

Serviço de Pós-Graduação EESC/USP

EXEMPLAR REVISADO

Data de entrada no Serviço..... 14 / 12 / 04

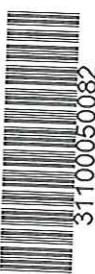
Ass.: 

Daniel Espanhol Razera

DETERMINADORES DE PITCH

Dissertação apresentada à Escola de Engenharia de São Carlos, da universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica.

DEDALUS - Acervo - EESC



Orientador: Prof. Dr. José Carlos Pereira

São Carlos
2004



FOLHA DE JULGAMENTO

Candidato: Engenheiro **DANIEL ESPANHOL RAZERA**

Dissertação defendida e julgada em 05-05-2004 perante a Comissão Julgadora:

Prof. Tit. **JOSÉ CARLOS PEREIRA** (Orientador)
(Escola de Engenharia de São Carlos/USP)

APROVADO

Prof. Dr. **EDUARDO TAVARES COSTA**
(Universidade Estadual de Campinas/UNICAMP)

APROVADO

Prof. Dr. **CARLOS DIAS MACIEL**
(Escola de Engenharia de São Carlos/USP)

Aprovado

Prof. Assoc. **MURILO ARAUJO ROMERO**
Coordenador do Programa de Pós-Graduação
em Engenharia Elétrica

Profa. Titular **MARIA DO CARMO CALIJURI**
Presidente da Comissão de Pós-Graduação

A Ilce e a Fernanda pelo amor e
pacienta para suportar a auséncia.

AGRADECIMENTOS

Ao Prof. Dr. José Carlos Pereira pela orientação, compreensão e apoio acadêmico.

À Escola de Engenharia de São Carlos – Universidade de São Paulo pela formação e apoio institucional.

Aos meus pais, José Espanhol Razera e Maria Olívia Petreca Razera, pelo apoio e incentivo incondicional.

Aos meus irmãos, Luís Carlos e Célia, e cunhados, Rita e Roni, pelo incentivo durante as visitas.

Ao Hélio e Eliane pela amizade e incentivo.

Ao Marcelo, Parê, Mori e Granato pelo companheirismo, paciência, discussões calorosas e ajuda para superar os problemas decorrente do dia-a-dia durante o decorrer do trabalho.

Aos colegas do SEL, dos quais não nomearei a todos, para não cometer o erro de esquecer de alguém.

Aos técnicos do departamento pela agilidade e interesse no desempenho de suas funções.

Aos funcionários da secretaria de graduação e pós graduação, pelo pronto auxílio relacionado ao programa de pós -graduação do instituto, em questões burocráticas, e amizade demonstrada.

Às bibliotecárias e funcionários da EESC, IQSC, ICMC e IFSC pelo auxílio bibliográfico.

À CAPES, pela bolsa institucional.

A Ilce e Fernanda para quem além de dedicar este trabalho também agradeço pela paciência, incentivo e amor.

E a todos que tornaram possível a realização deste trabalho.

RESUMO

Os parâmetros acústicos da voz abordados em diversas pesquisas de análise digital da voz, apresentam-se válidos para o uso em processo diagnóstico e terapêutico. O grupo de parâmetros de perturbação da voz necessita do conhecimento de todos os períodos do trecho de sinal de voz analisado, para ter seu valor calculado. Esta tarefa é desempenhada pelos determinadores de pitch, e a sua precisão determina a confiabilidade que se pode ter nos parâmetros calculados. Este trabalho visa estudar diversos métodos propostos ao longo dos anos e estabelecer qual destes tem a melhor precisão e robustez, quando utilizados com vozes patológicas. Estuda-se também algoritmos estimadores de pitch como uma ferramenta de auxílio para a correção e ajuste dos determinadores. Os resultados obtidos demonstram a necessidade de modificações externas e internas aos algoritmos determinadores e estimadores, para alcançarem a robustez e precisão desejada. Dois algoritmos determinadores, determinador por autocorrelação e por extração de harmônicas, mostraram-se dentro das metas estabelecidas e confirmam-se como os mais promissores em aplicações para obtenção de parâmetros acústicos da voz.

Palavras-chaves: Determinadores de pitch, voz, laringe, processamento digital da fala, engenharia biomédica.

ABSTRACT

Several researches of digital speech processing validate the use of acoustic parameters of the voice in diagnosis and therapeutic processes. Perturbation parameters need the knowledge of all the periods of the analyzed voice signal, to have their values calculated. This task is carried out by the pitch trackers and their precision determines the reliability off the evaluated parameters. The purpose of this work is to study several methods proposed along the years and to establish which algorithm has the best precision and robustness, when used with pathological voices. The pitch estimation is also studied as an aid tool for the correction and adjustment of the pitch trackers. The results demonstrate the need of external and internal modifications of the trackers and detector algorithms to reach the wanted robustness and precision. The algorithms for autocorrelaçao and for extraction of harmonics are confirmed as the most promising in applications for obtaining of acoustic parameters of the voice.

Keywords: pitch trackers, voice, larynx, digital speech processing, biomedical engineering

LISTA DE FIGURAS

Figura 1: Aparato vocal.	4
Figura 2: Pulmão e traquéia.	5
Figura 3: Vista frontal e em corte da laringe.	6
Figura 4: Movimento das pregas vocais.	9
Figura 5: Representação de um sinal de voz sonoro.	10
Figura 6: Representação de um sinal de voz não sonoro /s/.	11
Figura 7: Representação de um modelo completo do trato vocal.	12
Figura 8: Algoritmo de decomposição wavelet em árvore.	19
Figura 9: Tela do programa para síntese de sinais de voz.	25
Figura 10: Diagrama de blocos geral para os determinadores.	25
Figura 11A: Sinal de voz.	30
Figura 11B: Sinal filtrado e as marcas de períodos determinados	30
Figura 12: Determinação de pitch através do algoritmo de autocorrelação	31
Figura 13A: Sinal de voz	33
Figura 13B: Representação do sinal com as excursões mais significativas.	33
Figura 13C: Excursões de mesma polaridade que a maior.	33
Figura 13D: Resultado final com os pulsos indicando os períodos	33
Figura 14A: Representação do sinal por excursões	34
Figura 14B: Resultado com os pulsos indicando os períodos	34
Figura 15A: Função escala	36
Figura 15B: Função Wavelet	36
Figura 15C: Função filtro derivativa	36
Figura 16A: Sinal de voz	37
Figura 16B: Sinal filtrado pela função filtro derivativa	37
Figura 17: Interface gráfica do determinador semi-automático de pitch	38
Figura 18: Tela com mensagem de erro decorrente de uma marcação errada	39
Figura 19A: Sinal de voz da vogal \a\ sustentada.	41
Figura 19B: Autocorrelação da janela	41
Figura 20A: Sinal de voz analisado	42

Figura 20B: O sinal filtrado pela janela de Hann	42
Figura 20C: O Cepstrum real do sinal	42
Figura 20D: Janela de busca	42
Figura 21A: Sinal de voz	44
Figura 21B: Espectro do sinal de voz. .	44
Figura 21C: Filtragem pelo pente	44
Figura 21D: Soma dos valores da filtragem para cada valor de freqüência do pente, identificado o máximo da função	44
Figura 22A: Sinal de voz analisado	46
Figura 22B: AMDF do sinal , com os primeiros mínimos marcados.	46
Figura 23A: Sinal de voz analisado	47
Figura 23B: DWT para uma escala de 2^3 , com dois máximos locais	47
Figura 23C: DWT para uma escala de 2^4 , com 9 máximos locais	47
Figura 23D: DWT para uma escala de 2^5 , com os mesmos 9 máximos locais da escala anterior	47
Figura 24: Resultados dos estimadores para variação de Fo nas vozes sintetizadas	49
Figura 25: Resultados dos estimadores para variação de SNR	50
Figura 26: Resultados dos estimadores para variação do jitter.	51
Figura 27: Resultados dos estimadores para variação do Shimmer.	51
Figura 28: Resultados dos estimadores para vozes normais	53
Figura 29: Resultados dos estimadores para vozes patológicas masculinas	53
Figura 30: Resultados dos estimadores para vozes patológicas femininas.	54
Figura 31: Resultados dos determinadores para variação de Fo nas vozes sintetizadas	55
Figura 32: Resultados dos estimadores para variação de SNR	56
Figura 33: Resultados dos estimadores para variação do jitter.	56
Figura 34: Resultados dos estimadores para variação do Shimmer.	57
Figura 35: Resultados dos determinadores para vozes normais.	58
Figura 36: Resultados dos determinadores para vozes patológicas masculinas.	59
Figura 37: Resultados dos determinadores para vozes patológicas femininas.	60
Figura 38: Seqüência de períodos de uma voz patológica.	60
Figura 39: Seqüência de períodos de uma voz normal.	61
Figura 40: Pontos delimitadores para marcação dos períodos.	62
Figura 41: Períodos para diferentes delimitadores.	62

LISTA DE TABELAS

Tabela 1: Resultados dos estimadores para variação de Fo nas vozes sintetizadas.	69
Tabela 2: Resultados dos estimadores para variação de SNR.	69
Tabela 3: Resultados dos estimadores para variação do jitter.	70
Tabela 4: Resultados dos estimadores para variação do shimmer.	70
Tabela 5: Resultados dos estimadores para vozes normais	71
Tabela 6: Valores encontrados pelos estimadores para vozes masculina patológicas	71
Tabela 7: Valores encontrados pelos estimadores para vozes femininas patológicas	73
Tabela 8: Resultados dos determinadores para variação de Fo nas vozes sintetizadas	76
Tabela 9: Resultados dos determinadores para variação de SNR	76
Tabela 10: Resultados dos determinadores para variação do jitter	77
Tabela 11: Resultados dos determinadores para variação do shimmer	77
Tabela 12: Resultados dos determinadores para vozes normais	78
Tabela 13: Valores calculados pelos determinadores para vozes masculina patológicas	79
Tabela 14: Valores encontrados pelos determinadores para vozes femininas patológicas	81
Tabela 15: Valor dos períodos de um sinal de voz patológica	83

LISTA DE ABREVIATURAS

DWT - Transformada Wavelet Discreta

AMDF - Função Diferença de Amplitude Média

SIFT - Estimador por Filtragem Inversa Simples.

FFT - Transformada rápida de Fourier

Inf - Valor indeterminado

LISTA DE SÍMBOLOS

/a/ - Vogal a

/e/ - Vogal e

/i/ - Vogal i

F₀ - Freqüênci fundamental

Hz - Hertz

SUMÁRIO

RESUMO	I
ABSTRACT	II
LISTA DE FIGURAS	III
LISTA DE TABELAS	V
LISTA DE ABREVIATURAS	VI
LISTA DE SÍMBOLOS	VII
1 – Introdução	1
2 – Resumo Teórico	4
2.1 - Aparato vocal e formação da voz	4
2.1.1 - Pulmões e traquéia	5
2.1.2 – Laringe	6
2.1.3 - Cavidade supraglotal	8
2.2 – Vocalização	8
2.3 - O sinal de voz	9
2.4 - Modelo do trato vocal	11
2.5 - Processamento de sinais	13
2.5.1 - Transformada-z	13
2.5.2 - Transformada de Fourier	14
2.5.3 - Filtros Digitais	15
2.5.4 – Wavelet	16
2.5.5 – Autocorrelação	19
2.5.6 – Cepstrum	20
3 – Métodos e Procedimentos	23

3.1 – Sinais de voz	23
3.2 – Determinadores de pitch	25
3.2.1 - Determinador de Pitch por extração da harmonica fundamental	26
3.2.2 - Determinador de pitch por autocorrelação	30
3.2.3 - Determinador de Pitch através da análise da estrutura temporal	31
3.3.4 - Determinador de Pitch por Wavelets	34
3.3.5 - Determinador de Pitch semi-automático	37
3.3 – Algoritmos estimadores	40
3.3.1 - Estimador por Autocorrelação	40
3.3.2 - Estimador por Cepstrum	41
3.3.3 - Estimador por casamento de harmônicas (harmonics match)	43
3.3.4 - Estimador por Filtragem inversa simples	45
3.3.5 - Estimador por AMDF (Average Magnitude Diference Function)	46
3.3.6 - Estimador por Transformada Wavelet	47
4 – Resultados	49
4.1 - Teste com os Estimadores	49
4.2 - Analise dos determinadores	54
5 – Conclusão	63
6 – Bibliografia	65
APÊNDICE A	68

1 - Introdução

O estudo de diagnósticos de patologias da voz, através do processamento do sinal de voz (ROSA, 1998) caracteriza-se por ser não invasivo, já que analisa o sinal de voz de um indivíduo encontrando suas características determinantes. Trata-se de uma ferramenta excepcional para especialistas da área médica correlata (otorrinolaringologistas e fonoaudiólogos), já que possibilita desde a identificação de patologias até acompanhamento de um quadro de recuperação pós-cirúrgica. A determinação das características da voz é realizada pela extração de parâmetros desta e uma posterior análise destes parâmetros, determinando a normalidade ou não destes.

Um parâmetro isolado pode diferenciar com certa precisão vozes normais de patológicas (DAVIS, 1979), mas devido à complexidade do aparato vocal, torna-se difícil determinar o tipo de patologia. Diversos parâmetros podem ser extraídos da voz e sua combinação pode resultar num diagnóstico mais preciso.

Alguns parâmetros contêm informações redundantes quando comparados entre si, o que indica que estes podem ser combinados em grupos conforme o tipo de informação que podem fornecer ou da técnica de análise utilizada. O uso de um determinado conjunto de parâmetros pode melhorar a probabilidade de acertos nos diagnósticos, mas a falta deste conjunto pode incorrer em uma análise incompleta e reduzir a probabilidade de acertos. Assim, uma ampla gama de parâmetros deve ser estudada quando o objetivo final for a análise de vozes patológicas.

Para os parâmetros de perturbação da voz, o valor de cada período presente na amostra de voz é essencial para seu cálculo, conforme Andrade et. all.(2002). Esta tarefa torna-se difícil devido à característica da voz que é um sinal quase periódico, ou seja, durante a fonação o período sofre pequenas variações período a período, mesmo para vozes normais. Em vozes patológicas estas variações aumentam (LIBERMAN, 1963), dificultando ainda mais o trabalho de identificação dos períodos individualmente.

O objetivo deste trabalho é determinar através de testes objetivos qual algoritmo ou conjunto de algoritmos que possa ser utilizado para análise de vozes patológicas,

garantindo a robustez necessária para a determinação dos parâmetros da voz, que necessitem, para seu cálculo, do valor dos períodos presentes na amostra de voz.

O estudo dos algoritmos para determinação de pitch é uma pesquisa básica dentro da área de processamento de sinais de voz, e parte fundamental de muitas pesquisas ou aplicações comerciais. Estes algoritmos podem ser divididos em estimadores de valor médio para a freqüência fundamental, e os determinadores que calculam o valor de cada período de um sinal de voz sustentada. Vários artigos encontrados na literatura discorrem sobre o assunto, comparando algoritmos já testados ou apresentando um novo algoritmo. Anderson (1986) afirma que embora essas pesquisas venham explorando o assunto por vários anos, os melhores algoritmos apresentados atualmente não são significativamente melhores que os algoritmos de cinco ou 10 anos atrás, sendo necessário estatisticamente muitos algoritmos com novas técnicas para que algum destes obtenha melhores resultados que algoritmos já existentes.

A variabilidade de um sistema natural com todo o aparato vocal humano torna a tarefa de determinação dos períodos de uma voz uma tarefa extremamente difícil e complicada para uma abordagem lógica computacional. O uso de técnicas de processamento de sinais facilita a tarefa, mas não a torna mais simples.

Este trabalho aborda o mesmo tema com ênfase para uso posterior em aplicações comerciais, tendo-se, portanto um especial cuidado para a robustez do método, testando diversos algoritmos e recolhendo as melhores características, simplicidade e robustez, de cada um. A necessidade de confiabilidade no algoritmo é devido ao seu uso na determinação de pitch para vozes patológicas.

O termo pitch apresenta diferentes definições em diferentes áreas de estudo da voz, sendo indicado algumas vezes como freqüência fundamental, período de tom, ou mesmo qualidade vocal. A nomenclatura usada neste trabalho indica o termo pitch como sendo o valor do período médio do sinal de voz analisado, sendo o valor inverso deste igual a freqüência fundamental.

Hess (1983), em seu livro, apresenta um extenso trabalho a respeito dos estimadores e determinadores de pitch que utilizam técnicas clássicas, não abordando o uso de técnicas mais atuais como wavelet. O uso de wavelet em estimadores de pitch apresenta uma série de artigos atualmente, sendo seu uso em detrimento de técnicas mais clássicas como cepstrum ou autocorrelação, bastante explorado. Assim, é de grande importância o teste de algoritmos baseados na Transformada de Wavelet.

Nem sempre o algoritmo determinador de pitch é utilizado diretamente no sinal de voz. Um pré-processamento como uma filtragem pode ser necessário, bem como um pós-processamento para correção de possíveis erros de busca do início de cada período. Em certos algoritmos, este pós-processamento se acopla ao código, tornando-se parte do mesmo. O pré-processamento pode ser apenas uma filtragem ou mesmo uma redução da taxa de dados, conforme o algoritmo determinador trabalhar melhor. Para o pós-processamento deve-se determinar quão certos estão os valores identificados e corrigi-los se necessário. Esta tarefa torna-se difícil a medida que os erros relacionados com os valores dos períodos tornam-se pequenos, inviabilizando o uso de corretores de erros, ocorrendo a possibilidade de aumentar o valor dos mesmos. Entende-se por erro a diferença entre um valor encontrado pelo algoritmo e um valor adotado como padrão.

Neste trabalho, inicialmente faz-se uma abordagem da fisiologia do aparato vocal, detalhando seu funcionamento na produção de sons vocalizados. Uma breve explicação do sinal de voz e suas características importantes são abordadas e em seguida um modelo completo do aparato vocal é detalhado. A seguir abordam-se as técnicas de processamento digital de sinais utilizadas neste trabalho.

No capítulo seguinte detalham-se os métodos ou algoritmos determinadores e estimadores de pitch testados neste trabalho. Procura-se cobrir o assunto expondo as principais características de cada algoritmo e apresentar os sinais em cada etapa de processamento. Relatam-se a seguir as análises, realizadas com vozes, para estes algoritmos.

A seguir apresentam-se os resultados obtidos para cada algoritmo, objetivando a obtenção do melhor algoritmo determinador. Tais resultados são apresentados graficamente dentro das possibilidades dos dados processados durante os testes. A resposta visual facilita a compreensão e conclusão dos resultados de uma forma mais imediata.

Finalmente conclui-se sobre o trabalho em função dos resultados obtidos no capítulo anterior.

2 - Resumo Teórico

2.1 - Aparato vocal e formação da voz

Os órgãos ou estruturas responsáveis pela formação da voz, e denominados de aparato vocal, são os pulmões, traquéia, laringe e cavidade supraglotal (cavidades nasal e oral, língua, dentes e lábios) (figura 1). Cada órgão apresenta função distinta na produção da voz, embora trabalhem como uma unidade e qualquer variação em um destes órgãos influencia o sinal de voz gerado. É possível fazer uma subdivisão em três grupos do aparato vocal, baseado em suas funções na produção da voz e com a finalidade de facilitar a compreensão fisiológica do aparato vocal.

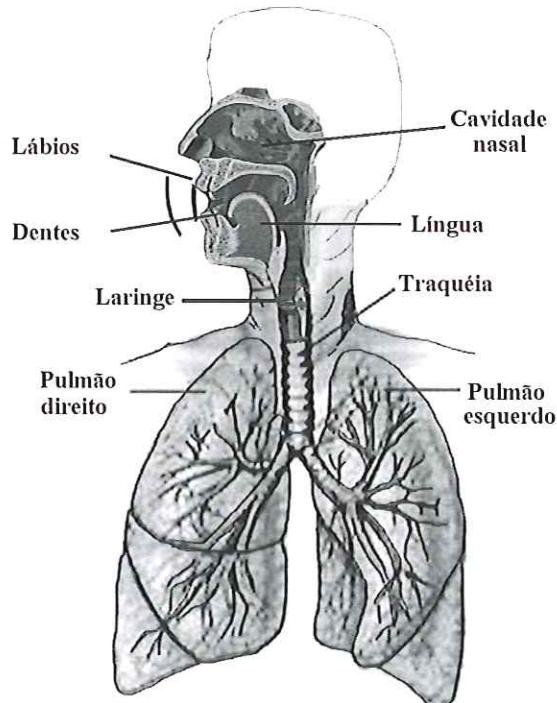


Figura 1: Aparato vocal.(Figura composta à partir das fontes). Fontes: O Corpo Humano-Guia Prático de Anatomia.Editora nova Cultural 1986, p. 97;
<http://www.corpohumano.hpg.ig.com.br/respiracao/respiracao2.html>

O primeiro grupo consta dos pulmões e traquéia e sua função é o escoamento de ar para a laringe. O segundo grupo consta apenas da laringe sendo esta a principal estrutura do sistema fonador, já que nesta se encontram as pregas vocais. O terceiro grupo consta da cavidade supraglotal que constitui um sistema de ressonância.

Cada grupo será detalhado a seguir neste capítulo.

2.1.1 - Pulmões e traquéia

Os pulmões são dois órgãos de estrutura esponjosa e têm forma de pirâmide com a base apoiando-se sobre o diafragma. Cada pulmão se compõe de lóbulos, o esquerdo consta de três lóbulos e o direito de dois. Tais lóbulos por sua vez, contêm os alvéolos, que são dilatações terminais dos brônquios. Por fora dos alvéolos há redes de capilares sanguíneos para a troca gasosa. A fixação dos pulmões na caixa torácica se dá através das pleuras que são membranas que recobrem os pulmões.

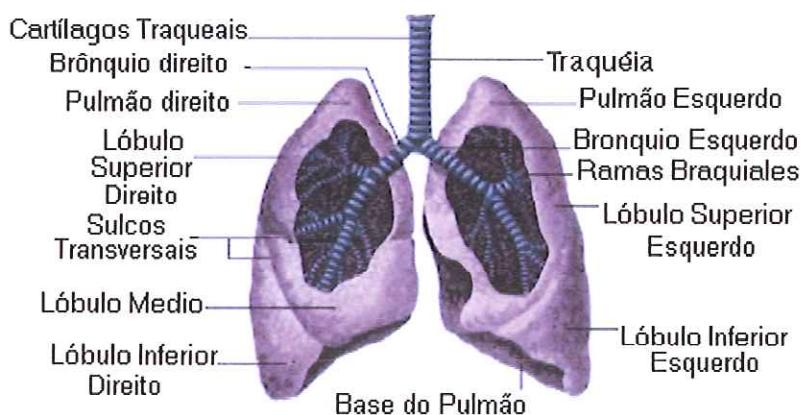


Figura 2: Pulmão e traquéia.

Fonte:<http://www.corpohumano.hpg.ig.com.br/respiracao/pulmao/pulmoes.html>

A função principal do pulmão é a hematose, na qual tanto o oxigênio como o dióxido de carbono atravessam a barreira sangue-ar, em forma passiva, por diferenças de concentração (difusão) entre as duas fases. Também participa na regulação da temperatura corporal e da produção da voz.

Esta última função apresentada é de interesse deste estudo, sendo efetuada na forma de geração de energia sob a forma de um escoamento do ar com determinada vazão. A vazão do ar é definida pela diferença de pressão do ar nos pulmões da pressão atmosférica fora do corpo humano, bem como pela geometria do trato vocal.

A traquéia é um conduto músculo-membranoso que tem a forma de um cilindro formado por anéis cartilaginosos em número variável de 12 a 16, unidos entre si por tecido fibroso. A traquéia está revestida no seu interior por uma mucosa, cujas células têm a característica de serem ciliadas: graças a esses cílios, finíssimos, ela se opõe à entrada de eventuais partículas estranhas. Estas são expulsas das vias respiratórias pela tosse. Para a geração da voz a traquéia trabalha como um duto transportador de ar para a laringe.

2.1.2 - Laringe

A laringe apresenta três funções: respiração, esfincteriana ou deglutição e fonatória. Agrupadas, a esfincteriana e deglutição evitam que corpos estranhos cheguem ao pulmão, função esta desempenhada principalmente pela epiglote. Outra função secundária e do interesse deste trabalho é a de emitir a voz. O uso de um sistema valvular para vocalização exigiu um desenvolvimento de intrincados controles neurais que permitem usar as pregas vocais para uma vocalização precisa, permitindo o uso na comunicação vocal entre os humanos. A qualidade vocal depende essencialmente da forma da própria laringe e pode variar na dependência das diversificações que afetam este órgão.

A laringe se encontra no pescoço acima da traquéia, e é um órgão constituído por cartilagens ligadas entre si por ligamentos e lâminas de aponeurose (semelhante a tendões e ligas que revestem órgãos internos), músculos, nervos e o osso hióide. A figura 3 mostra a representação de uma laringe em uma vista frontal e em corte.

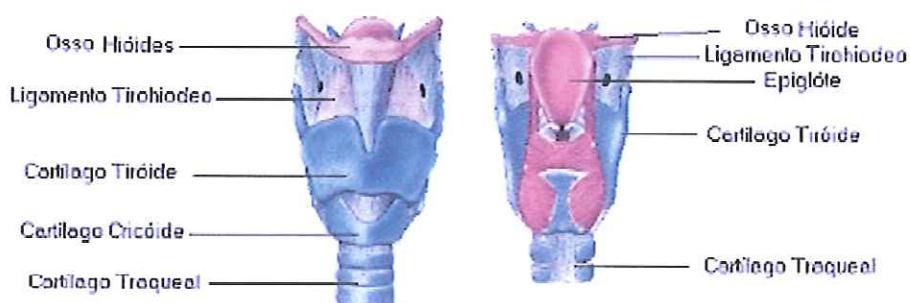


Figura 3: Vista frontal e em corte da laringe (Figura modificada da fonte). Fonte: <http://www.corpohumano.hpg.ig.com.br/respiracao/laringe/laringe2.html>

As cartilagens têm funções estruturais, cabendo a estas dar sustentação a laringe correspondendo assim a uma porção fixa, sendo movimentada pelos músculos e ligamentos. Existem cinco cartilagens principais: a cricóidea, a tireóidea, um par de aritenóideas e a epiglote. Além destas tem-se duas outras, as corniculadas, que têm pouca função na fonação, mas em conjunto com as cartilagens principais formam o “esqueleto” do sistema vibratório.

Os músculos dividem-se em extrínsecos e intrínsecos com relação a sua colocação no sistema de vocalização, sendo que os intrínsecos estão ligados somente às estruturas internas da laringe e os extrínsecos se ligam também à estruturas externas.

Os extrínsecos permitem a elevação ou abaixamento da laringe, além de mantê-la fixa e tem função apenas auxiliar na produção da voz.

Os músculos intrínsecos são em número de doze (seis pares) e agem diretamente sobre a fonação. Os cricoaritenóideos posteriores permitem abertura das pregas vocais, enquanto os cricoaritenóideos laterais produzem uma adução das pregas vocais. As fibras aritenóideas transversas auxiliam os cricoaritenóideos laterais, além de comprimir as pregas vocais. Os músculos aritenóideos oblíquos são envolvidos no fechamento glotal.

Os tiroaritenóideos formam o volume muscular da prega vocal, uma vez que esta ainda compreende a superfície interna da cartilagem aritenóidea e o ligamento vocal, sendo recoberto por uma membrana esbranquiçada, que efetivamente vibra com a variação da pressão. A tensão dos músculos tiroaritenóideos é regulável pela nossa vontade, através da transmissão das ordens necessárias para o nervo laríngeo inferior, que por sua vez, faz contrair ou relaxar o músculo. Em consequência, a fenda glótica, isto é, o espaço compreendido entre os bordos das cordas vocais, se alarga ou se restringe segundo o caso. É evidente então que o ar que passa pela glote provoca vibrações de intensidade diversa, a cada uma das quais corresponde uma nota musical ou um som elementar.

Outra estrutura presente é a epiglote que trabalha como uma válvula que, ao voltar-se para trás no ato de deglutir, forma uma tampa para a laringe, de maneira que os alimentos passem do esôfago para o estômago. Mais exatamente, a ação de válvula ocorre de modo que a laringe como um todo se eleva, enquanto a epiglote se abaixa sobre ele.

