

## UM EXEMPLO DO USO DA ABP NA DISCIPLINA DE PROCESSAMENTO DIGITAL DE SINAIS - ANÁLISE TEMPORAL DO SINAL DE VOZ

Glauber Inocência Feitosa de Carvalho – [glauber.inocencio@gmail.com](mailto:glauber.inocencio@gmail.com)  
Marla Vitória Melo Monteiro – [marlamonteiro19958@gmail.com](mailto:marlamonteiro19958@gmail.com)  
Anne Karoline Pontes de Macêdo – [akpmacedo@gmail.com](mailto:akpmacedo@gmail.com)  
Maria Eduarda Noberto dos Santos – [m.eduardanoberto@hotmail.com](mailto:m.eduardanoberto@hotmail.com)  
Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB) –  
Campus João Pessoa  
Av. Primeiro de Maio, 720, Jaguaribe  
55050-900 – João Pessoa – Paraíba

**Resumo:** Este trabalho trata do estudo e implementação de técnicas de processamento digital de sinais que são utilizadas em sistemas de resposta vocal, com a finalidade de estudar as formas de ondas da voz humana, utilizando as ferramentas computacionais GoldWave e Matlab em aspectos temporais. Foi utilizada a gravação da palavra "aplausos". Os sinais foram segmentados para observar as características do som, levando em consideração a frequência, energia e a entropia. Esta análise dá subsídios para desenvolver sistemas de reconhecimento de voz, detecção de início e fim de palavras além de permitir manipulações em sinais de voz.

**Palavras-chave:** Sinais de voz. Energia. Entropia. Matlab. GoldWave.

### 1 INTRODUÇÃO

Dentre as várias áreas que compõem o campo da comunicação por voz, a área da comunicação vocal homem-máquina é uma das mais interessantes e estimulantes. O desejo bem como a necessidade das pessoas se comunicarem com as máquinas através da forma mais natural de comunicação, a voz humana, tem dado grande impulso ao crescimento desta área (RABINER, 1978).

A voz está presente na maioria dos meios de comunicação e diante do avanço tecnológico em que o mundo se encontra, a digitalização da voz é uma técnica imprescindível. Para tanto, deve-se conhecer minuciosamente todas as características de cada pronúncia que a voz humana é capaz de realizar, para que seja possível desde uma conversa telefônica até um controle de ativação por sinal de voz.

Neste artigo, o objetivo é analisar como cada segmento dos sinais de voz se comporta com relação à energia, frequência e entropia, perceber que estas características são, em conjunto, únicas para cada fonema, se tornando a identidade de cada um e a informação necessária para digitalização e detecção de início e fim de palavras.

O processo de análise foi dividido em duas partes: a teoria e a prática, contando com os softwares GoldWave para gravação de sons e Matlab para processamento destes sinais,

através de algoritmos computacionais e implementação das funções espectrais e temporais dos sinais de voz.

Neste trabalho será apresentado o processamento de sinais de voz com intuito de fazer uma análise temporal tendo como metodologia de ação a Aprendizagem baseada em problemas (ABP).

## 2 APRENDIZAGEM BASEADA EM PROBLEMAS (ABP) APLICADA À DISCIPLINA DE PROCESSAMENTO DIGITAL DE SINAIS (PDS)

A ABP teve sua primeira aplicação como método de ensino na universidade de MacMaster no Canadá, no ano de 1969 no curso de medicina (SALES *et al.*, 2013). Tal método se fez necessário devido à percepção de que muitos estudantes concluíam o curso sem possuir as habilidades necessárias para atuação como profissional de medicina, visto que não conseguiam aplicar os conhecimentos teóricos adquiridos à obtenção de diagnósticos em casos reais (RIBEIRO, 2004), permitindo assim que aluno desenvolva sua autonomia para solução de problemas.

