4
Técnica de Combinação de Medidas de Verossimilhança Baseada no Espaço Nulo

Neste capítulo, é apresentada uma outra proposta que combina as respostas dos múltiplos classificadores em sub-bandas, cujo principal objetivo é melhorar ainda mais o desempenho do sistema de reconhecimento face a diferentes tipos de degradação, como os ruídos, por meio da escolha de pesos. Os resultados experimentais mostram a eficácia desse método quando comparado com outras técnicas apresentadas na literatura.

As técnicas de combinação dos múltiplos classificadores em sub-bandas empregadas até então utilizam os mesmos pesos (embora diferentes entre si no caso da proposta do Capítulo 3), referentes aos seus respectivos locutores modelados, em todas as vezes que qualquer pretenso locutor testa o sistema, e em qualquer tipo de degradação. Portanto, é de interesse que o conjunto de pesos varie automaticamente, já que as condições de teste se alteram, buscando melhorar o desempenho do sistema de reconhecimento. Nesse sentido, será proposto o emprego do cálculo do espaço nulo para gerar um conjunto de pesos que dependam da informação das energias nas bandas, e que poderão ser alternados quando um pretenso locutor testa o sistema. Esse cálculo permite que as informações das energias nas sub-bandas sejam representadas nesse conjunto de alternativas de pesos, que matematicamente formam uma base para o espaço nulo.

Este capítulo está organizado da seguinte forma: na Seção 4.1 são apresentados os conceitos de espaço nulo; na Seção 4.2 é apresentada a nova proposta que usa o cálculo do espaço nulo para a atribuição de pesos às respostas dos classificadores; na Seção 4.3 é apresentada a aplicação do treinamento em múltiplas condições num esquema de múltiplos classificadores em sub-bandas combinados pela técnica do espaço nulo; na Seção 4.4 é apresentado o uso de atributos dinâmicos no sistema de múltiplos classificadores combinados pela técnica do espaço nulo e na Seção 4.5, as conclusões do capítulo.

4.1
O Espaço Nulo

A álgebra linear fornece uma ferramenta matemática útil para expressar a dependência entre valores, organizados sob a forma de vetores. Por exemplo, dado um conjunto de vetores, o cálculo do espaço nulo [93] desse conjunto permite avaliar a dependência que esses vetores têm entre si. O Espaço nulo (ou *Kernel*) é o conjunto de vetores solução do caso homogêneo

$$\hat{\boldsymbol{A}}\boldsymbol{x} = \boldsymbol{0}, \tag{36}$$

onde $\hat{\boldsymbol{A}}$ é uma matriz de dimensão $m \times n$, 0 representa o vetor de zeros (caso homogêneo) e os vetores solução, representados por $\boldsymbol{x}$ de dimensão n, expressam a dependência entre as colunas de $\hat{\boldsymbol{A}}$. Quando existe pelo menos uma solução além da trivial, $\boldsymbol{x}$=0, as colunas de $\hat{\boldsymbol{A}}$ são linearmente dependentes. Nesse caso, os elementos do vetor $\boldsymbol{x}$ são ponderações aplicadas às colunas de $\hat{\boldsymbol{A}}$ tais que a soma assim ponderada resulte no vetor de zeros, isto é,

$$\sum_{i=1}^{n} x_i col_i = 0 \tag{37}$$

onde x_i representa o i-ésimo elemento do vetor $\boldsymbol{x}$ e col_i a i-ésima coluna de $\hat{\boldsymbol{A}}$. Os vetores múltiplos de $\boldsymbol{x}$ e qualquer soma vetorial envolvendo estes vetores são também soluções de $\hat{\boldsymbol{A}}\boldsymbol{x} = \boldsymbol{0}$.

Uma das maneiras de determinar $\boldsymbol{x}$ é através da eliminação gaussiana [93]. Porém, a maneira computacionalmente mais eficiente é a que utiliza a decomposição SVD (*Singular Value Decomposition*), que decompõe a matriz $\hat{\boldsymbol{A}}$ em duas outras matrizes ortogonais ($\boldsymbol{U}$ e $\boldsymbol{V}$) e em uma matriz diagonal ($\underline{\boldsymbol{S}}$). Essa decomposição pode ser expressa por

$$\hat{\boldsymbol{A}} = \boldsymbol{U}\underline{\boldsymbol{S}}\boldsymbol{V}^T \tag{38}$$

onde $\boldsymbol{V}^T$ é a transposta de $\boldsymbol{V}$, as colunas da matriz $\boldsymbol{U}_{mxm}$ são os auto-vetores de $\widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^T$ e as colunas da matriz $\boldsymbol{V}_{nxn}$ são os auto-vetores de $\widehat{\boldsymbol{A}}^T\widehat{\boldsymbol{A}}$. Além disso, os chamados valores singulares, que são os elementos da diagonal de $\underline{\boldsymbol{S}}$, são as raízes quadradas dos mesmos auto-valores não nulos de $\widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^T$ e de $\widehat{\boldsymbol{A}}^T\widehat{\boldsymbol{A}}$. Os vetores solução $\boldsymbol{x}$ desejados são as colunas de $\boldsymbol{V}$ cujos valores singulares associados são iguais a zero. Esses vetores são uma base ortonormal para o espaço nulo de $\widehat{\boldsymbol{A}}$.

A nova proposta apresentada neste capítulo consiste em usar os elementos dos vetores $\boldsymbol{x}$, obtidos da decomposição SVD, como conjunto de pesos a serem aplicados nas saídas dos classificadores para cada locutor. Esses pesos são calculados utilizando-se os valores das energias dos sinais passa-banda usados na fase de projeto (criação dos modelos) do sistema de reconhecimento de locutor. A Figura 13, abaixo, mostra um diagrama de blocos desse esquema de combinação de medidas de verossimilhança.

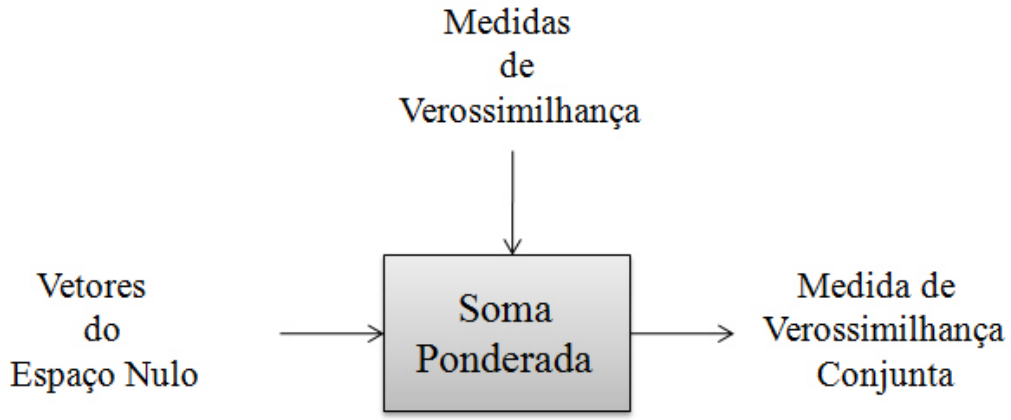

Figura 13 – Esquema de combinação de medidas de verossimilhança

4.2
A Proposta Baseada no Cálculo do Espaço Nulo

Nesta seção, é apresentada a utilização do cálculo do espaço nulo na combinação de classificadores em sub-bandas. Essa nova proposta baseia-se em

