

# CARACTERIZAÇÃO DE VOZES ENVELHECIDAS UTILIZANDO MÉTODOS DE CLUSTERIZAÇÃO

Marco Silva<sup>1</sup>, Jorge L. M. Amaral<sup>1</sup> e Marley M. B. R. Vellasco<sup>2</sup>

<sup>1</sup> Universidade do Estado do Rio de Janeiro - UERJ, Departamento de Engenharia Eletrônica e Telecomunicações, Av. Maracanã – Cep: 20559-900, Rio de Janeiro, RJ, Brasil.

<sup>2</sup> Pontifícia Universidade Católica do Rio de Janeiro - PUC, Departamento de Engenharia Elétrica, Rua Marquês de São Vicente, 225 - Cep: 22451-900, Rio de Janeiro, RJ, Brasil.  
lncs@springer.com

**Resumo.** O objetivo deste trabalho é analisar o processo de envelhecimento da voz através de técnicas de agrupamento (*clustering*), utilizando como atributos de entrada parâmetros extraídos dos sinais de voz e do sinal glotal. Os sinais de voz e glotal têm sido utilizados para identificar patologias diversas e a extração desses parâmetros são obtidos utilizando técnicas não invasivas. O sinal de voz foi obtido na gravação da voz dos locutores através de um microfone e o sinal glotal através da filtragem inversa do sinal das vozes gravadas. Dois métodos distintos foram usados nesta análise: Mapas Auto-organizáveis (SOM) – Kohonen e *Fuzzy c-Means*. A base de dados utilizada é constituída de vozes gravadas por locutores masculinos e femininos, de idades diferentes, na faixa etária de 15 a 100 anos. Os modelos de agrupamento permitiram classificar a idade vocal dos locutores, identificando modificações nos parâmetros acústicos da voz, independente da idade cronológica dos locutores.

**Palavras-chave:** Vozes, Agrupamento, Mapas de Kohonen, *Fuzzy*.

## 1 Introdução

O envelhecimento do ser humano é um processo natural que, com o avanço da idade, pode acarretar alguns problemas vocais. Esse assunto é de grande interesse na área de processamento de sinais, pois o envelhecimento torna o organismo vulnerável, o que reflete diretamente na produção da voz [1] [2] [3].

O envelhecimento vocal, também conhecido como presbifonia, consiste na deterioração da voz pela idade, sendo resultante de um processo de modificações anatômicas e fisiológicas ocorridas na laringe e que interferem diretamente na qualidade da voz [1] [14]. Uma série de modificações na voz aparece com a idade, e alguns fatores como problemas respiratórios, perda de elasticidade muscular e calcificação das cartilagens, contribuem para seu envelhecimento. Não se pode definir o envelhecimento somente pela cronologia, mas também pelas condições físicas e sociais nas quais se apresentam os indivíduos, ou seja, um locutor pode ter uma idade avançada e uma voz de ótima qualidade [3].

O sinal de voz é obtido diretamente na gravação da voz e o sinal glotal através da

filtragem inversa do sinal de voz. Os parâmetros extraídos desses sinais são denominados medidas acústicas da voz, e vêm sendo utilizados para obter informações sobre a saúde vocal, distinguindo aquelas que ocorrem naturalmente com a idade daquelas associadas a alguma patologia [1].

Como exemplo de aplicação desses parâmetros, [3] implementou uma rede neural com os parâmetros extraídos de um sinal de voz, de modo a identificar automaticamente locutores com vozes envelhecidas. A base de dados foi dividida em dois tipos de locutores masculinos e femininos.

Para este trabalho utilizou-se a mesma base de dados utilizando métodos de agrupamento para a formação de grupos.

Métodos de agrupamentos em grandes bancos de dados é um estudo da ciência denominada *data mining* ou *mineração de dados* [5]. Os métodos de agrupamentos podem ser divididos em hierárquicos e não hierárquicos. Métodos hierárquicos apresentam uma estrutura tipo árvore, sendo divididos em aglomerativos ou divisivos. Os métodos não hierárquicos agrupam  $n$  indivíduos em  $k$  grupos de tal forma que indivíduos de uma mesma classe sejam os mais semelhantes possíveis em cada cluster seja bem separado [6], e o número de cluster deve ser especificado.

