Modelização harmónica precisa de sons vozeados por humanos

Francisca Vieira de Brito

up201404169@fe.up.pt

Orientador: Prof. Dr. Aníbal João de Sousa Ferreira

15 de julho de 2019

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

ESTRUTURA DA APRESENTAÇÃO

1. **Introdução**
2. **Análise microscópica**
3. **Análise macroscópica**
4. **Análise, modelização e síntese harmónica**
5. **Transformação intencional das microvariações da frequência fundamental de sinais de voz falada**
6. **Conclusões e trabalho futuro**

1. INTRODUÇÃO

Enquadramento e objetivos

- **Projeto FCT "DyNaVoiceR"**

  - reconstrução da voz disfónica: conversão de voz sussurrada em voz artificial preservando a identidade sonora do próprio orador

  - facilitará a comunicação humano-com-humano e humano-com-máquina: discurso mais inteligível, natural e característico

- **Objetivo do trabalho desenvolvido**

  - modelização precisa das características da voz vozeada que refletem o funcionamento das pregas vogais, designadamente ao nível das microvariações da frequência fundamental (F0)

  - caracterizar, quer a sua importância ao nível da identidade sonora de um orador, quer o impacto sonoro que decorre se esta for alterada de forma intencional, por exemplo, por aplanamento ou por implantação das características de um outro orador

2. ANÁLISE MICROSCÓPICA

- **Estimação precisa da frequência de componentes sinusoidais individuais**

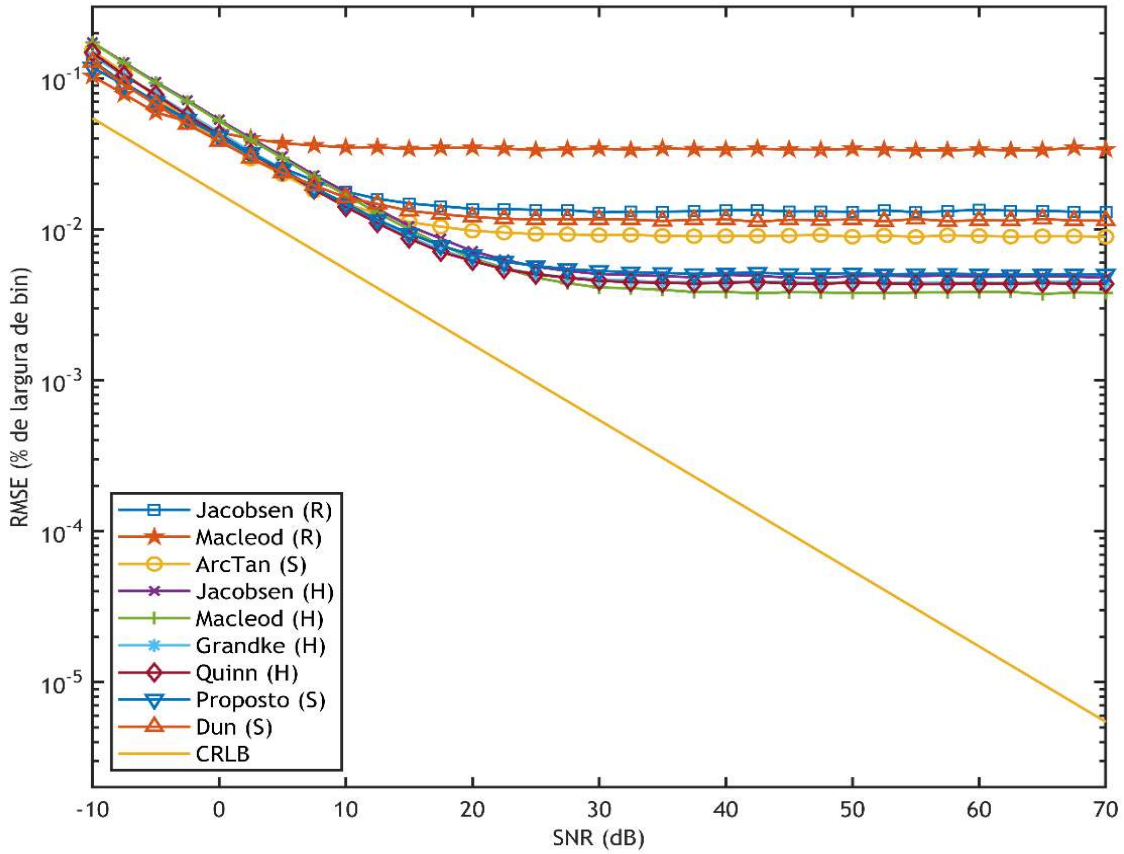


Figura 2 - RMSE (em % de largura de bin normalizada) dos oito estimadores de frequência abordados em função do SNR, quando o sinal sinusoidal é afetado por interferência de quatro harmónicos. O CRLB está também representado como referência.

