

## 5

### A Utilização da Técnica do Espaço Nulo e dos Atributos Baseados na Escolha de Coeficientes de Autocorrelações

Este capítulo apresenta uma nova proposta que consiste em empregar os atributos baseados na seleção de coeficientes de autocorrelações do sinal de voz considerando os múltiplos classificadores em sub-bandas e o espaço nulo, no reconhecimento de locutor. Essa estratégia visa observar se a remoção de alguns coeficientes de autocorrelações, geralmente associados aos ruídos, podem contribuir para a melhoria do desempenho do reconhecimento quando é utilizada a técnica do espaço nulo.

Os atributos AMFCC (*Autocorrelation Mel Frequency Cepstral Coefficients*) foram apresentados em [31] e usados para reduzir o efeito do ruído em reconhecimento de voz. A extração consiste das seguintes etapas: cômputo dos coeficientes de autocorrelações do sinal de voz em cada janela, exclusão das autocorrelações de baixos *lags* (índices), e aplicação das restantes autocorrelações no esquema de extração dos atributos MFCC, obtendo-se assim um vetor de atributos AMFCC a cada janela. Nesse esquema, as autocorrelações são usadas para gerar uma estimativa do espectro de potência do sinal de voz. Os coeficientes de autocorrelações de baixos *lags* são descartados porque o ruído que ocupa uma ampla faixa do espectro de frequências (ruído faixa larga) afeta mais a essas componentes. Ressalta-se que no processo de extração são utilizados somente os coeficientes de um lado da sequência de autocorrelações, já que é simétrica. No esquema de reconhecimento proposto, esses atributos são extraídos em cada sub-banda.

A organização deste capítulo é a seguinte: nesta Seção são descritos os fundamentos da proposta de reconhecimento que usa os atributos AMFCC e o espaço nulo; a Seção 5.1 apresenta os resultados experimentais; e a Seção 5.2, as principais conclusões deste capítulo.

#### 5.1

##### Resultados de Simulação

Nesta Seção, são apresentados os principais resultados experimentais de identificação de locutor independente do texto, empregando os atributos AMFCC.

O objetivo é avaliar o comportamento dos atributos AMFCC, quando comparados com outros esquemas. As condições experimentais são as mesmas dos capítulos anteriores. Na extração dos atributos AMFCC, são descartados o coeficiente de *lag* zero e os três primeiros coeficientes de baixos *lags* de um dos lados da sequência, na quarta sub-banda – a que tem as maiores frequências.

A proposta apresentada nesse capítulo utiliza o espaço nulo e os atributos AMFCC. O efeito da combinação usando pesos iguais (vetor de elementos unitários) foi incluído no conjunto de vetores produzidos pelo espaço nulo. Portanto, o novo conjunto tem  $n$  vetores ao invés de  $n-1$ . Isto foi realizado desta forma para incluir o efeito de independência condicional entre sub-bandas [38], já que busca-se reduzir o efeito do ruído antes da etapa de combinação das medidas de verossimilhança.

Os resultados experimentais expressos em termos das taxas de identificação estão apresentados nas Tabs. 5.1 a 5.6. Nessas tabelas, os esquemas que empregam os atributos AMFCC estão designados por “AMFCC (Soma, 4Sbs)”, “AMFCC (Prop1A e Soma, 4Sbs)”, “AMFCC (Prop1A e Soma, 4Sbs) delta”, “AMFCC (Prop1A e Soma, 4Sbs) delta-delta”, e “AMFCC (Prop1A e Soma, 4Sbs) delta, delta-delta”.

|  | Fábrica | Falatório | Carro | Branco |
|--|---------|-----------|-------|--------|
| Soma (4Sbs)                                    | 66,33   | 72,45     | 67,86 | 30,10  |
| Prop1A (4Sbs)                                  | 75,51   | 81,12     | 76,53 | 40,82  |
| Prop1A e delta (4Sbs)                          | 75,00   | 80,61     | 78,06 | 39,29  |
| Prop1A e delta-delta (4Sbs)                    | 74,49   | 80,57     | 79,08 | 37,76  |
| Prop1A e delta, delta-delta (4Sbs)             | 78,06   | 83,16     | 78,06 | 40,82  |
| AMFCC (Soma, 4Sbs)                             | 62,76   | 70,41     | 65,82 | 34,18  |
| AMFCC (Prop1A e Soma, 4Sbs)                    | 70,92   | 75,00     | 75,00 | 51,53  |
| AMFCC (Prop1A e Soma, 4Sbs) delta              | 70,41   | 76,53     | 72,96 | 56,12  |
| AMFCC (Prop1A e Soma, 4Sbs) delta-delta        | 71,94   | 73,98     | 75,00 | 52,04  |
| AMFCC (Prop1A e Soma, 4Sbs) delta, delta-delta | 75,00   | 77,55     | 72,96 | 55,61  |
| 1 GMM  | 70,41   | 76,35     | 73,47 | 27,51  |

