

3

Proposta do Atributo PAC-MFCC com Classificador em Sub-Bandas para o Reconhecimento de Locutor Independente do Texto em Presença de Ruído

Neste capítulo é proposto o atributo PAC-MFCC com Classificador em Sub-Bandas que combina o atributo PAC-MFCC com um classificador em sub-bandas com a finalidade de melhorar e acrescentar o desempenho da identificação no reconhecimento automático de locutor independente do texto, quando a voz é afetada por ruído com uma relação sinal ruído(SNR) baixa.

Um fato importante que deve ser levado em consideração com referência aos dados utilizados pelo sistema de reconhecimento de locutor é a distribuição não-uniforme das componentes do sinal de voz na frequência [23].

De fato, quando o sinal de voz, é degradado por um ruído seletivo em frequência, uma parte do espectro da voz, pode ainda carregar informações válidas [3].

Uma melhor maneira de aproveitar este fato é o de utilizar classificadores em sub-bandas, ao invés de utilizar toda a faixa de frequências, que dividem o sinal de voz em n sub-bandas de frequências e extraem atributos dos sinais passa-banda.

Os atributos de uma sub-banda são usados como entradas para um classificador aplicado naquela sub-banda. As técnicas que não se baseiam neste método de divisão do sinal em bandas utilizam o esquema de reconhecimento nos atributos extraídos considerando globalmente toda a faixa de frequências do sinal [23], como são as técnicas apresentadas no capítulo 2.

3.1

Fase AutoCorrelação com Coeficientes Cepstrais de Frequência Mel (PAC-MFCC)

O PAC-MFCC é uma técnica derivada da técnica Phase Autocorrelation (PAC) proposta por Ikbal [13] como uma técnica robusta ao ruído acústico. A seguir descreve-se esta técnica proposta para o reconhecimento automático de locutor independente do texto.

3.1.1

Autocorrelação

A extração de atributos em sistemas de reconhecimento de locutor dividem o sinal de voz $s(n)$ em uma sequência de quadros dado por :

$$\{s_0[n], s_1[n], \dots, s_t[n], \dots, s_{T-1}[n]\} \quad (3-1)$$

onde T é número total de quadros, $s_t[n]$ é dado por:

$$s_t[n] = \{s[Kt + 0], s[Kt + 1], \dots, s[Kt + N - 1]\} \quad (3-2)$$

N é o comprimento do quadro e K é o incremento do quadro, vetores são extraídos de cada um desses quadros assumindo que as características do sinal nesses quadros são estacionárias. Características extraídas dos quadros são comumente alguma transformação da magnitude do espectro. A magnitude do espectro pode ser obtida a partir da transformada discreta de Fourier(DFT) das amostras dos quadros e, em seguida tomando a magnitude dos coeficientes resultantes por várias frequências. A DFT assume cada quadro $s_t[n]$ como sendo parte de um sinal periódico $\tilde{s}_t[n]$ definido como:

$$\tilde{s}_t[n] = \sum_{k=-\infty}^{+\infty} s_t[n + kN] \quad (3-3)$$

Como é conhecido, o espectro da magnitude ao quadrado é a DFT da autocorrelação $R[n]$ da sequência periódica \tilde{s}_t sobre o comprimento do quadro; onde a equação da autocorrelação é dada por :

$$R[k] = \sum_{n=0}^{N-1} \tilde{s}_t[n] \tilde{s}_t[n + k], \quad k = 0, 1, \dots, N - 1. \quad (3-4)$$

A autocorrelação remove basicamente as diferenças de fase entre várias com-

ponentes senoidais no sinal de voz da $R[k]$. Outro ponto de vista é que $R[k]$ oferece uma medida da autocorrelação entre o espaço das amostras no intervalo de k , que é calculado como o produto escalar entre dois vetores no espaço dimensional N e dado por [13]:

$$x_0 = \{\tilde{s}_t[0], \tilde{s}_t[1], \dots, \tilde{s}_t[N-1]\} \quad (3-5)$$

$$x_k = \{\tilde{s}_t[k], \dots, \tilde{s}_t[N-1], \tilde{s}_t[0], \dots, \tilde{s}_t[k-1]\} \quad (3-6)$$

onde os coeficientes de autocorrelação são calculados utilizando o produto escalar entre x_0 e x_k , ou seja:

$$R[k] = x_0^T x_k \quad (3-7)$$

3.1.2 Fase AutoCorrelação

Na tentativa de reduzir a sensibilidade dos coeficientes de autocorrelação de ruídos externos presentes no sinal, Ikbal propôs uma nova medida de autocorrelação chamada Fase Autocorrelação(PAC).

