Draft of paper in: Revista de Investigação Operacional, 21 (1), Junho, 2001

VOICE RECOGNITION USING A FUZZY MULTIPLE ATTRIBUTE MODEL

Tiago M. L. M. Simas, Departamento de Informática, FCT - Universidade Nova Lisboa, 2825-114 Monte Caparica, tmsimas@clix.pt

Rita A. Ribeiro, Departamento de Informática, FCT - Universidade Nova Lisboa, 2825-114 Monte Caparica, rr@di.fct.unl.pt

Abstract: In this work we use a multiple attribute fuzzy decision making method to recognize the five Portuguese vowels. The objective of this preliminary study is to show the potential of the fuzzy method in voice recognition.

Resumo: Neste trabalho utiliza-se um método de decisão multi-atributo difuso (multiple attribute fuzzy decision making) para reconhecer as vogais da língua portuguesa. Trata-se de um estudo preliminar que serve para demonstrar as potencialidades de aplicação do método difuso no reconhecimento de voz.

Keywords: Multiple Attribute Fuzzy Decision Making, Formants, Evolving Peaks.

Título abreviado: Reconhecimento de Vogais com um Método de Decisão Difuso.

1. INTRODUÇÃO

Neste artigo apresenta-se uma aplicação de reconhecimento de vogais (voz) da língua Portuguesa utilizando um método de tomada de decisão multiatributo difusa [7]. Trata-se de um estudo preliminar, baseado no artigo de Sankar, e Dwijesh [8] mas tanto o método proposto para recolha e tratamento dos dados como o método de agregação são diferentes. Os resultados obtidos pela utilização do método multiatributo difuso escolhido [5] [6] [7] são discutidos para mostrar as potencialidades desta aproximação.

O trabalho desenvolveu-se em três fases distintas. Primeiro, a recolha de várias vozes pronunciando as vogais {a,e,i,o,u}. Segundo, o seu tratamento em termos de definição dos atributos (critérios) a ser utilizados no reconhecimento. Terceiro, a utilização do método de decisão multiatributo difuso para determinar qual o nível de reconhecimento das vogais. Este nível é dado em percentagem.

Para a recolha dos dados foram usadas tecnologias comerciais, vulgares na recolha dos dados, tais como um microfone unidireccional, impedância 600 Ω e frequência de resposta 20Hz-20kHz.

Para o tratamento da informação, tanto a nível da elaboração dos atributos, como da sua agregação foram usados o: MATLAB v5.3 e SOUND FORGE 4.5.

A seguir apresentam-se breves descrições dos fundamentos teóricos do reconhecimento de voz e do método difuso, aplicado ao reconhecimento de vogais. Depois de introduzido o tema e o método de resolução são discutidos os resultados obtidos.

2. RECONHECIMENTO DE VOZ

Do ponto de vista Físico uma vogal é uma propagação do som pelo nosso sistema Fisiológico (meio), Glote, Boca, Vias Nasais e Garganta. É gerado um impulso (através dos músculos) que provoca uma perturbação no ar e esta propaga-se da glote à boca, nariz e garganta.[1],[9] e [10].

Têm sido feitos estudos de forma a modular esta propagação a partir de conceitos físicos e verifica-se que se pode modelar o meio de propagação como sendo um modelo físico de pequenos tubos com

determinada impedância [2] e [10]. Estas considerações levam-nos a propor um modelo físico linear como aproximação. Assim, poder-se-à considerar o sistema como sendo linear e invariante com relação ao tempo (LTI) [9] e [10].
Considerando o nosso sistema como sendo LTI, este sistema fica descrito pela convulsão de duas funções:

$$y[k]=h[k]*u[k] \qquad (1)$$

onde y[k] é o output (resposta do sistema à perturbação inicial u[k], sendo h[k] a função de transferência, normalmente inferida de parâmetros físicos conhecidos. Esta função de transferência dá-nos a relação entre o sinal captado (output) e o sinal inicialmente gerado (impulso). O facto deste impulso passar por um meio físico provoca uma alteração e essa alteração é dada por h[k].
No caso em estudo só é conhecido y[k] à saída da boca, pois só usamos um microfone à saída da boca, não conhecemos nem h[k] nem u[k]. Existem trabalhos em que se tenta medir o u[k] e a partir de conceitos físicos inferir h[k], [10].Como os sinais recolhidos (voz) são digitais as nossas funções são séries temporais.
Assim, podemos seguir vários caminhos para a determinação de h[k]. Um deles é obter a função de transferência a partir de propriedades físicas inerentes ao sistema, [10]. Outro método é inferir um determinado u[k] e com base na equação das diferenças (1) determinar h[k]. Ou, equivalentemente, trabalhando no domínio das frequências usar a transformada-Z (para uma descrição mais detalhada ver [1] e [9]):

