

DESENVOLVIMENTO DE ALGORITMO DE PROCESSAMENTO DE SINAIS E CRIAÇÃO DE UM BANCO DE FONEMAS DO PORTUGUÊS BRASILEIRO

Anete Sayuri Kamimura Koga

anete.koga@gmail.com.br

Thaís Minato Takeuchi

thaisminato@gmail.com

Resumo: A ausência de implantes cocleares de fabricação nacional e o prejuízo no entendimento da fala por parte dos usuários brasileiros, causado pela utilização de implantes importados, motivou este trabalho. Seu propósito é desenvolver um algoritmo de processamento de sinais, a fim de construir um banco de fonemas do português brasileiro, que permitirá a avaliação da eficiência dos implantes existentes no mercado para falantes do português brasileiro e auxiliar na sua melhoria ou no desenvolvimento de um implante nacional.

A construção do banco de fonemas consiste basicamente na determinação dos principais formantes de cada fonema, isto é, as três frequências de maior intensidade dos gráficos gerados pela Transformada de Fourier das gravações dos fonemas feitas em formato WAV, às quais foi aplicado o comando *fft* no software Scilab.

Palavras chave: *cóclea, implante coclear, aparelho fonador, banco de dados, processamento de sinais.*

1. Introdução

O implante coclear, também conhecido como ouvido biônico, é um dispositivo eletrônico que permite a transmissão de sinais elétricos para o nervo auditivo. O usuário é capaz de detectar os sons através da decodificação destes sinais através do córtex cerebral.

Atualmente, existem quatro marcas aprovadas no Brasil: Cochlear (Cochlear Corp., Lane Cove, Austrália), MED-EL (MED-EL Corp, Innsbruck, Áustria), Advanced Bionics (Advanced Bionics Corp., Valencia, Califórnia, EUA) e Neurelec (MXM, Laboratoires SA, Côte d'Azur, França).

Esse tipo de implante funciona baseado em um banco de dados, no caso, um banco de fonemas que apresenta os principais formantes de cada fonema. Entretanto, devido às particularidades de cada língua, os implantes podem apresentar deficiências na compreensão da fala quando utilizados por falantes de outras línguas.

Em particular, o português do Brasil, possui características muito marcantes e únicas quando comparadas às demais línguas, como por exemplo, a presença das vogais nasais.

A Organização Mundial de Saúde estima que no Brasil, há 15 milhões de pessoas que sofrem de algum tipo de deficiência auditiva. Desses, 350 mil possuem deficiência severa, aos quais seria indicado o uso do implante coclear.

2. Transformada de Fourier

A Transformada de Fourier, em homenagem ao físico francês Joseph Fourier (1768-1830), é uma transformada integral que expressa uma função em termos de funções senoidais, decompondo-a em suas componentes elementares, seno e cosseno. Em outras palavras, é a soma de funções senoidais multiplicadas por coeficientes (amplitudes das ondas).

Essa ferramenta é um processo matemático extensamente utilizado hoje em dia em espectrometria na forma do engenhoso processo de cálculo digital simplificado, denominado FFT (Fast Fourier Transform).

Press et al (1994), enuncia que um processo físico pode ser descrito no domínio do tempo, por uma quantidade h como função do tempo, $[h(t)]$, ou no domínio da frequência, onde o processo é especificado por uma amplitude H como função da frequência, $[H(f)]$. É conveniente pensar em $[h(t)]$ e $[H(f)]$ como sendo duas representações diferentes da mesma função. Essa transição entre os domínios do tempo e da frequência é dada pelas Eq. (1) e Eq. (2).

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{2\pi i f t} dt \quad (1)$$

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{-2\pi i f t} df \quad (2)$$

A FFT corresponde à base para o desenvolvimento deste trabalho, pois é através dela, que os sons serão convertidos de funções no tempo para funções na frequência, permitindo a identificação dos principais formantes para a construção do banco de dados.

