Seleção de Atributos e Classificação Fonética de Sinais de Fala de Banda Limitada

Ênio dos Santos Silva e Rui Seara

***Resumo*— Este trabalho apresenta uma estratégia para seleção de atributos aplicada à classificação fonética de sinais de fala de banda limitada. Particularmente, a estratégia proposta visa o aprimoramento da classificação fonética em grupos fonéticos abrangentes (*broad group phonetic* - BGP) com ênfase na discriminação de fonemas pertencentes ao grupo fonético fricativo. Especificamente, os atributos característicos de sinais de fala são investigados através de métodos baseados em aprendizagem de máquinas e análise de componentes, tais como, informação mútua (*mutual information* - MI) e seleção de parâmetros baseada em correlação (*correlation-based feature selection* - CFS). Nesse contexto, diversos algoritmos do estado-da-arte são investigados, resultando em uma estratégia que consiste em identificar os atributos ótimos do sinal de fala de banda limitada que facilitem a distinção entre diferentes BGP. Os resultados de simulações permitem inferir acerca da seleção dos melhores atributos a serem considerados na composição dos classificadores fonéticos, comprovando a eficácia da estratégia proposta.**

***Palavras-Chave*— Classificação fonética, mineração de dados, reconhecimento de padrões, seleção de atributos.**

***Abstract*— This work presents a feature selection strategy applied to phonetic classification of band-limited speech signals. Particularly, the proposed strategy aims to improve the classification of broad group phonetic (BGP) with an emphasis on discrimination of phonemes that are relevant to the fricative phonetic group. Specifically, the features of the speech signal are investigated by using methods based on machine learning algorithms and component analysis, such as mutual information (MI) and correlation-based feature selection (CFS). Several state-of-the-art classifiers are investigated, resulting in a strategy that allows identifying the optimal band-limited speech signal features aiming to make easy the distinction between different BGP. The simulation results allow infering about the selection of the best features to be used in the composition of phonetic classifiers, confirming the effectiveness of the proposed strategy.**

***Keywords*— Phonetic classification, data mining, pattern recognition, feature selection.**

I. Introdução

A atividade de classificação de padrões é um processo inerente ao ser humano. Para distinguir diferentes objetos, usualmente, o ser humano armazena parâmetros marcantes, aqui chamados atributos característicos, de um determinado conjunto de objetos [1]. Nesse contexto, os métodos de seleção de atributos buscam identificar e reter apenas os atributos que mais contribuem para a discriminação em um determinado conjunto de objetos [1]–[3]

Na literatura, diversas estratégias de seleção de atributos vêm sendo propostas. Em [4], a seleção de atributos, provenientes da predição linear perceptual (*perceptual linear predictive* - PLP) e dos coeficientes cepstrais na escala mel (*mel-frequency cepstrum coefficients* - MFCC), tem como base o algoritmo *AdaBoost* [3]. Tais atributos são aplicados a sistemas de reconhecimento automático de fala baseados em classificadores de máquina de vetores de suporte (*support vector machine* - SVM). Em [5], classificadores SVM são utilizados para a classificação de subconjuntos fonéticos, recebendo como entrada os atributos de *spectral peak locations*, *spectral rolloff*, *spectral centroid*, dentre outros. Em [6], classificadores baseados em *k-nearest neighbor* (*k*-NN) são utilizados em uma classificação fonética hierárquica, operando com atributos nos domínios do tempo e da frequência, tais como os coeficientes de *dynamic time warping* (DTW) e os coeficientes MFCC, respectivamente.

Nos supramencionados trabalhos de pesquisa, os sinais considerados como referência são sinais de fala de banda larga, os quais são amostrados em uma taxa igual ou superior a 16000 Hz e contêm componentes de frequências superiores a 3400 Hz. Por essa razão, as estratégias supracitadas ([4]–[6]) apresentam desempenhos degradados quando utilizam sinais de banda limitada [7]. Em contraste, os algoritmos discutidos em [8] e [9] utilizam, como referência, sinais de fala de banda limitada oriundos da rede de telefonia pública. Nesse contexto, a busca pelo aprimoramento de algoritmos de classificação fonética que apresentem melhores desempenhos para sinais de banda limitada, utilizando uma quantidade reduzida de atributos, ainda é um tópico ativo de pesquisa na área.