$$Y(z) = Z\big(y[k]\big) = \sum_{k=0}^{\infty} y[k] z^{k} \qquad (2)$$

e atendendo que a transformada-Z da convolução é:

$$Z\big(h[k] * u[k]\big) = Z\big(h[k]\big) \cdot Z\big(u[k]\big) \qquad (3)$$

podemos obter:

$$Y(z)=H(z)U(z) \qquad (4)$$

onde Y(z)=Z(y[k]), U(z)=Z(u[k]) e H(z)=Z(h[k]).
Para inferirmos uma função de transferência, podemos considerar um modelo de amostragem estacionário, isto é, os pontos do sinal captados na janela de n pontos (foi utilizada uma janela de Hamming, [9] p.208), o sinal pode-se considerar estacionário. Assim, a função de transferência H(z) terá q-zeros e p-pólos e nesse caso o sinal y[k] poderá ser escrito na forma:

$$y[k] = \sum_{n=1}^{p} a_n y[k-n] + G \sum_{l=0}^{q} b_l u[k-l] \qquad (5)$$

onde G é o ganho da função u[k].
Aplicando a transformada Z a ambos os membros e atendendo que esta é linear, e considerando $b_0=1$ obtemos:

$$H(z) = G \cdot \frac{1 + \sum_{l=1}^{q} b_l z^{-l}}{1 - \sum_{k=1}^{p} a_k z^{-k}} \qquad (6)$$

A este método chama-se método AutoRegressivo (AR) [9]. Podendo assim escrever H(z) como a divisão de dois polinómios complexos;

$$H(z)=N(z)/D(z) \qquad (7)$$

Que correspondem respectivamente ao dominador e numerador de (6).

Como só conhecemos y(k) optou-se por usar o método AR, para a determinação dos coeficientes de N(z) e D(z) com um grau igual a n=fs/1000, onde fs é a frequência de amostragem, que neste caso foi de 11025Hz, e através do estudo de H(z) determinar os parâmetros que nos permitam tomar uma decisão sobre a classificação da vogal.
Sabemos que uma perturbação ao propagar-se num sistema tubular provoca frequências de ressonância [9], que correspondem aos pólos de H(z), isto é, aos zeros de D(z). Essas frequências de ressonância costumam denominar-se por formantes [9], e correspondem aos picos máximos da envolvente do espectro de frequências do sinal.
Assim, foram usadas as três primeiras formantes, designadas por, F_0, F_1 e F_2, sendo estas os atributos considerados no problema multiatributo difuso para o reconhecimento de vogais. Note-se que a primeira formante costuma estar próxima da frequência fundamental do sinal.
Foram então desenvolvidos vários algoritmos em MATLAB, que nos permitissem determinar as três primeiras formantes.

1- Módulo de aquisição do sinal e sua normalização;
2- Módulo para calculo das envolventes das formantes de acordo com o método já indicado em (5), (6) e (7) (designado por método 1);
3- Módulo para calculo das envolventes das formantes de acordo com o método de Ghael-Sandgathe [11], explicado em seguida (designado por método 2);
4- Módulo para determinação dos Picos da envolvente que correspondem às formantes, para ambos os métodos;
5- Módulo para tomada de decisão difusa multi-atributo.

Com base nos resultados da implementação destes algoritmos, verificou-se uma certa incerteza quanto a essa determinação, pois existem pólos que estão muito próximos e o algoritmo que determina os picos (máximos) da envolvente das formantes não os detecta, ou seja detecta uma média dos dois. Assim foi utilizado o método 2, de Ghael-Sandgathe, para uma melhor discriminação desses pólos e sua detecção, como podemos ver na figura 1.

