

1

Introdução

1.1

Motivação e Objetivos

A grande demanda por serviços de telecomunicações, principalmente de voz sobre IP (VoIP) [1], tem pressionado fabricantes de equipamentos e operadoras por um aumento da capacidade de transmissão de informação. Todo esse volume de dados necessita de uma largura de banda para ser trafegado que nem sempre está disponível. Em função disso, surge a necessidade de transmissão a baixas taxas, incluindo a voz, de modo a otimizar a utilização da banda disponível.

Para que essa redução na taxa de transmissão não resulte em perda de qualidade, é importante que se faça um estudo detalhado e uma avaliação criteriosa dos parâmetros do codificador de voz, visando uma obtenção mais precisa dos seus valores. Atualmente, os esquemas de codificação de voz que operam a baixas taxas utilizam, em geral, codificação linear preditiva, conhecida como LPC (Linear Predictive Coding), com base em um modelo de produção de fala. Nesse modelo, um sinal de excitação é aplicado a um filtro só de pólos (caracterizado pelos parâmetros LPC), que representa a envoltória espectral do sinal de voz [2][3]. Usualmente, os parâmetros LPC são transformados para LSF (Line Spectrum Frequencies), devido às propriedades atraentes destes últimos para processos de quantização e interpolação [4]. Devido à elevada carga computacional gerada no processamento desses parâmetros, e às baixas taxas de bits requeridas em diversas aplicações, os codificadores de voz transmitem LSF's em períodos de tempo mais longos (quadros tipicamente de 20 ms de voz) do que o desejável para se obter uma boa qualidade de voz reproduzida na saída do decodificador. Em função disso surge a necessidade de interpolar, usualmente de forma linear. Isso é feito de modo a obter valores intermediários dos parâmetros, em sub-quadros tipicamente de 5 ms, ao longo do intervalo de tempo entre uma transmissão do conjunto de parâmetros LSF e a anterior.

Em [5] foi proposto um método de ajuste dos coeficientes LSF do último sub-quadro, com base nas diferenças entre os valores interpolados e os referenciais (valores reais calculados para cada sub-quadro individualmente). Esse procedimento é feito com base em uma medida de distorção, em particular a distância LSF ponderada euclidiana, d_{LSF} [6].

Esta dissertação será focada principalmente em uma modificação do processo de interpolação, que consiste em ajustar os parâmetros LSF, com base em uma aproximação da distorção espectral, em vez da distância LSF ponderada euclidiana, d_{LSF} . Obviamente, o objetivo é que a modificação introduzida proporcione um melhor desempenho.

Na avaliação da performance será levada em conta a complexidade computacional, além das medidas de distorção usuais, DE média, outliers, tanto de 2 a 4%, quanto maiores que 4%, a já citada distância LSF ponderada euclidiana, d_{LSF} , e o teste PESQ, que é um método objetivo de avaliação perceptual da qualidade de voz, proposto na recomendação ITU-T P.862 [7][8]. É importante observar que as medidas de distorção são aplicadas nos coeficientes LPC, no caso da DE, e nos LSF's, no caso da d_{LSF} , visando avaliar distorções associadas aos parâmetros LSF's obtidos em cada método estudado. Os resultados foram obtidos antes e depois da quantização dos parâmetros. Tais medidas não permitem uma avaliação direta da qualidade de voz. O teste PESQ é o encarregado de avaliar a voz sintetizada, utilizando-se aqui um codificador recentemente proposto em [9].

1.2

Organização da Dissertação

O capítulo 2 inicialmente descreve o processo de codificação de voz de modo geral, discorrendo sobre a análise LPC, a conversão dos parâmetros LPC para LSF's, para depois entrar em detalhes do codificador escolhido para a realização das simulações.

O capítulo 3 considera o método de ajuste de coeficientes LSF, do ponto de vista da proposta original, em [5], ainda sem nenhuma modificação incluída. De início são tratados a formação dos parâmetros LSF, o processo de interpolação e a distância euclidiana ponderada (d_{LSF}), que é a medida de distorção utilizada

em [5]. São apresentadas também informações mais detalhadas sobre algumas das medidas utilizadas na avaliação de desempenho dos diversos métodos estudados, entre elas a distorção espectral (DE), os outliers, e o teste PESQ, que mede efetivamente a qualidade de voz. Em seguida, é apresentada uma descrição detalhada do ajuste de LSF's, tanto nos modos com e sem atraso. Logo após, são tratados os coeficientes de otimização, necessários ao processo de ajuste. Por último são apresentados resultados de simulações e conclusões à respeito do desempenho, fundamentadas através de tabelas e gráficos.

O capítulo 4 inicialmente apresenta uma relação [5] existente entre a distorção espectral (DE) e a d_{LSF} . É utilizado o método dos mínimos quadrados, comum no cálculo básico, para se chegar a uma equação que expressa a DE de modo aproximado em função da d_{LSF} . Além disso, também é feita a avaliação da qualidade dessa aproximação. Finalmente, chega-se à parte que pode ser considerada de maior destaque dessa dissertação, pois é a que tem o caráter mais inovador. Nela é apresentada uma nova formulação para o ajuste de LSF's, buscando minimizar a DE em sua forma aproximada, sem a necessidade de treinamento de coeficientes de otimização. Essa última parte é complementada com a apresentação de resultados de simulações.

O capítulo 5 apresenta conclusões gerais sobre o estudo realizado e sugestões para trabalhos futuros.