Em 1961, Donald D. Greenwood obteve empiricamente a função de Greenwood, que relaciona a posição das células ciliadas na cóclea e a frequência de estimulação nesta região. Com ela, é possível determinar a região apropriada para a colocação dos eletrodos. Assim, tem-se a localização das frequências ao longo da membrana basilar (Figura 1).

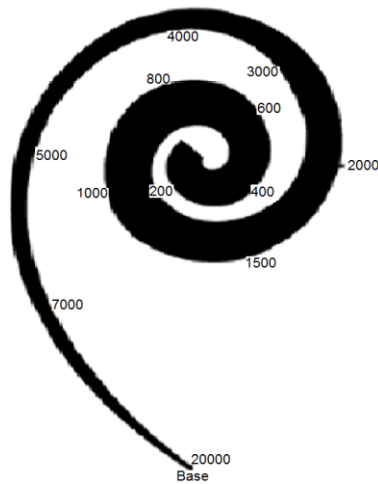


Figura 1 - Frequências na Membrana Basilar

3. Software

O Scilab foi o software utilizado no desenvolvimento deste trabalho, por ser um software aberto, isto é, distribuído gratuitamente via Internet desde 1994, disponível em 13 línguas, além de ser amplamente conhecido e utilizado em ambientes industriais e educacionais pelo mundo.

3.1. Aquisição do Sinal

Os sinais de áudio foram captados pelo próprio microfone do Notebook. Eles foram gravados com o uso do software livre Audacity. Este é um editor de áudio, capaz de gravar, reproduzir e importar/exportar sons nos formatos WAV, AIFF, MP3 e OGG. No caso, os sons foram exportados no formato WAV, para posterior processamento do sinal no programa de simulação numérica Scilab.

3.2. Código

```

clc
clear
[x,y] = loadwave('C:\Users\nome_usuario\Desktop\nome_arquivo.wav');
fx = fft(x(1:10000));
for i = 1:10000
fr(i) = (i-1)/(10000*(1/44100));
end
plot(fr,abs(fx));
mapsound(x(1,:))

```

Embora simples, este primeiro código de programação desenvolvido no Scilab fora suficiente frente aos objetivos iniciais do trabalho. Com ele, era possível importar um arquivo em formato WAV e gerar gráficos que ressaltavam suas principais frequências ou formantes.

Primeiramente, utilizou-se o comando “loadwave” para importar o arquivo WAV no Scilab. Em seguida, foi realizada uma Transformada de Fourier pelo comando “fft”, por um intervalo de 10 mil pontos, número este determinado empiricamente a fim de obter a melhor visualização dos dados procurados. Após a realização da FFT, o gráfico gerado apresentava a composição dos números imaginários do sinal, em valores absolutos. Foi também determinado um vetor de extensão do sinal, número de pontos amostrais, baseado na frequência em que o som foi gravado, 44100Hz, ou seja, 44100 pontos por segundo. Com isso, plotava-se o gráfico da Intensidade pela Frequência.

Outro comando colocado no final do programa é o “mapsound”. Seu princípio de funcionamento é o mesmo criado pela sequência de comandos anteriores. Ele aplica a FFT no sinal com um incremento de tempo de 0,1 segundo, e plota o espectrograma do som até o limite de 1500Hz. No espectro do som, há diferentes regiões que apresentam picos de intensidade (regiões de coloração mais intensa), sendo estas bandas de frequências (centros de energia) denominadas formantes (Matos et all). É possível observar que o resultado obtido pelo “mapsound” é exatamente o mesmo obtido com a sequência de comandos anteriores, porém com uma menor precisão de leitura.

4. Aplicação do Código para as Notas Musicais

Antes de realizar a gravação dos fonemas vocálicos, achou-se prudente fazer um prévio estudo com o uso de uma flauta doce, a fim de verificar se os resultados obtidos com a aplicação do código seriam coerentes. As sete notas musicais foram produzidas por uma flauta doce da marca Yamaha e gravadas no formato WAV pelo software Audacity.

