ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

DEPARTAMENTO DE ENGENHARIA MECÂNICA

Desenvolvimento de Algoritmo de Processamento de Sinais para Criação de um Banco de Fonemas do Português Brasileiro

Anete Sayuri Kamimura Koga

Thaís Minato Takeuchi

São Paulo

2011

ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

DEPARTAMENTO DE ENGENHARIA MECÂNICA

Desenvolvimento de Algoritmo de Processamento de Sinais para Criação de um Banco de Fonemas do Português Brasileiro

Trabalho de formatura a ser apresentado à Escola Politécnica da Universidade de São Paulo para conclusão de graduação em Engenharia Mecânica.

Anete Sayuri Kamimura Koga

Thaís Minato Takeuchi

Orientador: Prof. Dr. Raul Gonzalez Lima

São Paulo

2011

FICHA CATALOGRÁFICA

Koga, Anete Sayuri Kamimura

Desenvolvimento de algoritmo de processamento de sinais para criação de um banco de fonemas do português brasileiro / A.S.K. Koga; T.M. Takeuchi. – São Paulo, 2011.

99 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Mecânica.

1. Cóclea 2. Dispositivos eletrônicos 3. Aparelho fonador 4. Banco de dados 5. Processamento de sinais I. Takeuchi, Thaís Minato II. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Mecânica III. t.

Agradecimentos

Aos meus pais, que sempre proporcionaram condições favoráveis para que pudesse me dedicar aos estudos.

Ao meu namorado, pelo apoio incondicional na realização de todas as minhas conquistas.

À minha amiga e parceira Thaís e ao Prof. Dr. Raul Gonzalez Lima, sem os quais o desenvolvimento deste trabalho não teria sido possível.

Anete Sayuri Kamimura Koga

Aos meus pais e meu irmão André pelo amor e apoio incondicionais.

Ao meu namorado Marcus pela cumplicidade, companheirismo, compreensão e por estar sempre presente ao meu lado.

À minha companheira Anete S. K. Koga pela parceria, amizade e dedicação na elaboração deste trabalho.

Ao nosso orientador Prof. Dr. Raul Gonzalez Lima, pelo acompanhamento e auxílio ao longo do desenvolvimento deste trabalho.

Thaís Minato Takeuchi

Sumário

FICHA CATALOGRÁFICA....3
1. Resumo....10
2. Abstract....11
3. Implante Coclear....12
3.1. O Aparelho....13
3.2. Tipos de Implante Coclear....14
3.3. Marcas de Implantes Já Existentes no Mercado....15
3.3.1. Cochlear Corporation®....15
3.3.2. Medel®....16
3.3.3. Advanced Bionics®....17
3.3.4. Neurelec....18
4. A Fonética do Português Falado no Brasil....19
5. Som....20
6. Hardware....21
6.1. Microfone....21
6.2. Placa de Som....22
6.3. Sistema de Análise Acústica....23
6.4. Cabine Audiométrica....24
7. Processamento da Fala....25
7.1. Áudio Analógico....25
7.2. Áudio Digital....26
7.3. Estratégias de Processamento....27
7.3.1.Spectral Peak (SPEAK)....27
7.3.2. Continuous Interleaved Sampling (CIS)....28
7.3.3. Advanced Combination Encoder (ACE)....28
7.3.4. Estratégias Avançadas: Hi-ACE (Cochlear); Hi-CIS (Cochlear e MedEl) e HiRes (Clarion)....28
8. Transformada de Fourier....30
8.1. A Utilização da Transformada de Fourier....32
9. Localização das Frequências....33
10. Software....35
10.1. Aquisição do Sinal....35

10.2. Código 36
10.2.1. Descrição do Código 36
10.2.2. Código de Programação Utilizado 37
11. Aplicação do Código - Notas Musicais 39
11.1. Sons Musicais 39
11.2. Gráficos 40
11.3. Análise do Som da Flauta 43
11.4. Determinação das Formantes do Som da Flauta 44
12. Fonemas 47
12.1. Definições 47
12.2. Classificações 47
12.2.1. Vogais 47
12.2.2. Semivogais 48
12.2.3. Consoantes 49
12.3. Estrangeirismos 51
12.4. Regionalismo 51
13. Construção do Banco de Fonemas 52
13.1. Método 1 52
13.1.1. Resultados 54
13.2. Método 2 61
13.2.1. Resultados 61
13.3. Método 3 63
13.3.1. Resultados 64
13.4. Método 4 70
13.4.1. Comparação dos resultados com a referência 71
13.4.2. Comparação entre voluntário feminino e masculino 74
13.4.3. Vogais nasais e vogais orais 75
13.5. Método 5 79
13.5.1. Gráficos 80
13.6. Banco de Fonemas Limitado do Português Brasileiro 85
14. Análise dos Resultados 87
Bibliografia 89
ANEXO A 92
ANEXO B 95

ÍNDICE DE FIGURAS

Figura 1 - Esquema de um Implante Coclear 13
Figura 2 – Freedom 16
Figura 3 - Pulsar CI100 16
Figura 4 - HiRes 90K 17
Figura 5 - Paciente utilizando modelo Nucleus® 24 com microprocessador tipo "caixa" modelo SPrit 17
Figura 6 - Microfone 21
Figura 7 - Placa de som 22
Figura 8 - Analisador Portátil 23
Figura 9 - Calibrador Sonoro 23
Figura 10 - Cabine Audiométrica 24
Figura 11 - Ilustração com uma escala de 3 bits 27
Figura 12 - Frequências na Membrana Basilar 34
Figura 13 - Gravação em Flauta Doce da nota LA, feita em canal stereo e taxa de amostragem de 44,1kHz 35

ÍNDICE DE GRÁFICOS

Gráfico 1 - Onda dente de serra ____ *30*
Gráfico 2 - Componentes senoidais da onda dente de serra ____ *31*
Gráfico 3 - Gráfico da Nota DO, obtido através da FFT ____ *40*
Gráfico 4 - Gráfico da Nota DO, obtido através do "mapsound" ____ *40*
Gráfico 5 - Gráfico da Nota RE, obtido através da FFT ____ *41*
Gráfico 6 - Gráfico da Nota RE, obtido através do "mapsound" ____ *41*
Gráfico 7 - Gráfico da Nota LA, obtido através da FFT ____ *42*
Gráfico 8 - Gráfico da Nota LA, obtido através do "mapsound" ____ *42*
Gráfico 9 - Gráfico da Nota LA, com menor ampliação para visualização de demais frequências ____ *45*
Gráfico 10 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /BA/ /BE/ /BI/ /BO/ /BU/ ____ *64*
Gráfico 11 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /KA/ /KE/ /KI/ /KO/ /KU/ ____ *65*
Gráfico 12 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /BA/ /KA/ /FA/ ____ *66*
Gráfico 13 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /BE/ /KE/ /FE/ ____ *67*
Gráfico 14 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /BI/ /KI/ /FI/ ____ *68*
Gráfico 15 - Gráficos da intensidade x freqüência para a vogal /A/, voluntária feminina ____ *72*
Gráfico 16 - Gráficos da intensidade x freqüência para a vogal /A/, voluntário masculino ____ *72*
Gráfico 17 - Gráficos da vogal nasal /IM/, respectivamente, do voluntário masculino e feminino ____ *77*
Gráfico 18 - Gráficos da vogal nasal /UM/, respectivamente, do voluntário masculino e feminino ____ *78*
Gráfico 19 - Gráfico 3D referente à sílaba /BA/ ____ *80*
Gráfico 20 - Gráfico 3D referente à sílaba /BE/ ____ *81*
Gráfico 21 - Gráfico 3D referente à sílaba /BI/ ____ *81*
Gráfico 22 - Gráficos referentes ao fonema /B/, considerando de cima para baixo as sílabas: /BA/ /BE/ /BI/ /BO/ /BU/ ____ *82*
Gráfico 23 – Comparação entre os fonemas vocálicos /A/. Em cima: fonema inserido na sílaba /BA/. Em baixo: fonema analisado isoladamente. ____ *83*
Gráfico 24 - Sinal analógico original ____ *93*
Gráfico 25 - Sinal digital depois da decodificação analógico/digital – digital/analógico ____ *94*
Gráfico 26 - Gráfico da Nota DO, obtido através da FFT ____ *96*
Gráfico 27 - Gráfico da Nota RE, obtido através da FFT ____ *96*
Gráfico 28 - Gráfico da Nota MI, obtido através da FFT ____ *97*
Gráfico 29 - Gráfico da Nota FA, obtido através da FFT ____ *97*
Gráfico 30 - Gráfico da Nota SOL, obtido através da FFT ____ *98*
Gráfico 31 - Gráfico da Nota LA, obtido através da FFT ____ *98*
Gráfico 32 - Gráfico da Nota SI, obtido através da FFT ____ *99*

ÍNDICE DE TABELAS

Tabela 1 - Frequências obtidas experimentalmente 44
Tabela 2 - Frequências pré-estabelecidas 44
Tabela 3 - Diferença entre as frequências obtidas experimentalmente e padrão 46
Tabela 4 - Fonemas vocálicos orais 48
Tabela 5 - Fonemas vocálicos nasais 48
Tabela 6 - Fonemas semivocálicos 49
Tabela 7 - Fonemas consonantais com uma única forma de grafia 49
Tabela 8 - Fonemas consonantais com diferentes formas de grafia 50
Tabela 9 - Fonemas vocálicos orais 53
Tabela 10 - Fonemas vocálicos nasais 53
Tabela 11 - Fonemas consonantais 53
Tabela 12 - Resultados Método 1 para fonemas vocálicos orais 55
Tabela 13 - Resultados Método 1 para fonemas vocálicos abertos /e/ e /o/ 56
Tabela 14 - Resultados Método 1 para fonemas vocálicos nasais 57
Tabela 15 - Resultados Método 1 para sílabas da consoante B 58
Tabela 16 - Resultados Método 1 para sílabas da consoante K 59
Tabela 17 - Resultados Método 1 para sílabas da consoante D 60
Tabela 18 - Resultados Método 2 para sílabas da consoante B, com e sem repetição 62
Tabela 19 - Resultados Método 3 para algumas sílabas com vogal A 69
Tabela 20 - Resultados Método 3 para algumas sílabas com vogal E 69
Tabela 21 - Resultados Método 3 para algumas sílabas com vogal I 69
Tabela 22 - Resultados Método 3 para algumas sílabas com vogal O 69
Tabela 23 - Resultados Método 3 para algumas sílabas com vogal U 70
Tabela 24 - Comparação de resultado com referência, para vogal /A/ 73
Tabela 25 - Comparação de resultados, entre voluntário feminino e masculino 74
Tabela 26 - Comparação de resultados, entre vogais orais e nasais 75
Tabela 27 - Principal formante de cada fonema consonantal 85
Tabela 28 - As três principais formantes de cada fonema vocálico oral e nasal 86

1. Resumo

O implante coclear, também conhecido como ouvido biônico, é um dispositivo eletrônico que permite a transmissão de sinais elétricos para o nervo auditivo. Os usuários são capazes de ouvir através da decodificação destes sinais através do córtex cerebral.

Atualmente estes dispositivos são fabricados por apenas quatro fabricantes internacionais, portanto não são específicos para atender adequadamente as características fonéticas de todas as línguas faladas no mundo.

No português brasileiro há uma grande utilização das vogais nasais, diferenciando nossas características fonéticas das demais línguas. Assim, os implantes cocleares existentes no mercado causam um prejuízo no entendimento da fala quando utilizados por brasileiros.

O objetivo deste projeto é a criação de um algoritmo capaz de processar os dados coletados, e desenvolver um banco de fonemas limitado. Assim, futuramente, este estudo poderá ser aplicado de forma massiva para obtenção de um Banco de Fonemas completo do português brasileiro, com todas as suas variantes lingüísticas de região, sexo e idade.

Palavras-chave: implante coclear, vogais nasais, banco de fonemas, hardware, software.

2. Abstract

Cochlear implants, also known as bionic ear, are electronic devices that allow the transmission of electrical signals to the auditory nerve. Users are able to hear through the decoding of these signals through the cerebral cortex.
Currently these devices are manufactured by only four international manufacturers, so they are not specific to attend phonetic features of all languages spoken in the world.
In Brazilian Portuguese there is a great use of nasal vowels, differentiating our phonetic features of other languages. Thus, cochlear implants on the market cause a loss in understanding speech when used by Brazilians.
The objective of this project is to create an algorithm capable of processing the collected data, and develop a database of phonemes limited. Thus, in the future, this study can be applied to obtain a Complete Bank of Brazilian Portuguese phonemes, with all its linguistic variants of region, gender and age.

Key-words: cochlear implant, nasal vowels, phonemes database, hardware, software.

3. Implante Coclear

O implante coclear, popularmente conhecido como “ouvido biônico” é um dispositivo de alta tecnologia, que estimula eletricamente as fibras nervosas remanescentes em diferentes regiões da cóclea, permitindo a transmissão do sinal elétrico para o nervo auditivo [4]. Assim, o usuário é capaz de ouvir, através da decodificação deste sinal pelo córtex cerebral.

A Organização Mundial de Saúde estima que no Brasil, há 15 milhões de pessoas que sofrem de algum tipo de deficiência auditiva [5]. Desses, 350 mil possuem deficiência severa, aos quais seria indicado o uso do implante coclear. Porém, no mundo, atualmente pouco mais de 300 mil indivíduos beneficiam-se do mesmo.

O implante somente é recomendável para pacientes que tenham perda auditiva bilateral e que a recuperação parcial da audição não seja possível com aparelhos convencionais. A operação é feita em apenas um dos lados e indicada para aqueles com surdez completa ou que compreendem até 40% das sentenças.

Quanto mais cedo a cirurgia é realizada, melhores são os resultados esperados, sendo a idade ideal do paciente de 1 ano a 1 ano e 6 meses. Isso porque a privação sensorial auditiva leva a modificações morfológicas no sistema nervoso central. Assim, com o tempo, o cérebro vai perdendo a capacidade de decodificar os impulsos enviados pelo implante, já que as alterações sofridas pelos neurônios e a plasticidade cerebral perdem eficiência.

A privação para realização do implante é direcionada a pacientes que não possuem nervos auditivos, possuem a cóclea com má formação, ou apresentam infecções no ouvido.

Após a realização da cirurgia, o paciente deve estar vinculado a um determinado grupo ou instituição a fim de realizar periodicamente o mapeamento dos eletrodos, com intervalos de 3 a 6 meses [6]. Atualmente, este trabalho vem sendo delegado aos fonoaudiólogos.

O mapeamento consiste na determinação do nível de corrente elétrica que passará em cada eletrodo, favorecendo o desenvolvimento da fala e da audição do paciente. A cada retorno, esses níveis são alterados de acordo com o estabelecimento do estímulo mínimo perceptível e do máximo tolerável a

cada implantado. As alterações são necessárias, uma vez que o nervo auditivo se adapta progressivamente aos estímulos e o cérebro leva um tempo para interpretar estes sinais.

No Brasil, a cirurgia gratuita de implante coclear começou a ser realizada pelo SUS (Sistema Único de Saúde) no ano 2000 [3]. E desta data até fevereiro de 2011 atendeu 2993 pacientes, nos estados de São Paulo, Rio de Janeiro, Minas Gerais, Bahia, Ceará, Paraná, Rio Grande do Sul, Santa Catarina, Pará, Pernambuco, Rio Grande do Norte, além do Distrito Federal[1].

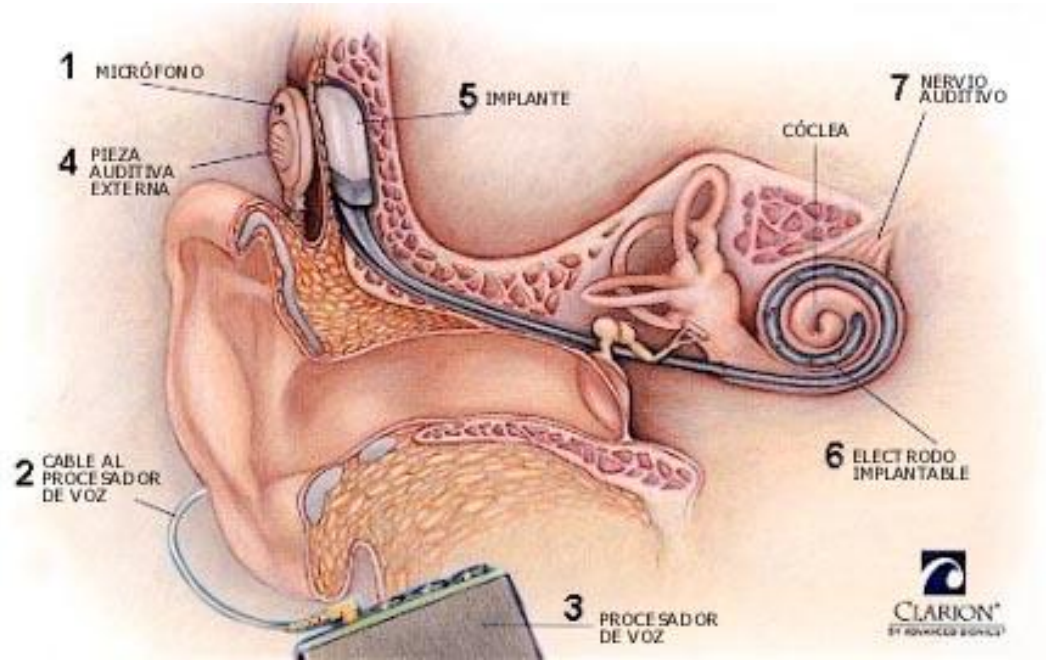

Figura 1 - Esquema de um Implante Coclear

3.1. O Aparelho

Os aparelhos são compostos por duas partes, uma externa e outra interna. A externa é utilizada pelo paciente conforme a sua vontade. Possui um microfone com a função de receber o som do ambiente, codificá-lo e enviá-lo para a parte interna. Possui também um processador de fala e uma antena transmissora. Por sua vez, a parte interna, que é cirurgicamente implantada, contém um receptor/estimulador e um conjunto de eletrodos. Sua função é receber o sinal

[1] http://idmed.uol.com.br/noticias/santa-catarina-passa-a-oferecer-implante-coclear-pelo-sus.html

que é emitido pela parte externa e transformá-lo em pulsos elétricos, os quais são enviados para o interior da cóclea pelo conjunto de eletrodos.

Os principais componentes de um implante coclear são [16]:

- Microfone Direcional: Localizado no compartimento retroauricular, capta as informações acústicas, que converte em sinais elétricos, e envia ao processador de fala.
- Processador de Fala: Digitaliza o som do sinal acústico de entrada para assim filtrar e analisar esses dados.
- Antena de Transmissão: Envia os sinais captados pelo processador de fala, em sinais FM, para o receptor interno.
- Receptor/Estimulador: Parte interna do implante, que se localiza debaixo da pele, onde são convertidos os sinais recebidos em energia elétrica, organizando e enviando-os para os eletrodos.
- Feixe de Eletrodos: Estimulam os nervos auditivos remanescentes na cóclea.

3.2. Tipos de Implante Coclear

Os implantes cocleares podem ser monocanais ou multicainais [7]. O primeiro estimula o nervo auditivo com um campo elétrico formado por dois eletrodos, um na cóclea e outro fora. Este tipo de estímulo permite que a corrente elétrica se espalhe, estimulando um grande número de células através de uma corrente elétrica baixa.

Implantes monocanais possibilitam aos pacientes o reconhecimento de sons externos e auxiliam na leitura labial, porém não permitem a diferenciação e o reconhecimento de frases e palavras sem o auxílio da leitura labial.

Através dos desenvolvimentos tecnológicos digitais e da informática, houve um avanço significativo da bioengenharia, impulsionando o desenvolvimento de sistemas mais complexos do estímulo sensorial e o desenvolvimento de processadores de fala que permitem estratégias mais sofisticadas.

Os sistemas multicanais têm como objetivo adicionar ao estímulo monopolar, um estímulo que proporciona campo elétrico mais específico, utilizando a propriedade tonotópica[2] da cóclea.
O multicanal Nucleus 22 foi o implante pioneiro e aprovado pela FAD (Food and Drug Administration) para uso clínico. Hoje, os sistemas multicanais já superaram os monocanais na diferenciação de frequências sonoras e na percepção da fala, substituindo-os.

3.3. Marcas de Implantes Já Existentes no Mercado[3]

No mercado, há uma concentração da produção e da distribuição de implantes cocleares em quatro fabricantes: Cochlear Corporation (Lane Cove, Austrália), Medel (Innsbruck, Áustria), Advanced Bionics (California, EUA) e Neurelec (Côte d'Azur, França).
Cada um deles possui suas características, onde há diferenças no material de fabricação e no formato. A semelhança entre eles é a utilização de um conjunto de eletrodos para estimular a cóclea.

3.3.1. Cochlear Corporation®

A Politec® Importação e comércio Ltda é o distribuidor no Brasil dos aparelhos produzidos pela empresa australiana Cochlear Corporation®.
Seus implantes possuem 22 canais internos de estímulo e possibilita a realização de telemetria neural ou NRT, possibilitando a verificação individual de cada eletrodo e suas respostas a cada estímulo. Os principais aparelhos deste fabricante são: Freedom , Nucleus® 24 Double Array, Nucleus® ABI(tronco cerebral).