3. ANÁLISE MACROSCÓPICA

- **Estimação precisa da frequência fundamental de sinais de fala que seguem uma estrutura harmónica**

Algoritmo de estimação da frequência fundamental	Erro relativo (%) para sinal FM sintético sem interferências harmónicas	Erro relativo (%) para sinal FM sintético com interferências harmónicas
Boersma (PRAAT)	0.12682 %	0.054429 %
YIN	0.14608 %	0.081841 %
SWIPE'	0.26859 %	0.74468 %
SearchTonal	0.066656 %	0.064038 %

Tabela 1 - Erros relativos de estimação de F0 obtidos com os diferentes métodos para sinais FM sintéticos sem e com interferências harmónicas, não afetados por ruído.

Algoritmo de estimação da frequência fundamental	Erro relativo (%) para sinal FM sintético sem interferências harmónicas	Erro relativo (%) para sinal FM sintético com interferências harmónicas
Boersma	0.12685 %	0.12686 %
YIN	0.14614 %	0.14609 %
SWIPE'	0.27162 %	0.27256 %
SearchTonal	0.32106 %	0.3219 %

Tabela 2 - Erros relativos de estimação de F0 obtidos com os diferentes métodos para sinais FM sintéticos sem e com interferências harmónicas, afetados por ruído.

Análise e modelização harmónica

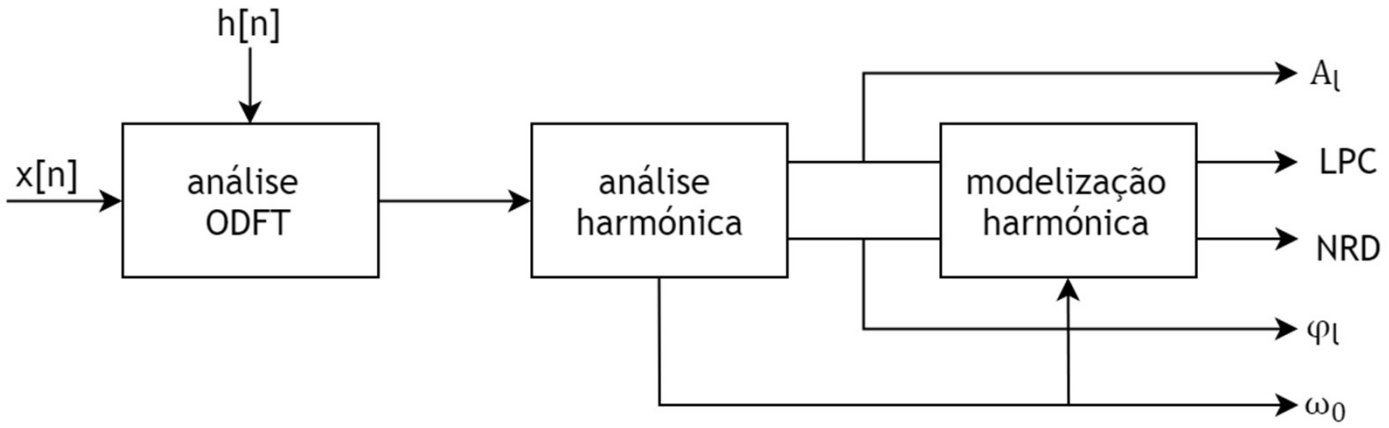


Figura 3 - Diagrama de blocos da análise e modelização harmónica de sinais de voz que contêm uma estrutura harmónica.