2.1.3 - Cavidade supraglotal

É representada neste trabalho pela porção superior do trato vocal e acima das pregas vocais, sendo assim, composta pela cavidade nasal e a oral, língua, dentes e lábios.

Esta cavidade muscular de diâmetro variável (continuamente mudando) funcionam como uma cadeia de ressonadores, respondendo, seletivamente, à diversas freqüências contidas nos sons produzidos pela laringe. Assim, se o trato vocal num determinado momento assume uma formação que é compatível a determinadas freqüências, digamos aos de freqüência próximas a 330, 800 e 2200 Hz, por exemplo, pode-se afirmar que estes são os primeiros formantes daquela configuração vocal. (FUKS, 2000).

Modificando-se os formantes do trato vocal, através de alterações em sua forma, pode-se alterar o som básico gerado pela laringe, para uma quantidade enorme e rica de timbres sonoros, mensuráveis e comparáveis.

2.2 - Vocalização

O pulmão, durante a expiração, provoca um escoamento do ar com determinada vazão através da laringe para o exterior do corpo humano pela boca ou narinas. Quando da respiração normal e tranqüila não ocorre vocalização, sendo que todas as estruturas relacionadas a voz permanecem relaxadas. Para a formação da voz, os músculos intrínsecos aproximam as pregas vocais quando a expiração se inicia, aumentando a pressão subglótica(Figura 4-1). A perda de energia das pregas e o aumento da pressão subglótica provocam a abertura das pregas vocais proporcionando a passagem de um fluxo de ar pela abertura glótica(Figura 4-2 ,4-3 e 4-4). A passagem do ar reduz um pouco a pressão subglótica e as pregas adquirem energia elástica ao se abrirem (Figura 4-5 ,4-6 e 4-7). O acúmulo de energia permite às pregas vencerem a pressão do ar, auxiliadas pelo efeito de sucção Bernoulli, e se aproximarem novamente (Figura 4-8 ,4-9 e 4-10). Este processo de abertura e fechamento, ciclo vibratório, repete-se até o fim da vocalização. O fechamento das pregas vocais inicia-se na parte inferior da mesma e se propaga para a parte superior, da mesma maneira que a abertura destas (figura 4).

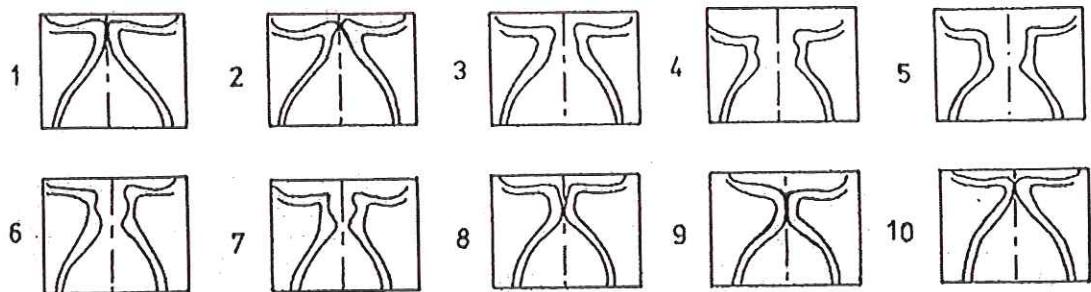


Figura 4: Movimento das pregas vocais. Fonte: HIRANO, 1981

O ciclo vibratório produz um sinal quase periódico que se propaga para a cavidade supraglotal que irá amplificar ou atenuar determinadas freqüências presentes neste sinal, formando assim um som inteligível.

A freqüência fundamental da voz relaciona-se diretamente ao número de ciclos vibratórios que ocorrem por segundo. Como a velocidade de fechamento e afastamento da pregas depende de seu comprimento e espessura, a freqüência fundamental também irá depender.

2.3 - O sinal de voz

O som é uma onda mecânica que se propaga em um meio (no caso o ar) e se movimenta em direção longitudinal. Uma forma inteligível de som é a voz humana que carrega informações que podem ser interpretadas por um ouvinte. Pode-se gerar a voz por três formas:

- Sons vocálicos:

São produzidos por pulsos de ar quase periódicos gerados na laringe pelo movimento cíclico das pregas vocais que excitam o trato vocal. São identificados principalmente pela emissão de vogais

- Sons fricativos:

Estes sons são produto da passagem turbulenta do ar através de alguma constrição formada no trato vocal (/r/ e /s/).

- Sons Plosivos:

São sons produzidos pelo fechamento completo do trato vocal, com um aumento da pressão anterior à obstrução e liberação abrupta desta (/d/, /t/, /p/, /b/).

Dos três possíveis modos de geração citados, os sons vocálicos são os que apresentam vibração das pregas vocais e ressonância do trato vocal completo. Portanto, para uma análise de todo o aparato vocal, este modo de geração de voz contém toda a informação fisiológica do sistema em questão.

Para sons sonoros existe um período de repetição chamado de período de tom ou *pitch*, sendo o inverso destes, denominada de freqüência fundamental. A figura 5 mostra um sinal de voz da vogal \a\ sustentada com três períodos completos distintos. Os sons não sonoros produzem um sinal sem periodicidade definida, conforme visualizado na figura 6.

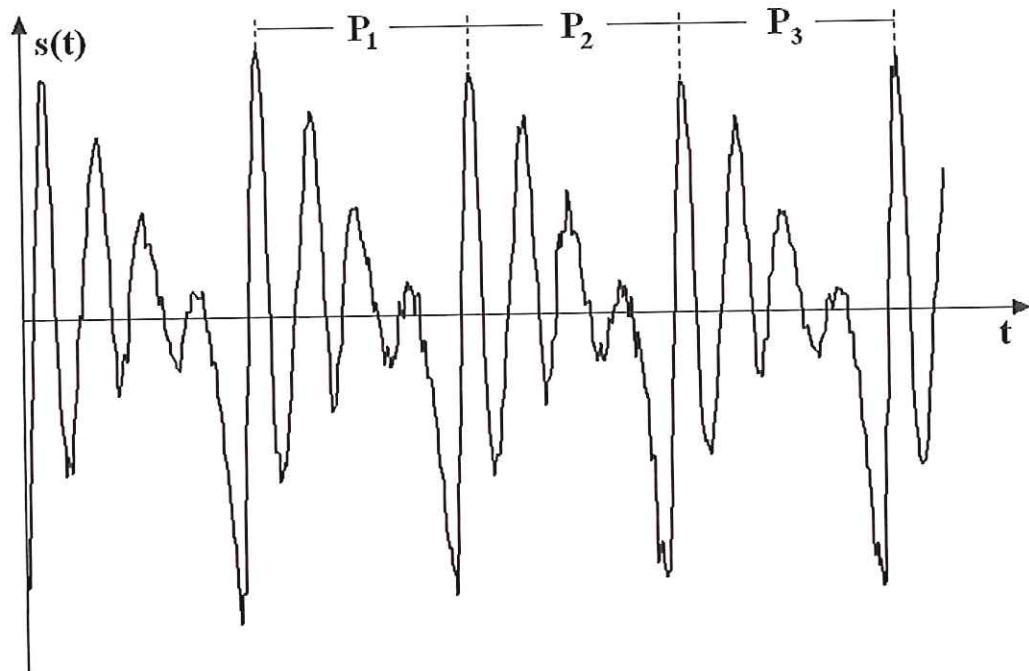


Figura 5: Representação de um sinal de voz sonoro

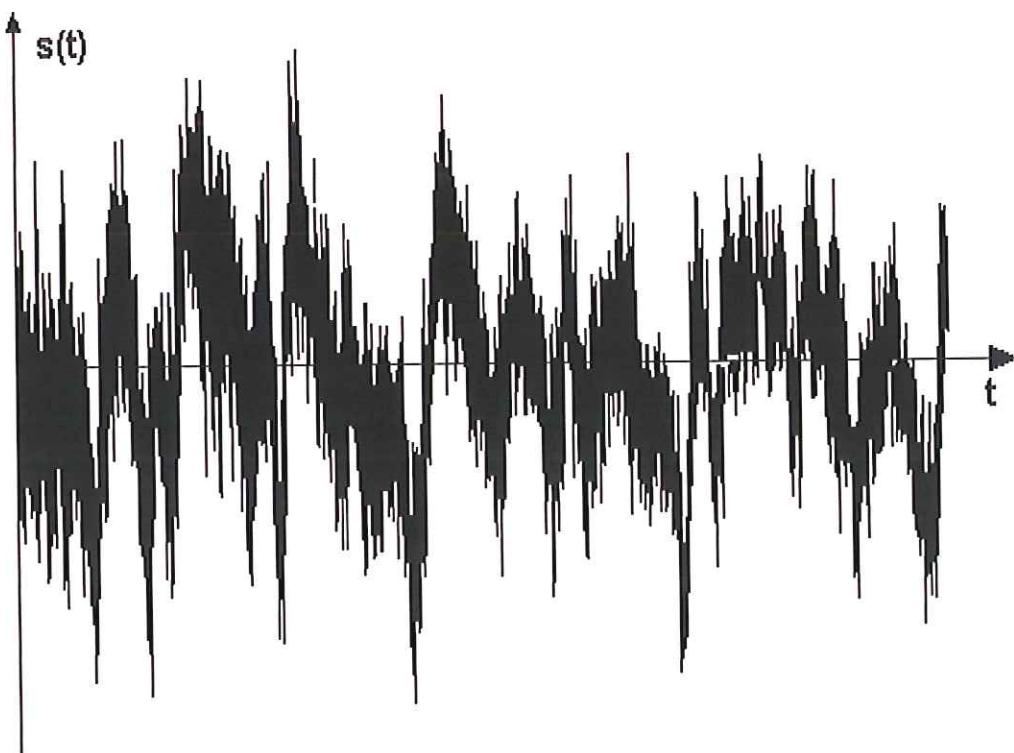


Figura 6: Representação de um sinal de voz não sonoro /s/

2.4 - Modelo do trato vocal

É extremamente útil conhecer importantes características físicas do sistema fonador, com a finalidade de obter um modelo matemático representativo deste sistema. O modelo de produção mais usual é o sistema linear variante no tempo e é baseado na premissa que o sinal de fala contém informações redundantes (MONTAGNOLI, 1998).

O modelo inclui todas as estruturas responsáveis pela produção da fala, citadas em detalhe anteriormente, podendo ser dividido funcionalmente em três estruturas distintas, apesar da inter-relação que ocorre fisicamente no ser humano. Assim tem-se um elemento produtor de excitação, caracterizado biologicamente pelo pulmão e suas estruturas de transporte de ar, um sistema produtor de oscilações periódicas, definido organicamente pelas pregas vocais e órgãos anexos, e um sistema produtor de ressonância, compreendendo os órgãos da cavidade supraglotal. Um modelo mais completo (ROSA, 1998), considerando também os efeitos de perda de energia é mostrado na figura 7.

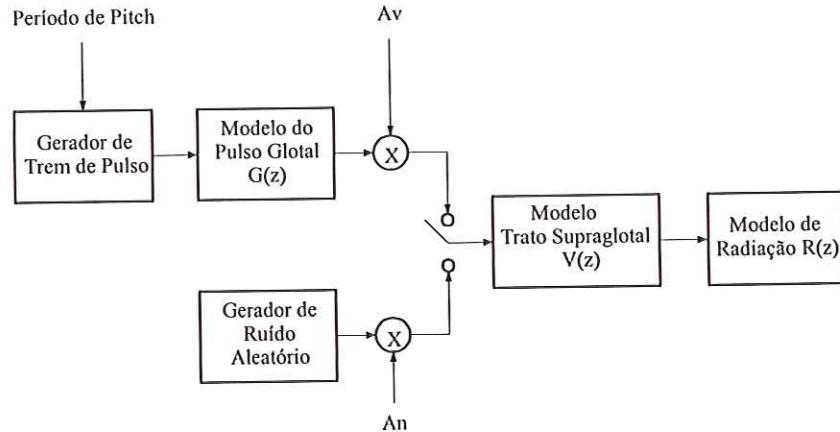


Figura 7: Representação de um modelo completo do trato vocal.

Os valores de ganho, A_v e A_n , levam em consideração os efeitos de maior ou menor pressão exercida pelo pulmão na fonação. O gerador de ruído aleatório representa a emissão de sinal não-vocálico. O modelo do pulso glotal é necessário para que se produza um sinal similar ao produzido pela vibração das cordas vocais de modo a excitar adequadamente o trato supraglotal artificial e produzir o sinal de voz desejado. Com todos os elementos agora definidos, todos os sistemas podem ser cascataeados, dando origem à uma única função de transferência que rege este modelo simplificado de trato vocal.

$$H(z) = G(z) \cdot V(z) \cdot R(z) \quad (1)$$

onde $G(z)$ representa o modelo do pulso glotal, responsável pela produção da onda quase periódica para sinais vocálicos e ruído branco para sinais não-vocálicos, $V(z)$ representa a estrutura moduladora do trato supraglotal, num modelo tudo-polo, e $R(z)$ produzindo a influência das perdas acústicas ao longo do trato vocal e principalmente devido à difração da onda sonora nos lábios.

O modelo apresenta certas limitações de uso, mas é adequado para sinais de voz que pouco variam durante o tempo, ou seja, vocálicos e não-vocálicos. Esta característica torna tal modelo útil para este trabalho principalmente para a filtragem inversa que irá tentar obter do sinal de voz o pulso glotal através da aplicação de um filtro equivalente ao inverso de $V(z)R(z)$.

2.5 - Processamento de sinais

Todo processamento de sinais de voz realizado neste trabalho foi digital. Assim, este tópico aborda o tema de processamento de sinais digitais, com ênfase para as técnicas que foram utilizadas no pré-processamento dos sinais de voz, bem como, nos algoritmos que serão posteriormente abordados.

O processamento digital de sinais envolve a transformação do sinal de uma forma de representação para outra mais conveniente aos propósitos desejados. Esta transformação geralmente possui uma seqüência entrada, $x(n)$, e uma saída, $y(n)=T[x(n)]$, sendo $T[]$ a transformada aplicada ao sinal de entrada. No processamento digital de sinais de voz, utiliza-se uma classe de transformadas cujo sistema é linear e invariante ao deslocamento, sendo este sistema completamente caracterizado por sua resposta ao pulso unitário.

2.5.1 - Transformada-z

O processamento, análise e projeto de sistemas lineares é imensamente facilitado pela sua representação em outros domínios como o da freqüência ou o plano complexo. A transformada-z é muito utilizada para a determinação da função do sistema, $H(z)$, e verificação de estabilidade e causalidade de sistemas, ou seja, a análise dos pólos e zeros da função do sistema.

A transformada-z de uma seqüência, $x(n)$, é definida como:

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \quad (2)$$

Em geral a transformada-z, $X(z)$, é uma série infinita de potências na variável z^{-1} , com a seqüência $x(n)$ sendo os coeficientes para a série de potências, e assim, a transformada só existe para valores de z para os quais a série converge. Deste modo, define-se a região de convergência da transformada-z como:

$$|X(z)| = \left| \sum_{n=-\infty}^{\infty} x(n)z^{-n} \right| < \infty \quad (3)$$

2.5.2 - Transformada de Fourier

Uma grande quantidade de ferramentas pode ser usada em análise de sinais. A mais utilizada e conhecida é a Transformada de Fourier, que quebra o sinal em suas componentes cosenóides para diferentes freqüências, ou seja, obtém-se o espectro do sinal.

É derivada da série de Fourier, para sinais periódicos, com a diferença de poder ser usada em sinais não periódicos. É definida como:

$$X(w) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t} dt \quad (4)$$

sendo $x(t)$ a seqüência representando um sinal discreto no tempo.

A transformada de Fourier é um produto interno do sinal por exponenciais complexas infinitas. Assim, para que a transformada de Fourier exista, o sinal $x(t)$ deve satisfazer as seguintes condições:

- Número finito de descontinuidades.
- Número finito de máximos e mínimos.
- Absolutamente somável.

O uso computacional da transformada de Fourier em processamento de sinais tem início em meados da década de 60 através do desenvolvimento de algoritmos de Transformada Rápida de Fourier (FFT).

Uma vez que a transformada permite visualizar o sinal no domínio da freqüência, torna-se uma ferramenta extremamente útil na análise de sinais da voz, onde componentes de freqüência podem determinar características específicas da voz.

2.5.3 - Filtros Digitais

Um filtro digital realiza por meios computacionais a ação de filtragem que deve ser executada num sinal. Normalmente o sinal é amostrado e digitalizado, antes de sofrer a filtragem.

O procedimento de projetos de filtros digitais normalmente baseia-se no uso de métodos de precisão analógica. Isto é feito para que se possa tirar o proveito da matemática bem compreendida de tempo discreto, mas de amplitude contínua (HAYKIN, 2001).

Genericamente, um filtro digital consiste em um sistema linear discreto no tempo e invariante ao deslocamento, com a propriedade da entrada e saída satisfazerem a equação linear da diferença dada por:

$$y(n) - \sum_{k=1}^N a_k y(n-k) = \sum_{r=1}^M b_r x(n-r) \quad (5)$$

Aplicando-se a transformada-z em ambos os lados da equação:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{r=0}^M b_r z^{-r}}{1 - \sum_{k=1}^N a_k z^{-k}} \quad (6)$$

A função sistema, $H(z)$, é em geral uma função racional de z^{-1} , a qual é caracterizada pela localização dos pólos e zeros no plano-z (OPPENHEIM e SCHAFER, 1989).

Os filtros digitais podem ser classificados pela duração da resposta ao impulso em filtros de resposta ao impulso de duração finita (FIR) e filtros de resposta ao impulso de duração infinita (IIR). Neste trabalho utilizaram-se os filtros FIR, por suas características de fase linear.

2.5.4 - Wavelet

Apesar de extremamente útil e rápida, a transformada de Fourier tem o inconveniente de não conservar nenhuma informação do tempo do sinal original, $x(t)$, no sinal transformado, $X(w)$, ou seja, no domínio da freqüência não temos informação do tempo, e vice-versa. Assim, novas ferramentas para análise tentam superar esta dificuldade.

A teoria Wavelet tem como base a representação de funções genéricas em termos de uma função primitiva com diferentes escalas e translações. Tal princípio já era aplicado em várias áreas de matemática, física e engenharia, desde o inicio do século, quando matemáticos perceberam que as técnicas de Fourier não eram muito adequadas para solução de muitos problemas.

Somente no início dos anos 80 a teoria de Wavelet começou a ganhar corpo. Em 1982, Jean Morlet usou as wavelets em seus trabalhos de geoexploração. Alex Grossman, Morlet e Yves Meyer estudaram a transformada Wavelet e perceberam que as técnicas da teoria de Calderón-Zygmund, em particular as representações de Littlewood-Paley, poderiam substituir as séries de Fourier em aplicações numéricas. Daí construíram as bases matemáticas da teoria Wavelet, com ênfase nas representações de sinais por “blocos construtivos”. Grossman e Morlet sugeriram o nome wavelet para os blocos construtivos, e o que antes se chamava teoria de Littlewood-Paley, passou-se a denominar Teoria Wavelet (JAWERTH e SWELDENS, 1994).

A atenção da comunidade de Processamento de Sinais foi atraída para a área quando Stéphan Mallat e Yves Meyer, introduzindo a noção de análise em multiresolução, estabeleceram as conexões com outras técnicas de processamento de sinais. A partir de então o número de contribuições teóricas e práticas à essa teoria cresceram, e continua evoluindo.

Tipos de Wavelets

Existem vários tipos de wavelets citados na literatura. O uso de uma ou outra está associado à aplicação. Regras de construção de wavelets estão sendo propostas por vários pesquisadores, segundo as restrições e necessidades que cada aplicação específica impõe. Isto nos leva a concluir que podemos gerar uma infinidade de wavelets diferentes, e particularmente construir um conjunto de wavelets adequado ao

processamento de um tipo de sinal ou aplicação específica, levando à obtenção de resultados melhores.

Condição de Admissibilidade

Para uma função ser considerada uma wavelet ela deve ser admissível.

Uma função $h(t) \in L^2(\mathbf{R})$ é admissível se:

$$C_g = \int_{-\infty}^{+\infty} \frac{|H(w)|^2}{|w|} dw < \infty \quad (7)$$

onde $H(w)$ é a transformada de Fourier de $h(t)$, e $L^2(\mathbf{R})$ é o conjunto de todos os sinais integráveis quadraticamente (i.e., de energia finita).

C_g é a constante de admissibilidade. A principal motivação em impor que C_g seja finita é que se torna condição para que a integral da transformada inversa converja, i.e., a transformada inversa só será possível se $C_g < \infty$.

Transformada de Wavelet

A transformada contínua de Wavelet é definida como:

$$X_{DWT}(a, b) = \int_{-\infty}^{+\infty} x(t) \psi_{a,b}(t) dt \quad (8)$$

onde:

$x(t)$ é o sinal que se deseja a transformada.

$\psi_{a,b}(t)$ é a função wavelet transladada e escalonada derivada da seguinte transformação:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (9)$$

a é o fator de escalonamento e b o fator de deslocamento

X_{DWT} são os coeficientes da transformada wavelet que representam a transformada wavelet para cada escala a .

Se a e b estão limitados para valores inteiros, tem-se a transformada wavelet discreta, cuja integral é aproximada por um somatório, tendo-se:

$$X_{DWT}(a,b) = \sum_{n=0}^{N-1} x(n)\psi_{a,b}(n) \quad (10)$$

Então, determinada uma função wavelet passa bandas $y(n)$ para um sinal finito amostrado discretamente $x(n)$, a transformada wavelet discreta funciona como um banco de filtros de constante Q e divide o sinal nas componentes da banda de passagem. Isto é útil na determinação de características do sinal localizadas em freqüência.

Algoritmo de decomposição em árvore

É inviável calcular a transformada para todos os valores de a e b no conjunto dos números reais, assim faz-se a seguinte restrição:

$$a=2^m \quad \text{e} \quad b=n2^m$$

onde m e n são números inteiros. Esse procedimento conduz a uma estrutura de escalas e translações chamada “diádica”, cuja translação e escalonamento é uma potência de dois. Nesse caso, o resultado da aplicação da TW sobre um sinal é um conjunto de coeficientes wavelet indexados por m (nível de escala) e n (índice de translação). Pode-se mostrar que dessa forma a informação do sinal é preservada e o número de coeficientes é igual ao número original de variáveis, como na TF. Além disso, torna-se possível obter os coeficientes de uma forma rápida e computacionalmente eficiente, através do algoritmo de decomposição em árvore proposto por Mallat (1992) e representado esquematicamente na Figura 8.

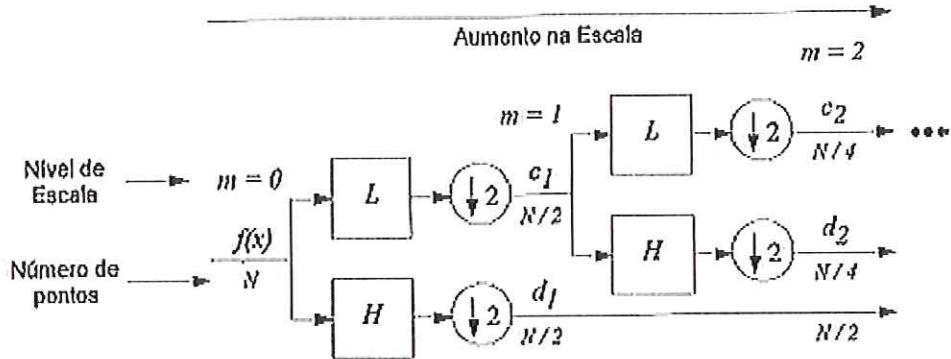


Figura 8: Algoritmo de decomposição wavelet em árvore.

Adotando-se que a freqüência fundamental da fala sonora para os homens e mulheres está na faixa de 30 a 500Hz, e sons não vocalizados normalmente são modelados como ruído branco, compostos principalmente por componentes de altas freqüências, a transformada wavelet pode ser aplicada devido a sua funcionalidade em separar ruídos da fala sonora nas diferentes escalas. Então, usando escalas que contenham informação de fala sonora, a análise pode ser executada para segmentação e determinação de pitch.

2.5.5 - Autocorrelação

A função de autocorrelação de um sinal é obtida pela seguinte equação:

$$\phi[k] = \sum_{m=-\infty}^{\infty} x[m]x[m+k] \quad (11)$$

se o sinal tem periodicidade P, então:

$$\phi[k+P] = \sum_{m=-\infty}^{\infty} x[m]x[m+k+P] = \sum_{m=-\infty}^{\infty} x[m]x[m+k] = \phi[k] \quad (12)$$

assim temos:

$\phi[k]$ tem um máximo em $k=0$.

A autocorrelação de um sinal periódico é periódica.

Estes dois fatos mostram que a função de autocorrelação terá picos para cada múltiplo inteiro do período P.

2.5.6 - Cepstrum

O Cepstrum é a transformada inversa de Fourier do logaritmo do espectro de potências de um sinal e pertence à área de processamento de sinais homomórficos. Foi introduzido inicialmente em 1963 por Bogert, Healy e Tukey (1989). Eles descobriram que o logaritmo do espectro de potências de um sinal contendo um eco possui uma componente aditiva periódica devido a este eco, e assim a Transformada de Fourier do logaritmo do espectro de potências poderia mostrar um pico referente ao atraso do eco. Eles chamaram esta função de Cepstrum, um anagrama das letras da palavra espectro (spectrum) e definiram um extenso vocabulário para descrever esta nova técnica de processamento de sinais. Atualmente, porém, somente os termos Cepstrum e quefrency têm sido freqüentemente utilizados.

Paralelamente, Oppenheim (1989) propôs uma nova classe de sistemas chamada "sistemas homomórficos". Embora não-lineares no senso clássico, estes sistemas satisfazem a generalização do princípio de superposição. O conceito de filtragem homomórfica é bastante geral, mas tem sido estudado de forma mais extensiva para a combinação de operações de multiplicação e convolução, pois muitos modelos de sinais envolvem estas operações.

A transformação de um sinal em seu Cepstrum é uma transformação homomórfica e o conceito de Cepstrum é uma parte fundamental da "teoria dos sistemas homomórficos" para processamento de sinais que tem sido combinados por convolução.

O Cepstrum complexo

Considerando uma seqüência $x[n]$ cuja transformada-z está na forma polar:

$$X(z) = |X(z)| e^{j\angle X(z)} \quad (13)$$

Desde que $x[n]$ seja estável, a região de convergência para $X(z)$ inclui o círculo unitário, e a transformada de Fourier de $x[n]$ existe e é igual a $X(e^{j\omega})$. O Cepstrum

complexo correspondente a $x[n]$ é definido como sendo uma seqüência estável, $\hat{x}[n]$, cuja transformada-z é:

$$\hat{X}(z) = \ln[X(z)] = \ln[|X(z)| e^{j\angle X(z)}] = \ln |X(z)| + j\angle X(z) \quad (14)$$

Sendo o operador \ln o logaritmo neperiano complexo. Apesar de qualquer base poder ser utilizada no logaritmo, o logaritmo natural é o mais utilizado.

O Cepstrum complexo existe se $\log [X(z)]$ possui uma representação convergente de séries de potências na forma:

$$\hat{X}(z) = \log[X(z)] = \sum_{n=-\infty}^{\infty} \hat{x}[n] z^{-n}, \quad |z|=1 \quad (15)$$

Assim, $\log [X(z)]$ deve possuir todas as propriedades da transformada-z de uma seqüência estável. Especificamente, a região de convergência para a representação das séries de potências do $\log [X(z)]$ deve ser da forma:

$$r_R < |z| < r_L$$

onde $0 < r_R < 1$ e $r_L > 1$. Se este é o caso, a seqüência dos coeficientes das séries de potência corresponde ao Cepstrum complexo de $x[n]$ e também é dado pela transformada-z inversa.

$$\hat{x}[n] = \frac{1}{2\pi j} \oint_C \log[X(z)] z^{n-1} dz \quad (16)$$

onde o contorno da integração está dentro da região de convergência. Como foi definido inicialmente, que $\hat{x}[n]$ é estável, a região de convergência inclui o círculo unitário, e o Cepstrum complexo pode ser representado utilizando-se a transformada inversa de Fourier.

$$\hat{x}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} [\log |X(e^{jw})| + j\angle X(e^{jw})] e^{jn w} dw \quad (17)$$

O uso do termo Cepstrum complexo neste contexto implica que o logaritmo complexo é utilizado na definição. Isto não significa que o Cepstrum complexo seja necessariamente uma seqüência de números complexos. Certamente, como vimos rapidamente, a definição escolhida para o logaritmo complexo assegura que o Cepstrum complexo de uma seqüência de números reais será também uma seqüência de números reais.

