

7

Conclusões e Sugestões para Trabalhos Futuros

Este trabalho abordou meios de melhorar o desempenho do reconhecimento de voz contínua na presença de ruído aditivo. Para isso, foram analisadas e testadas três técnicas. A primeira delas foi a extração de atributos do sinal de voz usando os métodos MFCC, SSCH e PNCC. A segunda foi a remoção de ruído do sinal de voz na via *wavelet denoising*, através da função de limiar. A terceira e última, chamada de *feature denoising*, melhorou os atributos extraídos usando um conjunto de redes neurais. Embora algumas dessas técnicas já sejam conhecidas na literatura, a comparação e a combinação entre elas trouxeram vários resultados interessantes e inéditos.

Inicialmente, avaliou-se apenas os métodos MFCC, SSCH e PNCC. Os SSCH, embora mais eficazes que os MFCC em presença de ruído, tiveram um desempenho inferior nos casos de sinais mais limpos. Já os PNCC foram claramente os mais robustos, com resultados melhores em todos os cenários analisados.

Em seguida, foi testado o *wavelet denoising*. Aplicando essa técnica com os MFCC, os resultados melhoraram bastante, confirmando os testes de outros autores. E, pela primeira vez, foi mostrado que também ela também aumenta bastante o desempenho dos SSCH. No entanto, os PNCC não foram muito beneficiados, embora suas taxas tenham se mantido as mais altas. Também é preciso observar que um dos parâmetros do *denoising* teve que ser ajustado manualmente, variando de acordo com a intensidade do ruído.

Finalmente, a técnica proposta de *feature denoising* também aumentou as taxas de acerto dos métodos de extração. E, ao contrário do que ocorreu com o *wavelet denoising*, o desempenho dos PNCC aumentou consideravelmente. Outra vantagem foi não precisar ajustar nenhum parâmetro durante os testes. Por isso, a combinação de *feature denoising* com PNCC se tornou a configuração mais bem sucedida neste trabalho.

Apesar dos resultados positivos, o sistema proposto ainda pode ser aperfeiçoado. Cada vez que se melhora uma das etapas individuais, aumenta-se um pouco as taxas de acerto no final. E talvez certas modificações possam tornar o encadeamento de *wavelet denoising*, extração de atributos e *feature denoising* a configuração mais eficaz.

Portanto, seguem abaixo algumas sugestões para trabalhos futuros.

- Muitos parâmetros foram determinados experimentalmente, como explicado na seção 6.3. Com alguns testes a mais, deve ser possível chegar a um novo conjunto de parâmetros que resultem em taxas de acerto mais altas.
- Existem bases de dados de voz maiores do que a TIMIT. Com mais amostras de voz, os HMMs de trifones seriam estimados com mais precisão. Da mesma forma, mais exemplos de frases tornariam o modelo de linguagem mais completo.
- Outras ferramentas de reconhecimento de voz poderiam ser testadas. Embora tenha sido muito útil, o HTK não faz uso de tecnologias computacionais modernas, tais como o aproveitamento da placa gráfica e dos múltiplos núcleos do processador.
- Existem outros métodos de extração de atributos na literatura. Talvez alguns deles sejam mais robustos que os PNCC.
- O parâmetro γ do *wavelet denoising* teve que ser ajustado manualmente de acordo com o nível de ruído. Um método automático poderia ser criado para estimar o valor mais apropriado, considerando as características do sinal de entrada.
- O conjunto de redes neurais poderia ser treinado com mais amostras de atributos e com outras razões sinal-ruído além de 10 dB e 20 dB. Dessa forma, o *feature denoising* estaria preparado para mais tipos de sinais corrompidos.