

Universidade Estadual Paulista
Instituto de Biociências, Letras e Ciências Exatas
Departamento de Ciência da Computação e
Estatística

Luis Fernando Teixeira Silva

Um sistema para reconhecimento de comandos falados
dependente do locutor

São José do Rio Preto - SP

2017

Luis Fernando Teixeira Silva

Um sistema para reconhecimento de comandos
falados dependente do locutor

Monografia apresentada ao Programa de
graduação em Ciência da Computação da
UNESP para obtenção do título de Bacharel.

Orientador: Prof. Dr. Rodrigo Capobi-
anco Guido

São José do Rio Preto - SP

2017

Ficha catalográfica elaborada pelo Serviço de Biblioteca do IBILCE/UNESP

Luis Fernando Teixeira Silva

titulo

titulo

titulo. / fulano de tal; orientador

Rodrigo Capobianco Guido. São José do Rio Preto, 2017.

xxx p.

Monografia (TCC

TCC

TCC, 2017.

1. Processamento de sinais. 2. Reconhecimento de locutor. 3. Acústica. 4. Energia. 5. Escala *Bark*. I. Capobianco Guido, Rodrigo, orient.

II. Título.

Dedico este trabalho a todos os meus familiares, em especial aos meus pais, Nilda, Luis Carlos e a minha irmã Ana Beatriz.

Dedico também esse trabalho para a minha namorada Cristiana Luiza.

Agradecimentos

Primeiramente gostaria de agradecer a meus pais e minha madrinha, pois sem o apoio deles, eu não conseguiria entrar nessa linda universidade.

Agradeço todos os meus familiares que me apoiaram ao longo de toda a minha jornada na faculdade, e principalmente a minha irmã, que nos momentos mais difíceis me apoiou a dar continuidade ao curso.

Gostaria de agradecer também minha namorada por ter me auxiliado nesses dois últimos anos de universidade e por me dar forças a concluir o curso nessa etapa final.

“No fim tudo dá certo, e se não deu certo é porque ainda não chegou ao fim.”

Fernando Sabino

Resumo

TAL, F. *titulo*. 2016. xxxp. TCC UNESP 2016.

Atualmente,

Palavras-chave: Processamento de sinais. Reconhecimento de locutor. Acústica. Escala *Bark*.

Abstract

TAL, F. *titulo*. 2016. xxxp. TCC UNESP 2017.

Nowadays, ...

Keywords: Signal processing. Speaker recognition. Acoustics. Bark scale.

Lista de Figuras

Lista de Tabelas

Tabela 2.1 - Estrutura de um arquivo <i>WAVE</i>	28
--	----

Lista de Abreviaturas

ABFF	Academia Brasileira de Fonoaudiologia Forense
CHMM	<i>Continuous Hidden Markov Models</i>
DTW	<i>Dynamic Time Warping</i>
DWT	<i>Discrete Wavelet Transform</i>
EER	<i>Equal Error Rate</i>
FIR	<i>Finite Impulse Response</i>
GMM-UBM	<i>Gaussian Mixture Models-Universal Background Models</i>
HMM	<i>Hidden Markov Model</i>
IEEE	<i>Institute of Electrical and Eletronic Engineers</i>
IIR	<i>Infinite Impulse Response</i>
LPCC	<i>Linear Predictive Cepstrum Coefficients</i>
MFCC	<i>Mel Frequency Cepstral Coefficients</i>
MLP	<i>Multiple Layer Perceptron</i>
NIST	<i>National Institute of Standards and Technology</i>
PCM	<i>Pulse Code Modulation</i>
RNA	Rede Neural Artificial
SVM	<i>Support Vector Machine</i>
TIMIT	<i>Texas Instruments and Massachusetts</i>
TTVNN	<i>Transição entre Trechos Vozeados e Não-Vozeados</i>
VQ	<i>Vector Quantization</i>
WAVE	<i>Waveform</i>
WPT	<i>Wavelet Packet Transform</i>
ZCR	<i>Zero Crossing Rate</i>

Sumário

1	Introdução	23
1.1	Introdução	23
1.2	Objetivo	23
1.3	justificativa	23
1.4	Motivação	24
1.5	Metodologia	24
1.6	Exequibilidade	25
1.7	Organização do trabalho	25
2	Revisão Bibliográfica	27
2.1	Fundamentação da Verificação de Locutores	27
2.2	Arquivos Acústicos no Formato <i>WAVE</i>	27
2.3	Energia	28
2.4	Vetores de Características	28
2.5	Níveis Críticos de Energia	29
3	Detalhamento do Trabalho Proposto	31
3.1	Considerações iniciais	31
4	Testes e Resultados	33
5	Conclusões e Trabalhos Futuros	35
	Referências	35
	Apêndice I - Gráficos das características extraídas	39

Capítulo 1

Introdução

1.1 Introdução

Petry (1) define o reconhecimento de locutores como...

1.2 Objetivo

Este trabalho tem como objetivo implementar um algoritmo computacional desenvolvido em C/C++ para reconhecer comandos falados de modo *off-line* com locutor prédefinido.

1.3 justificativa

justificativa...

