



MESTRADO ENGENHARIA INFORMÁTICA E MULTIMÉDIA

PROCESSAMENTO DE FALA

Vocoder

DATA: MAY 8, 2022

Docente:

Eng. Carlos Meneses

Realizado por:

Mihail Ababii - 46435

Índice de Conteúdos

| | |
|--------------------------------------|-----------|
| Lista de Figuras | ii |
| 1 Introdução | 1 |
| 2 Produção de fala | 1 |
| 3 Tratamento do sinal | 1 |
| 3.1 Obtenção dos Formantes | 2 |
| 3.2 Pré-ênfase | 2 |
| 4 Vocoder | 3 |
| 5 Codificação | 4 |

Lista de Figuras

| | | |
|---|---|---|
| 1 | Normalização da trama | 1 |
| 2 | Normalização da trama | 1 |
| 3 | Vocoder | 3 |
| 4 | Pulso glotal e respetiva derivada | 4 |
| 5 | Quantificação Uniforme | 4 |

1 Introdução

Este trabalho consiste na desconstrução de um sinal de voz, nas em componentes que possibilitem uma reconstrução o mais próximo possível do original, e a respetiva reconstrução através de um vocoder. Assim este relatório visa apresentar e explicar os diversos passos a tomar e os cuidados a tomar ao realizar este tipo de codificação.

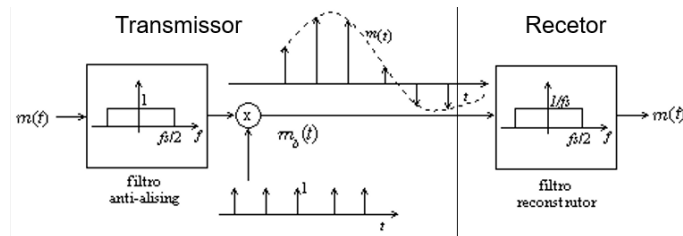


Figure 1: Normalização da trama

2 Produção de fala

O ar que sai dos pulmões quando expiramos passa pela laringe onde se se situam as pregas vocais, que podem ou não vibrar, criando assim, sons vosiados ou não vosiados. Sons vozeados (sonoros) produzidos com vibração das pregas vocais (quase-periódicos); Sons não vozeados (surdos) produzidos sem vibração das pregas vocais (ruidosos);

3 Tratamento do sinal

Antes de se começar a trabalhar com as tramas que constituem o sinal a enviar é importante normalizar as mesmas removendo a sua componente DC e aplicando uma janela de hamming.

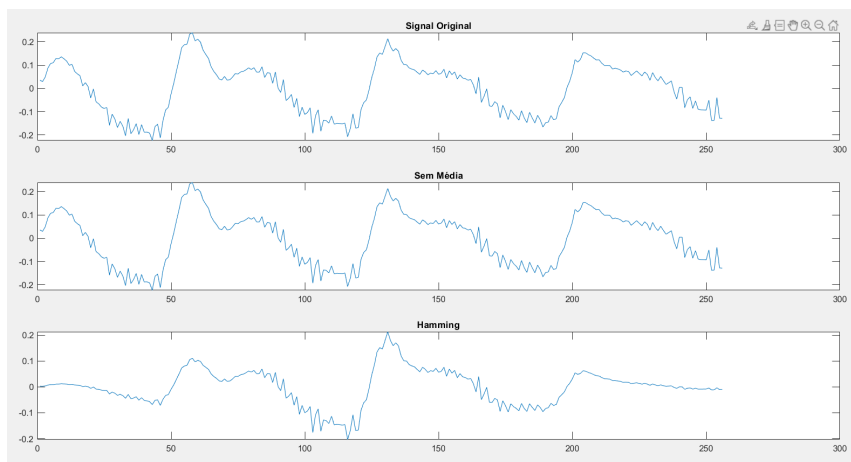


Figure 2: Normalização da trama

3.1 Obtenção dos Formantes

Numa segunda fase é necessário realizar uma estimação do pitch. Tal como indicado previamente, a autocorrelação de sinais vozeados é quase periódicos. A autocorrelação é a correlação cruzada de um sinal com o ele próprio. É uma ferramenta matemática para encontrar padrões de repetição, tal como a presença de um sinal periódico obscurecidos pelo ruído, ou para identificar a frequência fundamental em falta num sinal implícita pelas suas frequências harmónicas. Assim, caso, o valor máximo da autocorrelação normalizada for superior a um dado limiar (≈ 0.45), a trama é considerada como vozeada. É necessário procurar pelo máximo, apenas após alguns milissegundos, pois o primeiro máximo é a energia do sinal, sendo este sempre 1. É também importante referenciar que é procurado o primeiro máximo, ao invés do máximo absoluto.

