

1 Introdução

1.1 Motivação e Objetivos

A rede pública de telefonia comutada e os equipamentos que a tornam funcional estão implantadas em praticamente todas as localidades do mundo hoje. A disponibilidade de um telefone e o acesso a uma rede telefônica mundial de alta qualidade e baixo custo é considerado essencial em nossa sociedade moderna. Tudo que possa vir a comprometer esta rede é tratado com muito cuidado. Entretanto, uma mudança de paradigma começa a ocorrer quando mais e mais as comunicações se dão na forma digital e transportadas por redes de pacotes, tais como IP (*Internet Protocol*), Células ATM (*Asynchronous Transfer Mode*) e quadros de *frame relay*. O tráfego de dados está crescendo muito mais que o tráfego telefônico e há um interesse cada vez maior em transportar informações de voz sobre rede de dados alterando radicalmente o modo tradicional de transporte de dados sobre rede de voz.

O suporte às comunicações de voz utilizando o protocolo IP, mais conhecido como VoIP, ou "Voz sobre IP", torna-se um atrativo pelo baixo custo e pelo preço variável de enlaces de dados privados de acordo com a taxa de transmissão contratada pelos clientes. De fato, a busca por uma boa qualidade de telefonia em redes IP é um dos passos chave em direção à convergência das aplicações de voz, vídeo e comunicação de dados [1].

VoIP pode ser definida como a habilidade de se fazer chamadas telefônicas, como por exemplo operar todas as facilidades oferecidas hoje pela rede de telefonia convencional (rede pública de telefonia comutada) e enviar FAX em redes de dados baseadas em IP com um padrão aceitável de qualidade de serviço (QoS) [1].

Devido a esse crescimento gigantesco da Internet e dos sistemas de comunicações móveis celulares, as aplicações de processamento de voz nesses meios têm despertado um interesse crescente. Dentre as áreas de destaque

ressaltam-se a codificação e o reconhecimento de voz.

Tanto no caso de redes IP, como de telefonia móvel celular, os problemas relacionados ao projeto de codificadores e reconhecedores de voz eficientes são acentuados pelas altas taxas de erro de bits e perdas de pacotes, fora outros problemas usuais na concepção desses sistemas, como o ruído ambiente. Todos esses aspectos tornam-se mais sérios pelo fato dos esquemas de codificação de voz usualmente operarem a baixas taxas de bits.

Atualmente a maior parte dos algoritmos de codificação de voz a baixas taxas é baseada na técnica de análise por predição linear (ou análise LPC - *Linear Predictive Coding*), onde um sinal de excitação é aplicado a um filtro só de pólos, caracterizado pelos parâmetros LPC, que representa a informação da envoltória espectral do sinal de voz [2][3]. As frequências em linhas espectrais ou parâmetros LSF (*Line Spectrum Frequencies*) usualmente são escolhidos para representar os coeficientes LPC, uma vez que são mais adequados para procedimentos de quantização e interpolação [4].

Este trabalho trata do problema da transmissão da informação da envoltória espectral da voz e seu impacto sobre a qualidade da voz em codecs a baixas taxas, usando diferentes esquemas de quantização vetorial (QV) em redes IP. A análise se concentra na transmissão dos parâmetros LSF, pois estes consomem uma parcela significativa da taxa de bits total do codec, carregando a maior parte da informação. O desempenho dos quantizadores será avaliado utilizando a medida de distorção espectral, percentual de *outliers* e análise da qualidade de voz em codificadores a baixas taxas usando a medida PESQ (*Perceptual Evaluation of Speech Quality*) da recomendação ITU-T P.862 [5][6]. Esses estudos também serão realizados em cenários ruidosos, com diferentes tipos de ruído: branco, ruído de fábrica e ruído de falatório. Com o objetivo de reduzir as degradações causadas por esses ruídos, será avaliada, ainda, a vantagem da aplicação de uma técnica de supressão de ruído baseada em transformadas *wavelet*. Seu impacto na qualidade da voz codificada será avaliado através de testes de escuta informais, da recomendação ITU-T P.862 de avaliação perceptiva de qualidade da voz (PESQ) e testes subjetivos de comparação A/B.

1.2

Organização da Dissertação

O Capítulo 2 descreve o sistema sobre o qual serão realizados os estudos ao longo da dissertação, mostrando suas principais características e

os problemas envolvidos. Inicialmente é feita uma breve introdução à análise LPC e são apresentados diversos esquemas de quantização dos parâmetros LSF. Em seguida, é feita uma descrição do codec utilizado nas simulações, mostrando suas principais características. E, finalmente, o capítulo trata dos problemas envolvidos na transmissão e recuperação de LSFs em redes IP.

O Capítulo 3 apresenta e discute os resultados das simulações dos diferentes esquemas de quantização em codecs operando sobre redes IP. Nas simulações são utilizadas locuções tanto em ambientes livres de ruído como em presença de ruído branco, ruído de fábrica e ruído de falatório, com a RSR (Razão Sinal-Ruído) variando de -5dB a 30dB. A avaliação dos quantizadores é feita através da medida de distorção espectral e do percentual de *outliers*. Para análise da qualidade de voz é utilizada a medida PESQ da recomendação ITU-T P.862.

O Capítulo 4 apresenta , inicialmente, uma introdução a teoria *wavelet* e um algoritmo de supressão de ruído baseado em transformadas *wavelets*. Em seguida, são apresentados os resultados do impacto da utilização desse algoritmo na qualidade da voz codificada por um codec a baixa taxas de bits, operando em rede IP. A qualidade da voz é avaliada através do PESQ e de testes de comparação A/B.

Finalmente, o Capítulo 5 apresenta as conclusões da dissertação e sugestões para trabalhos futuros.