Na equação (3-7), se o espaço de amostras no intervalo de k são altamente correlacionados, x_0 será muito próximo a x_k no espaço dimensional N e assim resultará num alto valor do produto escalar [13]. Os dois vetores x_0 e x_k contem os mesmos componentes, então suas magnitudes $\|x\|$ são as mesmas. Se θ_k é o ângulo entre eles no espaço dimensional N , então (3-7) pode ser reescrito como :

$$R[k] = \|x\|^2 \cos(\theta_k) \quad (3-8)$$

Neste método proposto para o calculo da correlação é somente utilizado o ângulo entre os dois vetores, em lugar do produto escalar, como a medida de correlação, resultando num novo conjunto de coeficientes de correlação $P[k]$ definido por :

$$P[k] = \theta_k = \cos^{-1} \left(\frac{R[k]}{\|x\|^2} \right) \quad (3-9)$$

Esta nova medida de correlação é chamada como Fase AutoCorrelação (PAC),

onde o ângulo entre os vetores é utilizado como medida de correlação. A presença de ruído no sinal pode afetar $\|x\|$ e θ_k .

Das equações (3-8) e (3-9) podemos observar que os coeficientes de autocorrelação de $R[k]$ depende de $\|x\|$ e θ_k enquanto os coeficientes PAC $P[k]$ dependem só do θ_k . É exatamente por esse motivo que $P[k]$ é menos suscetível a ruídos externos comparado com $R[k]$ [13].

O coeficientes obtidos do PAC gera um equivalente do espectro chamado espectro PAC o qual é utilizado na obtenção do PAC-MFCC.

A diferença entre o PAC-MFCC e o MFCC se encontra na obtenção do espectro, enquanto o MFCC calcula o espectro por meio da DFT, o PAC-MFCC calcula o espectro utilizando o espectro PAC.

Nas figuras 3.1 e 3.2 mostram-se uma comparação entre os espectros obtidos de utilizar a FFT e o espectro PAC- para um quadro da palavra one. Destas figuras pode-se observar que ambos apresentam um espectro muito similar.

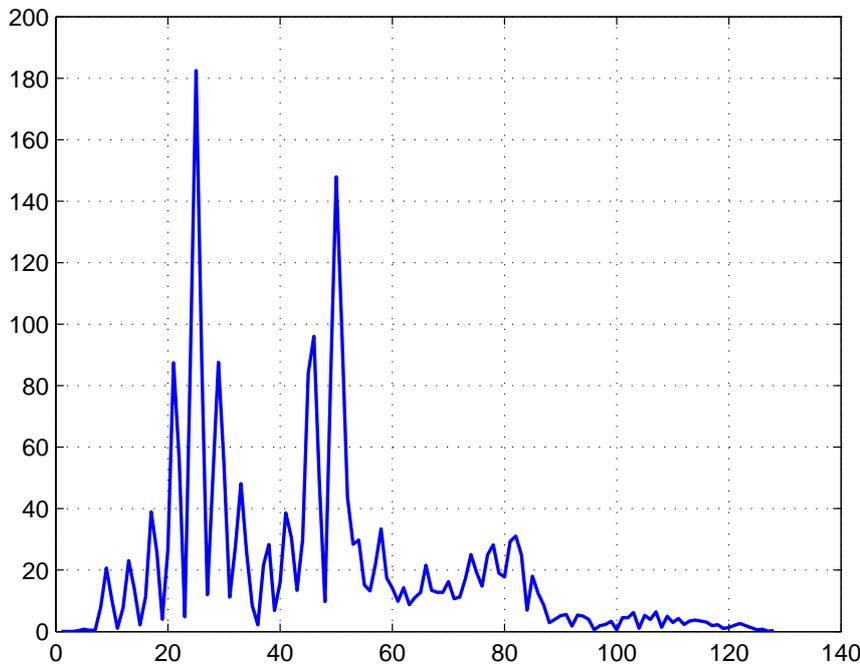


Figura 3.1: FFT de um quadro da palavra "one"