Tabela 5.1 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=15dB; com um GMM e sinal de teste sem ruído: 96,43%

O melhor resultado de 78,06% é obtido pelo esquema do espaço nulo, com MFCC e atributos dinâmicos, quando a voz de teste está corrompida por ruído de Fábrica. Quando a voz de teste está contaminada por ruído Falatório, o melhor desempenho de 83,16% também é obtido pelo esquema do espaço nulo com MFCC e atributos dinâmicos. Para o ruído de Carro, o melhor resultado de 79,08% é obtido também pelo esquema do espaço nulo usando MFCC e atributos delta-delta. Por outro lado, para o caso de ruído branco, o melhor resultado de 56,12% é obtido pela técnica do espaço nulo com AMFCC e atributos dinâmicos.

A Tab. 5.2 apresenta os resultados experimentais para os testes usando 5 segundos de voz em 15 dB de RSR.

|  | Fábrica | Falatório | Carro | Branco |
|--|---------|-----------|-------|--------|
| Soma (4Sbs)                                    | 52,55   | 59,69     | 57,65 | 26,02  |
| Prop1A (4Sbs)                                  | 60,71   | 68,37     | 65,31 | 35,71  |
| Prop1A e delta (4Sbs)                          | 64,80   | 68,35     | 64,29 | 36,73  |
| Prop1A e delta-delta (4Sbs)                    | 65,31   | 66,84     | 62,76 | 32,65  |
| Prop1A e delta, delta-delta (4Sbs)             | 65,31   | 69,90     | 66,33 | 33,67  |
| AMFCC (Soma, 4Sbs)                             | 52,04   | 59,18     | 52,55 | 29,59  |
| AMFCC (Prop1A e Soma, 4Sbs)                    | 61,73   | 64,80     | 61,22 | 46,43  |
| AMFCC (Prop1A e Soma, 4Sbs) delta              | 63,78   | 66,84     | 62,76 | 54,59  |
| AMFCC (Prop1A e Soma, 4Sbs) delta-delta        | 60,20   | 63,27     | 63,27 | 50,51  |
| AMFCC (Prop1A e Soma, 4Sbs) delta, delta-delta | 62,76   | 65,82     | 64,29 | 52,55  |
| 1 GMM  | 63,78   | 65,19     | 62,22 | 26,53  |

Tabela 5.2 - Desempenho de identificação (%) em 5s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=15dB

Da Tab. 5.2, pode-se notar que o melhor resultado de 65,31% é obtido pelos esquemas do espaço nulo com MFCC e atributos dinâmicos, quando a voz está corrompida por ruído de Fábrica. Quando a voz de teste está contaminada por ruído Falatório, o melhor resultado de 69,90% é também obtido pelo esquema do espaço nulo com atributos dinâmicos e MFCC. Para o ruído de Carro, o melhor resultado de 66,33% é novamente obtido pelo esquema do espaço nulo com atributos dinâmicos e MFCC. Finalmente, para o caso do ruído branco, o melhor

resultado de 54,59% é obtido pelo esquema do espaço nulo usando os atributos AMFCC com delta.

A Tab. 5.3 apresenta os resultados para os testes com 15 segundos de voz e 10 dB de RSR.

|  | Fábrica | Falatório | Carro | Branco |
|--|---------|-----------|-------|--------|
| Soma (4Sbs)                                    | 34,69   | 69,39     | 43,88 | 17,35  |
| Prop1A (4Sbs)                                  | 48,98   | 80,10     | 58,16 | 25,51  |
| Prop1A e delta (4Sbs)                          | 51,53   | 80,14     | 59,18 | 26,50  |
| Prop1A e delta-delta (4Sbs)                    | 57,14   | 80,10     | 66,33 | 24,49  |
| Prop1A e delta, delta-delta (4Sbs)             | 55,61   | 80,10     | 68,88 | 25,00  |
| AMFCC (Soma, 4Sbs)                             | 46,43   | 66,33     | 50,51 | 19,90  |
| AMFCC (Prop1A e Soma, 4Sbs)                    | 54,59   | 72,45     | 64,29 | 36,73  |
| AMFCC (Prop1A e Soma, 4Sbs) delta              | 56,12   | 76,02     | 63,78 | 41,84  |
| AMFCC (Prop1A e Soma, 4Sbs) delta-delta        | 57,14   | 75,00     | 62,24 | 41,33  |
| AMFCC (Prop1A e Soma, 4Sbs) delta, delta-delta | 58,67   | 75,00     | 63,78 | 44,90  |
| 1 GMM  | 34,18   | 63,27     | 33,67 | 10,20  |