representar nas colunas de $\widehat{\boldsymbol{A}}$ as energias de cada sub-banda. Assim fazendo, as colunas de $\widehat{\boldsymbol{A}}$ terão os valores das energias totais, calculadas no domínio do tempo, dos sinais passa-banda usados na fase de projeto do sistema de reconhecimento. Nessa estratégia, para cada locutor, $\widehat{\boldsymbol{A}}$ será computada na fase de treino, e serão calculados pela SVD os seus respectivos vetores de pesos. Esses vetores, obtidos das importantes informações de energia das sub-bandas são armazenados na memória para o posterior uso na fase de teste. Nesse esquema proposto os elementos dos vetores não nulos $\boldsymbol{x}$ são usados como pesos da seguinte forma

$$\underbrace{\begin{matrix} \mathbf{Sb1} & \mathbf{Sb2} & \mathbf{...} & \mathbf{Sbn} \\ \left(\mathbf{E_{T1}} \right. & \mathbf{E_{T2}} & \mathbf{...} & \left. \mathbf{E_{Tn}} \right) \end{matrix}}_{\widehat{A}} \begin{pmatrix} \mathbf{X_1} \\ \mathbf{X_2} \\ \vdots \\ \mathbf{X_n} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \end{pmatrix}$$

Vetor de pesos X (39)

$$\mathbf{Verossimilhança\ Total} = \begin{pmatrix} \mathbf{Veross.\ da\ Sb1} \end{pmatrix} \begin{pmatrix} \mathbf{X_1} \end{pmatrix} + \begin{pmatrix} \mathbf{Veross.\ da\ Sb2} \end{pmatrix} \begin{pmatrix} \mathbf{X_2} \end{pmatrix} + \mathbf{...} + \begin{pmatrix} \mathbf{Veross.\ da\ Sbn} \end{pmatrix} \begin{pmatrix} \mathbf{X_n} \end{pmatrix} \quad (40)$$

para *n* sub-bandas. Nessa ilustração Sb representa uma sub-banda dentre as *n*, E_T representa a energia total de uma sub-banda, dentre as *n*, na fase de projeto, e a verossimilhança total (ou conjunta) é a resultante da combinação das *n* verossimilhanças dos *n* classificadores empregando os *n* elementos de um vetor de pesos **X**. Cada locutor modelado possui um esquema de combinação como esse (pelo cálculo do espaço nulo obtém-se *n*-1 vetores de pesos a serem testados para cada locutor) mostrado em (40), e também os seus respectivos vetores de pesos.

Durante o teste de um segmento de voz, o sistema de reconhecimento que modela o locutor aplica cada uma das suas soluções $\boldsymbol{x}$, visando encontrar a que melhor corrobore para o reconhecimento, ou seja, aquela que produza a maior medida de verossimilhança. Convém observar que, cada uma das soluções $\boldsymbol{x}$ mencionadas constituem matematicamente uma base ortonormal para o espaço nulo de $\hat{\boldsymbol{A}}$, obtida da decomposição SVD dessa matriz de energias.

Sendo $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ e $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ vetores do $\mathbb{R}^n$, o produto interno [93] desses vetores é definido por

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i \ . \tag{41}$$

Portanto, uma interpretação vetorial dos esquemas de combinação permite perceber que os métodos Soma, Energia e os propostos nesse capítulo e no anterior, combinam as saídas dos classificadores fazendo o produto interno dessas saídas com um vetor referência contendo pesos. Quando a regra de combinação é uma simples soma, isto significa que o vetor referência é fixo, com elementos unitários (ou $u_i=1$), para qualquer condição de teste, ou seja,

$$v_{c_s} = \sum_{i=1}^{n} v_i \tag{42}$$

sendo v_{c_s} o valor da medida de verossimilhança conjunta e v_i o valor das medidas de verossimilhanças individuais de cada um dos *n* classificadores. Na apresentação da proposta do capítulo anterior, foi mostrado que a utilização de um vetor de referência, dependente do locutor, pode melhorar o desempenho do reconhecimento. É possível utilizar a expressão [93] do cosseno do ângulo $0 \le \phi \le \pi$ entre dois vetores não nulos do $\mathbb{R}^n$ que é definida por

$$\cos\phi = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \tag{43}$$

para representar a medida de verossimilhança conjunta em função do ângulo entre um vetor de pesos e o vetor das verossimilhanças individuais de cada sub-banda. Em (43) o denominador é o produto das normas dos vetores **u** e **v**. Usando esse conceito, a expressão (42) é representada por

$$v_{c_s} = \|\mathbf{v}\| \cos\phi_{uv} \tag{44}$$

onde ϕ_{uv} é o ângulo entre o vetor de pesos unitários e o vetor de verossimilhanças. Utilizando (43) é possível deduzir uma expressão para a soma ponderada, que resulta na medida de verossimilhança conjunta v_{c_ε} da proposta do capítulo anterior, em função dos valores dos cossenos de dois ângulos:

$$v_{c_\varepsilon} = \frac{\boldsymbol{\varepsilon} \cdot \mathbf{v}}{\boldsymbol{\varepsilon} \cdot \mathbf{u}} = \frac{\|\boldsymbol{\varepsilon}\| \|\mathbf{v}\| \cos\phi_{\varepsilon v}}{\|\boldsymbol{\varepsilon}\| \|\mathbf{u}\| \cos\phi_{\varepsilon u}} \tag{45}$$

sendo $\boldsymbol{\varepsilon}$ o vetor cujos elementos são os pesos obtidos calculando-se as energias das sub-bandas, $\phi_{\varepsilon v}$ o ângulo entre esse vetor e o vetor de verossimilhanças e $\phi_{\varepsilon u}$ o ângulo entre o vetor de pesos e um vetor **u** de componentes unitárias. Note-se que v_{c_ε} é proporcional aos valores dos cossenos, ou seja,

$$v_{c_\varepsilon} \propto \frac{\cos\phi_{\varepsilon v}}{\cos\phi_{\varepsilon u}} . \tag{46}$$

Semelhantemente, é possível deduzir uma expressão para a medida de verossimilhança conjunta v_{c_η} da proposta desse capítulo em função dos valores dos cossenos de dois ângulos, como se segue:

$$v_{c_\eta} = \frac{\boldsymbol{\eta} \cdot \mathbf{v}}{\boldsymbol{\eta} \cdot \mathbf{u}} = \frac{\|\boldsymbol{\eta}\| \|\mathbf{v}\| \cos\phi_{\eta v}}{\|\boldsymbol{\eta}\| \|\mathbf{u}\| \cos\phi_{\eta u}} \tag{47}$$

onde $\boldsymbol{\eta}$ é o vetor de pesos (um dos vetores ***x*** propostos do espaço nulo), $\phi_{\eta v}$ é o ângulo entre esse vetor e o vetor de verossimilhanças e $\phi_{\eta u}$ é o ângulo entre o vetor de pesos e um vetor **u** de componentes unitárias. Note-se que v_{c_η} é proporcional aos valores dos cossenos, ou seja,

$$v_{c_\eta} \propto \frac{\cos\phi_{\eta v}}{\cos\phi_{\eta u}}. \tag{48}$$

Na seção seguinte, será mostrado através de simulações, que o emprego de um conjunto de vetores ***x*** de referência associado a cada locutor, obtido pelo cálculo do espaço nulo, pode melhorar ainda mais o desempenho dos sistemas de reconhecimento em face de ruído.