O Cepstrum real

Diferente do cepstrum complexo, o cepstrum real, $c_x[n]$, de um sinal é definido como a transformada inversa de Fourier do logaritmo do módulo da transformada de Fourier.

$$c_x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |X(e^{jw})| e^{jnw} dw \quad (18)$$

Relacionando-se as definições apresentadas para o cepstrum complexo e real, pode-se verificar que $c_x[n]$ é a transformada inversa de Fourier da parte real de $\hat{X}(e^w)$. Consequentemente o cepstrum real é igual à conjugada simétrica parcial de $\hat{x}[n]$:

$$c_x[n] = \frac{\hat{x}[n] + \hat{x}^*[-n]}{2} \quad (19)$$

O Cepstrum real é muito útil em aplicações onde não se depende da fase, além de ser muito mais fácil de se implementar computacionalmente que o Cepstrum complexo.

3 – Métodos e Procedimentos

Este capítulo trata dos métodos analisados e os procedimentos adotados para a análise. Todos os algoritmos e programas desenvolvidos especificamente para o trabalho foram elaborados usando a linguagem interpretada do programa Matlab 6.0. Tal software executava sobre um sistema operacional Windows XP.

O uso deste programa comercial justifica-se pelo grande número de rotinas matemáticas e de processamento de sinais disponíveis para programação na biblioteca de funções do Matlab. Estas rotinas facilitam a implementação em detrimento do tempo de execução das rotinas, uma vez que se trata de uma linguagem interpretada. Dentre as vantagens de se usar tal programa para implementação dos algoritmos é a possibilidade de efetuar testes com saídas gráficas sem o peso de uma programação visual. O uso de rotinas gráficas facilitam a visualização dos resultados na forma de gráficos por um processo bastante simples, bem como, o desenvolvimento de programas com interface gráfica para o usuário. O uso de programas com interface gráfica facilitam a escolha do sinal de voz, a escolha de um trecho mais estável do sinal, a escolha das opções de filtros e métodos, e a visualização imediata de qualquer mudança efetuada.

3.1 – Sinais de voz

Os sinais de voz utilizados neste trabalho foram obtidos no banco de dados com sinais de vozes do Hospital Das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

Utilizou-se para os testes um total de 46 vozes com vogais sustentadas, sendo 23 masculinas e 23 femininas. Do total, 40 vozes contém alguma patologia, sendo estas 20 vozes masculinas e 20 femininas. As vozes normais contam com 3 vozes masculinas e 3 vozes femininas, perfazendo um total de 6 vozes normais. O maior número de vozes patológicas justifica-se da dificuldade de determinação dos períodos nestas vozes, o que

é objetivo deste estudo. O período médio ou pitch de cada voz foi obtido utilizando-se o determinador semi-automático, e estes valores servirão como padrão de comparação. Cada um dos sinais de voz coletados é composto por vogais /a/, /e/ e /i/ pertencentes a um indivíduo, com um número indeterminado de períodos para cada voz, não sendo este número menor que 100(cem). Assim, cada voz é composta por sinais separados para cada uma das vogais citadas, totalizando 138 sinais de voz para análise. Atribuiu-se um rótulo com índice crescente (Voz1..Voz46) para cada voz , para facilitar a elaboração dos testes.

Além dos sinais do banco de dados, utilizaram-se sinais de voz sintetizados, que pudessem refletir as variações da voz.

A voz humana apresenta *jitter*, *shimmer* e ruídos presentes em menor ou maior grau. A estes parâmetros une-se a diferença da freqüência fundamental de cada pessoa, o que totaliza um conjunto de quatro variáveis presentes em sinais de vozes reais. Em uma voz sintetizada é possível variar cada um desses parâmetros independentemente.

O objetivo de utilizar esses sinais de vozes sintetizadas é observar o comportamento dos algoritmos estimadores e determinadores com a variação de apenas um parâmetro de cada vez.

Para isto projetou-se um programa com interface gráfica que gerava um sinal de voz com cinqüenta períodos, sendo este número de períodos o mesmo para todos os sinais gerados para teste. Por meio de controles de barras é possível alterar separadamente a freqüência fundamental, a relação sinal ruído, *jitter* e *shimmer*; mantendo apenas o número de períodos inalterado. A figura 9 mostra uma tela do programa apresentando um sinal com variações no *jitter* e no *shimmer*.

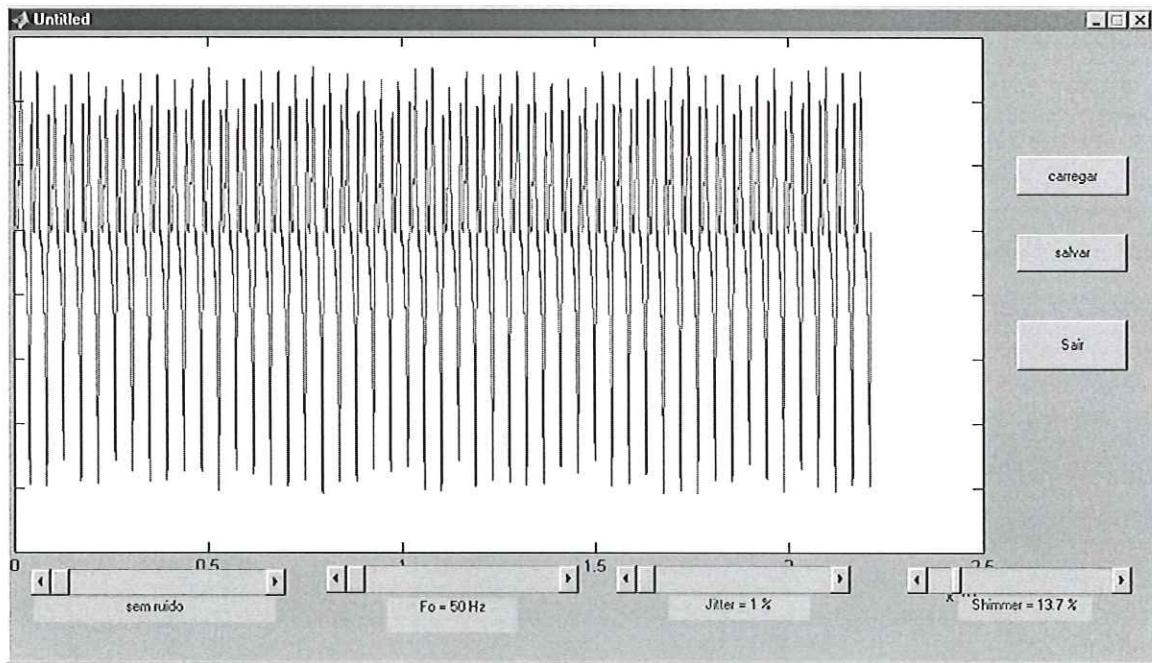


Figura 9: Tela do programa para síntese de sinais de voz.

3.2 – Determinadores de pitch

Os Determinadores de pitch no domínio do tempo foram um dos primeiros métodos de medida automática de pitch. Por trabalharem diretamente sobre o sinal de voz e sua relativa simplicidade quando comparados com os métodos que utilizam o espectro da voz, estes tiveram inicialmente implementações analógicas que posteriormente foram transportadas em código para os processos digitais.

Hess (1983) usa uma divisão clássica na qual faz uma representação geral dos determinadores em um diagrama de bloco como mostrado na figura 10.

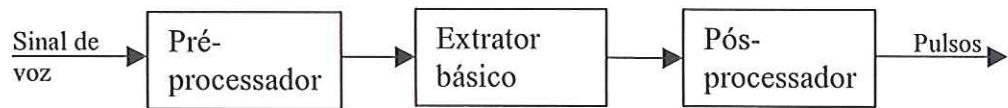


Figura 10: Diagrama de blocos geral para os determinadores.

Os algoritmos implementados neste trabalho contêm cada um dos blocos apresentados na figura 10. Entretanto, algumas funções destes algoritmos podem agrupar dois blocos, não permitindo uma separação tão nítida quando analisado o código do algoritmo.

O pré-processamento do sinal tem como objetivo facilitar o trabalho do extrator básico através de filtragem ou redução de dados (sub-amostragem) do sinal de voz. O Extrator básico irá determinar cada um dos períodos do sinal de voz sendo sua saída, normalmente, uma sucessão de pulsos (marcas) que servem para determinar um ponto significativo em cada período individualmente. O pós-processamento é mais orientado para o respectivo aplicativo que o determinador. Em dispositivos mais antigos este era principalmente uma conversão de tempo para freqüência, devido ao esforço analógico necessário. Com técnicas digitais a conversão requerida é uma tarefa simples e assim é possível o uso de técnicas mais complexas de pós-processamento como detecção e correção de erros, indicação de regularidade, ou suavização e edição do contorno obtido.

O uso de determinadores diretamente sobre o sinal de voz é possível considerando algumas características deste sinal e do trato vocal que o produziu. Hess (1983) destaca:

- a presença de um harmônico fundamental
- um padrão estrutural que se repete aproximadamente de período para período.

Assumindo o trato vocal como um sistema passivo linear (FANT, 1960), sua resposta de impulso consiste na soma de senóides exponencialmente amortecidas. Desta forma, encontra-se amplitudes altas no inicio de cada período e baixas amplitudes no fim.

Se um sistema linear for excitado por um trem de pulso, o sinal resultante pode exibir descontinuidades para esses momentos onde os pulsos individuais acontecem. Se isto não ocorre no próprio sinal, seguramente ocorrerá em uma derivada de mais alta ordem do sinal (HESS, 1983).

Este trabalho aborda algoritmos determinadores de pitch que se utilizam destas características para a determinação sobre o sinal direto ou indireto, sendo primeiramente abordados os algoritmos mais clássicos e posteriormente os mais atuais.

3.2.1 - Determinador de Pitch por extração da harmônica fundamental

Detecção da harmônica fundamental é a técnica clássica para detecção de pitch no domínio do tempo usando tecnologia analógica. Um filtro linear forma o pré-processador, seguido por um determinador básico relativamente simples. Há três tipos principais de extratores básicos:

1. determinador por análise de cruzamento por zero
2. determinador por análise de um limiar diferente de zero
3. determinador por análise de um limiar e histerese (dois extratores básicos por limiar)

O determinador por análise de cruzamento por zero é o mais simples de todos o que requer uma maior filtragem no pré-processamento. Tal filtragem pode ser linear ou não linear, sendo abordada neste trabalho apenas filtragem linear para este determinador.

A técnica de extração da harmônica fundamental necessita que exista uma primeira harmônica predominante na forma de onda. Isto restringe a aplicação destes determinadores para esses casos onde o sinal não é limitado em nenhuma faixa de freqüência, a menos que um pré-processamento não linear reconstrua a primeira parcela.

Como desvantagem, os determinadores que usam este princípio são em geral sensíveis à distorção em baixa freqüência no sinal. Em um sistema onde a situação ambiental é imprevisível (telefonia), este determinador foi substituído por outros dispositivos mais robustos. Para aplicações onde as condições ambientais são previsíveis (fonética e patologias), o princípio deste determinador é altamente atraente devido a sua simplicidade para implementação computacional.

Neste trabalho optou-se por usar o determinador por análise de cruzamento por zero, o que implica em um aumento na demanda de filtragem do sinal. Por tratar-se de um processamento digital de sinais, a filtragem torna-se uma tarefa relativamente simples quando inserida no algoritmo computacional. Isto permite o uso do determinador escolhido sem acarretar uma grande demanda de processamento na etapa de filtragem do sinal.

Para que o determinador funcione perfeitamente é requerido que o sinal que chega até este, sendo o mesmo sinal de saída do pré-processador, tenha dois e somente dois cruzamentos por zero para cada período. Levando-se em conta a grande variabilidade da freqüência fundamental de um grande número de vozes, esta tarefa não é tão simples de se executar. Um filtro passa-baixas pode garantir o pré-processamento desejado desde que sua freqüência de corte seja bem escolhida. Como citado, a variabilidade não permite que se defina um valor apenas para um conjunto de sinais de voz, ou seja, o valor deve ser definido em função do sinal que está sendo analisado.

Embora se possa determinar empiricamente a melhor freqüência de corte para o filtro do pré-processador, a discussão detalhada deste problema foi estudada por McKinney (1965).

Segundo McKinney, a seguinte condição é necessária e suficiente para garantir que dois e somente dois cruzamentos por zero ocorrerão por período:

$$RS1 = 20 \log \frac{A(1)}{\sum_{k=2}^M kA(k)} > 0(db) \quad (20)$$

onde RS1 é a relação (em dB) da amplitude do primeiro harmônico, A(1), e a soma aritmética das amplitudes das harmônicas mais relevantes, A(m), sendo estas multiplicadas pelos respectivos números de seus harmônicos. A amplitude do componente DC é suposta ser zero.

Se a equação 20 é violada, é possível que ocorra dois cruzamentos por zero por período para certas relações de fase entre o primeiro e os demais harmônicos, não sendo garantido para todas as relações de fase. Assumindo que todos os harmônicos são determinados em uma representação senoidal,

$$a_k(n) = A(k) \sin(kn + k\phi) \quad (21)$$

O pior caso é determinado quando os harmônicos de ordem mais alta tem fase contrária à do primeiro harmônico. Neste caso, a derivada do sinal no cruzamento por zero do primeiro harmônico (passando de valores negativos para valores positivos) é determinada por:

$$a'_k(n=0) = A(1) - \sum_{k=2}^M kA(k) \quad (22)$$

sendo que a componente DC tem fase zero. A derivada será positiva se a amplitude da primeira harmônica, A(1), for maior que a soma de todas as harmônicas relevantes multiplicada por seus respectivos índices harmônicos, ou seja, se a equação 20 for satisfeita. Se a derivada $a'(0)$ assume um valor menor que zero, um cruzamento

por zero é substituído por pelo menos três cruzamentos por zero, e o determinador não pode ser utilizado para o sinal analisado.

É possível definir um filtro linear cuja freqüência de corte permita filtrar o sinal retirando deste as harmônicas de mais alta ordem necessárias para satisfazer a equação 20 e assim garantir que ocorrerão apenas dois cruzamentos por zero em cada período, o que facilitaria a tarefa do determinador.

Neste determinador, para a tarefa de filtragem, foi utilizado um filtro passa-baixas FIR cuja freqüência de corte é determinada pela análise do espectro de uma parte do sinal (janela de 2048 pontos), por se tratar de sinais de voz sustentada tal janela é estatisticamente representativa. Esta freqüência de corte depende da freqüência fundamental determinada por um estimador. Este estimador foi determinado através dos testes realizados e demonstrados posteriormente. Usando a estimativa identificam-se quantas harmônicas devem ser eliminadas do sinal para satisfazer a equação 20 e assim determina-se a freqüência de corte do filtro.

O determinador irá identificar os cruzamentos por zero, no caso a passagem de valores negativos para valores positivos, do sinal filtrado. Como se tem a garantia de apenas dois cruzamentos, a tarefa é simples. No algoritmo usado cria-se uma onda quadrada do sinal e anota-se o tempo onde o sinal muda de zero para um. Este procedimento gera um vetor de posições referentes ao início de cada período.

Identificados os picos, relacionam-se estes ao sinal original e efetua-se uma correção de erro no pós-processamento. A escolha do cruzamento por zero justifica-se da necessidade deste trabalho de avaliar vozes patológicas cujas variações de *jitter* e *shimmer* são mais acentuadas que em vozes normais. Usar um cruzamento por zero cria uma referência cuja única interferência é o *jitter*, e facilita a tarefa de encontrar este ponto período a período, mesmo estes variando de tamanho durante todo o sinal.

O código de correção de erro no pós-processamento irá identificar os cruzamentos por zero no sinal mais próximos das marcas determinadas no sinal filtrado, bem como identificar os possíveis erros de falta ou excesso de marcas em um período.

A figura 11 mostra os sinais gerados por este determinador para a identificação dos períodos.

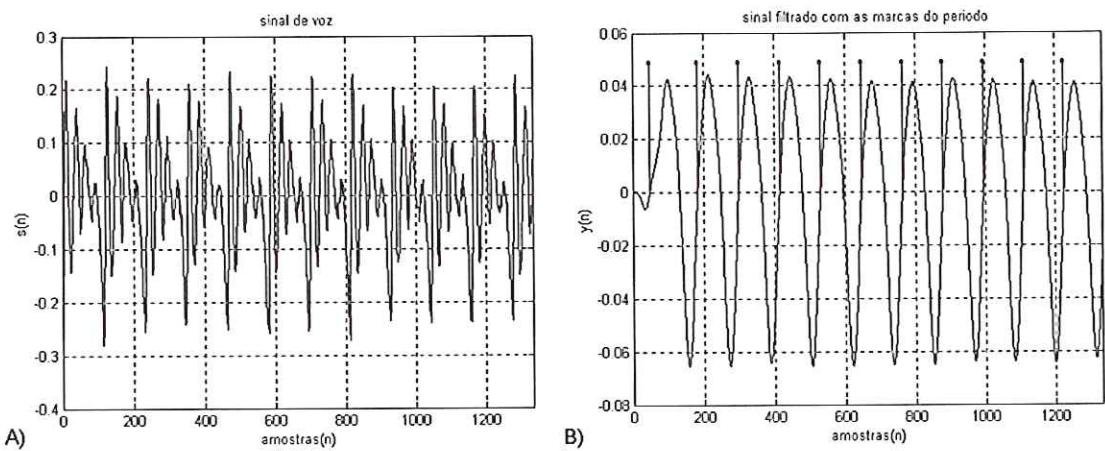


Figura 11: A) Sinal de voz. B) Sinal filtrado e as marcas de períodos determinados.

3.2.2 - Determinador de pitch por autocorrelação

Este método proposto neste trabalho utiliza-se da função de autocorrelação para determinar cada um dos períodos do sinal de voz.

Trabalha-se com uma janela cujo número de amostras depende da freqüência de amostragem, mas que contenha um tempo de aproximadamente 52 ms do sinal de voz. Isto garante para o pior caso(voz com freqüência fundamental de 50Hz) uma janela onde pelo menos dois períodos de voz estarão presentes, permitindo o uso da função de autocorrelação para identificar estes períodos.

A partir do início do sinal procura-se um cruzamento por zero que possa ser o início das marcações. Determinado o início da janela toma-se esta e efetua-se a autocorrelação. Determina-se o período e a posição que este ocorre na janela de autocorrelação, que está diretamente ligada ao sinal de voz, determinando um ponto que será simultaneamente o fim deste período e o início de um novo período. Toma-se uma nova janela com início no ponto determinado na janela anterior e novamente marca-se o fim deste período e consequentemente o início do período seguinte. Prossegue-se assim até o fim do sinal de voz, determinando todos os períodos do sinal.

Este método apresenta todas as deficiências do estimador por autocorrelação e assim testou-se o mesmo com e sem filtragem passa-baixas no sinal antes da autocorrelação, conforme resultados apresentados posteriormente. Outra mudança é o acréscimo de um corretor de erros conforme explicitado posteriormente neste capítulo. Este corretor visa melhorar a robustez do método eliminando falhas simples de falta de marcação.

Este determinador trabalha com a semelhança do sinal consigo mesmo, não tendo um evento (valor do sinal ou um pico) como indicador de início ou fim de período. Assim, a escolha do cruzamento por zero para início das marcas serve apenas para o uso de um corretor de erro futuramente. A figura 12 mostra o processo de determinação e algumas janelas representativas do funcionamento do algoritmo.

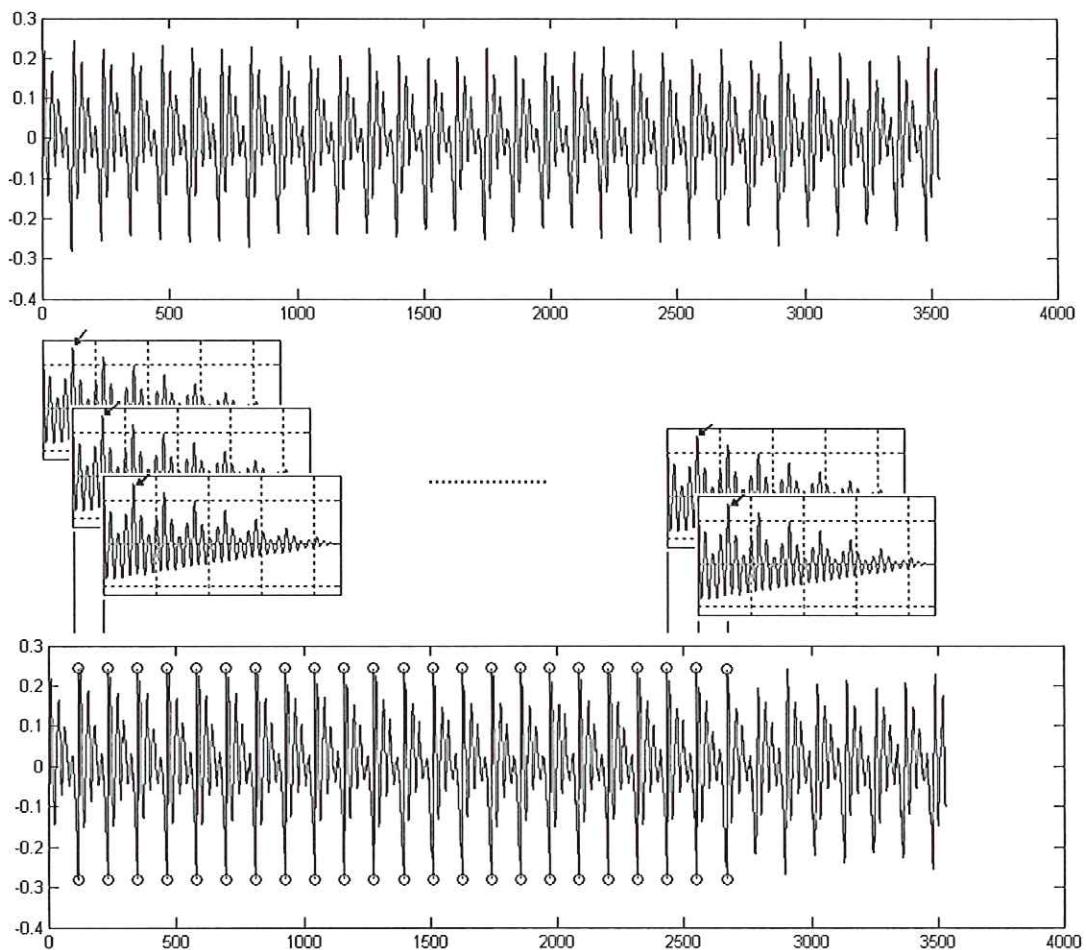


Figura 12: Determinação de pitch através do algoritmo de autocorrelação.

3.2.3 - Determinador de Pitch através da análise da estrutura temporal

Os métodos que utilizam a estrutura temporal tentam modelar o processo de reconhecimento que acontece na formação da mensagem enviada através do olho humano para o cérebro quando o pitch é visualmente determinado em uma figura do sinal de fala, pela semelhança entre os períodos. É um processo complexo de comparação de padrões que só são entendidos parcialmente e de nenhum modo

facilmente traduzidos para a forma algorítmica. Pode-se trabalhar com a envoltória (muito usado analogicamente) ou investigar diretamente a estrutura temporal através de algoritmos procurando e extraíndo pontos relevantes do sinal, dos quais a periodicidade é derivada.

No método usando a envoltória, o período de pitch é modelado por um decaimento exponencial, onde o início de cada período é determinado quando o sinal de voz torna-se maior que a função modelo. Ajustar o valor da constante de decaimento exponencial com a constante de decaimento temporal do sinal de voz é a grande dificuldade do método. O método de investigação direta aplica modelos heurísticos, o que garante uma grande liberdade no algoritmo de modelagem, embora sigam certos padrões de busca.

Basicamente, escolhe-se uma apropriada característica ou um evento no sinal de voz selecionando-os e eliminando os dados restantes; do sinal resultante, procura-se delimitar das características selecionadas, as que podem representar a periodicidade do sinal de voz, eliminando novamente dados não relacionados com estas características, o que se caracteriza como uma segunda redução de dados. A seguir conferem-se todas as marcas selecionadas, verificando se ocorre a formação de um padrão de periodicidade eliminando ou corrigindo os erros. A última parte do algoritmo pode ser relacionada com a operação realizada pelo corretor de erros. Para este trabalho, testaram-se dois determinadores desta categoria, o primeiro proposto por Miller (1974) e detalhado em Hess (1983), e o segundo uma variação deste algoritmo.

O primeiro determinador verifica a cada cruzamento por zero, a extensão da excursão entre as inflexões positivas e negativas, armazenando o valor da soma de todas as amostras deste cruzamento. Em seguida eliminam-se as excursões muito pequenas. Isto causa uma redução nos dados, o que implica em guardar o índice de cada uma das excursões. Procura-se em seguida uma segunda redução dos dados, eliminando as excursões com polaridade inversa da maior excursão. Após as reduções de dados, o algoritmo apresenta um sinal composto por pulsos de valores distintos, que ocorrem em tempos relacionados com o início dos períodos do sinal de voz analisado.

O segundo determinador é mais simples e também mais rápido computacionalmente. Este considera a polaridade do sinal, somando o valor absoluto de todas as amostras entre dois cruzamentos por zero, o que se caracteriza como a primeira redução de dados. A seguir os picos com menor amplitude são eliminados para uma nova redução de dados, e efetua-se a seleção dos picos mais representativos do sinal de

voz. Como no primeiro algoritmo, o índice de ocorrência do cruzamento por zero que identifica cada pico deve ser armazenado. Os picos restantes após a redução coincidem seus tempos de ocorrência com o início de cada período.

Para estes determinadores uma filtragem por um filtro passa-baixas possibilita uma melhora significativa na determinação correta dos períodos, devido a melhora na relação sinal ruído e a eliminação de harmônicas de mais alta ordem.

A figura 13 detalha as formas de onda para o primeiro algoritmo e a figura 14, as formas de onda do segundo algoritmo para o mesmo sinal de voz.

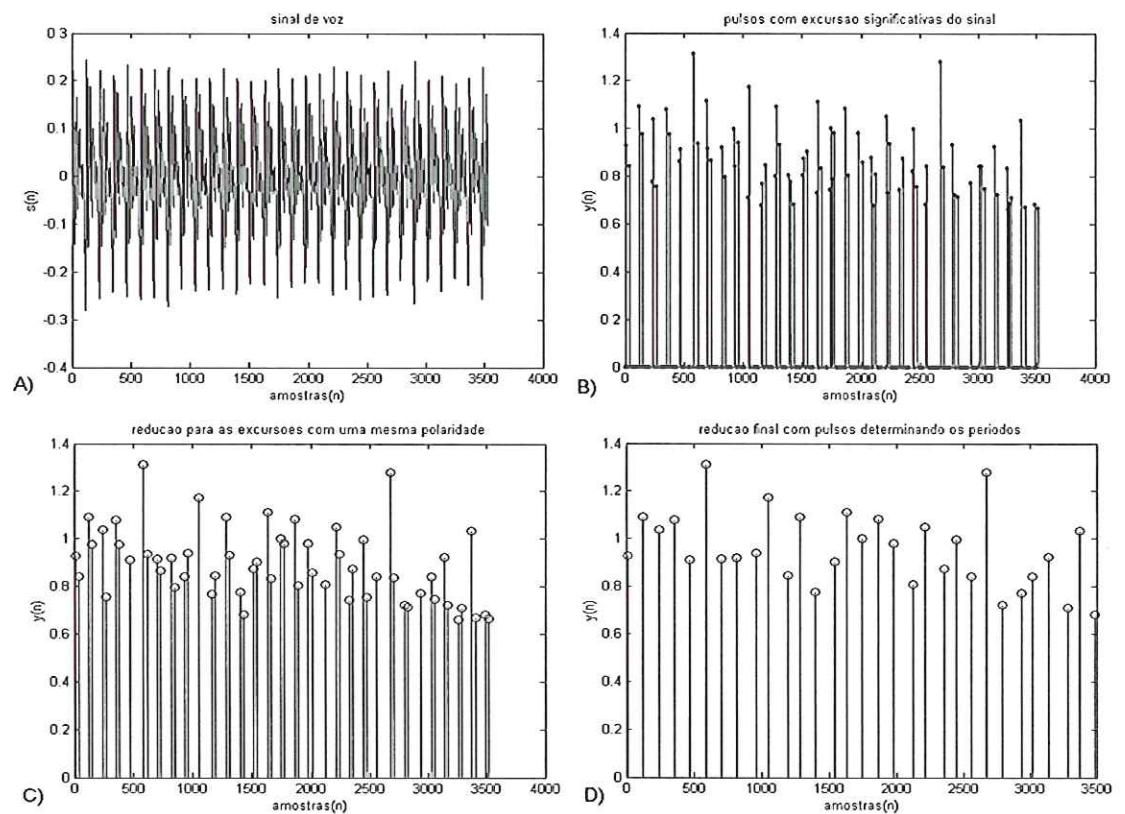


Figura 13: A) Sinal de voz. B) Representação do sinal com as excursões mais significativas. C) Excursões de mesma polaridade que a maior. D) Resultado final com os pulsos indicando os períodos.

