

Ref: CLME'2011_0706A

ANÁLISE ACÚSTICA VOCAL - DETERMINAÇÃO DO JITTER E SHIMMER PARA DIAGNÓSTICO DE PATOLOGIAS DA FALA

João Paulo Teixeira*, Débora Barbosa Ferreira, Susana Moreira Carneiro

Escola Superior de Tecnologia e Gestão
Instituto Politécnico de Bragança - Bragança, Portugal
*Email: joaopt@ipb.pt

RESUMO: *Algumas técnicas de processamento digital de sinais têm sido usadas para analisar desordens vocais provocadas por patologias na laringe. Os exames laringoscópicos usados na detecção dessas patologias são técnicas invasivas que causam desconforto ao paciente. A Análise acústica vocal das características temporais e espectrais dos sinais de fala pode ser utilizada como técnica auxiliar à laringoscopia, tanto para pré-diagnósticos de patologias, bem como no acompanhamento de tratamentos farmacológicos e pós-cirúrgicos. Neste artigo, descrevem-se algumas patologias do aparelho fonador que provocam alterações na qualidade da fala produzida. Depois descrevem-se os algoritmos implementados para determinação da frequência fundamental (F0) pelo método Cepstral e pelo método da auto-correlação. Descreve-se também o algoritmo aqui desenvolvido e implementado para determinação do jitter e do shimmer pelos seus parâmetros de Jitta, Jitt, rap, ppq5, para o caso do jitter e Shim, ShdB, apq3 e apq5, para o caso do shimmer. Estes parâmetros de jitter e shimmer são muitas vezes indiciadores de patologias.*

1.INTRODUÇÃO

É sabido que as características de produção do sinal da fala são distintas de pessoa para pessoa. Estas correspondem a uma combinação de características anatómicas, relativas ao aparelho fonador, e características adquiridas, sendo que estas últimas correspondem a diferentes hábitos e maneiras de falar adquiridos por cada locutor.

A existência de uma base de dados normativos que caracterizam a qualidade vocal tem a vantagem de: (a) servir como elemento discriminativo entre a voz normal e a patológica; (b) permitir a avaliação da voz e a sua monitorização do ponto de vista clínico e/ou profissional; e (c) diminuir o grau de subjectividade da análise perceptiva ao ser feita a sua correlação com os dados quantitativos [Guimarães, 2007].

A análise acústica vocal tem como objectivo quantificar e caracterizar um sinal sonoro. Quando usada no âmbito do estudo da voz, a análise acústica permite de forma não invasiva determinar e quantificar a qualidade vocal do indivíduo através dos diferentes parâmetros acústicos que compõem o sinal – periodicidade, amplitude, duração e composição espectral. Constitui assim um método de avaliação objectiva que nos permite, entre outras utilidades, detectar precocemente problemas vocais.

Existem ainda outros métodos para a análise fisiológica laríngea tais como a laringoscopia indirecta, endoscopia, estroboscopia, electroglotografia (EGG) e a avaliação perceptiva [Guimarães, 2007]. Estes métodos são actualmente utilizados, no entanto, têm algumas limitações, uma vez que alguns são de carácter subjectivo, como é o caso da avaliação perceptiva. Outros deles são invasivos e tornam-se um pouco

desconfortáveis para os pacientes podendo até provocar o vômito, como por exemplo a endoscopia.

Os dados encontrados através da análise acústica são complementares à análise perceptivo-auditiva. Ao mesmo tempo, a correlação entre estas duas avaliações e a avaliação laringoscópica, que tem como principal objectivo ampliar a imagem possibilitando avaliar a presença de lesões pequenas, permite-nos estender o leque de possibilidades diagnósticas e terapêuticas. Sendo assim, além da objectivação do sinal sonoro, a análise acústica permite ao avaliador captar as alterações vocais precoces, sendo também um óptimo recurso para promoção e prevenção da saúde vocal. A análise acústica é capaz de fornecer o formato da onda sonora permitindo-nos avaliar determinadas características tais como a frequência fundamental, definida como o número de vibrações por segundo produzidas pelas cordas vocais, as medidas de perturbação da frequência, o jitter, definido como sendo a perturbação da frequência fundamental ciclo a ciclo e ainda medidas da perturbação da amplitude, o shimmer, que é a variabilidade da amplitude ciclo a ciclo. Estes constituem os principais parâmetros acústicos utilizados actualmente na detecção de patologias. Existem ainda outras características do sinal sonoro capazes de nos fornecer informações importantes, tais como os formantes, as medidas de ruído, a intensidade, e o tempo máximo de fonação.

A frequência fundamental vocal (F0), medida mais frequentemente em Hertz, é o termo usado para referir o parâmetro físico resultante da vibração das cordas vocais por unidade de tempo no comportamento vocal sustentado ou em fala encadeada. Reflecte a eficiência do sistema fonatório, a biomecânica laríngea e a sua interacção com a aerodinâmica, sendo portanto um importante parâmetro na avaliação anatómica e funcional da laringe. Esta medida também usada para distinção entre locutores, uma vez que depende de características físicas do trato vocal tais como comprimento, tensão e massa.

Existe uma gama típica de valores desta frequência (F0) para os diferentes sexos e idades. Aceita-se por exemplo, que a F0 das mulheres é mais elevada do que a F0 dos homens, embora existam discrepâncias relativamente aos valores apresentados. No entanto estes valores não são estacionários uma vez que, além de variarem com o sexo e a idade, pensa-se que dependem também de factores como o estado de espírito da pessoa, o período do dia em que se enquadram (de manhã, à tarde e à noite), os hábitos de vida (álcool e tabaco), o uso profissional da voz (voz falada e cantada) e à disфонia.

As medidas da F0 mais referidas na literatura são a média, a mediana, o desvio padrão, o máximo e o mínimo.

As medidas obtidas em vários estudos realizados por diferentes autores relativas à média da F0, até à data, não distinguem significativamente a maioria dos indivíduos com voz patológica dos indivíduos com voz “normal”. Não foram encontradas diferenças estatisticamente significativas na média da F0 entre indivíduos com tumor laríngeo e os controlos saudáveis. O mesmo aconteceu na tentativa de distinção entre lesões benignas de massa, nódulos e edema em comparação com esses mesmos controlos saudáveis. Foram, no entanto, encontradas diferenças significativas na média da F0 entre indivíduos com paralisia das cordas vocais, nódulos e atrito vocal em comparação com controlos saudáveis.