Nas simulações com GMM, v apresenta componentes negativas. O vetor $\boldsymbol{\varepsilon}$ apresenta componentes positivas, já que é de energias, e $\boldsymbol{\eta}$ apresenta componentes positivas e negativas. Assim, em (44) tem-se $\phi_{\eta v} > \frac{\pi}{2}$, tornando v_{c_s} negativo. Além disso, em (45) tem-se $\phi_{\varepsilon u} < \frac{\pi}{2}$ e $\phi_{\varepsilon v} > \frac{\pi}{2}$, tornando v_{c_ε} negativo. O vetor **v** tende a ter as menores componentes, em magnitude, nas direções em que o vetor $\boldsymbol{\varepsilon}$ tiver as maiores em magnitude. As sub-bandas mais afetadas por ruídos são as que mais contribuem em magnitude para as componentes de **v**. Assim, **v** tende a ser perpendicular a $\boldsymbol{\varepsilon}$, devido ao ruído. Além disso, por definição [93], os vetores ***x*** (vetores do espaço nulo usados na proposta) são ortonormais entre si e são perpendiculares ao vetor de energias. Desse modo, em (47) o ângulo $\phi_{\eta v}$ tende a ser menor que $\frac{\pi}{2}$ e $\phi_{\eta u}$ tende a ser maior que $\frac{\pi}{2}$, tornando v_{c_η} negativo.

4.2.1
Resultados de Simulação

Os resultados experimentais estão apresentados nas Figuras 14 a 21 visando mostrar o desempenho da nova proposta baseada no espaço nulo ("Prop1A") quando comparada com os outros métodos de combinação. As mesmas condições experimentais usadas no capítulo anterior são utilizadas nesta seção.

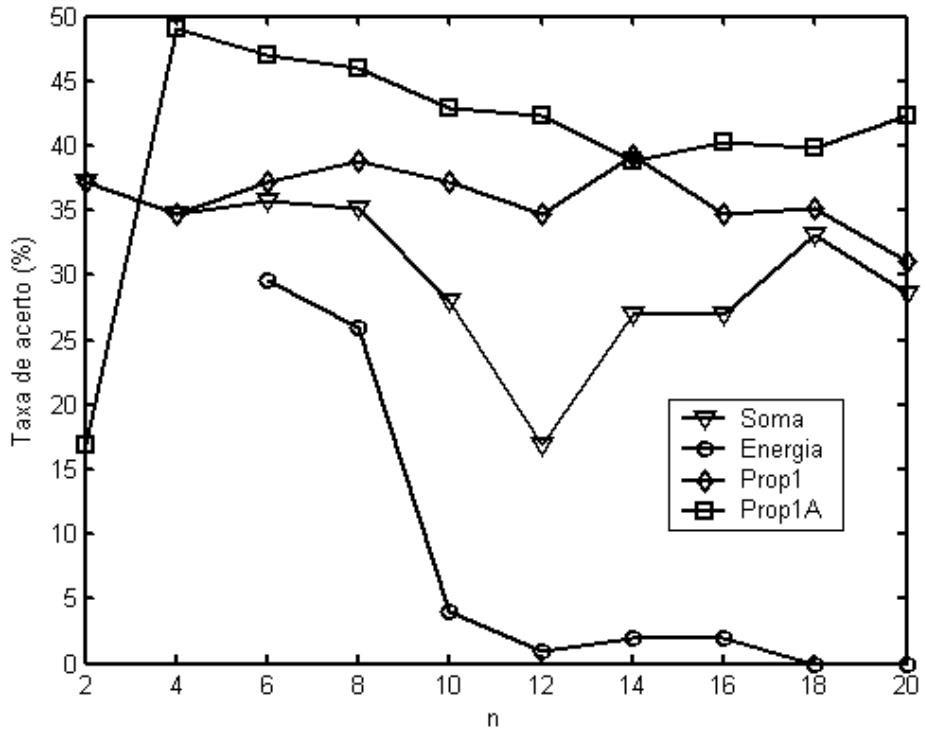

Figura 14 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica em RSR=10dB; com um GMM: 34,18%

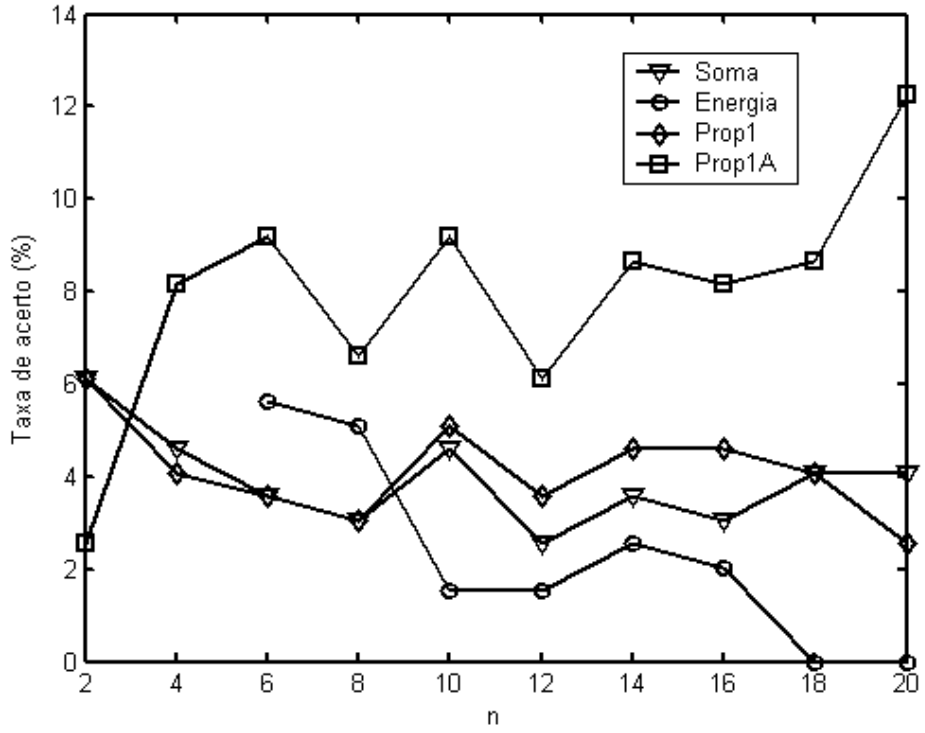

Figura 15 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica em RSR=0dB; com um GMM: 3,57%