Baseado em aprendizado competitivo e não-supervisionado [7], o SOM possui grandes vantagens com relação aos demais métodos: fornece uma visualização bidimensional dos clusters independente do tamanho da sua base de dados, menos esforço computacional, robusto na presença de dados ruidosos e *outliers*, não necessitam que o número de grupos seja fornecido e não necessita de comparação entre os indivíduos que formam a base de dados [11]. Assim, têm-se cada indivíduo pertencendo ao seu respectivo grupos.

Se o conjunto de dados possui informações de caráter impreciso, pode-se usar a teoria de Conjuntos *Fuzzy* com o objetivo de fornecer uma ferramenta matemática para tratamento dessas informações. A *Lógica Fuzzy* foi inicialmente construída a partir dos conceitos já estabelecidos de lógica clássica, mas permitindo que um indivíduo possa pertencer a mais de um cluster com diferentes graus de pertinência [6].

Essa pertinência pode assumir valores entre 0 (zero) e 1(um) ao invés de puramente verdadeiros ou falsos. A soma das pertinências deve ser igual a 1(um), e o algoritmo se inicia com uma suposição inicial sobre os centros de cada agrupamento, sendo esta na maioria das vezes incorreta. Para cada padrão é assinalado um grau de pertinência referente a cada agrupamento. Estes centros e os graus de pertinência vão sendo atualizados com o objetivo de minimizar uma função objetivo que representa a distância de cada padrão em relação ao centro de cada grupo, ponderado pelo grau de pertinência [6].

Este artigo está dividido da seguinte forma: na Seção 2 é apresentado o base de dados e a metodologia usada para extração dos parâmetros do sinal de voz e do sinal glotal. Na Seção 3, são apresentados, de forma resumida, os algoritmos de agrupamento usados no artigo: Mapa de Kohonen e *Fuzzy c-Means*. Na Seção 4 é apresentada a modelagem empregada, em ambos os modelos, para a análise do envelhecimento da voz. Na Seção 5 são apresentados e discutidos os resultados obtidos e, finalmente, as conclusões são alinhadas na Seção 6.

## 2 Base de Dados

Com o objetivo de analisar o envelhecimento da voz, foi criada uma base de dados composta por 52 registros de vozes masculinas e 70 registros de vozes femininas, da emissão da vogal sustentada /e/, em Português como falado na cidade do Rio de Janeiro.

A base de dados é formada por locutores voluntários, gravados em suas próprias casas, no consultório de uma fonoaudióloga, em casas de repouso e casa do Retiro dos Artistas [3].

Para cada registro de voz, a idade cronológica de cada locutor é conhecida. Para se ter uma idéia da distribuição dos locutores, a base de dados foi dividida em três grupos de faixas etárias: 15 a 30 anos de idade, 31 a 59 anos de idade e de 60 a 100 anos.

A distribuição de locutores nesses três grupos foi a seguinte:

Locutores Masculinos:

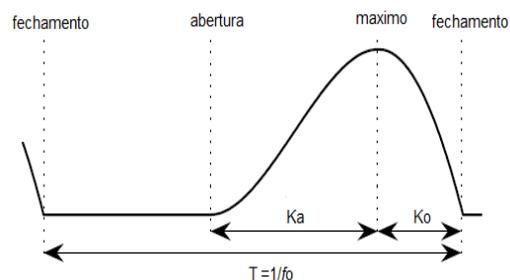
- 15 a 30 anos – 19 locutores;
- 31 a 59 anos – 19 locutores;
- 60 a 100 anos – 14 locutores.

Locutores Femininos:

- 15 a 30 anos – 15 locutores;
- 31 a 59 anos – 38 locutores;
- 60 a 100 anos – 17 locutores.

### 2.1 Extração dos parâmetros dos sinais de voz

Os parâmetros usados nesse trabalho são extraídos dos sinais de voz e glotal, obtido através da filtragem inversa do sinal de voz. A Fig. 1 ilustra um exemplo de um período completo do sinal glotal.