Para contornar as dificuldades de classificação em sinais de fala de banda limitada, em [7] e [8], são apresentadas análises de atributos do sinal de fala para diferentes larguras de banda. Dessa forma, atributos e grupos fonéticos são investigados visando o aprimoramento da classificação. Em [7], o uso de diferentes atributos durante o processo de classificação é proposto. Já em [8], assim como em [5], são propostas classificações com ênfase em grupos fonéticos fricativos, os quais são caracterizados por apresentar energia espectral concentrada em altas frequências (frequências superiores a 3400 Hz que estão além das consideradas aqui para os sinais de banda limitada).

A investigação de grupos fonéticos também é discutida em [10], em que um classificador fonético, construído a partir de redes neurais de múltiplas camadas (*multilayer perceptron* - MLP), é proposto com base na especialização de classificadores em grupos fonéticos abrangentes (*broad group phonetic* - BGP) e na seleção de atributos de tempo e frequência. Tais classificadores enfatizam a capacidade de discriminação de quatro diferentes BGP (vogais, nasais, fricativas e oclusivas) para, posteriormente, discriminar os fonemas pertencentes a cada BGP estimado. Dessa maneira, melhores taxas de desempenho são obtidas quando comparadas com a classificação isolada de fonemas. Para a redução de dimensionalidade, a medida de informação mútua (*mutual information* - MI) é aplicada como

Ênio dos Santos e Rui Seara, LINSE – Laboratório de Circuitos e Processamento de Sinais, Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brasil, e-mails: enio@linse.ufsc.br; seara@linse.ufsc.br. Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

uma função objetivo, selecionando os atributos usados como entrada dos classificadores. Todavia, mesmo com a utilização de uma técnica de seleção de atributos, a estratégia apresentada em [10] ainda necessita de uma grande quantidade de atributos para se obter um desempenho satisfatório na classificação fonética.

Neste artigo, visando a seleção dos atributos de maior relevância para a classificação fonética de sinais de fala de banda limitada[1], são investigadas as métricas de análise de componentes de MI e seleção de parâmetros baseada em correlação (*correlation-based feature selection* - CFS). Posteriormente, uma estratégia de classificação fonética hierárquica é proposta, tendo como base os seguintes algoritmos de classificação: árvores de decisão (J.48 e LMT), MLP, SVM e k-NN [1], [3]. Tal estratégia consiste em identificar os atributos que facilitem a distinção entre grupos fonéticos diferentes. Os resultados de simulação permitem inferir acerca da seleção dos melhores atributos a serem considerados na composição de diferentes classificadores fonéticos, ratificando a eficácia da estratégia proposta.

II. Classificação Fonética

O objetivo da classificação fonética é determinar a classe fonética produzida por um locutor a partir de uma amostra do sinal de fala. Para tal, é necessário o agrupamento de segmentos de sinais da fala em classes discriminativas. Tais classes representam um conjunto de características temporais e espectrais singulares. Essas singularidades são específicas de cada conjunto de sinais e proporcionam uma discriminação mais eficiente entre as diferentes classes fonéticas [6], [8].

Neste trabalho de pesquisa, com base em [5], [8] e [10], discutimos a classificação do sinal de fala em BGP e, devido ao cenário adotado (banda limitada), uma ênfase é dada ao subgrupo fonético fricativo. Então, usando o conceito de classificação hierárquica discutido em [6] e de classes BGP em [10], são propostas classes agrupadas hierarquicamente conforme apresentado na Tabela I, a qual será utilizada como referência para o processo de classificação fonética dos sinais de fala. A Tabela I mostra diferentes níveis e detalhes da classificação proposta.