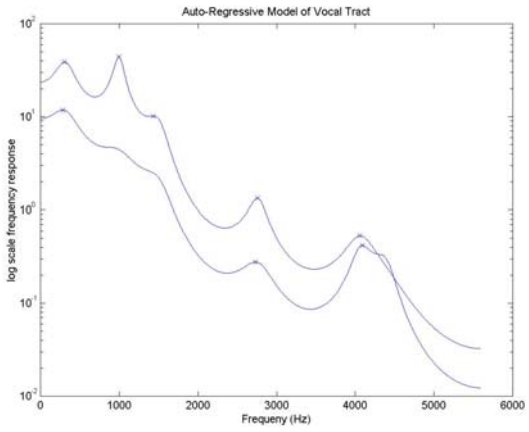

Figura 1- Dois picos detectados pelo método 2, de Ghael-Sandgathe (envolvente superior) que não são detectados pelo método 1 (envolvente inferior).

O método 2 aqui utilizado (Ghael-Sandgathe [11]) consiste em puxar pólos da zona exterior à do raio de convergência da transformada-Z, para o seu interior, recalculando assim os coeficientes dos polinómios complexos N(z) e D(z), conforme a equação (7), (ver com melhor detalhe [[11]).

3. DECISÃO MULTIATRIBUTO DIFUSA

Apresentaremos neste parágrafo uma breve introdução de lógica difusa (*Fuzzy Logic)* e de seguida a descrição do método de decisão multiatributo difuso (Multiple Atributte Fuzzy Decision Making), usado neste trabalho.

3.1. Breve descrição da lógica difusa

Em muitos casos em que temos um problema a modelar consideramos parâmetros que não podem ser definidos por um conjunto booleano (ou *Crisp*) bem delimitado, a tais conjuntos chamamos de difusos (ou *Fuzzy*) [3]. Consideremos um exemplo; quando pretendemos saber se um grupo de pessoas são altas ou baixas, podemos por exemplo dizer que uma pessoa é alta se a sua altura pertence ao conjunto C1=[1,80; 2,00]m, e baixa se pertencer ao conjunto C2=[1,00;1,80[m. Quando fazemos tal afirmação estamos a considerar que em conjuntos booleanos (ou *Crisp)*, uma pessoa que tenha altura de 1,79m é uma pessoa baixa e uma pessoa com 1,80m é alta. Neste tipo de abordagem uma pessoa com 1,00m é tão baixa como uma pessoa com 1,79m, quando esta última se aproxima mais de uma pessoa alta do que baixa. Para resolver este tipo de problemas, a utilização de conjuntos difusos em vez de *crisp* é apropriada. Assim, podemos considerar o mesmo tipo de problema em lógica difusa da seguinte forma: consideremos um conjunto D1=[1,60; 2,00] mas onde a cada altura corresponde uma determinada graduação ou grau de pertença μ, em que μ é uma função dos valores de D1, isto é, é a função $\mu(x):D1\rightarrow[0,1]$ (normalizada). Podemos ver na figura 2 um exemplo, onde uma pessoa é 0.5 alta (mais ou menos alta) para uma altura de 1,70m e 0.25 alta se tiver uma altura de 1,65m (pouco alta).

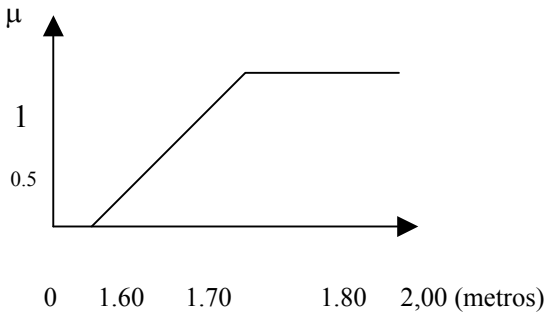

Figura 2 – Conjunto difuso "alto"

É de notar que podemos considerar um conjunto *Crisp* como um caso particular dum conjunto *Fuzzy* em que a função de pertença toma valores em $\mu(x):C\rightarrow\{0,1\}$.

