1 Introdução

Esta dissertação tem como principal objetivo analisar e propor esquemas eficientes de reconhecimento de voz distribuído em redes IP e redes de telefonia móvel celular.

Neste capítulo se fará uma breve introdução da problemática envolvida no reconhecimento de voz em redes IP e redes de telefonia móvel celular, visando preparar e introduzir o leitor ao assunto a ser tratado nos demais capítulos desta dissertação. Será apresentado, também, a estrutura dos capítulos e um breve resumo do conteúdo deste trabalho.

1.1. Sistemas de Reconhecimento de Voz em Ambiente Celular/Redes IP

O desenvolvimento tecnológico do mundo atual tem estimulado demanda cada vez maior por máquinas inteligentes. Dentro desse panorama, a área de reconhecimento automático de voz (RAV) é uma das que têm despertado maior interesse, apesar da grande complexidade envolvida em termos de projeto e de operação. Esse interesse crescente tem sido evidente tanto no âmbito das indústrias como dos centros de pesquisa no mundo inteiro.

Tendo em vista o crescimento gigantesco da Internet e dos sistemas de comunicações móveis celulares, as aplicações de processamento de voz nesses meios têm despertado interesse cada vez maior. Em particular, um problema importante nessa área diz respeito ao reconhecimento de voz em um sistema servidor, a partir de parâmetros acústicos calculados e quantizados no terminal do usuário. O servidor reconhece a voz de acordo com uma aplicação específica e envia de volta, ao usuário, informações relativas à ação tomada a partir do reconhecimento de voz.

Os parâmetros acústicos podem ser os especificados por um *codec* de voz – caso o serviço de voz seja também utilizado – ou podem ser os vetores de atributos que serão efetivamente empregados pelo reconhecedor de voz. Em

qualquer situação, esses parâmetros serão digitalizados, através de um esquema de codificação de baixas taxas, e transmitidos ao servidor em canais de comunicações que usualmente apresentam limitação de faixa, como a Internet e os sistemas de telefonia móvel. É exatamente por essa limitação que os parâmetros devem utilizar esquemas de compressão que sejam eficientes. Essa eficiência, entretanto, é medida no sentido de servir bem ao propósito de reconhecimento e não de qualidade de voz. Esse seria um cenário típico de aplicação da tecnologia de reconhecimento de voz à Internet e aos canais de telefonia móvel celular.

Devido à alta complexidade computacional e à grande quantidade de memória requerida em sistemas de RAV, se torna muito atraente a opção por sistemas de reconhecimento de voz distribuídos. Em sistemas desse tipo, o processamento é distribuído entre o terminal do usuário (telefone celular, computador pessoal) e o terminal de recepção em uma rede de comunicações (estação base em redes de telefonia móvel, servidor central em redes IP).

Os problemas relacionados ao projeto de reconhecedores de voz distribuídos para operação na Internet e em redes de telefonia móvel são acentuados pelas altas taxas de erro de bits e perdas de pacotes, fora outros problemas usuais na concepção de sistemas de RAV, como o ruído ambiente.

Além disso, os esquemas de codificação de voz usados operam a baixas taxas de bits e utilizam, em geral, codificação preditiva linear ou LPC (*Linear Predictive Coding*), com base em um modelo de produção da fala. Nesse modelo, um sinal de excitação é aplicado a um filtro só de pólos (caracterizado por parâmetros LPC), que representa a informação da envoltória espectral do sinal de voz. Usualmente os parâmetros LPC são transformados para LSF (*Line Spectral Frequencies*), devido às propriedades atraentes destes últimos para os processos de quantização e interpolação. No caso de sistemas de RAV distribuídos é preferível utilizar diretamente os parâmetros do *codec* do que extraí-los a partir do sinal decodificado, como será apresentado mais a frente nesta dissertação. A realização desse processamento envolve um grande número de aspectos e estratégias para concepção de reconhecedores de voz eficientes.

Para isso, conforme mencionado na caracterização do problema, diversos aspectos e estratégias deverão ser considerados. Primeiramente, os parâmetros LSF do *codec* não são necessariamente as melhores opções de atributos a serem usadas no RAV. Portanto, transformações desses parâmetros são estratégias

importantes a serem consideradas. Uma outra estratégia visada no projeto de reconhedores de voz no ambiente celular e de redes IP consiste em incorporar outros parâmetros já disponíveis no decodificador, de modo a melhorar o desempenho do reconhedor de voz. É de interesse, também, investigar atributos que sejam mais robustos em presença de ruído. Quando o serviço previsto é apenas o de reconhecimento, é importante examinar não só os novos conjuntos de atributos que sejam mais robustos, assim como novos métodos de codificação específicos para os atributos a serem empregados. Problemas relacionados ao comportamento do sistema em presença de erros no canal e de perdas de quadros, além da escolha do domínio de interpolação dos quadros, também são itens que devem ser examinados.

Dentre as diversas técnicas e problemas aqui apresentados, esta dissertação trata da obtenção de atributos de reconhecimento eficientes a partir dos parâmetros dos codificadores utilizados em ambientes celulares e de redes IP, bem como o melhor domínio de interpolação destes atributos.

Cabe ressaltar ainda que foi buscada a redução da complexidade computacional, sem comprometer o rendimento do sistema reconhedor de voz.

1.2. Organização do Trabalho

A importância e a necessidade dos sistemas de reconhecimento de voz em ambientes móveis celulares e de redes de voz sobre IP, motivaram, como apresentado, o desenvolvimento do presente trabalho, o qual será organizado da seguinte forma:

- Capítulo 2 – é feita uma introdução dos sistemas de reconhecimento de voz. Para isso são apresentados os sons da fala e suas classificações, unidades fonéticas utilizadas para reconhecimento, modelo de produção do sinal de voz e o reconhecimento automático de voz utilizando modelos de Markov escondidos.
- Capítulo 3 – descreve os sistemas de reconhecimento distribuído abordando as técnicas e parâmetros mais utilizados. Apresenta, ainda, o sistema de reconhecimento distribuído a ser analisado nesta dissertação,

bem como descreve as técnicas utilizadas para sua implementação e obtenção dos resultados apresentados nos Capítulos 5 e 6.

- Capítulo 4 – discorre sobre as etapas de pré-processamento da voz para reconhecimento e sobre a dedução matemática dos parâmetros de reconhecimento de voz utilizados nesta tese.
- Capítulo 5 – analisa os atributos em reconhecimento de voz distribuído, tendo como finalidade concluir qual dentre eles possui melhor desempenho de reconhecimento e em qual domínio é interessante interpolá-lo.
- Capítulo 6 – apresenta o codificador de voz ITU-T G.723.1, com o respectivo resultado de reconhecimento, utilizando os melhores parâmetros obtidos no capítulo anterior.
- Capítulo 7 – finaliza o trabalho com algumas conclusões gerais e sugestões para trabalhos futuros.
- Apêndice – consiste em um breviário de utilização do HTK, onde é apresentado sua finalidade e sua utilização em conjunto com o MATLAB®, pois o MATLAB® será o gerador dos atributos de reconhecimento para que o HTK possa realizar o reconhecimento (as interfaces entre os mesmos serão aqui apresentadas).