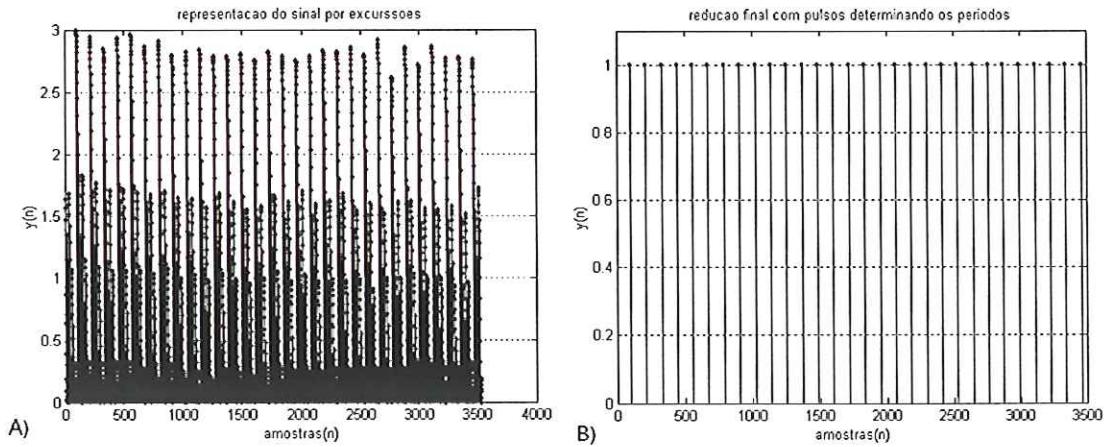


Figura 14: A) representação do sinal por excussoes. B) resultado com os pulsos indicando os períodos.

3.2.4 - Determinador de Pitch por Wavelets

Como apresentado anteriormente, a Transformada Wavelet faz parte de uma teoria relativamente nova quando a comparamos com outras técnicas como FFT, e tem se revelado uma ferramenta poderosa e vantajosa no processamento e análise de sinais para inúmeras aplicações. Para o uso de wavelets com sinais de vozes a escolha da função protótipo exige propriedades derivativas bem como a consideração da largura de banda de freqüência analisada.

O sinal de voz será filtrado, um número determinado de vezes, por uma função filtro formada a partir da função protótipo e função escala. O resultado deste filtro irá permitir identificar os instantes de fechamento glotal (WENDT e PETROPULU, 1996), que identificará o fim de um período e consequentemente o inicio de outro.

Implementou-se o algoritmo proposto por Wendt e Petropulu (1996). Neste algoritmo os autores determinam a função de escala, $\psi_{k_a}(t)$, e a função wavelet, $\varphi_{k_b}(t)$, como

$$\psi_{k_a}(t) = 2^{k_a/2} \psi(2^{k_a} t) \quad (23)$$

$$\varphi_{k_b}(t) = 2^{k_b/2} \varphi(2^{k_b} t)$$

sendo k_a e k_b determinados em função da freqüência de amostragem, F_a , do sinal, e os valores mínimo(30) e máximo(500) da frequência fundamental; conforme as igualdades abaixo:

$$2^{k_a} = \frac{F_a}{30}$$

$$2^{k_b} = \frac{F_a}{500}$$

a partir das funções escala e wavelet determina-se a função filtro derivativa, $\rho(t)$, como sendo:

$$\rho(t) = \psi_{k_a}(t) * \varphi_{k_b}(t) \quad (24)$$

Nesse trabalho utilizou-se como função de escala uma função linear constante e como função wavelet a função ‘haar’. A figura 15 mostra as duas formas de onda calculadas para uma freqüência de amostragem de 22050 Hz, bem como a função filtro resultante. As funções filtro, wavelet e derivativa foram obtidas para se trabalhar com um sinal amostrado, sendo o tempo substituído por amostras nas fórmulas acima.

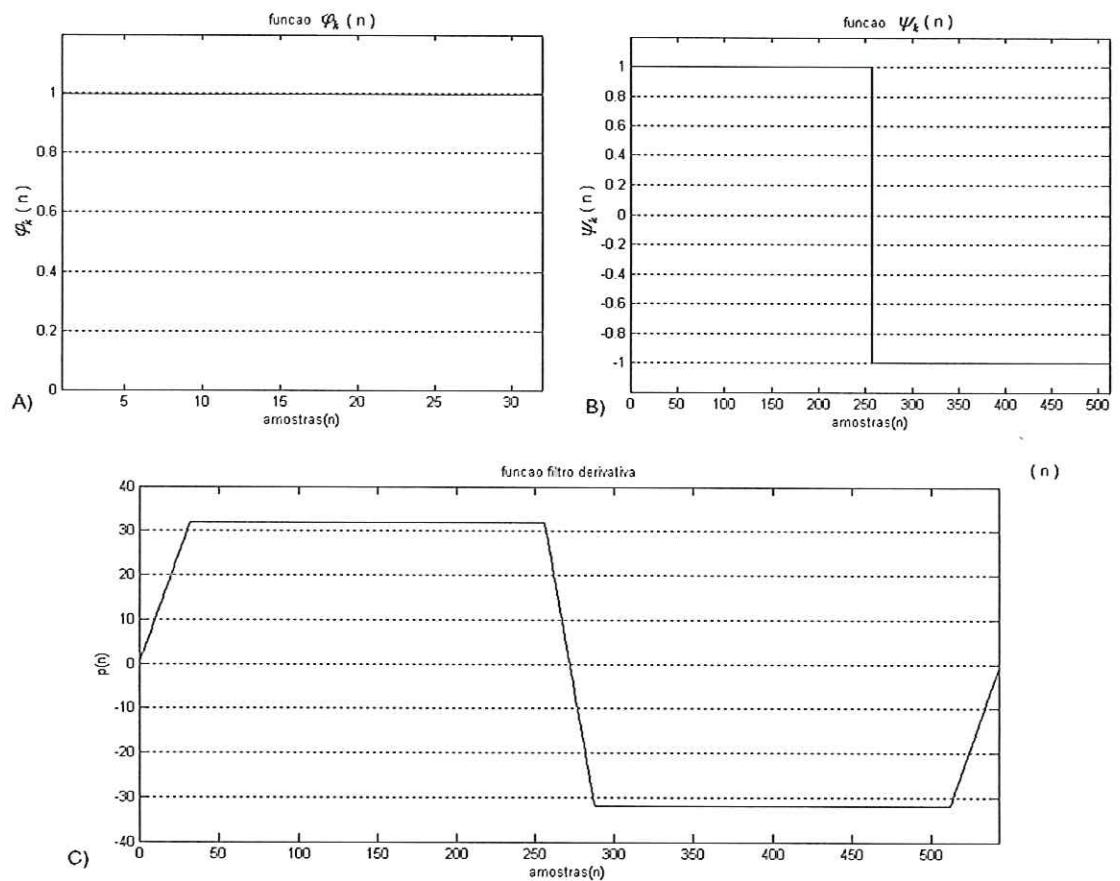


Figura 15: A) Função escala. B) Função Wavelet. C) Função filtro derivativa

Determinada a função filtro, efetua-se a convolução desta com o sinal de voz e obtém-se assim um sinal filtrado. Deste sinal filtrado elimina-se o início e o fim, para evitar um sinal modificado pelo janelamento que ocorre com a convolução. O sinal resultante permite a identificação dos picos e relacionando-os com os períodos do sinal de voz analisado é possível determinar o tempo de cada período, baseando-se na freqüência de amostragem. A figura 16 mostra em A) a forma de onda do sinal de voz e em B) o sinal filtrado com os períodos facilmente identificados através dos picos.

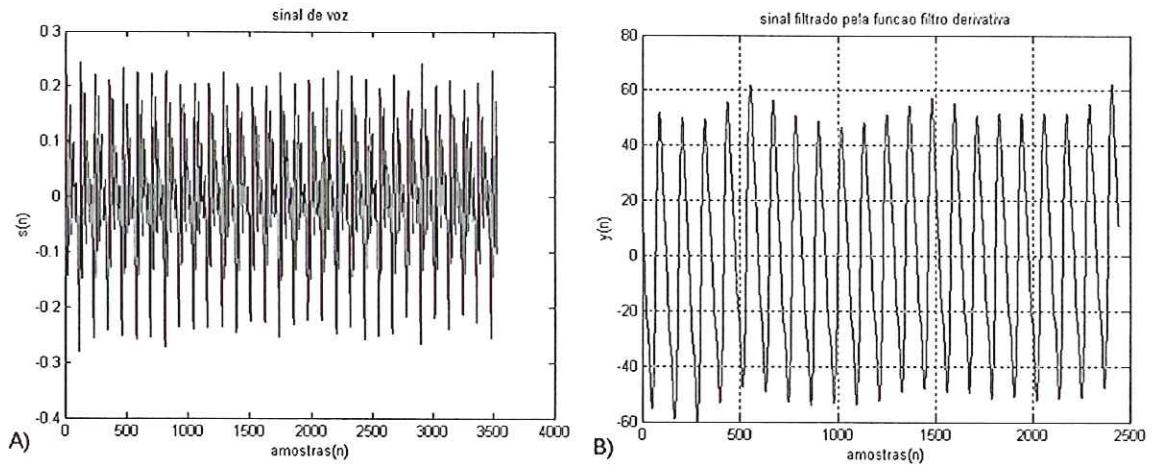


Figura 16: A) Sinal de voz. B) Sinal filtrado pela função filtro derivativa.

Este algoritmo apresenta, como principal característica, a velocidade de processamento em relação a uma técnica mais convencional de wavelet. Ao invés de trabalhar com a variação da escala e do deslocamento da função wavelet calculando os parâmetros a cada variação, este algoritmo constrói a função filtro a partir do valor máximo de escalonamento, k_a , e filtra o sinal, o que se torna extremamente simples na implementação computacional.

3.2.5 - Determinador de Pitch semi-automático

Para a melhor avaliação de outros determinadores e mesmo teste de robustez dos estimadores utilizados, teve-se a necessidade de se determinar com uma maior precisão os períodos presentes nos sinais de voz analisados. Esta tarefa, apesar de onerosa no que se refere ao tempo de execução, mostra-se indispensável para uma comparação entre os determinadores, pois isto só é possível com um valor de referência obtido através deste método semi-automático.

Para realizar este método construiu-se uma interface com o usuário que permitisse a visualização de um sinal de voz através de janelas, sendo o comprimento destas determinados pelo usuário até uma visualização confortável dos períodos. A figura 17 mostra a tela de um sinal tendo seus períodos determinados.

Utilizando-se de um mouse o usuário marca a localização de início ou fim de período, procurando manter sempre a mesma posição relativa ao período para a próxima marca, no próximo período. Esta marca pode ser determinada, preferencialmente, como sendo o cruzamento por zero de maior amplitude em cada período do sinal, com valores

passando de negativo para positivo. Como salientado anteriormente no início de cada período, as excursões são maiores que no fim dos períodos. Como se trata de um método de determinação manual semi-automático, é possível optar por uma inversão de sentido, ou seja, passagem por zero com valores passando de positivo para negativo.

O usuário da interface gráfica apenas indica a marca o mais próximo possível e o tipo de sentido, sendo o cruzamento por zero determinado por um algoritmo através de uma aproximação linear do sinal amostrado determinando o ponto de cada cruzamento. A aproximação linear visa reduzir pela metade o erro de amostragem presente neste tipo de marcação, pois nem sempre ocorre uma amostra exatamente no cruzamento do sinal por zero, tendo-se apenas a certeza do cruzamento.

Existe a possibilidade de se colocar a marca sem um ajuste linear, sendo esta marca determinada diretamente pelo usuário tornando o processo de determinação totalmente manual. Este tipo de marcação manual ocorre da necessidade de ignorar algumas variações que o sinal pode sofrer no cruzamento por zero conforme a figura 17, onde o último pico do sinal, apresentado na tela do programa, tem sua amplitude alterada significativamente a cada período.

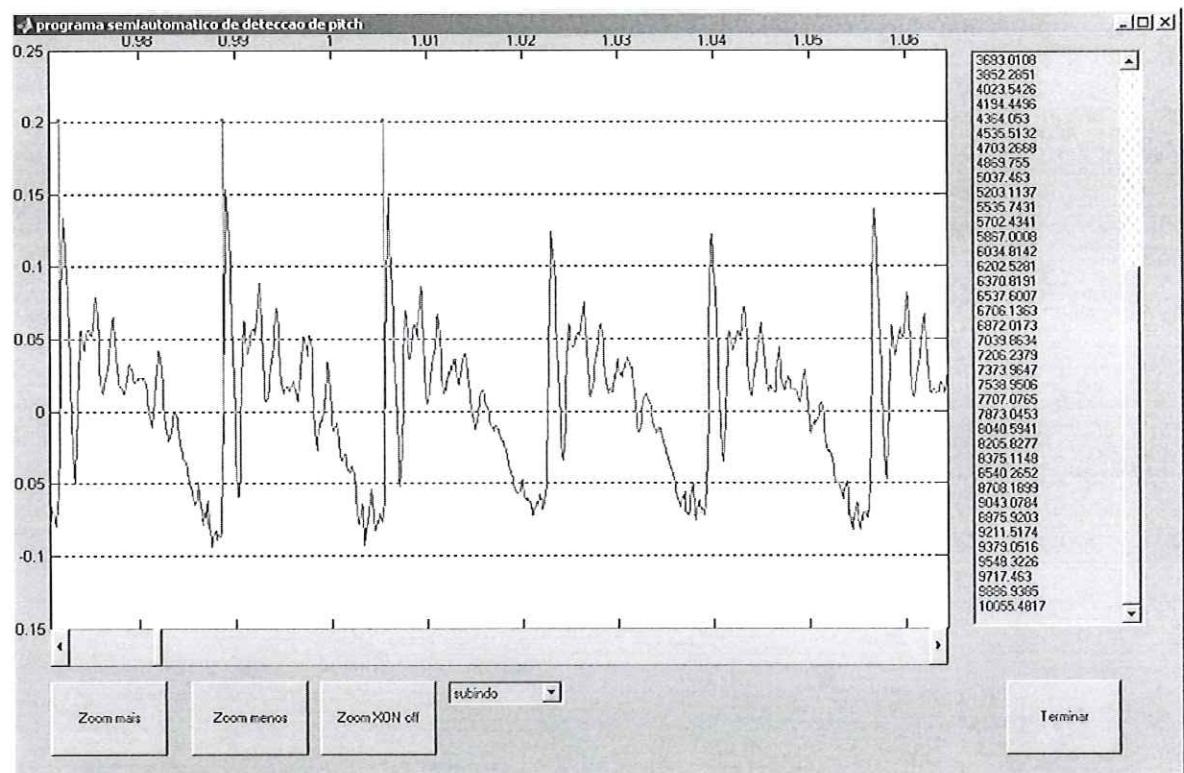


Figura 17: Interface gráfica do determinador semi-automático de pitch

As marcas podem ser apagadas após a colocação das mesmas, e a ordem de determinação não importa, podendo iniciar-se a marcação a partir do fim do sinal de voz. A falta de alguma marca gera um período muito mais longo que os outros, o que faz com que o programa avise o usuário sobre este problema solicitando uma verificação na localização determinada. Esta extensão para ajuste de erros presente no algoritmo visa melhorar a robustez do método eliminando possíveis erros humanos no uso do programa.

A figura 18 mostra uma tela do programa com uma janela de informação solicitando a verificação da marcação e ao fundo se identifica o erro na marcação. Este erro pode ser também de uma marcação incorreta para um cruzamento por zero o que caracteriza dois períodos desiguais, o que informa ao programa para solicitar uma verificação por parte do usuário.

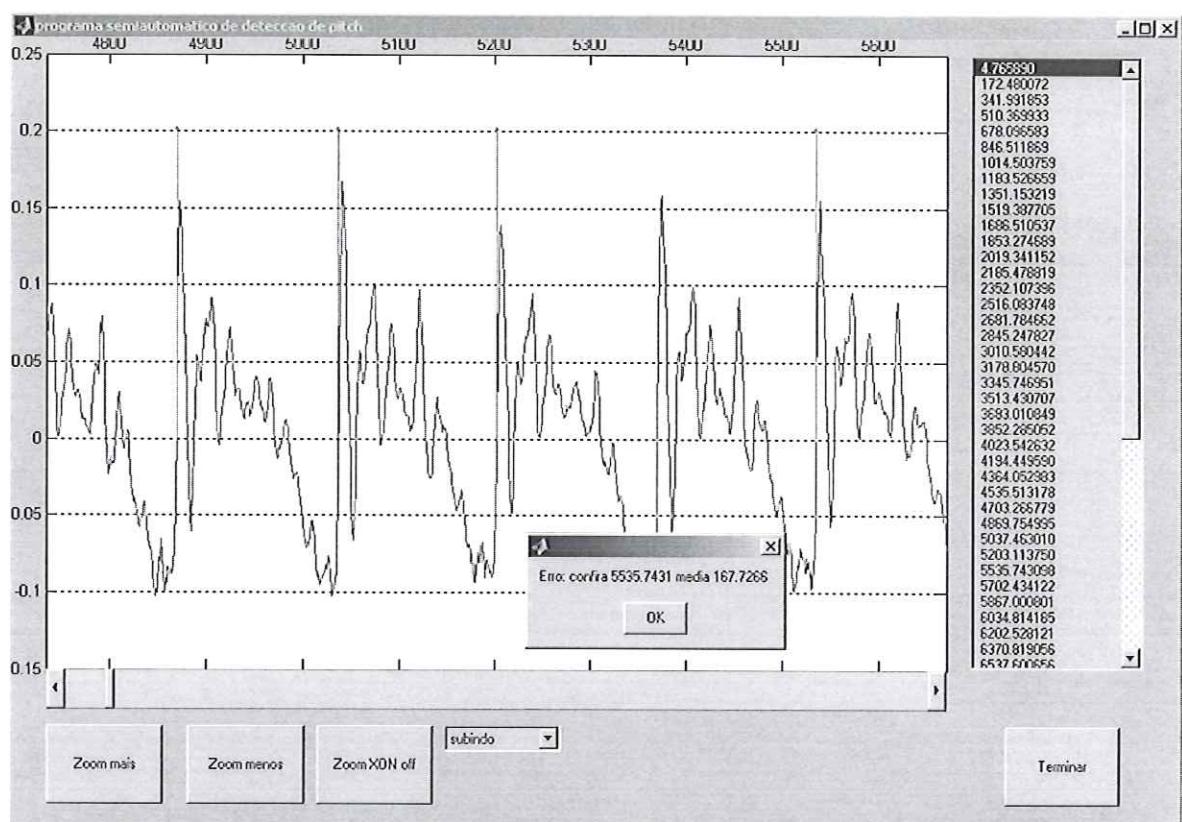


Figura 18: Tela com mensagem de erro decorrente de uma marcação errada.

A saída deste programa é armazenada em um arquivo de texto onde se encontra o nome do arquivo analisado, a freqüência fundamental média, pitch, e o tempo de cada período. Estes valores serão utilizados como um padrão para comparação dos outros métodos de determinação de pitch.

3.3 – Algoritmos estimadores

Os estimadores da freqüência fundamental são extremamente úteis quando não necessitamos do valor de cada período na voz, mas sim, de um valor médio ou aproximado. No processamento do sinal de voz o valor de pitch é usado em inúmeras aplicações como identificação de voz, análise e síntese de voz, diagnóstico de patologias, etc. Uma vantagem destes algoritmos é a velocidade de processamento. Neste trabalho, sua importância está no auxílio para futuras correções nas marcas de períodos realizadas pelos determinadores de pitch.

Nas subseções seguintes, detalham-se os estimadores implementados e testados neste trabalho. Uma explicação dos algoritmos bem como formas de ondas dos sinais observados serão expostas. Os testes realizados, bem como os resultados serão apresentados posteriormente a esta seção. Os testes realizados visam indicar qual, ou quais, estimadores podem ser utilizados para a tarefa de estimação, considerando-se como fator preponderante a robustez para um grande número de vozes.

3.3.1 - Estimador por Autocorrelação

Este método utiliza-se das características da função de autocorrelação, conforme exposto no tópico autocorrelação deste trabalho. Para sinais de voz utiliza-se a função de autocorrelação com janelas, de acordo com:

$$Rn[k] = \sum_{m=-\infty}^{\infty} s[m]w[n-m]s[m+k]w[n-k-m] \quad (25)$$

onde $w[n]$ é uma janela de tamanho N . É importante que pelo menos dois períodos de pitch completos estejam presentes em cada janela para permitir uma melhor estimativa do período. A necessidade de se trabalhar com janelas deve-se ao fato da função de autocorrelação sofrer um decaimento conforme o número de pontos do sinal de voz aumenta (figura 18). Isto ocorre devido ao fato do período de pitch mudar de valor período a período, isto é, o sinal não é perfeitamente periódico.

Percorrendo-se todo o sinal com janelas e calculando o valor do período do sinal em cada uma destas janelas, obtém-se um conjunto de valores dos quais se toma a mediana, para evitar valores esparsos, que pode ser usada como um valor estimado para

a freqüência fundamental do sinal de voz. A precisão do método está relacionada com o tamanho da janela e a precisão de identificar os picos obtidos da função de autocorrelação de cada janela (figura 19). Passar o sinal de voz por um filtro passa-baixas antes de processá-lo, melhora significativamente a precisão do resultado encontrado.

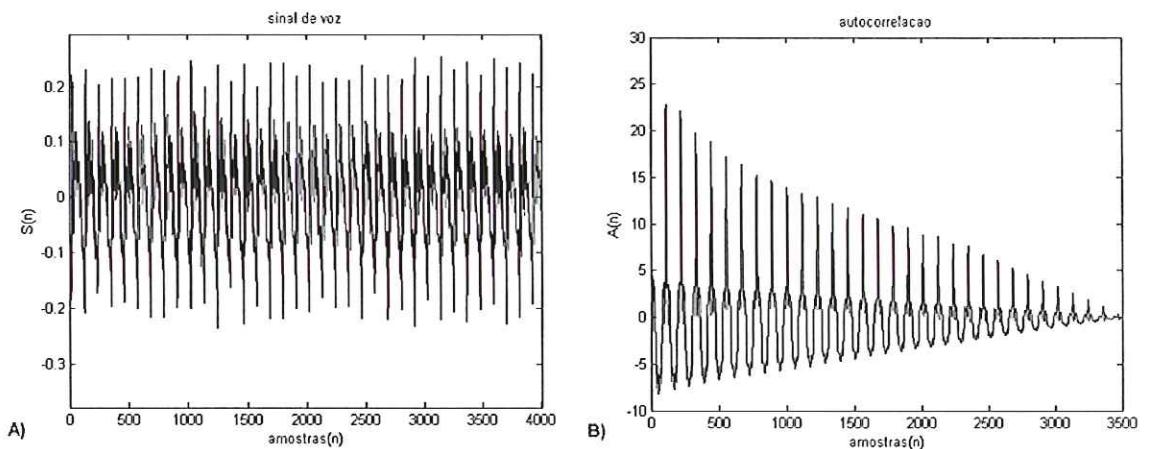


Figura 19: A) sinal de voz da vogal \a\ sustentada. B) Autocorrelação da janela.

3.3.2 - Estimador por Cepstrum

Este método usa o conceito do Cepstrum que permite separar as contribuições do trato vocal da excitação provocada pela vibração das pregas. Foi proposto inicialmente por Noll (1970) e apresenta uma forte dependência dos algoritmos da transformada de Fourier e a transformada inversa de Fourier. Com o surgimento do algoritmo FFT este método melhorou muito sua velocidade de processamento.

Utiliza-se o Cepstrum real para as análises e busca-se o máximo valor dentro de uma janela para determinar a freqüência fundamental.

Mais detalhadamente, o algoritmo busca o valor do Cepstrum real tomando a transformada inversa de Fourier do logaritmo do valor absoluto do espectro do sinal de voz. Para o cálculo do espectro utiliza-se uma janela de Hann de 51,2ms. O tempo é estimado em número de amostras usando a freqüência de amostragem do sinal de voz analisado. Em seguida é efetuada uma busca pela posição, q , do máximo valor do Cepstrum dentro de uma janela de busca, e que tenha um valor maior que um determinado limiar. Esta janela de busca é determinada em função dos valores máximo e mínimo da freqüência fundamental esperada para sinais de voz. Os valores de busca usados são 2 ms e 15 ms para o mínimo e máximo respectivamente. Encontrada esta

posição, q , determina-se a freqüência fundamental usando a freqüência de amostragem, por:

$$FO = \frac{F_{\text{amostragem}}}{q_{\text{valor máximo}}} \quad (26)$$

A figura 20 apresenta os sinais obtidos durante o processo de estimativa de pitch. No item A) tem-se o sinal de voz analisado, em B) o sinal filtrado pela janela de Hann em C) o Cepstrum real do sinal e em D) a janela de busca para o maior valor, sendo este indicado na figura.

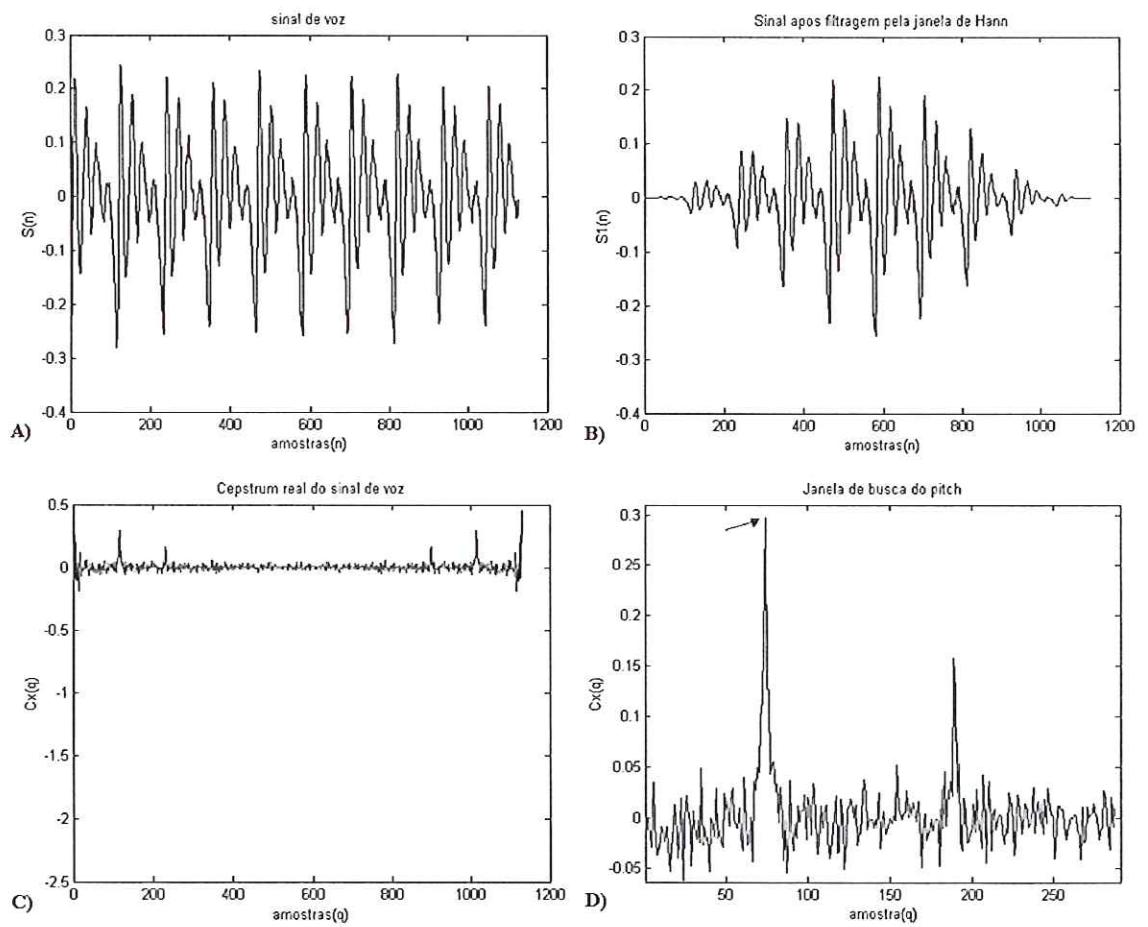


Figura 20: A) Sinal de voz analisado. B) O sinal filtrado pela janela de Hann. C) O Cepstrum real do sinal. D) Janela de busca.

