

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
COORDENAÇÃO DE ENGENHARIA ELETRÔNICA
CURSO DE ENGENHARIA ELETRÔNICA

HELIO RODRIGUES DA SILVA
JULIANO RODRIGUES DOURADO

**RECONHECIMENTO AUTOMÁTICO DE LOCUTOR UTILIZANDO
MODELO DE MISTURAS GAUSSIANAS TREINADO PELO
ALGORITMO DE MAXIMIZAÇÃO DA EXPECTATIVA**

TRABALHO DE CONCLUSÃO DE CURSO

TOLEDO
2018

HELIO RODRIGUES DA SILVA
JULIANO RODRIGUES DOURADO

**RECONHECIMENTO AUTOMÁTICO DE LOCUTOR UTILIZANDO
MODELO DE MISTURAS GAUSSIANAS TREINADO PELO
ALGORITMO DE MAXIMIZAÇÃO DA EXPECTATIVA**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Eletrônica da Universidade Tecnológica Federal do Paraná - UTFPR Campus Toledo, como requisito parcial para a obtenção do título de Bacharel em Engenharia Eletrônica.

Orientador: Alberto Yoshihiro Nakano
Universidade Tecnológica Federal do Paraná

TOLEDO
2018



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Toledo
Coordenação do Curso de Engenharia Eletrônica



TERMO DE APROVAÇÃO

Reconhecimento Automático de Locutor Utilizando Modelo de Misturas Gaussianas Treinado pelo Algoritmo de Maximização da Expectativa Nº 74

Reconhecimento automático de locutor utilizando modelo de misturas gaussianas treinado pelo algoritmo de maximização da expectativa

por

HELIO RODRIGUES DA SILVA
JULIANO RODRIGUES DOURADO

Esse Trabalho de Conclusão de Curso foi apresentado às **08h30 do dia 27 de junho de 2018** como **requisito parcial** para a obtenção do título de **Bacharel em Engenharia Eletrônica**. Após deliberação da Banca Examinadora, composta pelos professores abaixo assinados, o trabalho foi considerado **APROVADO**.

Jefferson Gustavo Martins
Universidade Tecnológica Federal do Paraná

Daniel Cavalcanti Jeronymo
Universidade Tecnológica Federal do Paraná

Alberto Yoshihiro Nakano
Universidade Tecnológica Federal do Paraná

Prof. Dr. Fábio Rizental Coutinho
Universidade Tecnológica Federal do Paraná

O termo de aprovação assinado encontra-se na coordenação do curso

Toledo, 27 de junho de 2018

Dedicamos este trabalho às nossas famílias que nos apoiaram continuamente para que esta etapa de aprendizagem fosse concluída.

AGRADECIMENTOS

Ao nosso orientador Prof. Dr. Alberto Yoshihiro Nakano por sua dedicação e orientação durante todo o desenvolvimento deste trabalho.

RESUMO

Desenvolver sistemas que possam reconhecer ou identificar indivíduos vem se tornando uma necessidade cada vez maior em aplicações que exigem a verificação e a garantia da identidade humana. Há vários sistemas que utilizam a biometria para reconhecer um determinado indivíduo, dentre estes, o reconhecimento automático de locutor que utiliza a fala como dado de reconhecimento. Neste trabalho, os parâmetros acústicos mel-cepstrais foram extraídos para modelagem de locutores por meio de ferramentas estatísticas. Para realizar a modelagem do trato vocal de um indivíduo, utilizou-se o modelo de misturas gaussianas (GMM, do inglês *Gaussian Mixture Model*). Os parâmetros do modelo GMM foram adaptados ou treinados pelo algoritmo de maximização da expectativa (EM, do inglês *Expectation Maximization*). Sendo assim, foram criados 40 modelos dos 40 locutores na etapa de treinamento e em seguida testados. Os testes realizados forneceram os resultados de máxima verossimilhança para a construção de matrizes de classificação. Por fim, em diversas aplicações práticas o reconhecimento de locutor se mostra promissor, compreendendo tarefas que possam vir a facilitar, agilizar e melhorar processos de verificação de identidade.

Palavras-chave: Reconhecimento de Locutor. Coeficientes Mel-Cepstrais. Modelo de Misturas Gaussianas. Maximização da Expectativa.

ABSTRACT

Developing systems that can recognize or identify individuals has becoming a growing need in applications that require verification and grant of human identity. There are several systems that use biometrics to recognize a certain individual, among them, automatic speech recognition that uses speech as recognition data. In this work, Mel Frequency Cepstral Coefficients were extracted to model speakers using statistical tools. For the modeling of the vocal tract of an individual, the Gaussian Mixture Model (GMM) was used. The parameters of the GMM model were adapted or trained by the Expectation Maximization algorithm (EM). Thus 40 models of 40 speakers were created and then tested. Speaker recognition is promising for many practical applications, including tasks that can facilitate, speed up and improve identity verification processes.

Keywords: Speech Recognition. Mel Frequency Cepstral Coefficients. Gaussian Mixtures Model. Maximizing Expectation.