A evidência científica mostra que os indivíduos com patologia apresentam, tendencialmente, uma extensão da F0 mais restrita e mais baixa. Por essas razões considera-se que as medidas de variabilidade da F0 são úteis para a avaliação do grau da patologia vocal.

Pela importância deste parâmetro, F0, apresentam-se neste trabalho dois métodos, baseado no Cepstro e na auto-correlação, usados para determinação automática das curvas de F0.

O jitter é uma medida de curto termo (ciclo a ciclo) e refere-se a pequenas variações involuntárias na frequência fundamental, que permite determinar o grau de estabilidade do sistema fonatório. O jitter altera-se principalmente com a falta de controlo de vibração de cordas vocais. As vozes de pacientes com patologias na voz apresentam frequentemente uma maior percentagem de jitter.

É aceitável que qualquer voz seja, de certa forma, instável face a factores de ordem neurológica, emocional e biomecânica. É, por isso, previsível a presença de um pequeno grau de perturbação e irregularidade no sinal vocal.

A maioria dos investigadores considera como valor típico “normal” a variação entre 0,5 e os 1,0% para as fonações sustentadas em jovens adultos [Guimarães, 2007].

Considerando que o jitter é uma medida da irregularidade percentual na altura da nota vocal (perturbação da F0), o shimmer é uma medida da irregularidade percentual na amplitude da nota vocal. É muitas vezes referida como a perturbação da amplitude. O shimmer, portanto, mede a variação na intensidade dos ciclos adjacentes de vibração das cordas vocais e altera-se com a redução da resistência glótica e lesões de massa nas cordas vocais, estando correlacionado com a presença de ruído à emissão e com a soprosidade. A Tabela 1 mostra os valores normativos dos parâmetros para a emissão do sinal sonoro “a”. Nesta tabela são apresentados valores para os parâmetros F0, Jitter, Shimmer e HNR (harmonic to noise ratio) [Murphy e Akande, 2005].

Tabela 1. Valores normativos dos parâmetros para a emissão do sinal sonoro “a”.

	Crianças	Mulher	Homem
Intensidade (dB)	69.7-78.9	-	-
F ₀ (Hz)	206-281	177-244	100-137
Jitter (%)	0.909-5.158	<1	<1
Shimmer (%)	0.393-1.013	<3	<3
HNR (dB)	-	>7	>7

Estes parâmetros de perturbação são obtidos por análise computacional da fonação vogal prolongada ou amostras de fala. Assim, a análise das medidas de perturbação da voz tem sido descrita como uma medida de fácil aplicação, indirecta, não-invasivo da função de vibração da laringe, que pode quantificar a regularidade e, consequentemente, a estabilidade da vibração das cordas vocais.

Pequenas irregularidades na onda acústica são consideradas uma variação normal associada com a função fisiológica do corpo e produção da voz. No entanto, os níveis de perturbação da voz demonstraram aumentar consideravelmente em casos de patologia da laringe, e, desta forma, discriminar desordens funcionais. Investigações em vozes que soam "normal", como em pacientes com refluxo laringo-faríngeo ou após a cirurgia à tiróide, podem levar a que os parâmetros de perturbação tenham alterações vocais imperceptíveis que não sejam facilmente detectáveis através da avaliação perceptiva ou visual. Neste caso opta-se por uma avaliação recorrendo ao jitter e shimmer como forma de obtenção dessas irregularidades.

Neste trabalho descrevem-se os algoritmos desenvolvidos para determinação destes parâmetros do jitter e do shimmer.

Na próxima secção descrevem-se algumas patologias da laringe.

2. PATOLOGIAS DA LARINGE

Existem vários tipos de exames utilizados como diagnóstico de distúrbios fonológicos. Inicialmente a avaliação vocal era realizada de forma subjectiva através da análise perceptiva auditiva. Porém, a falta de consenso entre examinadores experientes, mesmo com o uso de diferentes escalas de alterações vocais, tornou necessária a pesquisa de um tipo de avaliação objectiva, na qual a voz fosse analisada através de aparelhos que são capazes de medir vários parâmetros acústicos da voz.

Quando existe uma disfonia resultante de lesões nas cordas vocais, o processo de fonação é alterado porque os padrões de vibração durante a fase de abertura e fecho das cordas vocais são irregulares. Com base neste princípio, no presente trabalho são utilizadas técnicas de computação baseada em tecnologia sofisticada que nos permite fazer gravações de voz, que serão posteriormente analisadas pelo sistema de forma a obter medidas quantitativas das alterações da voz e, indirectamente, as características fisiológicas da laringe. Desta forma é possível detectar alterações de voz que são imperceptíveis auditivamente.

Há uma grande variedade de patologias que causam mudanças significativas nos padrões vibratórios, afectando a qualidade da produção vocal. Essas patologias podem derivar de infecções, tumores ou paralisias laríngeas.

2.1. Infecções laríngeas ou laringites

Laringite Catarral Aguda: É um processo inflamatório agudo que se instala sobre a mucosa da laringe com repercussões principalmente a nível das cordas vocais. Ocasionalmente causa congestão difusa desta mucosa com edema liso desta e diminuição da mobilidade do aparelho fonador. Pode ser desencadeada por inúmeros factores como choques térmicos, uso abusivo da voz, fumo, álcool, exposição a agentes químicos, resfriados comuns etc. Manifesta-se por uma rouquidão que pode ir até à afonia e é associada a um ligeiro edema sub-mucoso e uma hiper-secreção mucosa espessa.

Laringite crónica: A laringite crónica corresponde a uma inflamação persistente da mucosa laríngea, às vezes com muitos anos de evolução, normalmente provocada por infecções agudas repetidas. Afecta essencialmente as pessoas que estão em constante exposição a factores irritantes, nomeadamente o tabaco, álcool, ambientes repletos de fumo, pó e vapores irritantes, abuso ou má utilização da voz, entre outros. A inflamação provoca, além de uma tumefacção da mucosa laríngea, uma abundante produção de secreções, factores responsáveis pelas manifestações que dão sob a forma de rouquidão tenaz, permanente, embora variável.

2.2. Tumores da Larínge

Pólipos laríngeos: Os pólipos nas cordas vocais são formações não cancerosas que se desenvolvem devido ao uso excessivo da voz, a reacções alérgicas crónicas que afectam a laringe ou a uma inalação crónica de irritantes, como as emanções industriais ou o fumo do cigarro. É uma afecção muito frequente, sempre benigna, que surge por hipertrofia da mucosa da laringe e a sua localização mais frequente é sob a porção anterior de uma corda, ficando pendurado no espaço subglótico. Esta anomalia manifesta-se por uma rouquidão, e o exame põe em evidência uma pequena massa pediculada no bordo livre da corda vocal.

