1

Introdução

O reconhecimento automático de voz surgiu na metade do século XX, quando os avanços da eletrônica e o desenvolvimento da tecnologia na área computacional davam passos a uma nova era, que permitiria projetar novas ferramentas computacionais capazes de processar o sinal de voz emitido pelo ser humano e reconhecer a informação contida nesse sinal. Em outras palavras, criar sistemas de reconhecimento de padrões capazes de identificar uma mensagem dentro do sinal de áudio adquirido, com o fim de fazer uma interação oral homem-máquina em ambos os sentidos, visando substituir as interfaces tradicionais baseadas em teclados ou dispositivos similares.

Porém, as primeiras ferramentas e técnicas de processamento de voz propostas geravam baixas taxas de reconhecimento, em uma ordem muito inferior às taxas atribuídas aos seres humanos. Também utilizavam algoritmos de processamento de voz de baixa eficiência e alto custo computacional, devido a condições adversas no processo de comunicação, como efeitos do canal ou efeitos do meio ambiente, os quais indiscutivelmente geram problemas para o reconhecimento.

Em torno disso, a comunidade científica começou a estudar a possibilidade de desenvolver técnicas de reconhecimento que sejam mais robustas às condições adversas, permitindo obter processos muito mais eficientes e mais próximos ao diálogo entre dois humanos em situações reais [1].

Desde o início dos anos 80, com a utilização dos modelos ocultos de Markov (HMMs), conseguiu-se melhorar o reconhecimento de voz contínua, permitindo aumentar o tamanho do vocabulário, obtendo-se, assim, situações cada vez mais reais. Porém, as condições adversas, indesejáveis no reconhecimento de voz contínua ainda são difíceis de serem evitadas completamente do sinal de voz contínua. Portanto é bem provável a presença do ruído de fundo como ruído de rua, ruído de motor, etc.

É por isso que na atualidade têm se destinado muitos recursos às técnicas de tratamento do sinal em ambientes ruidosos, com o objetivo de evitar a disparidade entre as condições de treinamento e as de reconhecimento, dando origem às chamadas técnicas robustas de processamento de voz.

Elas podem ser divididas em três categorias principais, segundo seu enfoque principal [2]:

- Técnicas de realce de fala.
- Técnicas de compensação de atributos.
- Técnicas de adaptação de modelos.

É importante ressaltar que, embora muito se tenha aprendido a respeito da forma como implementar sistemas robustos, nas últimas décadas o reconhecimento de voz continua sendo uma tarefa difícil, devido às distorções frente a condições adversas e ao aumento de palavras no dicionário.

1.1 Motivação

Na atualidade têm se desenvolvido técnicas que melhoram a robustez dos sistemas de reconhecimento, Porém as taxas de reconhecimento dessas técnicas estão longe das atribuídas aos seres humanos, fazendo com que nos últimos anos a maior parte da pesquisa em sistemas de reconhecimento esteja centrada no reconhecimento de voz contínua.

Visando prover diversas soluções às necessidades de comunicação atual, esta pesquisa foca em dois pontos de vista importantes à hora de desenvolver um sistema completamente eficiente, os quais são:

- Funcional
- Econômico

Do ponto de vista funcional, procura-se desenvolver algoritmos que sejam capazes de emular a comunicação oral, conseguindo que mediante a devida sequência de funções cumpra o objetivo para o qual foi projetado, de modo que melhore seu rendimento, facilitando assim o trabalho do usuário.

Com relação ao fator econômico procura-se uma forma de implementar algoritmos ainda mais eficientes que contribuam às necessidades do mundo moderno, acelerando as operações e reduzindo custos, garantindo a maior confiabilidade e melhorando assim o grau de precisão e desempenho que possa-se alcançar.

Devido à análise anterior, nasce a motivação do presente trabalho, abordando os problemas correspondentes às distorções do sinal de voz devido às condições de ruído desenvolvendo algoritmos que sejam capazes de fornecer

maior robustez e que proporcionem o reconhecimento de voz com suficiente qualidade.

1.2

Objetivos da dissertação

Esta dissertação de mestrado centra-se no objetivo fundamental de misturar técnicas de robustez existentes, com o objetivo de propor melhoras que permitam com que o reconhecimento de voz seja mais robusto às condições adversas, especialmente em presença ruído acentuado, e que os sistemas sejam computacionalmente mais eficientes.

Conforme mencionado na caracterização do problema, diversos aspectos e estratégias relativos aos sistemas de reconhecimento de voz contínua foram considerados. Para tal, propõe-se alcançar os seguintes objetivos que serão alcançados com esta dissertação:

- Pesquisa bibliográfica e estudo teórico sobre sistemas de reconhecimento de voz e métodos de robustez desenvolvidos até a atualidade, os quais estabelecem uma perspectiva diferente à hora de se enfrentar a problemática do reconhecimento de voz em condições adversas.
- Revisar as estratégias utilizadas para o reconhecimento de voz.
- Revisar as técnicas e procedimentos de extração de atributos do sinal de voz.
- Analisar o efeito do ruído sobre o sinal de voz e no reconhecimento. Este é um ponto importante, já que é o desempenho dos reconhecedores de voz diminui quando o ambiente acústico distorce o espaço de representação do sinal de voz.
- Simulação dos algoritmos de robustez de voz cobrindo uma vasta gama de níveis de ruído, com o propósito de avaliar a eficiência e a capacidade de cada um deles para eliminar os efeitos do ruído sobre o sinal de voz.
- Apresentação e discussão dos resultados experimentais.

1.3

Estrutura da dissertação

Esta dissertação foi organizada em sete capítulos, sendo o primeiro esta introdução.

No capítulo 2 faz-se uma breve apresentação dos fundamentos e conceitos básicos do reconhecimento de voz. Também terá uma introdução aos subsistemas típicos de um sistema de reconhecimento de voz, apresentando sua arquitetura básica, e se fornecerá uma breve descrição dos modelos ocultos de Markov.

O capítulo 3 apresenta as técnicas e os procedimentos de extração dos atributos do sinal de voz.

O capítulo 4 faz uma apresentação e descrição do sistema de reconhecimento de voz em presença de ruído. Serão apresentados os conceitos básicos das técnicas de robustez e serão analisadas as ferramentas de software e condições de trabalho utilizadas para a elaboração do projeto. Além disso, este capítulo faz uma introdução das soluções de robustez propostas segundo duas estratégias a saber *Técnicas de realce de fala* e *Técnicas de compensação de atributos*.

Os capítulos 5 e 6 estão dedicados por completo às propostas que foram elaboradas para cumprir integralmente os objetivos propostos neste trabalho de reconhecimento robusto em presença de ruído. São descritos cada um dos métodos baseados em *Técnicas de Pre-extração de atributos* e *Técnicas de Pos-extração de atributos*, além de serem apresentados e discutidos os testes realizados e os resultados obtidos.

Finalmente, no capítulo 7 é apresentada uma série de conclusões e contribuições do trabalho e também são relacionadas possíveis linhas de trabalhos futuros.