

Produção e Análise do Sinal de Fala

Tecnologias de Reabilitação Aplicações de Processamento de Sinal

JPT

1

Resumo

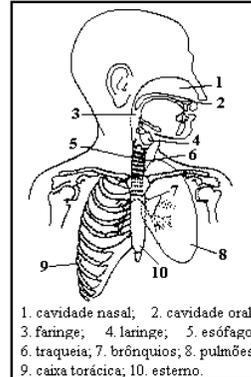
- ▶ **Aparelho Fonador Humano**
 - Modelo de Produção da Fala
 - Frequência Fundamental – F0
 - Formantes
- ▶ **Sinal de Fala**
 - recolha (filtro anti-aliasing, Fa, estéreo, quantificação, microfone, sala insonorizada, armazenamento)
 - características (vozeado, não vozeado, oclusão, silêncio)
- ▶ **Ferramentas de Análise**
 - Amplitude Média Deslizante
 - Janelas
 - Energia Média Deslizante
 - Taxa de Passagem por Zero

JPT

2

Aparelho Fonador Humano

- ▶ Forte ligação ao sistema de respiração.
- ▶ alvéolos - brônquios - traqueia - laringe - faringe - cavidade nasal e/ou oral.
- ▶ Na laringe situam-se as cordas vocais.
- ▶ O espaço entre as cordas vocais é a glote.



JPT

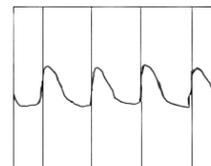
3

Aparelho Fonador Humano

- ▶ Um período de abertura e fecho das cordas vocais chama-se impulso glotal ou período fundamental.
- ▶ Cavidades supraglotais são formadas pela faringe e cavidades nasal e oral.
- ▶ Trato vocal - formas tomadas pelas cavidades supraglotais.



Glote



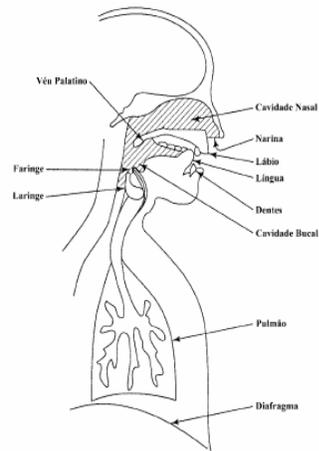
Impulso Glotal

JPT

4

Aparelho Fonador Humano

- ▶ As cavidades supraglotais formam um conjunto de ressoadores que favorecem a passagem de algumas frequências e a atenuação de outras, consoante a forma que assumem.
- ▶ Frequências formantes – frequências favorecidas pelo trato vocal.
- ▶ Anti-formantes – frequências atenuadas pelo trato vocal.

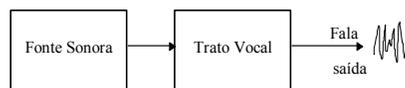


JPT

5

Modelo de Produção de Fala

- ▶ Modelo de Engenharia – separa a fonte sonora da operação de filtragem.
- ▶ Fonte sonora – excitação acústica causada pela vibração das cordas vocais ou pelo simples fluxo de ar.
- ▶ Operação de filtragem – realizada pelo trato vocal.



Modelo de Engenharia

JPT

6

Frequência Fundamental (F0)

- ▶ É o valor da frequência de vibração das cordas vocais para os sons vozeados.
- ▶ Gamas típicas para F0:

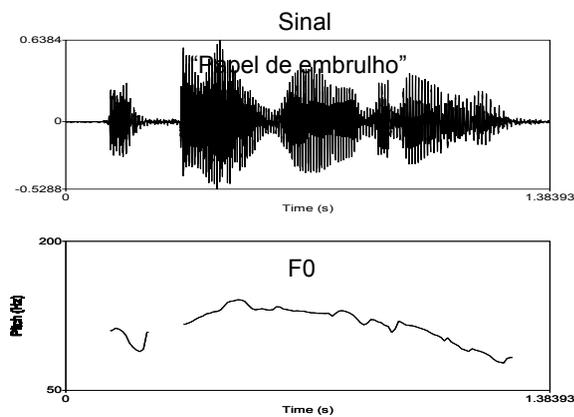
Homens	80 a 200 Hz
Mulheres	90 a 300 Hz
Crianças	200 a 500 Hz

- ▶ Cada orador faz variar este parâmetro ao longo do seu discurso para dar a entoação pretendida.