**Fig. 1.** Pulso glotal.

Parâmetros extraídos do sinal glotal:

- Duração da fase de fechamento (**Ko**) [4];
- Duração da fase de abertura (**Ka**) [4];
- Quociente de abertura (**OQ**) [4];
- Quociente de velocidade (**SQ**) [4];
- Quociente de fechamento (**CIQ**) [4];
- Quociente de amplitude (**AQ**) [4];
- Quociente de amplitude normalizado (**NAQ**) [2] [3];
- Quociente de abertura (**Oqa**) [3];
- Quociente de quase abertura (**QoQ**) [4];
- Diferença de harmônicos (**DHI2**) [4];
- Fator de grandeza de harmônicos (**HRF**) [4];

Parâmetros do sinal de voz:

- Frequência Fundamental (**Fo**) [3];
- *Jitter* [3];
- *Shimmer* [3];
- *Harmonics-to-Noise Ratio (HNR)* [3].

### 3 Algoritmos de Clusterização

#### 3.1 Mapas auto-organizáveis (SOM – *Self Organizing maps*)

O SOM é uma rede neural artificial, baseada em aprendizado competitivo e não-supervisionado, que define um mapeamento de um espaço de entrada contínuo para um conjunto discreto de vetores de referência (neurônios), os quais são dispostos em uma camada bidimensional. Cada neurônio tem a mesma dimensão do espaço de entrada, e para cada padrão de entrada, um neurônio é escolhido vencedor, utilizando o critério de maior similaridade. O mapa é organizável de tal forma que os neurônios fisicamente próximos no mapa tendem a representar o mesmo grupo, uma vez que o ajuste de pesos do neurônio vencedor (e seus vizinhos) é tal que faz com que o vetor de pesos se aproxime do vetor apresentado na entrada [7].

Em uma rede Kohonen, o neurônio com maior resposta a uma determinada entrada, tem os pesos de suas sinapses atualizados de maneira que aumente sua resposta diante daquele padrão de entrada, caracterizando assim o treinamento competitivo da rede. Os pesos do neurônio vencedor, juntamente com os pesos dos seus neurônios vizinhos, são ajustados de acordo com a (1) [7]:

$$m_i(t+1) = m_i(t) + h_{c_i}(t)[x(t) - m_i(t)] \quad (1)$$

Onde  $x(t)$  é o padrão de entrada e  $hci(t)$  inclui os fatores taxa de aprendizado e função de vizinhança em torno do neurônio vencedor [7], que reduzem com as épocas de treinamento. Uma vez que a rede tenha sido treinada, o arranjo de neurônios do SOM reflete características estatísticas importantes do espaço de entrada.

O processo de aprendizagem deste modelo acontece em duas fases: ordenação e convergência. Parâmetros importantes na fase de ordenação são a taxa de aprendizagem, que decresce a cada atualização, e a função e vizinhança. Na fase de convergência, ou ajuste fino, recomenda-se que o número de iterações seja de no mínimo 500 vezes o número de processadores, e a vizinhança ao redor de um neurônio é mantida igual a 1 (um) podendo decrescer até 0 (zero) [12].

Para a realização dos experimentos com a rede SOM utilizou-se a ferramenta *SOMToolbox* [13], que é uma implementação do Mapa Auto-organizável de Kohonen no Matlab®. Conforme será apresentado na Seção 5, esta ferramenta apresenta o resultado através de um mapa de cores, onde cores “quentes” representam distâncias maiores e cores “frias” distâncias menores [12].

### 3.2 Fuzzy c-Means

O objetivo de um agrupamento tradicional é atribuir cada ponto de uma base de dados a um único cluster. Um agrupamento difuso atribui diferentes graus de pertinência a cada ponto, e este é compartilhado entre vários grupos. Cria-se o conceito de fronteiras difusas, diferente do tradicional. Bezdek [11] afirma que o modelo de fronteira bem definida normalmente não reflete a descrição de dados reais, sendo necessário o desenvolvimento de um algoritmo de agrupamento difuso [10], denominado *Fuzzy c-Means*.