Papilomas da laringe: São considerados como uma forma de laringite crônica hipertrófica e como lesões pré-cancerosas. O papiloma da laringe é uma lesão que origina a proliferação exofítica do epitélio escamoso e estroma. O papiloma pode-se tornar maligno e ao invadir o estroma pode atingir a sua vascularização e provocar metástases. O papiloma da criança é uma afecção absolutamente diferente, provavelmente viral, rara, sem perigo de degenerescência, e que tende a curar-se na época da puberdade.

2.1. Paralisias Laríngeas

A paralisia vocal pode ocorrer unilateralmente ou bilateralmente, pode ainda ser provisória/temporária ou permanente.

Paralisia Unilateral Isolada ou paralisia recorrential: É causada devido à imobilidade de uma das cordas vocais, ou antes, de toda a hemilaringe correspondente, em particular da aritenóide. É a paralisia mais frequente e manifesta-se por rouquidão um pouco particular ou voz bitonal. Trata-se de uma mistura de sons graves ou agudos muito característicos.

Paralisia bilateral: Ocorre quando as duas cordas paralisam e é dominada pela dispneia respiratória intensa que muitas vezes origina.

Paralisias associadas: A paralisia laríngea pode associar-se a paralisias dos outros nervos cranianos.

3. DETERMINAÇÃO DA FREQUENCIA FUNDAMENTAL

Existem vários métodos que nos permitem a extracção dos parâmetros do sinal de fala que se baseiam no modelo genérico de produção de fala (Figura 1) e na parametrização por formantes do trato vocal. Neste projecto utilizaram-se dois métodos distintos para extracção da frequência fundamental que são o Método Cepstral e o Método da Autocorrelação. O Software de análise, Praat – doing phonetics by computer [Boersma e Weenink, 2011], foi usado para comparação dos resultados obtidos pelos algoritmos implementados.

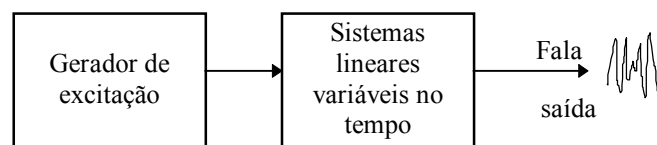


Figura 1- Modelo genérico de produção de fala.

3.1. Determinação da Frequência Fundamental pelo Método do Cepstro

A análise Cepstral de um sinal de fala permite-nos trabalhar com o sinal da glote (excitação) e do trato vocal (ressonância) separadamente, pelas suas propriedades homomórficas, separando as características do filtro do trato vocal da sequência de excitação, o que facilita o estudo das alterações nas cordas vocais. Dentro das

propriedades matemáticas envolvidas no processo, salienta-se principalmente as transformadas de Fourier e funções logarítmicas que resultaram numa função chamada Cepstral ou Cepstro.

O método do Cepstro é então uma operação matemática que consiste em extrair a Transformada de Fourier do espectro do sinal na forma de logaritmos. O nome “cepstrum” obtém-se ao inverter a ordem das primeiras quatro letras de “spectrum”. O significado físico desta transformação pode ser interpretado como uma informação do ritmo das mudanças nas diferentes bandas do espectro de um sinal qualquer. O cepstrum também pode ser interpretado como uma função que satisfaz o princípio geral da sobreposição com a convolução como entrada e a adição como a saída.

Este método permite-nos separar os efeitos da fonte e do meio de transmissão. Esta separação é possível uma vez que estes parâmetros constituem famílias harmónicas diferentes e consequentemente situam-se a que frequências diferentes.

Se tivermos um sinal composto por duas sinusóides, uma variando lentamente no tempo e a outra rapidamente, isto é uma de baixa frequência e a outra de frequência elevada; a correspondente à baixa frequência apareceria na parte “baixa” do espectro e a correspondente à alta frequência na parte “alta”. É esta lógica que está na base da separação das variações rápidas do espectro devido à fonte das variações lentas do filtro. Assim, se considerarmos o espectro como um sinal (no tempo) e aplicarmos a DFT então a parte devida à fonte deverá aparecer nas frequências elevadas e a relativa ao filtro nas frequências baixas.

O sinal da fala $s(t)$ é composto por um sinal de excitação $e(t)$ aplicado ao filtro do trato vocal, com uma resposta impulsional $v(t)$, de um ponto de vista no domínio temporal, como resume a Figura 2.

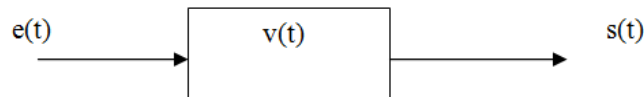


Figura 2. Modelo simples de produção da fala no domínio temporal.

$s(t)$ representa a convolução de $e(t)$ com $v(t)$, $s(t) = e(t)*v(t)$, que no domínio da frequência fica $S(w) = E(w).V(w)$, que representam as transformadas de Fourier das funções contínuas no tempo.

O sinal de excitação $E(w)$ e o filtro do trato vocal, $V(w)$, aparecem combinados multiplicativamente e a sua separação é possível usando a função logarítmica, obtendo:

$$\log [S(w)] = \log [E(w)] + \log [V(w)]$$

Portanto, a propriedade aditiva do espectro logarítmico representa a soma do espectro da fonte com o do filtro, o que nos fornece um caminho para os separar.

Sumariamente, o processo de estimação do Cepstro resume-se a uma sequência de passos como evidencia a Figura 3.

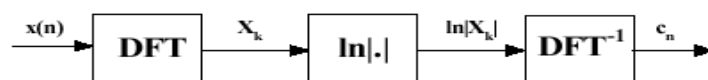


Figura 3. Diagrama de fluxo de estimação do Cepstro [Teixeira, 1995].

Depois de separados podemos eliminar cada uma das partes por um processo de filtragem, se eliminarmos a parte de frequências mais elevadas e voltarmos a efectuar uma DFT teremos um espectro “suave” com “apenas” as características devidas ao filtro (no caso da fala o filtro do tracto vocal).

A forte componente periódica no espectro logarítmico de um sinal de fala vocalizada num intervalo de frequências equivalente ao inverso do período fundamental T , aparece no Cepstro como um pico.

O eixo horizontal da função Cepstral tem dimensões temporais e o nome de quefrências. Para a fala vocalizada pode ser feita uma clara distinção entre a componente de excitação e a contribuição do trato vocal, que aparece como um aglomerado de componentes aos baixos valores de quefrência afastado da componente do período fundamental que aparece com valores mais altos de quefrência.