Síntese harmónica no domínio das frequências

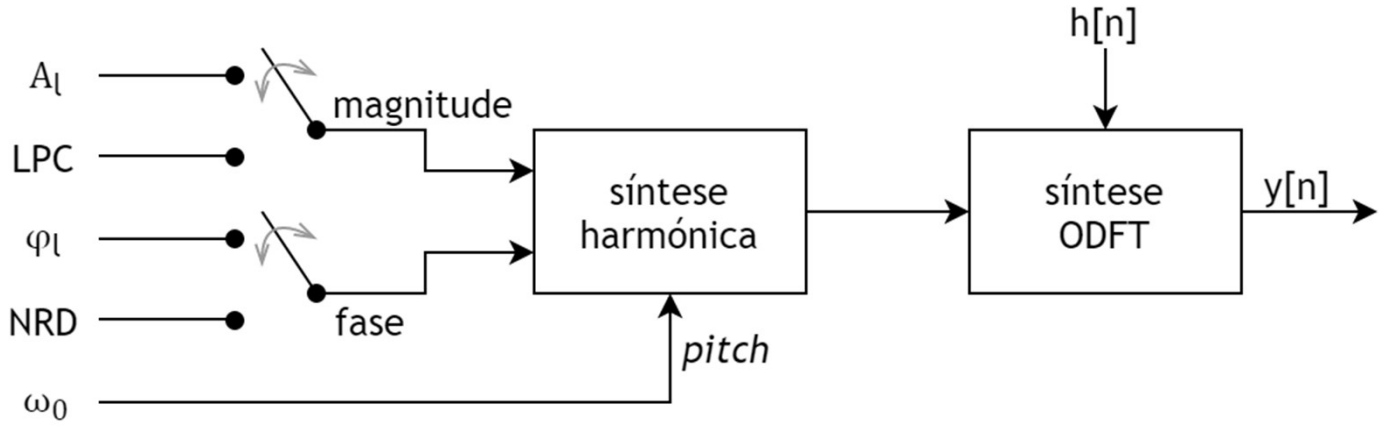


Figura 4 - Diagrama de blocos da síntese harmónica baseada em segmentos, utilizando os parâmetros exatos de magnitude (A_l) e fase (φ_l) ou, alternativamente, as aproximações correspondentes dadas pelo modelo harmónico LPC e pelo modelo NRD invariante ao deslocamento.

4. ANÁLISE, MODELIZAÇÃO E SÍNTESE HARMÓNICA

- **Cenários de teste de síntese harmónica no domínio das frequências**
  - síntese utilizando as magnitudes e fases harmónicas exatas (tais como estimadas), ou seja, utilizando os parâmetros A_l e φ_l
  - síntese utilizando o modelo harmónico LPC e as fases harmónicas exatas (tais como estimadas) φ_l
  - síntese utilizando o modelo harmónico LPC e o modelo NRD médio do orador
  - síntese utilizando o modelo harmónico LPC, o modelo NRD médio do orador e fase sintética para a frequência fundamental

- A utilização dos modelos paramétricos de magnitude e fase, modelos LPC e NRD, respetivamente, não compromete a qualidade auditiva do sinal processado

5. TRANSFORMAÇÃO INTENCIONAL DAS MICROVARIAÇÕES DE F0 DE SINAIS DE VOZ FALADA

Transformações de sinal realizadas

- aplanamento das microvariações da frequência fundamental
- implantação dos padrões extraídos de um orador num outro

Sinais de voz falada utilizados

- 8 sinais com duração de 1.45 segundos
- vogais orais sustentadas /a/ da palavra “água” e /i/ da palavra “ilha”
- produzidas por dois oradores do género feminino (SPF1 e SPF2) e dois oradores do género masculino (SPM1 e SPM2)

Participantes nos testes de perceção auditiva

- 17 ouvintes: 10 do género feminino e 7 do género masculino
- idade média de 31 anos (mínimo de 16 anos e máximo de 58 anos)