A técnica de *fuzzy clustering* é empregada para que os  $n$  vetores  $x_j$  sejam agrupados em  $c$  clusters [5]. Os centros assim como os graus de pertinência de cada padrão são atualizados, com o objetivo de minimizar uma função custo que representa a distância de cada padrão em relação ao centro de cada agrupamento, ponderado pelo grau de pertinência do padrão (8).

$$j(u_{ij}, v_k) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2 \quad (2)$$

Onde  $m \in [1, \infty)$  é o coeficiente *fuzzy* responsável pelo grau de fuzzificação dos elementos de  $x_j$  e  $v_i$  é o centróide do  $k$ -ésimo cluster. Maiores detalhes podem ser encontrados [5] [6]. Neste trabalho foi utilizado o *Fuzzy c-Means* no Matlab®.

## 4 Modelagem

O modelo SOM foi utilizado devido ao fato de possibilitar a representação de um conjunto de dados com muitos atributos, em um mapa de baixa dimensão. Por ser um

modelo baseado em aprendizado não-supervisionado o número de grupos é determinado pelo algoritmo, o que torna o modelo interessante.

O modelo Fuzzy possibilita considerar que indivíduos pertençam a diferentes grupos com diferentes graus de pertinência. Assume-se que o número de agrupamentos é conhecido, determinam-se valores para o coeficiente *fuzzy*, e inicializa a matriz de pertinência com valores aleatórios. Ocorre uma otimização da função custo com a atualização dos graus de pertinências e centros dos grupos.

#### 4.1 Mapas auto-organizáveis – SOM

A tabela foi montada com os 18 parâmetros extraídos dos sinais de voz e vocal, dos 52 locutores masculinos e 70 locutores femininos, utilizando a ferramenta SOMToolbox [13].

O treinamento da rede pode ser feito de duas maneiras: automático ou especificando passo-a-passo as respectivas fases. A rede foi treinada das duas maneiras e o melhor resultado foi especificando passo-a-passo um resultado de mapa obtido no treinamento automático. No treinamento passo-a-passo foram feitos diversos testes para diferentes tipos de mapas: 10x10, 5x10, 15x10, 15x15. No treinamento automático os resultados dos mapas foram 4x9 para o masculino e 8x5 para o feminino.

No treinamento passo-a-passo a fase de ordenação o  $hci(t)$  começa com um valor alto e vai diminuindo proporcionalmente, por um ciclo de  $t=1000$  iterações, e na fase de ajuste fino o valor de vizinhança foi mantido igual a 1 e a taxa de aprendizado igual a 0,01. O número de iteração nessa fase, que geralmente é 500 vezes o número de processadores, para os locutores masculinos (36 processadores) foram 18.000 iterações e os locutores femininos (40 processadores) foram 20.000 iterações.

#### 4.2 Fuzzy c-Means

Para o método *Fuzzy c-Means*, que foram realizados diversos testes para avaliar o melhor valor para o coeficiente *fuzzy* ( $m$ ), assim como o número ideal de clusters. O critério utilizado para obter o melhor valor de  $m$  foi realizar testes com diferentes valores de  $m$  para ambos os sexos, variando de 1,2 até 4. Segundo [10], quando  $m \rightarrow 1$  o *Fuzzy c-Means* converge para o método *K-means* (caso *crisp*) enquanto que quando  $m \rightarrow \infty$ , os centros dos agrupamentos ficam mais próximos do centróide do conjunto de dados e a variância de cada agrupamento se torna maior, tornando o agrupamento mais *fuzzy*. O número de clusters foi testado para 2 a 5 grupos. Este será detalhado posteriormente na seção 5.

Para este trabalho, valores de  $m$  acima de 1,8 apresentaram um agrupamento mais *fuzzy*, logo, a faixa de utilizada para o coeficiente *fuzzy* foi 1,2 a 1,6.