A título de demonstração, serão apresentados apenas os gráficos gerados para a nota musical LA, Gráficos 1 e 2, a partir dos quais foi possível determinar suas principais frequências e compará-las com a literatura. Os resultados foram coerentes, validando o código para este caso.

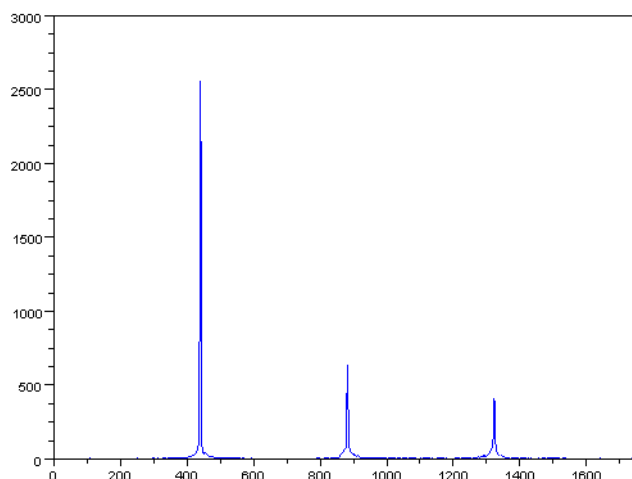


Gráfico 1 - Gráfico (Intensidade X Frequência) referente à nota musical LA, obtido através da FFT

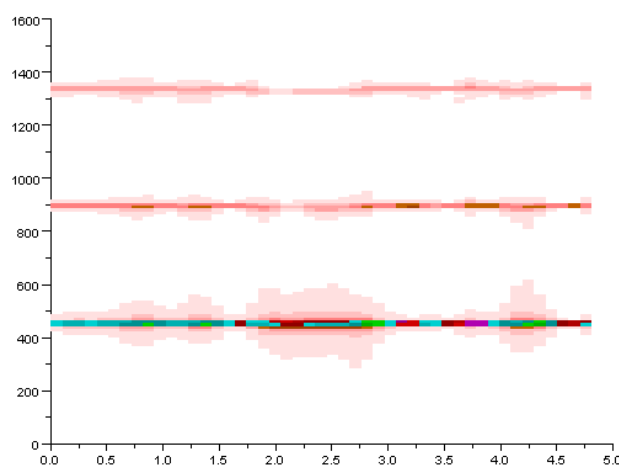


Gráfico 2 - Espectrograma referente à nota musical LA, obtido através do comando “mapprogram”

5. Fonemas do Português Brasileiro

O idioma português falado no Brasil é formado por 34 fonemas (Tabela 1), sendo 13 vocálicos (pronunciados sem obstáculos à passagem de ar), 19 consonantais (produzidos mediante a resistência que os órgãos bucais impõem à passagem de ar) e 2 semivocálicos (fonemas que se juntam a uma vogal, formando com esta uma única sílaba).

Tabela 1 - Listagem dos 34 fonemas constituintes da língua portuguesa falada no Brasil

Vocálicos orais	Vocálicos nasais	Semivocálicos	Consonantais (uma grafia)	Consonantais (diferentes grafias)
/a/	/ã/	/y/	/p/	/R/
/á/	/ẽ/	/w/	/b/	/z/
/e/	/ĩ/		/m/	/s/
/é/	/õ/		/f/	/j/
/i/	/ũ/		/v/	/x/
/o/			/t/	/g/
/ó/			/d/	/k/
/u/			/n/	
			/ɲ/	
			/l/	
			/ʎ/	
			/ɾ/	

6. Construção do Banco de Fonemas

Analogamente ao que foi feito para as notas musicais, foi realizada a análise dos fonemas vocálicos, dando início à tentativa de construção de um banco limitado de fonemas do português brasileiro. Limitado devido à restrição dos voluntários, residentes e nascidos em São Paulo. Portanto, não se abrangeu outros sotaques e faixas etárias que compõem a diversidade do país.