TABELA I
Distribuição de Classes Fonéticas

Classificação hierárquica			
Classes	Nível de classificação		
	1º	2º	3º
	BGP	Vozeamento	Ponto de articulação
C_{l1}	Vogal	–	–
C_{l2}	Nasal		
C_{l3}	Oclusiva		
C_{l411}	Fricativa	Vozeado	Alveolar
C_{l412}			Velar
C_{l413}			Pós-alveolar
C_{l414}			Labiodental
C_{l421}		Não-vozeado	Alveolar
C_{l422}			Glotal
C_{l423}			Pós-alveolar
C_{l424}			Labiodental

III. Extração e Seleção de Atributos

Para o aprimoramento da classificação fonética, é necessária a realização de procedimentos de mineração de dados a fim de evidenciar as características relevantes das amostras do sinal de fala [2], [4]. A técnica de seleção de atributos tem como objetivo evidenciar os atributos mais relevantes em uma tarefa de classificação, possibilitando a exclusão de atributos redundantes e de atributos que apresentem pouca contribuição (no sentido de MI e CFS) à tarefa de classificação.

A. Extração de Atributos

Assim, como em [5], [7] e [11], os seguintes atributos vetoriais e escalares são elencados:

- Taxa de cruzamento por zero

$$x_{zcr} = \frac{\sum_{k=2}^{N} \frac{1}{2} |\operatorname{sgn}[s(k)] - \operatorname{sgn}[s(k-1)]|}{(N-1)} \quad (1)$$

onde N representa o número de amostras por quadro. $\operatorname{sgn}[s(k)] = 1$, quando $s(k) \geq 0$. $\operatorname{sgn}[s(k)] = -1$, quando $s(k) < 0$.

- Energia normalizada do quadro

$$x_{nrp}(m) = \frac{\log_{10} E(m) - \log_{10} E_{min}(m)}{\log_{10} \bar{E}(m) - \log_{10} E_{min}(m)} \quad (2)$$

com $E(m) = \sum_{k=0}^{N-1} s^2(k)$, $E_{min}(m) = \min_{\mu=0}^{N_{min}} E(m-\mu)$ e $\bar{E}(m) = \alpha \bar{E}(m-1) + (1-\alpha)E(m)$, onde m representa o quadro atual. O fator de suavização é ajustado para $\alpha = 0,96$ e a janela de busca mínima é de $N_{min} = 200$.

- Índice de gradiente

$$x_{gi} = \sum_{k=2}^{N} \frac{\Psi(k)|s(k) - s(k-1)|}{\sqrt{\frac{1}{N} E(m)}} \quad (3)$$

com

$$\Psi(k) = \frac{1}{2}|\psi(k) - \psi(k-1)|. \quad (4)$$

A variável $\psi(k)$ representa o sinal do gradiente $\psi(k) = \operatorname{sgn}[s(k) - s(k-1)]$ [7].

- Estimativa da *kurtosis* local

$$x_k = \log_{10} \frac{1}{N} \sum_{k=0}^{N-1} s^4(k) - 2 \log_{10} \frac{1}{N} E(m). \quad (5)$$

- Centroide

$$x_c = \frac{1}{M} \frac{\sum_{i=1}^{M} |S(e^{j\Omega_i})|. \log_2(i)}{\sum_{i=1}^{M} |S(e^{j\Omega_i})|} \quad (6)$$

onde $S(e^{j\Omega_i})$ denota o i-ésimo coeficiente da transformada discreta de Fourier do sinal de entrada.

- Centroide espectral

$$x_{sc} = \frac{\sum_{i=0}^{M/2} i.|S(e^{j\Omega_i})|}{(\frac{M}{2}+1) \sum_{i=0}^{M/2} |S(e^{j\Omega_i})|}. \quad (7)$$

- *Flatness* espectral

$$x_{sf} = \frac{\sqrt[M]{\prod_{i=0}^{M-1} |S(e^{j\Omega_i})|^2}}{\frac{1}{M} \sum_{i=0}^{M-1} |S(e^{j\Omega_i})|^2}. \quad (8)$$

- Uniformidade

$$u = -\sum_{i=1}^{M} \left(\frac{|S(e^{j\Omega_i})|}{\sum_{i=1}^{M} |S(e^{j\Omega_i})|} \right) \log_M \left(\frac{|S(e^{j\Omega_i})|}{\sum_{i=1}^{M} |S(e^{j\Omega_i})|} \right). \quad (9)$$

[1] Aqui, a limitação de banda é associada a sinais amostrados à taxa de 8000 Hz, contendo componentes de frequências entre 0 e 3400 Hz.