Este processo é repetido em todo o sinal da voz, através do deslocamento da janela de Hann. Os valores obtidos para o pitch são armazenados e determina-se uma média destes valores como sendo a estimativa do pitch encontrada para o sinal de voz.

Este método pode apresentar problemas quando o sinal de voz contiver muito ruído conforme os teste efetuados e apresentados posteriormente.

3.3.3 - Estimador por casamento de harmônicas (harmonics match)

Este estimador trabalha no domínio da freqüência, ou seja, busca a freqüência fundamental do sinal utilizando-se do espectro do mesmo. O princípio de funcionamento deste método é a busca dos picos determinados pela freqüência fundamental e seus harmônicos. Através da comparação do espectro com um sinal composto por um trem de pulsos com freqüência de repetição variável, estima-se o valor do pitch. O trem de pulsos é denominado pente espectral.

Dado um pente espectral, $P(m,q)$:

$$P(m,q) = \begin{cases} 1 & m = kq; \text{ onde } k = 1, 2, 3, \dots, N_p \\ 0 & \text{caso contrário.} \end{cases} \quad (27)$$

onde N_p é o número de impulsos presentes no pente, e q representa a freqüência para a qual serão geradas harmônicas no pente. Ou seja, o sinal gerado terá pulsos apenas nos N_p múltiplos da freqüência desejada.

O espectro do sinal de voz, $E(n)$, que está sendo analisado é multiplicado por este pente, $P(m,q)$, sendo q a freqüência que gera o pente espectral. Efetua-se a somatória do sinal resultante da multiplicação, $E(n).P(m,q)$, para todas as freqüências do espectro e armazena-se estes valores em um sinal $A(q)$. Assim, do espectro teremos apenas N_p valores somados para cada valor de q , cujo resultado é um sinal com as somas para todos os valores de q . Ou seja:

$$A_c(q) = \sum_{m=1}^{N/2} E(m)P(m,q) \quad (28)$$

$N/2$ representa a freqüência máxima do espectro. Neste trabalho limitou-se esta soma em função do número de harmônicas do pente. Para este trabalho determinou-se

que uma boa faixa de freqüência para a busca seria abaixo de 500 Hz, limitando assim o valor de q e suas harmônicas.

O valor de q onde o sinal A_C indicar um máximo é tomado como sendo a estimativa para a freqüência fundamental.

Na figura 21 demonstra-se o funcionamento do estimador onde no item A) tem-se o sinal de voz a ser analisado e em B) o seu espectro. O item C) mostra como ocorre a filtragem pelo pente, variando-se a freqüência do mesmo e somando o sinal filtrado, gerando o sinal A_C . No sinal gerado busca-se o valor da abcissa q onde ocorre o máximo do sinal, item D, que corresponde à freqüência fundamental do sinal, identificada pelo espectro.

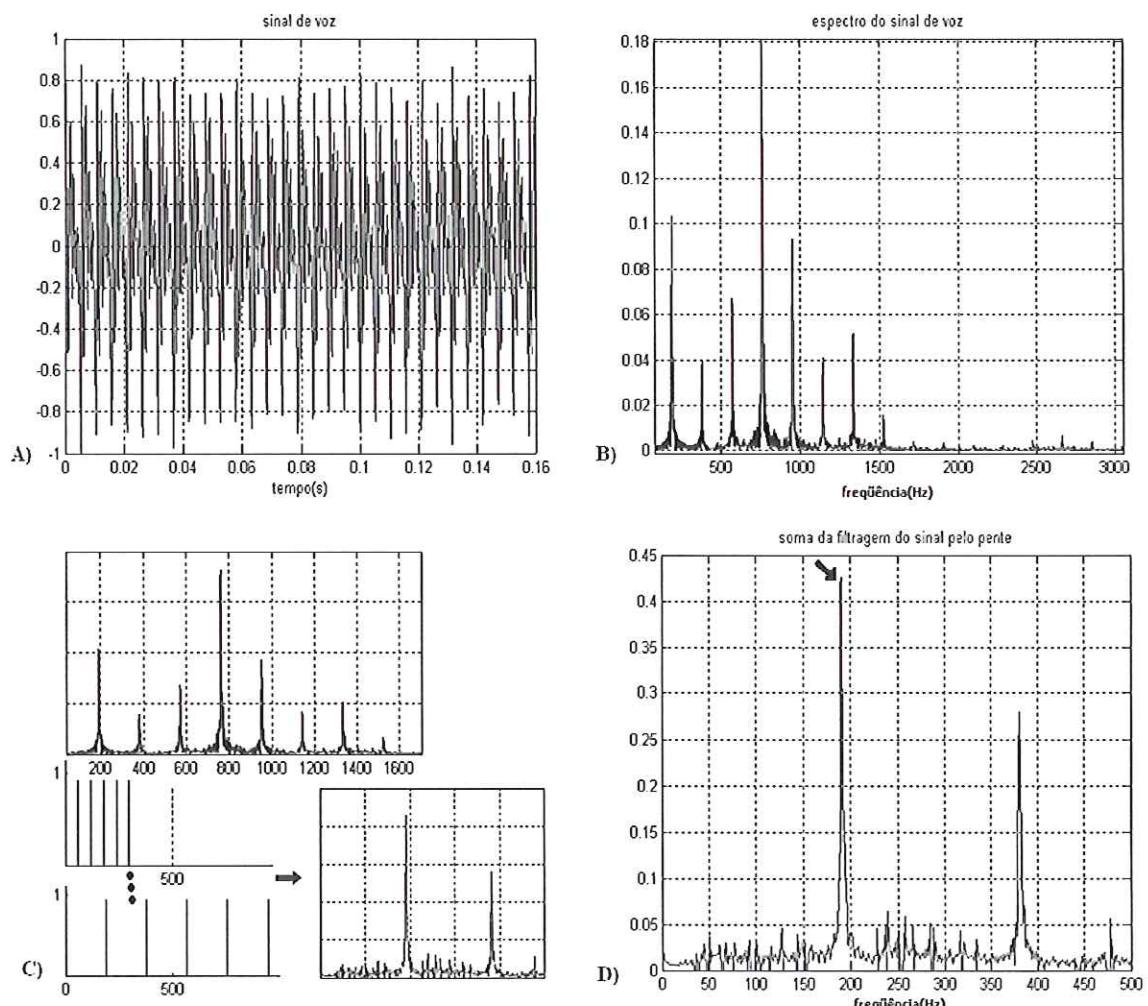


Figura 21: A) sinal de voz. B) espectro do sinal de voz. C) Filtragem pelo pente. D) Soma dos valores da filtragem para cada valor de freqüência do pente, identificado o máximo da função.

Constatou-se que para certos casos ocorria a duplicitade do pico máximo, impedindo uma estimativa correta. Estes picos resultam da filtragem para freqüências submúltiplas da freqüência fundamental. Para evitar este erro considera-se o valor da soma, somente para valores onde a primeira parcela da soma tem um valor maior que 5% do máximo valor do sinal $A_c(q)$ desta freqüência q . Isto faz com que não se tenham picos representativos para sub-harmônicas, melhorando a robustez do método.

3.3.4 - Estimador por AMDF (Average Magnitude Diference Function)

Se a autocorrelação busca semelhança entre sinais ou partes destes, este estimador usa uma função de diferenças sobre o sinal. Na verdade após o cálculo das diferenças procuram-se os mínimos valores que permitirão encontrar a estimativa para a freqüência fundamental.

A função AMDF de um sinal $x(n)$ é definida como (HESS, 1983):

$$AMDF(d) = \frac{1}{K} \sum_{n=0}^{K-d-1} |x(n) - x(n+d)| \quad (29)$$

Esta função terá valores mínimos para n igual a zero e quando o deslocamento d for igual ao período T_0 . Estes mínimos são exatamente zero quando o sinal $x(n)$ é exatamente periódico, enquanto que para os sinais de voz, os mínimos terão seus valores apenas próximos de zero, o que indica o fato de a voz não ser perfeitamente periódica.

Computacionalmente o algoritmo é bastante simples e facilmente implementado, sendo os sinais obtidos mostrados na figura 22. Os mínimos são encontrados e a distância entre os mesmos serve como estimativa da freqüência fundamental do sinal de voz analisado. Assim como o método por autocorrelação, este método é efetuado por janelas que percorrem o sinal e a mediana dos valores encontrados é tomado como estimativa da freqüência fundamental do sinal de voz.

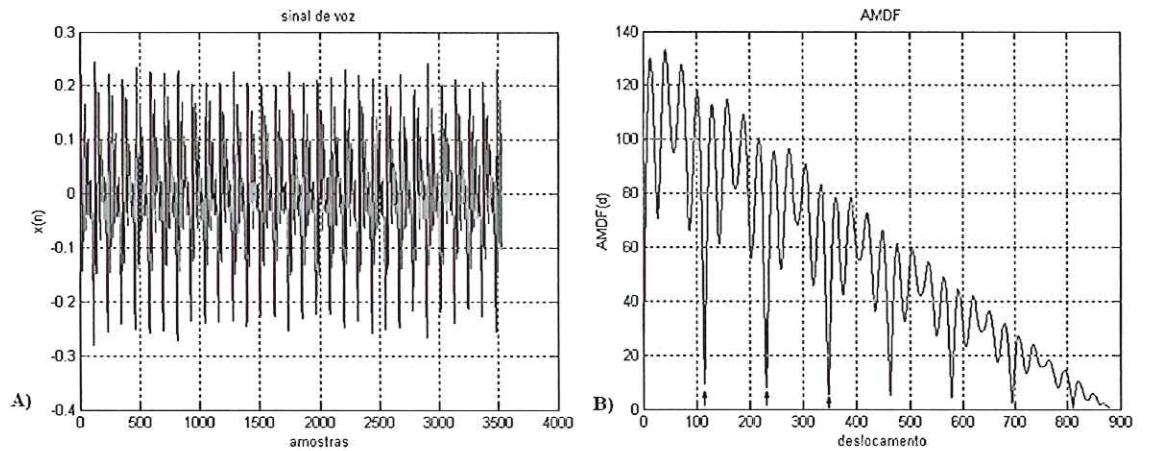


Figura 22: A) Sinal de voz analisado. B) AMDF do sinal, com os primeiros mínimos marcados.

3.3.5 - Estimador por Transformada Wavelet

Diversos algoritmos para estimação de pitch utilizando a transformada Wavelet foram propostos na literatura. Dentre estes, implementou-se o algoritmo proposto por Kadambe (1992). Este algoritmo detecta o fechamento glotal dentro do sinal de voz através da aplicação da transformada Wavelet discreta (DWT). O escalonamento da função primitiva, no caso uma Haar 4, é efetuado em potência de dois.

Inicialmente, faz-se um janelamento do sinal, onde cada janela tem um comprimento de 50ms. O algoritmo como foi proposto limita a busca dentro de uma faixa de escalonamento de 2^3 até 2^5 , com uma condição de parada determinada pela comparação dos sinais obtidos. Para cada nível de decomposição da DWT é obtido um sinal, em que se busca identificar os máximos locais maiores que um limiar de 80% do máximo deste sinal. A cada nível compararam-se os máximos locais do sinal atual com os do sinal anterior. Se os máximos forem os mesmos encerra-se a decomposição e calcula-se o pitch pela distância entre os picos adjacentes. Limita-se a decomposição até o quinto nível.

A figura 23 mostra o sinal para os três níveis de decomposição nos itens B), C) e D) com a indicação dos máximos locais.

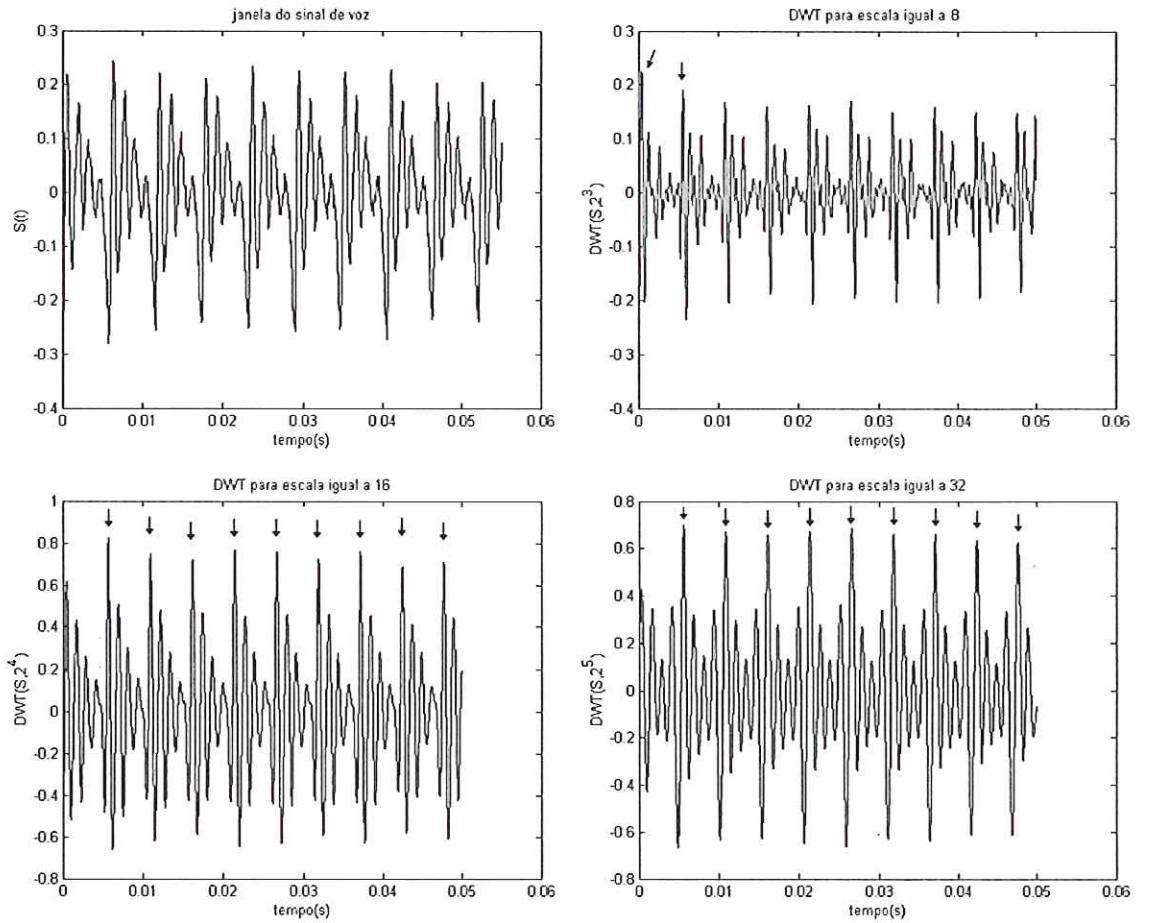


Figura 23: A) Sinal de voz analisado. B) DWT para uma escala de 2^3 , com dois máximos locais C) DWT para uma escala de 2^4 , com 9 máximos locais D) DWT para uma escala de 2^5 , com os mesmos 9 máximos locais da escala anterior.

3.4 – Procedimento de testes dos Algoritmos

Os testes, cujos resultados são expostos no capítulo seguinte, servem para especificar o algoritmo determinador de pitch mais robusto dentre os estudados para vozes patológicas. Os valores padrão para os sinais de voz testados foram tomados como sendo o valor determinado pelo determinador semi-automático.

Os testes iniciais foram realizados para os algoritmos de estimação, pois estes têm apenas uma saída, sendo um procedimento de verificação simples, ou seja, compara-se o valor estimado com o padrão. Estes testes devem selecionar um ou dois algoritmos que serão utilizados para auxiliar os determinadores no processamento dos sinais de voz.

Inicialmente testaram-se os algoritmos estimadores com sinais sintetizados com variação de um parâmetro por vez (freqüência fundamental, SNR, *jitter* e *shimmer*). Apenas os algoritmos que apresentaram melhores resultados foram testados com vozes naturais.

Em seguida realizaram-se testes para os algoritmos determinadores. Novamente utilizaram-se as vozes sintetizadas para os testes iniciais. Verificou-se inicialmente a freqüência fundamental calculada através destes algoritmos quando comparada ao valor padrão. Uma comparação dos valores encontrados período a período foi efetuada para os determinadores que foram mais robustos no teste inicial.

Prosseguindo, realizaram-se os testes com as vozes naturais, sendo efetuados apenas os testes de freqüência fundamental e valores encontrados para os períodos da voz.

O uso de vozes naturais normais e patológicas visa o objetivo final deste trabalho, ou seja, o uso destes algoritmos em vozes naturais que tenham patologias. Isto não descarta o uso do algoritmo em vozes normais, mas sim acrescenta mais exatidão às medidas realizadas nestas. A comparação dos períodos calculados objetiva ser uma apresentação final dos resultados. Os valores dos períodos podem ser comparados visualmente e demonstrando a robustez dos determinadores selecionados.

Os resultados dos testes encontram-se no capítulo seguinte a este.

4 - Resultados

Este capítulo apresenta os resultados encontrados nos testes realizados para os estimadores e determinadores. Os resultados foram organizados em tabelas individuais para cada um dos testes e impressas no apêndice A. Para cada análise realizada apresentam-se os resultados na forma de gráficos baseados nas tabelas.

4.1 – Testes com estimadores

O primeiro teste foi realizado com sinais sintetizados, variando apenas a freqüência fundamental, mantendo os outros possíveis parâmetros sempre com o mesmo valor. Usaram-se sinais sem ruído e com *jitter* e *shimmer* nulos. Isto permitiu verificar se a faixa de freqüência influencia na eficiência dos estimadores. Os dados coletados foram organizados na tabela 1 do apêndice A. O gráfico destes dados esta na figura 24.

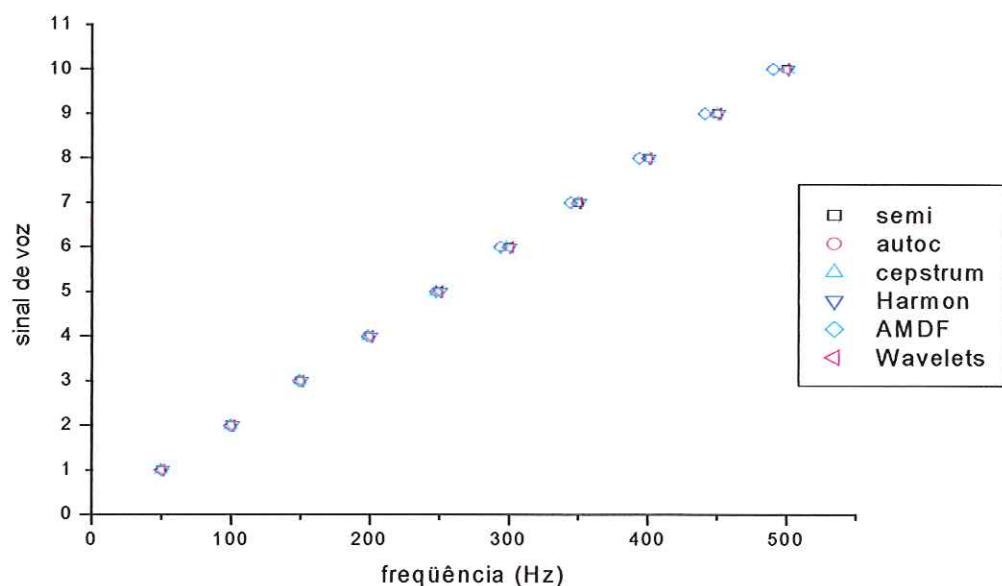


Figura 24: Resultados dos estimadores para variação de Fo nas vozes sintetizadas.

Apesar da baixa resolução do gráfico é possível perceber que os algoritmos estimaram a freqüência fundamental de cada um dos sinais com uma robustez satisfatória. Ou seja, não ocorreu nenhum erro grave como a impossibilidade de estimação do algoritmo ou um erro de mais de 3 %. Segundo a tabela 1 do apêndice A, o maior erro foi de 2,55% para o algoritmo estimador por casamento de harmônicas.

Para análise da variação dos demais parâmetros no sinal de voz, utilizou-se um sinal de voz sintetizada com freqüência fundamental fixa de 200 Hz.

Primeiramente, variou-se a SNR, verificando os valores estimados bem como as falhas dos algoritmos estimadores. Um valor igual a zero indica que não foi possível determinar a estimativa de F_0 . Os dados estão na tabela 2 e a visualização gráfica do resultado está na figura 25

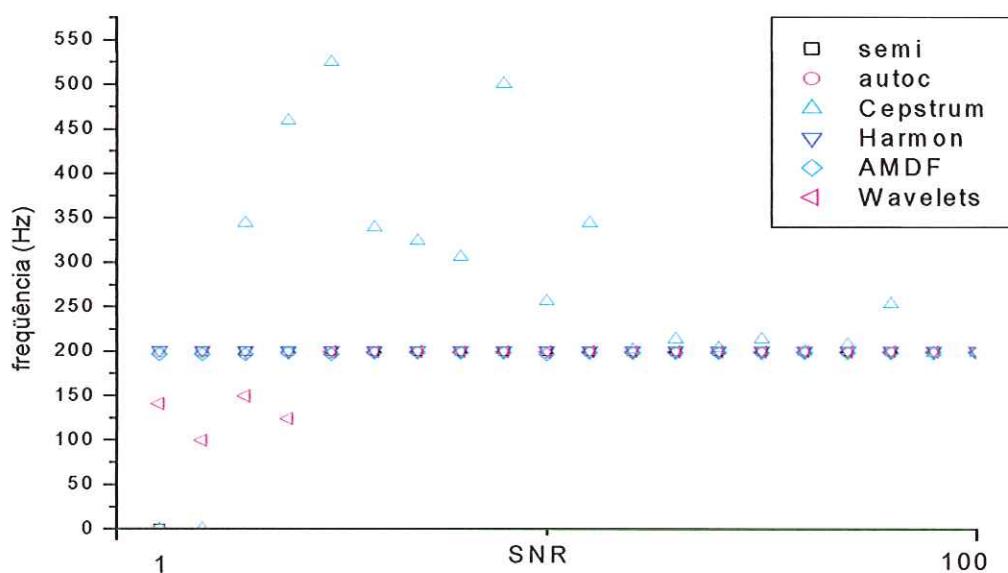


Figura 25: Resultados dos estimadores para variação de SNR.

Observa-se que a SNR afetou mais significativamente os métodos por cepstrum e wavelet, sendo o algoritmo por cepstrum afetado por uma SNR mais alta e de forma mais intensa que o método por wavelet. Os outros métodos não sofreram tanta influencia devido principalmente a envolverem algum processo de filtragem por passa-baixas em sua concepção. Conforme mostrado na tabela avaliou-se apenas valores para SNR maiores ou iguais 1 (um) e o valor máximo igual a 100 (cem) sendo a última análise com a voz sem ruído.

A seguir efetuou-se o mesmo procedimento para o *jitter* e *shimmer* sendo os dados apresentados na tabela 3 e 4 e os gráficos nas figuras 26 e 27 respectivamente.

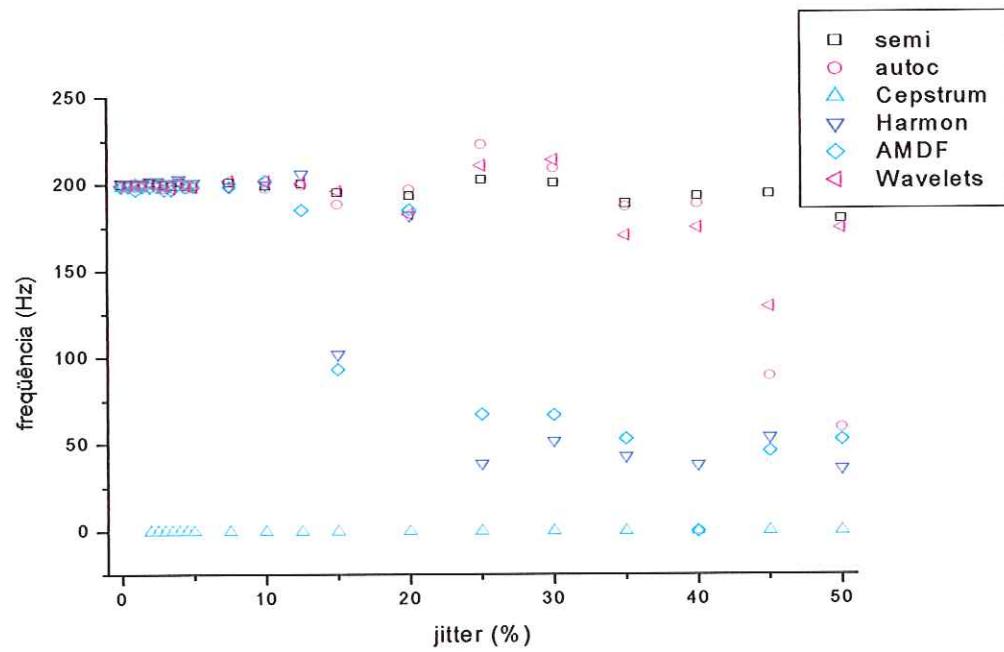


Figura 26: Resultados dos estimadores para variação do *jitter*.

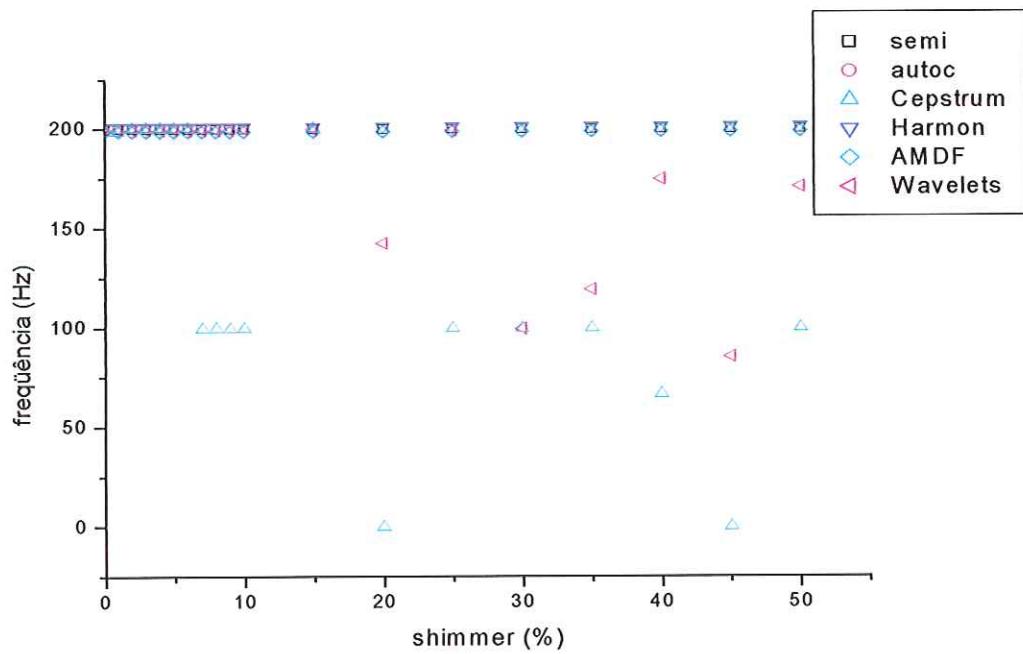


Figura 27: Resultados dos estimadores para variação do *Shimmer*.