Tal como na lógica booleana, na lógica difusa existem operadores de agregação entre proposições, tais como $\wedge$, $\vee$, e $\Rightarrow$, para definirem as operações de intersecção, união, e implicação [3]. Os operadores de agregação mais comuns caiem nas famílias das t-normas e t-conormas para, respectivamente, a intersecção e união (para mais tipos de operadores de agregação ver referência [3]). As t-normas e t-conormas são compostas por vários operadores sendo os mais conhecidos os seguintes: sejam A e B dois conjuntos difusos com funções de pertença μ_A e μ_B, respectivamente,

$$
\begin{aligned}
&t-normas\\
&\mu_{A\wedge B}=\mu_A\cdot\mu_B\\
&\mu_{A\wedge B}=\min(\mu_A,\mu_B)\\
&t-conormas\\
&\mu_{A\vee B}=\mu_A+\mu_B-(\mu_A*\mu_B)\\
&\mu_{A\vee B}=\max(\mu_A,\mu_B)
\end{aligned}
\qquad (8)
$$

Neste trabalho os operadores de agregação escolhidos foram o produto para a intersecção e o max para a união, pois empiricamente parecem ser mais apropriados para a classificação e reconhecimento das vogais. Note-se que mais estudos seriam necessários para garantir que outros operadores, como os paramétricos ou geométricos [3], não teriam melhor comportamento no reconhecimento das vogais.

3.2. Método Multiatributo Difuso

O método de decisão utilizado no reconhecimento de vogais, foi o multiatributo difuso (Multiple Atributte Fuzzy Decision - MAFD) [6]. Quando falamos de tomada de decisão multiatributo difusa [7], estamos a falar de problemas onde temos um conjunto de alternativas, conhecidas, as quais são classificadas com base em atributos/critérios pré-definidos. O processo de classificação é feito pela atribuição de valores a cada atributo, em relação a cada alternativa, que representam o nível de satisfação do atributo/critério para uma dada alternativa. No caso difuso, os atributos correspondem a conjuntos difusos, e portanto usam-se valores de pertença para definir o nível de satisfação dos atributos.

Depois agregam-se os atributos para cada alternativa, com um operador de agregação, para determinar a classificação final de cada alternativa. Este processo corresponde à intersecção dos atributos com o operador produto. Finalmente as alternativas são ordenadas e a que tiver melhor classificação é escolhida como *óptimo*. Este processo usa o operador max, das t-conormas. Neste trabalho as alternativas são as vogais (a,e,i,o,u), e os atributos considerados são as formantes *fuzzificadas*. Cada vogal tem um valor de pertença para os atributos que representa o grau de satisfação atingido na formante.

Usaram-se dois conjuntos de alternativas, as primeiras sendo as vogais e as segundas os dois métodos de cálculo das envolventes das formantes, o método 1 com os três primeiros picos da envolvente inicial e o método 2 com os três primeiros picos da envolvente determinada. A classificação das vogais é feita a partir das duas primeiras formantes, F1 e F2. Feita uma análise prévia dos dados, verificou-se uma certa dificuldade em determinar as duas primeiras formantes, pois os métodos usados por vezes não conseguem diferenciar bem as três primeiras formantes F0, F1 e F2, detectando um único pico que será um ponto que representa F0 e F1. Por forma a contornar este problema considera-se então para cada método os dois primeiros picos das envolventes como primeira e segunda formante respectivamente e um outro par que considera como primeira formante e segunda formante o segundo pico e o terceiro da envolvente, respectivamente, por forma a ultrapassar essa incerteza. Após a consideração anterior podemos ver na figura 3 a distribuição dos dados com todos os exemplos. Na figura 4 apresenta-se o mesmo estudo mas na banda 1.5kHzx3.5kHz onde se vê melhor as zonas de maior densidade onde se encontram as vogais. Comparando os resultados da figura 4 com outro trabalho realizado na área [4], figura 5, verifica-se que as zonas de maior densidade se aproximam das dos resultados obtidos no outro trabalho (figura 5).

Especificamente, os atributos considerados para cada uma das alternativas (a,e,i,o,u) foram a *fuzzificação* dos parâmetros F1 e F2. De acordo com a distribuição apresentada na Figura 4, que representam os dados obtidos experimentalmente e que representam o par ordenado (F1,F2) que caracteriza cada uma das vogais, e por comparação com a distribuição da Figura 5, de um outro trabalho já realizado (ver referência [4]), achou-se conveniente fazer como aproximação o uso de funções de pertença trapezoidais e triangulares.

