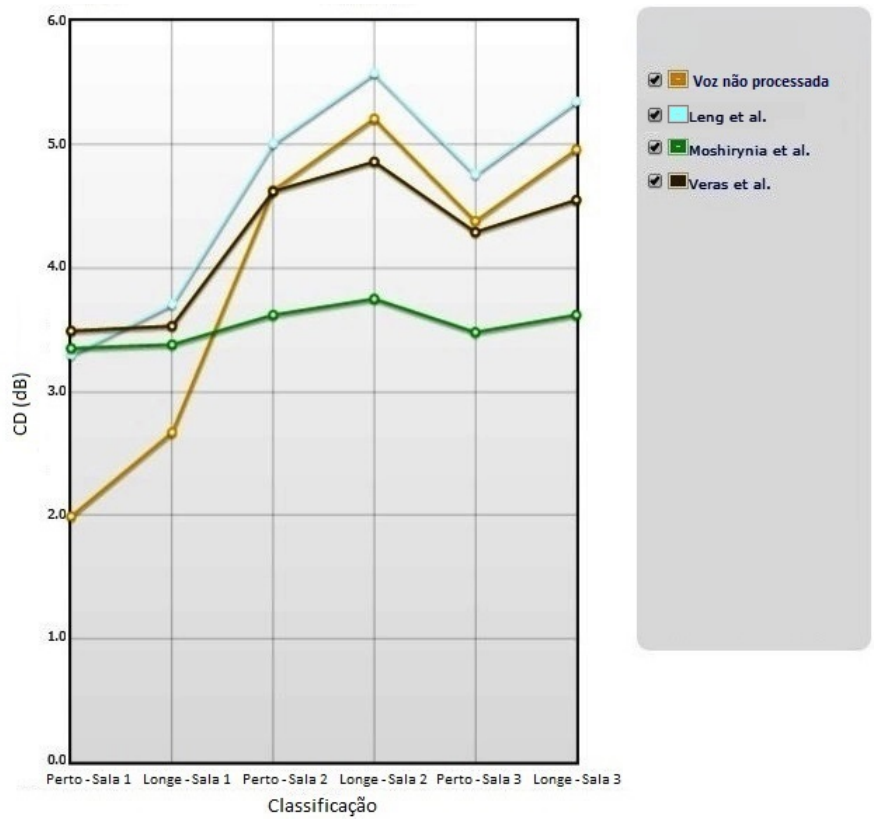

Figura 6.1: Métrica CD obtida através de algoritmos que utilizam configurações restritas. Fonte [18].

6.3.2 LLR

O log da razão de verossimilhança (LLR) similarmente a CD também possui um melhor desempenho nos ambientes grandes como pode ser conferido na Figura 6.2 ; dado que quanto menor o valor de LLR, melhor para o sinal.