Na Figura 4, é apresentado o Cepstro de um segmento de fala onde o pico correspondente ao período fundamental está próximo da quefrência de 10 ms, correspondendo à frequência fundamental de 100Hz, separado das componentes do trato vocal às baixas quefrências.

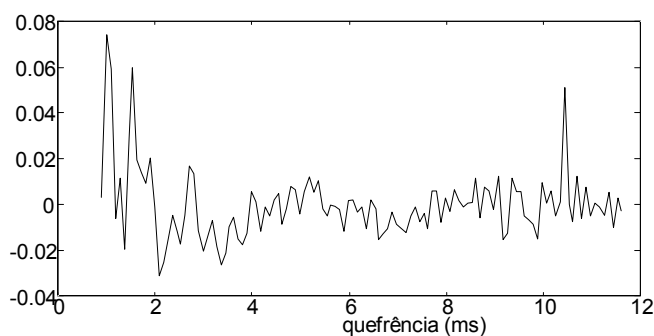


Figura 4. Cepstro de um segmento de fala [Teixeira, 1995].

A função de transferência do trato vocal e a função de excitação da voz aparecem em partes separadas da escala de quefrências, pelo que podem ser separadas as duas funções, ou removida uma delas por um processo de “lifteragem” (*liftering*), facilitando o estudo individualizado das modificações ocorridas na excitação e da parte de ressonância.

O Cepstro é constituído por um conjunto de valores Cepstrais discretos, que são o conjunto de valores de saída do processo final de IDFT (Inverse Discrete Fourier Transform). Então podemos aplicar um “lifter” com uma função de janela rectangular. Contudo, a transformada de Fourier da janela rectangular indica que tal processo gera lobos laterais indesejados às componentes requeridas no domínio das frequências logarítmicas, pelo que é requerida uma função de uma janela mais gradual. Assim aplicando esta função de “filtragem” ao sinal do Cepstro e tomando a transformada de Fourier Discreta Inversa (IDFT) do sinal resultante, obtém-se uma versão alisada do espectro logarítmico do filtro do trato vocal.

Este alisamento Cepstral pode ser usado para estimar a curva do envelope espectral como na Figura 5.

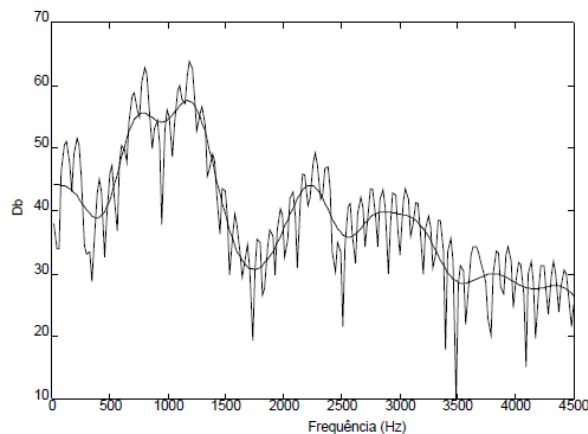


Figura 5. Espectro e envelope Cepstral (pelo método do cepstro) de um segmento de fala da vogal [a] [Teixeira, 1995].

Uma vez obtido o envelope espectral extrai-se as frequências, larguras de banda e amplitude das formantes.

Assim, o estudo do sinal acústico de vozes alteradas, através da aplicação da técnica de análise Cepstral na detecção de modificações no sinal relacionadas às alterações laríngicas e na identificação de modelos para a classificação destes distúrbios em estudo, permitirá obter uma ferramenta não-invasiva de diagnóstico.

Seguidamente é apresentado um excerto do código Matlab utilizado para a determinação da frequência fundamental.

```

a1=1; %posição inicial do vector do sinal y a ser analisado
N=length(y);
N1=512/2;
C1=floor(0.002*fs);
y=detrend(y); %remoção da eventual tendência linear do vector y
c=0;
while a1<N-2*N1
    x=y(a1:a1+511); %segmento do sinal y a ser analisado
    z=rceps(x); %Função do Matlab que determina o cepstro
    [z0,imax]=max(z(C1:512/2)); %Z0 corresponde ao "pico"
    imax=imax+C1; %índice máximo, onde se encontra o "pico"
    t0=imax/fs*1000; %passar para (ms)
    f0=1/t0*1000;
    a1=a1+N1;
    c=c+1;
    F0(c)=f0; %F0(c) é a frecuencia fundamental
end
F0correc=fcorrec(F0,1e-3); %Função que elimina os "outliers"

```

3.2. Determinação da Frequência Fundamental pelo Método da Autocorrelação

Para um sinal $x(t)$ estacionário, a autocorrelação, $r_x(\tau)$, em função do atraso τ é definida pela eq. 1:

$$r_x(\tau) = \int x(t) x(t + \tau) dt \quad (\text{Eq. 1})$$

Desta equação podemos inferir que, se estas duas partes estão correlacionadas, a autocorrelação do sinal total é dada pela soma das autocorrelações das suas partes.

A frequência fundamental, F_0 , de um sinal de fala $x(t)$, contínuo no tempo, pode ser determinada a partir dos valores máximos locais da autocorrelação de um segmento do sinal centrado em torno de t_{mid} .

O algoritmo de determinação da F_0 pelo método de Autocorrelação está resumido na Figura 6 para um determinado sinal $x(t)$.