- Estimativa de intensidade (*loudness*)

$$l = \log_2 \left(1 + \frac{1}{N} \sum_{k=1}^{N} |s(k)| \right). \quad (10)$$

- Estimativa da largura de banda

$$B = \sqrt{\frac{\sum_{i=1}^{M} [x_c - \log_2(i)]^2 . |S(\mathrm{e}^{j\Omega_i})|}{\sum_{i=1}^{M} |S(\mathrm{e}^{j\Omega_i})|}}. \quad (11)$$

- Os doze primeiros coeficientes LPC.
- Os dez primeiros coeficientes de reflexão.
- Os treze primeiros coeficientes MFCC, seguidos de suas primeiras e segundas derivadas (velocidade e aceleração) sobre uma janela temporal de três quadros sucessivos.

Desta forma, a etapa de extração de atributos resulta em vetores $\mathbf{x} = [x_1, \ldots, x_L]^{\mathrm{T}}$, obtidos a uma taxa r, onde $1/r$ representa a duração do quadro com $L = 71$ atributos e $r = 100$ Hz.

B. Seleção de Atributos

Neste trabalho, conforme [1], [2], [4] e [10], são investigados os métodos de seleção de atributos MI, avaliando atributos individualmente, e o método CFS, avaliando subconjuntos de atributos.

1) Informação Mútua: No método de MI, a função densidade de probabilidade $p(y)$ das classes fonéticas é estimada através de um conjunto de treinamento, representado por $\tau = \{(\mathbf{x}^1, y^1), \ldots, (\mathbf{x}^T, y^T)\}$, e a variável aleatória Y, associada às classes fonéticas, tem entropia

$$H(Y) = -\sum_{y \in Y} p(y) \log_2[p(y)] \quad (12)$$

onde T é o número total de quadros usados para o treinamento com $\mathbf{x} \supset x_i$ e $i \in \{1, \ldots, L\}$. Dessa forma, o conjunto de treinamento pode ser particionado de acordo com as funções densidades de probabilidade $p(x_i)$ e $p(y|x_i)$, estimadas através da frequência de ocorrência em τ. A entropia de Y, condicionada à observação da variável aleatória X_i, é expressa por

$$H(Y|X_i) = -\sum_{x_i \in X} p(x_i) \sum_{y \in Y} p(y|x_i) \log_2[p(y|x_i)] \quad (13)$$

e a informação mútua $I(X_i; Y)$ é computada como

$$I(X_i; Y) = H(Y) - H(Y|X_i). \quad (14)$$

O método consiste em calcular $I(X_i; Y)\ \forall\ i = 1, \ldots, L$, e então selecionar os A atributos com maiores valores de $I(X_i; Y)$.

2) Seleção de Parâmetros Baseada em Correlação: O ponto chave do método da CFS é a avaliação heurística de subconjuntos, examinando a utilidade de um atributo individual durante a predição da classe, juntamente com o nível de intercorrelação entre os demais atributos. A partir de (15), são atribuídos valores mais altos aos subconjuntos que contêm atributos fortemente correlacionados com a classe e apresentem baixa intercorrelação com os demais atributos. Assim,

$$Merit_s = \frac{k \bar{r_{cf}}}{\sqrt{k + k(k-1)\bar{r_{ff}}}} \quad (15)$$

onde $Merit_s$ é o 'mérito' heurístico do subconjunto de atributo A contendo k atributos, $\bar{r_{cf}}$, a correlação média atributo-classe e $\bar{r_{ff}}$, a intercorrelação média atributo-atributo. Para a aplicação de (15), é necessário o cálculo da correlação entre atributos. Para tal, em CFS, é utilizada a medida de incerteza simétrica para estimar o grau de associação entre os atributos X_i e X_j [1]. Portanto,

$$\bar{r_{ff}} = \left[\frac{H(X_i) + H(X_j) - H(X_i, X_j)}{H(X_i) + H(X_j)} \right]. \quad (16)$$

IV. Algoritmos de Classificação

A tarefa de classificação fonética pode ser estruturada em três blocos principais: pré-processamento, extração de atributos e classificação [3]. Costuma-se chamar os dois primeiros blocos de *front-end*, que tem como função reduzir o conjunto de dados a um conjunto específico de atributos que minimizem o erro no processo de classificação. O bloco seguinte - conhecido como classificador - tem a função de, a partir dos atributos apresentados, estimar a classe fonética a qual ele pertence.