LISTA DE FIGURAS

Figura 1 – Aparelho fonador.	5
Figura 2 – Representação em diagrama de blocos do comportamento de um trato vocal.	5
Figura 3 – Segmentação do sinal mostrando a sobreposição em quadros de 50%.	8
Figura 4 – Janelas Hamming, Hanning e Blackman.	8
Figura 5 – Diagrama de blocos ilustrando as etapas para extração dos atributos acústicos.	9
Figura 6 – Relação de frequência na escala MEL.	10
Figura 7 – Exemplo ilustrando o modelo de misturas de gaussianas unidimensional.	12
Figura 8 – Diagrama de blocos representando a sequência de atividades do projeto.	18

LISTA DE QUADROS

Quadro 1 – Classificação nBest - Sistema 1	39
Quadro 2 – Classificação nBest - Sistema 2	40
Quadro 3 – Classificação nBest - Sistema 3	41
Quadro 4 – Classificação nBest - Sistema 4	42
Quadro 5 – Percentuais de Classificação - Sistema 1	43
Quadro 6 – Percentuais de Classificação - Sistema 2	44
Quadro 7 – Percentuais de Classificação - Sistema 3	45
Quadro 8 – Percentuais de Classificação - Sistema 4	46

LISTA DE TABELAS

Tabela 1 – Matriz <i>n-best</i>	20
Tabela 2 – Matriz de confusão (%)	20

LISTA DE ABREVIATURAS E SIGLAS

DCT	<i>Discrete Cosine Transform</i>
EM	<i>Expectation Maximization</i>
GMM	<i>Gaussian Mixture Model</i>
MFCC	<i>Mel Frequency Cepstral Coefficients</i>
SLIT	Sistema linear invariante no tempo
SNR	<i>Signal-to-noise ratio</i>
A/D	Conversor Analógico Digital

LISTA DE SÍMBOLOS

\mathbf{X}	Conjunto de dados de entrada
Hz	hertz
kHz	quilo-hertz
n	Índice de tempo discreto
$p[n]$	Trem de pulsos gerado durante a passagem do fluxo de ar pela glote
$r[n]$	Ruído branco gaussiano gerado na glote
$x[n]$	Amostras dos sinal de voz
f_s	Taxa de amostragem ou frequência de amostragem
f_{max}	Frequência máxima de um sinal analógico
f_n	Frequência de Nyquist
z	Variável discreta (domínio da transformada z)
ms	milisegundos
$X[k]$	Transformada de Fourier Discreta da sequência de comprimento finita
N	Comprimento de cada quadro
f	Frequência do sinal
$\mu_{\mathbf{k}}$	Vetor de médias
D	Dimensão do vetor de médias
$\Sigma_{\mathbf{k}}$	Matriz de covariâncias
$\pi_{\mathbf{k}}$	Peso da k -ésima gaussiana em um modelo de mistura
\mathbf{x}	Vetor de coeficientes MFCCs
$P_r(\lambda_s \mathbf{X})$	Probabilidade <i>a posteriori</i> de um modelo gerar o conjunto de dados
\mathbf{X}	Conjunto de dados de entrada
$p(\mathbf{X} \lambda_s)$	Função densidade de probabilidade dos dados de entrada pertencerem a um modelo
λ_s	Modelo do locutor s
$p(\mathbf{X})$	Função densidade de probabilidade dos dados de entrada
$P_r(\lambda_s)$	Probabilidade de um dado locutor
\mathbf{x}_t	Dados de treinamento
\hat{s}	Representação de um locutor
γ_{tk}	Normalização da função densidade de probabilidade multidimensional

SUMÁRIO

1	INTRODUÇÃO	1
2	OBJETIVOS	2
2.1	OBJETIVO GERAL	2
2.2	OBJETIVOS ESPECÍFICOS	2
3	JUSTIFICATIVA	3
4	REVISÃO DA LITERATURA	4
4.1	Aparelho fonador	4
4.2	Amostragem de um sinal	6
4.3	Definições	6
4.3.1	Sinais determinísticos e estocásticos	6
4.4	Pré-processamento de um sinal	7
4.4.1	Pré-ênfase	7
4.4.2	Segmentação em quadros	7
4.4.3	Janelamento de um sinal	7
4.5	Parâmetros característicos	8
4.6	Procedimentos para a extração dos parâmetros	9
4.7	Reconhecimento de locutor utilizando modelo de misturas de gaussianas	11
4.7.1	Algoritmo Expectation Maximization	15
5	MATERIAIS E MÉTODOS	17
5.1	Procedimentos gerais	17
6	ANÁLISE E DISCUSSÃO DOS RESULTADOS	20
6.1	Aplicações	23
7	CONCLUSÃO	24
7.1	TRABALHOS FUTUROS	24
	Referências	25
	Apêndices	26
	APÊNDICE A TERMO DE CONFIDENCIALIDADE	27

APÊNDICE B	MODELO TCLE TCUIV	28
APÊNDICE C	PROJETO DETALHADO	33
APÊNDICE D	USO DAS INSTALAÇÕES	38
APÊNDICE E	Classificação nBest - Sistema 1	39
APÊNDICE F	Classificação nBest - Sistema 2	40
APÊNDICE G	Classificação nBest - Sistema 3	41
APÊNDICE H	Classificação nBest - Sistema 4	42
APÊNDICE I	Percentuais de Classificação - Sistema 1	43
APÊNDICE J	Percentuais de Classificação - Sistema 2	44
APÊNDICE K	Percentuais de Classificação - Sistema 3	45
APÊNDICE L	Percentuais de Classificação - Sistema 4	46
Anexos		47
ANEXO A	Frases Balanceadas	48

1 INTRODUÇÃO

Utilizar a voz humana como uma característica biométrica vem se tornando uma técnica segura a ser empregada nas mais diversas situações de controle e investigação como, por exemplo, no acesso a dispositivos pessoais e acesso a locais restritos. Esta segurança e simplicidade advêm da extração de características acústicas do sinal de fala de maneira não invasiva e seu emprego na modelagem de locutores por ferramentas estatísticas (CARDOSO, 2009).

A biometria é a parte da ciência que busca analisar e quantificar dados biológicos. Existem diversas formas, tais como impressões digitais, retinas e íris, reconhecimento de locutor (padrões de voz), padrões faciais e medições de mão. Para o reconhecimento de um indivíduo pela sua voz é necessário captar o som por um microfone e extrair dados do sinal captado. A partir destes dados criar modelos estatísticos que irão representar determinado locutor. Posteriormente, o processo de reconhecimento consiste em verificar qual o modelo mais provável que gerou uma sequência de teste. O reconhecimento de um locutor pode ser classificado em função do texto enunciado (dependente de texto ou independente de texto). Num sistema dependente de texto os locutores testados enunciam uma locução e o sistema irá comparar em tempo real o modelo de locução de teste com os modelos de locuções previamente armazenadas. Por sua vez, em um sistema de reconhecimento independente de texto, compara-se o modelo do locutor de teste com os modelos armazenados no sistema sem a necessidade de repetir frase(s) ou palavra(s) previamente armazenada(s) (CUADROS, 2007).