JPT

7

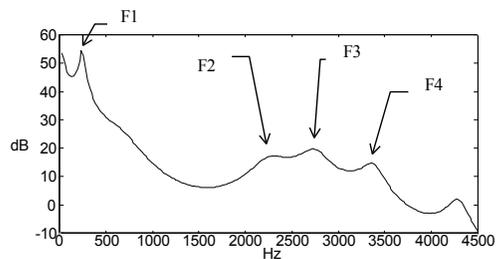
Frequência Fundamental (F0)



JPT

8

Formantes



Espectro alisado de um segmento de som da vogal [i]

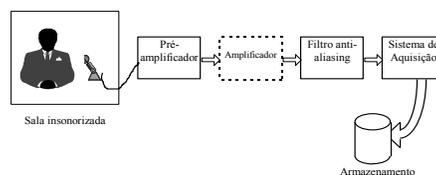
- ▶ Caracterizam o trato vocal.
- ▶ Importa também as Larguras de Banda (a -3 dB).

JPT

9

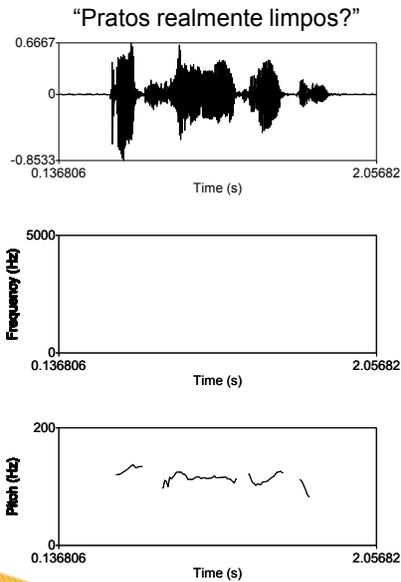
Sinal de Fala

- ▶ Recolha
 - Sala insonorizada/Câmara anecóica
 - Características do Microfone
 - unidirecional
 - multidirecional
 - função de transferência
 - relação sinal/ruído
 - Amostragem/Digitalização
 - Frequência de amostragem - F_a
 - Filtro anti-aliasing
 - estéreo/mono
 - resolução



JPT

10



Identificar:
 silêncio, sons vozeados,
 sons não vozeados,
 oclusão.
 Ver formantes

JPT

11

Média deslizando

$$M(n) = \frac{1}{N+1} \sum_{m=n-N/2}^{n+N/2} |x(m)|$$

- ▶ Corresponde a fazer deslizar uma janela dentro da qual se calcula a média, ao longo de uma sequência de amostras do sinal de fala.
- ▶ Essa janela normalmente é uma janela do tipo janela de Hanning para atenuar o efeito da convolução do sinal janelado com a própria janela, que nas frequências corresponde ao produto das transformadas. Sendo mais habitual a seguinte expressão:

$$M(n) = \frac{1}{N+1} \sum_{m=n-N/2}^{n+N/2} |x(m)| \cdot w(N+1)$$

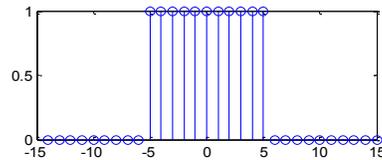
JPT

12

Janelas

- ▶ A $w[n]$ – mais óbvia é a janela rectangular:

$$w_R(n) = u\left[n + \frac{N-1}{2}\right] - u\left[n - \frac{N-1}{2}\right]$$



- ▶ Mas introduz um grande ‘overshoot’ nas discontinuidades de H_d (fenómeno de Gibbs).



JPT

13

Outras Janelas

- ▶ Hanning

$$w_N[n] = \left(0.5 + 0.5 \cos\left(\frac{2\pi n}{N-1}\right)\right) w_R[n]$$

- ▶ Hamming

$$w_M[n] = \left(0.54 + 0.46 \cos\left(\frac{2\pi n}{N-1}\right)\right) w_R[n]$$

- ▶ Blackman

$$w_B[n] = \left(0.42 + 0.5 \cos\left(\frac{2\pi n}{N-1}\right) + 0.08 \cos\left(\frac{4\pi n}{N-1}\right)\right) w_R[n]$$