Para definir quando um indivíduo pertence a um determinado grupo, valores limites de pertinência (*threshold*) foram estipulados. Para valores altos de  $m$  menor deve ser o valor de *threshold*. Se o maior valor de pertinência estiver abaixo desse *threshold*, significa que o indivíduo está com características de outros grupos. Quanto mais *crisp* ( $m \rightarrow 1$ ) mais próximo do valor 1 deve ser esse limiar.

A seguir os valores de  $m$  e seus respectivos *thresholds* testados para ambos os locutores:

$$m=1,2 - th=0,85; \quad m=1,4 - th=0,75; \quad m=1,6 - th=0,6$$

## 5 Resultados

Os mapas auto-organizáveis fornecem um resultado através de uma matriz-U (ou matriz de distância unificada), que permite visualizar as distâncias entre cada neurônio do mapa e seus vizinhos. É a técnica mais utilizada para visualizar clusters da rede SOM [12]. Conforme mencionado anteriormente, as cores fornecem a divisão dos grupos. Quanto mais a cor tende para o marrom, maior é a distância entre os neurônios, caracterizando a separação de grupos dos indivíduos. Por outro lado, quanto mais azul menor será a distância entre os neurônios, caracterizando a formação de um grupo. Assim, um dos usos mais comuns da matriz-U é servir como ponto de partida para visualização dos grupos.

As variáveis que são utilizadas como entradas do algoritmo são analisadas, e ajudam na caracterização dos grupos detectados na matriz-U.

Como a caracterização dos grupos na matriz-U é feita pela análise dos mapas dos parâmetros de entrada, observou-se que (Fig.2 e 3) dos 18 parâmetros utilizados como entrada alguns apresentaram mapas semelhantes e outros que nada contribuíram para a formação dos grupos. Isso ocorreu nos resultados para os dois grupos de locutores e estes parâmetros foram descartados. Os parâmetros descartados foram: pp, AQ, CIQ, OQ1, OQ2, Oqa e Qoq.

O parâmetro pp não contribuiu para a formação dos grupos, e os parâmetros AQ, CIQ, OQ1, OQ2, Oqa e Qoq têm mapas semelhantes ao NAQ, logo basta utilizarmos este para análise dos grupos na matriz-U.

A matriz-U permite visualizar os grupos formados, mas dentro do *SOMToolbox* [13] têm ferramentas que possibilitam gerar grupos com certo grau de detalhamento como número de vezes em que o neurônio foi ativado, proximidade dos neurônios, entre outros.

Analisando visualmente a matriz-U dos locutores, consegue-se ver a formação de três grupos para os locutores masculinos e dois grupos para os locutores femininos.

Utilizando o algoritmo *K-means* inserido no SOM, permite visualizar a formação dos grupos, a quantidade de vezes que o neurônio foi ativado e a distância entre os mesmos.

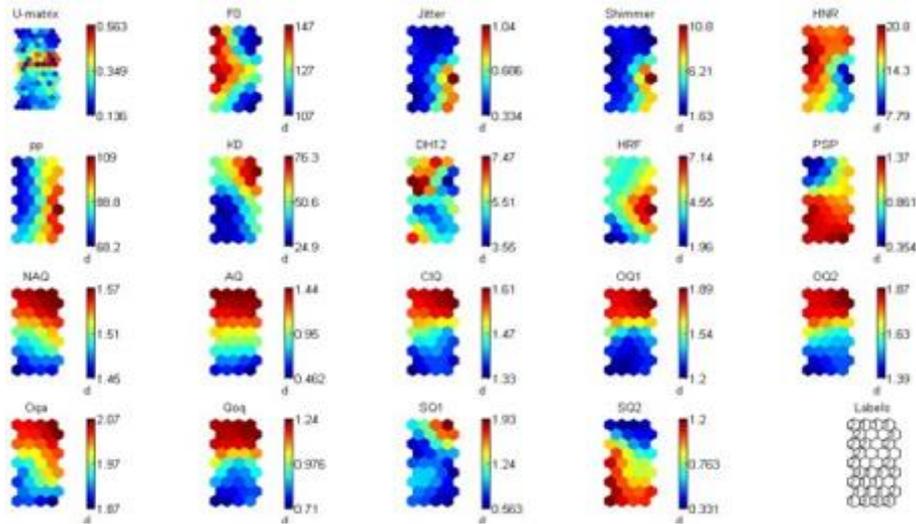


Fig. 2. Resultado locutores masculinos.