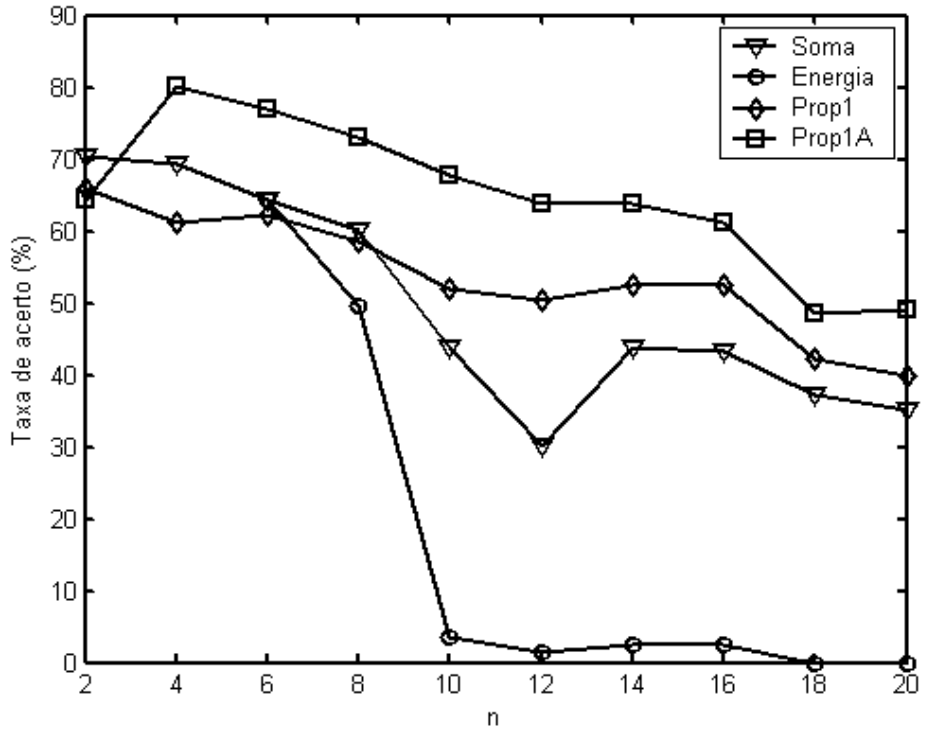

Figura 16 - Desempenho de identificação (%) em 15s de teste para ruído de Falatório em RSR=10dB; com um GMM: 63,27%

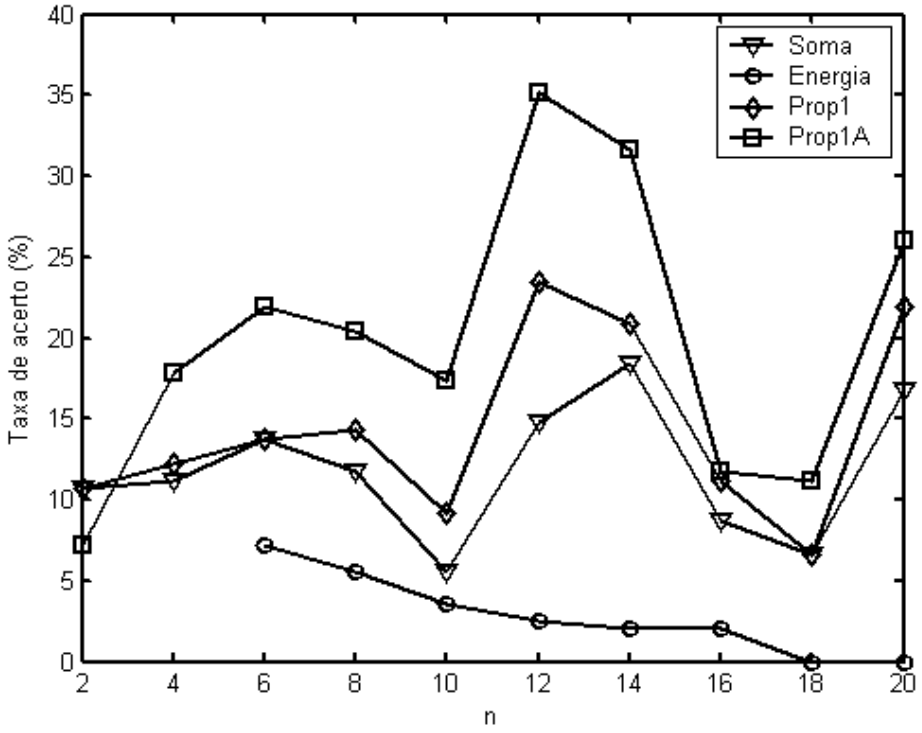

Figura 17 - Desempenho de identificação (%) em 15s de teste para ruído de Falatório em RSR=0dB; com um GMM: 10,71%

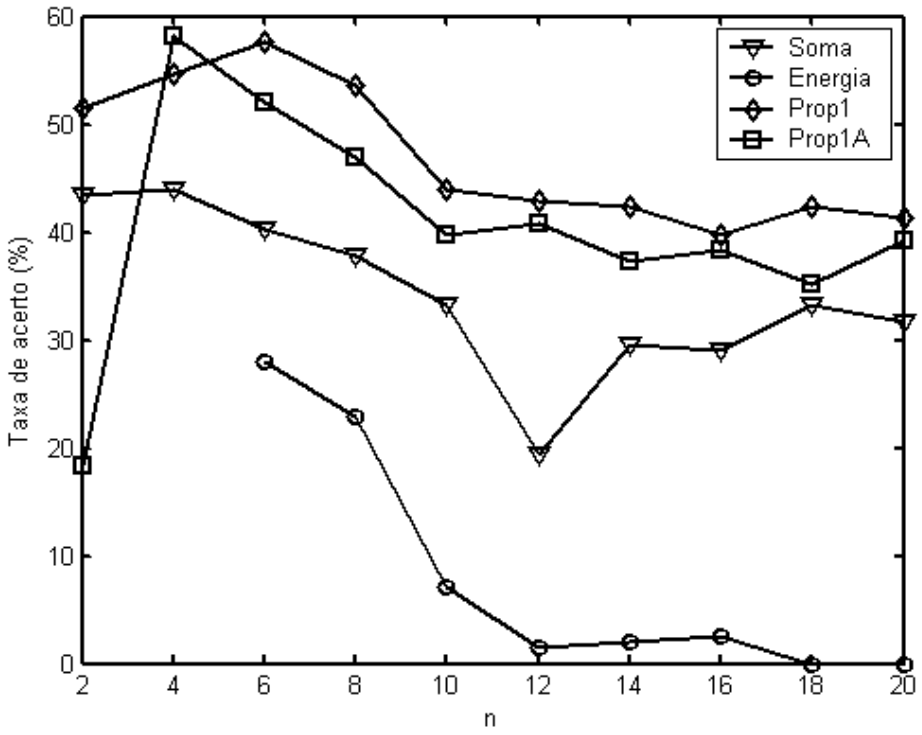

Figura 18 - Desempenho de identificação (%) em 15s de teste para ruído de Carro em RSR=10dB; com um GMM: 33,67%