Segundo Barrows *apud* Ribeiro (2004) o processo para a aplicação da ABP é dado da seguinte forma:

- Um problema é apresentado afim de que os alunos tentem resolvê-lo com base em seus próprios conhecimentos.
- Em seguida são levantadas questões sobre os aspectos do problema que não foram compreendidos.
- Tais questões são investigadas e compartilhadas com a equipe.
- Logo após, o conhecimento adquirido através da investigação das questões é incorporado ao contexto do problema a ser resolvido.
- Após a conclusão do trabalho e solução do problema, os alunos realizam uma avaliação do processo, dos seus colegas e deles mesmos de forma construtiva, com a finalidade de tornar mais eficaz o processo de aprendizagem e consolidação do conhecimento.

Nesse contexto, a tabela 1 mostra como adoção do elemento fundamental a proposta elaborada por Hadgraft e Pripic *apud* Ribeiro (2004), no formato 4-2-3-4-2 que auxiliará o professor na identificação do método mais adequado para a disciplina de Processamento Digital de Sinais, que irá guiá-lo para o formato ABP ideal (4-4-4-4-4).

Tabela 1: Elementos básicos da ABP

Passo	Problema	Integração	Trabalho em Equipe	Solução de Problemas	Aprendizagem Autônoma
1	Vários problemas por semana	Nenhuma ou pouca integração de conceitos.	Trabalho Individual	Nenhum método Formal de Problemas.	Professor fornece todo o conteúdo e os alunos concentram-se em aprender o que lhes foi dado

2	Um Problema por Semana	Alguma integração de conceitos	Alunos trabalham juntos em sala de aula, mas produzem trabalhos individuais	Método formal de solução de problemas	Professor o conteúdo, mas espera que os alunos investiguem alguns detalhes por conta própria
3	Mais de um Problema por Semestre. Cada um com duração de algumas semanas	Integração dos conceitos e habilidades na solução do problema	Relatório em conjunto, porém sem avaliação dos pares	Método formal para solucionar problemas, mas com orientação de tutores	Professor fornece um livro-texto como base, mas espera que os alunos utilizem esta e outras fontes
4	Um Problema por semestre	Grande integração, e interdisciplinaridade	Trabalho em equipe formal, encontro externo entre as equipes, avaliação por pares, relatório e apresentação dos resultados em conjunto	Método formal de solução, aprendizagem de problemas. Alunos aplicam este método sozinhos a cada novo problema	Professor fornece pouco ou nenhuma matéria. Alunos utilizam outras fontes para chegarem à compreensão do problema

Fonte: Elaborado pelo autor

Na disciplina de Processamento Digital de Sinais o método ABP é implementado pelo professor mostrando diversas áreas de conhecimento em que pode ser aplicado o processamento digital de sinais, como por exemplo: nas telecomunicações, com a utilização do processamento de imagem, áudio e dados; na proteção de direitos autorais através de *watermarking*; na medicina, com o processamento de sinais de voz, cardíacos, cerebrais, entre outros, para diagnóstico de doenças. Essa apresentação traz um esclarecimento sobre os conteúdos que serão abordados na disciplina e como eles podem ser utilizados fora de sala de aula, tornando a disciplina mais interativa e atrativa, melhorando inclusive a percepção do aluno no diagnóstico de problemas e levantamento de soluções.

### 3 CONCEITO DE SOM

O universo é um sistema dinâmico definido a cada momento pelas diversas formas que a energia apresenta. O som é uma dessas formas de energia, e como tal obedece às leis fundamentais da física. Dessa forma, Bonjorno (1992) descreve o som como aumentos e reduções periódicas da densidade do ar, ou seja, compressão e rarefação. Recorrendo à definição dada pela ABNT (1959), o som é "toda e qualquer vibração ou onda mecânica em um meio elástico dentro da faixa de audiofrequência".

Podemos acrescentar à descrição dada por Bonjorno (1992), na qual descreve que o som se propaga através de um meio elástico por uma onda longitudinal. Essas ondas transportam uma perturbação vibratória ao longo do espaço em todas as direções (em material elástico).