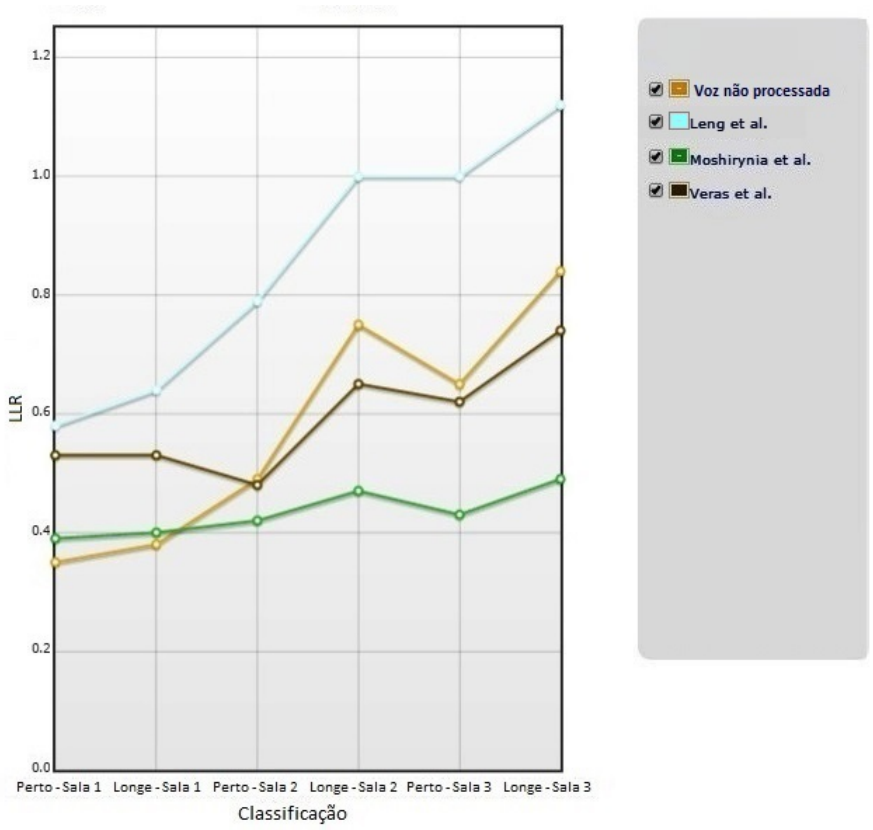

Figura 6.2: Métrica LLR obtida através de algoritmos que utilizam configurações restritas. Fonte [18].

6.3.3 FWSS

Para SNR ponderadas em frequência o comportamento do algoritmo é razoavelmente bom em todos os ambientes independente da dimensão, com um ganho praticamente constante como pode ser verificado na Figura 6.3. Dado que quanto maior o valor de $FWSS$, melhor para o sinal.

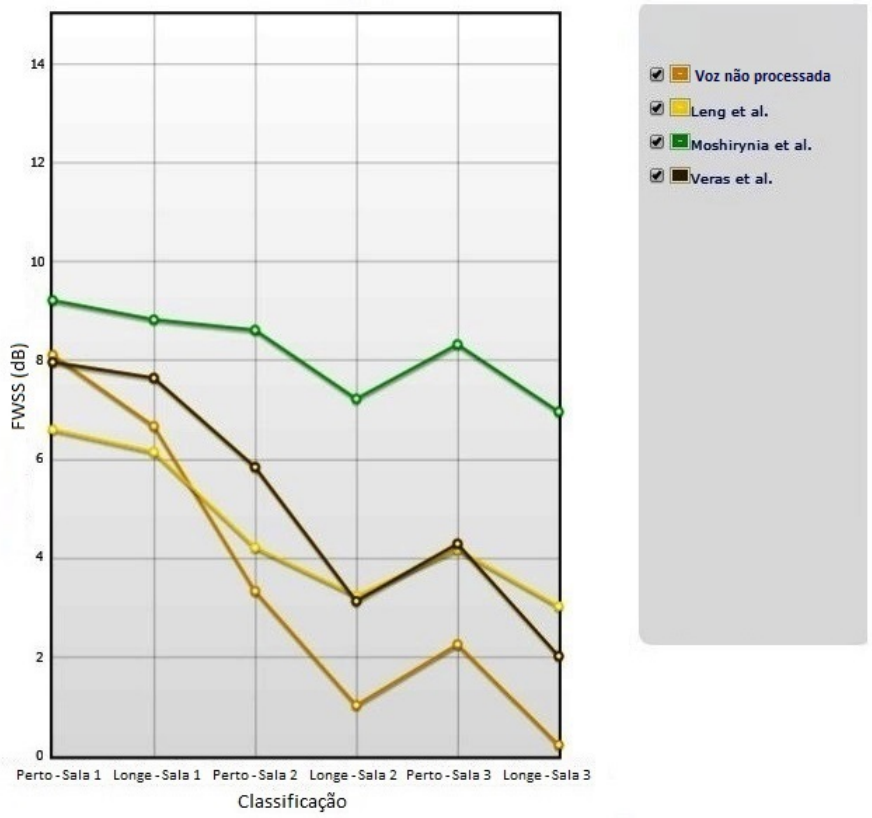

Figura 6.3: Métrica FWSS obtida através de algoritmos que utilizam configurações restritas. Fonte [18].

6.3.4 SRMR

No caso da Relação de energia de modulação de voz para reverberação (SRMR) o comportamento do algoritmo é razoavelmente bom nas 3 salas. Dado que quanto maior o valor de $SRMR$, melhor para o sinal.

É possível observar na Figura 6.4 que o ganho cresce conforme a dimensão da sala aumenta, por isso os sinais após o tratamento da desreverberação da sala 3 tem uma melhora mais significativa quando comparados aos sinais na sala 2, e por consequência da sala 1.