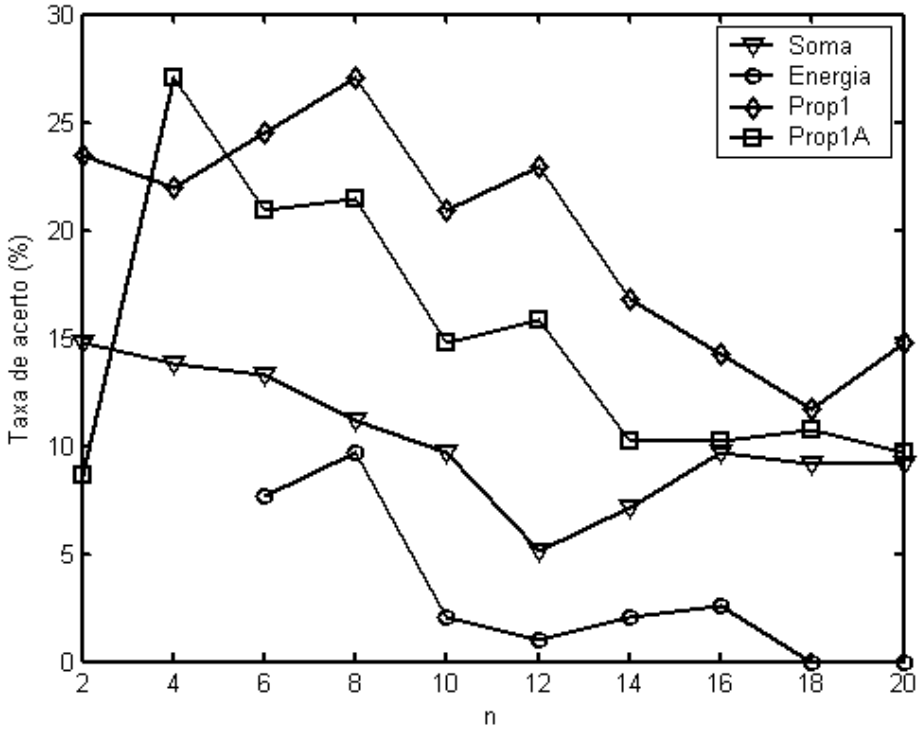

Figura 19 - Desempenho de identificação (%) em 15s de teste para ruído de Carro em RSR=0dB; com um GMM: 13,78%

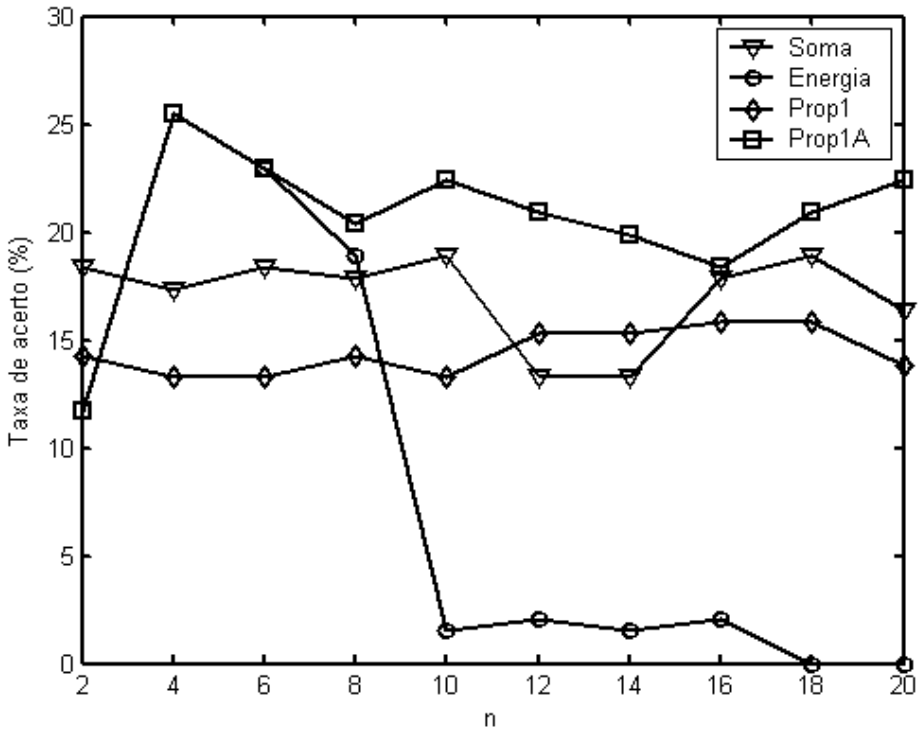

Figura 20 - Desempenho de identificação (%) em 15s de teste para ruído Branco em RSR=10dB; com um GMM: 10,20%

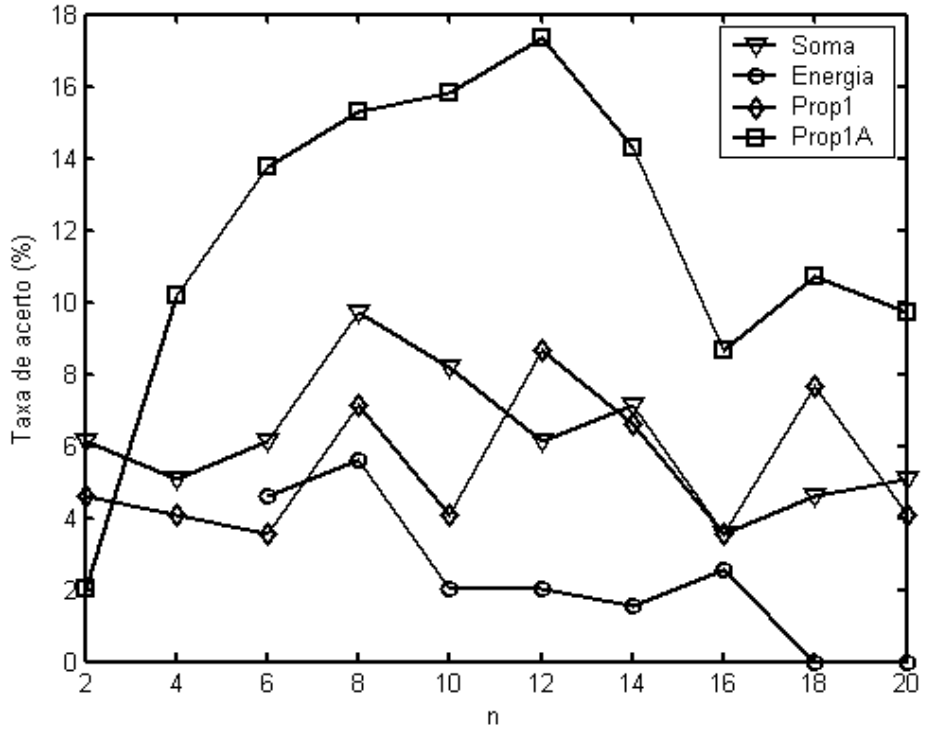

Figura 21 - Desempenho de identificação (%) em 15s de teste para ruído Branco em RSR=0dB; com um GMM: 2,55%

A Figura 14 mostra os resultados experimentais da técnica apresentada em [38], designada por "Soma", da apresentada em [39], designada por "Energia", da proposta do capítulo anterior (Prop1), e da nova proposta baseada no espaço nulo (Prop1A) usando 2, 4, 6, 8, 10, 12, 14, 16, 18 ou 20 sub-bandas e voz corrompida por ruído de Fábrica com 10dB de RSR. A Figura 15 mostra os resultados das mesmas técnicas para voz com ruído de Fábrica em 0 dB. As Figuras 16 e 17 apresentam os resultados dos métodos para voz corrompida por ruído Falatório com 10dB e 0dB, respectivamente. Nas Figuras 18 e 19 são mostrados os resultados para a voz com ruído de Carro em 10dB e 0dB. Finalmente, as Figuras 20 e 21 apresentam os resultados para voz corrompida por ruído gaussiano Branco com 10dB e 0dB, respectivamente.