#### **4 SINAIS DE VOZ**

Sinais de voz são sequências de sons e suas respectivas transições que representam simbolicamente a informação. A voz é uma característica humana intimamente relacionada com a necessidade do homem de se agrupar e se comunicar. A produção vocal é fruto de um trabalho em conjunto do sistema nervoso central e periférico, além da atuação de receptores sensoriais (BEHLAU, 2001).

A voz está associada à fala, na realização da comunicação verbal, e pode variar quanto à intensidade, altura, inflexão, ressonância, articulação e muitas outras características. Ela é produzida quando o ar passa pelas pregas vocais fazendo com que elas vibrem, essa vibração resulta na produção de um som, que por sua vez é modificado através das cavidades acima e abaixo das pregas vocais, tais cavidades recebem o nome de cavidades de ressonância (BEHLAU, 2001).

#### **5 O MECANISMO DE PRODUÇÃO DA FALA**

A voz humana é um sinal produzido por vibrações que são resultados das transformações que ocorrem em diferentes níveis: semântico, linguístico, articulatório e acústico. O trato vocal, nome genérico dado ao conjunto de cavidades e estruturas que participam diretamente da produção sonora, começa na abertura entre as cordas vocais ou glote e termina nos lábios.

O trato vocal assim consiste da faringe (a conexão entre o esôfago e a boca), laringe e termina na boca ou cavidade oral, de forma que órgãos como língua, dentes, e lábios também contribuam para a geração da voz humana. Estas devem ter uma tonicidade "ideal" para que sua movimentação ocorra normalmente. Uma respiração ideal provoca o alargamento da base dos pulmões, permitindo maior entrada de ar e menor contração da musculatura torácica, pescoço, mandíbula e língua (MATUCK, 2005).

No processo natural de produção da fala, alguns fatores são diretamente envolvidos com a qualidade e a inteligibilidade do som. Esses fatores contam com o correto funcionamento do cérebro, pulmões, ouvido, laringe, além de outras estruturas responsáveis pela produção e percepção da fala (SEARA, 2011).

A frequência é um dos parâmetros muito observados na voz, a frequência natural da voz é determinada, em grande parte, pelo comprimento das cordas vocais. As mulheres têm voz mais aguda que os homens porque suas pregas vocais são mais curtas de forma que quanto mais aguda é a voz, maior é a frequência (mulheres e crianças) e quanto mais grave for a voz menor a frequência (homens).

#### **6 TIPOS DE SONS DA FALA**

A voz é som, logo, possui todas as características sonoras do mesmo. Sobre a voz humana pode-se dizer que, "consiste em um som ou um conjunto de sons emitidos pelo aparelho fonador" (FERREIRA, 1986).

Os sons da voz são classificados como: sonoros, surdos, plosivos, fricativos e excitação mista. Os sons além de distinguirem-se pela presença ou ausência de vozeamento (quando há vibração das pregas vocais), também podem ser diferenciados por classes, dependendo do modo de articulação.

## 6.1 Sons sonoros

Na produção dos sons sonoros são obtidas ondas de pressão quase periódicas excitando as cordas vocais, atuando como um ressonador e modificando o sinal de excitação, produzindo frequências de ressonância que caracterizarão os diferentes sons sonoros.

## 6.2 Sons surdos

Na geração dos sons surdos, é produzida uma constrição em algum ponto do trato vocal (usualmente próximo ao final da boca), assim o ar adquire velocidade suficientemente alta para produzir turbulência gerando uma fonte de ruído de espectro largo (semelhante ao ruído branco) para excitar o trato vocal (RABINER, 1978).

## 6.3 Sons plosivos

Na geração dos sons plosivos, o ar é totalmente dirigido à boca, estando esta completamente fechada. Com o aumento da pressão, a oclusão é rompida bruscamente, gerando um pulso que excita o aparelho fonador. Com a excitação ocorre um movimento rápido dos articuladores em direção à configuração do próximo som (FECHINE, 2000).