Existem inúmeras maneiras de reconhecimento de um locutor. Neste trabalho, utilizou-se de Coeficientes Cepstrais de Frequência Mel, do inglês *Mel frequency cepstral coefficients* MFCC, como parâmetros extraídos. A modelagem de cada locutor foi feita por modelo de misturas de gaussianas, do inglês *Gaussian Mixture Model* GMM. Foi utilizado o algoritmo de Maximização da Expectativa, do inglês *Expectation Maximization* EM, que estima parâmetros para modelos estatísticos com base no conceito da máxima verossimilhança.

Há diversos softwares de cálculo numérico de alta performance para extração dos vetores característicos e para modelagem de locutores como, por exemplo, o GNU *Octave*. Um deles foi utilizado para implementação do trabalho. Este trabalho foi organizado em 7 capítulos, sendo que o Capítulo 1 é a Introdução. O Capítulo 2 tratará dos Objetivos deste trabalho, em um sentido amplo, e o Capítulo 3 das Justificativas para tal. O Capítulo 4 apresenta o Referencial Teórico correspondente a métodos, técnicas e ferramentas necessárias ao desenvolvimento deste trabalho. No Capítulo 5 descreveremos a Metodologia utilizada, no Capítulo 6 a Análise e Discussão dos Resultados e o Capítulo 7 apresenta a Conclusão e os Trabalhos Futuros.

2 OBJETIVOS

2.1 OBJETIVO GERAL

O objetivo deste trabalho é estudar o problema de reconhecimento de locutor independente de texto.

2.2 OBJETIVOS ESPECÍFICOS

Como objetivos específicos, têm-se:

- Gravar vozes de um grupo de locutores;
- Criar um banco de dados através das vozes gravadas;
- Extrair os parâmetros acústicos a partir do banco de dados;
- Com os parâmetros extraídos, criar os modelos estatísticos para os locutores;
- Realizar testes para o sistema de reconhecimento de locutor;
- Aplicação do sistema de reconhecimento de locutor para um caso hipotético.

3 JUSTIFICATIVA

Na sociedade moderna, desenvolver sistemas que possam reconhecer pessoas vem se tornando uma necessidade cada vez maior. Diversos setores da sociedade, por motivo de segurança, por exemplo, necessitam de certo rigor no controle do acesso de pessoas. Alguns destes setores são: bancos, indústrias, escolas, entre outros. Assim, para realizar este controle existem vários sistemas que recorrem a biometria para reconhecer um determinado indivíduo, dentre estes, tem-se o reconhecimento de locutor. Segundo (CARDOSO, 2009), “em termos de aplicações práticas, um sistema de verificação poderia ser empregado para confirmar a identidade de uma pessoa que tenta acessar sua conta em um banco, utilizando a própria voz como chave de acesso”. Assim, o reconhecimento de um locutor apresenta várias vantagens, tais como: ser um método não invasivo (pois apenas o som emitido pelo locutor é necessário para as fases treinamento e teste), a voz humana sofre lentas variações de suas características acústicas ao longo do tempo e o método possui erros muito baixos durante uma identificação em condições controladas (um ambiente silencioso). A fundamentação teórica do reconhecimento de locutor se desenvolveu muito nos últimos anos e a tendência é que se desenvolva ainda mais. Como afirma (CUADROS, 2007), “os primeiros trabalhos descrevendo máquinas que podiam reconhecer, com certo sucesso, a pronúncia de determinadas palavras datam de 1952. Uma grande quantidade de trabalhos sobre o assunto surgiu nos anos 60, graças as descobertas de propriedades da voz através do uso de espectógrafos e das novas facilidades que os computadores digitais vieram a oferecer”.

4 REVISÃO DA LITERATURA

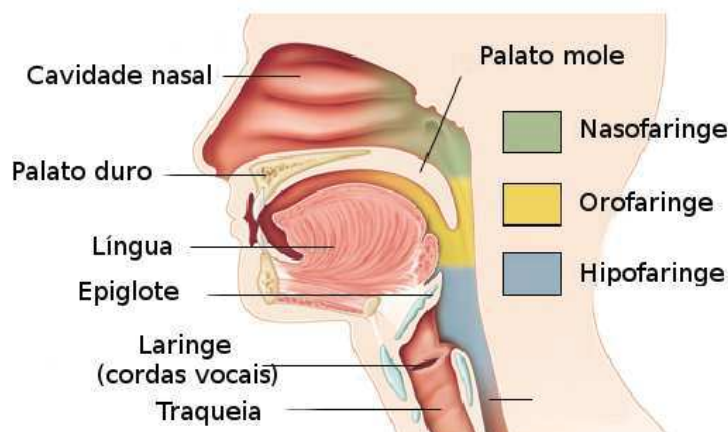
4.1 APARELHO FONADOR

Os seres humanos possuem características físicas que os diferenciam. São exemplos destas características, o sexo, a altura, a sua massa, a cor da pele, a cor do cabelo e a cor dos olhos, entre outras. Nas diversas situações da vida social, um indivíduo pode reconhecer ou ser reconhecido por suas características. Para isso, os seres humanos utilizam os seus cinco principais sentidos: olfato, audição, visão, paladar e o tato. De maneira geral, é por meio de seus sentidos que os indivíduos conseguem distinguir, reconhecer, comparar, identificar e classificar as diversas situações que ocorrem ao seu redor. Portanto, um ser humano possui como mecanismo de entrada de dados os seus sentidos e como mecanismo de saída as suas expressões (movimentos, fala). A fala é uma característica muito importante para os seres humanos, pode ser por meio dela, que conseguimos transformar os nossos pensamentos em linguagem e assim nos comunicarmos. Adicionalmente, o som emitido por um locutor possui diversas características físicas dentre as quais, a frequência. Assim, as frequências das locuções emitidas então compreendidas na faixa entre $0Hz$ e $7kHz$, conhecidas como limiares de fala. Tanto a frequência quanto outras características do som emitido pelos seres humanos podem ser afetados por diversos fatores, dentre os quais velocidade da locução, idade do locutor e estado de saúde do falante.