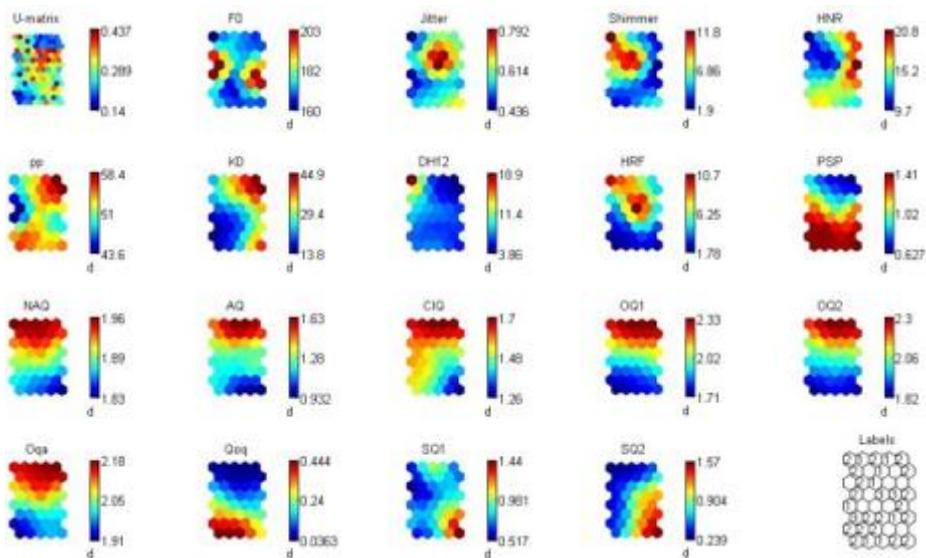
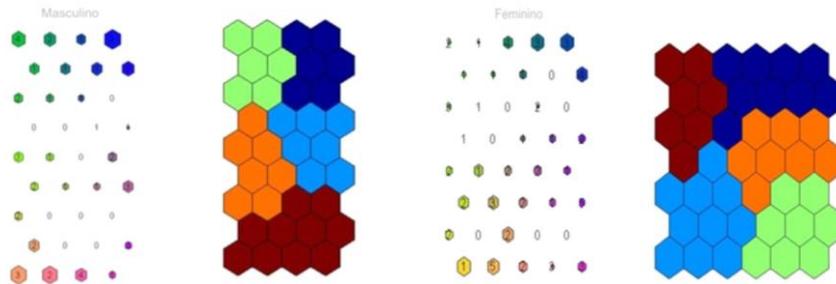


Fig. 3. Resultado locutores femininos.

Para utilizar esta ferramenta o número de grupos desejados pode ser fornecido ou o próprio algoritmo detecta automático. O número de iterações deve ser alto, algo em torno de 5000 iterações. O tamanho dos neurônios corresponde a maior proximidade entre eles, o número interno do neurônio a quantidade de vezes que este foi acionado e a cor a formação dos grupos [15].

Os resultados utilizando o *K-means* (Fig.4) para ambos os sexos permitindo ao algoritmo detectar o número de grupos automático. Os resultados apresentaram cinco grupos para ambos os sexos o que possibilitou dividir a base de dados em cinco faixas de características da voz.



**Fig. 4.** (a) Resultado *K-means* dos locutores masculino e locutores femininos (b).

No método *Fuzzy c-Means*, para analisar os resultados, utilizaram-se três locutores masculinos e femininos, que apresentavam ótimo estado de saúde e cujos parâmetros extraídos estavam de acordo com a literatura, para as análises dos graus de pertinência.

Para possibilitar uma comparação entre os métodos, os parâmetros que foram descartados no método SOM também foram para o método FCM utilizando a mesma base de dados.