Os sons vocálicos foram também gravados com o software livre Audacity, em formato WAV, pelo próprio microfone do Notebook, para a então análise através do software Scilab.

Vários métodos de processamento e análise de sinais foram testados. Porém, aqui somente serão apresentados aqueles que geraram resultados mais satisfatórios e, portanto de maior relevância.

6.1. Método 3

O Método 3 foi importante, pois através dele verificou-se que não seria possível aplicar no processamento dos sons vocálicos o código numérico de processamento de sinais aplicado de maneira satisfatória para análise dos sons musicais.

A gravação de sílabas com repetições sucessivas, refinou de forma considerável os picos das frequências nos gráficos. Porém, concluiu-se que apesar de conseguir identificar bem as formantes predominantes na sílaba, o Método 3 não distinguia o que seriam as formantes da consoante da mesma. Assim, a característica captada com este método foi a vogal que acompanhava a consoante, e não a mesma. Isso porque a vogal possui maior presença temporal na sílaba do que a consoante.

A seguir apresentam-se gráficos de três sílabas diferentes, porém com resultados muito semelhantes, de sinais gravados por uma voluntária do sexo feminino.

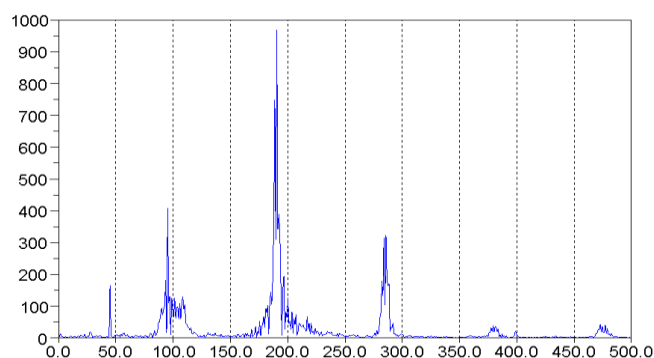


Gráfico 3 - Gráfico da intensidade x frequência para a sílaba /BE/

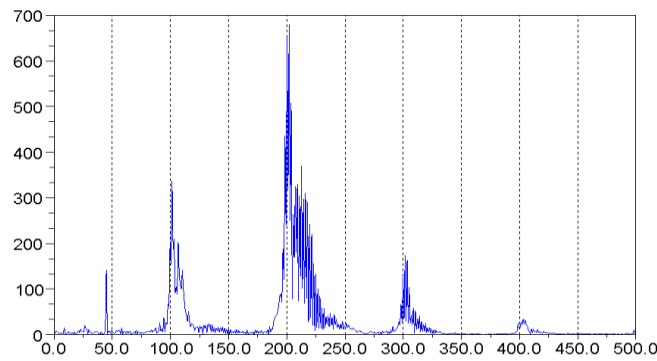


Gráfico 4 - Gráfico da intensidade x frequência para a sílaba /KE/

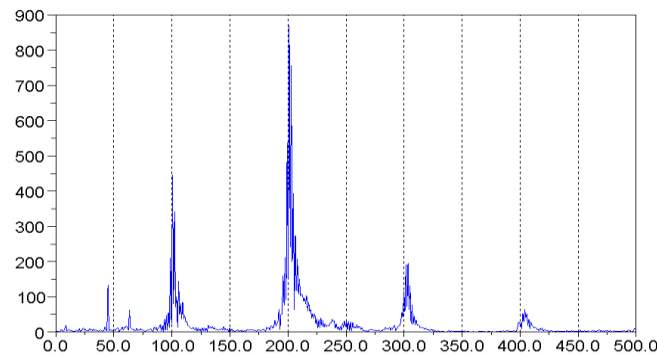


Gráfico 5 - Gráfico da intensidade x frequência para a sílaba /FE/

6.2. Método 4

Na tentativa de se obter resultados mais precisos na determinação das formantes das vogais, utilizamos um método denominado “zero padding”. Este consiste em acrescentar um “colchão de zeros” ao final do sinal. Com isso, são esperados gráficos que apresentem frequências com maior grau de clareza.