O melhor resultado da identificação (48,98%) para RSR=10dB e ruído de Fábrica, de acordo com a Figura 14, foi obtido pela técnica do espaço nulo proposta usando 4 sub-bandas. Para a maioria dos casos, esse esquema proposto teve um desempenho melhor do que os demais, para o mesmo número de sub-bandas usadas na decomposição do sinal. Quando a RSR é 0dB, como pode ser

visto na Figura 15, o melhor desempenho (12,24%) foi obtido usando 20 sub-bandas e a técnica do espaço nulo proposta, enquanto o sistema com um GMM alcançou 3,57%. Quando o sinal de voz é corrompido com ruído Falatório, o esquema com um GMM gerou um desempenho de 63,27% em 10dB e o esquema com o espaço nulo forneceu 80,10% com 4 sub-bandas, sendo esse o melhor resultado obtido para esse tipo de ruído nessa RSR, como podemos observar da Figura 16. Adicionalmente, quando a RSR é igual a 0dB, o melhor desempenho é obtido usando-se o método do espaço nulo proposto (35,20%) como visto na Figura 17. No caso do ruído de Carro (10dB), mostrado na Figura 18, mais uma vez o melhor resultado (58,16%) é obtido com o esquema do espaço nulo. O sistema com um GMM alcançou 33,67%. Pode ser visto na Figura 19 que para 0dB de RSR as duas técnicas propostas forneceram o melhor resultado (27,04%). Por outro lado, a técnica baseada no espaço nulo obteve esse desempenho usando apenas a metade do número de sub-bandas que a utilizada pela proposta apresentada no capítulo anterior.

Quando é considerado o ruído Branco (10dB), mostrado na Figura 20, o melhor desempenho foi novamente obtido pela proposta do espaço nulo (25,51%). Para 0dB de RSR, mostrado na Figura 21, o melhor resultado também foi alcançado pela técnica do espaço nulo (17,35%). Portanto, ressalta-se que o esquema do espaço nulo proposto é a melhor estratégia para ruído Branco, ao contrário do método proposto no capítulo anterior. Note-se, ainda, que em cinco dessas oito figuras, a técnica do espaço nulo obteve o melhor resultado utilizando apenas quatro sub-bandas. Nesses casos, isso é possivelmente devido à redução do número de sub-bandas que não contribuem para a identificação do locutor. Além disso, na maioria dos casos, a nova proposta supera em desempenho as demais para o mesmo número de sub-bandas usadas na decomposição dos sinais. A alternância dos vetores de ponderações na fase de teste permite encontrar aquela representação de pesos dependente do locutor (expressa pelas componentes do vetor X) que melhor combine as informações da identidade do locutor nas sub-bandas. Isso tem a tendência de reduzir o efeito do ruído e de aumentar a taxa de acerto. Essa nova estratégia tende a reduzir o esforço computacional e o uso de memória, comparativamente a outras técnicas simuladas neste capítulo, já que ela tende a apresentar os melhores resultados utilizando apenas quatro sub-bandas.

4.3
Combinação de Classificadores em Sub-bandas Usando o Espaço Nulo e Treinamento Com Múltiplas Condições

Esta Seção é dedicada à apresentação de uma nova proposta que combina a técnica do espaço nulo e o treinamento em múltiplas condições. Essa estratégia visa melhorar o reconhecimento da proposta que utiliza o espaço nulo, através de uma tentativa de compensar os modelos dos locutores para efeitos de ruídos.

O treinamento baseado em múltiplas condições consiste em treinar o classificador a partir de voz contaminada por ruído. Essa estratégia permite realizar uma compensação no modelo do locutor gerado pelo classificador na fase de projeto, para o efeito do ruído [94], com o objetivo de aprimorar as taxas de reconhecimento. Nesse esquema, o sistema de reconhecimento tira proveito das componentes espectrais com melhor casamento [95], entre o treino e o teste.

Uma técnica recente, apresentada em [94]-[96], emprega o treinamento em múltiplas condições e a exclusão de atributo. Nessa abordagem, um classificador é treinado com voz sem ruído e com voz contaminada por ruído branco para vários níveis de RSR (10, 12, 14, 16, 18, 20dB). Um banco de filtros Mel é utilizado na extração dos atributos. As saídas dos filtros Mel são descorrelatadas por um filtro passa-altas ($H(z)=1-z^{-1}$), cujas saídas correspondem às contribuições das sub-bandas empregadas nessa abordagem. No teste de cada janela de voz, são feitas todas as combinações (soma duas-a-duas, três-a-três, quatro-a-quatro, dentre outras) de verossimilhanças associadas a cada uma das sub-bandas e produzidas por um classificador GMM. A resposta do classificador para a janela testada é a medida de verossimilhança resultante da combinação que fornecer o maior resultado. Note-se que as sub-bandas que contribuem menos para o reconhecimento são simplesmente excluídas no teste. Esse método é, entretanto, de alta complexidade computacional.

A proposta desta Seção consiste em empregar o treinamento em múltiplas condições nos classificadores em sub-bandas que utilizam o espaço nulo na combinação das respostas. Nesse esquema, o sinal de treino em cada valor de RSR é usado para treinar um sistema de classificadores em sub-bandas que empregam o espaço nulo. Além disso, é treinado mais um desses sistemas usando voz sem ruído. Então, considerando os valores de RSR para o treinamento como sendo 10, 12, 14, 16, 18, 20dB, cada locutor será modelado por sete sistemas de classificadores em sub-bandas empregando o espaço nulo, como mostrado na Figura 22. O espaço nulo é obtido do sinal de treino sem ruído. A resposta conjunta final é a do sistema que produzir o maior valor de verossimilhança de saída.