[2] Ocorrência de processamento de informações auditivas ainda antes da consciência dos sons pelas estruturas superiores.
[3] Especificações e imagens retiradas do site <http://www.implantecoclear.org.br/textos.asp?id=3> do Grupo de Implante Coclear do Hospital das Clínicas e FMUSP.

Figura 2 – Freedom

Na Figura 2 há uma imagem do modelo Freedom, o mais avançado da Cochlear Corporation. Tem como vantagens o chip de plataforma aberta (microprocessador), que facilita a realização das atualizações ao longo dos anos quando novas tecnologias forem criadas, e eletrodo Softip, projetado para proteger as partes delicadas da cóclea durante a implantação cirúrgica.

3.3.2. Medel®

Figura 3 - Pulsar CI100

O implante PULSAR® CI100, presente na Figura 3 é produzido pela fabricante austríaca MED-EL®. Além do Pulsar® CI 100, a MED-EL® produz também outros dois modelos: o Sonata® TI100 e o Combi® 40+. Esta marca utiliza três tipos de microprocessadores: Opus 1, Opus 2 e o Tempo+.

3.3.3. Advanced Bionics®

Figura 4 - HiRes 90K

O Implante coclear HiRes® 90K é produzido pelo fabricante norte americano Advanced Bionics® e revendido pela Richards do Brasil, representante exclusivo no país.

Este implante possui um feixe com 16 eletrodos, cada um com 16 fontes independentes de energia. Permite a realização de telemetria intra operatória e apresenta 3 estratégias de codificação de fala, incluindo a mais nova delas, a Fidelity 120. Suas baterias são recarregáveis e têm durabilidade de 14 a 24 horas ininterruptas, sendo elas as Powercels®. Hoje, a empresa só revende os 2 últimos modelos de processador de fala, o retroauricular Harmony® e o tipo 'caixa' Platinum®. Ambos processam a estratégia Fidelity 120, o qual permite a criação de até 120 canais virtuais.

Figura 5 - Paciente utilizando modelo Nucleus® 24 com microprocessador tipo "caixa" modelo SPrit

3.3.4. Neurelec

Fabricante membro do grupo MXM, baseado em Sophia Antipolis Science Park, na Côte d'Azur, França.

Os implantes desta marca são comercializados no Brasil pela Widex, com registro da Anvisa (Agência Nacional de Vigilância Sanitária) desde 2010[4]. Sua estratégia de codificação é a utilização de 20 eletrodos de estimulação.

[4] http://ouvidobionico.org/2010/07/02/implantes-franceses-neurelec-comecam-a-ser-comercializado-no-brasil/

4. A Fonética do Português Falado no Brasil

A língua portuguesa é derivada do latim e com grande influência dos povos que habitaram a Península Ibérica, podendo-se citar os de origem árabe e germânica.

O português falado no Brasil ainda sofreu influência das línguas indígenas e africanas. Isso, somado aos fatores tempo e distância, explica as diferenças existentes entre as variedades faladas na Europa e no Brasil.

Apesar de ser uma única língua, é possível encontrar características marcantes na fala do português, o que a diferencia em toda a extensão do Brasil.

Além das diferenças fonético-fonológicas referentes à geografia, é possível citar também os fatores classe ou posição social, grau de escolaridade, idade, sexo e situação ou contexto social.

Somado às características já mencionadas, o português possui fonemas quase que exclusivos, como por exemplo, as vogais nasais, muito utilizadas, e que por serem particulares da língua podem prejudicar o funcionamento de aparelhos como o implante coclear, o qual depende de um banco de fonemas.

Esses fatos serviram de motivação para o presente projeto, com a finalidade de construção de um banco de fonemas do português falado no Brasil, a fim de avaliar o funcionamento de implantes cocleares quando utilizados por brasileiros e até auxiliar na sua melhoria.

5. Som

Quando uma série de perturbações é produzida, moléculas do meio (normalmente o ar) próximas a ela se deslocam, criando pequenas regiões de alta pressão. Ao aglomerarem-se para criar as zonas de alta pressão, as moléculas também criam zonas de baixa pressão. A essa alternância de zonas de alta e baixa pressão dá-se o nome de ondas. A partir de tais ondas, são gerados os sons e os ruídos.

O som pode ser definido por três parâmetros fundamentais: altura, intensidade e timbre. A altura representa sua frequência. A intensidade, sua potência. E o timbre é a identidade do som.

Já o ruído é a denominação dada ao som quando este possui intervalos irregulares entre as vibrações. Ainda não é possível excluir totalmente o ruído de um sinal reproduzido, mas uma forma de minimizar seu efeito na saída de um sinal de voz é utilizando a ferramenta Transformada de Fourier, a qual será apresentada no item 8 deste relatório e que foi a base para o desenvolvimento do software utilizado.

6. Hardware

Para a construção do hardware, foram levantados os componentes necessários, assim como suas especificações, fornecedores e preços, estes referentes a abril de 2010.

Porém, em comum acordo entre as autoras deste trabalho, juntamente com seu orientador, foi descartada a necessidade da construção do hardware, visto que os resultados obtidos através dos sons captados pelo microfone de um *Notebook*, foram suficientes para a análise e construção do banco de dados desejado. Entretanto, foi decidido manter este item no relatório.

Este aparelho seria constituído de um hardware composto por um microfone piezoelétrico, um condicionador de sinais e uma placa de aquisição dos mesmos e de um software capaz de converter os sinais analógicos em digitais.

O aparelho permitiria captar as formantes do fonema, com isso poder-se-ia identificar a intensidade relativa entre eles e criar a densidade espectral do banco de fonemas.

6.1. Microfone

Microfone Shure modelo SM 58 S.

Figura 6 - Microfone

Este microfone possui padrão de captação cardióide, que diferentemente dos unidirecionais ou bidirecionais, capta melhor os sons que incidem pela parte frontal (na direção do eixo central), e com menor qualidade os de incidência

lateral ou traseira. Assim, ele isola a fonte principal de som, e minimiza os ruídos de fundo.

Preço: R$ 671,20 na loja Hendrix World Music, localizada na rua Teodoro Sampaio, 719, Pinheiros, São Paulo.

Website: www.hendrixmusic.com.br

Telefone: (11) 3736-0345.

6.2. Placa de Som

Placa Sound Blaster modelo Creative X-Fi Xtreme Audio - 24 – Bit 7.1 PCI.

Figura 7 - Placa de som

A placa de som realiza conversão analógica-digital de 24 bits para entradas analógicas e para fontes digitais a 96kHz, além da conversão digital-analógica de 24 bits para saída estéreo a 192kHz. A taxa de amostragem para gravação de 16 a 24 bits varia de 8 a 96kHz.

Preço: R$ 314,00 no site www.ditecshop.com.br.

Telefone: (11) 3854-2311.

6.3. Sistema de Análise Acústica

- Analisador Portátil modelo 2250

Figura 8 - Analisador Portátil

Este aparelho possui dimensões de 300 X 93 X 50mm, pesando 650g. Possui um software de análise de frequência, baseando-se na análise FFT (Fast Fourrier Transform), além de gravar sons que podem posteriormente ser transferidos e analisados em um computador. A tela do aparelho apresenta um gráfico de intensidade em decibéis do som captado, facilitando o entendimento durante a coleta de dados.

- Calibrador Sonoro B&K 4231

Este aparelho é utilizado para realizar a calibração sonora de equipamentos de medição de ruído. Permite uma correção automática das características do microfone mediante uma compensação ativa entre a diferença da sensibilidade de pressão do microfone obtida com o calibrador e a sensibilidade com o campo sonoro exterior envolvente, onde se encontra exposto o microfone. Seus níveis de calibração podem ser realizados a 94dB ou 114dB, ambos a 1kHz.

Figura 9 - Calibrador Sonoro

- Cabo

Cabo de 10 vias para conexão do LEMO 1B de 10 pinos M para um LEMO 1B de 10 pinos F, com 2m de comprimento.

Preço: R$ 40.875,78 na loja Spectris do Brasil Instrumentos Eletrônicos Ltda, localizada na rua José de Carvalho, 55, Chácara Santo Antônio.

Website: www.bkhome.com ou www.bksv.com.

Telefone: (11) 5188-8186.

6.4. Cabine Audiométrica

Cabine modelo VSA 40.

Figura 10 - Cabine Audiométrica

A cabine para audiometria escolhida possui dimensões de 950 X 960 X 2000mm, tendo como características principais o piso de borracha pastilhada, os amortecedores de vibração e o revestimento acústico interno, para contribuir a uma melhor qualidade na coleta dos dados.

Pode-se ainda destacar que o isolamento acústico é de 40dB. E o visor é de duplo vidro isolante de dimensões 500 X 600 X 5mm, facilitando a observação dos candidatos durante o experimento.

Preço: R$ 3.001,00 na loja Vibrasom Tratamento Acústico, localizada na Estrada Sadae Takagi, 204, Bairro Cooperativa, São Bernardo do Campo.

Website: www.vibrasom.ind.br.

Telefone: (11) 4393-7900.

7. Processamento da Fala

As estratégias para o processamento da fala se diferenciam por um conjunto de regras ao converter os sinais sonoros em estímulos elétricos ao nervo auditivo [8]. O som a ser recebido é dividido em um determinado número de filtros pré-estabelecidos. Cada filtro é direcionado a um canal específico (par de eletrodos). Os filtros separam as principais características acústicas do som (duração, intensidade, frequência) que serão codificadas em um estímulo pulsátil sequencial. É vital que toda estratégia de codificação da fala esteja em adequada resolução espectral e resolução temporal para uma melhor percepção da fala.

7.1. Áudio Analógico

Um microfone é um dispositivo que converte uma forma física de energia, pressão acústica correspondente ao som (dB), numa variação correspondente de tensão elétrica (Volts) ou potencial elétrico. Quanto maior a pressão acústica, maior será a tensão elétrica correspondente no cabo do microfone.
Após ser representado por um sinal elétrico, o som pode ser processado por um sistema eletrônico analógico, como um amplificador, ou por um sistema digital, como um digitalizador ou gravador de CDs.
O uso do amplificador se faz necessário, uma vez que o sinal elétrico na saída do microfone é de baixa amplitude, assim é possível elevar sua tensão elétrica.
No caso do alto-falante, a lógica do seu funcionamento é inversa ao microfone. Na entrada do alto-falante apresenta-se uma variação de tensão elétrica (Volts) que vai ser convertida em variações de pressão acústica (dB) em sua saída, reproduzindo assim o som.
Há basicamente duas maneiras de efetuar uma gravação analógica: por registro magnético em fita magnética (cassete), ou por enrugamento da superfície de um disco (vinil). No caso da fita magnética, o som é armazenado de uma forma contínua sob a forma de variações de magnetismo na própria fita. E no caso de um disco de vinil, sua superfície é

"escavada", de modo a refletir a variação de amplitude do som que está sendo gravado em relevo na própria superfície do disco.

7.2. Áudio Digital

O som é um fenômeno contínuo no tempo, além de ser analógico. Para poder ser processado em um computador digital, o mesmo, deve ser convertido em eletricidade, filtrado e condicionado a níveis de tensão, antes de ser aplicado à placa de som de um computador, onde sofrerá a “digitalização”, ou seja, terá o sinal analógico convertido em sinal digital.

No caso do áudio digital, a onda de som é repartida em amostras individuais, regularmente espaçadas no tempo. A este processo denomina-se Conversão Analógico-Digital, ou Amostragem.

A taxa de amostragem, ou frequência de amostragem, é o número de amostras retiradas da forma de onda original, por segundo. Um sinal digital é definido apenas em alguns pontos no tempo (frequência de amostragem), e o mesmo possui apenas um determinado conjunto de valores discretos, em função do número de bits alocados para conversão, resultando em uma maior ou menor precisão na conversão do sinal analógico. Quanto mais elevada a taxa de amostragem, menor o erro de aproximação e melhor a aproximação à onda de som original. A taxa de amostragem limita a gama de frequências que o sinal a ser captado pode conter, sendo que o limite máximo para essa gama é metade do valor da taxa de amostragem, o que pode ser demonstrado pelo Teorema de Nyquist (Anexo A). Portanto, para um sinal com frequências até 8.000Hz, é necessário que a taxa de amostragem seja maior ou igual a 16.000Hz. Interessante notar que a frequência máxima que um ouvido humano pode captar é de cerca de 20.000Hz.

A sequência de amostras do som captado é representada numericamente em formato digital [11]. Este formato consiste num código binário, em que os números são representados sob a forma de uma sequência de bits. Um bit pode apenas apresentar o valor '0' (ligado) ou o valor '1' (desligado).

Para obtermos uma resolução equivalente à de um CD de áudio, são necessários 16 bits, o que significa 65.536 combinações numéricas possíveis

numa sequência de 0 e 1. Os valores de amplitude amostrados são sempre arredondados para o código binário mais próximo. Quanto maior o número de bits, maior a resolução, e o erro causado pelo arredondamento diminui.

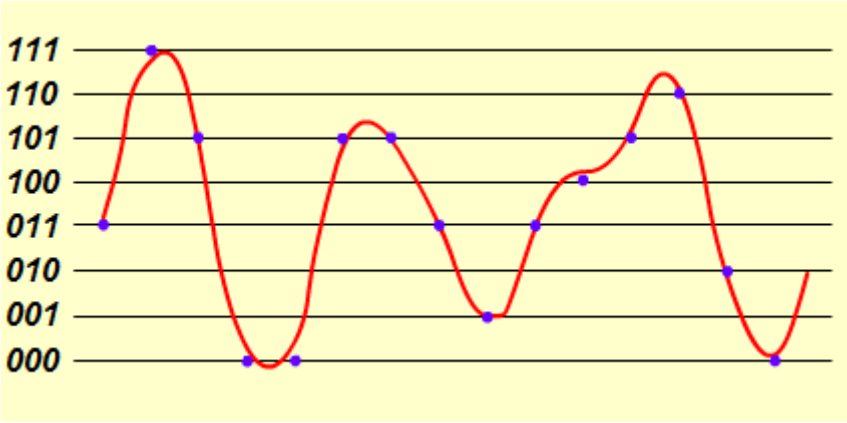

Figura 11 - Ilustração com uma escala de 3 bits

A gravação digital é efetuada armazenando os valores das amostras, bit a bit, em sequência ordenada. Por exemplo, na gravação em CD, cada bit é inscrito numa superfície refletora de luz, alterando a característica de reflexão do ponto de acordo com o bit correspondente a ele, '0' ou '1'.
Nos computadores, as placas de som (aquisição/reprodução) servem como meio intermediário entre os computadores e o exterior. Elas são capazes de fazer a conversão analógico-digital do sinal elétrico proveniente de um microfone, bem como a conversão digital-analógica, antes de o som ser reproduzido. As taxas de amostragem e número de bits que a placa usa para aquisição não são fixas e podem ser programáveis.

7.3. Estratégias de Processamento

7.3.1.Spectral Peak (SPEAK)

Estratégia utilizada no sistema multicanal Nucleus [2]. Nesta, um som entre 100 e 8.000Hz pode ser dividido em 20 bandas de frequência por um filtro. A estratégia escolhe de 1 a 8 máximas (pré-determinada na programação) a 250 pulsos por segundo por canal. A seleção prioriza a ênfase nas variáveis

espectrais. Por ter baixa velocidade, economiza maior quantidade de energia em relação às outras estratégias. É recomendada em pacientes que se adaptam melhor a uma baixa velocidade de estimulação ou que requerem fortes níveis de corrente na programação, economizando o consumo da bateria.

7.3.2. Continuous Interleaved Sampling (CIS)

Estratégia utilizada pelo sistema Nucleus e Med-El. Estimula múltiplos canais, porém não simultaneamente, ou seja, são pulsos intercalados. Utiliza de 4 a 12 canais fixos e a velocidade de estímulo é maior, chegando a 1.800Hz por canal. A estratégia CIS tem como prioridade a informação temporal ao invés da espectral. É recomendada para casos de inserção parcial do feixe de eletrodos, pois há poucos canais de estimulação disponíveis.

7.3.3. Advanced Combination Encoder (ACE)

Esta estratégia combina as vantagens das duas estratégias citadas anteriormente. Similar a SPEAK, o espectro de frequência pode variar entre 22 canais. Existe uma seleção dos canais estimulados de acordo com as máximas de maior amplitude. Como na estratégia CIS, a velocidade de estímulo é alta, podendo chegar a 1.800Hz por canal. Deste modo, oferece uma boa representação espectral com velocidades altas, aperfeiçoando também a resolução temporal. É uma estratégia flexível.

7.3.4. Estratégias Avançadas: Hi-ACE (Cochlear); Hi-CIS (Cochlear e MedEl) e HiRes (Clarion)

Estas permitem um aumento global da velocidade de estímulo. Na Hi-ACE até 36.000 pulsos por segundo, na Hi-CIS até 50.000 e na HiRes até 90.000. Melhoram a resolução temporal, facilitando a percepção de músicas, por

exemplo. Porém, é importante perceber que não são todos os pacientes que se beneficiam de velocidades altas, pois pode ocorrer “fadiga” do nervo em alguns casos (principalmente em pacientes com neuropatias ou com surdez pós-meningite).

8. Transformada de Fourier

Em meados do século XVII, o físico francês Jean-Baptiste Joseph Fourier (1768-1830) provou que qualquer forma de onda, independente da sua origem, é um somatório de ondas senoidais de diferentes frequências, amplitudes e fases [9].

A Transformada de Fourier, em homenagem ao físico J. Fourier, é uma transformada integral que expressa uma função em termos de funções senoidais, decompondo-a em suas componentes elementares, seno e cosseno [13]. Em outras palavras, é a soma de funções senoidais multiplicadas por coeficientes (amplitudes das ondas).

Define-se a Transformada de Fourier como sendo:

$$X(jw) = \int_{-\infty}^{\infty} x(t).\, e^{-jwt} dt$$

Nota-se que esta operação faz uma sobreposição de x(t) sobre a exponencial complexa (harmônica), de modo a obter a similaridade entre o sinal e a exponencial. Como exemplo, tem-se representada a seguir a onda dente-de-serra:

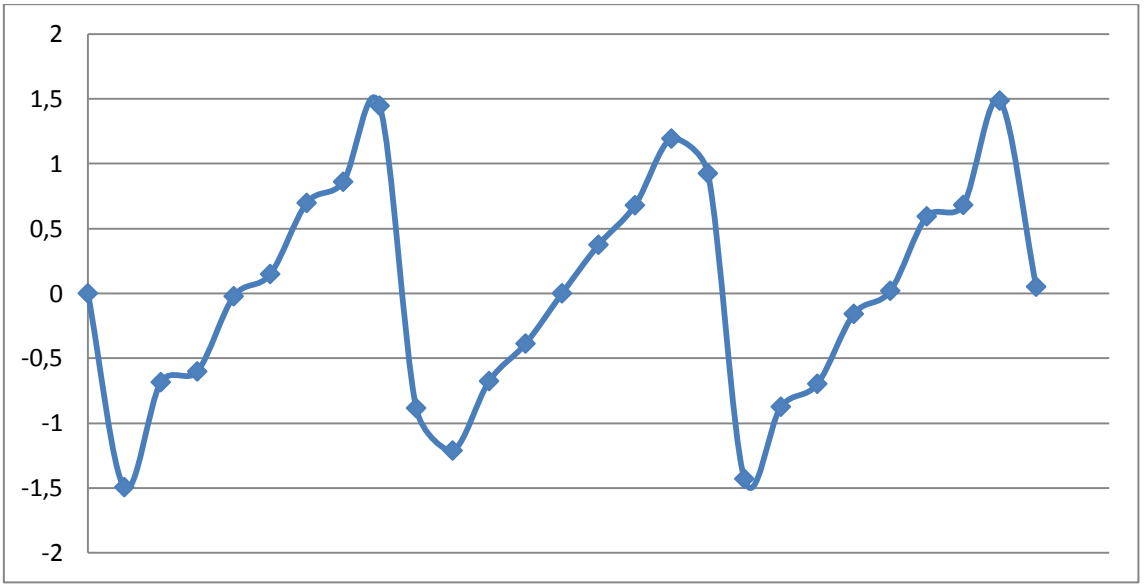

Gráfico 1 - Onda dente de serra

Que nada mais é do que o resultado da equação:

$$y = \sin(2\pi.440x) + \frac{\sin(2\pi.880x)}{2} + \frac{\sin(2\pi.1320x)}{3} + \frac{\sin(2\pi.1760x)}{4}$$

Ou seja, é a soma das ondas senoidais com diferentes amplitudes, frequências e fases, cujos gráficos estão representados abaixo:

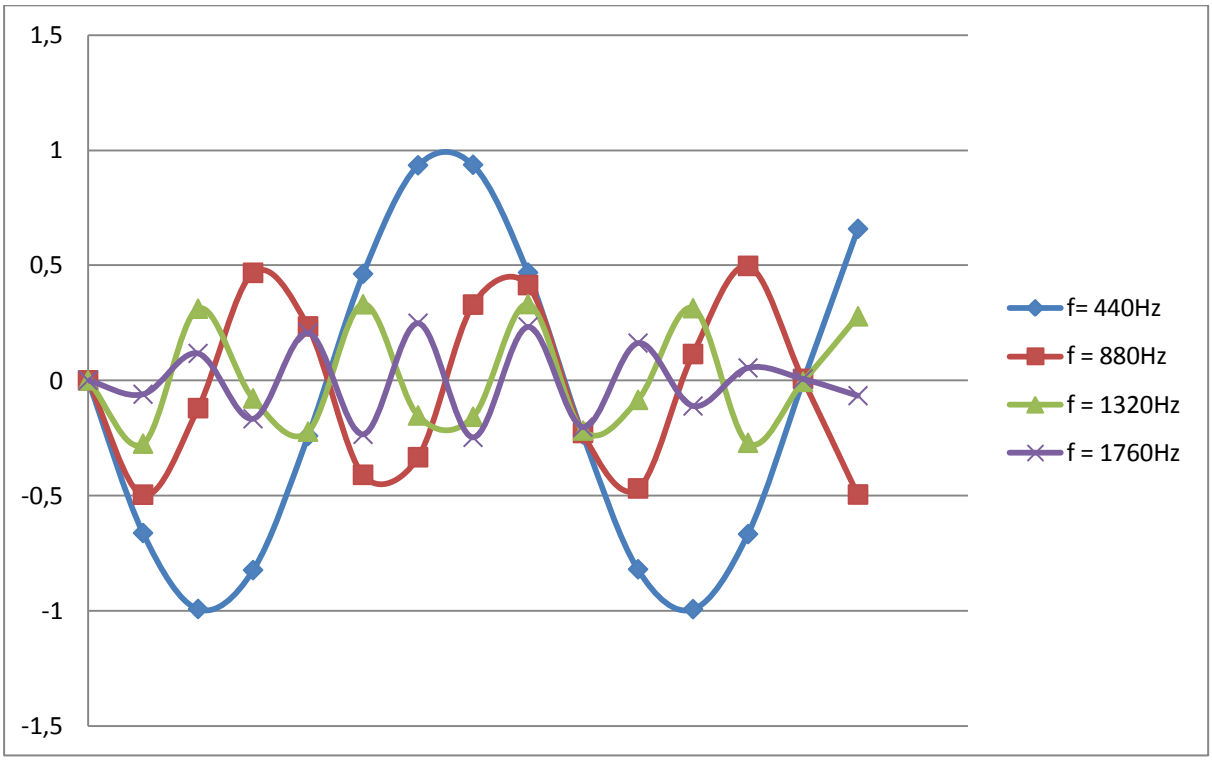

Gráfico 2 - Componentes senoidais da onda dente de serra

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).