1.4 Motivação

1.5 Metodologia

Para a elaboração deste projeto foi determinado os seguintes 11 comandos:

- Bom dia, Logan;
- Bom noite, Logan;
- Oi, Logan;
- Como está o tempo hoje?;
- vai chover?;
- Abrir calculadora;
- Ver notícias;
- Pesquisar;
- Alarme;
- Calendário;
- Sair;

sendo que posteriormente foi realizada a gravação de 10 áudios para cada um dos 11 comandos referidos, totalizando 110 arquivos de áudio no formato MPEG-4. Tais arquivos foram convertidos para o formato textitwave usando o programa *Audacity*. Vale ressaltar que todos os áudios foram gravados em um ambiente que proporciona-se um certo grau de isolamento sonoro, para assim se obter um som com menos ruído.

A partir dessa etapa inicial foi feita a extração dos dados brutos contidos nos arquivos *wave*. Para isso foi utilizada uma biblioteca fornecida pelo Prof.Dr.Rodrigo Capobianco Guido do Departamento de Ciência da Computação e Estatística (DCCE), IBILCE/Unesp. Essa biblioteca tem a função de separar o cabeçalho dos arquivos *wave*. A partir desse ponto, a biblioteca foi modificada para extrair os dados brutos dos arquivos e guarda-los em arquivos de texto.

1.6 Exequibilidade

Exequibilidade...

1.7 Organização do trabalho

O texto vindouro do presente trabalho está organizado da seguinte forma:

- No Capítulo 2 apresenta-se uma série de trabalhos publicados envolvendo a área de reconhecimento de locutores, mostrando como são inúmeras as possibilidades de se realizar essa tarefa. Expõe-se, também, os principais conceitos e teorias que estão relacionados com o trabalho que foi desenvolvido.

Capítulo 2

Revisão Bibliográfica

2.1 Fundamentação da Verificação de Locutores

Reconhecimento de locutores (??)...

2.2 Arquivos Acústicos no Formato *WAVE*

waveform audio file format é a abreviação de *WAVE* ou simplesmente *WAV*, que é um tipo de formato de arquivo de áudio que foi desenvolvido pela *Microsoft* em conjunto com a *IBM*. O formato *WAVE* é amplamente utilizado em uma variedade de trabalhos, sejam eles científicos ou profissionais, visto que o formato permite uma fiel representação dos dados digitalizados, uma vez que os dados digitalizados podem ser armazenados sem sofrer obrigatoriamente um processo de compressão, o que evita perdas. Porém, devido a essa característica o *WAV* ocupa muito mais espaço que os demais formatos de arquivos de áudios.

A Tabela 2.1 mostra a estrutura de arquivo *WAVE*.

Nesse trabalho foi utilizado o formato *WAV* de 16 bits PCM (*Pulse-code Modulation*) que não utiliza compressão, para se obter assim uma melhor qualidade na elaboração deste projeto

final.

Tabela 2.1 – Estrutura de um arquivo *WAVE*

Classe	Posição (bytes)	Tamanho (bytes)	Descrição
Cabeçalho	0	4	Apresenta o identificador do cabeçalho - "RIFF".
Cabeçalho	4	4	Tamanho do arquivo sem o identificado do cabeçalho.
Cabeçalho	8	4	Mostra o identificador <i>WAVE</i> .
Formato	12	4	Mostra o identificador do segundo bloco - "fmt".
Formato	16	4	Tamanho do bloco sem o identificador.
Formato	20	2	Mostra se o arquivo é do tipo PCM ou tem alguma compressão.
Formato	22	2	Mostra a quantidade de canais.
Formato	26	4	Apresenta o valor da taxa de amostragem.
Formato	30	4	Apresenta a taxa de bytes.
Formato	32	2	Demonstra a quantidade de bytes para uma amostra.
Formato	34	2	Demonstra a quantidade de bits para cada amostra.
Dados	36	4	Apresenta o identificador do terceiro bloco - "data".
Dados	40	4	Mostra o tamanho do bloco sem o identificador.
Dados	44	4	Demonstra os dados reais da música.

2.3 Energia

A energia...

2.4 Vetores de Características

Vetor de característica...

2.5 Níveis Críticos de Energia

Níveis de energia são...

Capítulo 3

Detalhamento do Trabalho Proposto

3.1 Considerações iniciais

Como primeira etapa para

Capítulo 4

Testes e Resultados

bla bla bla...

Capítulo 5

Conclusões e Trabalhos Futuros

Neste trabalho, ...

Referências

- 1 PETRY, A. *Reconhecimento automático de locutor utilizando medidas de invariantes dinâmicas não-lineares*. 2002. 155 p. Tese (Doutorado em Ciência da Computação)-Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.
 - 2 CAMPBELL, J. P. et al. Forensic speaker recognition: a need for caution. *IEEE Signal Processing Magazine*, v. 26, n. 2, p. 95-103, 2009. doi:10.1109/msp.2008.931100.
 - 3 ACADEFORD. Disponível em: <<http://www.acadefor.com.br/>>. Acesso em: 12 ago. 2014.
- numérico. São Paulo: Pearson Prentice Hall, 2007.

Apêndice I - Gráficos das características extraídas