

# 7

## Conclusões e Sugestões para Trabalhos Futuros

Neste último capítulo apresentam-se resumidamente as conclusões gerais obtidas a partir do projeto e desenvolvimento dos métodos de robustez propostos neste trabalho, com o objetivo de melhorar o desempenho do reconhecimento de voz contínua em presença do ruído. Embora algumas dessas técnicas já sejam conhecidas na literatura, a comparação e a combinação entre elas trouxeram vários resultados interessantes e inéditos.

### 7.1 Conclusões

No presente trabalho foi feita a pesquisa sobre alguns dos métodos de robustez do reconhecimento de voz contínua. Embora tenha se observado que não há métodos que atingem taxas de 100% no reconhecimento, conseguiram-se resultados mais que satisfatórios através das técnicas de *pré-extracção de atributos* e *pós-extracção de atributos*, chegando a um ponto de desenvolvimento tal que possa ser considerado que os objetivos estabelecidos no início desta dissertação como atingidos com sucesso. Esses objetivos são: eliminação do ruído, número reduzido de operações e sinal resultante idêntico ao original, sem degradação da inteligibilidade do discurso.

Em primeiro lugar foi utilizado um sistema de referência *Ref*, o qual refere-se a uma técnica de reconhecimento que simplesmente utiliza a extração de atributos do sinal de voz para o reconhecimento. Para isso, foram analisados dois métodos. O primeiro baseado em atributos MFCC, devido ao fato de ser uma das técnicas padrão mais utilizadas no sistema de reconhecimento, principalmente porque adapta-se bem às hipóteses utilizadas para estimar as distribuições de estados em HMM, convertendo-se no sistema de referência ideal para observar e comparar o rendimento dos métodos de robustez. O segundo através de atributos PNCC, pela robustez ao ruído superior em relação a outras técnicas alternativas de extração de atributos, como a mencionada MFCC. Mostrando claramente ser os mais robustos, com resultados melhores em todos os testes analisados.

A avaliação deste sistema *Ref* com apenas os métodos de extração de atributos, mostrou que os dois métodos são eficazes para reduzir os efeitos nocivos do ruído aditivo. Porém, o método PNCC foi claramente o mais próximo do comportamento do sistema auditivo humano, obtendo resultados mais robustos em todos níveis de sinal-ruído. Estes resultados, embora sejam melhoráveis, são suficientemente bons para desenvolver reconhecedores de voz.

A segunda parte do trabalho foi dedicada a aumentar a robustez do sistema de reconhecimento, realizando a pesquisa, implementação e análise de um conjunto de métodos, que sejam o mais simples possível, porém com um bom grau de robustez do reconhecedor na hora de este operar em condições adversas. Para isso este trabalho apresentou e discutiu quatro formas diferentes de melhorar o desempenho do sistema através de métodos que compensem o descasamento entre os dados de teste e os de treinamento em diferentes domínios: *Wavelet Denoising* e *Subtração espectral*, no domínio do espectro da magnitude, e *Mapeamento de histogramas* e *Filtro com redes neurais*, no domínio dos coeficientes cepstrais.

As principais conclusões deste trabalho são:

- Os testes feitos com os quatro métodos de robustez propostos apresentaram a eficiência de sua utilização no sistema de reconhecimento dependendo do tipo de parametrização utilizada.
- No caso dos métodos de *pre-extracção de atributos*, foram apresentados e discutidos dois algoritmos.

O primeiro foi a *Subtração espectral*. Seu principal atrativo é sua simplicidade matemática. Este método, aplicado com atributos MFCC, mostrou uma melhora considerável nas taxas de acerto do sistema, comparado com o sistema *Ref*. Porém, tem um rendimento limitado porque à medida que se aumenta a eliminação do ruído original, pode-se inserir um outro tipo de ruído, denominado ruído musical. Já para os atributos PNCC, as taxas de acerto não foram favoráveis, já que o método removeu informação importante do sinal, resultando em taxas de acerto piores que as de referência.

O segundo foi o *Wavelet denoising*. Diferentemente do que acontece com a *Subtração espectral* quando é aplicado com atributos MFCC, os resultados tiveram um desempenho menor em todos os níveis de sinal-ruído. Já os PNCC foram ligeiramente beneficiados. Embora mais eficazes que os MFCC em presença de ruído, tiveram um desempenho inferior no caso de sinais limpos.

É preciso dizer que para os métodos de *pré-extracção de atributos*, alguns parâmetros como  $\alpha$  e  $\beta$  da 5.1, tiveram que ser ajustados manualmente, variando de acordo com a intensidade do ruído.