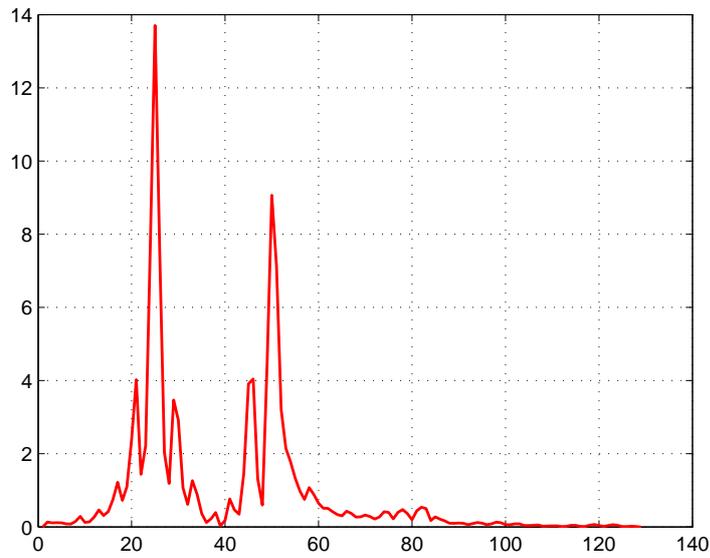


Figura 3.2: Espectro PAC de um quadro da palavra "one"

A figura 3.3 apresenta o diagrama de blocos do atributo PAC-MFCC.

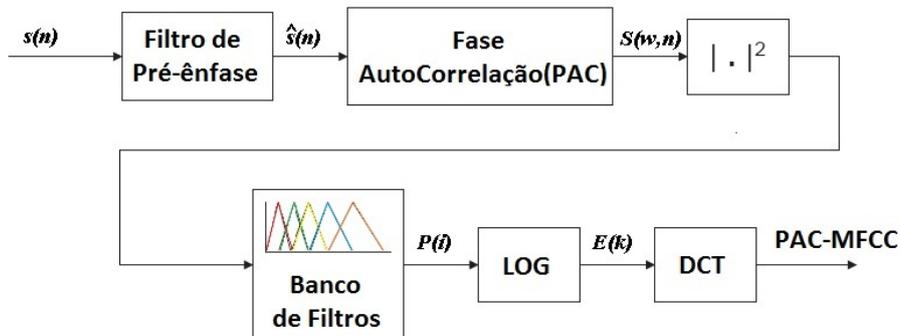


Figura 3.3: Diagrama em blocos do atributo PAC-MFCC

3.2 PAC-MFCC com Classificador em Sub-Bandas

Esta estratégia baseia-se na combinação do atributo PAC-MFCC com um Classificador em sub-bandas, pelo qual depois de ter sido explicado na seção anterior o atributo pac-mfcc ,se explicara a seguir dito atributo.

Esta técnica de reconhecimento de locutor, de forma similar as técnicas baseadas em sub-bandas, como as apresentadas em [12], [21] e [23], utilizam um banco de filtros para dividir o sinal de voz em bandas. Pelo qual para esta técnica foram usados filtros de Butterworth de ordem 6 espaçados na escala Mel. A partir desses filtros os coeficientes PAC-MFCC são extraídos do sinal passa-banda proveniente de cada sub-banda.

Este processo é realizado tanto na fase de projeto quanto na de teste do sistema de reconhecimento. Como o sinal de voz é dividido em n sub-bandas, n classificadores são aplicados a estas bandas, o que significa que cada banda tem seu próprio classificador. Na etapa do teste, as n saídas correspondentes aos n classificadores são somadas gerando uma medida de verossimilhança conjunta (ou resposta conjunta) [23]. Todo locutor modelado possui um esquema como este. O locutor é identificado se seu esquema produzir a maior verossimilhança conjunta comparada com a dos outros locutores modelados pelos classificadores.

No método proposto, o locutor é identificado se o seu sistema de reconhecimento produzir a maior soma de verossimilhanças comparada com a dos outros locutores. Nesta dissertação foi avaliado a técnica PAC-MFCC com Classificador em sub-bandas utilizando 2 e 4 sub-bandas.

Nas figuras 3.4 e 3.5 mostram-se o diagrama em blocos do atributo PAC-MFCC com Classificador em Sub-bandas.