# 7 CLASSIFICAÇÃO DOS SONS

## 7.1 Vogais

Os sons correspondentes às vogais são normalmente vozeados e produzidos com o trato vocal numa forma fixa. Nas vogas não há impedimento da passagem do ar, o fluxo de ar passa livre no trato vocal (SEARA, 2011).

## 7.2 Oclusivos ou explosivos

Quando é gerado um pulso que excita o aparelho fonador, a esse som é dado o nome de explosivo. Os oclusivos são sons produzidos pela constrição total do trato vocal, seguida da libertação da pressão acumulada. Estas últimas apresentam uma zona de oclusão com um silêncio quase total, enquanto que os segmentos oclusivos vozeados mantêm a periodicidade dos segmentos vizinhos (SEARA, 2011).

## 7.3 Fricativos

Fricativos são os sons nos quais o fluxo de ar é constringido em algum ponto do trato vocal, elevando-se a língua em direção ao palato, como por exemplo o lábio inferior contra os dentes superiores, tornando-se turbulento e produzindo um ruído de amplo espectro (RUSSO e BEHLAU apud COSTA, 2004).

# 8 FREQUÊNCIA FUNDAMENTAL

Habitualmente designada por frequência de *pitch*, está relacionada à percepção humana de sinais acústicos. A frequência fundamental depende da dimensão e espessura da glote. Para oradores do gênero masculino, a gama de vibração das pregas vocais situa-se nos 150-250 Hz (BEBER E CIELO, 2011), enquanto que para oradores do gênero feminino essa gama situa-se

nos 250 Hz (SCARPEL E FONSECA, 2014), ultrapassando os 250 Hz para as crianças (BRAGA, OLIVEIRA E SAMPAIO, 2009).

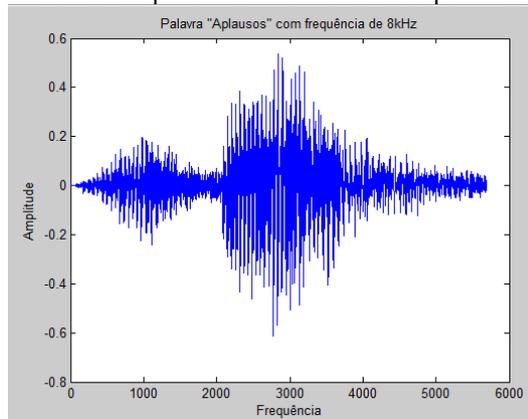
O espectro da voz humana típica ocupa a faixa de 7 - 8 kHz, mas a maior parte da energia e da inteligibilidade encontra-se na faixa de até 3,5 kHz, portanto pode-se utilizar um processo de amostragem de 8 kHz (COSTA, 2006).

## 9 ESTUDO COMPARATIVO DAS FREQUÊNCIAS DE AMOSTRAGEM

Neste trabalho, para a aquisição dos sinais de voz foi utilizado um microcomputador PC com placa de som e microfone comum. O software usado para gravação do som em mídias digitais foi o GoldWave na versão 4.26 e o formato de gravação escolhido foi o padrão WAV. Este formato de gravação é um dos mais utilizados para este tipo de aplicação e contém um cabeçalho de 44 bytes com informações sobre o próprio arquivo.

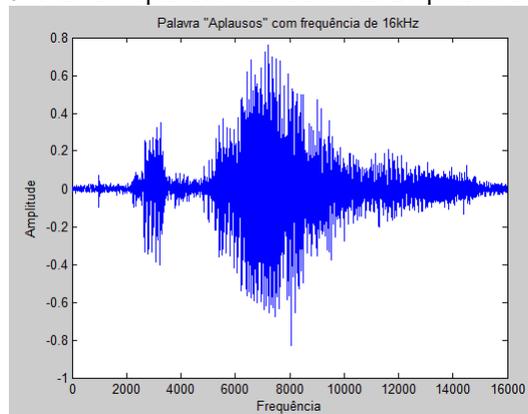
Com a obtenção do arquivo WAV contendo o sinal de voz na forma digital, pode-se agora manipulá-lo no ambiente de processamento digital de sinais implementado. A palavra gravada para análise foi ‘Aplausos’. Após a gravação de três sinais de voz distintos e cada um em três frequências de amostragem distintas, 8 kHz (Figura 4), 16 kHz (Figura 5) e 44,1 kHz (Figura 6) respectivamente, pôde-se observar a influência de cada frequência de amostragem nas características dos sinais, tanto como suas vantagens e desvantagens.