A produção da voz ocorre na laringe, onde se encontram as cordas vocais. Durante o processo de respiração o fluxo de ar entra e sai pelos pulmões sendo pressionado pelo diafragma. Quando falamos o fluxo de ar vindo dos pulmões passa pelas pregas vocais provocando vibrações. A abertura que se apresenta na laringe durante o processo de produção dos sons é denominada glote. As vibrações das cordas vocais resultam da ação de forças exercidas pela laringe pelo fluxo de ar vindo dos pulmões. No processo de produção de sons, as pregas vocais podem vibrar ou não. Se um sinal for gerado devido as vibrações das cordas vocais, diz-se que o som é sonoro. Porém, se durante o processo de produção dos sons as cordas vocais não vibrarem, diz-se que o som é surdo. Logo após os sinais serem gerados na laringe, ele passa pela faringe, boca e nariz, as quais funcionam como "amplificadores naturais". Estas estruturas são responsáveis pela articulação da fala e caracterizam o tipo de fala do indivíduo. Na Figura 1 estão representadas as partes constituintes do trato vocal de um indivíduo.

A explanação feita anteriormente sobre o processo de produção do som pode ser modelada por meio do diagrama de blocos da Figura 2 sendo que, no domínio de tempo discreto n , tem-se $p[n]$ que é um trem de pulsos gerados durante a passagem do fluxo de ar pela glote, quando esta está fechada. Estes pulsos são constituídos por uma componente fundamental e suas harmônicas e representam a componente sonora do sinal de voz. Por outro lado, $r[n]$ é um ruído branco gaussiano que é gerado na glote pela passagem de ar vindo dos pulmões, quando a mesma está aberta. Este ruído branco representa a componente sonora surda do sinal de voz. Logo, a

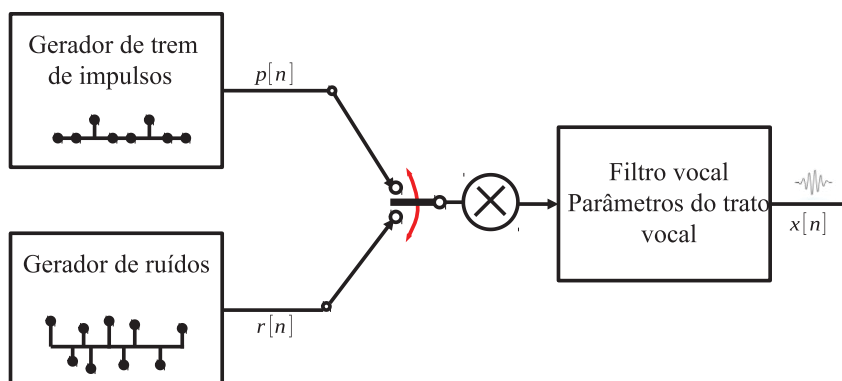
Figura 1 – Aparelho fonador.



Fonte: Adaptado da página <www.ericasitta.wordpress.com/2015/04/16/conheca-a-laringe>

saída do sistema descrito na Figura 2 são as amostras $x[n]$ que representam as ressonâncias que ocorrem no trato vocal. Segundo (OPPENHEIM; SCHAFER, 2013), “Nesse modelo, assume-se que amostras do sinal de voz são a saída de um sistema de tempo discreto variante no tempo, que modela as ressonâncias do sistema de trato vocal. O modelo de excitação do sistema comuta entre impulsos periódicos e ruído aleatório, dependendo do tipo de som a ser produzido”. Ainda em relação a produção de sons (SIQUEIRA, 2011) afirma, “Diz-se que o som é sonoro quando a corrente de ar que vem dos pulmões encontra as cordas vocais fechadas, fazendo-as vibrar. Por exemplo, na palavra ‘Bato’, percebe-se este som sonoro devido ao fonema /B/. E o som é surdo quando a corrente de ar que vem dos pulmões encontra as cordas vocais relaxadas (abertas), não ocorrendo vibração, por exemplo na palavra ‘Prato’ percebe-se este som surdo devido ao fonema /P/”.

Figura 2 – Representação em diagrama de blocos do comportamento de um trato vocal.



Fonte: Autoria própria.

4.2 AMOSTRAGEM DE UM SINAL

Para extrair e processar informação de uma locução é necessário digitalizar o som. Logo, deve-se extrair amostras deste sinal contínuo em intervalos fixos de tempo, o que é feito por um conversor Analógico Digital (A/D). O número de amostras por segundo de um sinal é denominado frequência de amostragem f_s . Para que um sinal seja recuperado com fidelidade, é necessário que o mesmo seja amostrado com no mínimo o dobro da frequência máxima do sinal analógico original f_{max} ,

$$f_s \geq f_n \geq 2f_{max}, \quad (1)$$

sendo que f_n é denominada frequência de Nyquist (OPPENHEIM; SCHAFER, 2013). Supondo que uma determinada locução tenha uma frequência máxima de $f_{max} = 8$ kHz, para que este sinal seja recuperado com fidelidade, é preciso ser amostrado no mínimo a $f_s = 16$ kHz. Caso o sinal seja amostrado com uma frequência abaixo da frequência de Nyquist ocorre o fenômeno de *aliasing*, o qual resulta em distorção devido a sobreposição no espectro.