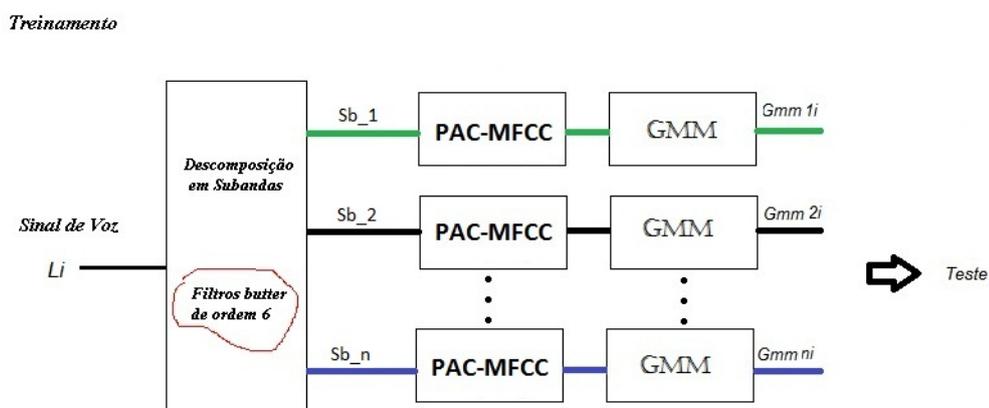


Figura 3.4: Diagrama em blocos da etapa de treinamento do atributo PAC-MFCC com Classificador em Sub-bandas

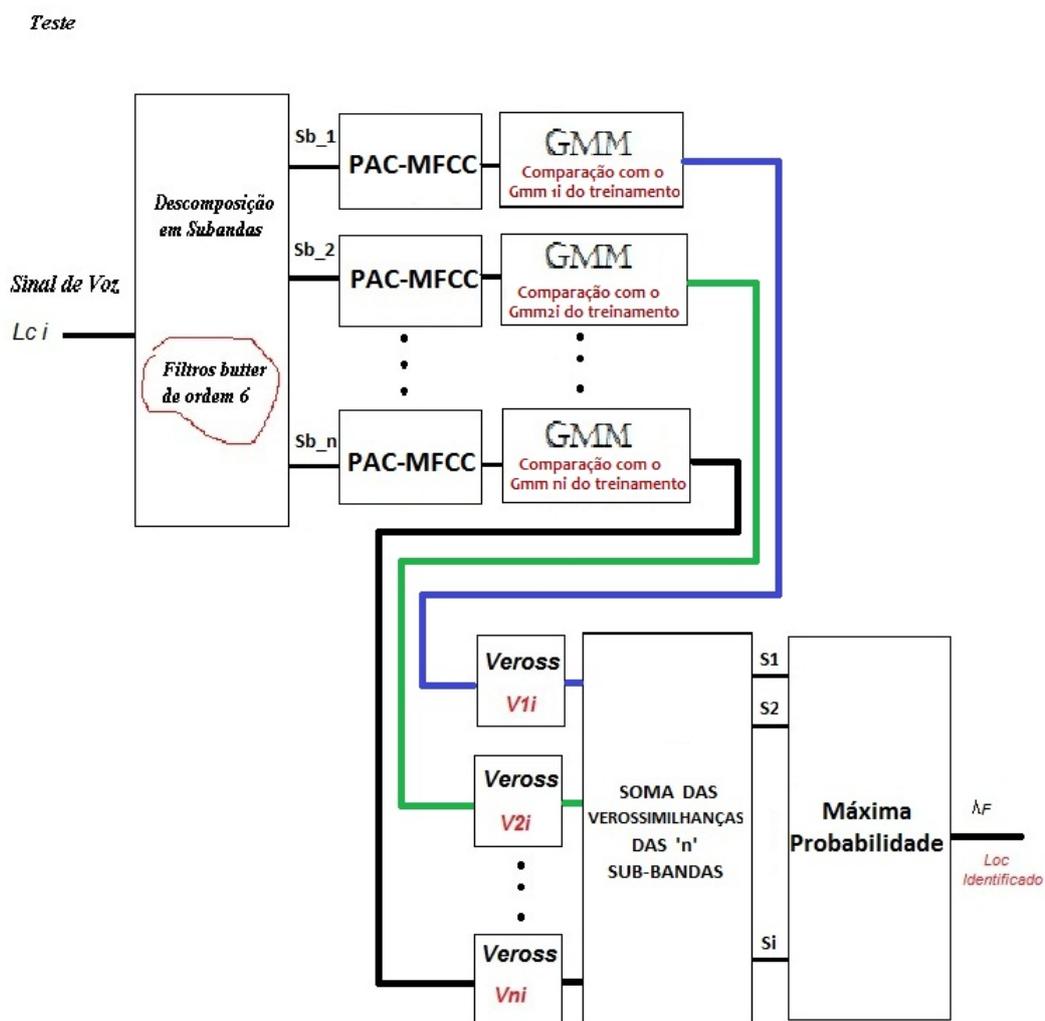


Figura 3.5: Diagrama em blocos da etapa de teste do atributo PAC-MFCC com Classificador em Sub-bandas