Neste trabalho, utilizamos o método de aprendizagem supervisionada e investigamos os seguintes algoritmos de classificação: árvores de decisão J.48 e LMT, MLP, SVM e k-NN. A descrição detalhada desses algoritmos é apresentada em [1] e uma breve descrição é dada a seguir.

1) Árvores de Decisão J.48 *e* LMT*:* Árvores de decisão realizam uma análise dos atributos consistindo na implementação de um conjunto de limiares que avaliam (como verdadeiro ou falso) os atributos em cada um dos nós da árvore. Após a avaliação através dos limiares, cada nó é ramificado em duas ou mais sub-árvores. Nos nós, são calculados resultados baseados nos valores individuais dos atributos. Cada resultado possível está relacionado com uma ramificação da árvore, apresentando as classes como resultado final dessa ramificação. No algoritmo J.48, os limiares são valores escalares calculados com base na avaliação específica de um único atributo, enquanto no algoritmo LMT os limares são funções de regressão lógica.

2) Rede Neural de Múltiplas Camadas (MLP)*:* Em termos práticos, as redes neurais são ferramentas não-lineares de modelagem estatística de dados. Elas podem ser utilizadas para modelar relações complexas entre entradas e saídas. Neste trabalho, utilizamos o algoritmo *backpropagation* MLP.

3) Máquinas de Vetores de Suporte (SVM)*:* O algoritmo SVM consiste em um método de aprendizado supervisionado que realiza a transformação de dados de entrada em um espaço de atributos de alta dimensionalidade e efetua a separação de classes através de superfícies de decisões (hiperplanos) nesse espaço. Dessa forma, os algoritmos SVM buscam a obtenção de hiperplanos ótimos capazes de separar diferentes classes de dados com as maiores margens de separação possíveis.

*4) k-Nearest Neighbor (k-*NN*):* O algoritmo k-NN realiza a estimação da densidade local dos atributos de treinamento. Para a estimação da classe de um novo atributo x_a, o algoritmo k-NN calcula os k-vizinhos mais próximos a x_a e classifica-o como pertencente à classe que ocorre com maior frequência dentre os seus k-vizinhos.

V. Treinamento dos Classificadores

Para implementar uma tarefa de classificação fonética e possibilitar o treinamento dos classificadores, torna-se necessária uma base de dados composta por arquivos de fala

em que cada sinal de referência é caracterizado por uma série de atributos que descrevam características fonéticas. Neste trabalho, utilizou-se o *corpora* de fala (foneticamente balanceado) disponibilizado em [12]. Visando a simulação de limitação em banda, os sinais de fala desse *corpora* foram submetidos a uma filtragem passa-baixas com banda passante de 0 a 3400 Hz. Dessa forma, 2 horas de áudio, de 13 locutores masculinos e 10 femininos, são utilizadas na etapa de treinamento. Para o treinamento dos classificadores, é usada a plataforma de código-livre WEKA (*Waikato Environment for Knowledge Analysis*) [1], que consiste em uma ferramenta de aprendizagem de máquina e mineração de dados do estado da arte. Os algoritmos de classificação enumerados na Seção IV são utilizados no WEKA e configurados conforme descrito na Tabela II (para mais detalhes veja [1]).

TABELA II
CONFIGURAÇÃO DOS CLASSIFICADORES NO WEKA

Classificador	Configuração
J.48	C=0,25; M=2
LMT	I=−1; M=15; W=0,0
MLP	L=0,3; M=0,2; N=500; V=S=0; E=20; H=a
SVM	S=G=R=0; K=2; D=3; N=0,5; M=40; C=1
k-NN	K=1; W=0

A. Arquitetura de Classificação

O sistema proposto de classificação hierárquica primeiro analisa a qual BGP cada quadro pertence. Caso o quadro analisado pertença ao grupo fonético fricativo, um novo classificador é utilizado para a detecção de classes vozeadas ou não-vozeadas. Finalmente, de acordo com o nível de vozeamento estimado do quadro, um terceiro classificador é aplicado para identificação do tipo de fonema fricativo. A Fig. 1 ilustra o diagrama de blocos do processo de classificação hierárquica proposto.