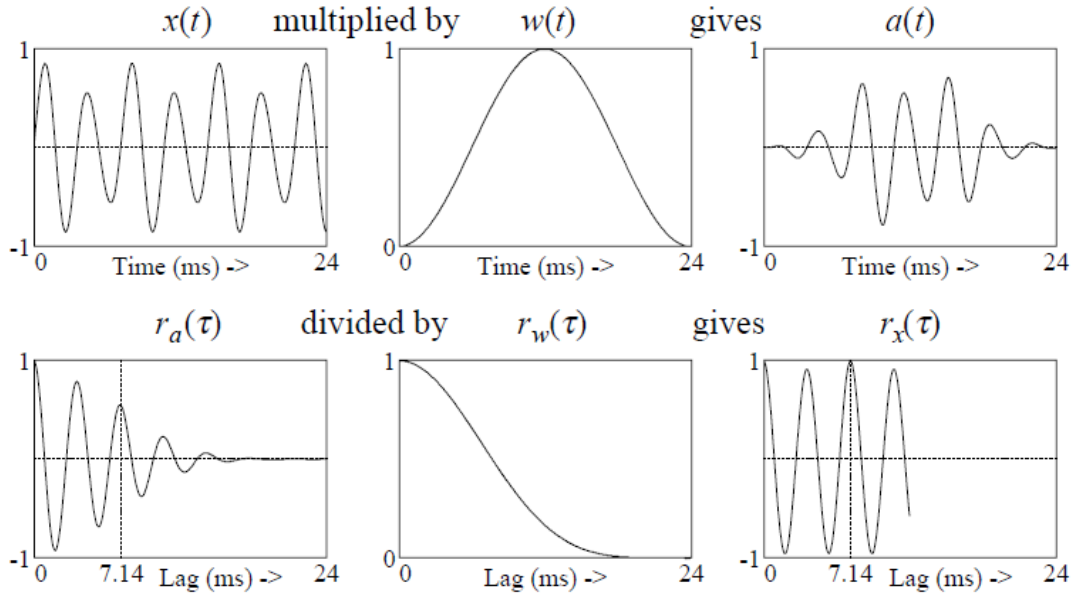


Figura 6. Representação dos sinais e operações de um exemplo de determinação da F_0 pelo Método de Autocorrelação.

Este algoritmo será resumidamente explicado passo a passo.

1ª etapa: Acha-se uma janela representativa do sinal $x(t)$ com uma duração T centrada em torno de t_{mid} . No caso da Figura 6 a janela tem um tamanho de 24ms. Esta janela deve conter pelo menos dois períodos fundamentais do sinal. Assumindo uma F_0 mínima de 80 Hz, dois períodos correspondem a $2 \cdot 1/80 = 25$ ms. Subtrai-se a esta janela a sua média, μ_x e multiplica-se o resultado por uma função de janela $w(t)$ (por exemplo uma janela de Hanning, Hamming, Blackman, Kaiser, etc.), para obter o sinal $a(t)$, como na eq. 2:

$$a(t) = \left(x \left(t_{mid} - \frac{1}{2}T + t \right) - \mu_x \right) w(t) \quad (\text{Eq. 2})$$

A função da janela $w(t)$ é simétrica em torno de $t=(1/2)T$ e zero fora do intervalo de tempo $[0, T]$. Escolheu-se para tal a janela de Hanning, dada pela eq. 3:

$$w(t) = \frac{1}{2} - \frac{1}{2} \cos \frac{2\pi t}{T} \quad (\text{Eq. 3})$$

2ª etapa: A autocorrelação normalizada, $r_a(\tau)$, do sinal $a(t)$ é uma função simétrica (eq. 4):

$$r_a(\tau) = r_a(-\tau) = \frac{\int_0^{T-\tau} a(t)a(t+\tau) dt}{\int_0^T a^2(t) dt} \quad (\text{Eq. 4})$$

3ª etapa: De seguida calcula-se a autocorrelação normalizada da janela de Hanning $r_w(\tau)$, de modo análogo à equação correspondente à da $r_a(\tau)$. A autocorrelação normalizada da janela de Hanning é dada pela eq. 5:

$$r_w(\tau) = \left(1 - \frac{|\tau|}{T}\right) \left(\frac{2}{3} + \frac{1}{3} \cos \frac{2\pi\tau}{T}\right) + \frac{1}{2\pi} \sin \frac{2\pi|\tau|}{T} \quad (\text{Eq. 5})$$

O resultado final da autocorrelação $r_x(\tau)$, do sinal original determina-se dividindo o valor da autocorrelação normalizada deste, pelo valor da autocorrelação normalizada da janela, (eq. 6):

$$r_x(\tau) = \frac{r_a(\tau)}{r_w(\tau)} \quad (\text{Eq. 6})$$

Para a implementação deste algoritmo utilizou-se a janela de Hanning. A autocorrelação foi determinada numericamente pela Transformada Rápida de Fourier. Isto é possível graças ao facto de a autocorrelação de um sinal poder ser obtida calculando a primeira transformada de Fourier do sinal, como na eq. 7:

$$\tilde{a}(\omega) = \int a(t) e^{-i\omega t} dt \quad (\text{Eq. 7})$$

E, de seguida, calcula-se a Transformada Inversa de Fourier da densidade de potência $|\tilde{a}(\omega)|^2$, que nos leva ao domínio de atraso (τ), como na eq. 8:

$$r_a(\tau) = \int |\tilde{a}(\omega)|^2 e^{i\omega\tau} \frac{d\omega}{2\pi} \quad (\text{Eq. 8})$$

No sinal resultante $r_a(\tau)$ pode-se determinar o período fundamental pela medição do tempo entre dois pontos sucessivos da curva com o valor máximo (muito próximo e 1). O seu inverso é a frequência fundamental, eq. 9.

$$F_0 = 1/T_0 \quad (\text{Eq. 9})$$

Apresentam-se de seguida as linhas de código em Matlab usadas para implementar este método:

```
N=floor(0.02*fs); %Define janelas de 20 ms
C1=floor(0.002*fs); %Define uma frequência mínima de 50 Hz
t1=(1:N)/fs*1000;
p=1;
d=0;
while p<length(y)-N
    y1=y(p:p+320).*hanning(320+1);
    p=p+N;
    X=fft(y1,640); %Transformada directa de Fourier
    rx=ifft(abs(X).^2); %Transformada inversa de Fourier
    [x0,imax]=max(rx(C1:N));
    imax=imax+C1;
    t0=imax/fs*1000;
    f0=1/t0*1000;
    d=d+1;
    F0corr(d)=f0;
end
```

3.3. Discussão dos Resultados da Determinação de F0

Para comparação dos resultados dos algoritmos implementados usamos o software Praat, no qual foi inserido um sinal de fala correspondente à frase “A Ana ama o Luís”, uma frase composta quase totalmente com sons vozeados e, portanto, sempre com valor de F0, excepto na parte final correspondente ao “s”. Esta frase permite-nos verificar se o algoritmo criado é sensível ou não às variações repetidas que a F0 pode adquirir. A curva de F0 obtida pelo Praat é apresentada na Figura 7, onde se pode ver o espectrograma do sinal com a escala representada à esquerda e a curva da F0 a azul com a escala do lado direito.