Figura 4: Palavra ‘aplausos’ amostrada na frequência de 8kHz



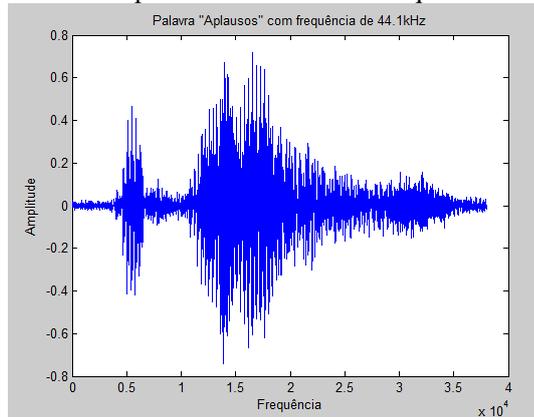
Fonte: Elaborada pelo autor

Figura 5: Palavra ‘aplausos’ amostrada na frequência de 16kHz



Fonte: Elaborada pelo autor

Figura 6: Palavra ‘aplausos’ amostrada na frequência de 44,1 kHz



Fonte: Elaborada pelo autor

Observa-se que à medida que a frequência aumenta, a qualidade do sinal melhora, pois a informação é bem maior e esta é, a priori, uma vantagem quando se deseja obter o máximo de informação de um sinal. Porém, com o aumento da frequência, a energia diminui e, conseqüentemente, a inteligibilidade torna-se menor, pois a maior parte desta está concentrada nas baixas frequências. Para estudar algumas das diversas características que um sinal de voz pode ter, os sinais gravados no software GoldWave foram então lidos no software Matlab e estudados nos casos em que a frequência de amostragem era de 8kHz fazendo uso da seguinte função:

```
[x, frequencia] = audioread('aplausos8kHz.wav')
plot(x);
```

Esta lê o sinal no formato WAV e retorna a frequência de amostragem para verificar se é a mesma em que foi gravada. O arquivo no formato WAV pode ser convertido também para formato de texto. Desta forma, pode-se obter um resultado de fácil manuseio no decorrer do programa.

## 10 ENERGIA DA VOZ

Pode-se dizer que existe uma concentração maior de energia acústica em determinadas faixas de frequências e isto acontece, geralmente, nas frequências mais baixas. A energia segmental é um parâmetro muito utilizado para diferenciação entre segmentos surdos e sonoros do sinal de voz, pois a amplitude dos sons surdos é menor que dos sonoros. Para calcular a energia do sinal de voz é necessário dividi-lo em intervalos de curta duração e calcular a energia de cada segmento utilizando a Equação 1 (COSTA, 2006)

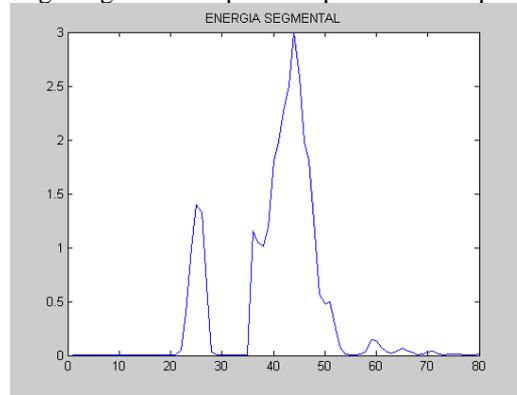
$$E_{seg} = NA \cdot E\{[s(n) - \mu_s(n)]\}^2 \quad (1)$$

Em que, NA representa o número de amostras,  $s(n)$  representa o sinal de voz e  $\mu_s(n)$  a média do sinal de voz. O próximo passo de análise é também a partir de uma análise espectral, a energia segmental do sistema.