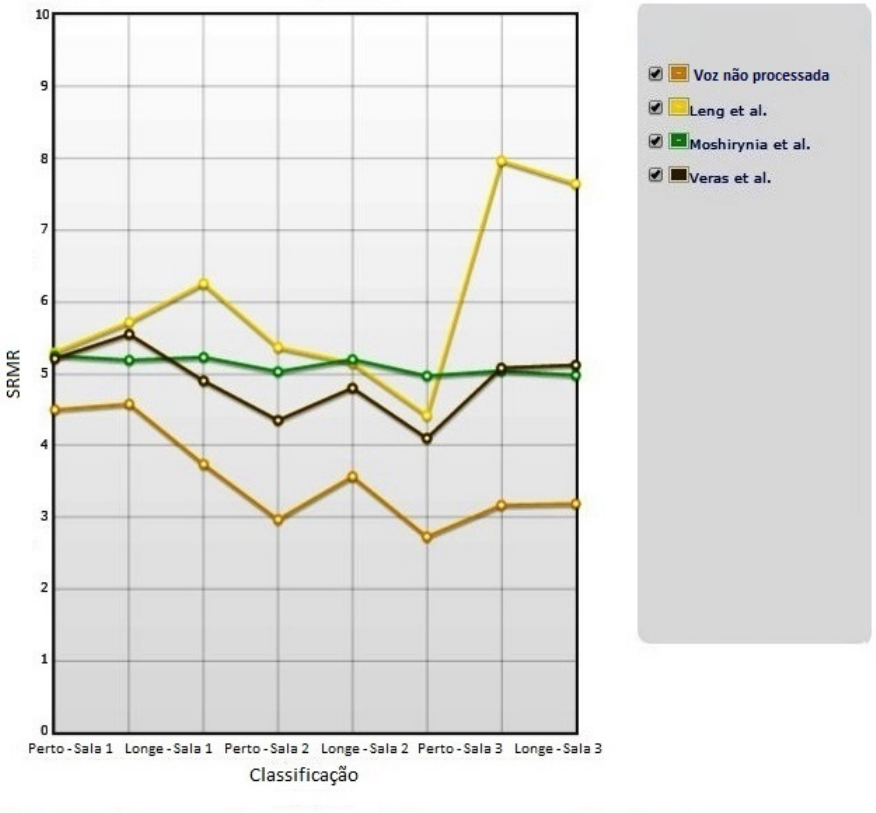

Figura 6.4: Métrica SRMR obtida através de algoritmos que utilizam configurações restritas. Fonte [18].

6.3.5 MUSHRA

Para a análise da qualidade subjetiva dos resultados gerados pelo processo de desreverberação foi utilizado pelos organizadores um teste conhecido como MUSHRA, que avalia dois aspectos: reverberação percebida e qualidade geral do áudio processado.

Esse teste é feito considerando o número de canais utilizados. Na Figura 6.5 temos os resultados para os grupos que utilizaram 1 canal. Neste gráfico é possível perceber que o algoritmo tem uma performance melhor que alguns e pior que outros.