O processo de amostragem compreende a etapa de quantização, no qual os valores das amostras são aproximados para níveis discretos pré-definidos. O número de níveis de quantização influencia diretamente a resolução de um sinal. Por exemplo, se um sinal for codificado a 3 bits, ele terá 2^3 níveis de quantização possíveis. Após o processo de quantização, o sinal amostrado é codificado na forma binária, de acordo com o número de bits que neste trabalho é de 16.

4.3 DEFINIÇÕES

Inicialmente, vamos definir alguns conceitos importantes para um melhor entendimento dos assuntos abordados nas próximas etapas deste trabalho de conclusão de curso.

4.3.1 SINAIS DETERMINÍSTICOS E ESTOCÁSTICOS

Um sinal é dito determinístico quando pode ser compreendido o seu comportamento para qualquer intervalo de tempo. Têm-se, como exemplos de sinais determinísticos, o seno e o cosseno. Um sinal é compreendido como estocástico quando sua função depende do tempo e uma ou mais variáveis aleatórias, ou seja, não há como saber o seu comportamento com precisão para quaisquer instantes de tempo. Pode-se obter informações de um sinal aleatório por meio de suas estatísticas. São exemplos de sinais estocásticos a voz, ruído branco e variações elétricas registradas em um eletroencefalograma, entre outros. Segundo (CUADROS, 2007), “Um sinal é dito estacionário quando suas características estatísticas não variam em função do tempo. A voz humana é um tipo de sinal denominado quase-estacionário, pois pode-se considerá-lo como estacionário em curtos intervalos de tempo”.

4.4 PRÉ-PROCESSAMENTO DE UM SINAL

Um sinal analógico pode conter ruídos. Os ruídos são tipos de sinais que interferem diretamente no processamento digital de sinais. A relação sinal-ruído, do inglês *signal-to-noise ratio* (SNR), mostra o quanto a medida de espectro do ruído está próximo da medida de espectro do sinal. Portanto, quanto maior a SNR, menor é a presença de ruídos no sinal e quanto menor a SNR mais o sinal é afetado pelo ruído.

4.4.1 PRÉ-ÊNFASE

A pré-ênfase é uma etapa fundamental para um bom resultado de reconhecimento de locutor. Esta fase consiste em passar o sinal de voz por um filtro passa-altas, cuja função é enfatizar as componentes de frequências mais elevadas e atenuar as frequências mais baixas haja vista o trato vocal de um ser-humano possui características de um filtro passa baixas. Muitas informações importantes estão presentes nas componentes de alta frequência, as quais também constituirão os coeficientes característicos do locutor. A função de transferência de um filtro de pré-ênfase de primeira ordem é dada por

$$H(z) = 1 - az^{-1}, \quad (2)$$

sendo a é uma constante de valor 0,97.

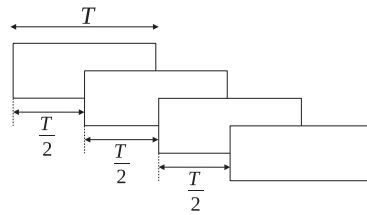
4.4.2 SEGMENTAÇÃO EM QUADROS

Visto que o sinal de voz é estocástico e quase-estacionário, podendo ser considerado como estacionário em pequenos intervalos de tempo, entre $10ms$ e $30ms$ (OPPENHEIM; SCHAFER, 2013). Um segmento do sinal de voz neste intervalo de tempo é considerado um sistema linear e invariante no tempo SLIT. Segundo (MOLAU et al., 2001), “a forma de onda da locução, amostrada a uma frequência de 8 kHz ou 16 kHz, é inicialmente filtrada (pré-ênfase) e, em seguida, segmentada em uma série de quadros, cada um com comprimento de T ms cada, e sobrepostos por um deslocamento de $\frac{T}{2}$ ms”. Porém, a segmentação ocasiona perdas de informações nas fronteiras de cada quadro devido as variações abruptas dos segmentos. Estas mudanças geram componentes de altas frequências que alteram o espectro do sinal de voz. Para contornar este problema, realiza-se então a sobreposição entre os quadros vizinhos conforme citado anteriormente. A Figura 3 ilustra de maneira simplificada o processo de segmentação e sobreposição do sinal.

4.4.3 JANELAMENTO DE UM SINAL

Somente a sobreposição de amostras adjacentes não é suficiente para reduzir os efeitos indesejáveis no espectro do sinal. É necessário convoluir em frequência cada quadro com uma função janela específica. O janelamento é um processamento realizado em cada um dos quadros

Figura 3 – Segmentação do sinal mostrando a sobreposição em quadros de 50%.



