Aplicação do Modelo de Misturas Gaussianas para a Classificação de Vogais

Andréia Seixas Leal[1], Leonardo Carneiro de Araujo[2]

***Resumo*—Este artigo apresenta um aplicação do modelo de misturas gaussianas para o reconhecimento das vogais do português brasileiro. Foram realizadas gravações de vogais pronunciadas por diferentes locutores Estas foram pré-processadas e as frequências extraídas. Um modelo de misturas gaussianas foi utilizado para agrupar as vogais, criando assim um classificador de vogais orais do Português Brasileiro.**

Palavras-Chave—Modelo de Mistura Gaussiana, EM, Reconhecimento de Vogais, Reconhecimendo de Voz.

***Abstract*— This article presents an application of the Gaussian Mixture Model for recognizing vowels of Brazilian Portuguese. By recording the signal voice of different speakers, training the signal and extracting the formants, a system with a mix of vowels was created. The goal is to get a similar distribution of the mixture using GMM and to make the recognition of vowels.**

Keywords—Gaussian Mixture Model (GMM), EM, Vowels Recognition, Speech Recognition.

I. INTRODUÇÃO

Um sistema de reconhecimento automático de fala independente do locutor é um sistema capaz de extrair informações necessárias de um sinal acústico de fala a fim de reconhecer as palavras pronunciadas por um locutor qualquer. Uma tarefa no âmbito de um sistema de reconhecimento de fala é identificar vogais e classificá-las.

O sinal de fala consiste em uma onda acústica produzida por movimentos voluntários das estruturas anatômicas ligadas ao trato vocal. O sinal de fala é analisado, sob o ponto de vista linguístico, como uma sequência de fones que, a grosso modo, são classificados como vozeados. Os sons vozeados, como os das vogais, são produzidos por pulsos de ar quase-periódicos gerando uma excitação que ressoará no trato vocal. Os picos de ressonância são conhecidos como formantes e usualmente denotados por Fi onde i é o i-ésimo formante. Os formantes carregam a principal informação na caracterização da vogal e portanto serão aqui utilizados na tarefa de classificação das vogais orais do Português Brasileiro (PB).

No PB existem sete (7) vogais orais em posição tônica: / a , e , ɛ , i , o , ɔ , u / [4]. Este trabalho tem como propósito realizar a classificação automática destas vogais utilizando para tanto os formantes extraídos de amostras de áudio gravadas de diferentes locutores.

Para o processamento do sinal de fala é necessária a captura, análise e representação das características do sinal de voz. Neste trabalho foi utilizada análise baseada no filtro LPC, *Linear Predictive Coding*. Para um trecho de sinal de áudio será encontrado o filtro LPC associado e extraídas as frequências de ressonância do filtro, que serão as frequências dos formantes.

O modelo de mistura gaussiana (*Gaussian Mixed Model*, GMM) é muito utilizado em reconhecimento de voz para realizar a inferência estatística e modelar dados. Iremos utilizar o GMM como modelo acústico subjacente para o sistema, no qual as características extraídas são modeladas como uma função densidade de probabilidade em mais de uma dimensão. No caso em questão, a dimensionalidade será determinada pelo número de formantes considerados. Os parâmetros que descrevem a mistura gaussiana são encontrados pelo algorítimo *Expectation Maximization* (EM) de forma iterativa.

II. ETAPA DE TREINO E CLASSIFICAÇÃO

Para tornar o modelo prático, no treino das vogais foram utilizados cinco diferentes locutores, para cada uma das cinco vogais, gravados com taxa de amostragem de 44kHz. Depois os sinais foram subamostrados em 8kHz e amostras de silêncio foram retiradas.

Na análise da voz, para extração dos formantes, é feito o procedimento de aquisição até a extração dos parâmetros do sinal utilizando LPC. A representação do primeiro e segundo formantes extraídos das vogais do primeiro locutor está representada na figura 1.