Essencialmente a Transformada de Fourier consiste em, dada uma quantidade h, como função do tempo [h(t)], determinar a correspondente quantidade H, como função da frequência [H(f)], geralmente com um número complexo que também indica a fase [12]. Resumindo, seu objetivo é ver uma informação temporal na frequência, ou vice-versa.

As aplicações referentes à FFT são inúmeras, incluindo filtragem, segmentação, reconhecimento de padrões, descrição de imagens, compressão e reconstrução

[10]. Assim, ela é uma ferramenta largamente empregada em processamento de sinais, sons e imagens.

8.1. A Utilização da Transformada de Fourier

A ferramenta FFT corresponde à base para o desenvolvimento deste trabalho, pois é através dela, que os sons (fonemas do português brasileiro gravados no formato WAV) serão convertidos de funções no tempo para funções na frequência, permitindo assim a identificação de suas principais formantes para a construção do banco de dados.

No item 10.2.2 será apresentado e detalhado o código utilizado no software Scilab para a realização da Transformada de Fourier, além da plotagem de seus respectivos gráficos.

Inúmeros softwares que realizam análises semelhantes são disponibilizados na Internet, porém foi optado por não utilizá-los, uma vez que a utilização do código no Scilab proporciona uma clara visualização da rotina empregada e auxilia a compressão passo-à-passo do que está sendo realizado.

9. Localização das Frequências

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 [14]. Com ela é possível determinar a região apropriada para a colocação dos eletrodos. A função é dada por:

$$f = \int_0^x \Delta f_{cb} = A(10^{ax} - K)$$

Onde:

f = frequência característica.

A = constante de escalonamento entre a frequência característica e a frequência máxima de audição da espécie.

a = inclinação da reta entre a frequência e a posição em que se mostrou constante em todas as espécies.

x = distância da posição apical até a região de interesse (em mm, ou relativo entre 0 e 1).

K = constante de integração que representa a divergência entre a natureza logarítmica da curva e a determinada pela mínima frequência audível pela espécie.

Para os seres humanos a equação fica:

$$f = 165{,}4 * (10^{2{,}1x} - 0{,}88)$$

a = 2,1 quando utilizando x relativo (entre 0 e 1) ou 0,6 para x em milímetros.

Exemplo:

Se o último eletrodo for implantado a uma distância de 25mm da base da cóclea, então para uma cóclea de tamanho médio de 35mm, o eletrodo está a 10mm do ápice e assim x=10/35. Pela função de Greenwood tem-se que f = 513Hz.

Assim, é possível encontrar a localização das frequências ao longo da membrana basilar (Figura 12).

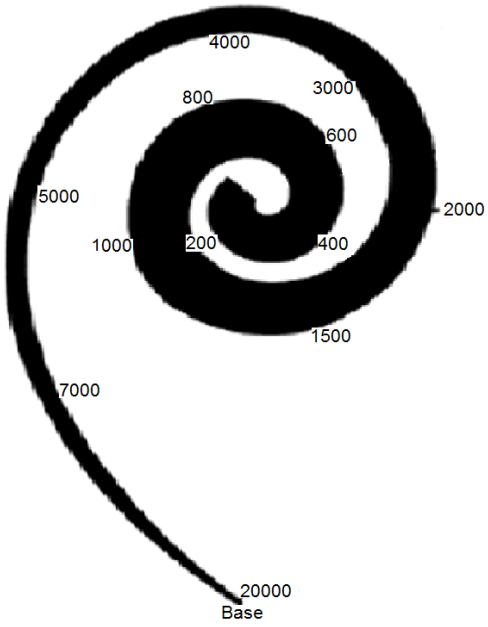

Figura 12 - Frequências na Membrana Basilar

10. Software

O Scilab[5] foi o software utilizado no desenvolvimento deste trabalho. Ele foi escolhido por ser um software aberto, isto é, distribuído gratuitamente via Internet desde 1994, disponível em 13 línguas, além de atualmente ser amplamente conhecido e utilizado em ambientes industriais e educacionais pelo mundo.

10.1. Aquisição do Sinal

Inicialmente, o sinal de áudio foi captado pelo próprio microfone do *Notebook*, a fim de possibilitar o desenvolvimento de um algoritmo para processamento do som. Este sinal foi gravado com o uso do software livre Audacity[6].

Audacity é um editor de áudio que pode 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. A Figura 13 apresenta um sinal de áudio gravado neste software.

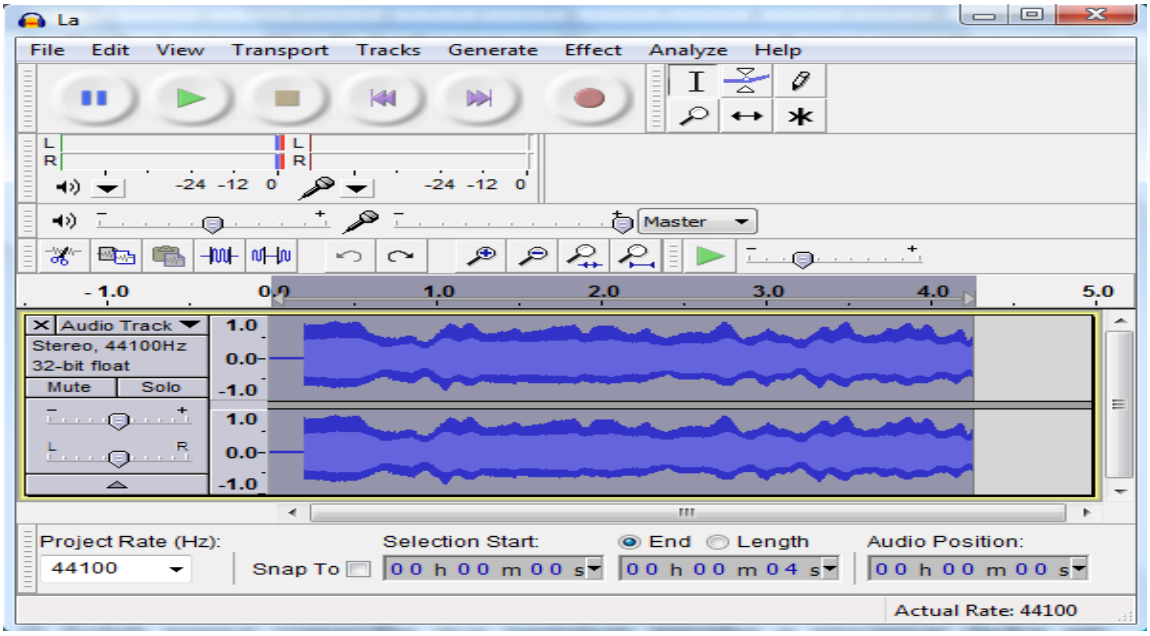

Figura 13 - Gravação em Flauta Doce da nota LA, feita em canal stereo e taxa de amostragem de 44,1kHz

[5] http://www.scilab.org/
[6] http://audacity.sourceforge.net/

10.2. Código

Embora simples, o código de programação utilizado no Scilab é suficiente e eficiente frente aos objetivos do trabalho. Com ele é possível importar um arquivo em formato WAV e gerar gráficos que ressaltam suas principais frequências ou formantes, permitindo consequentes análises e comparações, além da construção do banco de fonemas almejado.

O som é uma onda e por isso, como já visto no item 8, pode ser analisado através das teorias de Fourier. Para uma onda senoidal perfeita poderíamos facilmente determinar sua frequência olhando seu gráfico no tempo, para obter seu período e frequência correspondentes. Porém, para sinais de áudio, não temos apenas uma frequência, e sim uma combinação de várias frequências e ruídos, não sendo possível determinar sua frequência fundamental observando-os através do tempo.

Assim, parte-se para uma análise no domínio da frequência, utilizando a Transformada de Fourier (FFT). Esta transformada recebe um sinal de entrada no domínio do tempo, e retorna um sinal (número imaginário) no domínio da frequência.

O princípio da Transformada de Fourier é tentar compor o sinal de entrada por uma soma de sinais periódicos. Assim, através da função no domínio da frequência, podemos determinar o módulo e a fase para cada valor de frequência.

10.2.1. Descrição do Código

Primeiramente, para importar o arquivo WAV no Scilab, utilizou-se o comando *loadwave*. 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 apresenta a composição dos números imaginários do sinal, em valores absolutos.

Por fim, foi determinado um vetor de extensão do sinal, número de pontos amostrais, baseado na frequência em que o som foi gravado, 44.100Hz, ou

seja, 44.100 pontos em 1 segundo. Com isso, é possível plotar 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 faz a FFT do sinal com um incremento de tempo de 0,1 segundo, e plota o espectrograma do som até o limite de 1.500Hz.

É possível ver que o resultado obtido pelo “mapsound” é exatamente o mesmo obtido com a sequência de comandos anteriores, porém com uma menor precisão de leitura, como será mostrado no item 10.2. Mas ainda assim, será utilizado como base de comparação de resultados, sendo uma maneira diferente de visualizar os mesmos.

10.2.2. Código de Programação Utilizado

Este foi o primeiro código utilizado no processamento de sinais. Apesar de simples, mostrava-se eficiente para as necessidades iniciais. Porém, à medida que o trabalho foi avançando novos tipos de análises foram sendo requisitados e portanto, um refinamento deste código inicial foi progressivamente sendo desenvolvido e todos as etapas serão mostrada adiante.

```
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,:))
```

10.2.3. Mapsound

Seu princípio de funcionamento é o mesmo criado pela sequência de comandos anteriormente citados. Ele faz a FFT do sinal com um incremento de tempo de 0,1 segundo, e plota o espectrograma do som até o limite de 1.500Hz, caso não seja especificado outro valor para a frequência máxima (fmax).

O código implementado pelo comando *mapsound* está apresentado a seguir.

```
[lhs,rhs]=argn(0);
if ( rhs <= 5 ) then rate=0;end
if ( rhs <= 4 ) then simpl=1; end ;
if ( rhs <= 3 ) then fmax=3000; end ;
if ( rhs <= 2 ) then fmin=300; end ;
if ( rhs <= 1 ) then dt=0.1;  end ;
n=prod(size(w));
defaultrate=44100;
if rate==0; rate=defaultrate; end;
dp=dt*rate;
points=2^floor(log(dp)/log(2));
ind=fmin/rate*points/simpl:fmax/rate*points/simpl;
f=abs(fft(w(1:points),1));
f=sum(matrix(f,points/simpl,simpl),'c')';
M=f(ind);
i=1;
i=i+dp;
while i+points <= n;
  f=abs(fft(w(i:i+points-1),1));
  f=sum(matrix(f,points/simpl,simpl),'c')';
  M=[M;f(ind)];
  i=i+dp;
end;
[nl,nc]=size(M);
if nl==1 ; return;end
x_abs= (0:(nl-1))*(n/rate)/(nl-1);
y_abs= (0:(nc-1))/(nc-1);
y_abs= 2*(fmin*(1-y_abs) + fmax*(y_abs))
grayplot(x_abs,y_abs,-M)
```

11. Aplicação do Código - 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. E então aplicar o mesmo tratamento às demais gravações.

11.1. Sons Musicais

As sete notas musicais foram produzidas por uma flauta doce da marca Yamaha e gravadas no formato WAV pelo software Audacity.
A seguir, serão apresentados os gráficos das frequências relativas a um sinal sonoro, obtidos pela aplicação do código do item 10.2.2. Aqui, estão apenas os resultados das notas DO, RE e LA, para não prolongar demasiadamente este item. Os demais estão contidos no Anexo B.

11.2. Gráficos

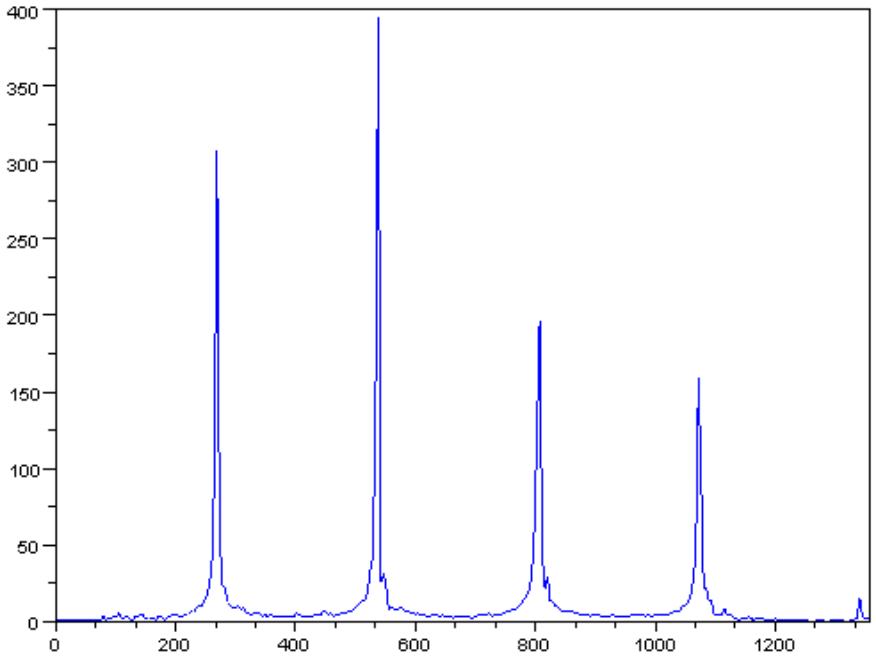

Gráfico 3 - Gráfico da Nota DO, obtido através da FFT

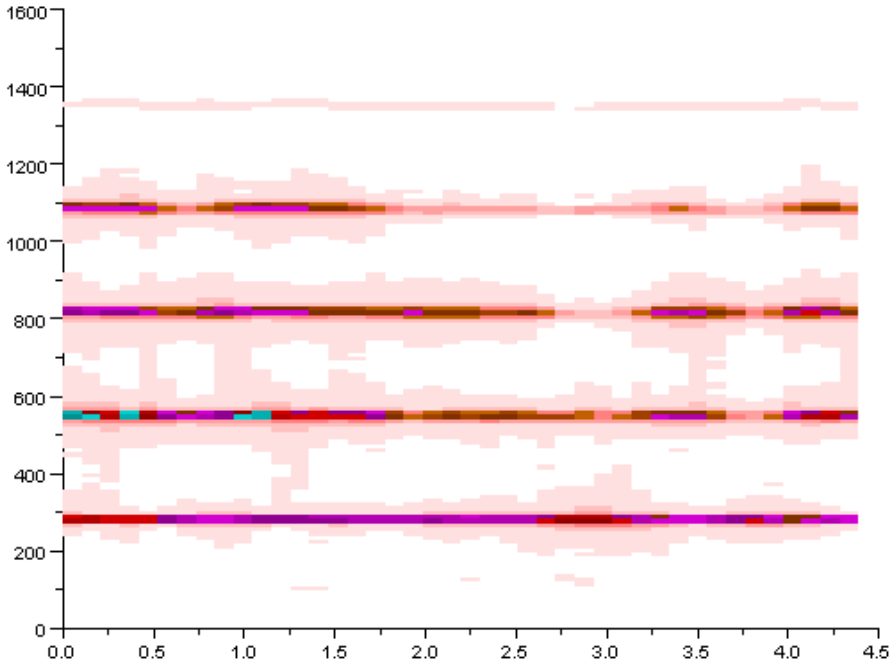

Gráfico 4 - Gráfico da Nota DO, obtido através do "mapsound"

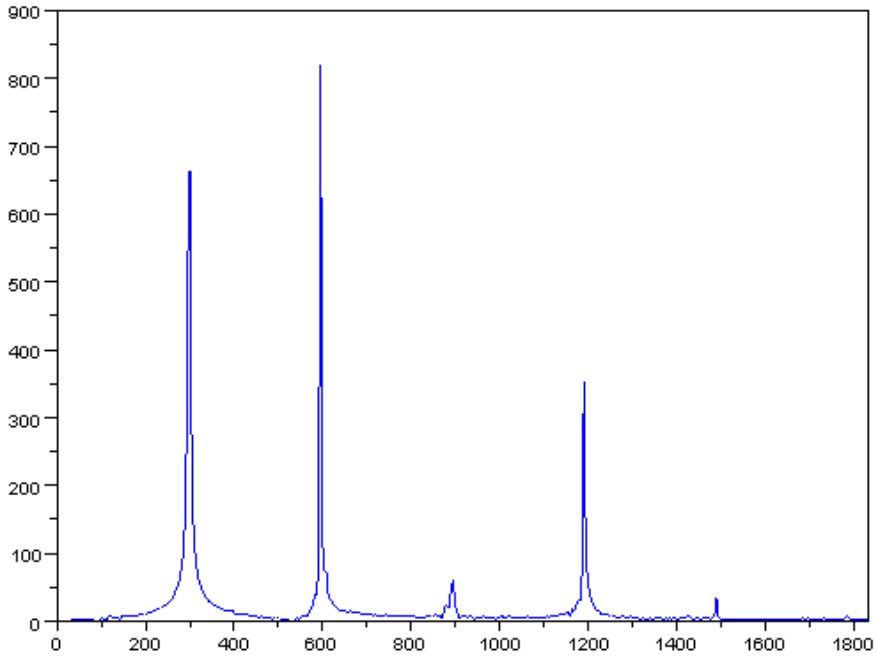

Gráfico 5 - Gráfico da Nota RE, obtido através da FFT

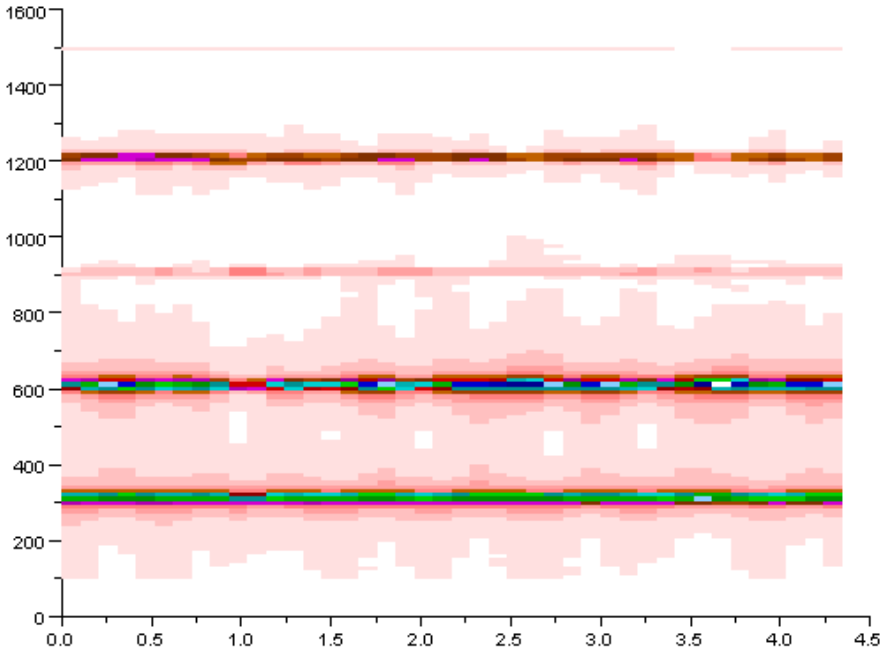

Gráfico 6 - Gráfico da Nota RE, obtido através do "mapsound"