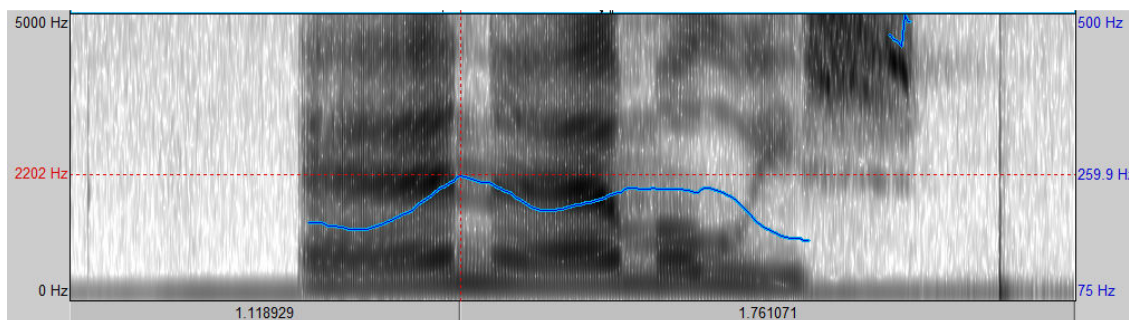


Figura 7. Curva da F0 para o sinal obtido pelo Praat.

O sinal acústico correspondente à locução da referida frase apresenta-se na Figura 8.

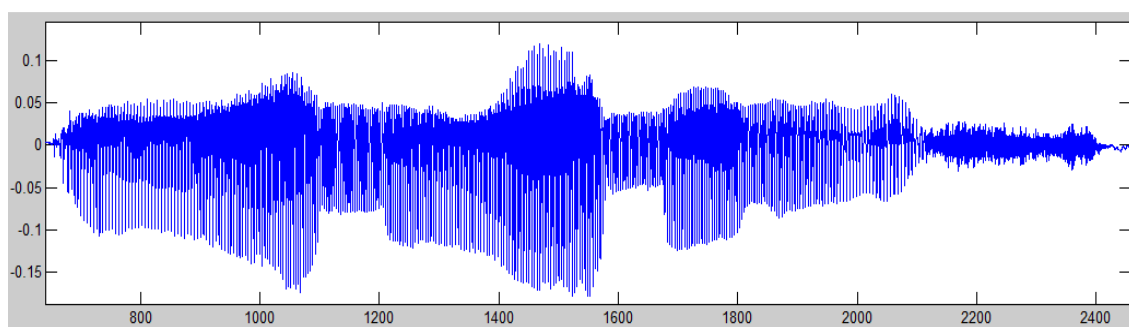


Figura 8. Onda correspondente à emissão da frase “A Ana ama o Luís”.

De seguida, aplicou-se o algoritmo criado de forma a visualizar a curva da F0 e assim poder verificar se os diferentes métodos estão de acordo com a curva da F0 obtida pelo Praat. Na Figura 9 pode-se visualizar a curva correspondente à frequência fundamental obtida através dos dois métodos distintos em Matlab: o Método da Autocorrelação, representado a azul, e o Método Cepstral, representado a vermelho.

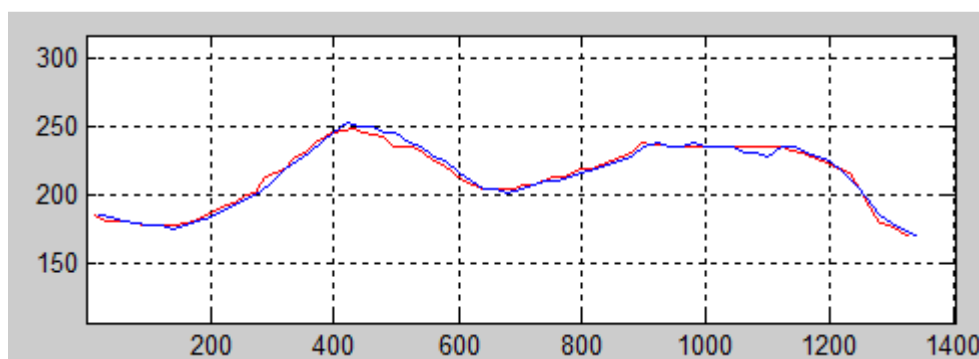


Figura 9. Resultado da curva de F0 obtidos através do método Cepstral (a vermelho) e através do método de Autocorrelação (a azul).

Analisando os gráficos obtidos verifica-se que os métodos usados e o algoritmo elaborado mostram-se eficazes pois as curvas de F0 são muito semelhantes às curvas obtidas pelo Praat.

Apresenta-se de seguida a Tabela 2 correspondente aos valores de F0 relativos ao sinal analisado obtidos através do Praat e pelo algoritmo criado.

Tabela 2. Tabela com os resultados obtidos pelo Praat e pelo algoritmo criado com os diferentes métodos para o sinal correspondente à emissão da frase “A Ana ama o Luís”.

	Praat (Hz)	Método cepstral (Hz)	Método da autocorrelação (Hz)
Media	219	216	215
Mediana	224	219	219
Desvio padrão	24	24	23
Mínimo	169	170	170
Máximo	260	296	254

4. DETERMINAÇÃO DO JITTER E DO SHIMMER

Para a determinação das medidas de perturbação jitter e shimmer baseamo-nos no facto de o jitter reflectir a variação nos períodos consecutivos e o shimmer reflectir a variação nas amplitudes consecutivas.

Assim, o algoritmo para determinar estes parâmetros terá que fazer a detecção de sincronismo com os impulsos glotais. Isto consiste na marcação realizada a cada impulso glotal, onde são detectados os “picos” máximos de cada impulso. Posto isto, serão determinados os índices dos respectivos picos máximos para poder calcular a diferença entre cada período, correspondente ao jitter.

Para o caso do shimmer, após calcular o sincronismo calcula-se a diferença entre as amplitudes sucessivas do sinal nesses instantes.

A determinação do sincronismo com os impulsos glotais [Teixeira, 95] é realizada aplicando uma média deslizante ao sinal de fala com um comprimento adequado de forma a evidenciar os períodos glotais. De seguida determinam-se os máximos de cada período glotal a que correspondem os instantes dos impulsos glotais. De seguida apresentam-se as linhas de código em Matlab usadas neste trabalho.

```
y=detrend(y);
n=length(y);

N=86*16000/fs; %Comprimento da janela da média (86 para uma Fs de 16
KHz)
espacamento=1;
M=zeros(1,n);
j=0;
M=fmedia(y,N,espacamento); % função da média deslizante
M=detrend(M);
n=length(M);
m=zeros(n,1);

%          ***** Determinação dos Impulsos Glotais *****
k=15;
```

```

while (k<n-15),
    while ((M(k+1)<M(k)) & (k<n-1)), k=k+1; end
    while ((M(k+1)>=M(k)) & (k<n-1)), k=k+1; end
    if (k<n-15),
        sina=y(k-15:k+15);
        [pico,indp]=max(sina);
        kj=k-15+indp-1;
        m(kj,1)=y(kj);
    end
end
%%%%%% Determinar os Índices dos Impulsos Glotais %
z=length(m);
v=zeros(1,z);
j=1;
for i=1:z
    if m(i)>0.02
        v(j)=i;
        j=j+1;
    end
end
v=v(1:j-1);

```

O vector m de saída contém picos nos índices correspondentes ao início do impulso glotal do sinal.