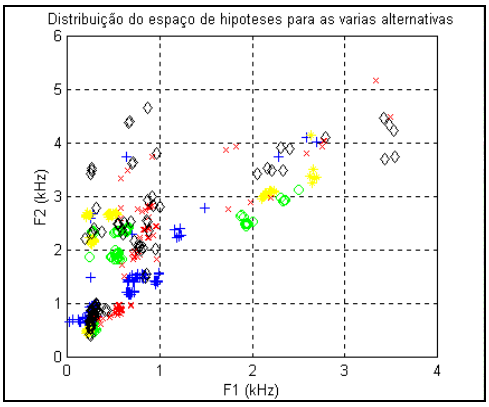

Figura 3- Estudo experimental com todos os exemplos

Legenda: 'a'='+', 'e'='o', 'i'='*', 'o'='×' e o 'u'='◊'

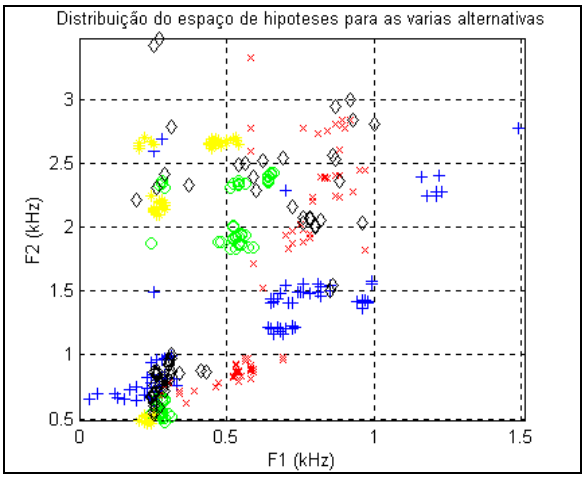

Figura 4- Estudo experimental com exemplos na banda 1.5kHz×3.5kHz

Legenda: 'a'='+', 'e'='o', 'i'='*', 'o'='×' e o 'u'='◊'

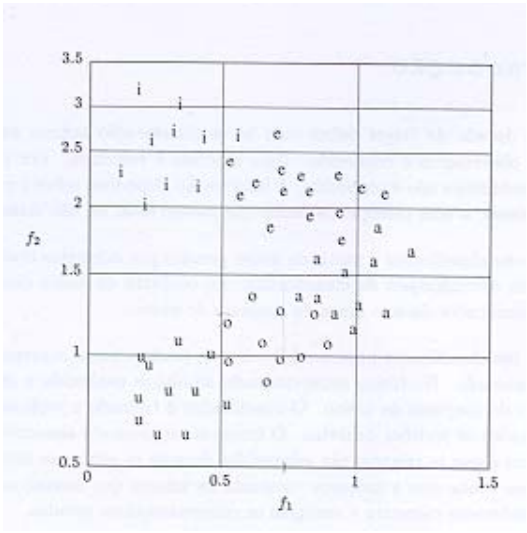

Figura 5- Resultados de outro estudo[4]

Legenda: Frequências das duas primeiras formantes em kHz.

A tabela 1 apresenta os conjuntos difusos correspondentes às funções triangulares e trapezoidais considerados para as formantes F1 e F2. Esta tabela foi estabelecida por inspecção dos dados, de acordo com a análise das distribuições referenciadas anteriormente e apenas aponta os pontos extremos das funções, i.e. os pontos onde os μ são 0 ou 1. As funções foram criadas de forma a dar maior grau de pertença às zonas de maior densidade da vogal, tentando-se arranjar um compromisso entre os dados obtidos experimentalmente, figura 4, e os dados da figura 5 [4]. É de dar especial atenção à função que representa o 'u', pois esta é triangular pelo facto de existir nessa zona grande densidade de dados de outras vogais, conforme se pode ver na figura 4. Este resultado deve-se ao facto de se ter incerteza em relação à consideração das duas primeiras formantes, conforme já explicado anteriormente.