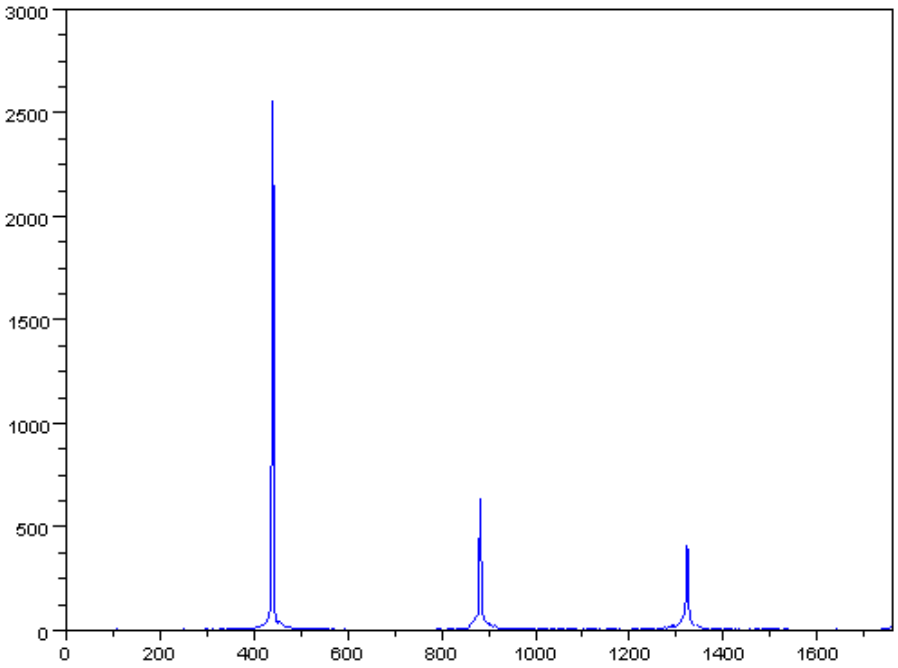

Gráfico 7 - Gráfico da Nota LA, obtido através da FFT

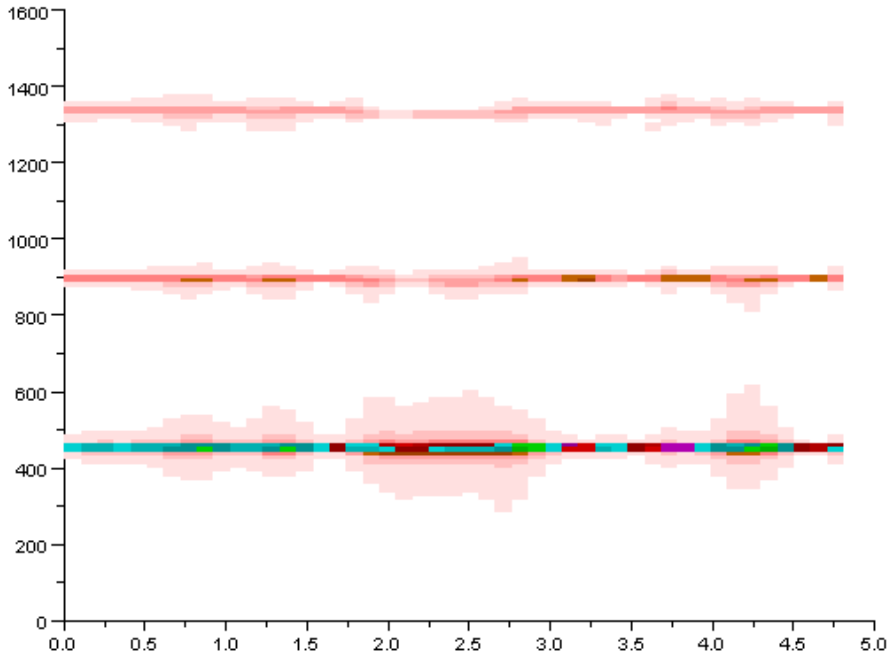

Gráfico 8 - Gráfico da Nota LA, obtido através do "mapsound"

Pode-se observar que as frequências obtidas nos gráficos 3, 5 e 7 são as mesmas dos gráficos correspondentes, 4, 6 e 8, respectivamente. Porém, nos gráficos obtidos através da FFT, tem-se a Intensidade pela Frequência, enquanto que nos gráficos obtidos pelo comando “mapsound” tem-se a Frequência pelo Espaço amostral. Além disso, as frequências de maior amplitude, ou seja, de maior intensidade, são mostradas nos espectogramas dos gráficos do comando *mapsound*, com uma tonalidade mais escura. Observou-se que tal tonalidade variava desde o rosa até vermelho, sendo que em frequências ainda mais fortes, os pontos apresentam coloração azulada ou esverdeada.

11.3. Análise do Som da Flauta

Quando um instrumento musical produz uma determinada nota, ele produz não apenas o som da frequência fundamental, que dá a qualidade sonora (timbre) da nota musical ouvida, mas também o de frequências harmônicas múltiplas da fundamental [17]. Se não fosse assim, não haveria diferença entre sons de instrumentos diferentes. A proporção dos harmônicos presentes no som produzido por cada tipo ou qualidade de instrumento é que os diferencia.

O som produzido por instrumentos de corda geralmente possuem todos os múltiplos da frequência fundamental como componentes, já alguns instrumentos de sopro, produzem sons apenas com harmônicos ímpares [18]. Como exemplo, o som produzido por um clarinete contém quase que somente harmônicos ímpares e muitos deles agudos. Já o som de uma flauta doce, possui praticamente, somente o primeiro e o segundo harmônicos, com um som fundamental mais forte, além de harmônicos fracos e em pouca quantidade.

É possível afirmar então que o som de cada instrumento é característico, justamente pelo fato da presença significativa ou não de harmônicos.

11.4. Determinação das Formantes do Som da Flauta

A Tabela 1 mostra as três principais frequências, obtidas a partir da observação dos gráficos do item 11.2 e demais apresentados no Anexo B.

Tabela 1 - Frequências obtidas experimentalmente

Notas	Frequências [Hz]		
	1º	2º	3º
DO	267	534	801
RE	298	600	1195
MI	330	670	1020
FA	350	700	1070
SOL	400	800	1200
LA	440	880	1330
SI	495	990	1490

Para permitir uma comparação, a Tabela 2 mostra as frequências de alguns harmônicos aceitos universalmente no mundo da música[7].

Tabela 2 - Frequências pré-estabelecidas

Notas	Frequências [Hz]		
	Índice 4	Índice 5	Índice 6
DO	261,63	523,25	1046,50
RE	293,66	587,33	1174,70
MI	329,63	659,26	1318,50
FA	349,23	698,46	1396,90
SOL	392,00	783,99	1568,00
LA	440,00	880,00	1760,00
SI	493,88	987,77	1975,50

Apesar do terceiro harmônico, de Índice 6, não ser uma das principais frequências na determinação do som da flauta, ela está presente na composição de seu som, em menor intensidade, como pode ser visto no

[7] http://www.phy.mtu.edu/~suits/notefreqs.html

Gráfico 9. Os gráficos mostrados no item 11.2 estão ampliados, para melhor visualização das frequências principais, e por isso não apresentam os demais harmônicos que o compõem.

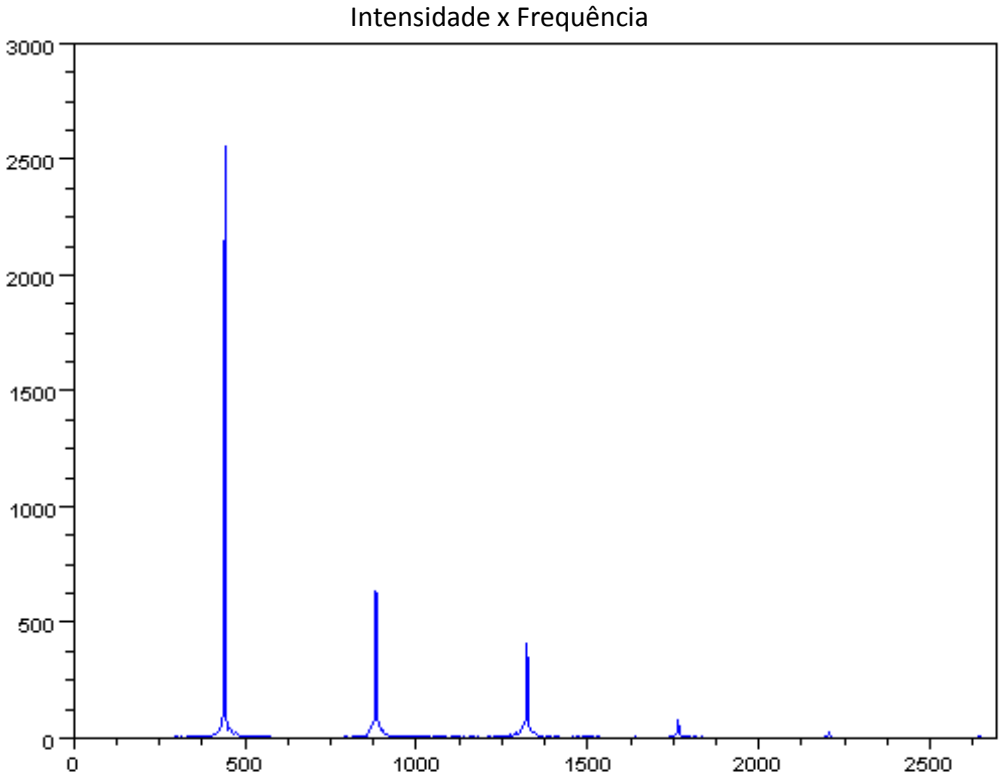

Gráfico 9 - Gráfico da Nota LA, com menor ampliação para visualização de demais frequências

É interessante notar que no Gráfico 9, a quarta frequência de maior intensidade, possui valor de 1.764Hz, que em comparação com a frequência de Índice 6 da Tabela 2, 1.760Hz, possui diferença de 0,23%, considerada insignificante.

Para o caso da flauta doce, os dois primeiros harmônicos são mais intensos que os demais, e portanto, essenciais na determinação das formantes do som da flauta e então, serão comparadas com valores da Tabela 2.

Da comparação entre as duas primeiras formantes (harmônicos da flauta) com as frequências padrão para todas as notas gravadas, tem-se a Tabela 3, a qual mostra, em porcentagem, a diferença entre o resultado obtido e o esperado.

Tabela 3 - Diferença entre as frequências obtidas experimentalmente e padrão

Notas	Frequências [Hz]				Diferença [%]	
	1º	2º	Índice 4	Índice 5	1º e Índ.4	1º e Índ.5
DO	267	534	261,63	523,25	2%	2%
RE	298	600	293,66	587,33	1%	2%
MI	330	670	329,63	659,26	0%	2%
FA	350	700	349,23	698,46	0%	0%
SOL	400	800	392,00	783,99	2%	2%
LA	440	880	440,00	880,00	0%	0%
SI	495	990	493,88	987,77	0%	0%

As frequências de maior intensidade são as formantes principais, ou seja, de maior relevância para a caracterização da análise dos fonemas. Assim, a partir do estudo realizado, mesmo sem a utilização de nenhum tipo de filtro, é possível afirmar que os resultados foram coerentes, já que a frequência fundamental e seus harmônicos foram identificados sem desvios relevantes quando comparados com a literatura.

Além disso, é interessante notar que o desvio encontrado, de até 2%, pode ser explicado por variações naturalmente encontradas entre as flautas.

12. Fonemas

Nesta seção será realizada uma breve descrição dos fonemas que compõem o português falado no Brasil e sua listagem, além de ressalvas para detalhes e particularidades da língua.

12.1. Definições

A fala é constituída pela emissão de sons gerados no aparelho fonador, o qual é constituído de: pulmões, brônquio, traquéia, laringe, glote, cordas vocais, faringe, úvula, boca e órgãos anexos e fossas nasais [21].
Cada indivíduo tem uma maneira singular de realizar esses sons no ato da fala, e essas peculiaridades na pronúncia de cada falante são estudadas pela fonética.
Descritos pela fonética, os fones são os sons pronunciados pelos falantes da língua, isto é, são a realização física ou real dos fonemas, que por sua vez, correspondem à realidade acústica registrada pelo aparelho auditivo, ou seja, "os sons elementares e distintivos que o homem produz quando, pela voz, exprime seus pensamentos e emoções"[8].

12.2. Classificações

O idioma português falado no Brasil é formado por 34 fonemas, sendo 13 vogais, 19 consoantes e 2 semivogais.

12.2.1. Vogais

São fonemas pronunciados sem obstáculos à passagem de ar, chegando livremente ao exterior.

[8] BECHARA, Evanildo. **Moderna gramática portuguesa**. 37 ed. rev., ampl. E atual. Conforme o novo Acordo Ortográfico. Rio de Janeiro: Nova Fonteira, 2009.

Tabela 4 - Fonemas vocálicos orais

Fonema	Representação	Exemplo
/a/	A	pano
/á/	A ou Á	amor, átomo
/e/	E	beijo
/é/	E ou É	berro, café
/i/	I	ilha
/o/	O	olho
/ó/	O ou Ó	cola, mói
/u/	U	uva

Tabela 5 - Fonemas vocálicos nasais

Fonema	Representação	Exemplo
/ã/	Ã	rã
	AM	campo
	AN	anta
/ẽ/	EM	sempre
	EM	sente
/ĩ/	IM	sim
	IN	cinto
/õ/	Õ	põe
	OM	pomba
	ON	ponta
/ũ/	UM	bumbo
	UN	mundo

12.2.2. Semivogais

São os fonemas que se juntam a uma vogal, formando com esta uma única sílaba. Apenas os fonemas /i/ e /u/ átonos funcionam como semivogais, e para que não sejam confundidos com as vogais i e u serão representados por [y] e [w] e chamados respectivamente de iode e vau.

Tabela 6 - Fonemas semivocálicos

Fonema	Representação	Exemplo
/y/	I	uivo
	E	área
	ÃE	mãe
	ÊM	têm
	ÉM	também
	EM	vivem
/w/	U	automático, frequente
	L	móvel
	ÃO	pão
	AM	falam

12.2.3. Consoantes

São os fonemas produzidos mediante a resistência que os órgãos bucais (língua, dentes, lábios) opõem à passagem de ar.

Tabela 7 - Fonemas consonantais com uma única forma de grafia

Fonema	Representação	Exemplo
/p/	P	pai
/b/	B	bola
/m/	M	mãe
/f/	F	faca
/v/	V	vaca
/t/	T	tio
/d/	D	dado
/n/	N	nada
/ñ/	NH	nhoque
/l/	L	lua
/λ/	LH	lhama
/r/	R	caro

Tabela 8 - Fonemas consonantais com diferentes formas de grafia

Fonema	Representação	Exemplo
/R/	RR	carro
	R	rosa
/z/	Z	azul
	S	rosa
	X	exato
/s/	S	seda, valsa
	SS	massa
	Ç	maçã
	C	cedo
	SC	descer
	SÇ	desça
	X	trouxe
	XC	excesso
/j/	J	jeito
	G	giz
/x/	X	xadrez
	CH	chave
/g/	G	gato
	GU	gueixa
/k/	C	cume
	QU	queixa

Para a construção das tabelas dos fonemas, foi utilizado um conjunto de grafemas adaptado à fala do português no Brasil, o que não corresponde integralmente ao Alfabeto Fonético Internacional [20].

Além disso, é possível afirmar que a definição dessa lista de fonemas não é precisa, contendo determinada arbitrariedade, causada por ocorrências de estrangeirismos e regionalismos na língua.

12.3. Estrangeirismos

Existem palavras dicionarizadas em português como *hardware* e *hub*, que apresentam *H* aspirado. Diante disso, seria de se esperar que esse fonema fosse incluído na lista dos fonemas do português, porém, isso faria com que a lista se ampliasse com uma profusão de fonemas de uso muito restrito no idioma.

O *H* aspirado é repelido pelos hábitos fonéticos da língua portuguesa, portanto sua utilização se restringe às palavras estrangeiras incorporadas recentemente ao idioma [19]. Outras palavras que apresentavam a característica do *H* aspirado no idioma de origem passaram por uma acomodação fonológica ao serem incorporadas ao português, perdendo este fonema aspirado.

Portanto, é possível afirmar que os fonemas que podem ser considerados genuínos do idioma são aqueles de uso amplo e não repelidos pelos hábitos fonéticos da comunidade.

Como a língua está em constante evolução e os hábitos mudam, é possível que fonemas atualmente repelidos, possam ser incorporados à língua com o passar do tempo.

12.4. Regionalismo

O Brasil é um país de dimensão continental, e portanto há inúmeras variações quanto à pronúncia da língua em diversas localidades do país

No Rio Grande do Sul, por exemplo, é comum pronunciar palavras como *sul* e *naval* utilizando uma variação do /l/ no final dessas palavras, em contraposição à utilização da semivogal *w* como é comum no resto do país.

Da mesma forma, os cariocas usam contrações específicas de /tx/ e /dj/ em palavras como noite e dia.

Essas características fonéticas poderiam ser consideradas como fonemas, mas não entram na tabela devido ao caráter regional.

13. Construção do Banco de Fonemas

Analogamente ao que foi feito para as notas musicais no item 11, nesta seção será realizada a análise dos fonemas vocálicos, assim como das diferenças entre as vogais orais e as vogais nasais, tão características da língua portuguesa.

Realizada as análises, inicialmente para as notas musicais produzidas por uma flauta doce, deu-se 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, neste trabalho não se abrangeu outros sotaques e faixas etárias que compõem a diversidade do país.

Os fonemas também foram gravados através do software livre Audacity em formato WAV pelo próprio microfone do *Notebook*, para a então análise através do software Scilab.

13.1. Método 1

Denominou-se Método 1, a etapa do processo na qual foram gravadas 12 vogais sustentadas (7 orais e 5 nasais), mais 15 sílabas, formadas por apenas três consoantes), em frequência e intensidade habituais. Esta pequena quantidade de variações facilitaria o estudo do processamento dos dados. Assim, foram escolhidos 4 voluntários, sendo dois do sexo feminino e a outra metade do sexo masculino, todos brasileiros, nativos da cidade de São Paulo, com histórico negativo de distúrbios de comunicação oral ou de audição.

As seguintes tabelas foram apresentadas aos voluntários, para um melhor entendimento de como cada uma das vogais deveriam ser pronunciadas. Cada voluntário pronunciou cada uma das 27 sílabas, uma única vez.

Tabela 9 - Fonemas vocálicos orais

Fonemas vocálicos orais		
Fonema	Representação	Exemplo
/a/	A ou Á	amor
/e/	E	beijo
/é/	E ou É	café
/i/	I	ilha
/o/	O	olho
/ó/	O ou Ó	cola
/u/	U	uva

Tabela 10 - Fonemas vocálicos nasais

Fonemas vocálicos nasais		
Fonema	Representação	Exemplo
/ã/	AM	rã
/ẽ/	EM	sempre
/ĩ/	IM	sim
/õ/	OM	pomba
/ũ/	UM	mundo

A Tabela 11, apresenta as consoantes que foram gravadas. Nota-se que para cada fonema consonantal, foram gravadas cinco sílabas, na tentativa de se obter algum padrão entre elas.

Tabela 11 - Fonemas consonantais

Fonemas consonantais			
Fonema	Representação	Exemplo	Sílabas
/b/	B	bola	ba.be.bi.bo.bu
/k/	C	comer	ka.ke.ki.ko.ku
/d/	D	dado	da.de.di.do.du

O código utilizado para análise, muito semelhante ao utilizado para a análise dos sons musicais, foi:

```
clc
clear
[x,y]=loadwave('C:\Users\nome_usuario\Desktop\ nome_arquivo.wav');
fx=fft(x(1:5000));
for i=1:5000
fr(i)=(i-1)/(5000*(1/44100));
end
plot(fr,abs(fx));
mapsound(x(1,:))
```

13.1.1. Resultados

Pode-se observar das tabelas a seguir que, utilizando o mesmo processamento de sinais utilizado para os sons musicais (Item 11), não há semelhança de resultados entre quaisquer dois indivíduos. Isso impossibilita a determinação de um padrão de frequências, seja entre sílabas, pessoas de mesma idade ou, de mesmo sexo.

Outro ponto a ser observado nos resultados é que as formantes foram colocadas em ordem decrescente de intensidade, ou seja, a primeira formante seria a frequüência de maior intensidade encontrada na amostra, e a terceira formante, a terceira frequência de maior intensidade.