Aplanamento das microvariações da frequência fundamental

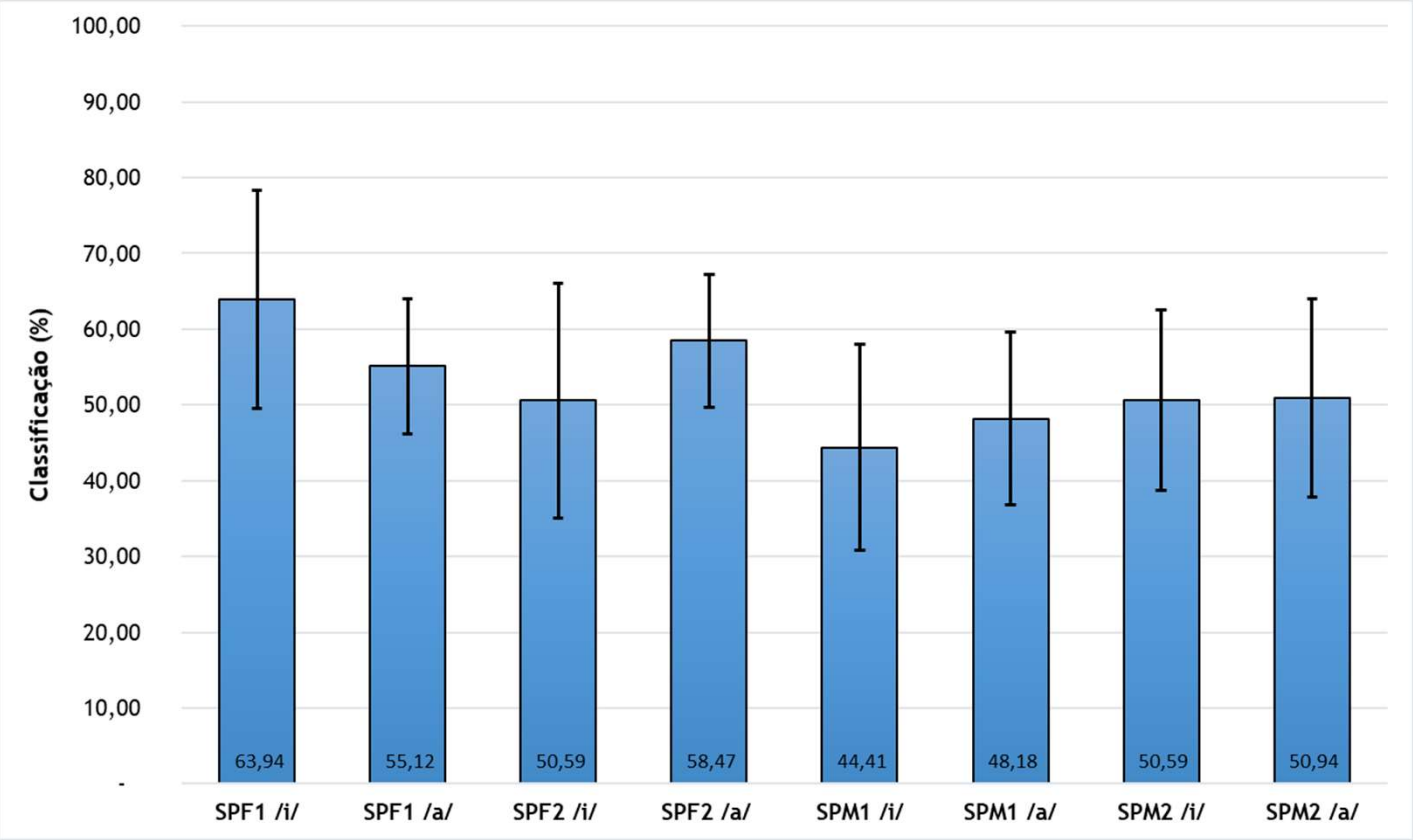


Figura 5 - Médias e intervalos de confiança (95%) dos resultados de avaliação subjetiva.

- Definir todo o trajeto de F0 como sendo o seu valor médio
- Avaliação das diferenças audíveis entre os sinais originais e as respetivas versões modificadas

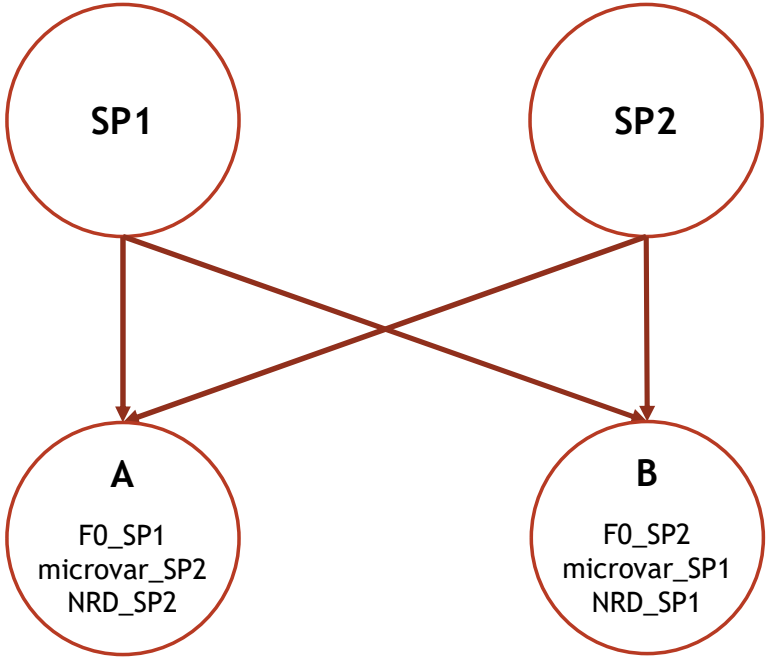
Escala de classificação percetual

- impercetíveis (80% a 100%);
- percetíveis, não descaracterizadoras (60% a 80%);
- ligeiramente descaracterizadoras (40% a 60%);
- descaracterizadoras (20% a 40%);
- muito descaracterizadoras (0% a 20%).