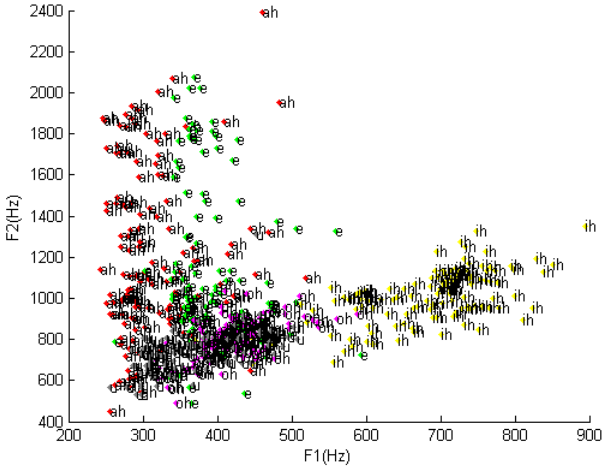

Fig. 1. Primeiro e Segundo formantes das vogais ‘ah’(amarelo), ’eh’(verde), ’ih’(vermelho), ’oh’(lilás), ’u’(cinza).

Feito isto, os dados e suas classificações prévias estão definidos e ao utilizar o método EM, a melhor mistura de gaussianas pode ser encontrada, representada figura 2.

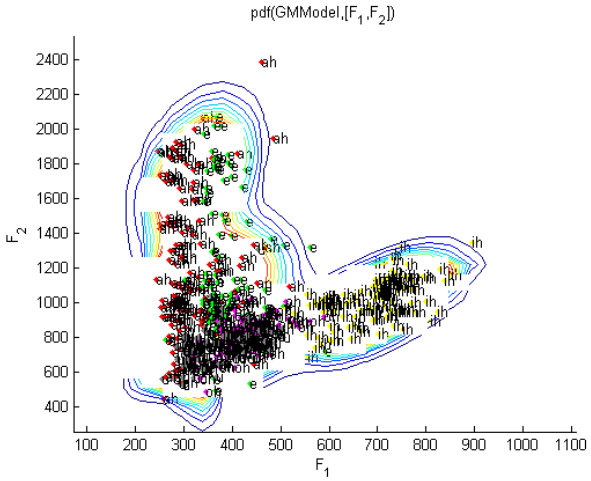

Fig. 2. Primeiro e Segundo formantes das vogais obtidos pela GMM.

O mesmo foi feito comparando o segundo e o terceiro formantes dos dados coletados dos locutores. Com estes resultados é possível verificar o percentual de acertos dos formantes obtidos pelo modelo com relação ao sinal de voz original do locutor.

III. CONCLUSÕES

Esta versão preliminar dos estudos, que foram recém-iniciados, permite a implementação de um sistema simplificado de classificação de vogais em reconhecimento automático de voz. Entretanto, já se identifica que amostras de sinais de voz com frequências muito próximas não nos possibilita resultados muito precisos e um maior número de locutores e amostras é o ideal para melhores resultados.

As diferenças do modelo implementado com relação aos dados podem estar correlacionadas com a utilização do LPC, em oposição a um método mais direto.

AGRADECIMENTOS

Agradeço ao meu professor orientador Leonardo Carneiro de Araújo pela paciência e à Coordenação Técnica do SBrT2015 pela iniciativa.

REFERÊNCIAS

[1] L. Rabiner e B.H. Juang, *Fundamentals of Speech Recognition*.PTR Prentice-Hall International, 1993.

[2] X. Huang, A. Acero e H.W. Hon, *Spoken Language Processing: a guide to theory, algorithm, and system development.* Prentice-Hall, 2001.

[3] L. Rabiner, *Introduction to Digital Speech Processing*. Now Publishers, 2007.

[4] J.M.Câmara Jr. *Estrutura da lingua portuguêsa.* Petrópolis: Vozes, 1994.

[1]Andréia Seixas Leal, Graduanda em Engenharia de Telecomunicações, Universidade Federal de São João del-Rei, Campus Alto Paraopeba, Ouro Branco-MG, Brasil, E-mail: andreia_sl@msn.com.

[2]Leonardo Carneiro de Araujo, Departamento de Engenharia de Telecomunicações e Mecatrônica, Universidade Federal de São João del-Rei, Campus Alto Paropeba, Ouro Branco-MG, Brasil, email: leolca@ufsj.edu.br