Assim, resumidamente, o seguinte código numérico: carrega o sinal que será analisado; lê seu tamanho; acrescenta um vetor de zeros de igual comprimento ao seu final; realiza a Transformada de Fourier; gera um gráfico da intensidade pela frequência da gravação captada.

```
-----
clc
clear
dt=1/44100.0;
[x,y]=loadwave('C:\Users\ nome_usuario\Desktop\ pasta\arquivo.wav');
lin=size(x);
df=1/(dt*lin(2));
completo=zeros(2*lin(2));
sinal=x(2,:);
for it=1:lin(2)
    completo(it)=sinal(it);
end
for it=lin(2)+1:2*lin(2)
    completo(it)=0.0;
end
fcompleto=fft(completo);
for i=1:2*lin(2)
    frequencia(i)=(i-1)*df;
end
plot(frequencia,abs(fcompleto));
-----
```

6.2.1. Comparação dos resultados com a referência

Nesta etapa, utilizou-se como referência a obra da Dra. Mara Behlau. O Gráfico 6 apresenta o resultado obtido com o código descrito anteriormente para a vogal /A/, gravado por uma voluntária do sexo feminino.

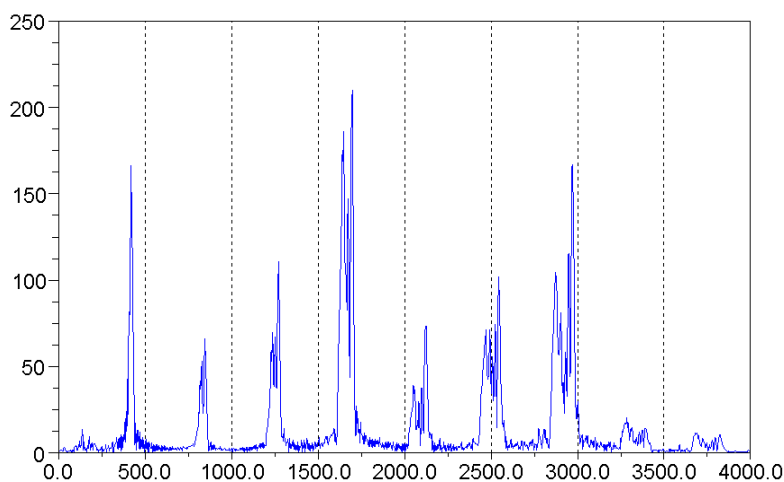


Gráfico 6 - Gráfico da intensidade x frequência para a vogal /A/, voluntária feminina

Na tabela abaixo, tem-se a comparação entre os dados retirados da referência e os obtidos pelos gráficos anteriores:

Tabela 2 - Comparação entre os dados obtidos pelo Método 4 e os da literatura

Mulheres	Formantes	/A/
Referência	F1	925
	F2	1.767
	F3	3.089
Experimental	F1	450
	F2	1.700
	F3	2.997
Diferença (%)	F1	51%
	F2	4%
	F3	3%
	Média (F1,F2,F3)	19%
	Média (F2,F3)	4%

A diferença percentual média entre os resultados obtidos em relação à referência, com a voluntária feminina, é de 19%. Porém, observa-se que se essa média desconsiderar a primeira formante, a qual mais destoa da referência, tanto no caso feminino quanto masculino, essa diferença diminui consideravelmente para 4%.

