

Processamento de Voz

V. C. Parro e-mail: vparro@maua.br

30 de março de 2020

Objetivos Processamento de Voz

Construção de um sistema de tratamento e reconhecimento de voz fundamentado em análise de Fourier, vogais e medidas de distância entre formantes.

1 Introdução

A decomposição de um sinal complexo em uma soma de sinais como forma de análise é amplamente utilizada na engenharia. Nesta experiência trataremos uma aplicação da decomposição espectral ou decomposição de Fourier aplicada à sinais de voz. Nos limitaremos a um conjunto de sons com o objetivo de ilustrarmos uma aplicação e, derivarmos a partir daí, de forma sistêmica, como seriam aplicações mais elaboradas.

A representação de um sinal complexo pode ser feita utilizando-se uma representação ou uma codificação adequada. A figura 1 ilustra a partitura da introdução da música *I can't get no satisfaction* da banda inglesa *Rolling Stones*. A partitura neste caso trata-se de uma codificação dos sons da música, resultado de uma ANÁLISE feita pelo compositor. Por outro lado, qualquer músico devidamente capacitado pode decodificar estes símbolos e sintetizar a música emitindo os sons correspondentes.

A idéia de análise e síntese ilustrada no exemplo anterior, está intrinsecamente colocada na transformada (análise) e anti-transformada (síntese) de Fourier. A proposta feita



Figura 1: Trecho da partitura de "*I can't get no satisfaction*" Jagger & Richards. Exemplo de codificação de elementos sonoros.

para a partitura pode ser ampliada para outros sinais, desde que sejamos capazes de identificar o conjunto de símbolos que os representam. Este mesmo conceito é encontrado em algoritmos derivados da análise por LPC (*Linear Predictive Coding*)(?),(?). Se achou interessante procure conhecer como funciona o skype ou o sistema celular no que tange a transmissão de Voz.

2 Leitura

Uma das propostas deste trabalho é proporcionar a possibilidade de aplicações para os assunto estudados. Neste sentido é importante que o estudante tenha autonomia para ler e interpretar um texto técnico. Além das referências indicadas no final deste texto, dois outros são indicados como base:

- Vogais no Brasil: Características dos sons das vogais do português falado no Brasil - Inatel.
- Estimação de formantes: <http://biochaves.com/metodo-de-estimacao-de-formantes-por-preditor-linear-de-ordem-variavel/>

3 O trabalho

1. A turma deve coletar 3 amostras aleatórias, para cada aluno de cada uma das 5 vogais (duração estimada de 2 segundos para cada amostra).
2. Os resultados devem ser compartilhados entre as duplas para formar um banco de dados de sinais. Sugere-se a criação de um repositório comum (one drive, por exemplo).
3. Para que seja simples a manipulação dos arquivos sugere-se uma nomenclatura que possa ser facilmente indexada (**RA a 1.wav -RA vogal index.tipo**, por exemplo).
4. **Análise temporal:** para cada vogal deve ser estudado sinal no tempo. Você consegue observar uma periodicidade? **(0.5 ponto)**
5. **Análise de Fourier:** para cada vogal deve ser estudada a série de Fourier. Você consegue justificar sua resposta anterior pela análise de Fourier?**(0.5 ponto)**
6. **Formantes:** estudando os textos indicados na Seção 2 determine as formantes para cada vogal e plote em um gráfico bi dimensional, onde: eixo X - formate $f_1[Hz]$ e formante $f_2[Hz]$. **(2.0 pontos)**.
7. **Detector automático:** crie uma função que detecta automaticamente as duas primeiras formantes de uma vogal. **(2.0 pontos)**
8. **Reconhecimento de vogais:** estabeleça um critério de como identificar automaticamente duas vogais. Para testar esta parte, solicite a pessoas próximas que pronunciem as vogais, na mesma condição do item 01 e verifique o nível de acerto de seu algoritmo (3.0 pontos).

9. **Reconhecimento de uma palavra:** considerando, por exemplo, a palavra Mauá, elabore um software que seja capaz de reconhecer a sequência de vogais. O resultado esperado desse software é: a + u + a e indique: 3 vogais, para o exemplo indicado. (2. pontos)

🌀 Entrega 🌀

- O trabalho deverá ser feito em dupla e entregue em código do tipo Matlab/Octave (**main.m**) pelo Moodle até 15 de Junho de 2020.
- O código deverá ser comentado e executado a partir de um exemplo *default* de arquivo ou arquivos de som. Todos os arquivos devem estar comprimidos e os respectivos caminhos resolvidos. O arquivo **main.m** deve resolver internamente todas as pendências.
- Não é necessário relatório, o código deve ser suficiente para representar o que foi executado.

Referências

- [1] Rabiner, L. Schafer, R.W. ? Digital Processing of Speech Signals ? Prentice Hall, 1978.
- [2] Rabiner, L. Juang, B. ? Fundamentals of Speech Recognition ? Prentice-Hall, 1993.