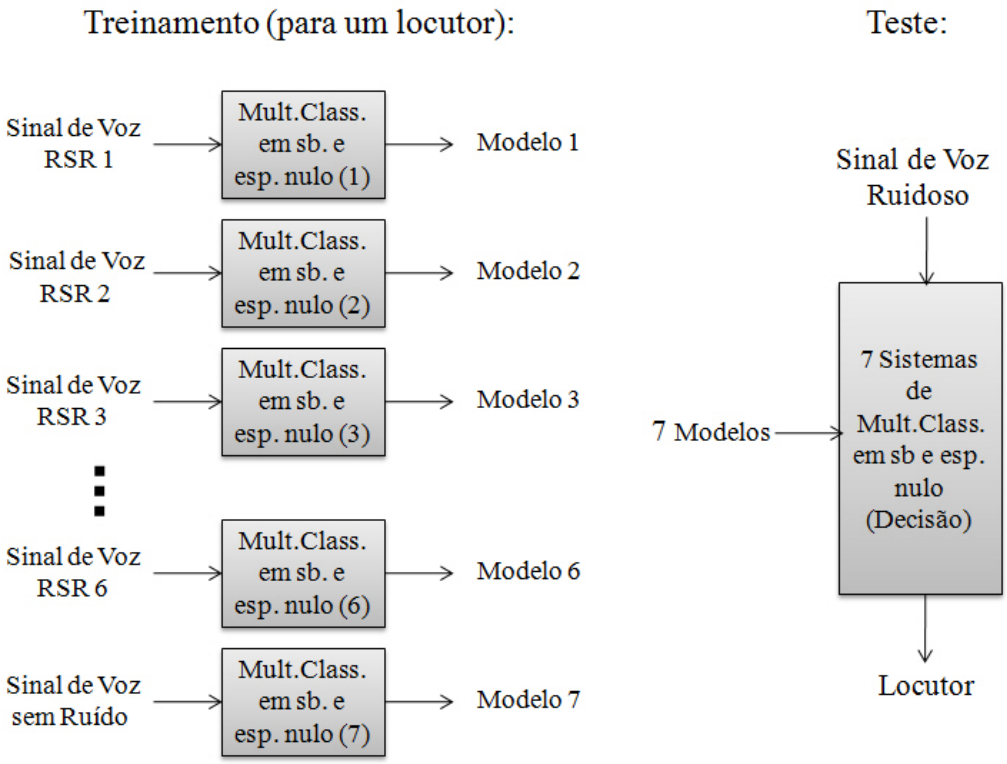

Figura 22 – Esquema de combinação utilizando o treinamento com múltiplas condições

Ressalta-se que nesta proposta nenhuma sub-banda é desprezada, e não são feitas quaisquer combinações de medidas de verossimilhanças, reduzindo-se bastante a complexidade em relação ao esquema apresentado em [94]-[96].

4.3.1
Resultados de Simulação

Resultados de simulação são apresentados para a identificação de locutor independente do texto, com o objetivo de mostrar o desempenho da nova proposta quando comparada com outras técnicas. As condições experimentais são as mesmas empregadas anteriormente, com a diferença de que no treino, além dos sinais limpos, são empregados sinais contaminados por ruído gaussiano branco, para valores de RSR: 10, 12, 14, 16, 18 e 20dB. Os sinais das duas sessões restantes da base KING, compostas por quatro segmentos de 15 segundos (para cada locutor e sem silêncio), são corrompidos por ruído com 10dB de RSR e utilizados para o teste. A proposta aqui apresentada é denominada "CRSR Esp. Nulo". O desempenho da identificação usando apenas um GMM (sem decomposição em sub-bandas) é de 96,43% para o teste com voz sem ruído. Esse desempenho cai severamente (10,2%) em ambientes ruidosos, como será apresentado a seguir.

Conforme mostrado na Tab. 4.1, quando a voz de teste está contaminada por ruído gaussiano branco, a técnica que apresenta os melhores resultados (72,45% usando 6 sub-bandas, 73,47% usando 4 sub-bandas) é a que emprega a combinação do espaço nulo com o treinamento em múltiplas condições. Esses resultados são bem superiores ao obtido com espaço nulo sem múltiplas condições (25,51%). A proposta também apresenta os melhores resultados (58,67% usando 6 sub-bandas, 60,71% usando 4 sub-bandas) quando o sinal de teste está contaminado com ruído de carro.

Em % acerto	Fábri-ca	Fala-tório	Carro	Branco
Soma(4Sbs)[38]	34,69	69,39	43,88	17,35
Prop1 (4Sbs)	34,69	61,22	54,59	13,27
Prop1A (4Sbs)	48,98	80,10	58,16	25,51
CRSR Esp. Nulo (6Sbs)	42,86	67,86	58,67	72,45
CRSR Esp. Nulo (4Sbs)	46,43	63,78	60,71	73,47
1GMM	34,18	63,27	33,67	10,20

Tabela 4.1 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=10dB; para um GMM e sinal de teste sem ruído: 96,43%

Os resultados obtidos indicam que acrescentar no esquema de combinação das respostas dos classificadores e na geração de modelos, as informações relacionadas ao comportamento espectral do sinal ruidoso, ou seja, a distribuição do ruído (treinamento com múltiplas condições) e das contribuições do sinal de voz na frequência (espaço nulo), pode melhorar muito o desempenho do reconhecimento para ruídos de teste semelhantes aos usados no treino. Na abordagem dessa tese, isso foi feito através da utilização de uma base para o espaço nulo obtida das energias das sub-bandas dos sinais sem ruído, e pelo treinamento em múltiplas condições usando ruído branco. No caso do ruído colorido é necessário que seja feita uma melhor modelagem de sua distribuição em frequência para que o treinamento em múltiplas condições ajude a melhorar o desempenho. Ou seja, o ruído a ser usado no treino deve ter comportamento espectral o mais semelhante possível do ruído presente no sinal que testará o sistema, de forma a melhorar o casamento entre o modelo treinado e o sinal de teste. Apesar do ruído branco não ser semelhante em todo o espectro ao de carro, ele contribui favoravelmente, dentro das sub-bandas (naquelas mais importantes para a identificação do locutor), para melhorar o reconhecimento.

4.4
Combinação de Classificadores em Sub-bandas Usando o Espaço Nulo e Atributos Dinâmicos

Nesta Seção, é apresentada uma nova proposta que consiste em empregar os atributos dinâmicos no esquema de reconhecimento baseado no espaço nulo. Essa estratégia visa observar se os atributos delta e delta-delta, extraídos em cada sub-banda, podem contribuir para a melhoria do desempenho quando é usado o método de combinação que baseia-se no espaço nulo.

Os atributos delta e delta-delta complementam a informação instantânea ou estática obtida dos atributos MFCC. Além disso, esses atributos dinâmicos serão empregados visando remover a informação local invariante no tempo proveniente do ruído. A estratégia consiste em alocar em um mesmo vetor de características, os atributos MFCC e seus respectivos atributos dinâmicos (delta

ou delta-delta, ou delta e delta-delta). Isso será feito para cada vetor de atributos extraídos do sinal de voz.

4.4.1
Resultados de Simulação

Nessa Seção, são apresentados resultados experimentais de identificação de locutor independente do texto com o objetivo de mostrar o comportamento do esquema que emprega os atributos dinâmicos, quando comparado com outros métodos. As mesmas condições experimentais das simulações anteriores foram aqui utilizadas. Adicionalmente, foi feito o reconhecimento utilizando-se quatro segmentos de 5 segundos de sinal de fala, com o objetivo de observar o comportamento dos esquemas nessa condição. Foram usados 20 atributos MFCC (com seus 20 respectivos atributos delta ou delta-delta, ou delta e delta-delta, quando for desejado incluir atributos dinâmicos), extraídos de janelas de 20 ms de voz (usando janela de Hamming com superposição de 50%).