A	μF1	μF2
a	Trap[0.5 0.7 0.9 1.4]	Trap[0.8 1.2 1.5 1.7]
e	Trap[0.3 0.6 0.7 1.0]	Trap[1.5 1.9 2.3 2.7]
i	Trap[0.15 0.3 0.4 0.6]	Trap[1.9 2.3 2.4 3.1]
o	Trap[0.4 0.55 0.65 0.8]	Trap[0.6 0.8 0.9 1.3]
u	Triag[0.1 0.3 0.6]	Triag[0.4 0.6 1.1]

Tabela 1- Funções de Pertença

Legenda: os valores apresentados são em kHz.

Por exemplo, Trap[0.5 0.7 0.9 1.4] representa a função trapezoidal apresentada na figura 6.

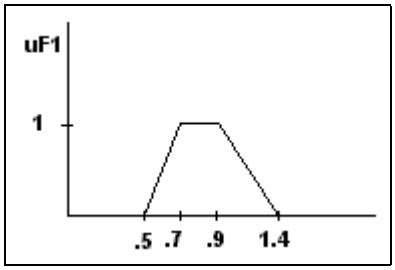

Figura 6- Exemplo de uma função de pertença

Como referido, neste trabalho são utilizados dois tipos de operadores de agregação max() para a selecção das alternativas e produto() para combinação dos atributos. Muitos outros operadores poderiam ter sido usados como se viu em 3.1 e se poderá ver em [3]. De seguida passa-se a descrever o processo pelo qual foi desenvolvido o trabalho.
Assim, para cada um dos métodos, método 1 e 2 para determinação das formantes, o cálculo para a agregação foi feito seguindo os seguintes passos:

(1) para cada envolvente das formantes escolhem-se dois vectores v_1=[F1 F2], tendo como coordenadas, os dois primeiros picos da envolvente e v_2=[F1' F2'], tendo como coordenadas o segundo e terceiro pico da mesma envolvente.

(2) para cada uma das alternativas calcula-se o grau de pertença dos vectores de acordo com a tabela-1, $\mu_{F_1}; \mu_{F_2}; \mu_{F_1'}; \mu_{F_2'}$.

(3) depois faz-se a conjugação das coordenadas, usando o operador de agregação
$$\begin{cases} v_1 = \cap(\mu_{F_1}, \mu_{F_2}) = \mu_{F_1} \cdot \mu_{F_2} \\ v_2 = \cap(\mu_{F_1'}, \mu_{F_2'}) = \mu_{F_1'} \cdot \mu_{F_2'} \end{cases}$$

(4) a seguir procede-se à sua disjunção, usando o operador de agregação max; $\mu A_{i1}=\max(v_1.v_2)$e assim obtém-se o grau de pertença da primeira envolvente, método 1, para cada uma das alternativas (a,e,i,o,u). Repete-se o mesmo processo para o método 2, correspondente à segunda envolvente e obtemos, $\mu A_{i2}=\max(v_1.v_2)$

(5) finalmente agregam-se os graus de pertença obtidos em (4) e (5) resultando num novo grau de pertença, $\mu A_i=\max[\mu A_{i1}\ \mu A_{i2}]$. A decisão óptima sobre qual a vogal que foi reconhecida é então obtida de acordo com:
$D=\max(\mu A_I)$

A título de exemplo temos:

Obtemos para uma determinada pessoa os 3 primeiros picos das envolventes que poderão corresponder às 3 primeiras formantes, do sinal da vogal 'a':

Método 1:	**Método 2**
F1=0.301 KHz	F1=0.280 kHz
F2=0.980 kHz	F2=2.692 kHz
F3=1.410 kHz	F3=4.027 kHz

Tabela 2- Três primeiros picos das envolventes

Com estes valores são formados os vectores com as respectivas coordenadas, segundo o passo (1)

Método 1	**Método 2**
v_1=[0.301;0.980]	v_1=[0.280;2.692]
v_2=[0.980;1.410]	v_2=[2.692;4.027]

Tabela 3- Aplicação do primeiro passo

Para cada alternativa são calculados os valores de pertença para cada formante considerada, segundo o passo (2), para cada um dos métodos usados.