Tabela 5.3 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=10dB

Da Tab. 5.3, verifica-se que a melhor taxa de identificação de 58,67% é obtida pelo esquema do espaço nulo com AMFCC e atributos dinâmicos, quando a voz de teste está com ruído de Fábrica. Para o caso do ruído Falatório, o melhor desempenho de 80,14% é obtido pelo esquema do espaço nulo usando os atributos MFCC e delta. Para o ruído de Carro, o melhor resultado de 68,88% é obtido pelo esquema do espaço nulo usando os atributos MFCC com delta e delta-delta. Finalmente, para o caso do ruído branco, o melhor resultado de 44,90% é obtido pelo esquema do espaço nulo com AMFCC e atributos dinâmicos.

A Tab. 5.4 apresenta as taxas de identificação para os testes usando 5 segundos de voz e 10 dB de RSR.

|  | Fábrica | Falatório | Carro | Branco |
|--|---------|-----------|-------|--------|
| Soma (4Sbs)                                    | 33,80   | 47,45     | 41,02 | 17,23  |
| Prop1A (4Sbs)                                  | 40,51   | 57,14     | 58,03 | 22,45  |
| Prop1A e delta (4Sbs)                          | 42,06   | 57,16     | 57,14 | 22,96  |
| Prop1A e delta-delta (4Sbs)                    | 52,04   | 59,69     | 57,10 | 21,94  |
| Prop1A e delta, delta-delta (4Sbs)             | 53,06   | 59,18     | 58,16 | 22,45  |
| AMFCC (Soma, 4Sbs)                             | 41,33   | 46,43     | 42,35 | 15,82  |
| AMFCC (Prop1A e Soma, 4Sbs)                    | 52,04   | 55,10     | 53,57 | 33,67  |
| AMFCC (Prop1A e Soma, 4Sbs) delta              | 53,06   | 54,08     | 55,10 | 39,29  |
| AMFCC (Prop1A e Soma, 4Sbs) delta-delta        | 47,96   | 55,10     | 57,14 | 39,80  |
| AMFCC (Prop1A e Soma, 4Sbs) delta, delta-delta | 53,57   | 54,59     | 55,61 | 42,86  |
| 1 GMM  | 33,76   | 55,24     | 33,06 | 9,69   |

Tabela 5.4 - Desempenho de identificação (%) em 5s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=10dB

A Tab. 5.4 mostra que o melhor resultado de 53,57% é obtido pelo esquema do espaço nulo usando os atributos AMFCC com delta e delta-delta, quando a voz de teste está corrompida por ruído de Fábrica. Quando a voz de teste está contaminada por ruído Falatório, o melhor resultado de 59,69% é obtido pelo esquema do espaço nulo com MFCC e atributos delta-delta. Para o ruído de Carro, o melhor resultado de 58,16% é também obtido pelo esquema do espaço nulo empregando os MFCC e atributos dinâmicos. Finalmente, para o caso do ruído branco, o melhor resultado de 42,86% é obtido pela técnica do espaço nulo utilizando os atributos AMFCC com delta e delta-delta.

A Tab. 5.5 apresenta os resultados para os testes usando 15 segundos de voz e 5 dB de RSR.

|  | Fábrica | Falatório | Carro | Branco |
|--|---------|-----------|-------|--------|
| Soma (4Sbs)                                    | 19,90   | 29,08     | 40,31 | 13,27  |
| Prop1A (4Sbs)                                  | 29,08   | 42,35     | 51,53 | 16,33  |
| Prop1A e delta (4Sbs)                          | 30,61   | 42,35     | 52,55 | 14,29  |
| Prop1A e delta-delta (4Sbs)                    | 31,12   | 43,37     | 48,98 | 15,82  |
| Prop1A e delta, delta-delta (4Sbs)             | 30,61   | 39,29     | 53,06 | 17,86  |
| AMFCC (Soma, 4Sbs)                             | 18,37   | 28,57     | 33,67 | 12,24  |
| AMFCC (Prop1A e Soma, 4Sbs)                    | 27,04   | 34,18     | 46,94 | 28,06  |
| AMFCC (Prop1A e Soma, 4Sbs) delta              | 27,55   | 34,69     | 45,41 | 34,69  |
| AMFCC (Prop1A e Soma, 4Sbs) delta-delta        | 30,10   | 38,27     | 44,39 | 26,02  |
| AMFCC (Prop1A e Soma, 4Sbs) delta, delta-delta | 28,57   | 34,69     | 47,96 | 34,69  |
| 1 GMM  | 16,84   | 41,33     | 32,14 | 3,06   |