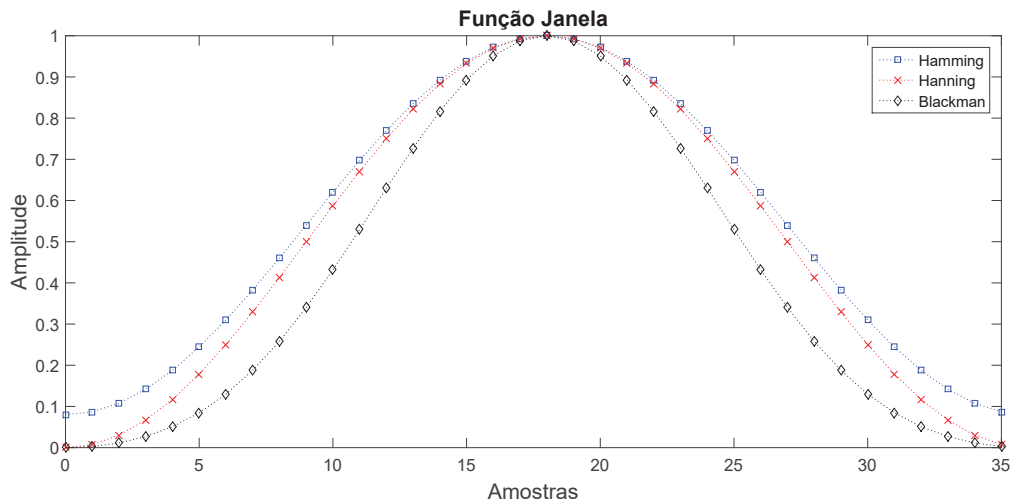
Fonte: **Autoria Própria.**

obtidos na etapa de segmentação e sua função é suavizar as bordas de cada quadro, além de acentuar as informações contidas no centro dos quadros. Para isso, podem ser utilizadas diversas funções janela como, por exemplo, Hamming, Hanning e Blackman que estão representadas na Figura 4. As mais utilizadas para tratamento de sinais de voz são as janelas de Hamming e a de Hanning. A janela de Hamming (OPPENHEIM; SCHAFER, 2013), empregada neste trabalho é definida por

$$w[n] = \begin{cases} 0,54 - 0,46 \cos(2\pi n/M), & 0 \leq n < M \\ 0, & \text{caso contrario,} \end{cases} \quad (3)$$

sendo que n é número de amostras e M é o comprimento de cada quadro.

Figura 4 – Janelas Hamming, Hanning e Blackman.



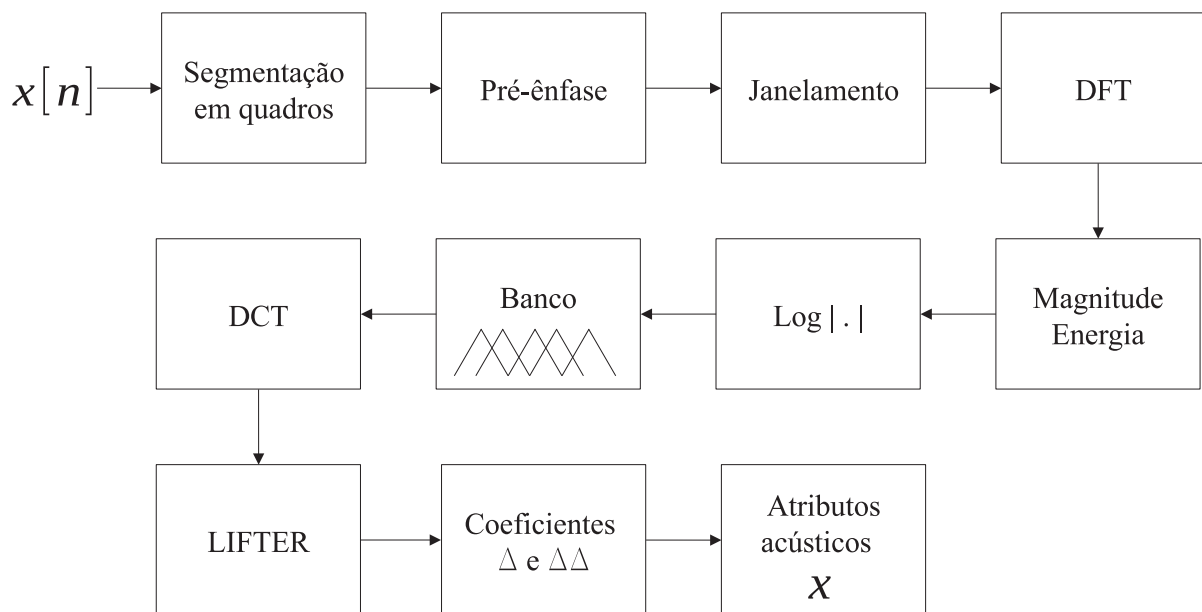
Fonte: **Autoria própria.**

4.5 PARÂMETROS CARACTERÍSTICOS

Durante a produção da voz, muitas informações estão presentes no sinal. Estas informações podem ser parametrizadas para posterior modelagem do trato vocal do locutor. Muitas características do som emitido pelo locutor podem ser percebidas por um ouvinte, tais como:

dialeto, sotaque e estado de saúde. Porém, outras informações que constituem o sinal da fala são melhores compreendidas por uma análise mais criteriosa no domínio da frequência. São exemplos destes parâmetros: frequência fundamental, timbre e intensidade. Para este projeto, será abordado a extração de parâmetros característicos MFCCs. Divide-se o processo de extração dos coeficientes MFCCs a partir do sinal de voz amostrado, conforme Figura 5. Uma visão geral sobre cada processo será fornecida na Seção 4.6.

Figura 5 – Diagrama de blocos ilustrando as etapas para extração dos atributos acústicos.



Fonte: **Autoria própria.**

4.6 PROCEDIMENTOS PARA A EXTRAÇÃO DOS PARÂMETROS

Em todas as etapas analisadas anteriormente, os sinais de voz obtidos dos locutores estão no domínio do tempo. Neste domínio é possível analisar as variações de amplitude do sinal que foi segmentado em cada quadro. Porém, para se obter os vetores de coeficientes característicos ou atributos dos sons emitidos pelos locutores, é necessário transformar os sinais de voz para um domínio cuja as representações dos sinais sejam mais distintas. Transforma-se então, o sinal de voz para o domínio da frequência. Para realizar essa mudança, será utilizado neste projeto a Transformada de Fourier Discreta (TFD), (OPPENHEIM; SCHAFER, 2013) dada por

$$X[k] = \begin{cases} \sum_{n=0}^{N-1} x[n]e^{-j(2\pi k/N)}, & 0 \leq k \leq N-1 \\ 0, & \text{c.c.}, \end{cases} \quad (4)$$