A figura 26 mostra claramente que a variação do *jitter* é a grande responsável pela falha na estimativa de Fo por parte dos algoritmos. Fora o método por cepstrum

que falha com pouco *jitter* os demais funcionam bem até 10%. O *shimmer* afetou apenas os algoritmos por cepstrum e por wavelet.

Pelos dados apresentados os algoritmos por autocorrelação, casamento de harmônicas e AMDF comportaram-se de forma satisfatória, estimando os valores para todas as vozes, mesmo que contento um erro. Realizaram-se então os testes com sinais de voz natural, sendo inicialmente com vozes normais e posteriormente com vozes patológicas.

As tabelas 5, 6 e 7 no Apêndice A apresentam os resultados encontrados para a estimativa de Fo, respectivamente para vozes normais, patológicas masculinas e patológicas femininas. Também nestas tabelas encontram-se os valores da diferença normalizada entre os valores estimados pelos algoritmos e o valor padrão. Esta diferença normalizada pode ser visualizada como porcentagem, e é calculada como o módulo da subtração entre o valor estimado subtraído do valor padrão, sendo o resultado dividido pelo valor padrão. Deste modo obtém-se valores normalizados e que podem ser comparados entre si para determinar a eficiência de cada algoritmo na estimação de Fo.

Este procedimento é apenas um artifício para uma melhor visualização gráfica dos resultados. As figuras 28, 29 e 30 apresentam respectivamente os valores desta diferença normalizada para vozes normais, vozes patológicas masculinas e vozes patológicas femininas. Esses gráficos apresentam uma visualização da região próxima de zero, o que faz alguns pontos não serem apresentados. O objetivo inicial é determinar o algoritmo mais robusto, e pela tabelas 5, 6 e 7 este algoritmo é o método por autocorrelação.

Os gráficos apresentados visam mostrar esta melhor robustez em relação aos outros algoritmos, melhorando a resolução dos mesmos em torno dos pontos representativos dos valores calculados para o algoritmo por autocorrelação.

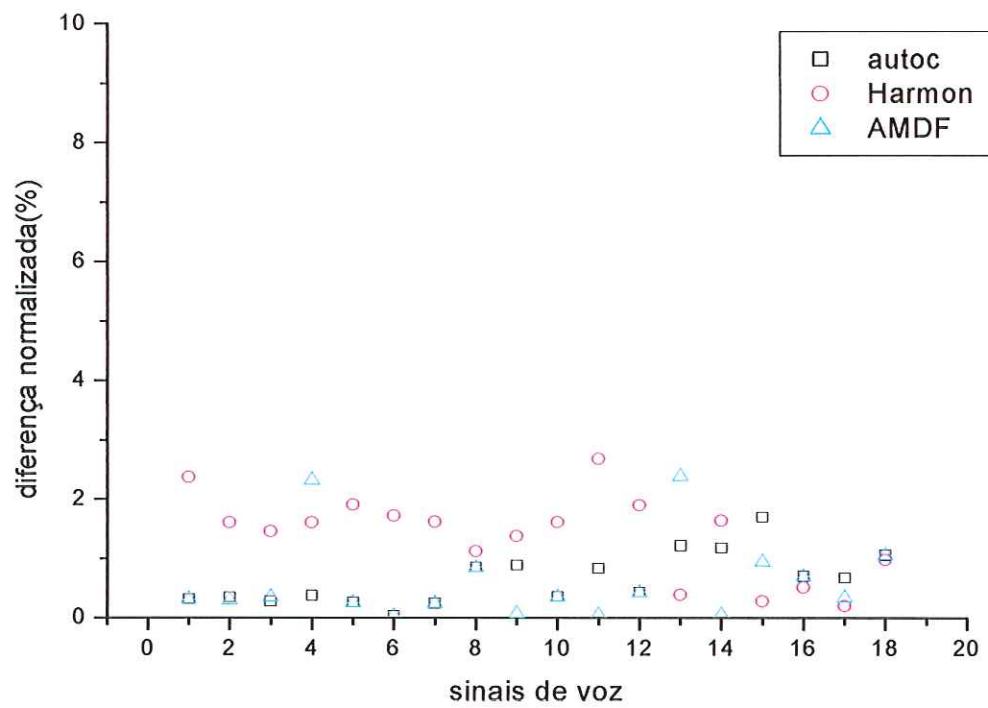


Figura 28: Resultados normalizados dos estimadores para vozes normais.

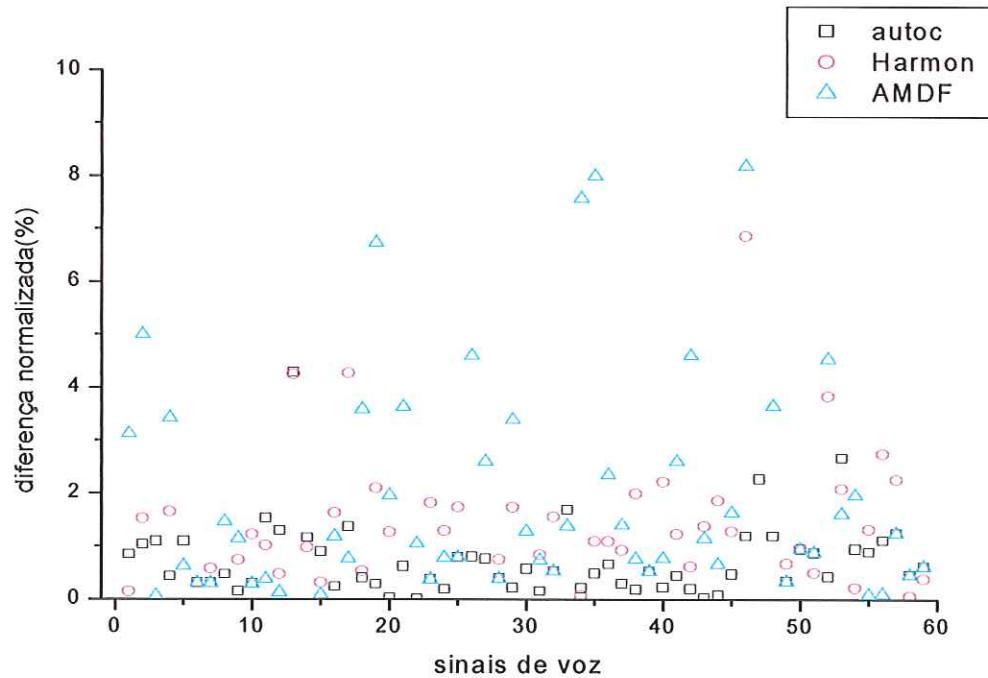


Figura 29: Resultados normalizados dos estimadores para vozes patológicas masculinas.

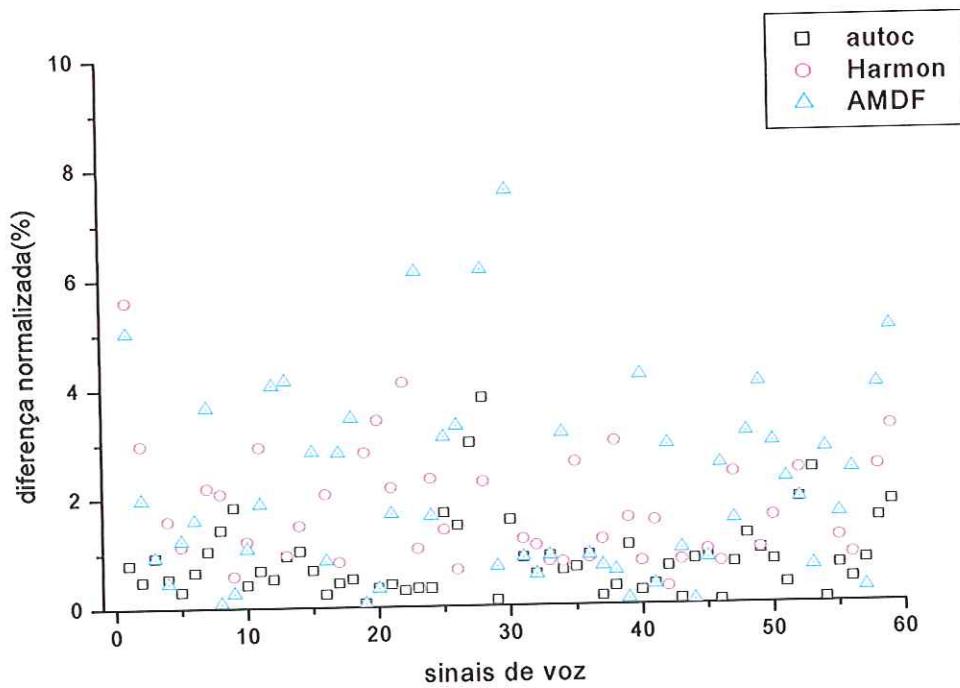


Figura 30: Resultados normalizados dos estimadores para vozes patológicas femininas.

O resultado obtido com vozes normais demonstra concordar com o esperado. Os valores encontrados para estas vozes está muito próximo do determinado pelo método semi-automático. Isto se deve ao fato de que para vozes com pequenas variações(vozes normais) de *jitter* e *shimmer* os algoritmos funcionam melhor que em vozes com grandes variações(vozes patológicas) de *jitter* e *shimmer*. Para as vozes patológicas o algoritmo por autocorrelação se mostra bem mais robusto e por isto este algoritmo será utilizado pelos determinadores.

4.2 – Testes com determinadores

Primeiramente estipula-se que a freqüência fundamental calculada pelos algoritmos determinadores dever ser bem próxima da freqüência calculada para o padrão encontrado pelo método semi-automático. Em uma primeira etapa utilizaram-se sinais de voz sintetizados. Estes testes iniciais servem para encontrar os melhores algoritmos determinadores, antes de testá-los com vozes naturais.

Os testes seguem a mesma seqüência usada para os estimadores. Assim, analisa-se primeiramente a faixa de freqüência. Os dados obtidos encontram-se na tabela 8, e o

respectivo gráfico na figura 31. O valor igual a zero significa que o determinador falhou não sendo possível encontrar períodos que determinassem uma média. Estes erros referem-se a implementação do algoritmo adotado que em alguns casos gera uma divisão por zero.

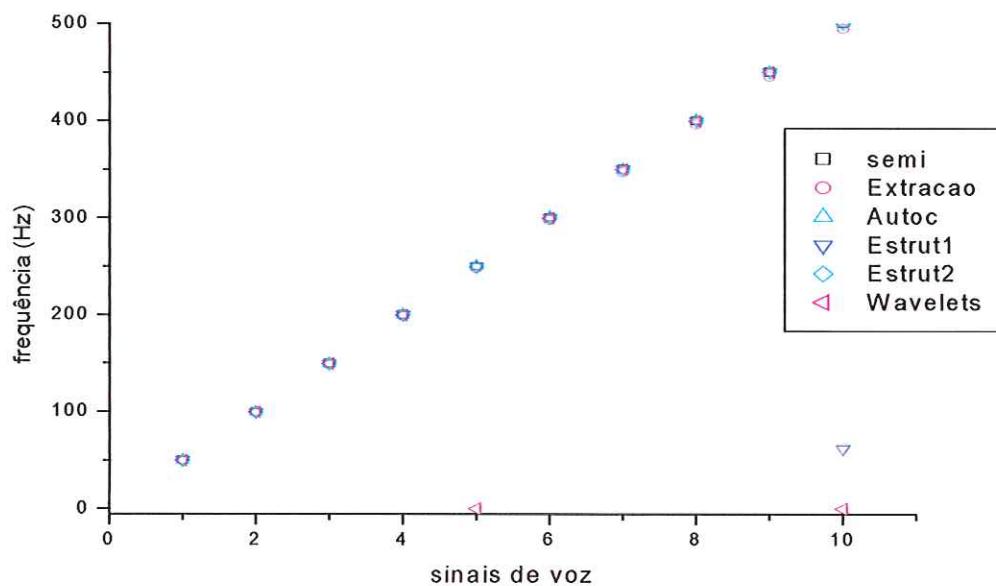


Figura 31: Resultados dos determinadores para variação de Fo nas vozes sintetizadas.

Verifica-se que os algoritmos por Wavelet e Estrutura temporal 1, falharam para vozes com Fo igual a 500 Hz, e o algoritmo de wavelet também falhou para fo de 250 Hz. Estes algoritmos não apresentam as qualidades procuradas por este trabalho, sendo assim posteriormente descartados. A seguir testou-se a influência do ruído nos algoritmos determinadores, com os dados armazenados na tabela 9 e seu gráfico na figura 32.

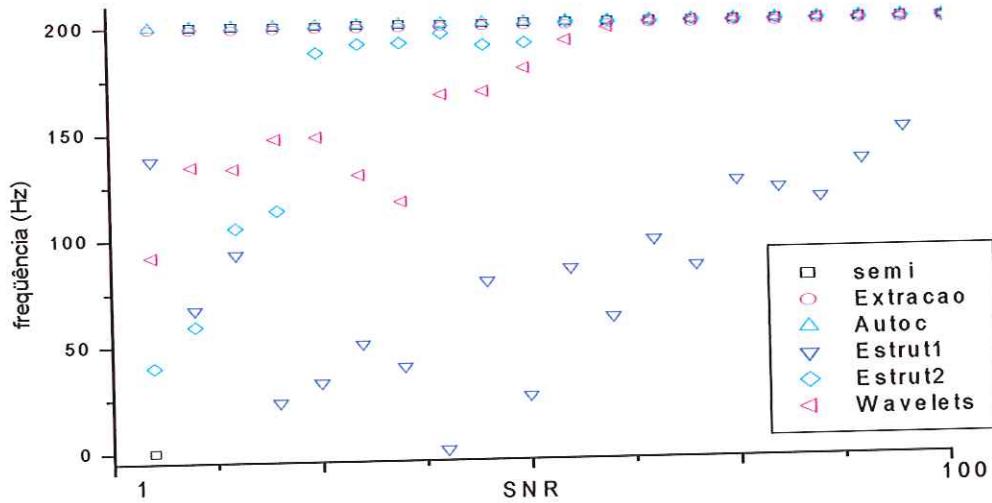


Figura 32: Resultados dos estimadores para variação de SNR.

Novamente os algoritmos de Wavelet e Estrutura temporal tiveram um desempenho inferior aos outros algoritmos, demonstrando novamente sua inadequação para o uso com vozes patológicas. Em vozes patológicas o nível de energia do ruído tende a ser maior que em vozes normais.

Finalmente avaliou-se o *jitter* e *shimmer*. Com os dados das tabelas 10 e 11, confeccionou-se respectivamente os gráficos das figuras 33 e 34.

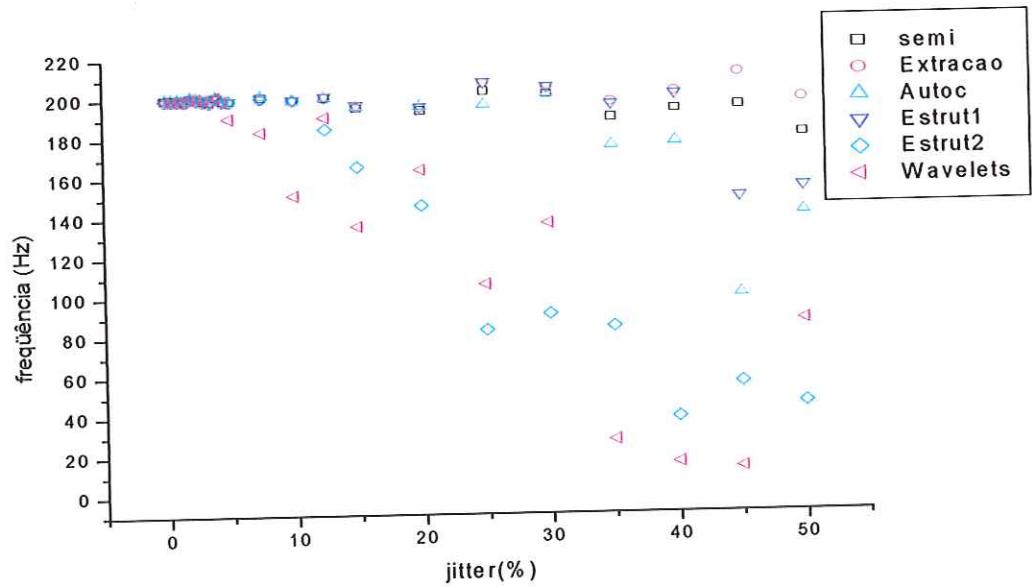


Figura 33: Resultados dos estimadores para variação do *jitter*.

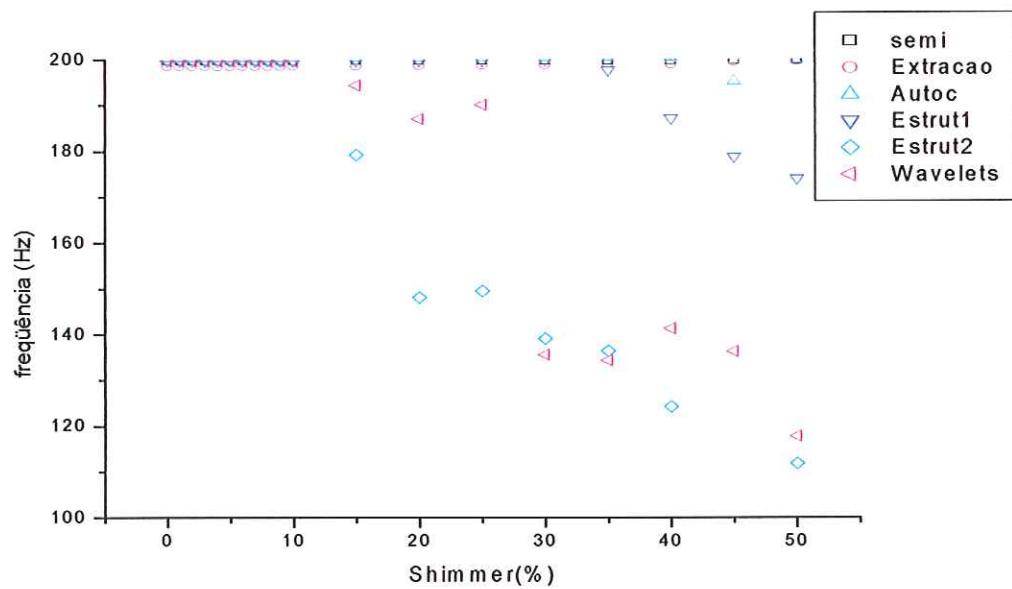


Figura 34: Resultados dos estimadores para variação do *Shimmer*.

Observando os gráficos das figuras 33 e 34 verifica-se que os algoritmos por Wavelets e Estrutura temporal novamente falharam com mais intensidade que os algoritmos por Autocorrelação e Extração de Harmônicas. Assim, apenas os algoritmos por autocorrelação e Extração de Harmônicas serão testados com vozes naturais.

Para as vozes naturais também se verifica a robustez e precisão do valor calculado para Fo. Assim, comparam-se os valores calculados pelos algoritmos com o valor padrão determinado pelo método semi-automático.

A figura 35 apresenta os resultados para vozes normais, a figura 36 para vozes patológicas masculinas e a figura 37 as patológicas femininas. Novamente os resultados são apresentados graficamente na forma de diferenças normalizadas.

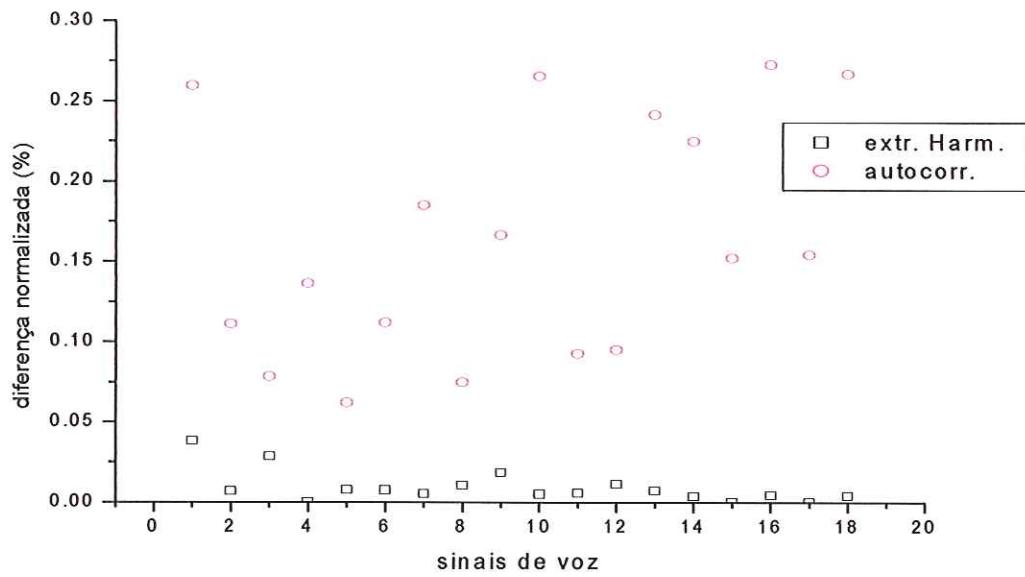


Figura 35: Resultados dos determinadores para vozes normais.

Conforme a figura 35, a precisão dos algoritmos por autocorrelação e extração de harmônicas é bastante satisfatória para vozes normais. Este resultado se deve principalmente à pouca variabilidade no tempo das vozes utilizadas, o que é uma característica das vozes normais. O algoritmo por extração de harmônicas se mostrou mais preciso neste teste, e em alguns casos as médias foram idênticas, pois temos diferenças iguais a zero. Tal teste ressalta o motivo da escolha destes algoritmos após os testes anteriores com vozes sintetizadas.

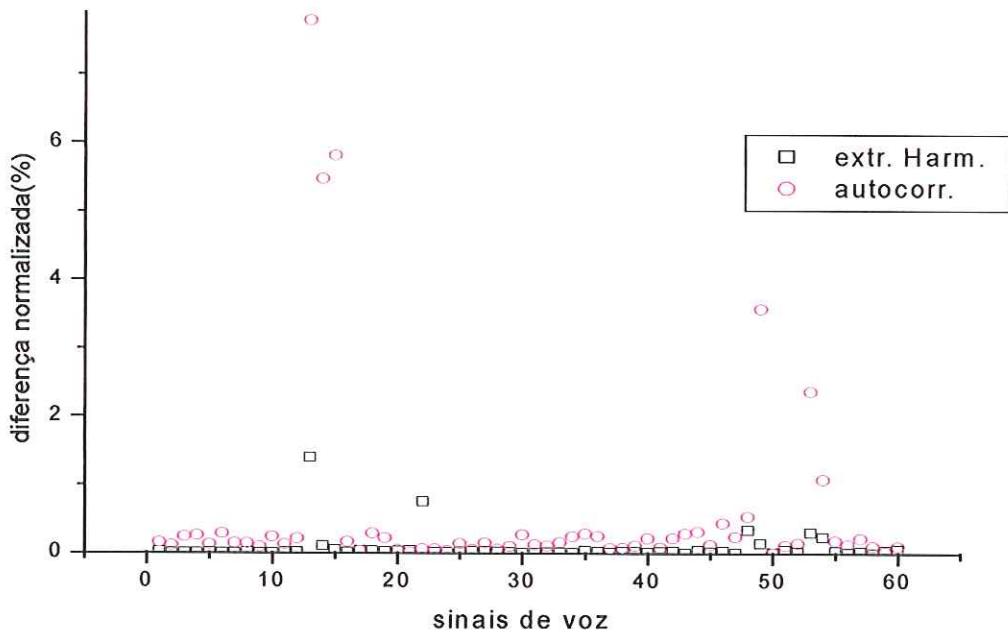


Figura 36: Resultados dos determinadores para vozes patológicas masculinas.

Apesar da baixa resolução, ocasionada por pontos dispersos, o gráfico da figura 36 apresenta novamente o algoritmo por extração de harmônicas como mais preciso que um algoritmo por autocorrelação. Vale salientar que uma boa aproximação da média significa uma tendência a acertar o valor de cada período com maior precisão, assim o algoritmo por extração de harmônicas se mostrou bastante favorável. Os sinais de voz utilizados são todos masculinos e apresentam alguma patologia diagnosticada e registrada no banco de vozes. As vozes patológicas tendem a ter uma variabilidade maior no período (*jitter*) e na amplitude (*shimmer*), além de algumas patologias terem como característica um aumento acentuado no ruído. Estes fatores fizeram com que a diferença normalizada aumentasse, chegando em alguns casos a valores de 7%. Para o algoritmo por extração de harmônicas o maior valor foi 1,4 % sendo a maioria dos valores próximos de zero.

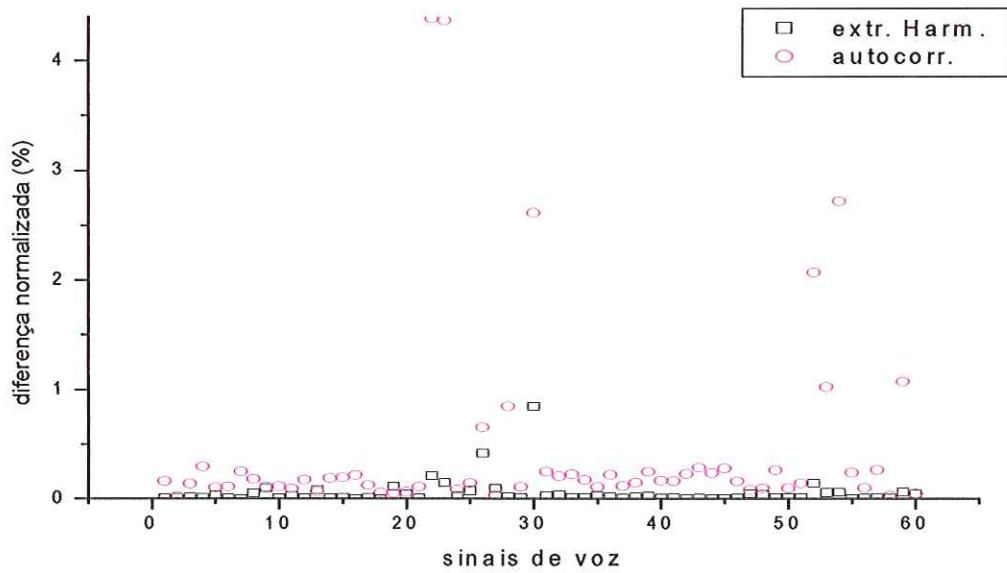


Figura 37: Resultados dos determinadores para vozes patológicas femininas.

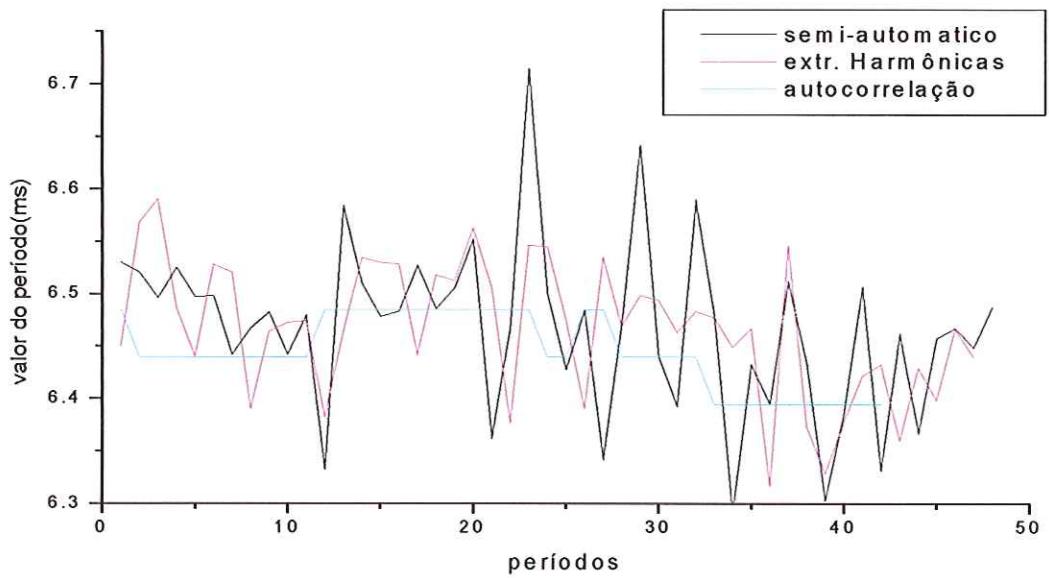


Figura 38: Seqüência de períodos de uma voz patológica.