Tabela 12 - Resultados Método 1 para fonemas vocálicos orais

Voluntário		No	1	2	3	4
		Sexo	F	F	M	M
		Idade	22	24	22	54
A	Formantes	1a	450	360	620	385
		2a	675	90	390	298
		3a	775	695	335	600
E	Formantes	1a	225	1190	220	90
		2a	1100	80	30	185
		3a	1440	270	1015	1025
I	Formantes	1a	230	100	125	90
		2a	110	195	65	175
		3a	1380	395	1470	1050
O	Formantes	1a	225	270	235	87
		2a	425	370	120	255
		3a	350	90	175	170
U	Formantes	1a	250	90	115	105
		2a	375	180	180	300
		3a	125	370	55	200

Tabela 13 - Resultados Método 1 para fonemas vocálicos abertos /e/ e /o/

Voluntário		No	1	2	3	4
		Sexo	F	F	M	M
		Idade	22	24	22	54
É	Formantes	1a	210	350	320	355
		2a	320	255	800	265
		3a	1000	980	1210	70
Ó	Formantes	1a	355	280	275	350
		2a	470	370	325	265
		3a	240	465	50	440

Tabela 14 - Resultados Método 1 para fonemas vocálicos nasais

Voluntário		No	1	2	3	4
		Sexo	F	F	M	M
		Idade	22	24	22	54
AM	Formantes	1a	255	275	240	90
		2a	750	820	120	275
		3a	300	100	60	375
EM	Formantes	1a	220	1195	240	95
		2a	1370	270	1000	270
		3a	1055	95	115	1000
IM	Formantes	1a	230	90	120	100
		2a	120	180	180	200
		3a	340	270	60	300
OM	Formantes	1a	240	270	215	275
		2a	370	345	55	325
		3a	125	180	275	80
UM	Formantes	1a	220	100	175	330
		2a	315	180	235	220
		3a	440	280	118	105

Tabela 15 - Resultados Método 1 para sílabas da consoante B

Voluntário		Nº	1	2	3	4
		Sexo	F	F	M	M
		Idade	22	24	22	54
BA	Formantes	1ª	390	90	245	300
		2ª	95	140	300	375
		3ª	685	270	600	70
BE	Formantes	1ª	90	80	240	70
		2ª	190	160	170	15
		3ª	1080	250	110	150
BI	Formantes	1ª	98	80	175	75
		2ª	270	160	115	10
		3ª	180	45	57	140
BO	Formantes	1ª	240	90	235	75
		2ª	100	175	55	160
		3ª	425	260	175	240
BU	Formantes	1ª	108	95	220	75
		2ª	30	170	170	150
		3ª	42	260	50	375

Tabela 16 - Resultados Método 1 para sílabas da consoante K

Voluntário		Nº	1	2	3	4
		Sexo	F	F	M	M
		Idade	22	24	22	54
KA	Formantes	1ª	420	435	270	300
		2ª	315	345	330	100
		3ª	100	685	595	750
KE	Formantes	1ª	115	190	176	90
		2ª	30	90	232	150
		3ª	225	300	115	10
KI	Formantes	1ª	110	95	119	90
		2ª	210	180	178	130
		3ª	320	270	60	20
KO	Formantes	1ª	235	460	230	340
		2ª	120	185	177	100
		3ª	450	275	115	80
KU	Formantes	1ª	205	100	180	95
		2ª	100	190	65	355
		3ª	305	340	245	195

Tabela 17 - Resultados Método 1 para sílabas da consoante D

Voluntário		Nº	1	2	3	4
		Sexo	F	F	M	M
		Idade	22	24	22	54
DA	Formantes	1ª	302	345	240	320
		2ª	405	260	288	150
		3ª	800	780	52	80
DE	Formantes	1ª	200	195	170	70
		2ª	100	95	225	150
		3ª	1100	280	60	10
DI	Formantes	1ª	113	95	54	140
		2ª	215	180	112	150
		3ª	15	1610	164	220
DO	Formantes	1ª	220	275	225	48
		2ª	95	180	170	395
		3ª	20	80	55	190
DU	Formantes	1ª	205	185	180	95
		2ª	102	90	60	190
		3ª	20	360	115	10

13.2. Método 2

Frente à inconsistência dos resultados obtidos com o Método 1, decidiu-se realizar uma nova tentativa na obtenção dos dados de determinadas sílabas.
Com a finalidade inicial de apenas validar este novo método, foram gravados apenas 5 sílabas (Tabela 18), utilizando dois voluntários do sexo feminino.
Neste processo, primeiramente o voluntário pronunciou cada uma das sílabas uma única vez, como realizado no Método 1, em seguida fez-se a gravação dos mesmos repetindo-os cinco vezes seguidas. Assim, esperava-se que as dispersões de resultados obtidos com o Método 1 diminuíssem.
Para esta análise de resultados, a programação foi adaptada para se obter resultados mais refinados. Assim, a amostra passou a ser repartida em 30mil pontos, ao invés de 10mil, como mostrado no Código a seguir:

```
clc
clear
[x,y]=loadwave('C:\Users\nome_usuario\Desktop\nome_arquivo.wav');
fx=fft(x(1:30000));
for i=1:30000
fr(i)=(i-1)/(30000*(1/44100));
end
plot(fr,abs(fx));
```

13.2.1. Resultados

Os resultados estão apresentados na Tabela 18. Dela, pode-se fazer a seguinte análise: quando as gravações foram realizadas sem repetição da sílaba, houve uma diferença de 24% em média, entre os resultados da primeira para a segunda voluntária. Já para as gravações com repetição, essa diferença cai para 16%, mostrando uma melhora de 30% nos resultados. Isso porque a repetição reforça no resultado final as frequências pronunciadas repetidas vezes.

Tabela 18 - Resultados Método 2 para sílabas da consoante B, com e sem repetição

Sem repetição						Com repetição					
Voluntário		Nº	1	2	Diferença (%)	Voluntário		Nº	1	2	Diferença (%)
		Sexo	F	F				Sexo	F	F	
		Idade	23	23				Idade	23	23	
BA	Formantes	1ª	2300	2495	8%	BA	Formantes	1ª	2445	2525	3%
		2ª	1680	1225	37%			2ª	1815	1262	44%
		3ª	565	230	146%			3ª	610	229	166%
BE	Formantes	1ª	1165	1023	14%	BE	Formantes	1ª	1245	1210	3%
		2ª	582	522	11%			2ª	618	610	1%
		3ª	229	194	18%			3ª	229	229	0%
BI	Formantes	1ª	1200	1230	2%	BI	Formantes	1ª	1420	1325	7%
		2ª	600	618	3%			2ª	697	662	5%
		3ª	230	229	0%			3ª	230	230	0%
BO	Formantes	1ª	1358	1050	29%	BO	Formantes	1ª	1250	1245	0%
		2ª	592	510	16%			2ª	625	627	0%
		3ª	230	195	18%			3ª	229	230	0%
BU	Formantes	1ª	1112	1305	15%	BU	Formantes	1ª	1375	1360	1%
		2ª	556	653	15%			2ª	685	680	1%
		3ª	235	318	26%			3ª	230	229	0%
Média das diferenças					24%	Média das diferenças					16%

13.3. Método 3

Podemos observar que a tabela obtida pelo segundo método possui resultados mais semelhantes entre os participantes, do que os resultados obtidos com a aplicação do primeiro método. Assim, conclui-se que a gravação dos fonemas utilizando repetições sucessivas, apresentou um refino considerável dos resultados.

Posto isso, partiu-se para uma terceira tentativa de gravação, fazendo com que o número de repetições das sílabas aumentasse para dez. Nos gráficos do item 13.3.1 apresentam-se a comparação de alguns dos resultados obtidos com o método 3, com sinais gravados por um voluntário do sexo feminino, de 19 anos.

O código numérico utilizado para tal apresenta-se abaixo, e nada mais é do que o código utilizado desde o começo do estudo, porém com uma maior discretização do sinal, para obtenção de um gráfico com mais clareza na distinção entre as principais formantes.

```
clc
clear
 [x,y]=loadwave('C:\Users\ nome_usuario\Desktop\ pasta\arquivo.wav');
fx=fft(x(1:50000));
for i=1:50000
fr(i)=(i-1)/(50000*(1/44100));
end
plot(fr,abs(fx));
```

13.3.1. Resultados

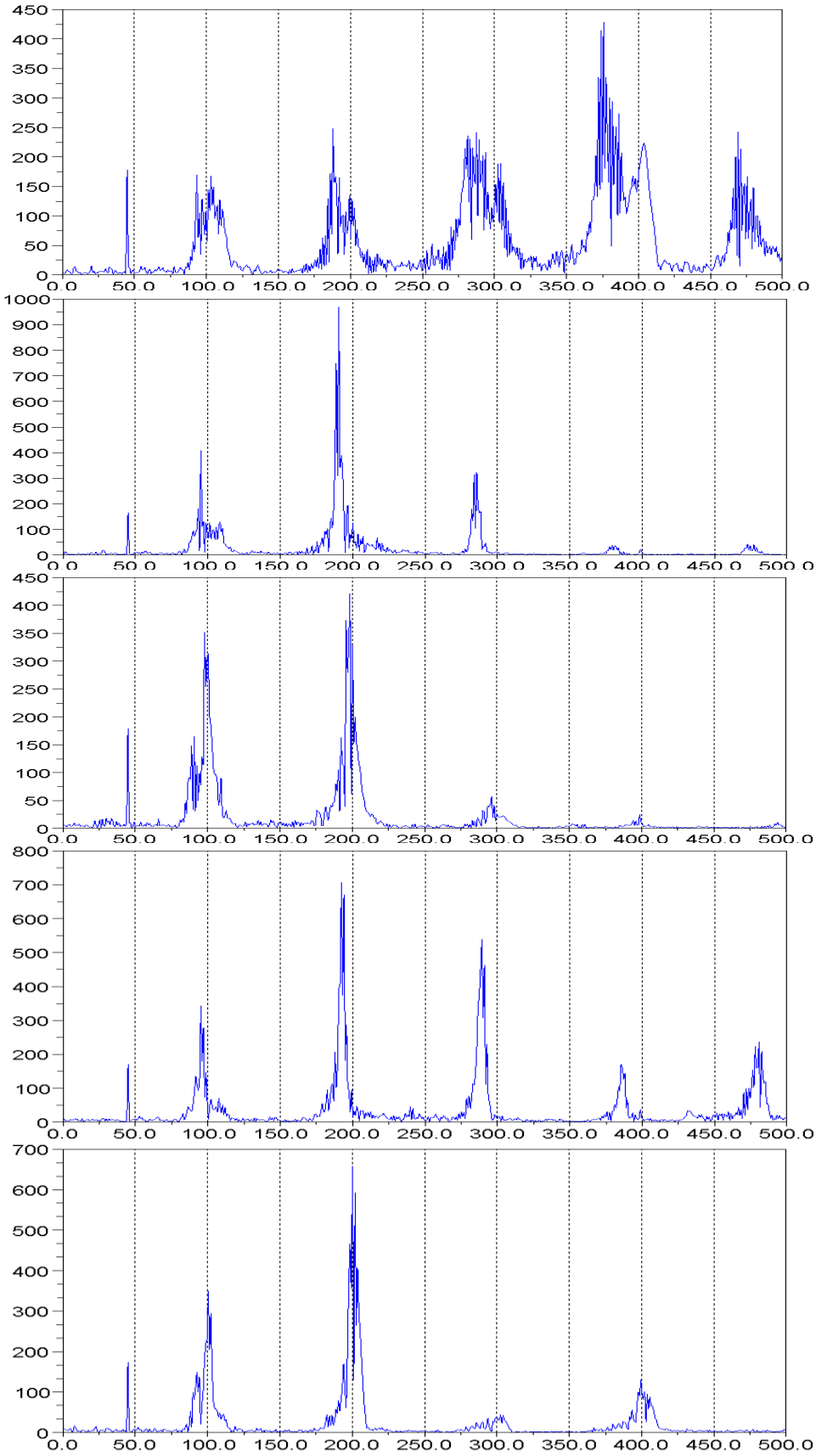

Gráfico 10 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /BA/ /BE/ /BI/ /BO/ /BU/

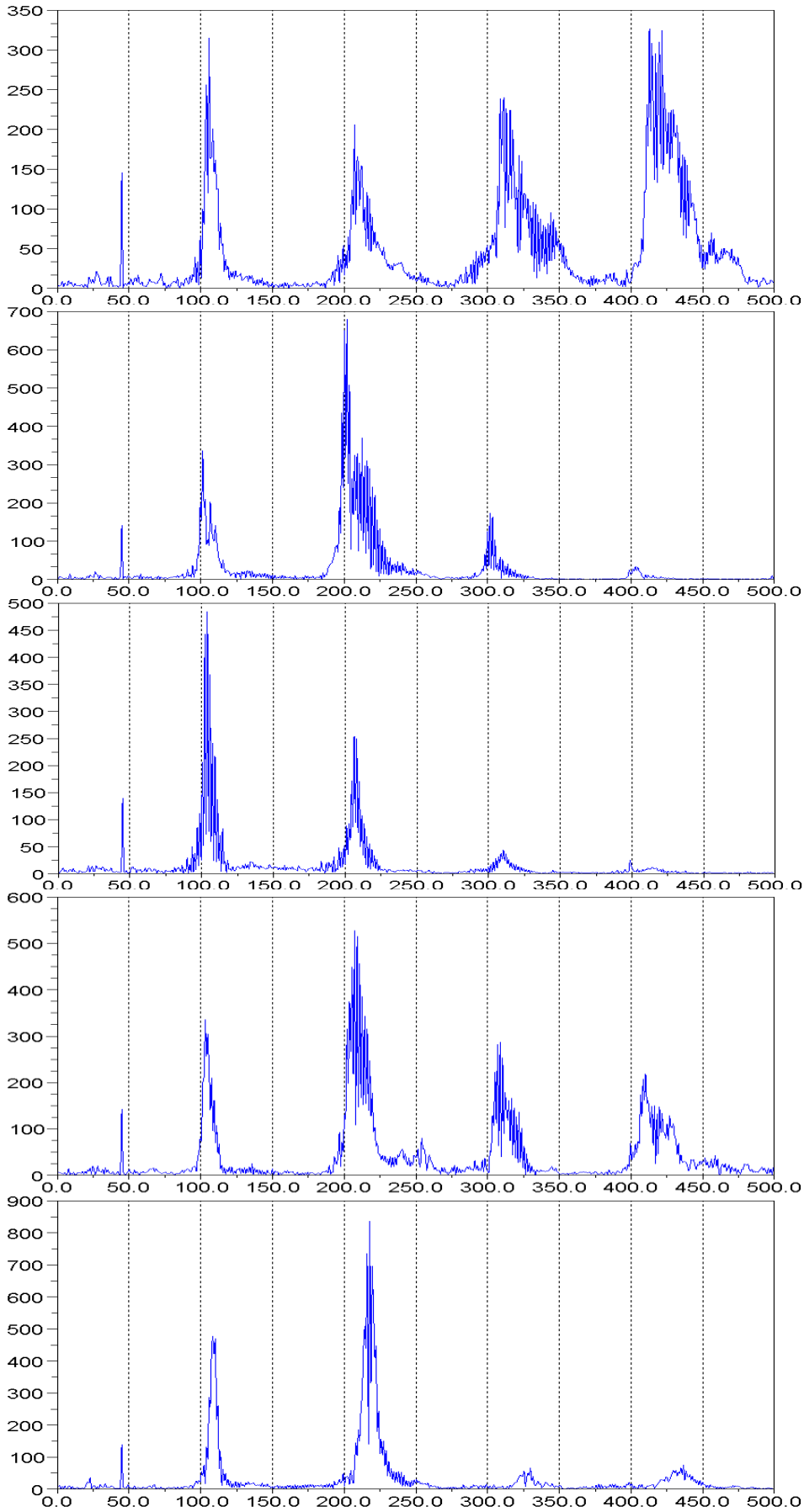

Gráfico 11 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /KA/ /KE/ /KI/ /KO/ /KU/

Dado a semelhança entre os gráficos obtidos para uma mesma vogal, nos Gráficos 10 e 11 resolveu-se analisá-los de forma a tentar identificar alguma diferença entre sílabas formadas pela mesma vogal com consoantes diferentes. Os gráficos a seguir mostram algumas dessas comparações:

