



ALGORITMOS PARA A CLASSIFICAÇÃO AUTOMÁTICA DE ÁUDIO EM VÍDEO

Por Maria Isabel
isabel@dsc.ufcg.edu.br

1

SUMÁRIO

- Motivação
- Etapas da implementação
 - 1) Pré- Processamento
 - 2) Processamento do Sinal
 - 3) Extração de Características
 - 4) Treinamento e Classificação
- Considerações Finais

MOTIVAÇÃO

1. Este trabalho consiste em implementar um sistema automático de classificação do áudio.
2. A classificação manual do áudio é bastante trabalhosa.

ETAPAS DE DESENVOLVIMENTO DO SISTEMA

- 1) Pré- Processamento
 - a) Catálogo manual
- 2) Processamento do Sinal
 - a) Pré-Ênfase
- 3) Extração de Características
- 4) Treinamento e Classificação

PRÉ-PROCESSAMENTO

- Catalogação manual de 2 horas de áudio.
 - Classificação em Fala, Música e Outros.

PROCESSAMENTO DO SINAL

- Esta etapa inclui aquisição do sinal, pré-ênfase, janelamento.
- Aquisição: O ideal seria que os sons tivessem sido coletados em um ambiente relativamente silencioso e sem ruídos.
- O áudio foi retirado usando o programa VirtualDub;
- Formato dos vídeos utilizados: WAV

PROCESSAMENTO DO SINAL

- Taxa de Amostragem: 44Hz (dois canais)
- Resolução: 16 bits
- Janelamento do sinal: Janela de **Hamming** com superposição de 50%.
- Tamanho dos **segmentos** : 20 ms.

PROCESSAMENTO DO SINAL

○ Segmentação:

- Divisão do sinal em segmentos, selecionados por janelas com duração definida.

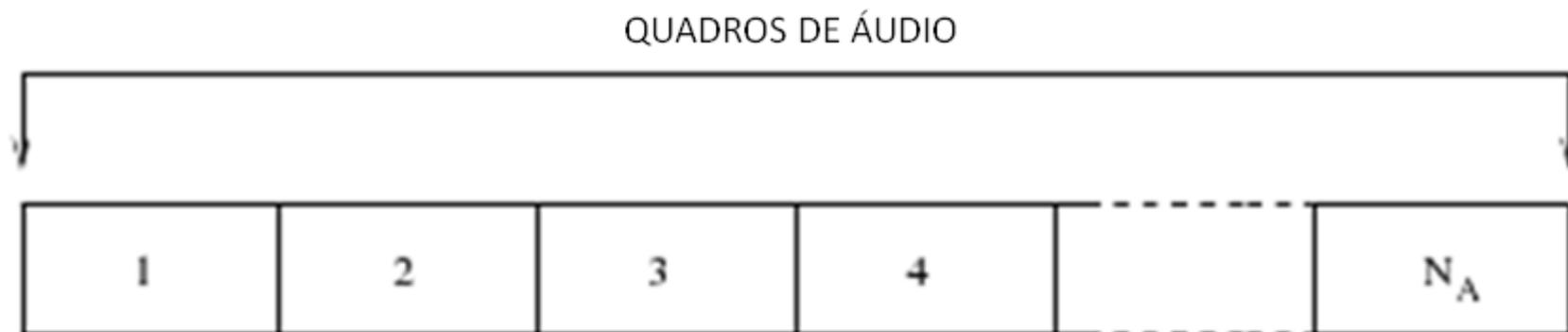


Figura 1: Sinal de áudio segmentado

PROCESSAMENTO DO SINAL

- Os segmentos são escolhidos dentro dos limites de estacionariedade do sinal.
 - Sinais estacionários
 - Sinais não-estacionários

PROCESSAMENTO DO SINAL

- Tipos de Janelamento:
 - Janelamento Retangular
 - **Janelamento de Hamming**
 - Janelamento de Hanning