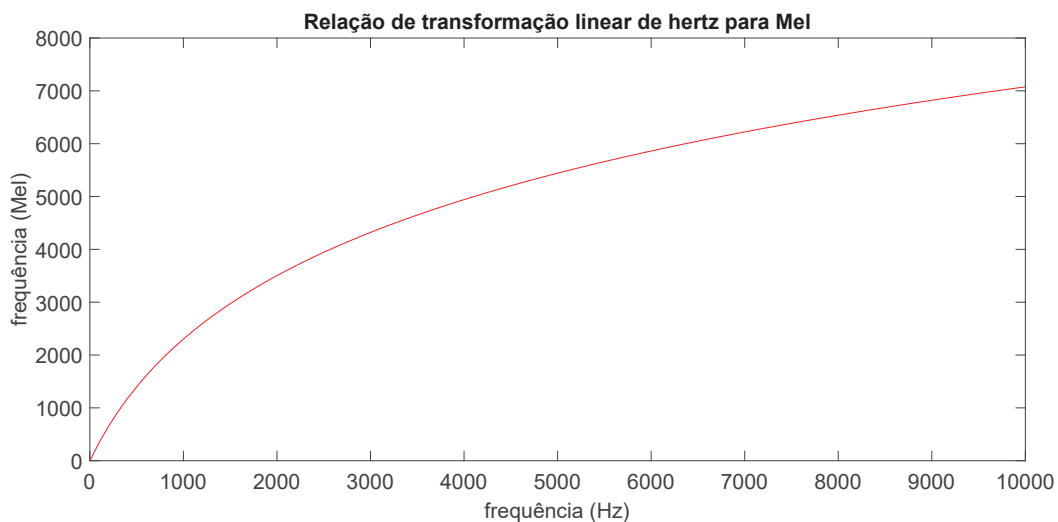
sendo que $X[k]$ é a TFD da sequência de comprimento finita ou amostras de tempo discreto $x[n]$ e N é o comprimento de cada quadro. Após a conversão do sinal para o domínio da frequência, trabalha-se com a magnitude ou energia do sinal distribuída pelo espectro e a esses valores é

aplicado o logaritmo natural. Em seguida há a aplicação de um banco de filtros triangulares na escala MEL (OPPENHEIM; SCHAFER, 2013). Essa escala visa modelar as características de aquisição do som pelo ouvido humano e é expressa por

$$Mel(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \quad (5)$$

sendo que f é a frequência obtida com a TFD. Com isso, têm-se que a relação entre a escala Mel com a frequência em Hz é indicado na Figura 6.

Figura 6 – Relação de frequência na escala MEL.



Fonte: **Autoria própria.**

Ainda como processo de tratamento dos sinais obtidos, utiliza-se da Transformada discreta do cosseno, do inglês *Discrete Cosine Transform* DCT. Esta visa eliminar redundâncias das informações obtidas, resultando na compressão ou redução da quantidade de dados a serem armazenados. Como afirma (GORDILLO, 2013), “Esta redução é feita por meio de uma propriedade da DCT conhecida como compactação da energia, concentrando os valores mais significativos nos primeiros termos do vetor, e descartando os últimos, melhorando assim a eficiência computacional”. Uma vez utilizada a DCT, é aplicado a um filtro passa-baixas no domínio do cepstrum denominado lifter, o qual visa eliminar as componentes de alta quefrências a partir de um certo limiar, relativas ao sinal de excitação provenientes da glote. O termo quefrência é a unidade análoga a frequência porém, no domínio do cepstrum conforme (MACHADO, 2013). Por fim, são obtidos os MFCCs. Adicionalmente, por meio dos coeficientes estáticos MFCCs são obtidos os coeficientes dinâmicos energia, delta e delta-delta. Estes, por sua vez, melhoram os resultados nas etapas de treinamento e teste do reconhecimento de locutor. Como afirma (GORDILLO, 2013) e relação aos dois últimos coeficientes, “A ideia principal da extração de atributos é captar as mudanças temporais bruscas presentes no espectro. Devido a isto, utilizam-se além, dos coeficientes extraídos até agora, chamados coeficientes “estáticos”, os coeficientes delta e delta-delta, chamados coeficientes “dinâmicos”, que capturam essas mudanças e incorporam

informação relativa à transição dos coeficientes estáticos entre quadros vizinhos”. Similarmente, a partir dos coeficientes estáticos “log-energia”, que segundo (GORDILLO, 2013) “carrega muita informação do meio de transmissão”, obtém-se os coeficientes dinâmicos “delta log-energia” e “delta-delta log-energia”. Isto posto, estes coeficientes dinâmicos podem ser obtidos analiticamente, pela seguinte expressão, conforme (YOUNG et al., 2006):

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta (c_{t+\theta} - c_{t-\theta})}{2 \sum_{\theta=1}^{\Theta} \theta^2} \quad (6)$$

sendo que, d_t é o coeficiente delta calculado no tempo t a partir dos coeficientes estáticos $c_{t+\theta}$ e $c_{t-\theta}$ e Θ é o atraso utilizado para calcular os respectivos coeficientes dinâmicos visto que, para este trabalho foi adotado $\Theta = 2$. Resumidamente, o número de coeficientes, somando 42 coeficientes por quadro, estão listados a seguir:

- coeficientes estáticos - 13 MFCCs e 1 log-energia;
- coeficientes dinâmicos - 13 delta MFCCs e 1 delta log-energia;
- coeficientes dinâmicos - 13 delta-delta MFCCs e 1 delta-delta log-energia.