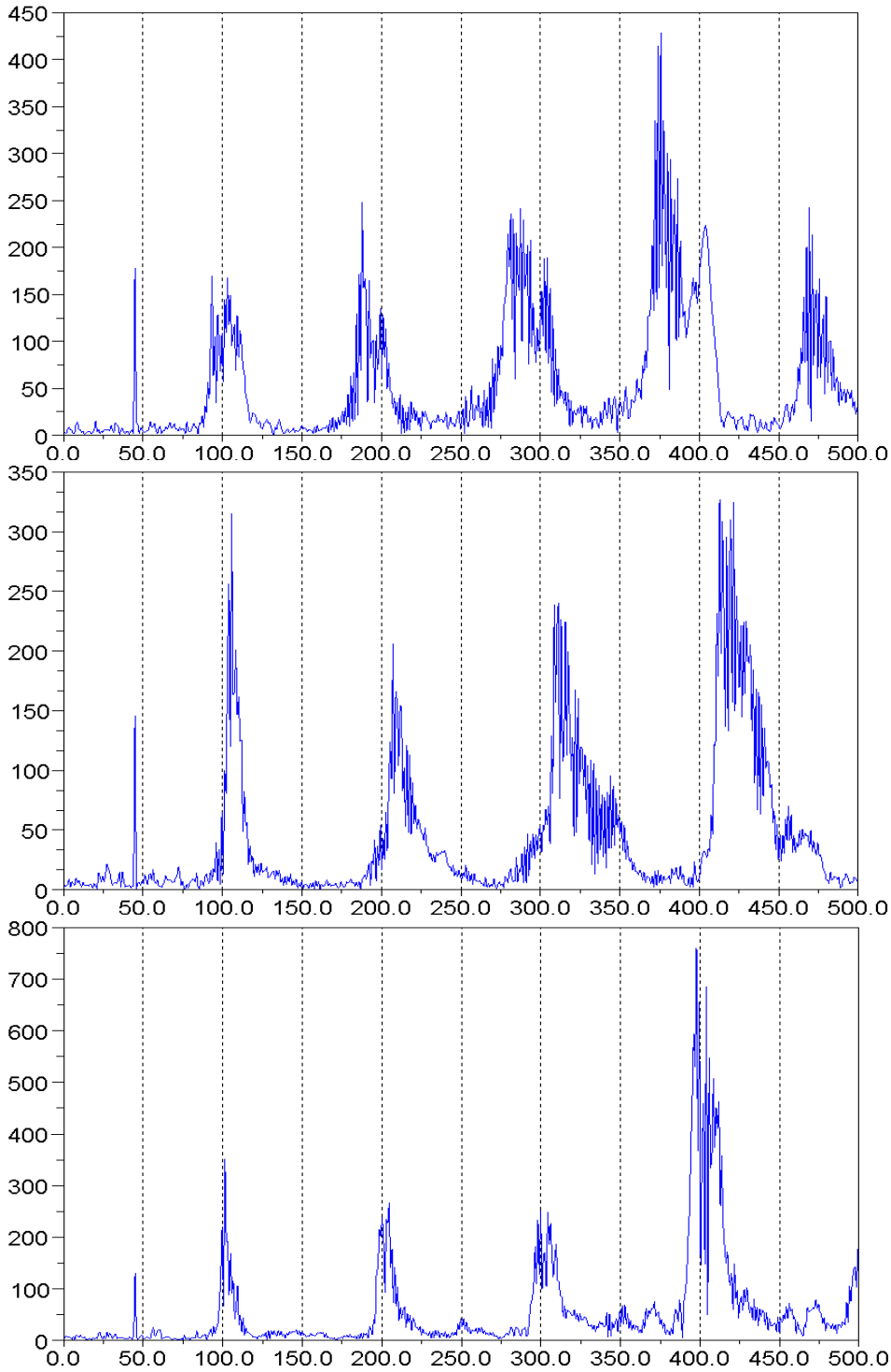

Gráfico 12 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /BA/ /KA/ /FA/

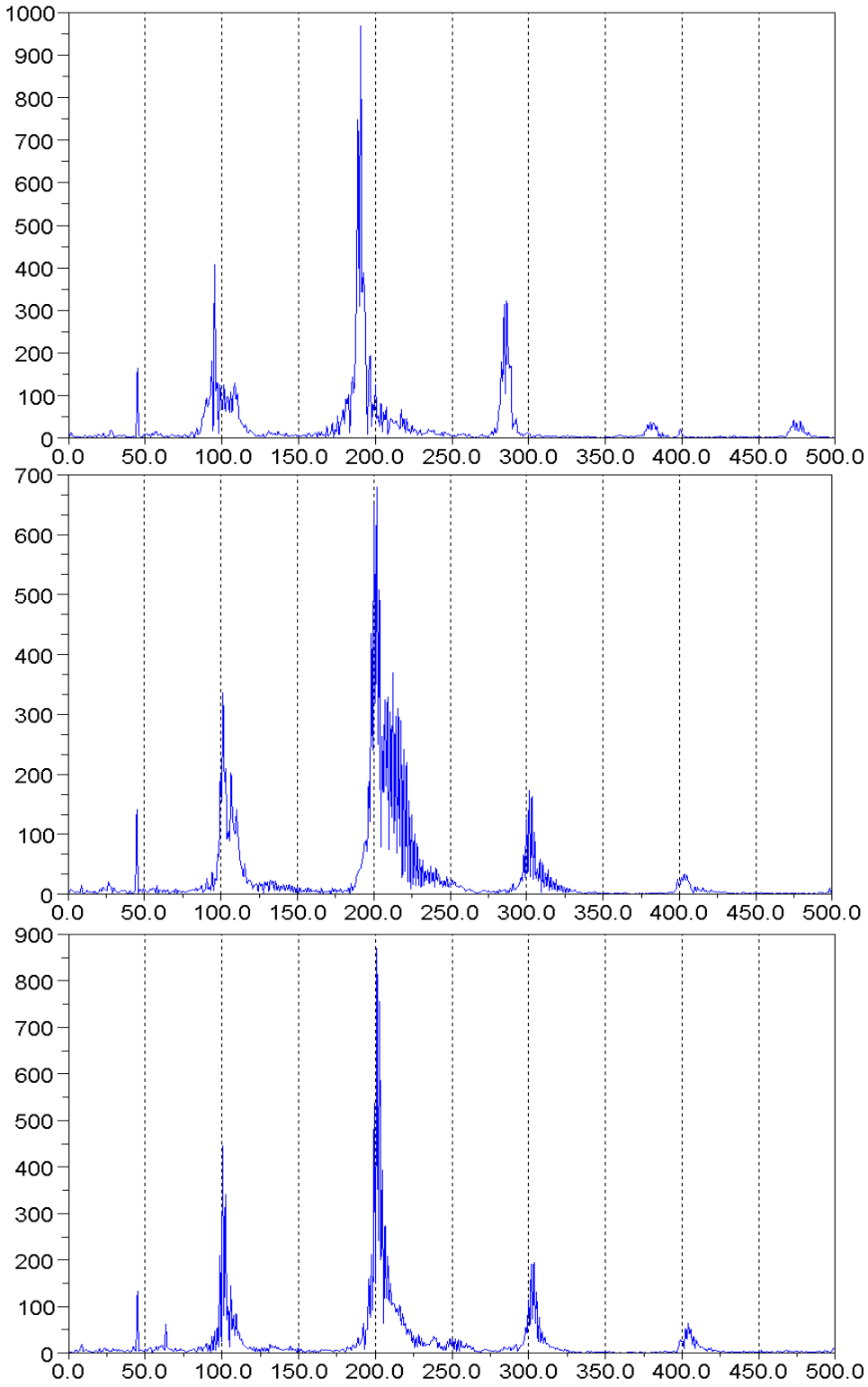

Gráfico 13 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /BE/ /KE/ /FE/

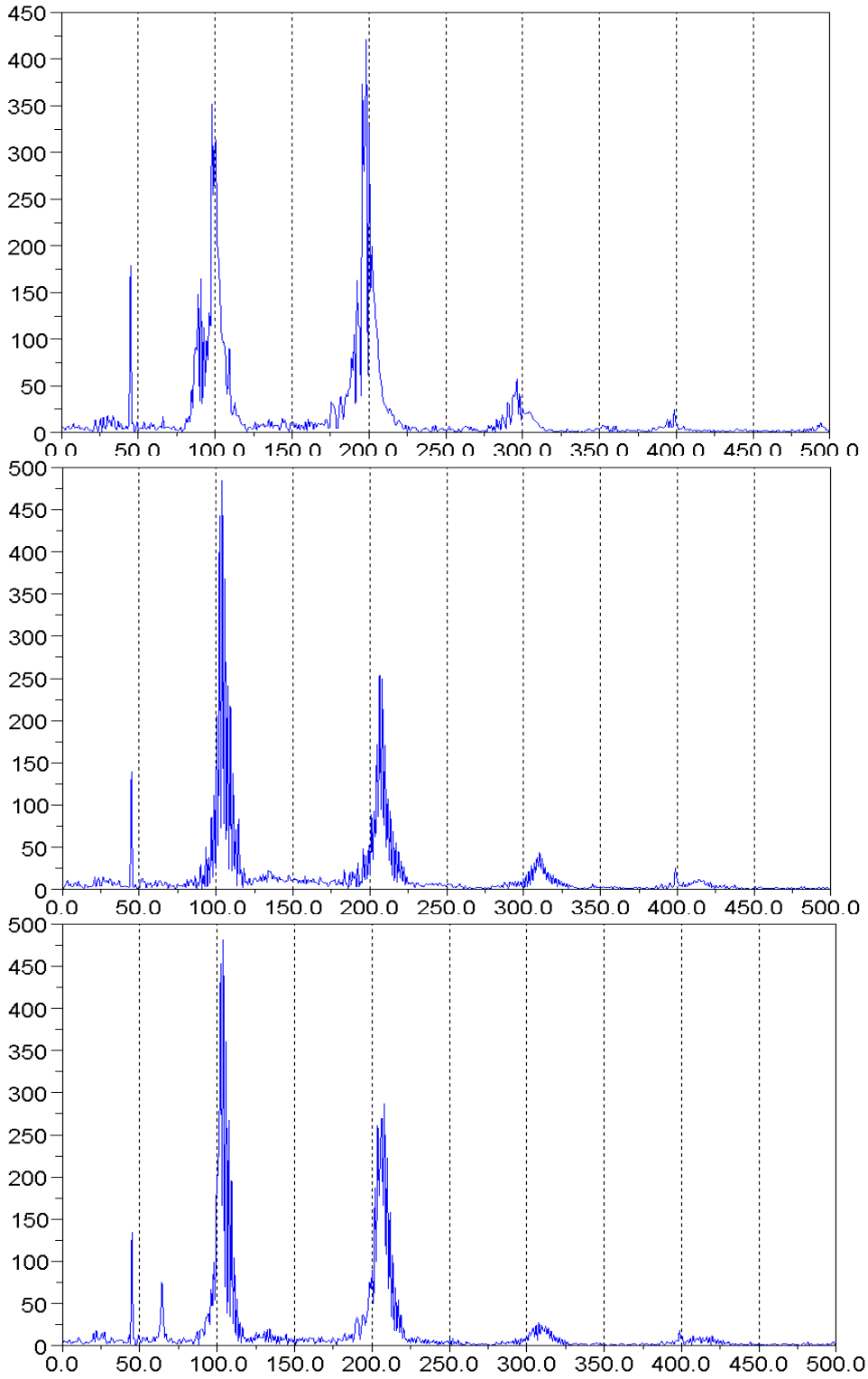

Gráfico 14 - Gráficos da intensidade x freqüência para as sílabas, de cima para baixo: /BI/ /KI/ /FI/

Apesar de muito semelhantes, vemos que o gráfico do /BA/ fica deslocado para a esquerda em relação ao do /FA/, enquanto o do /KA/ fica à direita do gráfico do /FA/, ou seja, com frequências mais altas. Além disso, as frequências com maiores intensidades, apesar de parecidas, diferem de sílaba para sílaba. Suas principais formantes resumem-se na tabela a seguir:

Tabela 19 - Resultados Método 3 para algumas sílabas com vogal A

Sílaba	Formantes (Hz)		
/BA/	190	280	375
/KA/	105	310	420
/FA/	100	200	395

Já no Gráfico 13, as sílabas /KE/ e /FE/ quase não apresentam diferença, sendo suas principais formantes:

Tabela 20 - Resultados Método 3 para algumas sílabas com vogal E

Sílaba	Formantes (Hz)		
/BE/	95	190	275
/KE/	100	200	300
/FE/	100	200	300

No Gráfico 14, todas as sílabas possuem gráficos quase idênticos e o mesmo acontece para os demais gráficos aqui não apresentados, mas com resultados registrados a seguir:

Tabela 21 - Resultados Método 3 para algumas sílabas com vogal I

Sílaba	Formantes (Hz)		
/BI/	48	100	200
/KI/	48	102	203
/FI/	48	102	205

Tabela 22 - Resultados Método 3 para algumas sílabas com vogal O

Sílaba	Formantes (Hz)		
/BO/	95	195	280
/KO/	102	205	305
/FO/	102	205	305

Tabela 23 - Resultados Método 3 para algumas sílabas com vogal U

Sílaba	Formantes (Hz)		
/BU/	45	100	200
/KU/	45	105	220
/FU/	45	105	222

Podemos concluir com o Método 3, que o código utilizado, apesar de conseguir identificar bem as formantes predominantes na sílaba, não distingue o que seriam as formantes da consoante da mesma. Assim, vemos que a característica principal captada com este método é a vogal que acompanha a consoante, e não esta. Isso porque a vogal possui maior presença temporal na sílaba do que a consoante.

13.4. 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 freqüência 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));
```

13.4.1. Comparação dos resultados com a referência

Nesta etapa, utilizou-se como referência a obra da especialista Dra. Mara Behlau [1]. Esta obra reúne o conteúdo científico do Curso de Especialização em Voz do Centro de Estudos da Voz - CECEV, de São Paulo, organizado e desenvolvido pela coordenadora, Prof. Dra. Mara Behlau.

O Gráfico 15 apresenta o resultado obtido com o código descrito anteriormente para a vogal A, gravado por uma voluntária do sexo feminino de 23 anos. Já o Gráfico 16, apresenta a mesma vogal, gravada por um voluntário do sexo masculino de 22 anos.

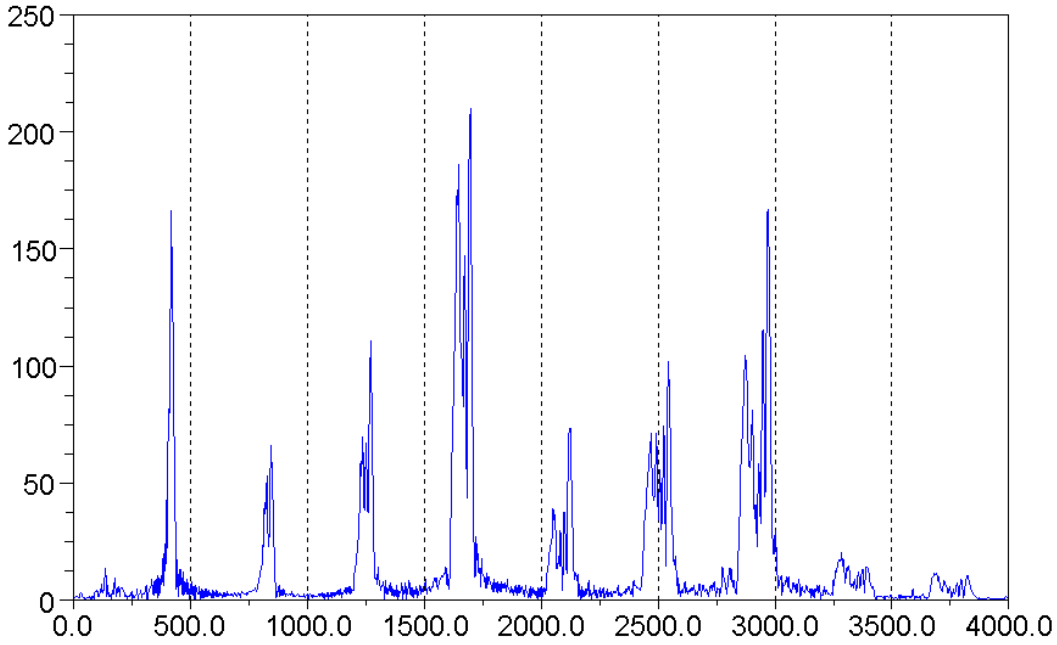

Gráfico 15 - Gráficos da intensidade x freqüência para a vogal /A/, voluntária feminina

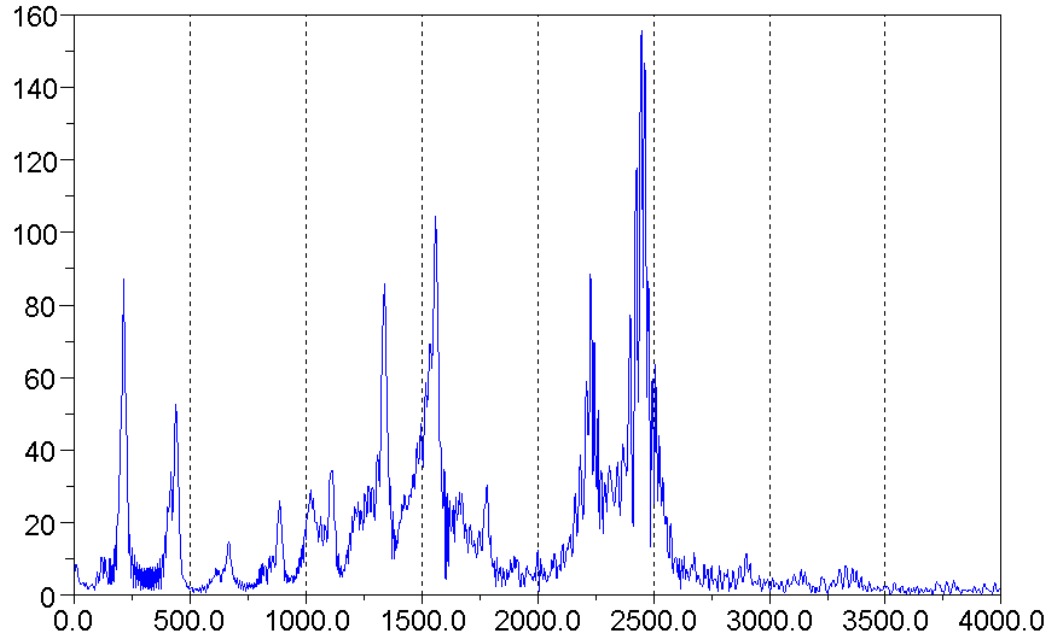

Gráfico 16 - Gráficos da intensidade x freqüência para a vogal /A/, voluntário masculino

A Tabela 24 apresenta a comparação entre os dados retirados da referência e os obtidos pelos Gráficos 15 e 16:

Tabela 24 - Comparação de resultado com referência, para vogal /A/

Homens	Formantes	/A/	Mulheres	Formantes	/A/
Referência	F1	730	Referência	F1	925
	F2	1.273		F2	1.767
	F3	2.425		F3	3.089
Experimental	F1	240	Experimental	F1	450
	F2	1.550		F2	1.700
	F3	2.450		F3	2.997
Diferença (%)	F1	67%	Diferença (%)	F1	51%
	F2	22%		F2	4%
	F3	1%		F3	3%
	Média (F1,F2,F3)	30%		Média (F1,F2,F3)	19%
	Média (F2,F3)	11%		Média (F2,F3)	4%

A diferença percentual média é de 30% entre os resultados obtidos para o voluntário masculino e a referência, enquanto que para a voluntária feminina, essa diferença cai para 19%. Porém, notou-se que desconsiderando a primeira formante, a mais destoante em relação à referência em ambos os casos, essas diferenças diminuem consideravelmente para 11% e 4%, respectivamente.

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.

Devido à semelhança na maioria dos resultados, admitiu-se que pequenas diferenças seriam resultado da comparação considerando apenas um voluntário de cada sexo, e não uma amostra. Assim, era esperada uma

variação, já que a análise de um único indivíduo deverá apresentar resultados diferentes quanto comparados aos resultados médios de 180 indivíduos (número de participantes da pesquisa coordenada pela Dra. Behlau), estes apresentando maior confiabilidade.

Cerca de 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.

Assim, 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.

13.4.2. Comparação entre voluntário feminino e masculino

Com a aplicação do Método 4 de processamento de sinais, obteve-se a seguinte tabela de resultados:

Tabela 25 - Comparação de resultados, entre voluntário feminino e masculino

	Formantes	/A/	/Ê/	/I/	/Ô/	/U/	/É/	/Ó/
Masculino	F1	240	450	400	650	400	1.150	450
	F2	1.550	650	650	850	700	1.350	1.100
	F3	2.450	900	980	1.570	1.700	3.600	1.300
Feminino	F1	450	450	450	450	450	1.200	450
	F2	1.700	800	850	900	950	1.470	1.250
	F3	2.997	1190	1400	1800	2100	3850	1700
Diferença (Fem, Masc)		302	147	223	160	233	140	183

Dela, observa-se que a diferença média de resultados entre os voluntários feminino e masculino é de 198Hz. Segundo a obra de referência, para estas mesmas vogais, essa diferença é de 221Hz. Como a mesma obra, no final do estudo, conclui que a diferença média de frequências entre homens e mulheres é de 100Hz, pode-se dizer que o resultado é razoavelmente satisfatório, dentro

dos padrões descritos anteriormente, porém com uma pequena melhora de 19% na diferença de resultados entre os sexos.

13.4.3. Vogais nasais e vogais orais

Aplicando novamente o Método 4, segue os resultados obtidos para as gravações de vogais nasais, feitos pelos mesmos voluntários que gravaram as vogais orais. Assim, pode-se seguir com uma comparação entre estes dois tipos de vogais que são presentes no português brasileiro.

Tabela 26 - 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 freqüência 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, caracterizando as formantes deste tipo de vogal.

Outro ponto interessante, é que os gráficos das vogais nasais, tiveram formatos muito semelhantes entre os voluntários. Os Gráficos 17 e 18 exemplificam com as vogais /IM/ e /UM/ essa semelhança:

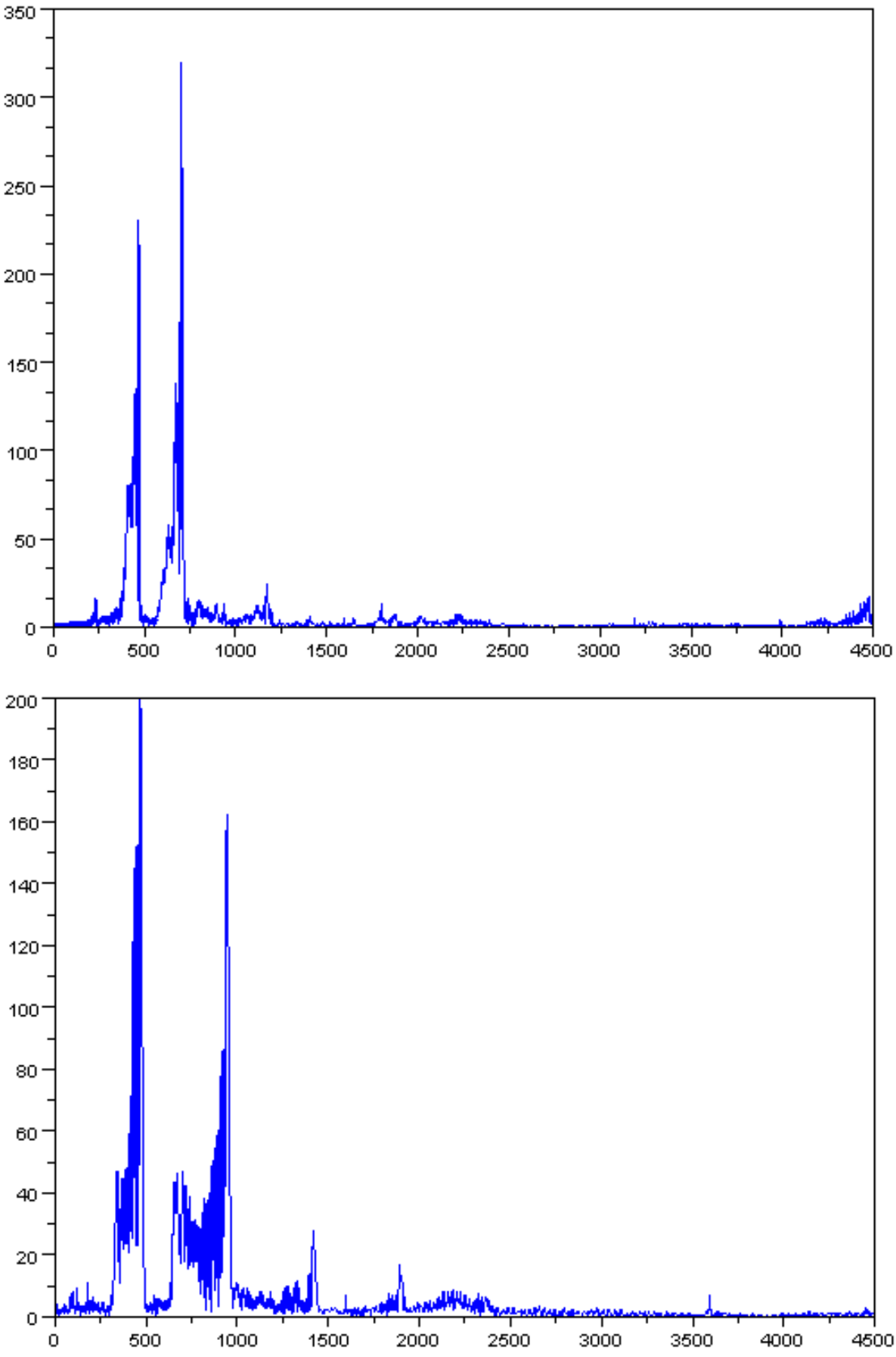

Gráfico 17 - Gráficos da vogal nasal /IM/, respectivamente, do voluntário masculino e feminino

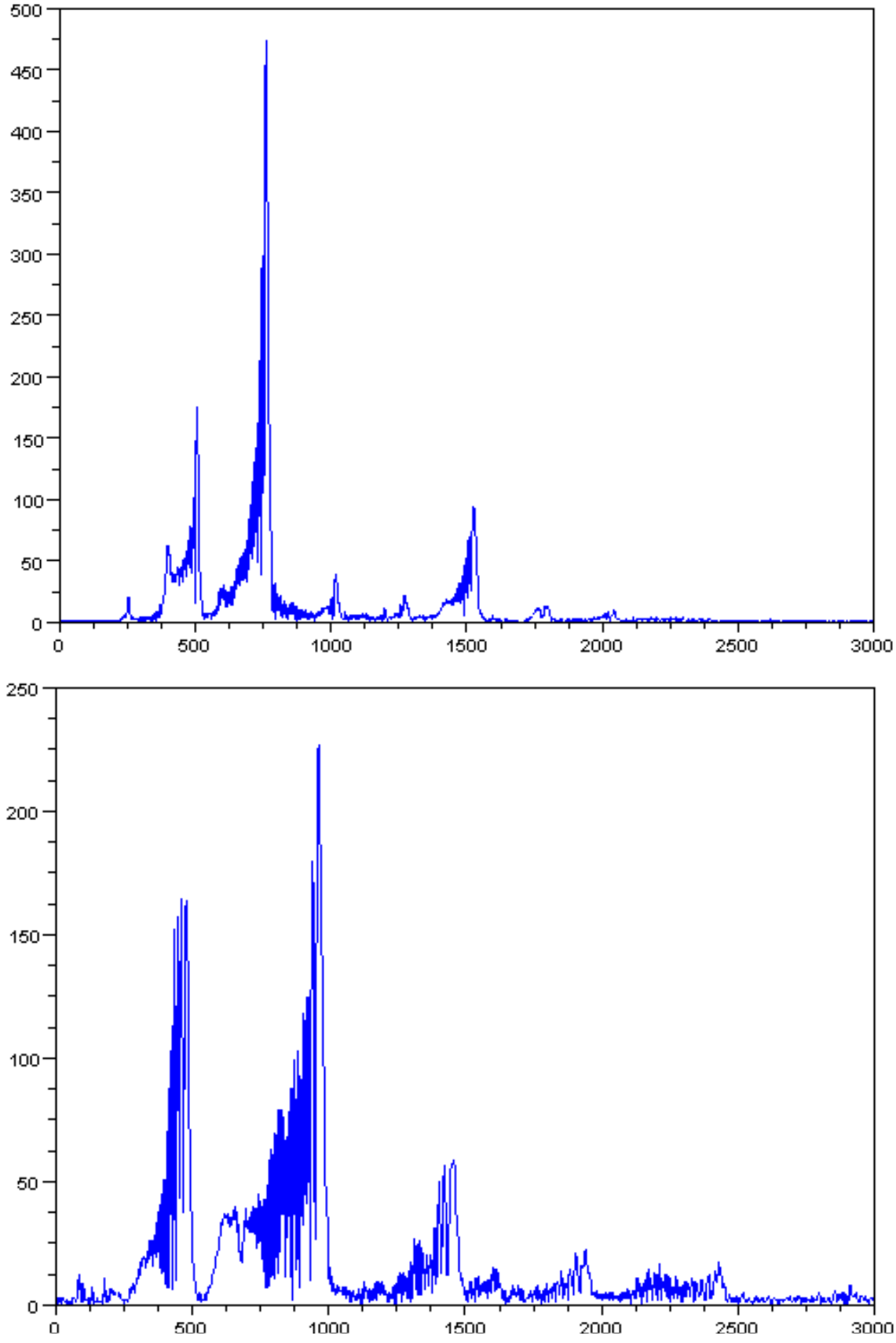

Gráfico 18 - Gráficos da vogal nasal /UM/, respectivamente, do voluntário masculino e feminino

13.5. Método 5

Finalmente, para obtenção das formantes das consoantes, criou-se uma nova análise, já que o Método 3 mostrou ser incapaz de fazer o mesmo.

Este método utilizou como base de análise todas as sílabas constituintes do português brasileiro e consistiu 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 Rápida 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 cada uma dessas 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 formante da sílaba.

A fim de verificar as principais frequências das consoantes viu-se a semelhança da primeira seção do eixo x, para todas as sílabas formadas por uma mesma consoante.

Para compor o Banco de fonemas utilizou-se uma média da principal frequência de todas as sílabas com a mesma consoante.

O código numérico utilizado para tal apresenta-se a seguir.