O método da autocorrelação tem inconvenientes:

- Janelas de dimensão pelo menos duas vezes o máximo período
- Diminuição do número de parcelas do somatório com o aumento do atraso k , com as seguintes repercussões
 - diminuição da periodicidade da função de autocorrelação;
 - maior sensibilidade às variações entre períodos;
 - maior sensibilidade ao ruído;
 - enfatiza o primeiro formante;

3.2 Pré-ênfase

As vibrações criadas pela voz viajam pelo meio aereo, que é inerentemente um meio barulhento e como ruído tem distribuição triangular, o que significa que o ruído é mais alto em frequências mais altas, é necessario realizar um pré-ênfase para aumentar o nível de altas frequências sobre o piso de ruído.

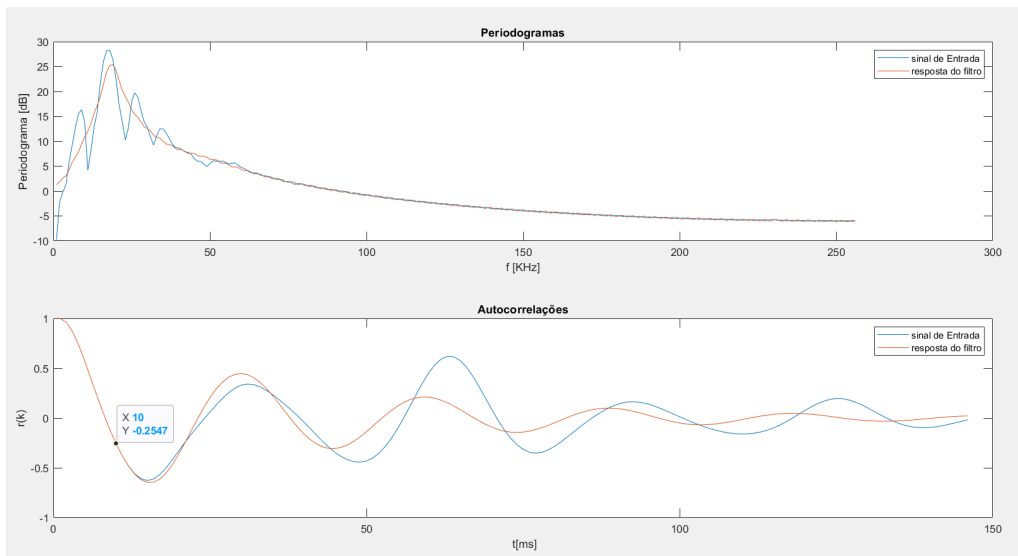
Após a obtenção e pré-ênfase do pitch, é necessário estimar a autocorrelação sobre o novo sinal, obtendo um sistema com p incógnitas, cuja solução é representada pela equação abaixo.

$$\begin{bmatrix} R(0) & R(1) & R(2) & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & R(p-2) \\ R(2) & R(1) & R(0) & \dots & R(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ R(p-1) & R(p-2) & R(p-3) & \dots & R(0) \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_p \end{bmatrix} = - \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ \dots \\ R(p) \end{bmatrix}$$

Após ter os a_k e a autocorrelação da trama, é possível calcular o ganho de predição, por forma a impor que a energia da trama é igual à da resposta impulsiva, ou seja, a resposta impulsiva do filtro até à ordem p , tem a mesma autocorrelação que a do sinal original.

$$G = \sqrt{R_s[0] + \sum_{k=1}^p a_k R_s[k]}$$

Assim sendo, com os dados adquiridos, é possível utilizar os mesmos para obter a resposta impulsional.