4.1. Determinação do Jitter

Existem vários tipos de medidas tanto do jitter como do shimmer. Relativamente ao jitter pode-se encontrar os seguintes [Boersma, P e Weenink, D.]:

Jitter (local): Representa a diferença média absoluta entre dois períodos consecutivos, dividido pelo período médio, equação 10. Chama-se a esse parâmetro *jitt* e tem 1.040% como limite para detectar patologias.

$$jitt = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (\text{Eq. 10})$$

Onde T_i representa o comprimento de cada período de F_0 e N é o número de períodos de F_0 .

Jitter (local, absoluto): Representa a diferença média absoluta entre dois períodos consecutivos e chama-se a este parâmetro *jita*, equação 11. Este tem como limite para detectar patologias o valor de 83.200 μs .

$$jitta = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| \quad (\text{Eq. 11})$$

Jitter (rap): Representa a média relativa da perturbação, isto é, a diferença média absoluta de um período e a média desse período com os seus dois vizinhos, dividido pelo período médio, equação 12. Este tem como limite para detectar patologias o valor de 0.680%.

$$rap = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| T_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (\text{Eq. 12})$$

Jitter (ppq5): Representa o quociente de perturbação num período de cinco pontos, ou seja, a diferença média absoluta entre um período e a média desse mesmo período com os seus quatro vizinhos, equação 13:

$$ppq5 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} |T_i - (\frac{1}{5} \sum_{n=i-2}^{i+2} T_n)|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (\text{Eq. 13})$$

Segue-se um excerto das linhas de código utilizado para a determinação do Jitter.

```
sizeV=j-1,
for i=1:sizeV-1
    periodo(i)=(v(i+1)-v(i))/fs;

end

for i=1:sizeV-2
    jitter(i)=(periodo(i+1)-periodo(i));
end

MP=mean(periodo),

jita=mean(abs(jitter)), %segundo
jitt=(jita/MP)*100, %(%)

%RAP
for i=1:sizeV-3
    md3(i)=sum(periodo(i:i+2))/3;
    vetor3(i)=periodo(i+1)-md3(i);
end
RAP=(mae(vetor3)/MP)*100,

%ppq5
for i=3:sizeV-3
    md5(i)=mean(periodo(i-2:i+2));
    vetor5(i)=periodo(i)-md5(i);
end
PPQ5=(mae(vetor5)/MP)*100,
```

4.2. Determinação do Shimmer

Relativamente ao shimmer podem-se encontrar das seguintes formas [Boersma, P e Weenink, D.]:

Shimmer (local): Representa a diferença média absoluta entre as amplitudes de dois períodos consecutivos, dividido pela amplitude média, equação 14. Chama-se a esse parâmetro shim e tem 3.810% como limite para detectar patologias.

$$Shim = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (\text{Eq. 14})$$

Shimmer (local, dB): Representa a diferença média absoluta do logaritmo de base 10 da diferença entre dois períodos consecutivos e chama-se a este parâmetro SHDB, equação 15. Este tem como limite para detectar patologias o valor de 0.350 dB.

$$ShdB = \frac{1}{N-1} \sum_{i=1}^{N-1} |20 * \log(A_{i+1}/A_i)| \quad (\text{Eq. 15})$$

Onde A_i representa a amplitude do impulso glotal i , e N o número de impulsos glotais.

Shimmer (apq3): Representa o quociente de perturbação da amplitude de três pontos, ou seja, a diferença média absoluta entre a amplitude de um período e a média de amplitudes dos seus dois vizinhos, dividido pela amplitude média, equação 16:

$$apq3 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - (\frac{1}{3} \sum_{n=i-2}^{i+2} A_n)|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (\text{Eq. 16})$$

Shimmer (apq5): Representa o quociente de perturbação da amplitude de cinco pontos, ou seja, a diferença média absoluta entre a amplitude de um período e a média das amplitudes dele e dos seus quatro vizinhos mais próximos, dividido pela amplitude média, equação 17:

$$apq5 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} |A_i - (\frac{1}{5} \sum_{n=i-2}^{i+2} A_n)|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (\text{Eq. 17})$$

O algoritmo de determinação do Shimmer é apresentado em seguida, onde é determinada a média deslizando e calculados os “picos” do sinal. Refira-se que se volta a determinar o sincronismo com os impulsos glotais de forma alternativa, mas usando o mesmo princípio da média deslizando. Os picos dos impulsos glotais são os máximos do sinal numa vizinhança muito próxima da posição dos picos dos impulsos glotais.

```
% Determinar a media deslizando do sinal de entrada y
sinal=x;
janela=150;

espacamento=1;
n=length(sinal);
media=zeros (fix (n/espacamento), 1);
for i=janela/2:espacamento:n-janela/2,
    media (i/espacamento+1)=sum (abs (sinal (i-
janela/2+1:i+janela/2) .*hanning (janela)))/janela;
end;
for j=1:round (janela/2/espacamento),
media (j)=sum (abs (sinal (1:j+janela/2) .*hanning (j+janela/2)))/(j+janela/
2);
end;
v=(i/espacamento+2);
f=(fix (n/espacamento));
for j=v:f,
    media (j)=sum (abs (sinal ((j-1)*espacamento-janela/2+1:n) .*hanning (n-
((j-1)*espacamento-janela/2))))/(n+1-((j-1)*espacamento-janela/2+1));
end
%para calcular os picos do sinal y
m=zeros (n, 1);
ms=zeros (1, n);
```

```

picossinal=zeros(1,n);
k=21;
j=1;
w=1;
while (k<n-21),
    while ((media(k+1)<media(k)) & (k<n-1)), k=k+1; end;
    while ((media(k+1)>=media(k)) & (k<n-1)), k=k+1; end;
    if (k<n-1),
        m(k,1)=media(k);
        picos(j)=media(k);
        [picossinal(j),ind]=max(sinal(k-20:min(k+20,length(sinal))));
        ms(ind+k-21)=max(sinal(k-20:min(k+20,length(sinal))));

        j=j+1;
    end;
end;
z=ms>0.01;
q=sum(z);
ms_valores=zeros(1,q);
for i=1:n
    if ms(i)>0.02,
        ms_valores(w)=ms(i);
        w=w+1;
    end;
end
plot(1:n,x,'k',1:n,media*2,'b',1:n,ms,'r');
for i=1:q-1
    shimmer(i)=ms_valores(i)-ms_valores(i+1);
end

MA=mae(ms_valores),
Shim=(mae(abs(shimmer))/MA)*100,
ShdB=mae(abs(20*log(ms_valores(i+1)/ms_valores(i)))),

%apq3
for i=2:j-2
    md3(i)=mean(ms_valores(i-1:i+1));
    vetor3(i)=ms_valores(i)-md3(i);
end
apq3=(mae(v
%apq5
for i=3:j-3
    md5(i)=mean(ms_valores(i-2:i+2));
    vetor5(i)=ms_valores(i)-md5(i);
end
apq5=(mae(vetor5)/MA)*100,
etor3)/MA)*100,