Os resultados experimentais expressos em termos de taxa de reconhecimento estão apresentados nas Tabs. 4.2 a 4.5. Nessas tabelas, os esquemas que empregam os atributos dinâmicos estão representados por "Prop1A e delta", "Prop1A e delta-delta", e "Prop1A e delta, delta-delta". A Tab. 4.2 apresenta as taxas de identificação para os testes usando segmentos de 15 segundos de sinal de fala e 15 dB de RSR.

	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
1 GMM	70,41	76,35	73,47	27,51

Tabela 4.2 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=15dB; para um GMM e sinal de teste sem ruído: 96,43%

O melhor resultado de 78,06% foi obtido pelo esquema do espaço nulo com atributos dinâmicos delta e delta-delta, quando a voz de teste está corrompida

por ruído de Fábrica. Quando a voz de teste está corrompida por ruído Falatório, o melhor desempenho de 83,16% também é obtido pelo esquema do espaço nulo utilizando os atributos dinâmicos delta e delta-delta. Para o ruído de Carro, o melhor resultado de 79,08% é obtido pelo esquema proposto com atributos delta-delta. Finalmente, para o caso de ruído branco, o melhor resultado de 40,82% é obtido pelas técnicas do espaço nulo com e sem atributos dinâmicos.

A Tab. 4.3 apresenta as taxas de reconhecimento para os testes usando 5 segundos de voz com 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
1 GMM	63,78	65,19	62,22	26,53

Tabela 4.3 - Desempenho de identificação (%) em 5s de teste para ruído de Fábrica, Falatório, Carro e Branco com RSR=15dB

Da Tab. 4.3 é observado que o melhor resultado de 65,31% é obtido pelos esquemas propostos utilizando atributos dinâmicos, quando a voz está corrompida por ruído de Fábrica. Quando a voz de teste está contaminada por ruído Falatório, o melhor desempenho de 69,90% é também obtido pelo esquema do espaço nulo com atributos dinâmicos. Para o ruído de Carro, o melhor resultado de 66,33% é novamente obtido pelo esquema do espaço nulo com atributos dinâmicos. Finalmente, para o caso do ruído branco, o melhor resultado de 36,73% é obtido pelo esquema proposto usando os atributos delta.

A Tab. 4.4 apresenta as taxas de reconhecimento 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
1 GMM	34,18	63,27	33,67	10,20

Tabela 4.4 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco com RSR=10dB

Nota-se da Tab. 4.4, que a melhor taxa de reconhecimento de 57,14% é obtida pelo esquema proposto usando os atributos delta-delta, quando a voz de teste está contaminada por ruído de Fábrica. Quando a voz de teste está contaminada por ruído Falatório, o melhor desempenho de 80,14% é obtido pelo esquema proposto usando os atributos delta. Para o ruído de Carro, o melhor resultado de 68,88% é conseguido pelo esquema proposto utilizando os atributos delta e delta-delta. Finalmente, para o caso do ruído branco, o melhor resultado de 26,50% é obtido pelo esquema proposto usando os atributos delta.

A Tab. 4.5 apresenta a taxa de reconhecimento 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
1 GMM	33,76	55,24	33,06	9,69

Tabela 4.5 - Desempenho de identificação (%) em 5s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=10dB

Da Tab. 4.5 é verificado que o melhor resultado de 53,06% é obtido pelo esquema proposto usando os atributos dinâmicos, quando a voz de teste está contaminada por ruído de Fábrica. Quando a voz de teste está contaminada por ruído Falatório, o melhor desempenho de 59,69% é conseguido pelo esquema proposto usando os atributos delta-delta. Para o ruído de Carro, o melhor resultado de 58,16% é também obtido pelo esquema do espaço nulo empregando atributos dinâmicos. Finalmente, para o caso do ruído branco, o melhor resultado de 22,96% é conseguido pela técnica proposta utilizando os atributos delta.

As Tabs. 4.2 a 4.5 mostraram que a técnica que utiliza os atributos dinâmicos fornece os melhores resultados em todos os casos. Particularmente, para 15 dB, o melhor desempenho, na maioria dos casos, é obtido quando os atributos delta e delta-delta estão ambos presentes no mesmo vetor de atributos. Porém, isso não ocorre para o caso de 10 dB.

Note-se que, em algumas situações, os atributos dinâmicos não melhoram o reconhecimento, como, por exemplo, em [77], [97]. Porém, quando o sinal de teste é menos afetado por ruído, a inclusão de vários atributos dinâmicos pode contribuir para aumentar o desempenho do sistema devido às informações dinâmicas adicionais. Em ambientes sem ruído, os atributos delta e delta-delta são usualmente mais adequados para identificação de locutor dependente do texto e para aplicações que requeiram a redução do descasamento provocado por efeitos do canal de comunicações [6]. Além disso, pode ser visto que quando a voz de teste está severamente contaminada por ruído (RSR=10dB), a contribuição devido ao MFCC fica bem pequena. Por outro lado, a contribuição devido à inclusão dos atributos dinâmicos tende a aumentar o desempenho do reconhecimento. O ruído penaliza mais a contribuição dos atributos MFCC (que não é robusto) do que a dos atributos dinâmicos.

4.5
Conclusões

Neste capítulo, foi apresentada uma nova proposta, baseada no espaço nulo, para a combinação das respostas dos classificadores em sub-bandas. As simulações mostraram que essa estratégia proposta fornece melhores resultados que as demais face a diferentes tipos de ruído, inclusive no caso do ruído branco. Também foi apresentada uma abordagem para reconhecimento de locutor independente do texto, empregando o espaço nulo na combinação das respostas dos classificadores em sub-bandas e o treinamento em múltiplas condições. Mostrou-se que a técnica baseada no espaço nulo pode tirar proveito da compensação realizada pelo treinamento em múltiplas condições, no sentido de melhorar o desempenho do reconhecimento quando os sinais de treino estão contaminados com ruído branco e os de teste com ruído branco ou de carro. Os resultados reforçam a idéia de que considerar no esquema de combinação das respostas dos classificadores e na geração de modelos, as informações relacionadas ao comportamento espectral do sinal, pode contribuir significativamente para melhorar o desempenho do sistema de reconhecimento.

Para os testes com os demais tipos de ruído colorido, necessita-se que seja feita uma melhor escolha do tipo (ou tipos) de ruído usado para treinar o sistema.

Este capítulo também propôs a inclusão de atributos dinâmicos (delta e delta-delta) nos vetores de entrada dos classificadores em sub-bandas, com o propósito de observar como o desempenho das técnicas de reconhecimento, usando o espaço nulo, podem ser melhoradas na identificação de locutor independente do texto, em ambientes ruidosos. Foram feitos experimentos com atributos dinâmicos extraídos em sub-bandas de frequências e aplicados ao sistema de múltiplos classificadores em sub-bandas que usa o espaço nulo. Os resultados obtidos mostram que a inclusão de atributos dinâmicos é capaz de aumentar o desempenho do sistema de reconhecimento.

Vários resultados experimentais de identificação de locutor independente do texto foram aqui apresentados, usando uma base de vozes bem extensa e diferentes tipos de ruído ambiente, com o propósito de verificar o desempenho dos sistemas de reconhecimento.