Método 1				
Vogal	μF_1	μF_2	$\mu F_1'$	$\mu F_2'$
a	0	0.45	0.84	1
e	0.003	0	0.07	0
i	1	0	0	0
o	0	0.8	0	0
u	1	0.24	0	0

Tabela 4- Aplicação do segundo passo ao método 1

		Método 2		
Vogal	μF_1	μF_2	$\mu F_1'$	$\mu F_2'$
a	0	0	0	0
e	0	0.02	0	0
i	0.87	0.58	0	0
o	0	0	0	0
u	0.9	0	0	0

Tabela 5- Aplicação do segundo passo ao método 2

Usando o operador de intersecção produto, os atributos são agregados de forma a resultar uma classificação por alternativa, segundo o passo (3).

	Método 1		**Método 2**	
Vogal	v_1	v_2	v_1	v_2
a	0	0.84	0	0
e	0	0	0	0
i	0	0	0.51	0
o	0	0	0	0

u	0.24	0	0	0

Tabela 6- Aplicação do terceiro passo

Depois procede-se à selecção das alternativas, em cada método, usando o operador max, segundo os passo (4) e (5).

Método 1 (μA_1)				
a	e	i	o	u
0.84	0.00	0.00	0.00	0.24

Tabela 7- Aplicação do quarto e quinto passo ao método 1

Método 2 (μA_2)				
a	e	i	o	u
0.00	0.00	0.51	0.00	0.00

Tabela 8- Aplicação do quarto e quinto passo ao método 2

Finalmente procede-se à escolha do óptimo, ou seja do melhor valor obtido nos dois métodos, segundo o passo (6).

μA_I				
a	e	i	o	u
0.84	0.00	0.51	0.00	0.24

Tabela 9- Aplicação do sexto passo

Então a vogal com melhor reconhecimento é:

'a' com um valor de reconhecimento de 0.84

Este método é bastante interessante pois permite ter uma medida de comparação entre as várias alternativas, que com alguma análise posterior poderá ser utilizado no reconhecimento do orador.
Fazendo uma breve análise comparativa entre o método 1 e 2, os resultados de sucesso parciais são apresentados na tabela 10.

Alternativa	**Método 1**	**Método 2**
a	52%	48%
e	43%	57%
i	50%	50%
o	54%	46%
u	80%	20%

Tabela 10- Comparação entre métodos

Verificamos que os métodos se complementam de forma a atingir um grau de sucesso razoável. O resultado mais díspar, é a vogal 'u', no entanto não se avançam de momento justificações pois pensamos fazer um estudo estatístico mais aprofundado no futuro.

4. RESULTADOS

Foram recolhidos dados de quatro pessoas, dois Homens e duas Mulheres, de idades compreendidas entre os 30 e 38 anos, para cada um dos indivíduos foram recolhidos 15 amostras de cada vogal, perfazendo um total de 300 dados que correspondem a 600 exemplos. A frequência de amostragem, foi de fs=11025Hz, com uma amplitude de 16bits. Para a aquisição dos dados (sons) foi utilizado o software "Sound Forge v4.5".

Note-se que neste tipo de modulação é importante o factor idade, pois uma criança tem uma fisiologia diferente de um adulto e o mesmo para grandes variações de idade. Estas foram as razões da escolha de população efectuada.
Os resultados obtidos, tanto para os resultados correctos como para os mal classificados, utilizando o método multi-atributo difuso são apresentados na tabela 11 e 12.

Alternativa	%Resultados Correctos	Maior Valor de Pertença	Menor Valor de Pertença
a	100%	1.00	0.57
e	58%	0.88	0.59
i	87%	0.67	0.25
o	62%	1.00	0.0
u	45%	0.81	0.0

Tabela 11- Tabela de resultados correctos

Na tabela 12 apresentam-se os graus de pertença dos dois primeiros classificados, no caso do exemplo ser mal classificado, observando-se que se usou como elementos representativos as classificações em que o primeiro classificado tinha maior grau de pertença.