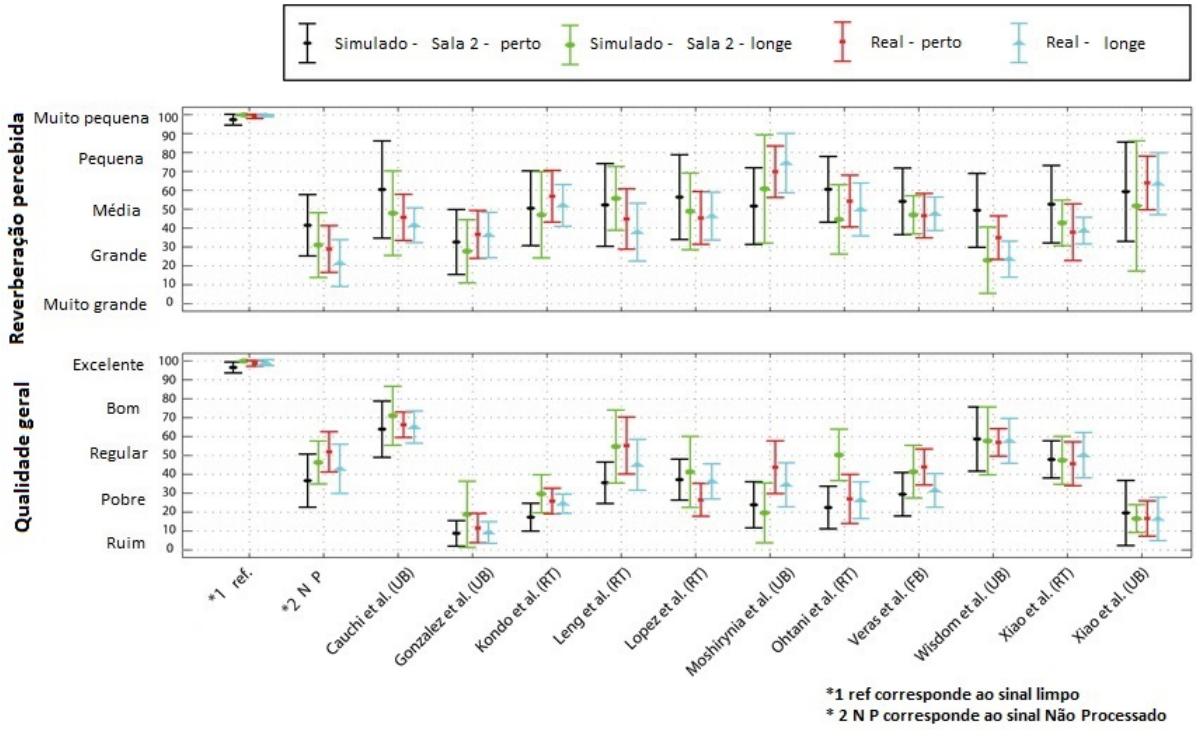

Figura 6.5: MUSHRA para avaliar as métricas de percepção. Fonte [18].

6.3.6 WER

Além das características já mencionadas adotadas pela minha equipe, para o caso do reconhecedor ainda há mais duas que podem ser usadas para diferenciar dos outros grupos.

O Modelo acústico escolhido foi o limpo, entre as opções ainda haviam Multi-condições e um próprio modelo que poderia ser desenvolvido por cada grupo. Já para o reconhecedor de voz, a equipe poderia escolher entre utilizar o próprio reconhecedor ou o modelo oferecido pelos organizadores que poderia ser com CMLLR - Constrained Maximum Likelihood Linear Regression ou sem essa ferramenta (opção escolhida pelo meu grupo).

Na Figura 6.6 verificamos que nosso algoritmo tem uma melhor performance na maior sala comparado ao outro programa que utiliza as mesmas configurações; dado que quanto menor o valor de WER, melhor para o sinal.