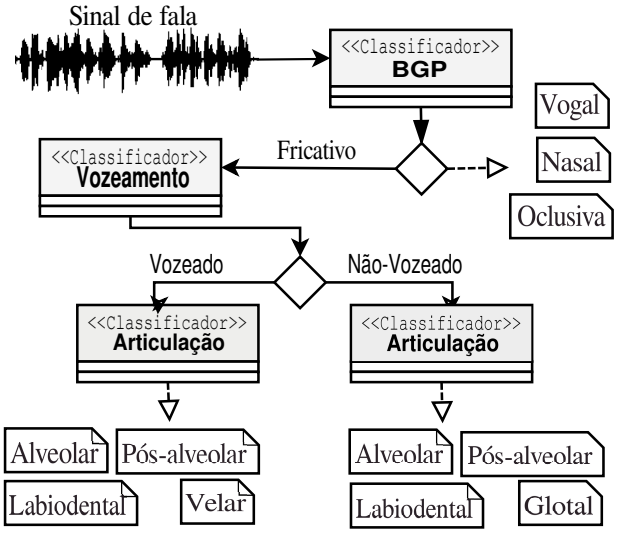

Fig. 1. Diagrama de blocos do processo de classificação.

Para cada nível hierárquico de classificação, um subconjunto dos atributos de entrada $\mathbf{x} = [x_1, \ldots, x_L]^{\mathrm{T}}$ é utilizado. Esses subconjuntos são previamente analisados e definidos, conforme as métricas de seleção de atributos descritas na Seção III-B, constituindo a etapa de *front-end* dos classificadores. Para efeitos de comparação, um sistema de classificação direta dos grupos apresentados na Tabela I é obtido usando o algoritmo k-NN.

VI. Resultados e Análise de Desempenho

Com o intuito de analisar a eficácia da seleção de atributos para os diferentes níveis hierárquicos, os conjuntos de atributos selecionados pelas técnicas MI e CFS são testados nos classificadores, discutidos na Seção IV, para cada grupo fonético. Os testes realizados avaliaram o desempenho dos classificadores para 47 minutos de sinais de fala, considerando 6 locutores masculinos e 3 femininos.

A Fig. 2 mostra a evolução do erro de classificação dos algoritmos k-NN em relação ao número de atributos selecionados, bem como a matriz de confusão para cada nível de classificação. Assim, os diferentes níveis hierárquicos podem ser avaliados isoladamente. Dessa forma, os atributos mais relevantes são selecionados para comporem o *front-end* de cada classificador. Nota-se, da Fig. 2, que a seleção de atributos utilizando a técnica CFS apresenta um melhor desempenho quando comparada com a técnica MI. Neste trabalho, tal comportamento é semelhante para todos os classificadores elencados na Seção IV. Por essa razão, as demais discussões aqui realizadas estão baseadas na técnica CFS.

A Tabela III apresenta o desempenho dos classificadores e destaca os mais eficientes. Assim, os classificadores SVM e k-NN, nos diferentes níveis hierárquicos, compõem a estratégia proposta de classificação. Além dos classificadores de melhor desempenho, são também apresentados (na Tabela III) os números de atributos de cada nível hierárquico, selecionados a partir da CFS.

Da Tabela III, obtém-se o desempenho geral da estratégia proposta para classificação. Nesse contexto, as taxas de erro de classificação dos subgrupos fricativos $E_{rro}(C_{l4vf})$ (considerados na Tabela I) são calculadas como segue: $E_{rro}(C_{l4vf}) = 1 - P(C_{lBGP})P(C_{lvoze})P(C_{lfri})$, onde $P(C_{lBGP})$ representa a probabilidade do quadro pertencer ao grupo fonético fricativo, $P(C_{lvoze})$, a probabilidade do quadro pertencer ao subgrupo vozeado ou não-vozeado e $P(C_{lfri})$, a probabilidade do quadro pertencer ao subgrupo fonético correspondente ao ponto de articulação (veja Tabela I). Dessa forma, a estratégia proposta de classificação hierárquica obteve taxa de erro de classificação de $21,72\%$ usando 67 atributos (vale aqui ressaltar que apenas alguns atributos LPC e MFCC não foram selecionados). Em contraste, o sistema padrão de classificação direta, usando todos os 71 atributos, obteve taxa de erro de classificação de $23,02\%$.