Como se pode observar na Figura 7, é entre os segmentos 45 e 50 que se concentra a maior parte na energia, isto porque é entre os pontos 2500 e 3000 onde ocorre o som mais forte da palavra na frequência de 8kHz. Analogamente, percebe-se que a energia nos

primeiros e últimos segmentos é praticamente nula, pois nestes pontos ocorre o silêncio e sons surdos, respectivamente.

Figura 7: Energia segmental da palavra aplausos na frequência de 8kHz



Fonte: Elaborada pelo autor

## 11 ENTROPIA DOS SINAIS - ENTROPIA DE SHANNON

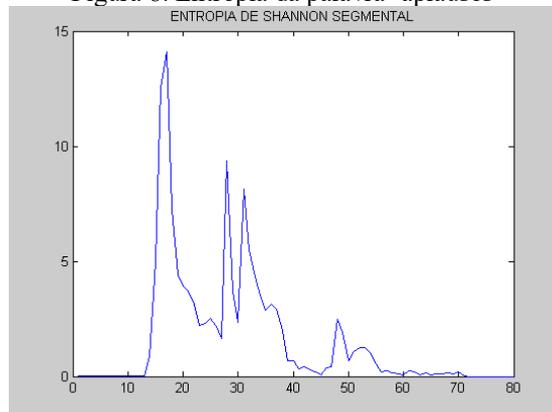
Entropia é uma grandeza termodinâmica geralmente associada ao grau de desordem. Ela mede a parte da energia que não pode ser transformada em trabalho. É uma função de estado cujo valor cresce durante um processo natural em um sistema fechado.

Logo a entropia de Shannon pode ser definida como uma medida de incerteza probabilística associada a uma distribuição de probabilidade. A noção de entropia está ligada ao grau de desorganização existente na informação e pode ser calculada pela Equação 2 (ARTUSO, 2011):

$$H = - \sum_{i=1}^n p_i \cdot \log_2(p_i) \quad (2)$$

Em que H representa a entropia, p é a função de probabilidade e i é o número de classes da função de probabilidade.

Figura 8: Entropia da palavra ‘aplausos’



Fonte: Elaborada pelo autor

Ao analisar de forma segmental a entropia de Shannon do sinal, observa-se, a partir da Figura 8, que o maior índice de entropia está no início da pronúncia da palavra, a vogal /a/, já

que a mesma, neste caso, retorna uma maior compreensão em comparação ao resto na palavra pronunciada e traz uma grande quantidade de informação a ser transmitida.

Até o presente momento, a comparação na palavra aplausos, do início da pronúncia com o fim é de grande importância porque estes casos representam um bom exemplo comparativo por serem de características opostas. Assim, a entropia de Shannon também traz características marcantes na pronúncia do /s/ no final, pois este som pode ser confundido com um ruído e, assim, tem uma baixa entropia.

## 12 CONSIDERAÇÕES FINAIS

A utilização do método ABP fez com que o processo de aprendizagem e assimilação dos conteúdos propostos na disciplina de Processamento Digital de Sinais fosse mais eficaz. Tal metodologia permitiu aos alunos o contato com um problema real, aperfeiçoando nos discentes a habilidade de solucionar de problemas, que é essencial não apenas para o profissional de Engenharia Elétrica, mas para profissionais de qualquer outra área também.

Por meio do estudo das características de energia, entropia e análise de frequência fundamental realizados no presente trabalho pode-se perceber que é possível fazer a identificação de início e fim de palavras de forma eficaz para ambos os sexos. Tal análise faz com que as perspectivas de implementação de comando de voz não se limitem apenas a observar o conteúdo de uma só palavra, mas uma complexidade do conjunto em geral.

## REFERÊNCIAS

ARTUSO, A.R. **Entropias de Shannon e Rényi aplicadas ao reconhecimento de padrões.** Artigo, Revista CIATEC – UPF, v.3, n.2, p 56-72, Paraná, 2011.