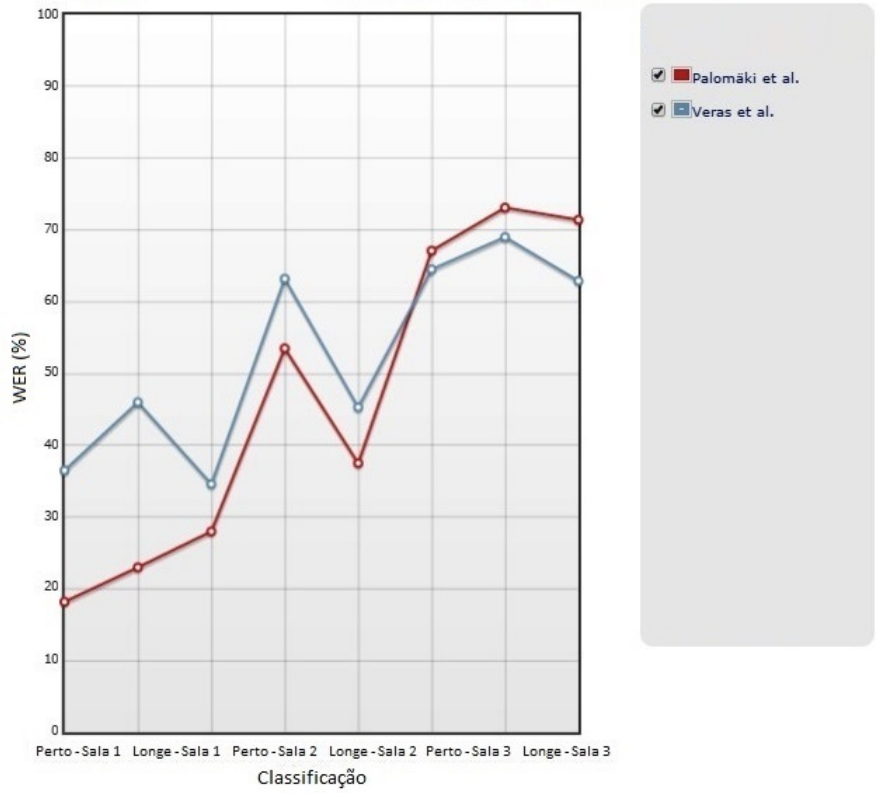

Figura 6.6: Métrica WER obtida através de algoritmos que utilizam configurações restritas. Fonte [19].

Uma possível explicação para esse comportamento deve-se a presença de alguns artefatos inseridos no sinal pelo processo de desreverberação, como por exemplo cliques. Essa reação ainda precisa ser melhor investigada para que se possa esclarecer com mais detalhes sua origem, e assim buscar métodos para combatê-la.

Capítulo 7

Conclusão

7.1 Análise do trabalho

Este estudo fez uma validação do algoritmo de desreverberação baseado em subtração espectral para um conjunto de sinais com diferentes características. As medidas foram baseadas na métrica QAreverb, SRMR, LLR, FWSS e outras mais.

O trabalho começa no Capítulo 2 fazendo uma descrição do fenômeno da reverberação, suas principais causas e principalmente o seu quase sempre indesejado efeito. O capítulo mostra também sua formulação matemática que depende essencialmente de três variáveis: tempo de reverberação (T_{60}), variância espectral da sala (σ_r^2) e razão de energia direta sobre reverberante (E_{dr}). Neste mesmo capítulo as variáveis mencionadas são explicadas em detalhes, e também é visto como calculá-las.

No Capítulo 3 foi apresentado o sistema QAreverb e seu variante QAreverb cego. Foram explicados ainda os 5 estágios do modelo: pré-processamento, desconvolução, cálculo dos parâmetros, cálculo da métrica Q e mapeamento na escala Q_{mos}. Além disso o capítulo compara as diferenças entre o modelo QAreverb e QAreverb cego no que diz respeito a forma de se obter as principais variáveis. É mostrado como conseguir o tempo de reverberação (T_{60}), a variância espectral da sala (σ_r^2) e a razão de energia direta sobre reverberante (E_{dr}) em um sistema sem sinal de referência. Nessa abordagem, a proposta é calcular os parâmetros T_{60}, σ_r^2 e E_{dr} de forma parcial, ao dividir o sinal reverberado $s_r(n)$ em vários pequenos trechos no espectro. Em seguida, é feito um tratamento estatístico com essas medidas parciais

que por fim geram o valor desejado. É importante ressaltar que este procedimento propaga erros estatísticos que devem ser considerados no valor final da medida, uma forma adotada para minimizar essa diferença e que tem se mostrado eficiente é o uso do mapeamento. A ideia é que a medida varie no mesmo intervalo dinâmico da base de referência através de um ajuste com dois coeficientes determinados durante o treinamento utilizados em uma equação de primeira ordem.

No Capítulo 4 é detalhado como se dá o processo de desreverberação feito pelo algoritmo de subtração espectral. O algoritmo tem um conceito simples, como o próprio nome indica, a ideia central do programa é subtrair do sinal a parcela correspondente à reverberação. Já que esta operação é feita no domínio da frequência, podemos entendê-lo como uma subtração espectral. O processo pode ser dividido em 6 fases: janelamento, FFT, divisão em módulo e fase, subtração, espectro da frequência e IFFT. Nesta seção também são apresentados os 4 parâmetros ajustáveis do algoritmo: ϵ que é um limiar inferior para o valor da porcentagem de reverberação no sinal, a que é o tamanho da janela usada para dividir o sinal em trechos, ζ que define a influência das componentes tardias no sinal e ρ que é o número de deslocamentos necessários para se chegar a componente tardia partindo-se do inicio do sinal. Esses parâmetros são de suma importância para um melhor desempenho do algoritmo e por isso devem ser ajustados pra cada base. No final do capítulo é feita uma breve descrição da base utilizada no treinamento desse projeto.