```
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
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));
```

13.5.1. Gráficos

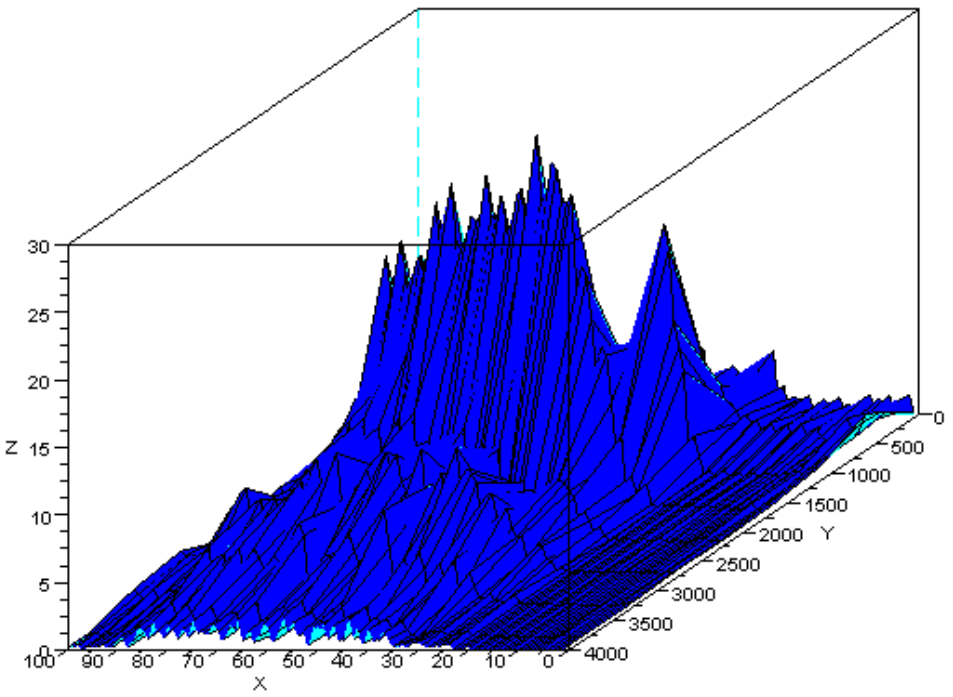

Gráfico 19 - Gráfico 3D referente à sílaba /BA/

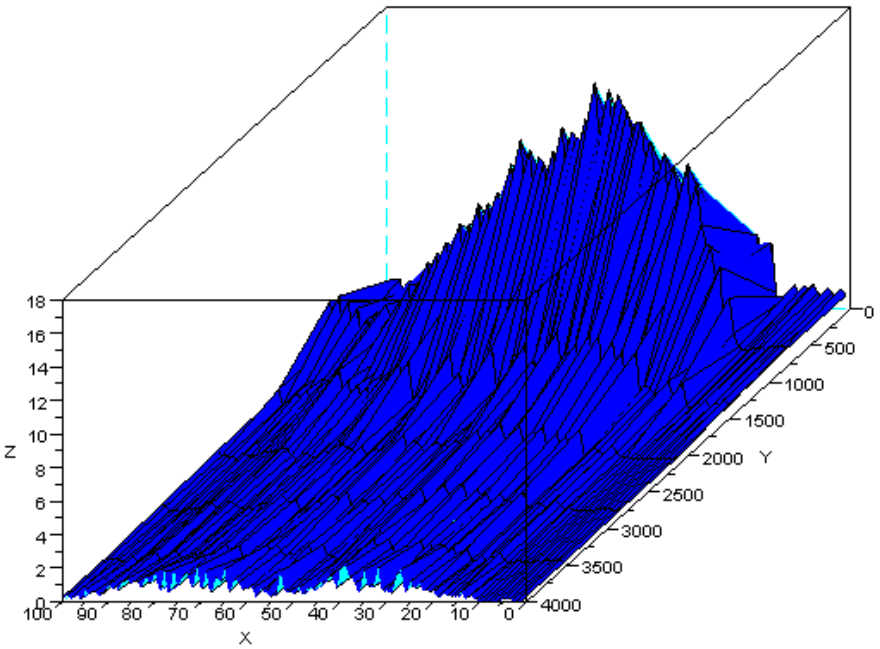

Gráfico 20 - Gráfico 3D referente à sílaba /BE/

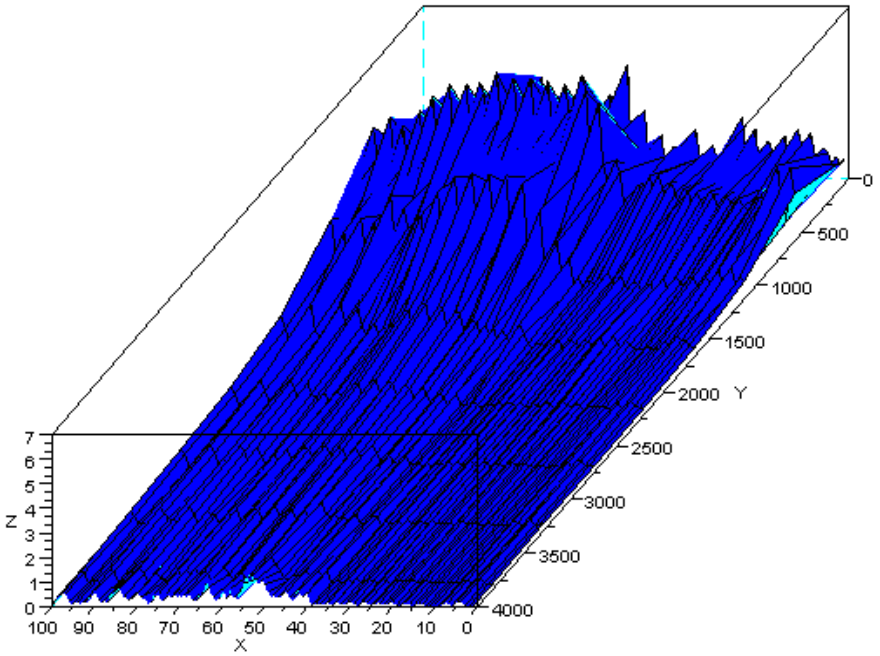

Gráfico 21 - Gráfico 3D referente à sílaba /BI/

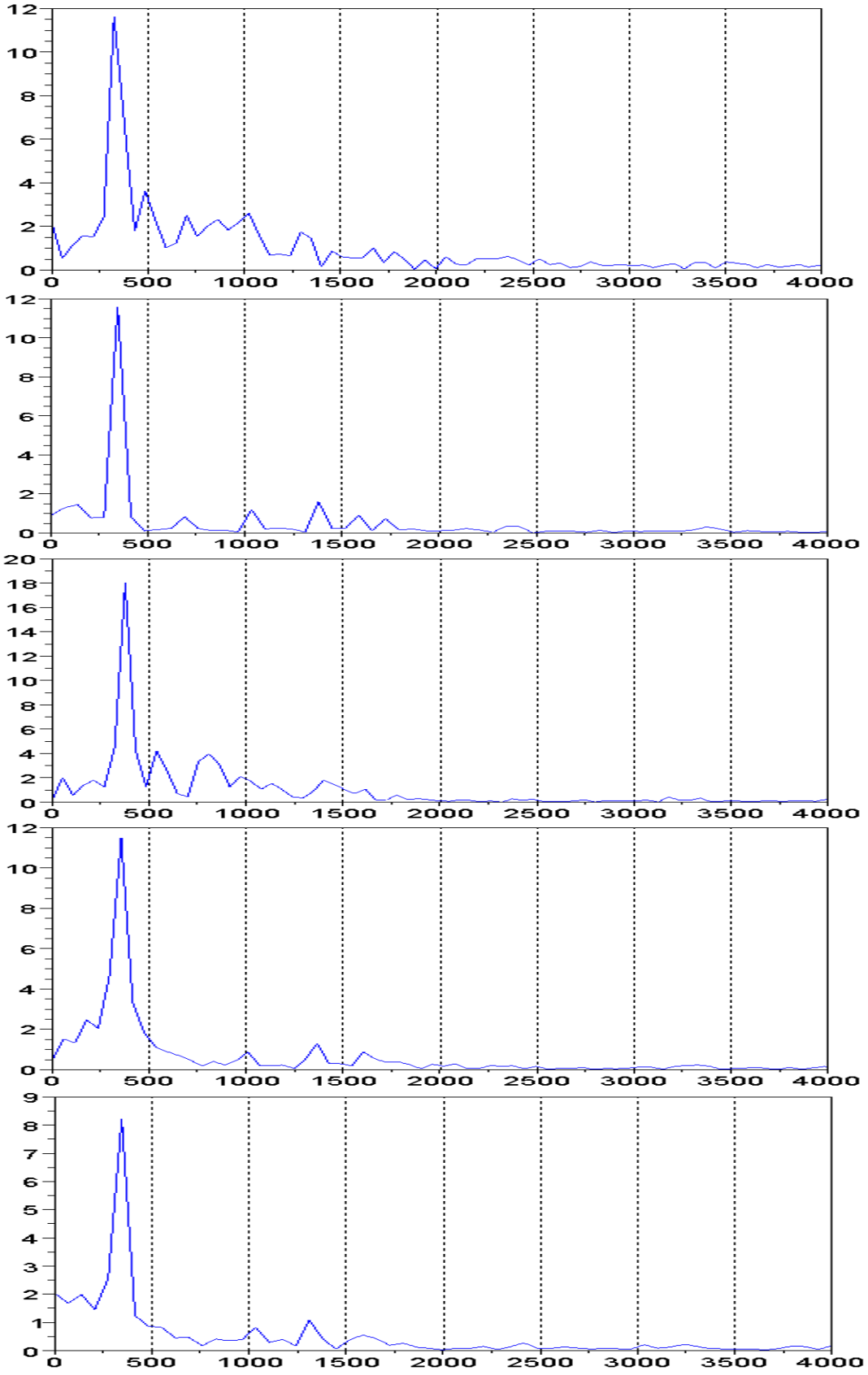

Gráfico 22 - Gráficos referentes ao fonema /B/, considerando de cima para baixo as sílabas: /BA/ /BE/ /BI/ /BO/ /BU/

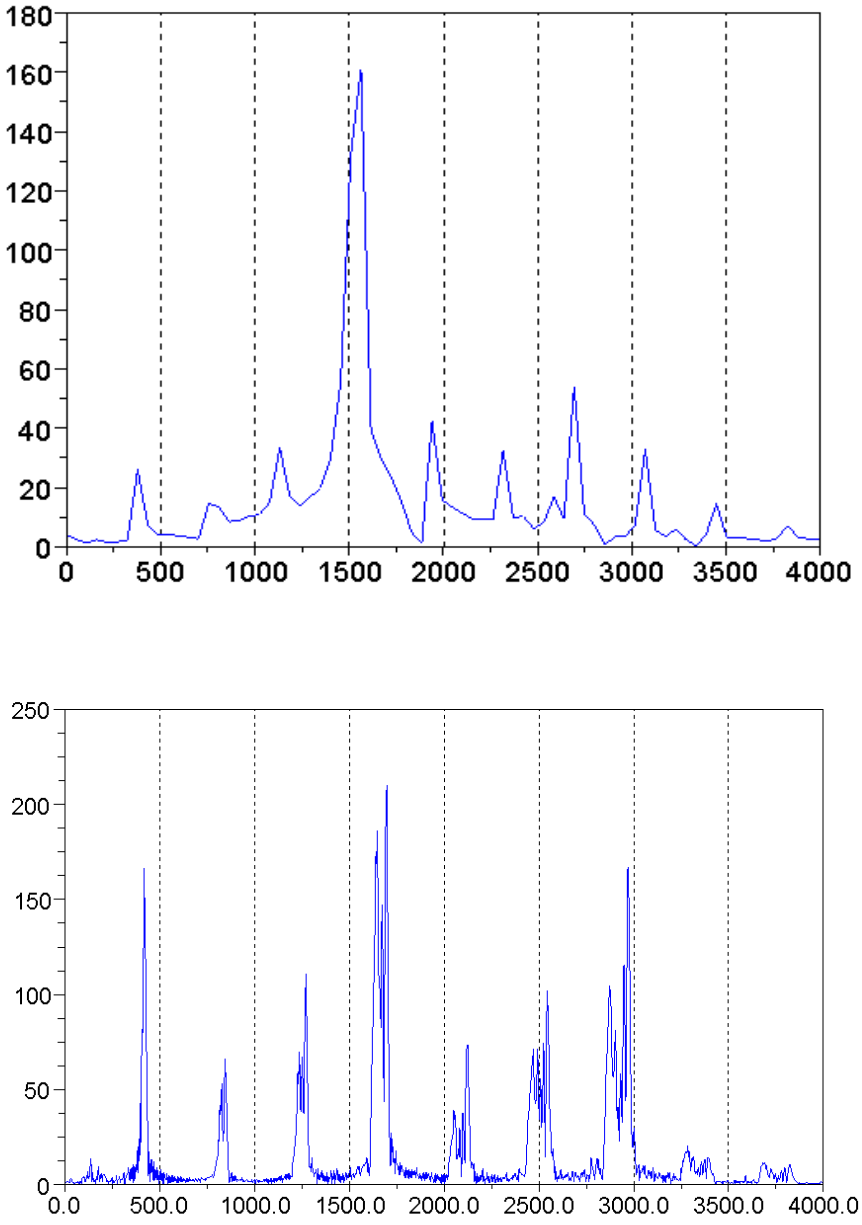

Gráfico 23 – Comparação entre os fonemas vocálicos /A/. Em cima: fonema inserido na sílaba /BA/. Em baixo: fonema analisado isoladamente.

A partir dos gráficos apresentados anteriormente foi então possível construir o banco de fonemas almejado, contendo todos os fonemas constituintes da língua portuguesa falada no Brasil.

O Gráfico 22 ressalta a semelhança entre os trechos iniciais das gravações de áudio referentes às sílabas formadas pelo fonema /B/. Porém, é possível observar 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, os quais estão apresentados a seguir no item 13.6.

Vale ressaltar que apesar de ser possível identificar um pico bem definido no gráfico, a intensidade da frequência observada para o fonema consonantal é consideravelmente menor que as intensidades observadas para o fonema vocálico.

O Gráfico 23 está apresentado no intuito de validar a análise realizada, pois demonstra a semelhança entre o trecho final da sílaba /BA/, que espera-se que corresponda basicamente ao fonema vocálico em questão (gráfico de cima) e o gráfico da FFT aplicada ao fonema vocálico isolado (gráfico de baixo).

Apesar da comparação feita entre a voz masculina e a feminina na reprodução dos fonemas vocálicos, considerou-se desnecessária a mesma avaliação para os fonemas consonantais, devido ao desvio percentual baixo entre eles.

13.6. Banco de Fonemas Limitado do Português Brasileiro

Tabela 27 - Principal formante de cada fonema consonantal

Fonema	Frequência (Hz)
/p/	1000
/b/	340
/m/	450
/f/	289
/v/	460
/t/	250
/d/	395
/n/	285
/ñ/	470
/l/	438
/λ/	370
/r/	510
/R/	1475
/z/	330
/s/	200
/j/	385
/x/	240
/g/	388
/k/	650

Tabela 28 - As três principais formantes de cada fonema vocálico oral e nasal

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
Masculino	Formantes	/É/	/Ó/			
Orais	F1	1150	450			
	F2	1350	1100			
	F3	3600	1300			
Feminino	Formantes	/É/	/Ó/			
Orais	F1	1470	450			
	F2	3850	1250			
	F3	140	1700			

14. Análise dos Resultados

Muito estudo e pesquisa foram realizados durante o desenvolvimento deste trabalho. Somado a isso, a comparação com a literatura específica demonstrou a coerência dos resultados obtidos.

Apesar disso, deve-se ressaltar que os resultados podem sofrer alterações, inclusive devido às mutações inerentes à língua portuguesa falada no Brasil. E porque este estudo e os de muitos especialistas, como os da Dra. Mara Behlau, ainda não foram capazes de incorporar todas as variantes da língua, principalmente os regionalismos e sotaques presentes na língua, devido ao país possuir dimensões continentais.

Como visto no item 10, primeiramente comprovou-se a eficácia do tratamento que seria dado a todas as gravações, com o uso de notas musicais. Porém, com o passar do tempo, foram detectadas falhas na aplicação do tratamento escolhido. Estas foram sendo sanadas, durante o item 12.

Devido ao fato de não existir um Banco de Fonemas Nacional, não houve base de comparação para confirmar ou desprezar os resultados obtidos. Apenas tivemos base para dizer se estávamos caminhando no mesmo sentido que alguns outros pesquisadores da área.

Realizamos comparações com um estudo realizado, na década de 90, por uma renomada especialista na área da voz, Dra. Mara Behlau. Em seguida, foi realizada uma breve comparação entre os fonemas vocálicos e orais. O objetivo disso foi mostrar que as vogais nasais, características 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 brasileiros. Essa diferença de frequências foi mostrada na Tabela 26, 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 13.5, concluindo o trabalho com um esboço do que pode vir a ser o Banco de Fonemas do Português Brasileiro.

Devido à limitação do tempo de nossa pesquisa, não aplicamos os métodos desenvolvidos em uma grande amostra de voluntários, o que daria maior

credibilidade ao nosso Banco. Porém, acredita-se que este encontra-se encaminhado, e somado às considerações de variações considerando a idade dos voluntários e características como os regionalismos seriam os próximos passos para alcançar o objetivo final deste estudo.

Bibliografia

[1] BEHLAU, M. Voz – O Livro do Especialista. Editora Revinter. 2005. 1ª Edição. 576 p.

[2] GIL, B. D.; CARDOSO, E. A.; CONDÉ, V. G. Modelos de Análise Linguística. Editora Contexto. 2009. 256 p.

[3] Site da Sociedade Brasileira de Otologia sobre a Campanha da Saúde Auditiva. Disponível em: <http://www.saudeauditiva.org.br/artigos/artigos_detalhe.asp?id=12>. Acesso em: outubro/2010.

[4] BEVILACQUA, M. C.; COSTA FILHO, O. A. Implante Coclear. Disponível em: <http://www.implantecoclear.com.br>. Acesso em: outubro/2010.

[5] Site do Grupo de Implante Coclear do Hospital das Clínicas e FMUSP. Disponível em: <http://www.implantecoclear.org.br>. Acesso em: novembro/2010.

[6] Trabalho sobre Implantes Cocleares disponibilizado pela FORL (Fundação de Otorrinolaringologia). Disponível em: <http://www.forl.org.br/pdf/seminarios/seminario_1.pdf>. Acesso em: novembro/2010.