Alternativa	Resultado mal classificado com as duas melhores classificações com maior grau de pertença para o primeiro classificado
a	N/A
e	0.89/u+0.88/e; 0.94/i+0.66/e
i	0.52/u+0.42/i
o	0.74/u+0.63/o;1.0/e+0.68/u;0.75/e+0.43/o
u	1.0/e+0.73/u; 0.92i; 0.44/i+0.34/u

Tabela 12- Tabela de resultados mal classificados

Na tabela 11, o resultado alto da vogal ‘a’, justifica-se pelo facto de se encontrar numa região onde F1 e F2 são altos, logo mais fácil a detecção dos picos das formantes e não se dispersa muito. Em relação ao ‘u’ era de esperar um valor baixo pois este situa-se em zonas de muito baixa frequências.
É curioso observar a dispersão dos dados na figura 3, e comparar com a tabela de resultados tabela 11, pois inicialmente os resultados pareciam bastante dispersos e a decisão parecia ser difícil. No entanto, através do método aplicado, a decisão mostrada estatisticamente na tabela 11, revela um bom poder de decisão, o que mostra que o método poderá ser eficaz para este género de problema.
Como se pode ainda ver na tabela 11, os valores de pertença obtidos também apresentam alguma disparidade. Enquanto para a vogal "a" foi de 1.00 o valor máximo de pertença, para a vogal “e” foi de 0.88, vogal “i” de 0.67, vogal “o” de 1.00 e para a vogal “u” de 0.81. Comparando os valores de pertença máximos com os valores estatísticos verifica-se que para que todas as vogais à excepção do “a” se encontram um pouco dispersas como se pode ver nas figuras 3 e 4. Para melhorar estes resultados deveriam ser considerados os seguintes aspectos: utilizar novas medidas de decisão; melhorar o algoritmo de detecção de picos da envolvente; ou melhorar a qualidade da aquisição dos dados. No entanto, dadas as limitações de hardware, o método de decisão apresentado parece ser bastante viável face ao ruído do sinal. A solução para melhores resultados depende mais do método utilizado para a detecção dos picos da envolvente e será necessária uma análise estatística mais aprofundada, considerando uma amostra maior, médias de reconhecimento, desvios padrões etc., para se poderem tirar mais ilações sobre os valores de pertença obtidos e qual o seu significado neste contexto. Isto fará parte de trabalho futuro.

5. CONCLUSÕES

Este trabalho representa uma primeira experiência do uso de um método de decisão difuso em problemas de reconhecimento de vogais Portuguesas. Para ser feito um estudo mais rigoroso, esta aproximação terá de aumentar significativamente o número de dados, e considerar no mínimo 30 pessoas, de varias idades, e uma maior quantidade de dados por pessoa, o que não foi possível neste primeiro trabalho.

Contudo, pode-se concluir que o método multi-atributo de decisão difuso parece ter grande potencial para o reconhecimento de voz, uma vez que é capaz de lidar, de uma forma simples, com a incerteza intrínseca deste tipo de problemas. O exemplo apresentado assim o demonstra.

BIBLIOGRAFIA

[1] Chen, C. T., System and signal analysis (2ªED), Saunders Colledge Publishing, New York, 1994.
[2] Halliday, D., Resnick, R., Fundamentals of physics (3ªED), John Wiley & Sons, New York, 1988.
[3] Klir, G.J. e T.A. Folger, Fuzzy sets, uncertainty, and Information, Prentice-Hall, New York, 1988.
[4] Marques, J.S., Reconhecimento de padrões, IST Press, Portugal, 1999.
[5] Ribeiro, R. A., Fuzzy Evaluation of Thermal Quality Of Buildings, Computer-Aided Civil and Infrastructure Engineering, 14 (1999), 155-162.
[6] Ribeiro, R. A., Baldwin, J F., A Multiple Attribute Fuzzy Decision Support System: Two Applications, Fuzzy logic and Soft computing, (1995), 452-461.
[7] Ribeiro, R. A., Fuzzy multiple attribute decision making: A review and new preference elicitation techniques, Fuzzy Sets and Systems 78 (1996) 155-181.
[8] Sankar, K. P., Dwijesh, D M., Fuzzy sets and decisionmaking approaches in vowel and speaker recognition, IEEE Trans. Systems, Man and Cybernetics, (1977) pp. 625-629.
[9] Shaughnessy, D. O., Speech Communication Human and Machine, Addison-Wesley, New York, 1990.
[10] http://speech.llnl.gov/thesis/index.htm
[11] http://umamitoba.ca/linguistics/russell/138/sec4/formants.htm