Tabela 5.5 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=5dB

Da Tab. 5.5 observa-se que para ruído de Fábrica o melhor resultado de 31,12% é obtido pelo esquema do espaço nulo usando MFCC com delta-delta. Para ruído Falatório, o melhor resultado de 43,37% é o do esquema do espaço nulo com MFCC e delta-delta. Para ruído de Carro, o melhor resultado de 53,06% é o do espaço nulo com os MFCC e atributos dinâmicos. Já com ruído branco, o melhor resultado, de 34,69%, é o da técnica do espaço nulo com AMFCC e atributos dinâmicos.

|  | Fábrica | Falatório | Carro | Branco |
|--|---------|-----------|-------|--------|
| Soma (4Sbs)                                    | 15,82   | 26,02     | 37,24 | 10,71  |
| Prop1A (4Sbs)                                  | 28,06   | 41,33     | 43,88 | 15,31  |
| Prop1A e delta (4Sbs)                          | 27,04   | 39,80     | 48,98 | 14,29  |
| Prop1A e delta-delta (4Sbs)                    | 23,47   | 39,29     | 43,37 | 14,29  |
| Prop1A e delta, delta-delta (4Sbs)             | 29,08   | 39,25     | 45,41 | 14,80  |
| AMFCC (Soma, 4Sbs)                             | 16,33   | 27,04     | 32,65 | 10,71  |
| AMFCC (Prop1A e Soma, 4Sbs)                    | 26,53   | 34,15     | 38,78 | 25,51  |
| AMFCC (Prop1A e Soma, 4Sbs) delta              | 27,04   | 34,10     | 40,31 | 29,59  |
| AMFCC (Prop1A e Soma, 4Sbs) delta-delta        | 27,55   | 33,67     | 39,29 | 23,98  |
| AMFCC (Prop1A e Soma, 4Sbs) delta, delta-delta | 28,06   | 33,94     | 43,88 | 33,16  |
| 1 GMM  | 16,79   | 39,80     | 32,05 | 3,01   |

Tabela 5.6 - Desempenho de identificação (%) em 5s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=5dB

A partir da Tab. 5.6, pode ser visto que o melhor resultado de 29,082% é obtido pelo esquema do espaço nulo usando os MFCC e atributos dinâmicos, quando a voz de teste está afetada por ruído de Fábrica. No caso do ruído Falatório, o melhor resultado de 41,33% é obtido pelo esquema do espaço nulo com MFCC e sem atributos dinâmicos. No caso do ruído de Carro, o melhor resultado de 48,98% é obtido pelo esquema do espaço nulo empregando os MFCC e atributos dinâmicos. Já para o caso do ruído branco, o melhor resultado de 33,16% é obtido pela técnica do espaço nulo utilizando AMFCC e atributos dinâmicos.

As Tabs. 5.1 a 5.6 mostraram que quando a voz usada no teste está contaminada com ruído branco, o melhor desempenho é obtido pela técnica proposta que utiliza o espaço nulo e os AMFCC com atributos dinâmicos. O ruído branco tende a se concentrar nas autocorrelações de baixos *lags*. Já para ruídos coloridos, as técnicas que empregam os atributos MFCC apresentam geralmente os melhores resultados. Nesse caso, as autocorrelações de baixos *lags* são menos afetadas pelos efeitos dos ruídos coloridos.

## 5.2

### Conclusões

Neste capítulo, foi proposta a utilização dos atributos AMFCC no esquema de múltiplos classificadores empregando a técnica do espaço nulo. Esse tipo de atributo foi inicialmente usado em reconhecimento de voz, porém nesta tese foi examinada a sua utilização em identificação de locutor. Os resultados experimentais mostraram que o uso dos AMFCC com atributos dinâmicos proporciona significativa melhoria na identificação de locutor quando a voz de teste está contaminada por ruído branco.