O Capítulo 5 apresenta a proposta do desafio internacional REVERB Challenge no qual tanto o algoritmo de subtração espectral como o reconhecedor de voz tiveram a chance de ser testados. Os sinais da base fornecida possuem diferentes características quanto ao número de canais, distância entre locutor-microfone, dimensões da sala e origem que pode ser real ou simulada. Nessa parte do trabalho são apresentadas três configurações possíveis para o processamento dos sinais que são: lote completo de testes, lote dividido de testes ou ainda tempo real. Nesse capítulo também são mostradas quais métricas os organizadores sugerem que sejam utilizadas para medir a eficiência da desreverberação. As métricas buscam no domínio do tempo ou no domínio da frequência quantificar a qualidade do processo de desreverberação, seja de forma objetiva ou perceptiva. As métricas utilizadas para avaliar estritamente a melhoria do sinal de voz foram: Q_{mos}, Distância Cepstral (CD), Razão do log da verossimilhança (LLR), SNR ponderadas em frequência

(FWSS) e Relação de energia de modulação de voz para reverberação (SRMR). Já para avaliar a performance do reconhecedor de voz foi utilizada a métrica Razão de palavras erradas (WER). E por fim o custo computacional foi medido através das métricas ATime e RTime.

O Capitulo 6 mostra os resultados obtidos nas métricas propostas pelo desafio para os sinais de acordo com as classificações entre simulados ou reais, nas bases de desenvolvimento ou avaliação. O texto também compara os valores encontrados com os resultados das outras equipes participantes do REVERB Challenge. As métricas avaliadas foram: CD, LLR, FWSS, SRMR, WER, além de um teste de avaliação subjetiva do sinal chamado MUSHRA feito diretamente pelos organizadores. Um fato interessante observado foi que para os sinais reais, o algoritmo de desreverberação surtiu um efeito melhor do que quando comparado a sinais simulados. Essa característica não deixa de ser útil, já que nas principais aplicações não há o sinal de referência.

7.2 Prosseguimento do projeto

Uma possível forma de continuar o trabalho seria buscar um novo estimador de variância $\hat{\sigma}^2$ que não fosse tão dependente do T_{60} e da E_{dr}, evitando assim a propagação de erros para essas variáveis.

Outra possibilidade é fazer um novo treinamento no algoritmo, subdividindo os sinais em grupos mais específicos como por exemplo quanto à distância locutor - microfone ou até mesmo a origem do sinal. A ideia é buscar valores para os quatro parâmetros de ajuste ϵ, a, ζ e ρ que gerem resultados ainda melhores nas métricas que estão sendo otimizadas.

Adicionalmente, ainda explorando a questão do treinamento, pode-se variar quais métricas serão escolhidas paras serem otimizadas. Algumas candidatas são as métricas Q_{mos}, $SRMR$ e $PESQ$. Na verdade, o ideal seria otimizar múltiplas medidas simultaneamente buscando não exatamente um valor ótimo para cada uma individualmente e sim um valor intermediário que produzisse resultados melhores considerando todas.

Ainda se podem testar outros algoritmos para o processo de desreverberação. Inclusive pode-se considerar os que foram apresentados durante o desafio REVERB

Challenge pelas outras equipes como por exemplo algoritmos de desreverberação baseados em programação esparsa ou predição linear.

Uma outra melhoria a ser implementada refere-se a busca de uma solução para os artefatos inseridos nos sinais durante o processo de desreverberação. Esses erros foram detectados pelo reconhecedor pois comprometem a inteligibilidade do sinal e por isso acarretaram na redução da nota WER.

Referências Bibliográficas

[1] NEELY, S. T., ALLEN, J. B., "Invertibility of a room impulse response".In:*J. Acoust. Soc. Am.*, vol. 66, no. 1 165-169, Jul 1979

[2] PREGO, T. M.*"Acerca da reverberação em sinais de voz: quantificação perceptual e aperfeiçoamento de algoritmos de desreverberação.* Rio de Janeiro : Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Tese de Doutorado, 2012.

[3] VERAS, J DO C. S., PREGO, T. DE M., LIMA, A. A. DE, FERREIRA, T. N., NETTO, S. L . *Speech quality enhancement based on spectral substraction* . Proc. Reverberation Challenge, Florence, Italy, pp. 1-5, May 2014.