[7] Artigo sobre Estratégia de Codificação de Fala do Implante Coclear disponibilizado pelo Grupo de Implante Coclear do Hospital das Clínicas e FMUSP.

[8] Trabalho sobre Processamento do sinal de som disponibilizado pela Universidade Federal de Santa Catarina. Disponível em: <http://www.eps.ufsc.br/disserta96/tafner/cap5/cap5.htm>. Acesso em: março/2011.

[9] Site da Wikipedia sobre Filtro de transformada de Fourier. Disponível em:

<http://pt.wikipedia.org/wiki/Filtro_de_Transformada_de_Fourier>. Acesso em: fevereiro/2011.

[10] Trabalho sobre a Transformada de Fourier e suas aplicações disponibilizado pela Universidade Federal de Campina Grande. Disponível em: <http://www.dsc.ufcg.edu.br/~pet/atividades/ciclo_seminarios/tecnicos/2010/TransformadaDeFourier.pdf>. Acesso em: janeiro/2011.

[11] Site do Prof. Luiz Netto sobre as características do som. Disponível em: <http://caraipora.tripod.com/fourier.htm>. Acesso em: fevereiro/2011.

[12] Material sobre Transformada de Fourier do Prof. Ulysses Sodré da Universidade Estadual de Londrina. Disponível em: <http://www.mat.uel.br/matessencial/superior/pdfs.tfourier.pdf>. Acesso em: janeiro/2011.

[13] Material sobre Transformada de Fourier disponibilizado pela Universidade Católica de Pelotas. Disponível em: <http://paginas.ucpel.tche.br/~tst/fourier.html>. Acesso em: março/2011.

[14] Estudo de Transformada de Fourier em Basic do Departamento de Química e da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo. Disponível em: <http://www.scielo.br/pdf/qn/v23n3/2830.pdf>. Acesso em: fevereiro/2011.

[15] Estudo sobre o Teorema de Nyquist e Áudio Digital disponibilizado pela Escola de Comunicações e Artes da Universidade de São Paulo. Disponível em: <http://www.eca.usp.br/prof/iazzetta/tutor/audio/a_digital/a_digital.html>. Acesso em: março/2011.

[16] Blog do Grupo SULP (Surdos Usuários da Língua Portuguesa) a respeito de implantes cocleares. Disponível em: <http://sulp-surdosusuariosdalinguaportuguesa.blogspot.com/2011/05/implante-coclear-pelo-sus-tambem-em.html>. Acesso em: maio/2011.

[17] Trabalho da Universidade Católica de Brasília a respeito da relação existente entre a música e a matemática. Disponível em: < http://www.ucb.br/sites/100/103/TCC/22006/DyegoRaphaelAlvesdosSantos.pdf >. Acesso em: maio/2011.

[18] Site da empresa Organia Engenharia Musical Tecnologia Audiovisual sobre estudo e análises de flauta. Disponível em: < http://www.organia.com.br/sites/regis/an_flauta.pdf>. Acesso em: maio/2011.

[19] Site So Portugues a respeito de fonemas. Disponível em: <http://www.soportugues.com.br/secoes/fono/fono1.php>. Acesso em: abril/2011.

[20] Site Wikipedia sobre fonemas e suas classificações. Disponível em: <http://pt.wikipedia.org/wiki/Fonema>. Acesso em: abril/2011.

[21] Site Wikipedia sobre a fonologia da língua portuguesa. Disponível em: <http://pt.wikipedia.org/wiki/Fonologia_do_portugu%C3%AAs>. Acesso em: abril/2011.

ANEXO A

TEOREMA DE NYQUIST

O Teorema de Nyquist propõe que a taxa de amostragem deve ser pelo menos duas vezes a maior frequência que se deseja registrar [15]. Este valor é denominado frequência de Nyquist.

Esta proposta pode ser explicada pela necessidade de manter uma determinada qualidade após a digitalização de um sinal analógico de áudio qualquer, e a posterior operação inversa, que é a decodificação deste sinal em analógico novamente.

A conversão analógico-digital captura amostras do sinal analógico ao longo do tempo na frequência que corresponde à taxa de amostragem. Cada amostra é convertida em um número (em bits), levando em consideração seu nível de tensão.

Durante a codificação digital-analógico os bits são convertidos de volta em tensões. A forma de onda resultante da codificação digital-analógico não será perfeita, uma vez que ela não possuirá todos os pontos do sinal analógico original, apenas alguns deles. Assim, o conversor digital-analógico conectará todos os pontos capturados pelo conversor analógico-digital e qualquer valor que existira originalmente entre esses pontos será descartado.

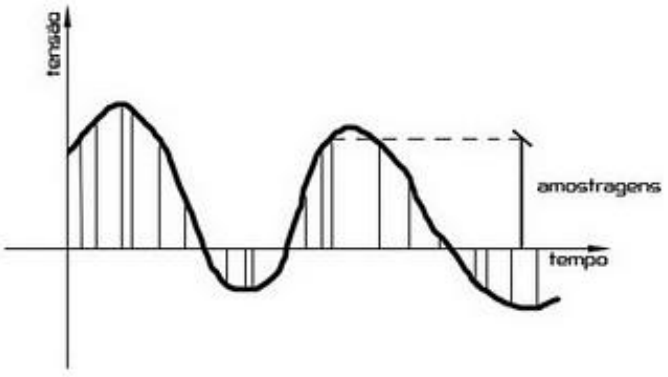

Gráfico 24 - Sinal analógico original

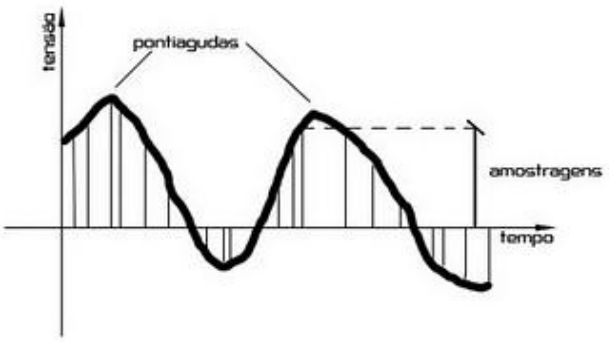

Gráfico 25 - Sinal digital depois da decodificação analógico/digital – digital/analógico

As Figuras 1 e 2 mostram, respectivamente, um sinal analógico qualquer e como ele ficaria após ser convertido para digital e para analógico novamente.

O formato do gráfico original é consideravelmente mais arredondado nas pontas que o sinal decodificado.

É possível concluir que quanto maior a taxa de amostragem, maior será a qualidade do sinal digital convertido. Analogamente, quanto maior a taxa de amostragem, o sinal convertido de digital para analógico novamente será um sinal mais próximo do verdadeiro.

Portanto, para generalizar uma taxa de amostragem, a fim de garantir uma qualidade mínima para o sinal decodificado, é comum a utilização da frequência de Nyquist.

ANEXO B

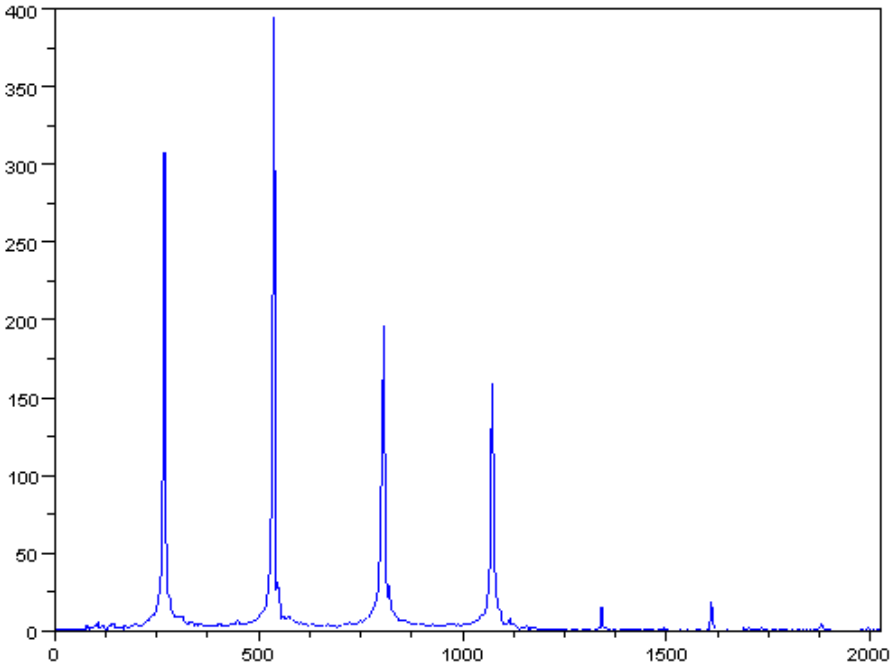

Gráfico 26 - Gráfico da Nota DO, obtido através da FFT

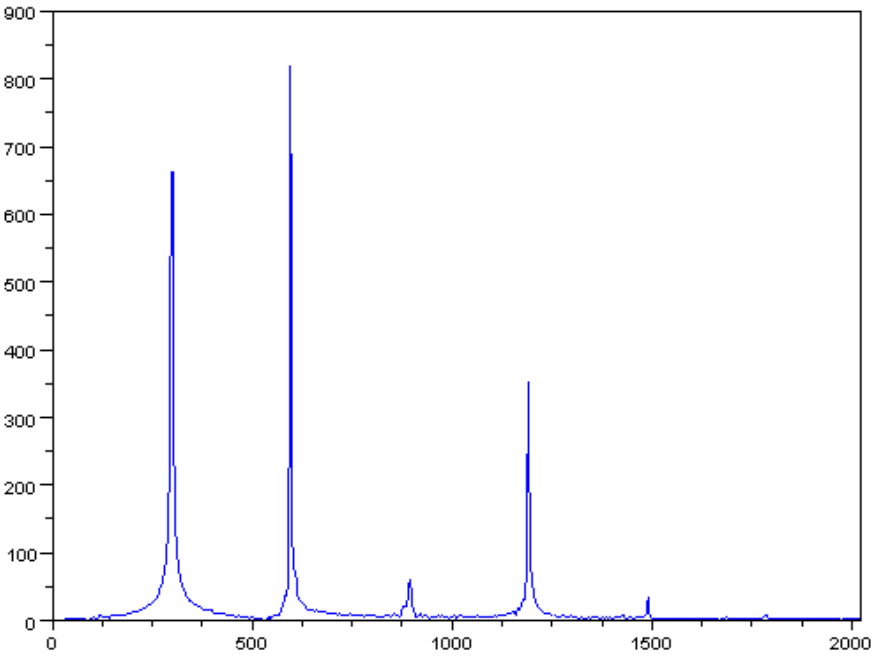

Gráfico 27 - Gráfico da Nota RE, obtido através da FFT

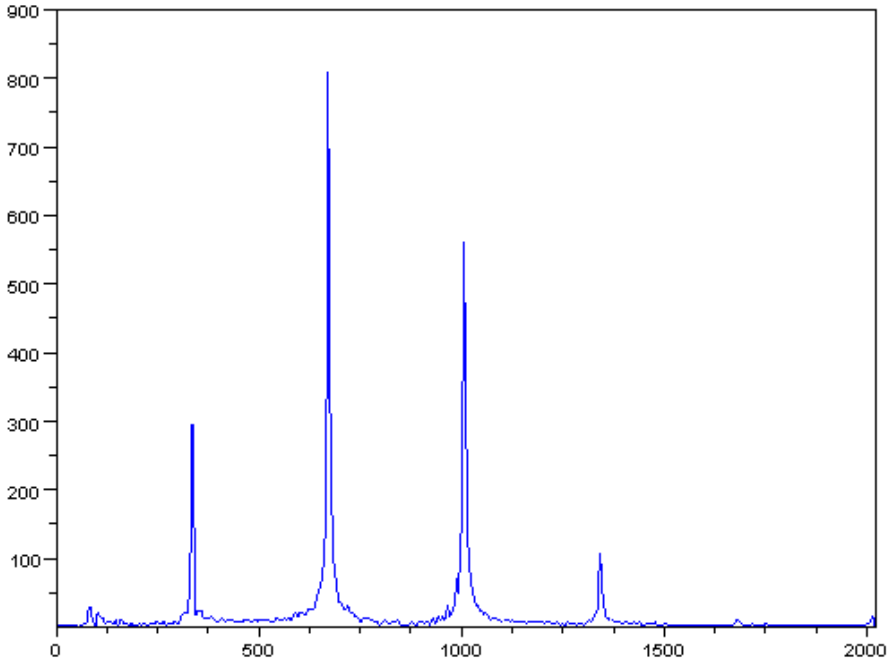

Gráfico 28 - Gráfico da Nota MI, obtido através da FFT

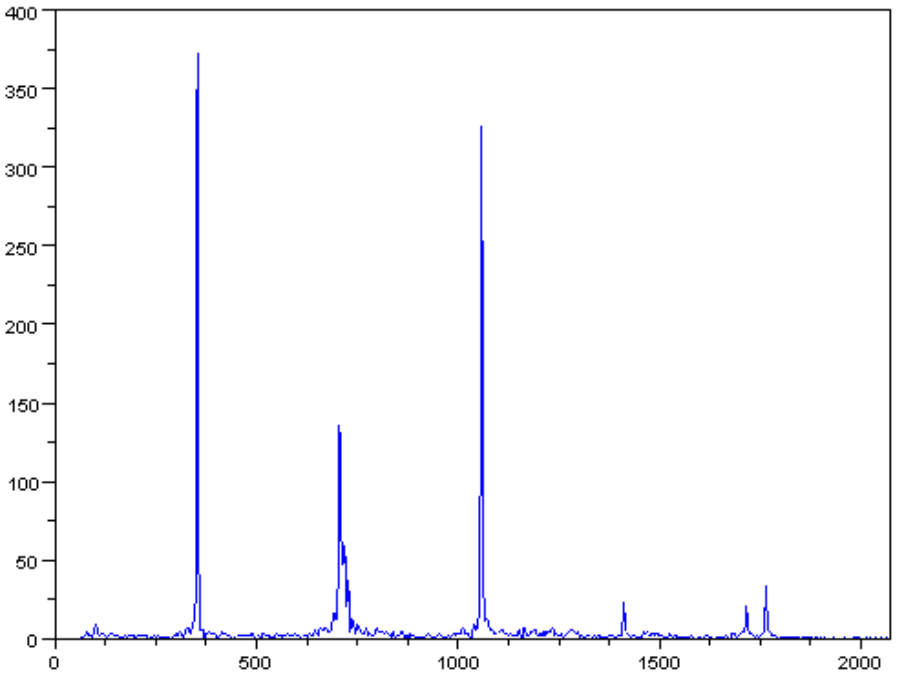

Gráfico 29 - Gráfico da Nota FA, obtido através da FFT

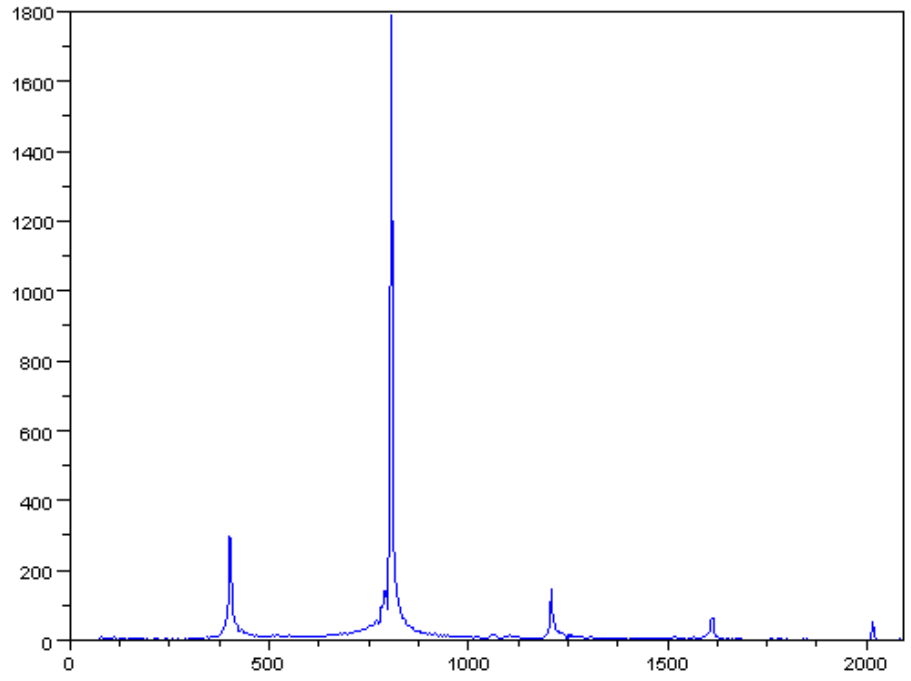

Gráfico 30 - Gráfico da Nota SOL, obtido através da FFT

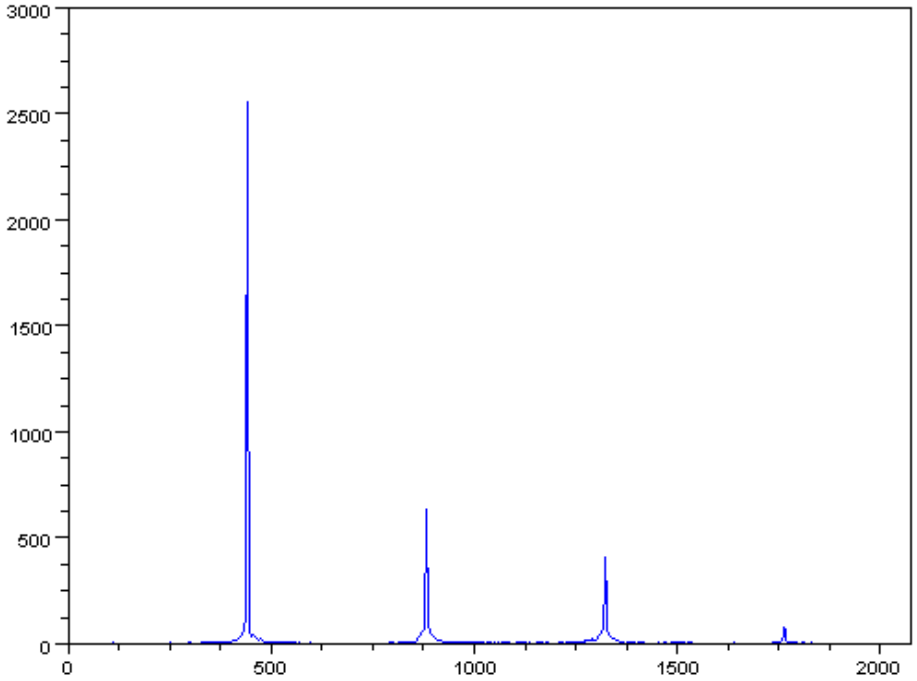

Gráfico 31 - Gráfico da Nota LA, obtido através da FFT

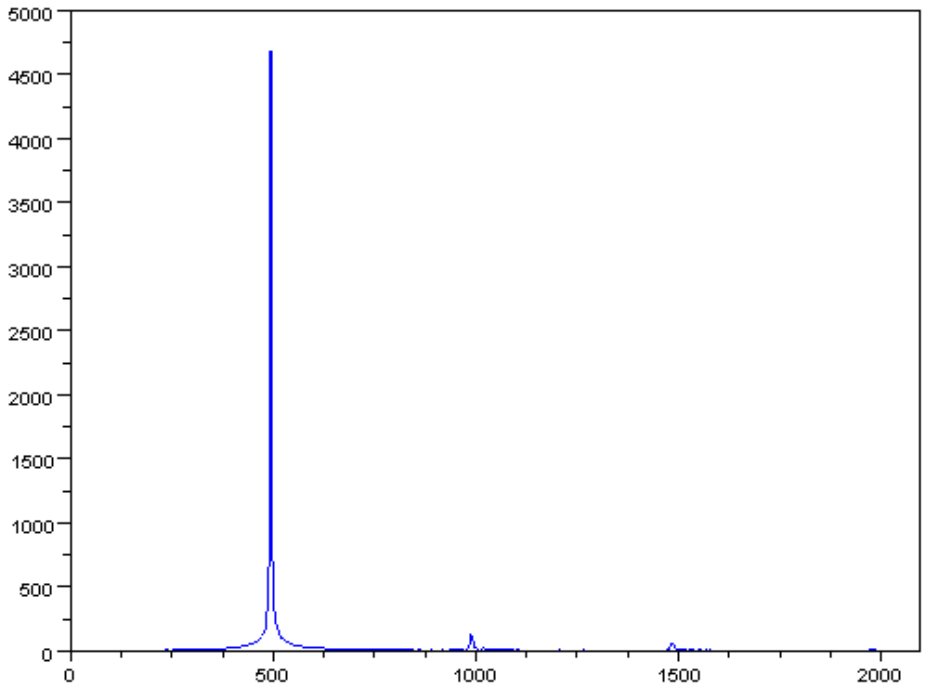

Gráfico 32 - Gráfico da Nota SI, obtido através da FFT