```

4.3. Discussão dos Resultados da Determinação do Jitter e do Shimmer

Tanto para o algoritmo do Jitter como para o do Shimmer começou-se por implementar uma função que detecta o sincronismo do período fundamental para a análise síncrona com a F0. O vector de saída dessa função contém picos nos índices correspondentes ao início do impulso glotal do sinal, ou seja, esta função retorna um vector com a mesma dimensão do sinal mas só com os picos. Na Figura 10 é possível observar que a função que detecta o sincronismo (representada na figura a vermelho) calcula correctamente o pico máximo do impulso glotal. A azul está representada a curva do sinal alisado por

uma função de média deslizante, que poderia desviar ligeiramente a posição do máximo do sinal.

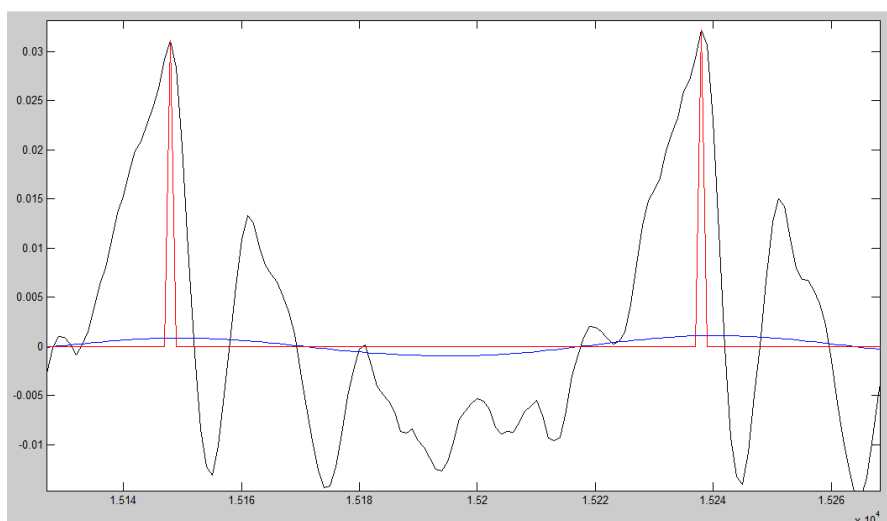


Figura 10. Imagem onde podemos verificar que o sincronismo corresponde, agora, ao pico máximo do impulso glotal.

Seguidamente são apresentadas as tabelas relativas aos resultados obtidos através do algoritmo criado em Matlab para o Jitter e o Shimmer para o “sinal1”, uma gravação da emissão da vogal “a” de forma sustentada por uma das autoras do trabalho. Para a análise feita ao “sinal1” o valor de N (que representa a janela de alisamento utilizado) aplicado foi de $(86 * 16000 / fs)$, ver Tabela 3.

Tabela 3. Tabela de valores do Praat e do Matlab para o jitter do “sinal1”.

“sinal1_cort”	Matlab
Jitt (%)	0.6540
Jitta (s)	36.504e-6
Rap (%)	0.4247
ppq5 (%)	0.4429

Para o Shimmer, os valores obtidos para o “sinal1” encontram-se na Tabela 4.

Tabela 4. A tabela seguinte é representativa dos valores do Praat e do Matlab para o shimmer do “sinal1”.

“sinal1”	Matlab
Shim (%)	4.0196
ShdB (dB)	0.1135
Apq3 (%)	2.1765
Apq5 (%)	2.3506

Por comparação com os valores obtidos por outros pacotes de software os resultados são muito semelhantes.

4. CONCLUSÃO

Dada a relevância do registo da F0 e de parâmetros como jitter e o shimmer para um primeiro diagnóstico de algumas patologias descritas associadas ao aparelho fonador humano, apresentou-se aqui a implementação de dois métodos descritos na literatura para determinação automática da F0, com resultados que se podem considerar dentro de uma margem de erro aceitável quando comparados com os resultados obtidos pelo pacote de software Praat. Apresentou-se também o algoritmo e a respectiva implementação de um método desenvolvido neste projecto para determinar os parâmetros associados aos jitter e ao shimmer nas suas várias medidas como sejam o Jitt, Jitta, RAP e ppq5, para o jitter e Shim, ShdB, Apq3 e Apq5 para o shimmer. Os valores obtidos para estes parâmetros são da mesma ordem dos resultados obtidos pelo mesmo Praat.

REFERÊNCIAS

Guimarães, Isabel, A Ciência e a Arte da Voz Humana, Março de 2007, Escola Superior de Saúde de Alcoitão.

Teixeira, João Paulo, Modelação Paramétrica de Sinais Para Aplicação em Sistemas de Conversão Texto-Fala, Dissertação de Mestrado, Outubro de 1995, FEUP, Porto.

Paul Boersma e David Weenink, Praat: doing phonetics by computer, <http://www.fon.hum.uva.nl/praat/>, University of Amsterdam.

Murphy, Peter e Akande, Olatunji, Cepstrum-Based Estimation of the Harmonics-to-noise Ratio for Synthesized and Human Voice Signals. Em *Nonlinear Analyses and Algorithms for Speech Processing*, Barcelona 2005, LNAI 3817, Springer.

Zwetsch, I. C.; Fagundes, R. D. R; Russomano, T. e , Scolari, D. : Processamento digital de sinais no diagnóstico diferencial de doenças laringeas benignas. *Scientia Medica*, Porto Alegre: PUCRS, v. 16, n. 3, jul./set. 2006