Como visto anteriormente, no FCM é necessário especificar a quantidade de grupos. Por ter dividido inicialmente a base de dados em três tipos de faixas etárias, o primeiro valor sugerido para o número de grupos foi igual três. Utilizando três grupos, alguns locutores apresentaram valores de pertinência próximos para todos os grupos formados. Esses locutores foram classificados como *outliers* (pontos fora do padrão) por não apresentar o valor limite de pertinência sugerido e impossibilitar a definição do locutor a um grupo.

Sendo *Fuzzy* um método difuso que permite um indivíduo possa pertencer a mais de um cluster com diferentes graus de pertinência, esses locutores apresentam a transição de sua idade vocal. Mas o que se pretende com esse trabalho é utilizar a idade vocal dos locutores na formação dos grupos. Assim sendo, variou-se a quantidade de grupos até que todos os locutores estivessem definidos a algum grupo. O resultado que classificou todos os locutores foi utilizando cinco grupos, para um coeficiente *fuzzy*  $m=1,2$  com *threshold*=0.

A divisão dos grupos, masculino e feminino, ficaram da seguinte forma:

- grupo 1 – 15 a 25 anos - Jovem
- grupo 2 – 26 a 40 anos - Jovem→Adulto

- grupo 3 – 41 a 55 anos - Adulto
- grupo 4 – 56 a 75 anos - Adulto→Idoso
- grupo 5 – 76 a 100 anos - Idoso

As Tabelas 1 e 2 mostram os resultados com três e cinco grupos. Com três grupos (n=3) vêm-se os locutores que apresentaram transição entre grupos e com cinco grupos (n=5) esse locutores foram agrupados. Nas faixas etárias 1 e 3 os locutores *outliers* masculinos foram agrupados no grupo 4 enquanto que os da faixa etária 2 foram para os grupos 4 e 5 (transição de Adulto→Idoso e Idoso respectivamente). Os locutores *outliers* femininos das faixas etárias 2 e 3 foram agrupados no grupo 5 (Idoso).

**Tabela 1.** Resultados locutores Masculinos (m=1,2).

n=3		n=5	
Faixa Etária 1	19 locutores	Faixa Etária 1	19 locutores
Jovens	7	Jovens	4
Adultos	3	Jovens - Adultos	1
Idosos	7	Adultos	7
Outliers	2	Adultos - Idosos	4
-	-	Idosos	3
-	-	Outliers	0

Faixa Etária 2	19 locutores	Faixa Etária 2	19 locutores
Jovens	8	Jovens	4
Adultos	6	Jovens - Adultos	7
Idosos	3	Adultos	4
Outliers	2	Adultos - Idosos	3
-	-	Idosos	1
-	-	Outliers	0

Faixa Etária 3	14 locutores	Faixa Etária 3	14 locutores
Jovens	1	Jovens	0
Adultos	5	Jovens - Adultos	4
Idosos	6	Adultos	4
Outliers	2	Adultos - Idosos	2
-	-	Idosos	4
-	-	Outliers	0

**Tabela 2.** Resultados locutores Femininos (m=1,2).

<b>n=3</b>		<b>n=5</b>	
Faixa Etária 1	15 locutores	Faixa Etária 1	15 locutores
Jovens	1	Jovens	2
Adultos	6	Jovens - Adultos	3
Idosos	8	Adultos	4
Outliers	0	Adultos - Idosos	5
-	-	Idosos	1
-	-	Outliers	0

Faixa Etária 2	38 locutores	Faixa Etária 2	38 locutores
Jovens	6	Jovens	5
Adultos	13	Jovens - Adultos	7
Idosos	18	Adultos	11
Outliers	1	Adultos - Idosos	9
-	-	Idosos	6
-	-	Outliers	0

Faixa Etária 3	17 locutores	Faixa Etária 3	17 locutores
Jovens	0	Jovens	3
Adultos	9	Jovens - Adultos	2
Idosos	7	Adultos	1
Outliers	1	Adultos - Idosos	10
-	-	Idosos	1
-	-	Outliers	0

## 6 Conclusões

Este trabalho apresentou uma metodologia que explora as características e propriedade dos métodos mencionados anteriormente, para realizar visualização de agrupamentos de dados. Os resultados foram bem significativos, uma vez que foi possível formar grupos com os parâmetros extraídos dos sinais de voz e glotal.