[4] T. de M. Prego, A. A. de Lima and S. L. Netto. *Perceptual Improvement of a Two-Stage Algorithm for Speech Dereverberation.* Proc. InterSpeech, Lyon, France, pp. 1360-1364, Sep. 2013.

[5] MOURJOPOULOS, J., HAMMOND, J. "Modelling and enhancement of reverberant speech using an envelope convolution method". In:*Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing (ICASSP)*, pp. 1144- 1147, Boston, USA, Apr 1983.

[6] SCHROEDER, M. R. "New method of measuring reverberation time", *J.Acoust. Soc. Am.*, v. 37, n. 3, pp. 409 - 412, Mar 1965.

[7] LUNDEBY, A., VIGRAN, T. E., BIETZ, H., et al. "Uncertainties of measurements in room acoustics", *Acustica*, v. 81, n. 4, pp. 344-355, Jul 1995.

[8] ANTSALO, P., MAKIVIRTA, A., VALIMAKI, V., et al. "Estimation of modal decay parameters from noisy response measurements."In: Proc. Conv.*Audio Engineering Society*, pp. 867-878, Amsterdam, Netherlands, May 2001.

[9] KARJALAINEN, M., ANTSALO, P., MAKIVIRTA, A., et al. "Estimation of modal decay parameters from noisy response measurements", *J. Audio Eng. Soc.*, v. 50, n. 11, pp. 867-878, Nov 2002.

[10] JETZ, J. J. "Critical distance measurement of rooms from the sound energy spectral response",*J. Acoust. Soc. Am.*, v. 65, n. 5, pp. 1204-1211, May 1979.

[11] KUSTER, M. "Reliability of estimating the room volume from a single room impulse response", *J. Acoust. Soc. Am.*, v. 124, n. 2, pp. 982-993, Aug 2008.

[12] GRIESINGER, D. "The importance of the direct to reverberant ratio in the perception of distance, localization, clarity, and envelopment". In:*126th AES Convention*, Munich, Germany, Preprint 7724, May 2009.

[13] RATNAM, R., JONES, D. L., WHEELER, B. C., et al. "Blind estimation of reverberation time", *J. Acoust. Soc. Am.*, v. 114, n. 5, pp. 2877-2892, Nov 2003.

[14] RATNAM, R., JONES, D. L., W. D. O'BRIEN, J. "Fast algorithms for blind estimation of reverberation time", *IEEE Signal Processing Letters*, v. 11, n. 6, pp. 537-540, Jun 2004.

[15] VIEIRA, J. "Automatic estimation of reverberation time". In:*116th AES Convention*, Berlin, Germany, Preprint 6107, May 2004.

[16] HABETS, E. A. P., GANNOT, S., COHEN, I. "Late reverberant spectral variance estimation based on a statistical model", IEEE Signal Processing Letters, v. 16, n. 9, pp. 770-773, Sep 2009.

[17] "Documentation about the room impulse responses and noise data used for the REVERB challenge SimData". REVERB Challenge. `http://reverb2014.dereverberation.com/tools/Document_RIR_noise_recording.pdf` (Acesso em 19/02/2016)

[18] "Results for the SE task". REVERB Challenge. `http://reverb2014.dereverberation.com/result_se.html` (Acesso em 19/02/2016)

[19] "Results for the ASR task". REVERB Challenge. `http://reverb2014.dereverberation.com/result_asr.html`(Acesso em 19/02/2016)

[20] JUNIOR, J. A. A., MALVAR, H. S. "Criptoanálise de sinais de voz cifrada por permutação de segmentos temporais baseada em distâncias cepstrais", 11º Simpósio Brasileiro de Telecomunicações, Set 1993.

[21] González, D. R., Arias S. C., Lara, J. R. C. *Single channel speech enhancement based on zero phase transformation in reverberated environments* . Proc. Reverberation Challenge, Florence, Italy, pp. 1-5, May 2014.

[22] Gray Jr., Augustine H. e Markel, John D. *Distance Measures for Speech Processing* in "IEEE Transactions on Acoustic, Speech, and Signal Processing", vol. ASSP-24, n°. 5, pp 380-391, Oct 1976.