Pelo gráfico da figura 38 observa-se que os algoritmos identificam os períodos próximos aos valores padrões, mas não acompanham a variação destes padrões na mesma intensidade. Como a diferença dos valores em sua maioria está no mesmo nível de grandeza que o erro por amostragem do sinal de voz ($45 \mu\text{s}$), tal valor encontrado tem um grau de incerteza muito grande.

No gráfico da figura 39 pode-se encontrar o valor dos períodos de um trecho de voz normal. Verifica-se que os valores estão mais próximos, devido a maior estabilidade da voz normal.

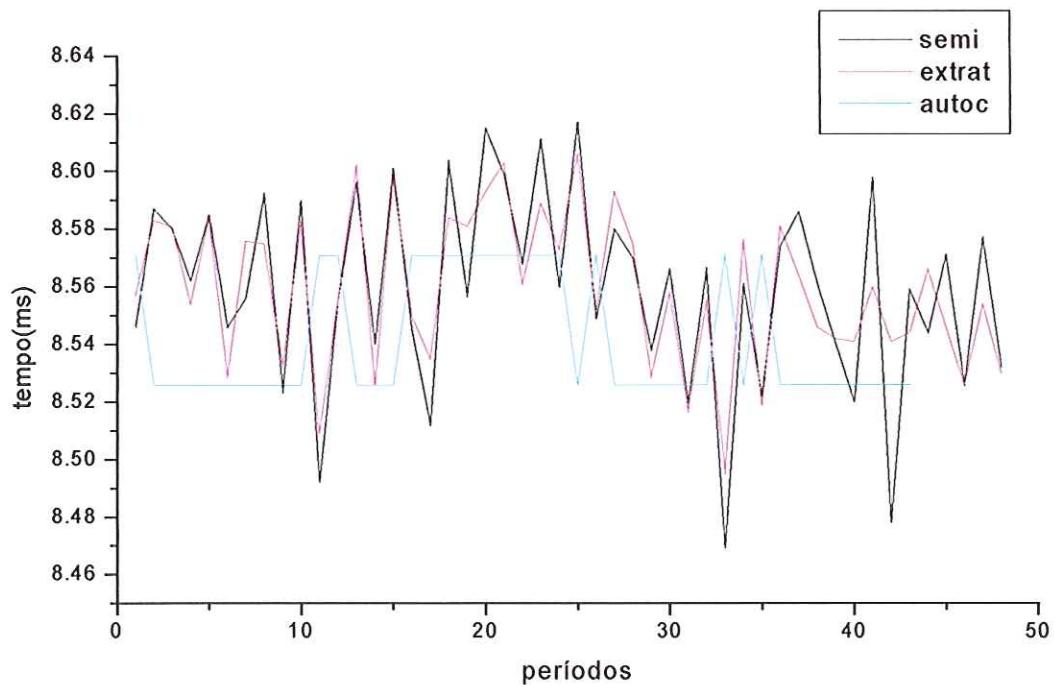


Figura 39: Seqüência de períodos de uma voz normal.

Dos testes realizados observou-se que a referência utilizada para delimitar os períodos (cruzamento por zero, pico, vale, etc.), influencia o valor dos períodos. Esta constatação explica a diferença entre alguns algoritmos, devido ao ponto de referência adotado.

Seguindo a referência indicada pela figura 40, efetuou-se a marcação semi-automática para o sinal de voz já utilizado no gráfico 38, determinando seus períodos para as referências dos limites destes. Utilizando os pontos determinados estabeleceu-se uma tabela com os períodos para marcações com cruzamento por zero em sentido ascendente (XA) e descendente (XD), em picos (P) e vales (V), sendo estes os de maior intensidade no período. Elaborou-se a tabela 16 com os valores dos períodos e os resultados são apresentados visualmente no gráfico da figura 41.

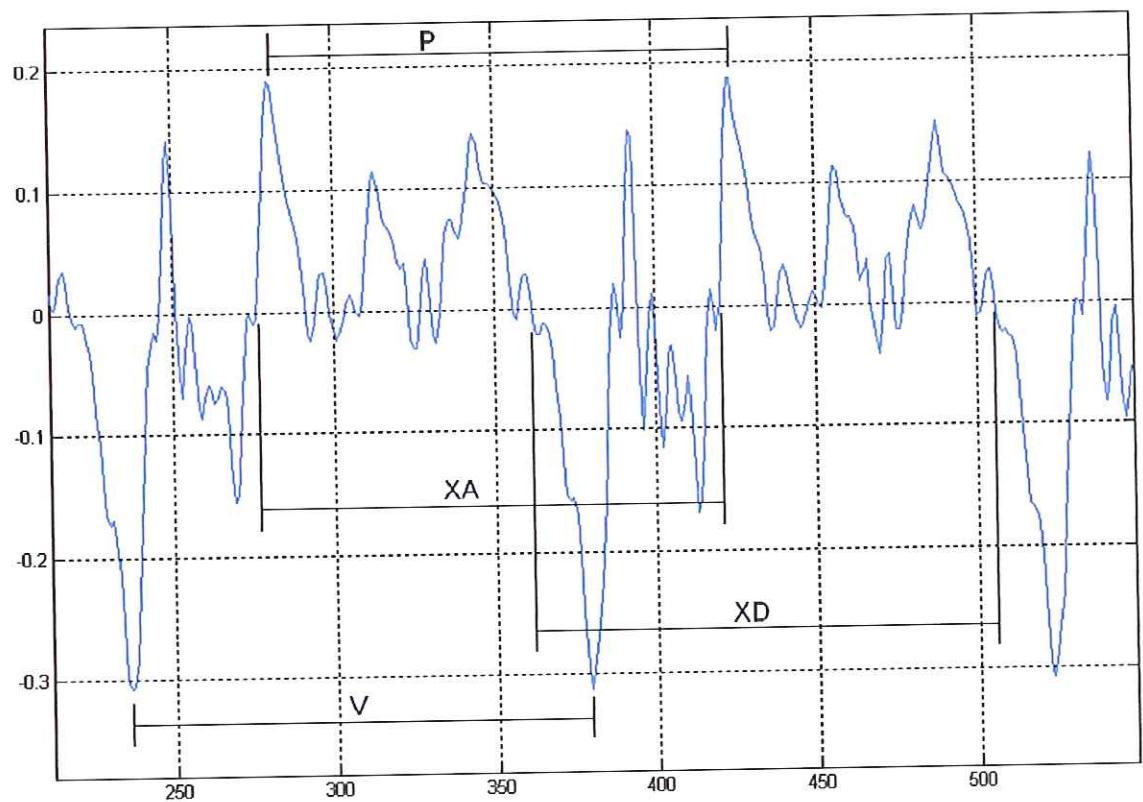


Figura 40: Pontos delimitadores para marcação dos períodos.

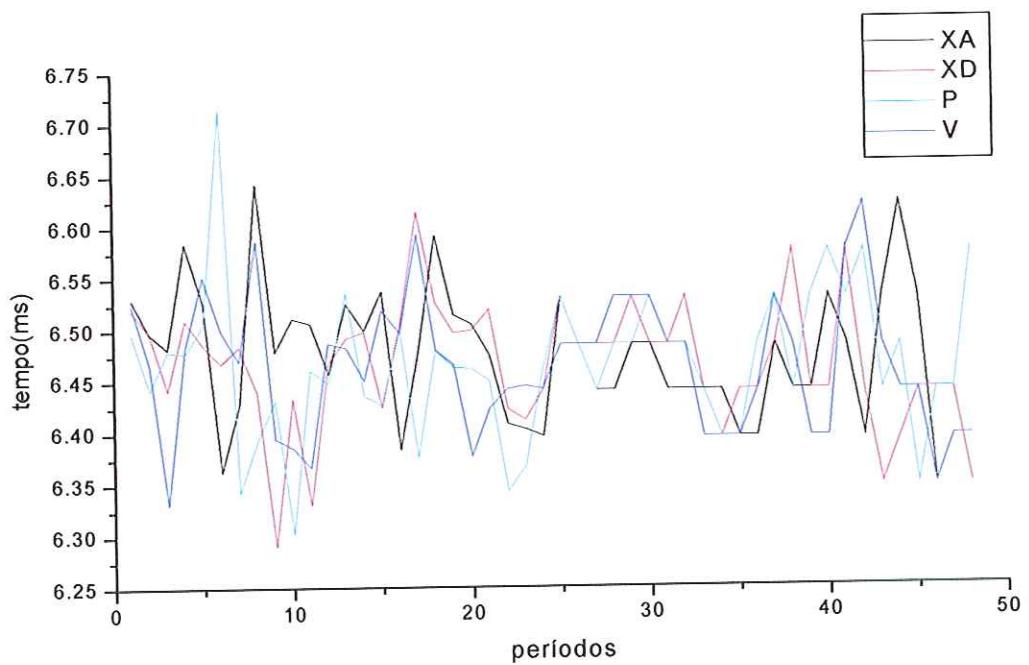


Figura 41: Períodos para diferentes delimitadores.

5 - Conclusão

Dos algoritmos determinadores analisados, os algoritmos por extração de harmônicas e por autocorrelação demonstraram-se os mais promissores no quesito robustez, e apresentando margem para melhorias na precisão. Este fato deve ser bem estudado a fim de permitir uma melhora e não adicionar mais erros.

Os algoritmos por estrutura temporal são muito afetados pelo tamanho do sinal de voz bem como para sinais com variações mais acentuadas no período e principalmente na amplitude, descartando-os como opção para o uso em determinadores utilizados em vozes patológicas, onde as variações citadas normalmente ocorrem.

Apesar do algoritmo por wavelet apresentar uma boa margem de acertos dentro do número de vozes utilizadas, o mesmo necessita de melhorias principalmente na identificação dos picos e no pós-processamento para melhorar a robustez.

Os estimadores testados foram e serão utilizados para este trabalho apenas como instrumentos de ajustes para os determinadores, sendo o algoritmo de autocorrelação o mais robusto dos testados. Este fato demonstra que o uso de semelhança do sinal em contrapartida com um evento(cruzamento por zero ou um máximo local) pode ser uma boa escolha quando se deseja um algoritmo robusto.

Quando se avaliaram as vozes sintetizadas, notou-se que o grande entrave para os algoritmos foi o aumento do jitter. Este fator causou um aumento do erro nos algoritmos de determinação e estimação mostrando que o mesmo afeta muito a robustez. Essas considerações ressaltam que o jitter deve ser uma preocupação constante no desenvolvimento de algoritmos para detecção e estimação de pitch para vozes patológicas onde o jitter é alto em relação a uma voz normal.

Quando se analisou o ruído para vozes sintetizadas, verificou-se que o mesmo afetou mais seriamente alguns algoritmos, sendo estes descartados, enquanto que outros algoritmos tiveram apenas uma piora na precisão, mantendo a robustez. Este fato implica no uso de filtros passa-baixa ou passa-faixa para os algoritmos melhorarem a

precisão. Observou-se que os algoritmos que tinham estas filtragens em seus procedimentos ficaram imunes ao aumento do ruído, considerando um nível de ruído real.

A filtragem dos sinais por um filtro passa-baixas foi necessária para a obtenção da melhora dos algoritmos determinadores e estimadores, o que eliminou a interferência de ruídos e harmônicas de mais alta ordem no sinal de voz facilitando o uso dos algoritmos.

Quanto às vozes, não se identificaram características das vozes femininas ou masculinas que permitissem uma melhor taxa de acertos por parte dos algoritmos testados. Isto foi comprovado pelas vozes sintetizadas.

Observou-se que, como esperado, as vozes normais permitiam uma melhor exatidão nas medidas.

A diferença entre os valores dos períodos padrões e os determinados por extração de harmônicas, é resultado da filtragem necessária para o funcionamento do algoritmo. Ocorre uma mudança no delimitador e este apesar de encontrar valores próximos dos padrões, carrega uma diferença. Outro fator observado é que a filtragem também acarreta na diminuição do jitter para os períodos determinados.

6 – Bibliografia

- ANDERSON, M. D. **Pitch determination of speech signals.** Dissertação (mestrado) - Massachusetts Institute of Technology, 1986
- ANDRADE, L. M. O., VIEIRA, J. M., RAZERA, D. E., GUERRA, A. C., PEREIRA, J. C. **Medidas de Perturbação da voz: um novo enfoque.** Revista Fonoaudiologia-Brasil, Brasília-DF, v.2, n.2, p.39-46, 2002.
- BORGET, B. P., HEALY, M. J. R., TUKEY, J. W. - apud OPPENHEIM, A. V.; SCHAFER, R. W. - **Discrete-Time Signal Processing**, New Jersey, Prentice Hall, 1989.
- DAVIS, S.B. **Acoustic Characteristics of Normal and Pathological Voices.** In Lass, N.J. (ED) Speech and language: Advances in Basic Research and Practice, vol.1, New York Academic Press, 271-335, 1979
- FUKS, L. SUNDBERG, J. **Using respiratory inductive plethysmography for monitoring professional reed instrument performance.** Medical Problems of Performing artists. Hanley & Belfus, Inc., Philadelphia, PA, 1999.
- HANSELMAN, D., LITTLEFIELD, B. **Mastering Matlab 5: A comprehensive Tutorial and Reference.** New Jersey, Prentice Hall, 1998.
- HAYKIN, S. **Sinais e sistemas.** Trad. de José Carlos Barbosa dos Santos. Porto Alegre, Bookman, 2001.
- HESS, W. **Pitch Determination of Speech Signals,** Springer, Berlin, 1983

HIRANO, M. **Clinical Examination of Voice**, Springer, New York, 1981

JAWERTH, B., SWELDENS, W. **An Overview of Wavelet based Multiresolution Analysis**, SIAM Rev. , 1994, 36:377-412.

KADAMBE, S., BOURDEAUX-BARTELS, G. F. **Aplication of the Wavelet transform for pitch detection of speech signals**. IEEE Trans. Inf. Theory, 1992,38(2):917-24.

LIEBERMAN, P. **Some acoustics Measures of the Fundamental Periodicity of Normal and Pathologic Larynges**, Journal of the Acoustics Society of America, 35, 344-353, 1963.

MALLAT, S., ZHONG, S. **Characterization of signals from multiscale edges**. IEEE Trans. Patt. Rec. Mach. Int., 14(7):710–732, 1992.

MANFREDI, C., D'ANIELLO, M., BRUSCAGLIONI, P., ISMAELLI, A. **A comparative analysis of fundamental frequency estimation methods with application to pathological voices**. Medical Engineering & Physics.,2000, 22, 135-147.

MARKEL, J.D. **The SIFT Algorithm for Fundamental Frequency Estimation**, IEEE Transaction on Audio and Electroacoustics, Vol. AU-21, n° 5, pp.367-377, December, 1972.

MONTAGNOLI, A. N. **Análise residual do sinal de voz**. Dissertação(mestrado) – Escola de Engenharia de São Carlos-Universidade de São Paulo. São Carlos, 1998

NOLL, A. M. **Cepstrum Pitch Determination**, J. Acoust. Soc. Amer., vol. 41, no. 2, pp. 293-309, 1970.

OPPENHEIM, A. V.; SCHAFER, R. W. **Discrete-Time Signal Processing**, New Jersey, Prentice Hall, 1989.

RABINER, L. R., SCHAFER, R. W. **Digital processing of speech signals.** New Jersey, Prentice Hall, 1978.

ROSA, M.O. **Análise Acústica da Voz para Pré-Diagnóstico de Patologias da Laringe.** Dissertação (mestrado) – Escola de Engenharia de São Carlos -Universidade de São Paulo. São Carlos, 1998

WENDT, C., PETROPULU, A.P. **Detection and Speech Segmentation Using the Discrete Wavelet Transform,** Proc. International Symposium on Circuits and Systems, ISCAS, Atlanta, 1996, v.2 , p.45-8.

APÊNDICE A – Tabelas das análise realizadas para os algoritmos

Tabela 1: Resultados dos estimadores para variação de Fo nas vozes sintetizadas.

Sinal de Voz	Semi Automático	Autoc.	Cepstrum	Casament Harmôn	AMDF	Wavelets
vozfo050	49.73	49.89	49.66	51.00	49.66	49.77
vozfo100	99.75	99.77	99.77	101.00	99.32	99.62
vozfo150	149.77	148.99	150.00	151.00	148.99	149.66
vozfo200	199.80	198.65	198.65	201.00	198.65	199.77
vozfo250	249.82	247.75	247.75	251.00	247.75	250.00
vozfo300	299.84	294.00	297.97	301.00	294.00	300.47
vozfo350	349.86	344.53	350.00	351.00	344.53	349.63
vozfo400	399.89	393.75	400.91	401.00	393.75	400.05
vozfo450	449.90	441.00	450.00	451.00	441.00	450.00
vozfo500	499.93	490.00	501.14	501.00	490.00	499.51

Tabela 2: Resultados dos estimadores para variação de SNR.

Sinal de Voz	Semi Automático	Autoc.	Cepstrum	Casament Harmôn	AMDF	Wavelets
Voz snr 1	0	198.65	0	201.00	196.88	141.17
Voz snr 2	199.83	198.65	0	201.00	196.88	100.00
Voz snr 3	199.83	196.88	344.53	201.00	196.88	150.00
Voz snr 4	199.78	198.65	459.38	201.00	198.65	124.58
Voz snr 5	199.75	197.76	525.00	201.00	196.88	199.77
Voz snr 6	199.84	198.65	339.23	201.00	198.65	199.55
Voz snr 7	199.80	198.65	324.26	201.00	198.65	200.00
Voz snr 8	199.76	198.65	306.25	201.00	198.65	199.55
Voz snr 9	199.78	198.65	501.14	201.00	198.65	199.77
Voz snr 10	199.80	197.76	256.40	201.00	196.88	199.77
Voz snr 20	199.80	198.65	344.53	201.00	198.65	200.00
Voz snr 30	199.81	198.65	202.29	201.00	198.65	200.00
Voz snr 40	199.81	198.65	214.08	201.00	198.65	200.00
Voz snr 50	199.80	198.65	204.17	201.00	198.65	199.77
Voz snr 60	199.81	198.65	214.08	201.00	198.65	199.77
Voz snr 70	199.81	198.65	200.45	201.00	198.65	200.00
Voz snr 80	199.79	198.65	208.02	201.00	198.65	199.77
Voz snr 90	199.82	198.65	253.45	201.00	198.65	200.00
Voz snr 100	199.79	198.65	198.65	201.00	198.65	199.77
Sem ruido	199.80	198.65	198.65	201.00	198.65	199.77

Tabela 3: Resultados dos estimadores para variação do jitter.

Sinal de Voz	Semi Automático	Autoc.	Cepstrum	Casament Harmôn	AMDF	Wavelets
Jitter 0	199.80	198.65	198.65	201.00	198.65	199.77
Jitter 0.5	199.83	198.65	198.65	201.00	198.65	200.00
Jitter 1	199.93	198.65	200.45	201.00	196.88	199.77
Jitter 1.5	199.59	198.65	198.65	201.00	198.65	200.00
Jitter 2	200.82	198.65	0	202.00	198.65	200.91
Jitter 2.5	200.26	198.65	0	202.00	198.65	199.55
Jitter 3	199.53	197.76	0	201.00	196.88	198.87
Jitter 3.5	198.99	196.88	0	201.00	196.88	198.20
Jitter 4	201.25	199.55	0	203.00	200.45	201.83
Jitter 4.5	199.20	197.76	0	201.00	198.65	199.55
Jitter 5	199.07	198.65	0	201.00	198.65	198.65
Jitter 7.5	200.81	198.65	0	199.00	198.65	202.29
Jitter 10	199.07	197.76	0	200.00	202.29	202.06
Jitter 12.5	200.19	199.55	0	206.00	185.29	200.91
Jitter 15	194.92	188.46	0	102.00	93.43	196.22
Jitter 20	193.12	196.88	0	182.00	185.29	182.88
Jitter 25	202.50	222.73	0	39.00	67.43	210.75
Jitter 30	200.67	209.00	0	52.00	67.02	213.82
Jitter 35	188.59	186.86	0	43.00	53.26	170.36
Jitter 40	192.65	188.46	0	38.00	0	175.00
Jitter 45	194.07	89.27	0	54.00	46.23	129.40
Jitter 50	179.66	59.68	0	36.00	52.75	174.54

Tabela 4: Resultados dos estimadores para variação do shimmer.

Sinal de Voz	Semi	Autoc.	Cepstrum	Casament Harmôn	AMDF	Wavelets
Sem shimmer	199.80	198.65	198.65	201.00	198.65	199.77
Shimmer 1	199.80	198.65	198.65	201.00	198.65	199.77
Shimmer 2	199.80	198.65	200.45	201.00	198.65	199.77
Shimmer 3	199.80	198.65	200.45	201.00	198.65	199.77
Shimmer 4	199.80	198.65	200.45	201.00	198.65	199.55
Shimmer 5	199.80	198.65	200.45	201.00	198.65	199.77
Shimmer 6	199.80	198.65	200.45	201.00	198.65	199.55
Shimmer 7	199.80	198.65	99.77	201.00	198.65	199.77
Shimmer 8	199.80	198.65	99.77	201.00	198.65	199.55
Shimmer 9	199.80	198.65	99.77	201.00	198.65	199.55
Shimmer 10	199.80	198.65	99.77	201.00	198.65	200.23
Shimmer 15	199.80	198.65	200.45	201.00	198.65	200.00
Shimmer 20	199.80	198.65	0	201.00	198.65	142.63
Shimmer 25	199.80	198.65	99.77	201.00	198.65	199.32
Shimmer 30	199.80	198.65	99.77	201.00	198.65	99.62
Shimmer 35	199.80	198.65	99.77	201.00	198.65	119.40
Shimmer 40	199.80	198.65	66.62	201.00	198.65	174.80
Shimmer 45	199.80	198.65	0	201.00	198.65	85.47
Shimmer 50	199.79	198.65	99.77	201.00	198.65	170.93

Tabela 5: Resultados dos estimadores para vozes normais

Voz	Vo g	Semi Fo(std)	Autoc	Autoc. desv.(%)	Harmôn.	Harmôn. desv(%)	AMDF	AMDF desv(%)
Voz41	\a\	130.89 (1.51)	130.47	0.32	134.00	2.38	130.47	0.32
	\e\	143.68 (1.58)	143.18	0.35	146.00	1.61	144.12	0.31
	\i\	139.95 (2.78)	139.56	0.28	142.00	1.46	140.45	0.36
Voz42	\a\	117.11 (1.21)	116.67	0.38	119.00	1.61	119.84	2.33
	\e\	128.54 (1.32)	128.2	0.26	131.00	1.91	128.20	0.26
	\i\	133.69 (1.43)	133.64	0.04	136.00	1.73	133.64	0.04
Voz43	\a\	188.93 (2.00)	188.46	0.25	192.00	1.62	188.46	0.25
	\e\	186.89 (1.93)	185.29	0.86	189.00	1.13	185.29	0.86
	\i\	216.01(1.28)	214.08	0.89	219.00	1.38	216.18	0.08
Voz44	\a\	195.83 (2.00)	195.13	0.36	199.00	1.62	195.13	0.36
	\e\	172.37 (2.05)	170.93	0.84	177.00	2.69	172.27	0.06
	\i\	178.60 (2.37)	177.82	0.44	182.00	1.90	177.82	0.44
Voz45	\a\	268.94 (3.89)	265.66	1.22	270.00	0.39	262.50	2.39
	\e\	275.47 (4.65)	272.22	1.18	280.00	1.64	275.63	0.06
	\i\	295.16 (3.67)	290.13	1.70	296.00	0.28	297.97	0.95
Voz46	\a\	219.86 (1.29)	218.32	0.70	221.00	0.52	218.32	0.70
	\e\	226.53 (1.72)	225.00	0.68	227.00	0.21	227.32	0.35
	\i\	239.64 (1.60)	237.10	1.06	242.00	0.98	237.10	1.06

Tabela 6: Valores encontrados pelos estimadores para vozes masculina patológica

Voz	V og	Semi Fo(std)	Autoc	Autoc. desv(%)	Harmôn.	Harmôn. desv(%)	AMDF	AMDF desv(%)
Voz1	\a\	261.62(4.63)	259.41	0.84	262.00	0.15	253.45	3.12
	\e\	256.10(4.76)	253.45	1.03	260.00	1.52	268.90	5.00
	\i\	259.23(6.43)	256.40	1.09	89.00	65.67	259.41	0.07
Voz2	\a\	165.27(3.67)	164.55	0.44	168.00	1.65	170.93	3.42
	\e\	190.54(2.38)	188.46	1.09	65.00	65.89	191.74	0.63
	\i\	184.32(1.84)	183.75	0.31	63.00	65.82	183.75	0.31

Voz	Vog	Semi Fo(std)	Autoc	Autoc. desv.(%)	Harmôn.	Harmôn. desv(%)	AMDF	AMDF desv(%)
Voz3	\a\	225.69(1.40)	225.00	0.31	227.00	0.58	225.00	0.31
	\e\	223.78(2.26)	222.73	0.47	113.00	49.50	220.50	1.47
	\i\	225.33(1.97)	225.00	0.15	227.00	0.74	222.73	1.15
Voz4	\a\	243.02(2.37)	242.31	0.29	246.00	1.23	242.31	0.29
	\e\	257.38(2.52)	253.45	1.53	260.00	1.02	256.40	0.38
	\i\	259.76(3.11)	256.40	1.29	261.00	0.48	259.41	0.13
Voz5	\a\	168.15(23.94)	160.95	4.28	161.00	4.25	80.77	51.97
	\e\	174.29(10.47)	172.27	1.16	176.00	0.98	84.16	51.71
	\i\	179.43(7.30)	177.82	0.90	180.00	0.32	179.27	0.09
Voz6	\a\	212.53(1.88)	212.02	0.24	216.00	1.63	210.00	1.19
	\e\	237.84(5.93)	234.57	1.37	248.00	4.27	239.67	0.77
	\i\	243.30(4.92)	242.31	0.41	242.00	0.53	234.57	3.59
Voz7	\a\	139.08(2.15)	138.68	0.29	142.00	2.10	129.71	6.74
	\e\	144.17(1.53)	144.12	0.03	146.00	1.27	147.00	1.96
	\i\	173.35(2.44)	172.27	0.62	58.00	66.54	167.05	3.63
Voz8	\a\	117.28(0.63)	117.29	0.01	236.00	101.23	116.05	1.05
	\e\	115.89(0.68)	115.45	0.38	118.00	1.82	115.45	0.38
	\i\	132.29(1.34)	132.04	0.19	134.00	1.29	131.25	0.79
Voz9	\a\	185.22(5.28)	183.75	0.79	182.00	1.74	183.75	0.79
	\e\	191.63(3.38)	190.09	0.80	97.00	49.38	200.45	4.60
	\i\	209.63(4.35)	208.02	0.77	106.00	49.43	204.17	2.60
Voz10	\a\	175.68(3.36)	175.00	0.39	177.00	0.75	175.00	0.39
	\e\	181.14(3.39)	180.74	0.22	178.00	1.73	175.00	3.39
	\i\	203.46(2.61)	202.29	0.58	104.00	48.88	206.07	1.28
Voz11	\a\	129.91(1.45)	129.71	0.15	131.00	0.84	128.95	0.74
	\e\	131.95(1.68)	131.25	0.53	134.00	1.55	131.25	0.53
	\i\	138.02(5.33)	135.69	1.69	69.00	50.01	136.11	1.38
Voz12	\a\	160.12(2.76)	159.78	0.21	160.00	0.07	147.99	7.58
	\e\	158.27(3.26)	157.50	0.49	160.00	1.09	170.93	8.00
	\i\	217.62(2.73)	216.18	0.66	220.00	1.09	222.73	2.35