A estratégia proposta de classificação, composta pelos atributos e classificadores em destaque na Tabela III, é examinada através da comparação de suas matrizes de confusão fonética, mostradas na Fig. 2, com a matriz obtida pelo sistema padrão de classificação direta, mostrada na Fig. 3. A Fig. 4 apresenta os desempenhos individuais do sistema padrão e o da estratégia proposta para cada classe fonética.

TABELA III
DESEMPENHO DOS CLASSIFICADORES E SELEÇÃO DE ATRIBUTOS

Classes	Taxa de erro de classificação(%)				
	J.48	LMT	MLP	**SVM**	**k-NN**
I BGP	30,61	24,40	23,50	22,91	**18,36**
II Vozeamenento	10,10	05,72	05,60	**03,54**	03,78
III Fricativa vozeada	32,96	24,96	22,51	20,81	**16,86**
IV Fricativa não-vozeada	24,43	17,51	15,56	**13,13**	16,08

Seleção de atributos usando CFS				
Classes	I	II	III	IV
Números de Atributos	**53**	**55**	**56**	**52**

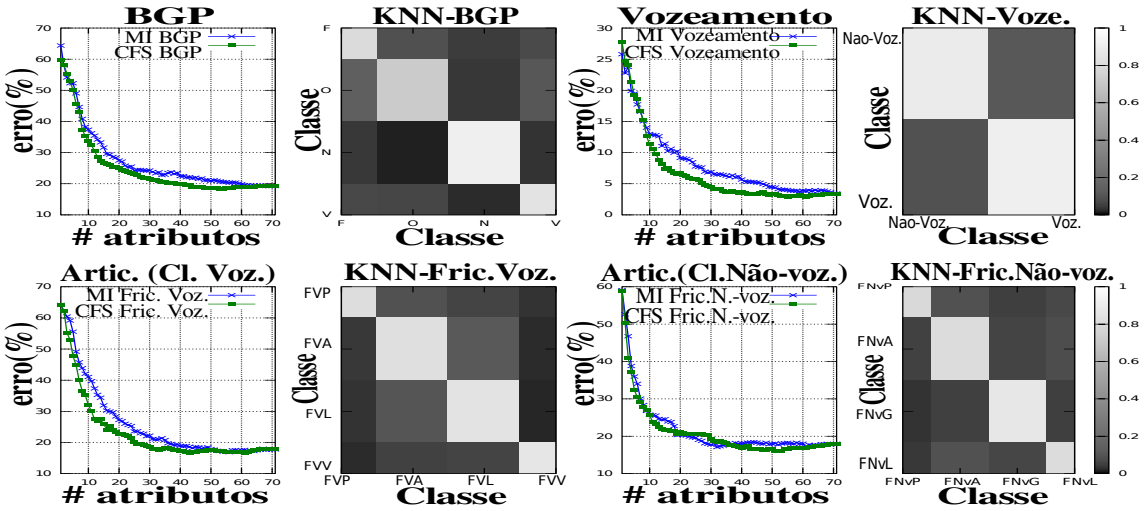

Fig. 2. Seleção de atributos e classificação fonética.

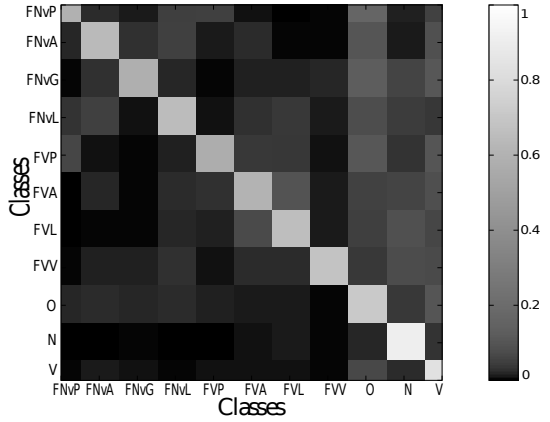

Fig. 3. Matriz de confusão do classificador k-NN padrão.