Tal como podemos observar nos gráficos acima podemos observar que o periodograma da resposta impulsional ainda que não exata aproxima-se ao periodograma do sinal de entrada, assumindo a mesma forma. É também importante verificar que a autocorrelação de ambos os sinais coincidem até pelo menos p , sendo neste caso $p=10$, verificando que os valores obtidos são os desejados.

4 Vocoder

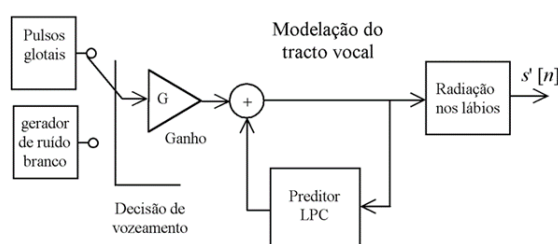


Figure 3: Vocoder

O vocoder é responsável pela reconstrução do sinal a partir dos dados previamente adquiridos. É primeiramente necessário que gerar um pulso glotal a partir do pitch, sendo que para compensar a radiação dos lábios, é utilizado a derivada do pulso glotal.

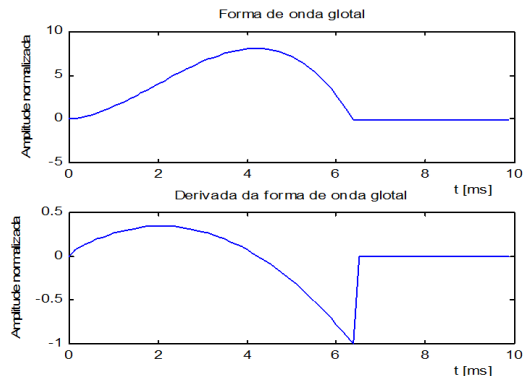


Figure 4: Pulso glotal e respetiva derivada

Ao utilizar a derivada dos pulsos glotais, os ganhos de predição e os ak 's é possível reconstruir o sinal original, com menos qualidade, mas perceptível e sem muita distorção. Visto que cada trama é constituída por múltiplos pulsos decidiu-se realizar uma interpolação dos ganhos, por forma a suavizar o sinal de saída.

5 Codificação

Este processo está muitas vezes associado com a armazenagem dos audios, de forma mais compacta, pelo que é importante codificar e armazenar os dados adquiridos, neste caso num ficheiro binário.

Como o pitch pode variar entre 20 e 146 ou 0 para o caso de a trama não ser vozeada, ao subtrair 19 os valores do pitch passam a variar entre 1 e 127, permitindo uma binarização direta para 7 bits.

Para quantificar os ak 's é utilizado o LSF – Line Spectrum Frequencies, que são usadas para representar coeficientes de predição linear (LPC) para transmissão em um canal. Os LSPs têm várias propriedades (por exemplo, menor sensibilidade ao ruído de quantização) que os tornam superiores à quantização direta de LPCs.

Para quantificar o ganho, é utilizada uma quantificação uniforme, onde são calculados valores de decisão, valores entre os quais se decide por um valor de quantificação, e de quantificação, valor discreto que a amostra quantificada pode tomar, a amostra é aproximada por um valor de quantificação. Sendo a codificação realizada para 5 bits.

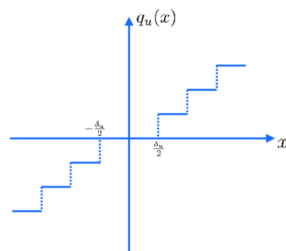


Figure 5: Quantificação Uniforme

Após quantificar os dados e guardar os mesmos num ficheiro binário, é necessário reverter a quantificação dos mesmos e atribuí-los ao vocoder, que irá reconstruir o sinal. Ao ouvir este sinal é possível perceber que sofreu algumas perdas devido à quantificação, no entanto a qualidade, continua bastante próxima do original. Mais importante, é o facto de o tamanho do ficheiro binário ser consideravelmente inferior ao ficheiro áudio original, demonstrando a capacidade de compressão das técnicas utilizadas.