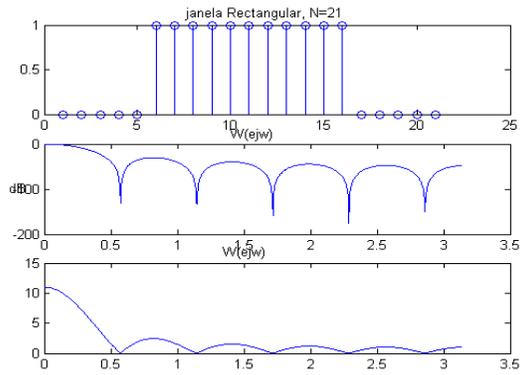


JPT

14

Janelas

► Rectangular

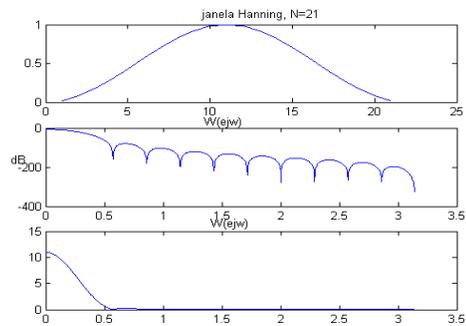


JPT

15

Janelas

► Hanning

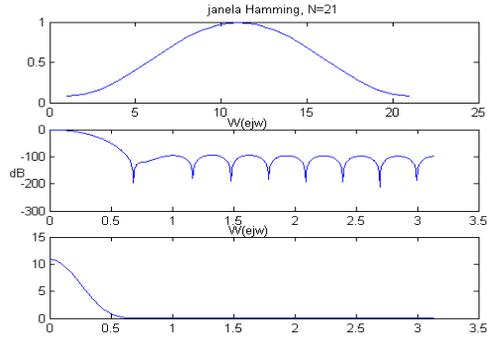


JPT

16

Janelas

► Hamming

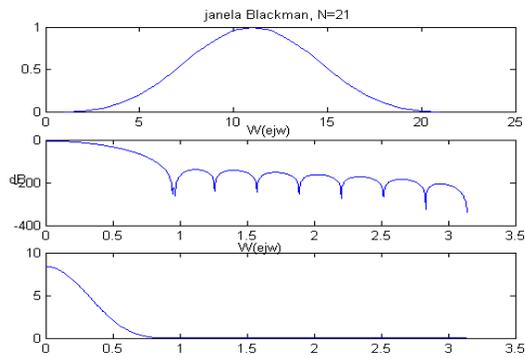


JPT

17

Janelas

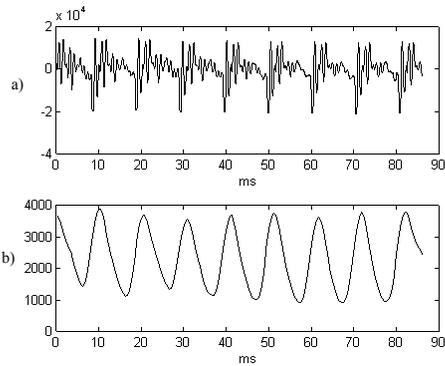
► Blackman



JPT

18

Média deslizante – Determinação de F0

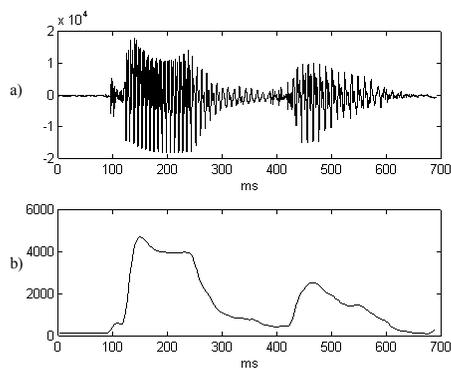


- ▶ Aplicação da média deslizante a um sinal de fala vocalizado correspondente à vogal /a/. Janela deslizante com comprimento 150 e espaçamento entre amostras de 10, para uma Fa de 11025Hz.

JPT

19

Média deslizante – Determinação de sons vozeados



- ▶ Amplitude média deslizante aplicada à palavra “tudo”. Janela de comprimento 350 e espaçamento de 50. Fa de 11025Hz.

JPT

20

Energia Média Deslizante

$$E(n) = \frac{1}{N+1} \sum_{m=n-N/2}^{n+N/2} [x(m) \cdot w(N+1)]^2$$

- ▶ Semelhante à amplitude média deslizante, mas mais adequado para distinguir segmentos de fala vozeada ou não vozeada e silêncio.
- ▶ Um pouco mais pesada computacionalmente que a amplitude média deslizante.
- ▶ Tal como a amplitude média deslizante pode ser determinada com um espaçamento entre amostras.