BEBER, B.C.; CIELO, C.A. **Características vocais acústicas de homens com voz e laringe normal.** Artigo, Revista CEFAC, v.13, n.2, p. 340-351, Rio Grande do Sul, 2011.

BEHLAU, M. AZEVEDO, R. **Voz: o livro do especialista.** Livro, Volume 1, Editora Revinter, São Paulo, 2013.

BONJORNO, Clinton. **Física 2.** 1992.

BRAGA, J.N; OLIVEIRA, D.S.F; SAMPAIO, Tânia M.M. **Frequência fundamental da voz de crianças.** Artigo, Revista CEFAC, v.11, n.1, p. 119-126, 2009.

COSTA, S.L.N.C. **Processamento Digital de Sinais aplicado à oralização de surdos.** UNIVERSIDADE FEDERAL DE CAMPINA GRANDE. Projeto de Pesquisa, 2004.

COSTA, S. L. N. C; PASSOS, M. G.; SILVA, P.L.; NETO, B.G.A.; FECJINE, J. M. **Um ambiente para Processamento digital de sinais aplicado à comunicação vocal homem-máquina.** Artigo, Revista Princípia, Instituto Federal de Educação Ciência e Tecnologia da Paraíba, João Pessoa - PB, Brasil, n.14, p. 25 – 31, 2006.

FECHINE, J. M. **Reconhecimento automático de identidade vocal utilizando modelagem híbrida: paramétrica e estatística.** Tese de doutorado, Universidade Federal de Campina Grande, Campina Grande - PB, Brasil, 2000.

FERREIRA. **Novo dicionário da língua portuguesa.** 1986.

MATTOS, R.S.;VEIGA, A. **Otimização de entropia: implementação computacional dos princípios maxent e minxent.** Artigo, Pesquisa Operacional, v.22, n.1, Rio de Janeiro, 2002.

MATUCK, G. R. **Processamento de sinais de voz na identificação de padrões comportamentais por redes neurais artificiais.** Disponível em: <<http://mtc-m16.sid.inpe.br/col/sid.inpe.br/iris@1916/2005/09.27.11.38/doc/processamento%20sinais.pdf>>. Acesso em 12 fev. 2018, às 17:35h.

RABINER, R.W; SCHAFER, L. **Digital Processing of Speech Signals.** Livro, Editora Prentice-Hall, New Jersey, 1978.

RIBEIRO, L.R.C.; ESCRIVÃO FILHO, E.; MIZUKAMI, M. G. N. **Uma Experiência com a PBL no Ensino de Engenharia Sob a Ótica dos Alunos.** Artigo, Revista de Ensino em Engenharia, v 23, n.1, p.63-71, 2004.

SCARPEL, R.D.; FONSECA, M.D.L. **Parâmetros acústicos de vozes de mulheres na pós-menopausa.** Artigo, Revista Brasileira de Geriatria e Gerontologia, v.17, n.4, p.741-750, Rio de Janeiro, 2014.

SEARA, Izabel C; NUNES, Vanessa G; VOLCÃO, Cristiane L. **Fonética e fonologia do português brasileiro.** Apostila, UFSC, Santa Catarina, 2011.

## **AN EXAMPLE OF THE USE OF ABP IN THE DIGITAL PROCESSING DISCIPLINE OF SIGNALS - TEMPORAL ANALYSIS OF THE VOICE SIGN**

***Abstract:** This work deals with the study and implementation of digital signal processing techniques that are used in vocal response systems, with the purpose of studying the waveforms of the human voice, using the GoldWave and Matlab computational tools in temporal aspects. The recording of the word "applause" was used. The signals were segmented to observe the characteristics of the sound, taking into account frequency, energy and entropy. This analysis provides insights to develop voice recognition systems, word start and end detection, and allow manipulation of voice signals.*

***Key words:** Speech signals. Energy. Entropy. Matlab. GoldWave.*