- Para o caso dos métodos de *pos-extracção de atributos*, foram apresentados dois métodos de mapeamento de atributos capazes de tirar vantagem das relações não lineares entre dois espaços acústicos. Estes métodos se enquadram na categoria de compensação de atributos e permitem melhorar o rendimento do reconhecedor, apresentando diferenças apreciáveis, tanto em MFCC como PNCC, ao contrário do que ocorre com a técnica de *pré-extracção de atributos*. Os métodos implementados foram: o *Filtro com redes neurais* e o *Mapeamento de histogramas*.

Aplicando o FRN com os MFCC, os resultados melhoraram consideravelmente. Porém, os resultados obtidos quando o FRN é usado com atributos PNCC não são satisfatórios. Isto é, o FRN remove o ruído à custa de distorcer, de alguma forma, o sinal limpo, o que pode indicar que possivelmente o critério utilizado para fazer o treinamento não é adequado.

Por outro lado, a aplicação do método MAP requer menção especial, já que em comparação com os outros métodos, mostrou ser eficiente na hora de tornar o sinal mais robusto, aumentando consideravelmente o desempenho do sistema. Além de ser um mecanismo de baixo custo computacional que proporciona melhorias sob qualquer nível de razão sinal-ruído e diferentes parametrizações do sinal de voz. Concluindo assim, que a utilização de MAP tornou-se o método mais bem sucedido neste trabalho.

A vantagem dos métodos de *pós-extracção de atributos* foi não precisar ajustar nenhum tipo de parâmetro manualmente durante os testes.

- Uma vez analisados os resultados dos métodos agindo isoladamente, procurou-se testar e analisar a mistura desses métodos duas a duas, no mesmo sistema de reconhecimento. A primeira mistura implementada foi *Subtração espectral* e *Wavelet denoising*. Verificou-se que seu emprego com coeficientes MFCC é vantajosa em relação ao uso isolado de *Subtração Espectral* e *Wavelet denoising*, apresentando as melhores taxas de reconhecimento do sistema. Entretanto, quando a mistura é usada com atributos PNCC, a vantagem ocorre só para o método isolado *Subtração espectral*, fazendo que não seja recomendável usar essa dupla com os atributos PNCC.

A segunda mistura testada engloba *Mapeamento de histogramas* e *Filtro com redes neurais*. Observam-se melhorias significativas em relação ao uso

isolado de MAP quanto no emprego isolado de FRN, quando são usados coeficientes MFCC. No entanto, para coeficientes PNCC, observou-se melhoras significativas em relação ao emprego isolado do método FRN. Já que a mistura dos métodos evidenciou como o FRN prejudica a compensação imposta pelo método MAP.

A fim de obter sistemas robustos de reconhecimento de voz, foram feitos experimentos de reconhecimento aplicando os métodos propostos de forma isolada e misturando-os em diferentes domínios. Obtendo-se como conclusão final que a combinação de *Mapeamento de histogramas* com atributos PNCC tornou-se a configuração mais bem sucedida neste trabalho.

## 7.2 Sugestões para trabalhos futuros

Finalmente esta dissertação deixa as portas abertas a futuras revisões e melhorias, afim de aperfeiçoar ainda mais o desempenho do sistema.

- Estender o estudo das técnicas de robustez propostas em testes em que o ruído não seja adicionado artificialmente à voz limpa, permitindo assim uma maior flexibilidade do sistema em relação ao universo do falante.
- Definir testes com outros bancos de dados maiores do que a TIMIT, que possam comprovar os resultados obtidos.
- Apesar dos bons resultados obtidos com os atributos MFCC e PNCC, deixa-se como sugestão avaliar um outro tipo de parametrização nas diferentes técnicas que possa fazer ainda mais robusto o sistema.
- Misturar as técnicas de *pré-extração de atributos* e *pós-extração de atributos* a fim de aproveitar os benefícios que cada uma delas oferece em separado.
- Avaliar os resultados dos diferentes métodos através de diferentes tipos de ruído, já que neste trabalho só se avalia o efeito do ruído aditivo branco.
- Com relação às redes neurais, poderia-se experimentar outras possíveis arquiteturas que se misturem com o MAP, afim de estabelecer redes como estimadores de funções de densidade de probabilidade.
- Não obstante, os bons resultados fornecidos utilizando a função de referência gaussiana no caso do MAP, poderia se tentar experimentar com outras funções de referência, a fim de encontrar resultados eventualmente melhores.