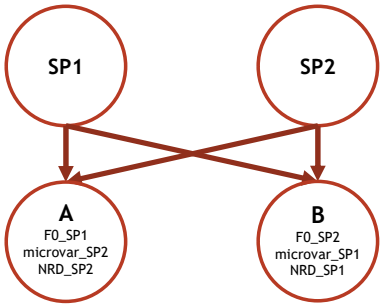
Implantação dos padrões extraídos de um orador num outro

- Implantação das microvariações de F0 e do modelo de fase NRD médio de um outro orador na F0 média de um orador
  - os sinais correspondem à **mesma vogal** e os **oradores são do mesmo género**

- Avaliação do quanto as características dos sons originais dos oradores são percetíveis nas versões modificadas (A e B)



5. TRANSFORMAÇÃO INTENCIONAL DAS MICROVARIAÇÕES DE F0 DE SINAIS DE VOZ FALADA



Escala de classificação percetual

- totalmente percetíveis, sons iguais (80% a 100%);
- muito percetíveis (60% a 80%);
- percetíveis (40% a 60%);
- pouco percetíveis (20% a 40%);
- nada percetíveis, sons totalmente diferentes (0% a 20%).

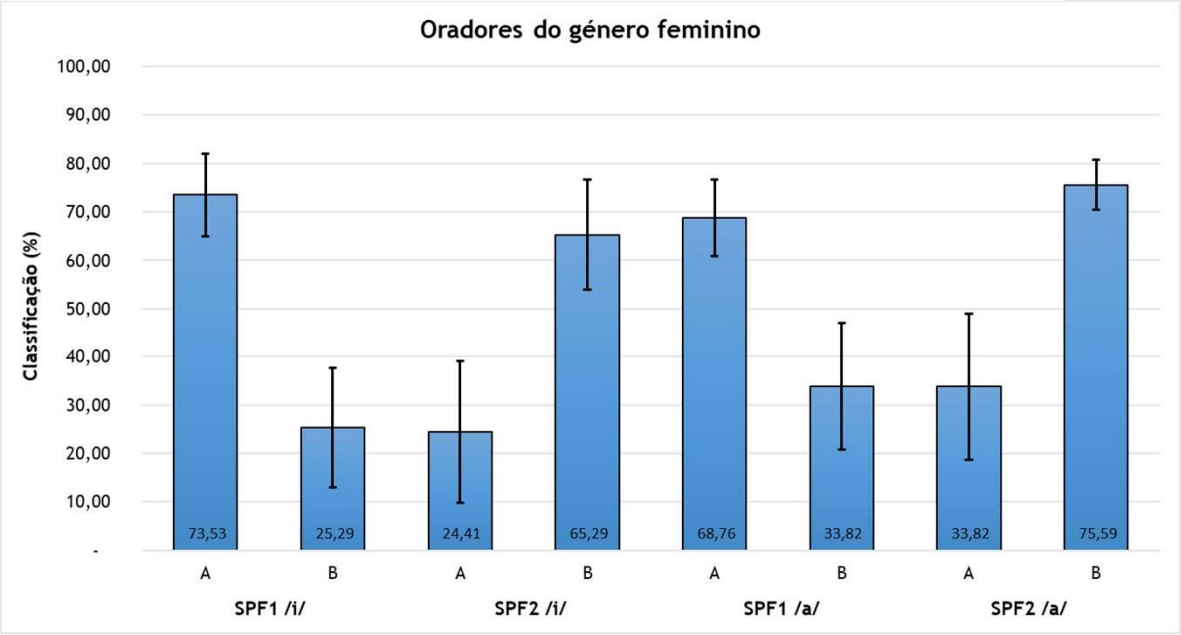


Figura 6 - Médias e intervalos de confiança (95%) dos resultados de avaliação subjetiva para os oradores do género feminino.