A primeira formante da vogal é a frequência que caracteriza seu som. Assim, como o estudo não teve participação de fonoaudiólogos, pode-se atribuir a diferença obtida para a primeira formante, ao fato da pronúncia da vogal ter sido realizada de maneira equivocada na visão de profissionais da área. A orientação de um especialista poderia ter fornecido valores diferentes para as frequências fundamentais, frente à especificação da maneira exata como a vogal deveria ser pronunciada e articulada, de acordo com suas particularidades.

Até 10% de diferença, como obtido com este método, é a variação da frequência fundamental entre indivíduos, caracterizando o timbre de voz de cada pessoa. Logo, uma variação aceitável.

Este último método foi aplicado na construção do banco simplificado de fonemas, para a obtenção das principais formantes de todas as vogais, orais e nasais, como mostrado na Tabela 3.

Tabela 3 - Comparação de resultados, entre vogais orais e nasais

Masculino	Formantes	/AM/	/EM/	/IM/	/OM/	/UM/
Nasais	F1	700	450	450	450	500
	F2	950	700	700	700	700
	F3	1150	950	1300	950	1520
Masculino	Formantes	/A/	/Ê/	/I/	/Ô/	/U/
Orais	F1	240	450	400	650	400
	F2	1.550	650	650	850	700
	F3	2.450	900	980	1.570	1.700
Feminino	Formantes	/AM/	/EM/	/IM/	/OM/	/UM/
Nasais	F1	400	450	450	500	470
	F2	900	950	950	950	950
	F3	1450	1400	1450	1450	1450
Feminino	Formantes	/A/	/Ê/	/I/	/Ô/	/U/
Orais	F1	450	450	450	450	450
	F2	1.700	800	850	900	950
	F3	2.997	1190	1400	1800	2100

Primeiramente, é interessante citar que apesar das diferenças entre as vogais orais e nasais de um mesmo voluntário, claramente, há semelhança de valores das formantes das vogais nasais de um mesmo voluntário. Assim, para o voluntário do sexo masculino, a frequência de 460Hz, em média, apareceu em todas as vogais, exceto /AM/, enquanto a frequência de 700Hz apareceu em todas as vogais nasais. Para a voluntária feminina, as frequências de 455Hz e 940Hz, em média, apareceram em todas as vogais nasais, possivelmente caracterizando as formantes deste tipo de vogal.

6.3. Método 5

Para obtenção das formantes das consoantes, criou-se uma nova análise, onde foram utilizadas todas as sílabas do português brasileiro. Este método consiste na divisão dos dados em 100 partes ao longo do tempo, permitindo a análise temporal das variações das principais frequências. Para cada divisão foi realizada a Transformada de Fourier, e considerou-se que a parte inicial da amostra representava o fonema consonantal e a parte final, o fonema vocálico.

Para visualização do método foram construídos gráficos tridimensionais, onde o eixo x representa as 100 divisões, o eixo y mostra os valores das frequências e o eixo z, a intensidade. E para a determinação das principais frequências dos fonemas consonantais, plotou-se o gráfico corresponde a valores baixos de x (x aproximadamente 10), enquanto que para valores maiores de x, foi possível observar que o gráfico correspondente se assemelhava à vogal que formava a sílaba.

O código numérico utilizado nesta seção foi:

```

-----
clc
clear
[x,y]=loadwave('C:\Users\ nome_usuario\Desktop\ pasta\arquivo.wav');
comprimento=length(x(1,:));
col=round(comprimento/100)-1;
M=zeros(100,col);
linha=1:100;
z=1;
for i=1:100
    for j=1:col
        M(i,j)=x(1,z);
        z=z+1;
    end
end

```

```

end
deltaf=44100/col;
for i=1:col
    frequencia(1,i)=2*(i-1)*deltaf;
end
for i=1:100
    Mfft(i,:)=fft(M(i,:));
end
plot3d(linha, frequencia, abs(Mfft));

```

O gráfico a seguir representa um dos resultados obtidos com essa programação, no caso para a sílaba /BA/.