Voz	V og	Semi Fo(std)	Autoc	Autoc. desv(%)	Harmôn.	Harmôn. desv(%)	AMDF	AMDF desv(%)
Voz13	\a\	122.86(0.89)	122.50	0.29	124.00	0.93	124.58	1.40
	\e\	128.44(1.29)	128.20	0.19	131.00	1.99	127.46	0.76
	\i\	133.54(0.93)	132.83	0.53	46.00	65.55	132.83	0.53
Voz14	\a\	110.55(0.76)	110.80	0.23	113.00	2.22	109.70	0.77
	\e\	109.65(0.81)	109.16	0.45	111.00	1.23	112.50	2.60
	\i\	113.30(1.19)	113.08	0.19	114.00	0.62	108.09	4.60
Voz15	\a\	128.23(1.27)	128.20	0.02	130.00	1.38	129.71	1.15
	\e\	130.57(0.83)	130.47	0.08	133.00	1.86	129.71	0.66
	\i\	130.33(0.76)	129.71	0.48	132.00	1.28	128.20	1.63
Voz16	\a\	168.44(11.30)	166.42	1.20	180.00	6.86	182.23	8.19
	\e\	169.64(14.82)	165.79	2.27	94.00	44.59	188.46	11.09
	\i\	208.57(20.17)	206.07	1.20	53.00	74.59	216.18	3.65
Voz17	\a\	223.49(8.64)	222.73	0.34	225.00	0.68	222.73	0.34
	\e\	244.63(1.85)	242.31	0.95	247.00	0.97	242.31	0.95
	\i\	241.79(5.01)	239.67	0.88	243.00	0.50	239.67	0.88
Voz18	\a\	238.12(9.77)	237.10	0.43	229.00	3.83	227.32	4.54
	\e\	240.98(26.44)	234.57	2.66	246.00	2.08	237.10	1.61
	\i\	229.50(12.21)	227.32	0.95	230.00	0.22	225.00	1.96
Voz19	\a\	218.13(5.27)	216.18	0.89	221.00	1.32	218.32	0.09
	\e\	268.64(6.57)	265.66	1.11	276.00	2.74	268.90	0.10
	\i\	282.62(5.21)	279.11	1.24	289.00	2.26	279.11	1.24
Voz20	\a\	154.91(1.60)	154.20	0.46	155.00	0.06	154.20	0.46
	\e\	173.33(1.36)	172.27	0.61	174.00	0.39	172.27	0.61
	\i\	205.42(1.41)	204.17	0.61	208.00	1.26	3150.0	1433.44

Tabela 7: Valores encontrados pelos estimadores para vozes femininas patológicas

Voz	V og	Semi Fo(std)	Autoc	Autoc. desv(%)	Harmôn.	Harmôn. desv(%)	AMDF	AMDF desv(%)
Voz21	\a\	154.35(5.37)	153.13	0.84	163.00	0.15	162.13	3.12
	\e\	178.68(2.37)	177.82	1.03	184.00	1.52	182.23	5.00
	\i\	207.99(2.81)	206.07	1.09	107.00	65.67	206.07	0.07

Voz	Vog	Semi Fo(std)	Autoc	Autoc. desv(%)	Harmôn.	Harmôn. desv(%)	AMDF	AMDF desv(%)
Voz22	\a\	219.50(2.66)	218.32	0.44	223.00	1.65	220.50	3.42
	\e\	210.62(1.38)	210.00	1.09	213.00	65.89	208.02	0.63
	\i\	213.42(1.41)	212.02	0.31	108.00	65.82	210.00	0.31
Voz23	\a\	337.59(9.16)	334.09	0.31	345.00	0.58	350.00	0.31
	\e\	338.92(6.26)	334.09	0.47	346.00	49.50	339.23	1.47
	\i\	350.96(11.30)	344.53	0.15	353.00	0.74	350.00	1.15
Voz24	\a\	164.01(1.90)	163.33	0.29	166.00	1.23	165.79	0.29
	\e\	160.29(2.83)	159.21	1.53	165.00	1.02	163.33	0.38
	\i\	164.18(2.52)	163.33	1.29	82.00	0.48	157.50	0.13
Voz25	\a\	271.42(9.03)	268.90	4.28	274.00	4.25	282.69	51.97
	\e\	234.51(5.10)	232.11	1.16	231.00	0.98	459.38	51.71
	\i\	249.42(4.36)	247.75	0.90	123.00	0.32	242.31	0.09
Voz26	\a\	138.14(1.83)	137.81	0.24	141.00	1.63	136.96	1.19
	\e\	135.88(1.79)	135.28	1.37	137.00	4.27	132.04	0.77
	\i\	142.08(1.72)	141.35	0.41	73.00	0.53	147.00	3.59
Voz27	\a\	110.87(2.74)	110.80	0.29	114.00	2.10	110.80	6.74
	\e\	138.28(1.77)	137.81	0.03	143.00	1.27	137.81	1.96
	\i\	152.68(1.85)	152.07	0.62	156.00	66.54	155.28	3.63
Voz28	\a\	110.47(2.68)	110.80	0.01	115.00	101.23	94.23	1.05
	\e\	136.56(2.11)	136.11	0.38	138.00	1.82	128.20	0.38
	\i\	143.65(2.10)	143.18	0.19	147.00	1.29	146.03	0.79
Voz29	\a\	320.47(7.86)	315.00	0.79	316.00	1.74	310.56	0.79
	\e\	344.29(26.33)	339.23	0.80	342.00	49.38	355.65	4.60
	\i\	349.67(21.05)	339.23	0.77	112.00	49.43	310.56	2.60
Voz30	\a\	143.25(19.90)	137.81	0.39	140.00	0.75	134.45	0.39
	\e\	137.09(1.89)	136.96	0.22	71.00	1.73	136.11	3.39
	\i\	162.32(7.98)	159.78	0.58	54.00	48.88	150.00	1.28
Voz31	\a\	234.16(3.78)	232.11	0.15	237.00	0.84	232.11	0.74
	\e\	233.44(1.60)	232.11	0.53	236.00	1.55	232.11	0.53
	\i\	250.00(0.94)	247.75	1.69	252.00	50.01	247.75	1.38

Voz	Vog	Semi Fo(std)	Autoc	Autoc. desv(%)	Harmôn.	Harmôn. desv(%)	AMDF	AMDF desv(%)
Voz32	\a\	203.60(3.07)	202.29	0.21	202.00	0.07	210.00	7.58
	\e\	203.68(4.54)	202.29	0.49	209.00	1.09	386.84	8.00
	\i\	358.91(9.59)	355.65	0.66	362.00	1.09	355.65	2.35
Voz33	\a\	188.75(2.37)	188.46	0.29	191.00	0.93	190.09	1.40
	\e\	210.71(3.89)	210.00	0.19	217.00	1.99	212.02	0.76
	\i\	218.54(6.14)	216.18	0.53	222.00	65.55	218.32	0.53
Voz34	\a\	235.15(2.49)	234.57	0.23	237.00	2.22	245.00	0.77
	\e\	235.40(2.07)	234.57	0.45	239.00	1.23	234.57	2.60
	\i\	255.19(2.04)	253.45	0.19	256.00	0.62	247.75	4.60
Voz35	\a\	204.36(1.48)	204.17	0.02	206.00	1.38	202.29	1.15
	\e\	195.00(1.98)	193.42	0.08	100.00	1.86	195.13	0.66
	\i\	215.88(1.52)	214.08	0.48	218.00	1.28	214.08	1.63
Voz36	\a\	188.56(2.09)	188.46	1.20	190.00	6.86	183.75	8.19
	\e\	199.23(3.27)	197.76	2.27	204.00	44.59	202.29	11.09
	\i\	214.73(4.03)	212.02	1.20	73.00	74.59	208.02	3.65
Voz37	\a\	268.30(2.88)	265.66	0.34	271.00	0.68	279.11	0.34
	\e\	267.75(3.33)	265.66	0.95	272.00	0.97	275.63	0.95
	\i\	283.69(3.28)	282.69	0.88	144.00	0.50	290.13	0.88
Voz38	\a\	197.18(13.28)	193.42	0.43	202.00	3.83	193.42	4.54
	\e\	201.81(10.46)	196.88	2.66	105.00	2.08	200.45	1.61
	\i\	206.22(11.64)	206.07	0.95	71.00	0.22	212.02	1.96
Voz39	\a\	209.48(1.83)	208.02	0.89	212.00	1.32	206.07	0.09
	\e\	226.00(2.03)	225.00	1.11	228.00	2.74	220.50	0.10
	\i\	231.49(1.87)	229.69	1.24	79.00	2.26	232.11	1.24
Voz40	\a\	148.31(4.35)	146.03	0.46	152.00	0.06	154.20	0.46
	\e\	162.77(11.22)	159.78	0.61	168.00	0.39	170.93	0.61
	\i\	166.66(7.77)	164.55	1.27	84.00	49.60	183.75	10.25

Tabela 8: Resultados dos determinadores para variação de Fo nas vozes sintetizadas.

Sinal de Voz	Semi	Extração de harmônicas	Autoc.	Estrut. temp. 1	Estrut. temp. 2	wavelet
vozfo050	49.73	49.62	49.87	49.73	49.73	49.73
vozfo100	99.75	99.33	99.77	99.75	99.76	99.75
vozfo150	149.77	148.88	150.00	149.77	149.77	149.76
vozfo200	199.80	198.64	200.45	199.81	199.79	199.81
vozfo250	249.82	248.12	250.57	249.81	249.82	0
vozfo300	299.84	297.55	301.06	299.93	299.89	299.89
vozfo350	349.86	346.94	350.00	349.89	349.78	350.00
vozfo400	399.89	396.19	400.91	399.87	399.89	399.72
vozfo450	449.90	445.46	450.00	449.81	450.00	450.00
vozfo500	499.93	494.67	501.14	62.17	500.00	0

Tabela 9: Resultados dos determinadores para variação de SNR.

Sinal de Voz	Semi	Extração de harmônicas	Autoc.	Estrut. temp. 1	Estrut. temp. 2	wavelet
Voz snr 1	0	199.08	200.10	137.63	40.22	92.03
Voz snr 2	199.83	198.69	200.41	67.27	59.35	134.42
Voz snr 3	199.83	198.64	200.50	93.20	105.51	133.40
Voz snr 4	199.78	198.59	200.32	23.34	113.60	147.22
Voz snr 5	199.75	198.85	200.45	32.40	187.48	147.92
Voz snr 6	199.84	198.63	200.50	50.46	191.17	129.85
Voz snr 7	199.80	198.65	200.45	39.43	191.43	116.99
Voz snr 8	199.76	198.67	200.45	0	195.69	166.86
Voz snr 9	199.78	198.73	200.45	78.93	189.87	168.15
Voz snr 10	199.80	198.68	200.45	25.06	190.61	178.79
Voz snr 20	199.80	198.60	200.45	84.36	200.04	191.52
Voz snr 30	199.81	198.64	200.45	61.40	200.01	197.23
Voz snr 40	199.81	198.67	200.45	97.38	199.98	199.81
Voz snr 50	199.80	198.63	200.45	85.18	200.30	199.86
Voz snr 60	199.81	198.67	200.45	124.78	200.03	199.81
Voz snr 70	199.81	198.63	200.45	121.08	200.14	199.81
Voz snr 80	199.79	198.64	200.45	116.19	200.18	199.81
Voz snr 90	199.82	198.66	200.45	133.63	199.95	199.81
Voz snr 100	199.79	198.64	200.45	148.20	199.98	199.81
Sem ruido	199.80	198.64	200.45	199.81	199.79	199.81

Tabela 10: Resultados dos determinadores para variação do jitter.

Sinal de Voz	Semi	Extração de harmônicas	Autoc.	Estrut. temp. 1	Estrut. temp. 2	wavelet
Jitter 0	199.80	198.64	200.45	199.81	199.79	199.81
Jitter 0.5	199.83	198.68	200.45	199.85	199.83	199.81
Jitter 1	199.93	198.76	200.59	199.93	199.90	199.95
Jitter 1.5	199.59	198.44	199.79	199.60	199.61	199.63
Jitter 2	200.82	199.68	201.71	200.81	200.88	200.94
Jitter 2.5	200.26	199.12	201.00	200.31	200.23	200.85
Jitter 3	199.53	198.49	199.98	199.61	199.64	199.94
Jitter 3.5	198.99	197.88	199.98	199.07	199.03	199.62
Jitter 4	201.25	200.18	201.77	201.33	201.24	201.54
Jitter 4.5	199.20	198.26	199.45	199.33	199.28	199.42
Jitter 5	199.07	198.15	199.41	199.23	199.31	190.43
Jitter 7.5	200.81	199.64	201.83	200.87	200.61	183.43
Jitter 10	199.07	198.60	199.20	199.67	199.56	151.28
Jitter 12.5	200.19	199.98	200.59	201.02	184.53	190.49
Jitter 15	194.92	195.18	195.62	196.23	165.39	135.47
Jitter 20	193.12	194.38	195.59	195.20	145.58	163.52
Jitter 25	202.50	206.48	195.84	206.94	82.64	105.63
Jitter 30	200.67	203.71	200.96	204.58	90.38	135.98
Jitter 35	188.59	196.55	174.68	195.49	83.81	26.77
Jitter 40	192.65	201.42	176.27	200.34	37.94	15.16
Jitter 45	194.07	210.55	99.44	148.33	55.14	12.48
Jitter 50	179.66	197.36	140.33	153.08	44.72	86.21

Tabela 11: Resultados dos determinadores para variação do shimmer.

Sinal de Voz	Semi	Extração de harmônicas	Autoc.	Estrut. temp. 1	Estrut. temp. 2	wavelet
Shimmer 0	199.80	198.64	200.45	199.81	199.79	199.81
Shimmer 1	199.80	198.64	200.45	199.81	199.79	199.81
Shimmer 2	199.80	198.64	200.45	199.81	199.79	199.81
Shimmer 3	199.80	198.66	200.45	199.81	199.79	199.81
Shimmer 4	199.80	198.63	200.45	199.81	199.79	199.81
Shimmer 5	199.80	198.64	200.45	199.81	199.79	199.81
Shimmer 6	199.80	198.67	200.45	199.81	199.79	199.81
Shimmer 7	199.80	198.67	200.45	199.81	199.79	199.81
Shimmer 8	199.80	198.66	200.45	199.81	199.79	199.81
Shimmer 9	199.80	198.66	200.45	199.81	199.79	199.81
Shimmer 10	199.80	198.71	200.45	199.81	199.79	199.81
Shimmer 15	199.80	198.71	200.45	199.81	179.29	194.48
Shimmer 20	199.80	198.78	200.45	199.81	148.17	187.11
Shimmer 25	199.80	198.83	200.45	199.81	149.55	190.19
Shimmer 30	199.80	198.91	200.45	199.81	139.15	135.66
Shimmer 35	199.80	199.15	200.45	197.70	136.48	134.43

Sinal de Voz	Semi	Extração de harmônicas	Autoc.	Estrut. temp. 1	Estrut. temp. 2	wavelet
Shimmer 40	199.80	199.02	200.41	187.14	124.32	141.37
Shimmer 45	199.80	199.61	195.27	178.80	96.16	136.39
Shimmer 50	199.79	199.66	199.84	173.98	111.92	117.87

Tabela 12: Resultados dos determinadores para vozes normais

Vozes	Vogal	Semi-automático F0(desvio padrão)	Extração	Extração desv (%)	Autoc	Autoc. desv(%)
Voz41	\a\	130.89 (1.51)	130.84	0.32	131.23	2.38
	\e\	143.68 (1.58)	143.67	0.35	143.84	1.61
	\i\	139.95 (2.78)	139.91	0.28	140.06	1.46
Voz42	\a\	117.11 (1.21)	117.11	0.38	117.27	1.61
	\e\	128.54 (1.32)	128.53	0.26	128.62	1.91
	\i\	133.69 (1.43)	133.68	0.04	133.84	1.73
Voz43	\a\	188.93 (2.00)	188.92	0.01	189.28	0.19
	\e\	186.89 (1.93)	186.91	0.01	187.03	0.07
	\i\	216.01(1.28)	216.05	0.02	216.37	0.17
Voz44	\a\	195.83 (2.00)	195.13	0.36	199.00	1.62
	\e\	172.37 (2.05)	170.93	0.84	177.00	2.69
	\i\	178.60 (2.37)	177.82	0.44	182.00	1.90
Voz45	\a\	268.94 (3.89)	265.66	1.22	270.00	0.39
	\e\	275.47 (4.65)	272.22	1.18	280.00	1.64
	\i\	295.16 (3.67)	290.13	1.70	296.00	0.28
Voz46	\a\	219.86 (1.29)	218.32	0.70	221.00	0.52
	\e\	226.53 (1.72)	225.00	0.68	227.00	0.21
	\i\	239.64 (1.60)	237.10	1.06	242.00	0.98

Tabela 13: Valores calculados pelos determinadores para vozes masculina patológicas

Voz	Vog	Semi-automático	Extração	Extração Desvio (%)	Autoc.	Autoc. Desvio (%)
Voz1	\a\	261.62(4.63)	261.57	0.02	262.03	0.16
	\e\	256.10(4.76)	256.11	0.00	256.39	0.11
	\i\	259.23(6.43)	259.23	0	259.86	0.24
Voz2	\a\	165.27(3.67)	165.28	0.01	165.7	0.26
	\e\	190.54(2.38)	190.52	0.01	190.79	0.13
	\i\	184.32(1.84)	184.32	0	184.85	0.29
Voz3	\a\	225.69(1.40)	225.7	0.00	226.02	0.15
	\e\	223.78(2.26)	223.75	0.01	224.1	0.14
	\i\	225.33(1.97)	225.33	0	225.54	0.09
Voz4	\a\	243.02(2.37)	243.02	0	243.61	0.24
	\e\	257.38(2.52)	257.34	0.02	257.72	0.13
	\i\	259.76(3.11)	259.73	0.01	260.32	0.22
Voz5	\a\	168.15(23.94)	165.79	1.40	155.06	7.78
	\e\	174.29(10.47)	174.11	0.10	164.75	5.47
	\i\	179.43(7.30)	179.35	0.04	169	5.81
Voz6	\a\	212.53(1.88)	212.51	0.01	212.89	0.17
	\e\	237.84(5.93)	237.75	0.04	237.92	0.03
	\i\	243.30(4.92)	243.22	0.03	244.01	0.29
Voz7	\a\	139.08(2.15)	139.05	0.02	139.39	0.22
	\e\	144.17(1.53)	144.13	0.03	144.24	0.05
	\i\	173.35(2.44)	173.28	0.04	173.37	0.01
Voz8	\a\	117.28(0.63)	118.16	0.75	117.35	0.06
	\e\	115.89(0.68)	115.88	0.01	115.96	0.06
	\i\	132.29(1.34)	132.28	0.01	132.34	0.04
Voz9	\a\	185.22(5.28)	185.19	0.02	185.48	0.14
	\e\	191.63(3.38)	191.57	0.03	191.73	0.05
	\i\	209.63(4.35)	209.57	0.03	209.94	0.15
Voz10	\a\	175.68(3.36)	175.65	0.02	175.78	0.06
	\e\	181.14(3.39)	181.12	0.01	181.32	0.10
	\i\	203.46(2.61)	203.45	0.00	204.01	0.27

Voz	Vog	Semi-automático	Extração	Extração Desvio (%)	Autoc.	Autoc. Desvio (%)
Voz11	\a\	129.91(1.45)	129.91	0	130.07	0.12
	\e\	131.95(1.68)	131.94	0.01	132.09	0.11
	\i\	138.02(5.33)	138	0.01	138.24	0.16
Voz12	\a\	160.12(2.76)	160.11	0.01	160.51	0.24
	\e\	158.27(3.26)	158.21	0.04	158.72	0.28
	\i\	217.62(2.73)	217.57	0.02	218.16	0.25
Voz13	\a\	122.86(0.89)	122.85	0.01	122.95	0.07
	\e\	128.44(1.29)	128.41	0.02	128.54	0.08
	\i\	133.54(0.93)	133.51	0.02	133.69	0.11
Voz14	\a\	110.55(0.76)	110.56	0.01	110.79	0.22
	\e\	109.65(0.81)	109.63	0.02	109.74	0.08
	\i\	113.30(1.19)	113.27	0.03	113.55	0.22
Voz15	\a\	128.23(1.27)	128.23	0	128.6	0.29
	\e\	130.57(0.83)	130.52	0.04	130.98	0.31
	\i\	130.33(0.76)	130.3	0.02	130.49	0.12
Voz16	\a\	168.44(11.30)	168.49	0.03	169.18	0.44
	\e\	169.64(14.82)	169.63	0.01	170.05	0.24
	\i\	208.57(20.17)	209.27	0.34	209.69	0.54
Voz17	\a\	223.49(8.64)	223.17	0.14	215.5	3.58
	\e\	244.63(1.85)	244.6	0.01	244.64	0.00
	\i\	241.79(5.01)	241.68	0.05	242.07	0.12
Voz18	\a\	238.12(9.77)	238.07	0.02	238.47	0.15
	\e\	240.98(26.44)	240.26	0.30	235.27	2.37
	\i\	229.50(12.21)	228.95	0.24	227.02	1.08
Voz19	\a\	218.13(5.27)	218.19	0.03	218.53	0.18
	\e\	268.64(6.57)	268.68	0.01	268.99	0.13
	\i\	282.62(5.21)	282.57	0.02	283.23	0.22
Voz20	\a\	154.91(1.60)	154.91	0	155.07	0.10
	\e\	173.33(1.36)	173.27	0.03	173.37	0.02
	\i\	205.42(1.41)	205.31	0.05	205.63	0.10

Tabela 14: Valores encontrados pelos determinadores para vozes femininas patológicas

Voz	Vog	Semi-automático	Extração	Extração Desvio (%)	Autoc.	Autoc. Desvio (%)
Voz21	\a\	154.35(5.37)	154.36	0.01	154.6	0.16
	\e\	178.68(2.37)	178.66	0.01	178.72	0.02
	\i\	207.99(2.81)	207.96	0.01	208.28	0.14
Voz22	\a\	219.50(2.66)	219.48	0.01	220.15	0.30
	\e\	210.62(1.38)	210.55	0.03	210.84	0.10
	\i\	213.42(1.41)	213.41	0.00	213.66	0.11
Voz23	\a\	337.59(9.16)	337.59	0	338.44	0.25
	\e\	338.92(6.26)	338.75	0.05	339.54	0.18
	\i\	350.96(11.30)	350.62	0.10	351.35	0.11
Voz24	\a\	164.01(1.90)	164	0.01	164.2	0.12
	\e\	160.29(2.83)	160.25	0.02	160.44	0.09
	\i\	164.18(2.52)	164.17	0.01	164.47	0.18
Voz25	\a\	271.42(9.03)	271.2	0.08	271.61	0.07
	\e\	234.51(5.10)	234.5	0.00	234.96	0.19
	\i\	249.42(4.36)	249.4	0.01	249.91	0.20
Voz26	\a\	138.14(1.83)	138.14	0	138.44	0.22
	\e\	135.88(1.79)	135.87	0.01	136.05	0.13
	\i\	142.08(1.72)	142.08	0	142.16	0.06
Voz27	\a\	110.87(2.74)	110.99	0.11	110.81	0.05
	\e\	138.28(1.77)	138.33	0.04	138.37	0.07
	\i\	152.68(1.85)	152.69	0.01	152.85	0.11
Voz28	\a\	110.47(2.68)	110.24	0.21	115.31	4.38
	\e\	136.56(2.11)	136.36	0.15	142.52	4.36
	\i\	143.65(2.10)	143.62	0.02	143.77	0.08
Voz29	\a\	320.47(7.86)	320.24	0.07	320.93	0.14
	\e\	344.29(26.33)	342.86	0.42	342.04	0.65
	\i\	349.67(21.05)	349.34	0.09	349.79	0.03
Voz30	\a\	143.25(19.90)	143.23	0.01	142.04	0.84
	\e\	137.09(1.89)	137.1	0.01	137.24	0.11
	\i\	162.32(7.98)	163.69	0.84	158.08	2.61

Voz	Vog	Semi-automático	Extração	Extração Desvio (%)	Autoc.	Autoc. Desvio (%)
Voz31	\a\	234.16(3.78)	234.09	0.03	234.74	0.25
	\e\	233.44(1.60)	233.37	0.03	233.92	0.21
	\i\	250.00(0.94)	249.98	0.01	250.56	0.22
Voz32	\a\	203.60(3.07)	203.62	0.01	203.95	0.17
	\e\	203.68(4.54)	203.63	0.02	203.9	0.11
	\i\	358.91(9.59)	358.86	0.01	359.7	0.22
Voz33	\a\	188.75(2.37)	188.74	0.01	188.97	0.12
	\e\	210.71(3.89)	210.68	0.01	211.02	0.15
	\i\	218.54(6.14)	218.49	0.02	219.08	0.25
Voz34	\a\	235.15(2.49)	235.16	0.00	235.54	0.17
	\e\	235.40(2.07)	235.38	0.01	235.78	0.16
	\i\	255.19(2.04)	255.19	0	255.77	0.23
Voz35	\a\	204.36(1.48)	204.35	0.00	204.94	0.28
	\e\	195.00(1.98)	195	0	195.46	0.24
	\i\	215.88(1.52)	215.88	0	216.48	0.28
Voz36	\a\	188.56(2.09)	188.57	0.01	188.86	0.16
	\e\	199.23(3.27)	199.14	0.05	199.39	0.08
	\i\	214.73(4.03)	214.65	0.04	214.93	0.09
Voz37	\a\	268.30(2.88)	268.27	0.01	269	0.26
	\e\	267.75(3.33)	267.73	0.01	268.01	0.10
	\i\	283.69(3.28)	283.66	0.01	284.09	0.14
Voz38	\a\	197.18(13.28)	196.91	0.14	193.11	2.06
	\e\	201.81(10.46)	201.7	0.05	199.75	1.02
	\i\	206.22(11.64)	206.1	0.06	200.62	2.72
Voz39	\a\	209.48(1.83)	209.48	0	209.98	0.24
	\e\	226.00(2.03)	225.99	0.00	226.22	0.10
	\i\	231.49(1.87)	231.47	0.01	232.1	0.26
Voz40	\a\	148.31(4.35)	148.32	0.01	148.35	0.03
	\e\	162.77(11.22)	162.67	0.06	161.03	1.07
	\i\	166,66(7.77)	166.59	0.04	166.74	0.05

Tabela 15: Valor dos períodos de um sinal de voz patológica

Períodos	Semi-automático	Extração	Autoc.
1	6.509	6.47084	6.48526
2	6.527	6.55735	6.43991
3	6.503	6.57788	6.43991
4	6.49	6.49514	6.43991
5	6.498	6.45099	6.43991
6	6.506	6.52785	6.43991
7	6.433	6.50582	6.43991
8	6.365	6.41387	6.43991
9	6.599	6.45996	6.43991
10	6.456	6.47725	6.43991
11	6.471	6.46678	6.43991
12	6.465	6.40293	6.48526
13	6.342	6.46343	6.48526
14	6.622	6.51564	6.48526
15	6.496	6.5185	6.48526
16	6.409	6.51524	6.48526
17	6.608	6.45521	6.48526
18	6.504	6.51839	6.48526
19	6.489	6.49926	6.48526
20	6.533	6.55485	6.48526
21	6.531	6.5105	6.48526
22	6.49	6.40253	6.48526
23	6.52	6.53815	6.48526
24	6.497	6.53888	6.43991
25	6.5	6.48224	6.43991
26	6.425	6.40316	6.48526
27	6.485	6.53178	6.48526
28	6.477	6.46695	6.43991
29	6.548	6.5009	6.43991
30	6.422	6.48643	6.43991
31	6.465	6.47018	6.43991
32	6.469	6.49117	6.43991
33	6.383	6.4688	6.39456
34	6.498	6.45664	6.39456
35	6.313	6.46245	6.39456
36	6.366	6.33298	6.39456
37	6.438	6.5223	6.39456
38	6.39	6.38543	6.39456
39	6.398	6.34248	6.39456
40	6.397	6.37737	6.39456
41	6.405	6.41945	6.39456
42	6.387	6.42745	6.39456
43	6.455	6.36304	-
44	6.387	6.42569	-
45	6.423	6.39025	-
46	6.584	6.46118	-

Períodos	Semi-automático	Extração	Autoc.
47	6.442	6.4369	-
48	6.452	-	-