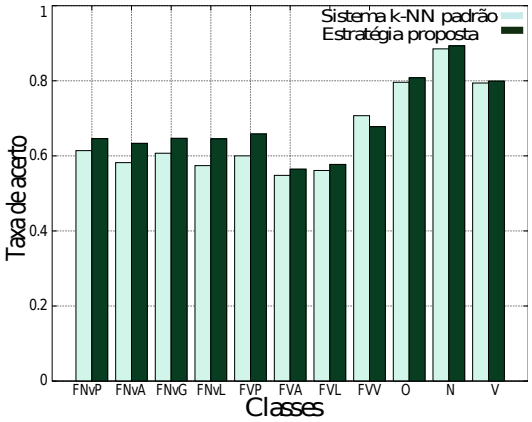

Fig. 4. Desempenho individual dos classificadores.

VII. Conclusões e Comentários Finais

Neste trabalho, uma nova estratégia para classificação fonética de sinais de fala de banda limitada foi apresentada. Essa estratégia foi baseada na seleção de atributos aplicados a uma classificação fonética hierárquica. A partir da técnica de seleção de atributos, foram apresentados os atributos mais relevantes para a tarefa de classificação. A estratégia de classificação hierárquica permitiu a utilização de diferentes algoritmos em cada nível hierárquico. Assim, diversos algoritmos de classificação foram investigados e integrados à estratégia proposta. A implementação de algoritmos especialistas em cada nível proporcionou a redução do número de atributos, assim como a redução no erro de classificação. A estratégia proposta exibiu um desempenho superior ao do sistema padrão de classificação direta. Os resultados de simulação apresentados permitiram uma análise acerca da seleção de atributos discriminativos para cada nível hierárquico, confirmando a eficácia da estratégia de classificação proposta.

Referências

[1] M. A. Hall, I. Witten, and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.

[2] M. A. Hall and G. Holmes, "Benchmarking attribute selection techniques for discrete class data mining," *IEEE Trans. On Knowledge and Data Engineering*, vol. 15, no. 6, pp. 1437–1447, Nov. 2003.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley-Interscience, 2001.

[4] A. Klautau, "Mining speech: automatic selection of heterogeneous features using boosting," in *Proc. Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, College Station, TX, USA, Apr. 2003, pp. 6–10.

[5] A. Frid and Y. Lavner, "Acoustic-phonetic analysis of fricatives for classification using svm based algorithm," in *Proc. IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, Eliat, Israel, Nov. 2010, pp. 17–20.

[6] H. Hamooni and A. Mueen, "Dual-domain hierarchical classification of phonetic time series," in *Proc. IEEE Int. Conf. on Data Mining (ICDM)*, Shenzhen, China, Dec. 2014, pp. 160–169.

[7] P. Jax and P. Vary, "Feature selection for improved bandwidth extension of speech signals," in *Proc. Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Montreal, Canada, May 2004, pp. 697–700.

[8] E. Silva and R. Seara, "Extensão artificial de largura de banda para sinais de fala usando classificação fonética," in *Anais XXXI Simpósio Brasileiro de Telecomunicações (SBrT)*, Fortaleza, CE, Brasil, Set. 2013, pp. 1–5.

[9] J.-W. Lee, J.-Y. Choi, and H.-G. Kang, "Classification of stop place in consonant-vowel contexts using feature extrapolation of acoustic-phonetic features in telephone speech," *The J. of the Acoust. Soc. of America*, vol. 131, no. 2, pp. 1536–1546, Dec. 2012.

[10] P. Scanlon, D. P. E. Ellis, and R. B. Reilly, "Using broad phonetic group experts for improved speech recognition," *IEEE Trans. Audio, Speech and Lang. Process.*, vol. 15, no. 3, pp. 803–812, Mar. 2007.

[11] M. Algabri1, M. Alsulaiman, G. Muhammad, M. Zakariah, M. Bencherif, and Z. Ali, "Voiced and unvoiced classification using fuzzy logic," in *Proc. Int'l Conf. IP, Comp. Vision, and Pattern Recognition (IPCV)*, Las Vegas, NV, USA, Jul. 2015, pp. 416–420.

[12] C. A. Ynoguti and F. Violaro, "A brazilian portuguese speech database," in *Anais XXXI Simpósio Brasileiro de Telecomunicações (SBrT)*, Rio de Janeiro, RJ, Brasil, Set. 2008, pp. 15–20.