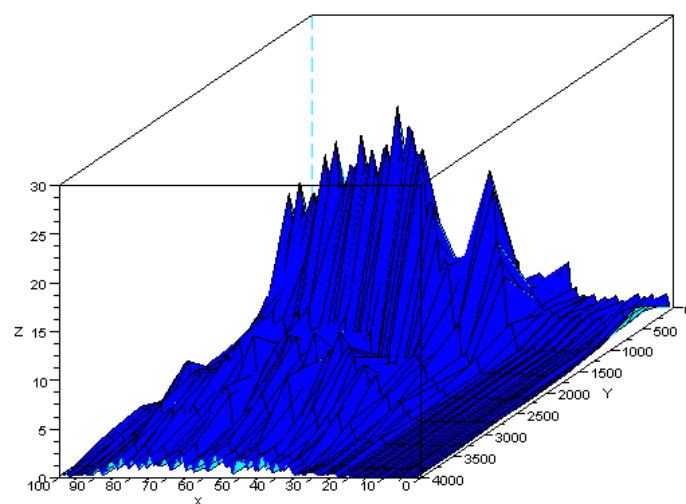


Gráfico 7 - Gráfico 3D referente à sílaba /BA/

A fim de verificar as principais frequências das consoantes, foi plotado um gráfico 2D da primeira seção do eixo x, para todas as sílabas formadas por uma mesma consoante, como mostram os Gráficos 8, 9 e 10. Para compor o Banco de fonemas utilizou-se uma média da principal frequência de todas as sílabas com a mesma consoante.