JPT

21

Aplicações

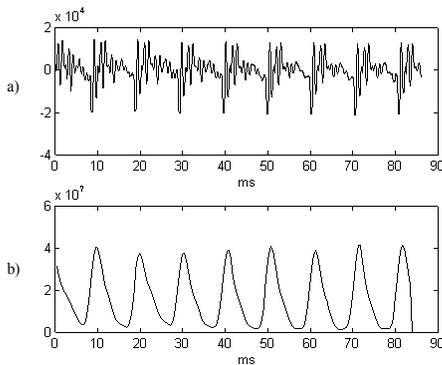
1. Determinação de sincronismo do período fundamental – mais adequada a amplitude média deslizante.
2. Determinação de vozeamento/não vozeamento ou silêncio – mais adequada a energia média deslizante.
3. Determinação de início e fim de palavra – mais adequada a energia média deslizante.



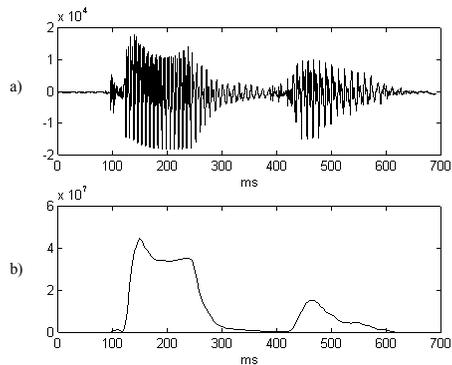
JPT

22

Energia Média Deslizante



EMD - /a/ - janela de 150 amostras e espaçamento de 10 amostras (Fa=11.025KHz)



EMD - /tudu/ - janela de 350 amostras e espaçamento de 50 amostras (Fa=11.025KHz)

JPT

23

Taxa de Passagem por Zero

$$Z(n) = \frac{1}{N+1} \sum_{m=n-N/2}^{n+N/2} |\text{sign}(x(m)) - \text{sign}(x(m-1))| \times w(N+1)$$

Em que

$$\text{sign}(x(m)) = \begin{cases} 1 & \text{se } x(m) \geq 0 \\ -1 & \text{se } x(m) < 0 \end{cases}$$

- ▶ A taxa de passagem por zero do sinal é muito diferente entre segmentos vozeados e segmentos não vozeados. E mesmo em zonas de silêncio. Pelo que esta taxa é um importante contributo para determinar o tipo de segmento.

JPT

24

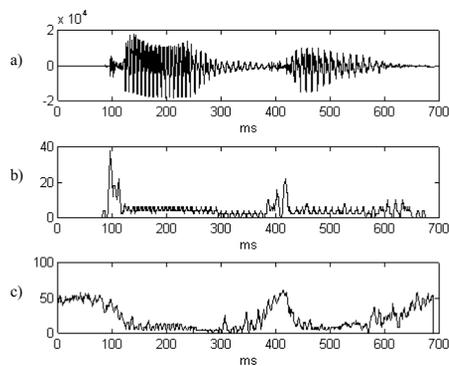
Taxa de Passagem por Zero

- ▶ Como na aquisição do sinal de fala ocorre, por vezes um pequeno *offset* que faz alterar a razão de passagens por zero do sinal, é sugerido que se use um filtro passa alto com uma frequência de corte aos 70 Hz para eliminar esse offset.
- ▶ Eu sugiro que a taxa de passagem por zero seja aplicada não ao sinal directamente mas à sua derivada $d(n)$, em que $d(n)=x(n)-x(n-1)$.

JPT

25

Taxa de Passagem por Zero



- a) locução de /tudu/
- b) tpz aplicada ao sinal
- c) tpz aplicada à derivada do sinal

É possível identificar as zonas de silêncio, sons vozeados e não vozeados

JPT

26

Trabalho

- ▶ Desenvolver um algoritmo que determine a AMD, EMD e TPZ a um sinal de fala dado e represente o seu resultado. O comprimento da janela e o espaçamento devem poder ser alterados pelo utilizador.
- ▶ Aplicar os algoritmos da AMD e EMD a um sinal ECG.
- ▶ Aplicar os algoritmos AMD, EMD e TPZ a um sinal de Fala.