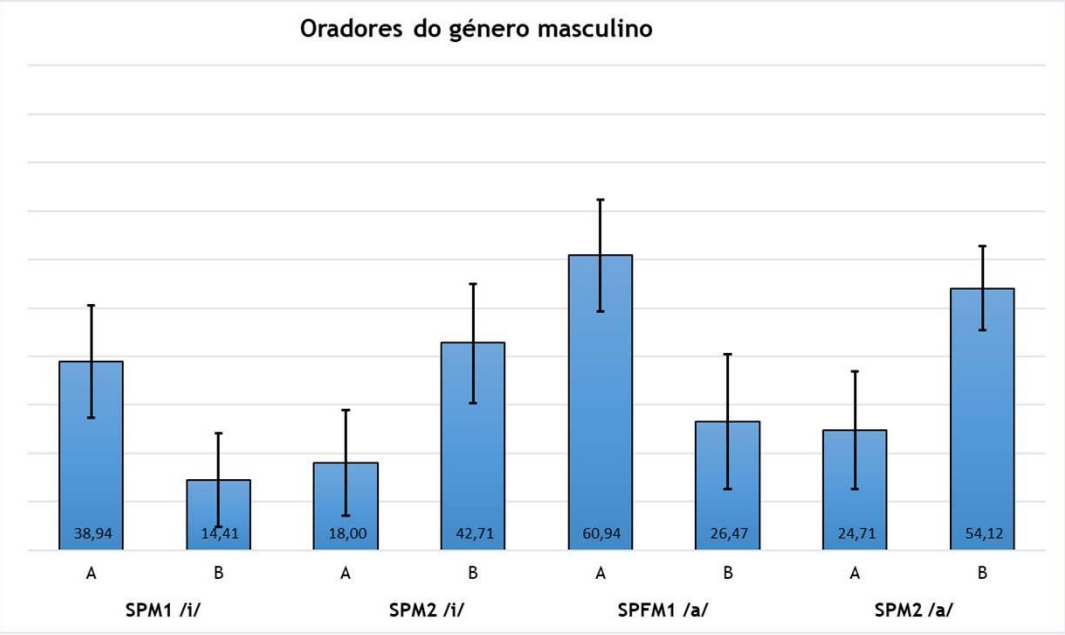


Figura 7 - Médias e intervalos de confiança (95%) dos resultados de avaliação subjetiva para os oradores do género masculino.

6. CONCLUSÕES E TRABALHO FUTURO

Conclusões do trabalho

- **Análise microscópica de estimação precisa da frequência de componentes sinusoidais individuais**

  - o estimador proposto tem o melhor desempenho entre todos os estimadores que utilizam a janela sinusoidal

- **Análise macroscópica de estimação precisa da frequência fundamental de sinais de fala que seguem uma estrutura harmónica**

  - o algoritmo SearchTonal tem um bom desempenho na estimação da frequência fundamental de estruturas harmónicas e faz parte de uma *framework* que permite realizar a análise microscópica, a análise macroscópica, bem como a modelização e síntese harmónica, pelo que será o algoritmo utilizado

- **Análise, modelização e síntese harmónica**

  - os modelos paramétricos de magnitude e fase, modelos LPC e NRD, respetivamente, podem ser utilizados na síntese harmónica de sinais no domínio das frequências, pois não comprometem a qualidade auditiva dos sinais processados

- **Transformação intencional das microvariações da frequência fundamental de sinais de voz falada**

  - a transformação intencional das microvariações da frequência fundamental de sinais de voz falada por aplanamento destas retira naturalidade à voz

  - a frequência fundamental média da voz é uma característica percetual predominante na captação da atenção do ouvido humano, pelo que as microvariações da frequência fundamental não são capazes de modificar a assinatura sonora de um dado orador quando implantadas sobre o valor médio da frequência fundamental de um outro orador

6. CONCLUSÕES E TRABALHO FUTURO

Trabalho futuro

- **Efetuar melhorias no algoritmo SearchTonal**

  - pois este é induzido em erro quando existe presença de ruído

- **Realizar um estudo que abranja todas as vogais produzidas por todos os oradores da base de dados da qual o projeto FCT dispõe**

  - de modo a se poder realizar uma avaliação mais alargada do impacto que tem a transformação intencional das microvariações da frequência fundamental de sinais de voz falada

- **Realizar os mesmos testes para palavras em contexto real**

  - utilizar de vogais em contexto de fala natural, isto é, vogais retiradas de palavras, pois a utilização de vogais sustentadas isoladas representa um cenário ideal na análise das transformações de sinal descritas e analisadas anteriormente

FIM

Obrigada pela atenção!