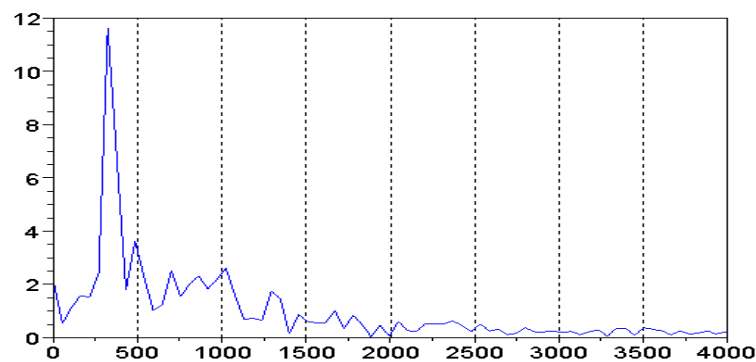


Gráfico 8 - Gráfico referentes ao fonema /B/, considerando a sílaba /BA/

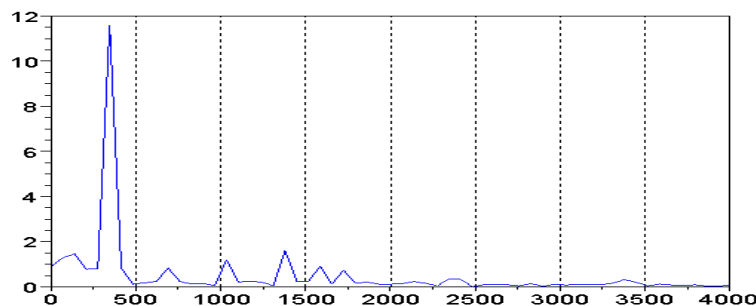


Gráfico 9 - Gráfico referentes ao fonema /B/, considerando a sílaba /BE/

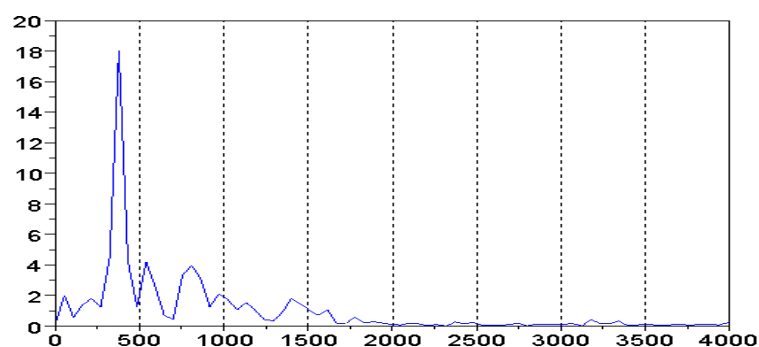


Gráfico 10 - Gráfico referentes ao fonema /B/, considerando a sílaba /BI/

A comparação entre os Gráficos 8, 9 e 10 ressalta a semelhança entre os trechos iniciais das gravações de áudio referentes às sílabas formadas pelo fonema /B/. Porém, vê-se a dificuldade na determinação das três principais formantes características do fonema, uma vez que apenas um pico bem definido é observado. Portanto, foi decidido determinar apenas a principal formante de cada fonema consonantal, isto é, a frequência de maior intensidade.

Analogamente ao realizado neste caso, foram feitas as análises para os demais fonemas consonantais. Então, construiu-se um esboço do banco de fonemas almejado, contendo os fonemas constituintes da língua portuguesa do Brasil.

7. Resultados e Análises

Muito estudo fora realizado antes e durante o desenvolvimento deste trabalho. Por isso, acredita-se que os resultados obtidos até o momento sejam coerentes, apesar de sempre poder sofrer alterações à medida que a pesquisa avança.

Devido ao fato de não existir um Banco de Fonemas Nacional, não havia como dizer se os resultados obtidos estavam ou não corretos. Apenas tivemos base para dizer se estávamos caminhando no mesmo sentido que alguns pesquisadores da área.

Realizamos comparações com um estudo realizado, na década de 90, por uma renomada especialista na área da voz, Dra. Mara Behlau. E depois, fizemos uma breve comparação entre os fonemas vocálicos e orais. O objetivo disso era mostrar que as vogais nasais, característica do português brasileiro, são diferentes das demais línguas, para as quais os implantes cocleares são produzidos. Assim, há um prejuízo no entendimento da fala quando o implante é utilizado por um brasileiro. Essa diferença de frequências foi mostrada na Tabela 3, que resume as formantes das vogais orais e das vogais nasais.

Por fim, criou-se um método de análise dos sinais a fim de captar as formantes de uma consoante, item 6.3, concluindo o trabalho com um esboço do que pode vir a ser o Banco de Fonemas do Português Brasileiro.

8. Agradecimentos

Agradecemos ao nosso orientador Prof. Dr. Raul Gonzalez Lima pelas instruções, auxílio e acompanhamento. E agradecemos também ao Dr. Silvio Penteado pelo auxílio fornecido como conhecedor do assunto abordado.

9. Referências

- Behlau, M., 2005. "Voz – O Livro do Especialista". Editora Revinter. 1a Edição. 576 p.
- Matos, P. and Segundo, P. S.. "Estudo de técnicas de processamento digital de sinais para análise das vogais da língua ka'apór". Instituto de Estudos Superiores da Amazônia (IESAM).

ALGORITHM DEVELOPMENT FOR SIGNAL PROCESSING AND CREATION OF A DATABASE OF BRAZILIAN PORTUGUESE PHONEMES

Anete Sayuri Kamimura Koga

anete.koga@gmail.com.br

Thaís Minato Takeuchi

thaisminato@gmail.com

Abstract: The absence of national cochlear implants and the speech understanding loss by the Brazilian users, caused by the use of foreign implants, motivated this paper. The objective is to develop a signal processing algorithm, aiming to create a Brazilian Portuguese phonemes database, which will allow the efficiency evaluation of the existing implants on the market for Brazilian Portuguese speakers, assist in their improvement or the development of a national implant.

The phonemes database development consists, basically, on the main formers determination for each phoneme, in other words, the three highest intensity frequencies on the graphics generated by the Fourier Transform of the phonemes records in WAV format, to which was applied the Scilab *fft* command.

Keywords: *cochlea, cochlear implant, vocal tract, database, signal processing.*