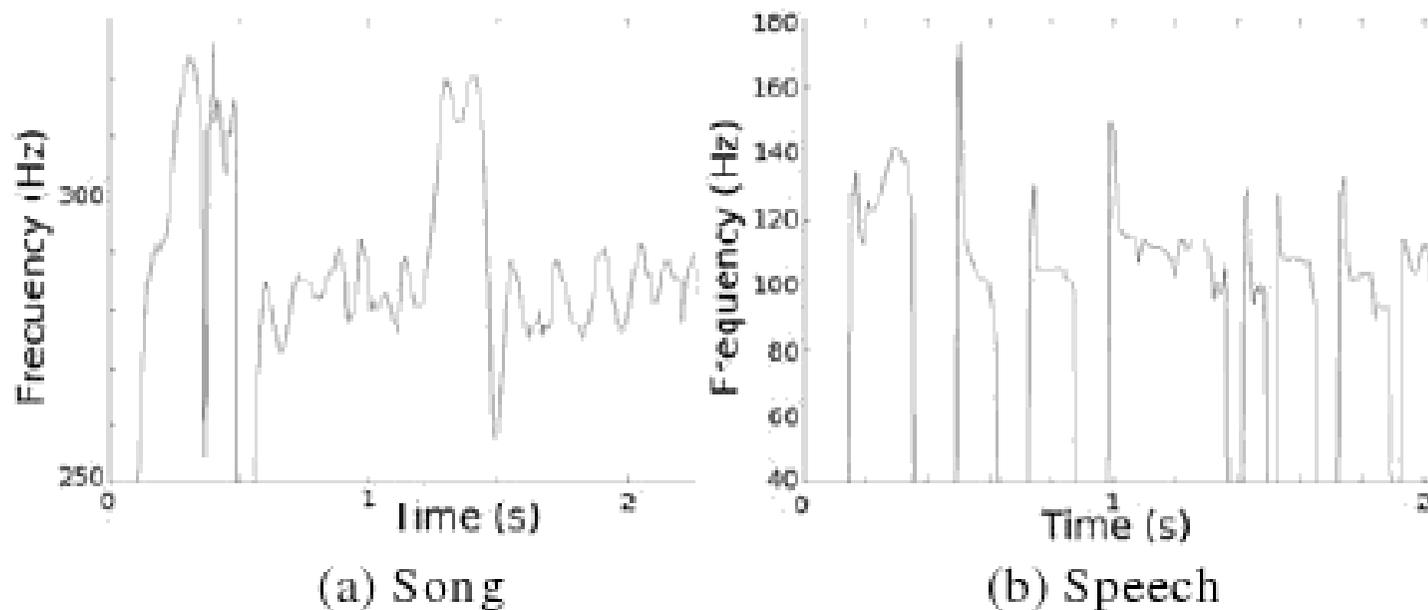
- Por que usamos Janelamento de Hamming?

EXTRAÇÃO DE CARACTERÍSTICAS

- **Vibrato;**
- **Coeficiente Harmônico;**

VIBRATO

É a variação da frequência de um instrumento ou da voz.



Frequência Fundamental (Hz) x Tempo (s)

VIBRATO

- Como medir isso?
- A função de entrada recebe um vetor com as frequências fundamentais e sobre elas é aplicada a Transformada de Fourier.

- Transformada de Fourier:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

$$f(t) = \mathcal{F}^{-1}(F(\omega)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega.$$

COEFICIENTE HARMÔNICO

O espectro é dividido em três séries trigonométricas, a idéia é buscar a mais importante e medir a sua importância, o coeficiente harmônico.

A voz cantada tem maior coeficiente harmônico maior que a fala.

Pode ser calculado, usando a combinação espectral e temporal:

COEFICIENTE HARMÔNICO

Correlação Temporal:

$$R^T(\tau) = \frac{\sum_{n=0}^{N-\tau-1} [\bar{s}(n) \cdot \bar{s}(n + \tau)]}{\sqrt{\sum_{n=0}^{N-\tau-1} \bar{s}^2(n) \cdot \sum_{n=0}^{N-\tau-1} \bar{s}^2(n + \tau)}}$$

Correlação Espectral:

$$R^S(\tau) = \frac{\sum_{\omega=0}^{N-\omega_\tau-1} [S(\omega) \cdot S(\omega + \omega_\tau)]}{\sqrt{\sum_{\omega=0}^{N-\omega_\tau-1} S^2(\omega) \cdot \sum_{\omega=0}^{N-\omega_\tau-1} S^2(\omega + \omega_\tau)}}$$

COEFICIENTE HARMÔNICO

Combinação das autocorrelações:

$$R(\tau) = \beta \cdot R^T(\tau) + (1 - \beta)R^S(\tau)$$

Maximizandoc $R(\tau)$, encontraremos o coeficiente harmônico:

$$H_\alpha = \max_{\tau} R(\tau) = R\left(\frac{1}{F_0}\right)$$

CONSIDERAÇÕES FINAIS

Além das características citadas, ainda há outras bastante robustas, como coeficientes LPC, Cepstral e Mel cepstral;

Há muitos trabalhos realizados sobre o tema.

Taxa de acerto: 63% (ainda muito baixa)

REFERÊNCIAS

- Tese de Joseana Macedo Fechine, disponível em: <http://www.dsc.ufcg.edu.br/~joseana/>
- Singing Voice Characterization for Audio Indexing, disponível em: <http://www.eurasip.org/Proceedings/Eusipco/Eusipco2007/Papers/c31-d02.pdf>
- Wikipédia, disponível em: http://pt.wikipedia.org/wiki/Transformada_r%C3%A1pida_de_Fourier

DÚVIDAS?



Algoritmos para a classificação de áudio em Vídeo
Ciclo de Seminários Técnicos

OBRIGADA!