Utilizando o método *K-means* juntamente com o SOM alguns indivíduos foram classificados em grupos diferentes independente da sua pré-classificação, de acordo com a idade. Alguns locutores apresentaram boa saúde vocal para sua faixa etária e outros não [3].

O método *Fuzzy c-Means* permitiu visualizar clusters identificando as faixas etárias, utilizando os parâmetros do sinal glotal, e locutores cuja saúde vocal estava em transição. Foi possível classificar todos locutores quando o número dos grupos foi alterado de três para cinco, o que era o nosso objetivo.

Apesar de *Fuzzy c-Means* não ser um método de inferência, depois dos grupos formados, pode-se criar regras utilizando o resultado obtido para permitir a classificação de locutores.

A utilização dos parâmetros dos sinais de voz e glotal, que geralmente são usados para identificar patologias diversas, mostraram-se bons discriminantes na criação dos grupos, identificando a idade vocal dos locutores independente da idade cronológica.

## References

1. I. S. Rosa, “Análise acústica da voz de indivíduos na terceira idade”. Dissertação de mestrado - Universidade de São Carlos 2005.
2. J. M. Silva, “Um estudo comparativo entre o sinal eletroglotográfico e o sinal de voz”, Dissertação de mestrado - Universidade Federal Fluminense, UFF 2008.
3. M. Silva, “Uma contribuição para caracterização do sinal de voz envelhecida”. Dissertação de mestrado - Universidade Federal Fluminense, UFF 2010.
4. H. Pulakka, “Analysis of Human Voice Production Using Inverse Filtering, High-Speed Imaging, and Electroglottography”. University of Technology Helsinki. 2005.
5. M. N. Vale, “Agrupamentos de Dados: Avaliação de Métodos e Desenvolvimento de Aplicativos para Análise de Grupos”. Dissertação de Mestrado – Pontifícia Universidade Católica do Rio de Janeiro - PUC - 2005.
6. N. L. C. Júnior, “Clusterização baseada em Algoritmos Fuzzy”. Dissertação de Mestrado – Universidade Federal de Pernambuco, Recife 2006.
7. S. Laine, “Using Visualization, Variable Selection and Feature Extration to Learn from Industrial Data”, Dissertation for the degree of Doctor, University of Technology Helsinki. 2003.
8. E. A. Botter, C. L. N. Júnior, T.Yoneyama “Redes Neurais auto-organizáveis para classificação de sinais eletrocardiográficos atriais”. *Integração*, N: 40 pp.51-56, 2005.
9. A. Kohler, M. Ohrnberger, F. Scherbaum, “Unsupervised feature selection and general pattern discovery using Self-Organizing Maps for gaining insights into the nature of seismic wavfields”. *Computer & Geosciences* 35, pp: 1757-1767, 2009.
10. G. Fung, “A Comprehensive Overview of Basic Clustering Algorithms”, Citeseer, 2001 - <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.5037&rep=rep1&type=pdf>.
11. J. C. Bezdek, “Pattern Recognition with Fuzzy Objective Function Algorithms”, New York, 1981.
12. R. J. Sassi, “Uma Arquitetura Híbrida para Descoberta de Conhecimento em Bases de Dados: Teoria dos Rough Sets e Redes Artificiais Mapas Auto-Organizáveis”. Tese de Doutorado – Escola Politécnica da Universidade de São Paulo – São Paulo 2006.
13. SOMToolbox,  
[HTTP://www.cis.hut.fi/projects/somtoolbox/package/somtoolbox2\\_Mar\\_17\\_2005.zip](http://www.cis.hut.fi/projects/somtoolbox/package/somtoolbox2_Mar_17_2005.zip)
14. S. E. Linville, “The Sound of Senescence”, *Journal of Voice*, Vol.10, No.2, pp.190-200, 1996
15. M. A. Loureiro, “Timbre de um instrumento musical: caracterização e respresentação”, *Per Musi – Revista Acadêmica de Música*, No.14, pp.57-81, 2006.