4.7 RECONHECIMENTO DE LOCUTOR UTILIZANDO MODELO DE MISTURAS DE GAUSSIANAS

A modelagem estatística é fundamental para representar diversos eventos ou fenômenos que ocorrem diariamente. Encontra-se aplicação da estatística num simples lançamento de dado, na previsão do tempo, tratamento de dados estatísticos, reconhecimento de padrões e aprendizado de máquinas, entre outros. A voz humana como dito anteriormente é um evento estocástico. Logo, é necessário uma modelagem estatística para representar as diversas informações presentes no sinal da fala. Estes dados descritivos obtidos nas fases preliminares do processamento de voz serão utilizadas para compor o modelo do trato vocal de um dado locutor. A voz humana pode ser utilizada para reconhecer um locutor porém, não é uma tarefa trivial, haja vista o sistema de reconhecimento deve ser capaz de reconhecer padrões e construir modelos para cada um dos locutores na fase de treinamento e, posteriormente, na fase de teste ser capaz de comparar parâmetros extraídos de um locutor com os modelos previamente armazenados. Assim, o modelo que apresentar maior verossimilhança em relação a um dado valor limiar, será considerado o locutor.

A função densidade de probabilidade gaussiana

$$\mathcal{N}(\mathbf{x}|\mu_{\mathbf{k}}, \Sigma_{\mathbf{k}}) = \frac{1}{2\pi^{\frac{D}{2}} |\Sigma_{\mathbf{k}}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{k}}) \Sigma_{\mathbf{k}}^{-1} (\mathbf{x} - \mu_{\mathbf{k}}) \right\}, \quad (7)$$

é uma função estatística empregada em diversas áreas do conhecimento, representada por um vetor de médias $\mu_{\mathbf{k}}$ com dimensão D , e uma matriz de covariâncias $\Sigma_{\mathbf{k}}$ com dimensão $D \times D$. Em muitas situações reais, apenas uma distribuição gaussiana não consegue representar de maneira satisfatória determinados conjuntos de dados. Nesses casos, quando se utiliza um número maior

de gaussianas, a representação se torna mais adequada. A partir disso surgiu o modelo de misturas de gaussianas dada por

$$p(\mathbf{x}|\lambda) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k), \quad (8)$$

que consiste em um número definido de gaussianas sobrepostas e ponderadas por um peso π_k , de tal maneira que a soma de suas contribuições ou pesos seja unitária

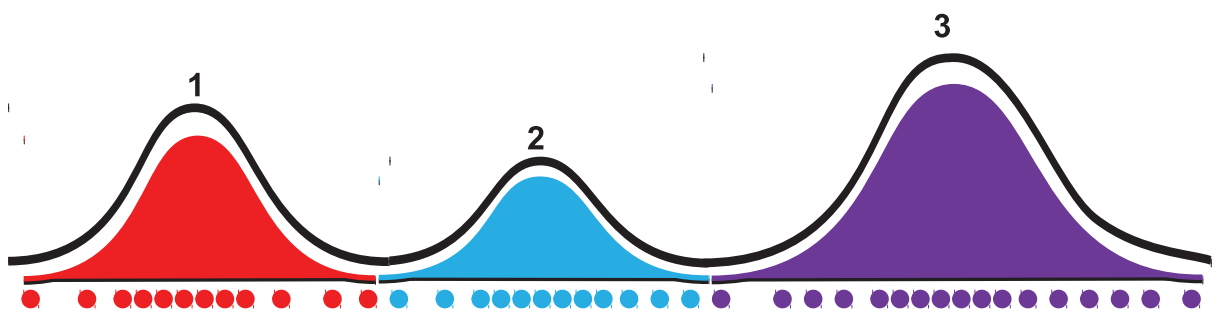
$$\sum_{k=1}^K \pi_k = 1. \quad (9)$$

Sendo assim, segundo (BISHOP, 2006), “o modelo de mistura de gaussianas é uma combinação linear, ou superposição de componentes gaussianas”. Na equação (8) o vetor \mathbf{x} são os coeficientes MFCCs e cada uma das componentes da mistura possui um peso, um vetor de médias e uma matriz de covariâncias. Vale observar que os vetores \mathbf{x} e μ_k possuem dimensão D , e a matriz de covariância Σ_k é de dimensão $D \times D$. Isto posto, um modelo de mistura de gaussianas pode ser representado por

$$\lambda = \{\pi_k, \mu_k, \Sigma_k\}. \quad (10)$$

A voz humana é um tipo de sinal quase-estacionário quando analisado em quadros de 10 a 30 ms. Os vetores característicos que constituem cada quadro obtido nas fases preliminares apresentam características que são comuns. Estes valores serão utilizados nos modelos de misturas de gaussianas, conforme a ilustração unidimensional dada na Figura 7.

Figura 7 – Exemplo ilustrando o modelo de misturas de gaussianas unidimensional.



Fonte: **Autoria Própria.**

Nesta ilustração pode-se notar que existem três agrupamentos de círculos coloridos distribuídos ao longo do eixo unidimensional. Como dito anteriormente, apenas uma gaussiana não conseguiria representar de maneira precisa estes agrupamentos, os quais podem ser representados cada um por uma gaussiana. Aplicando-se o modelo de mistura de gaussianas, cada conjunto possuirá uma média e um desvio padrão que representará a gaussiana. Na Figura 7 vê-se que a gaussiana 3 possui maior peso do que as gaussianas 1 e 2. Ao se realizar a mistura

destas gaussianas para representar estes dados, obtém-se a curva superior dada pela soma de cada componente que representa o conjunto de dados.

Diante dos conceitos abordados anteriormente, como construir um modelo de misturas de gaussianas que represente um dado locutor? Ou seja, encontrar os parâmetros ótimos da equação (10), que representam cada uma das gaussianas que constituem o modelo. Para responder a essa pergunta, considere inicialmente o Teorema de Bayes

$$P_r(\lambda_s|\mathbf{X}) = \frac{p(\mathbf{X}|\lambda_s)}{p(\mathbf{X})} P_r(\lambda_s), \quad (11)$$

tal que $P_r(\lambda_s|\mathbf{X})$ representa a probabilidade *a posteriori* de um determinado conjunto de dados conhecidos pertencerem a um dado modelo, $p(\mathbf{X}|\lambda_s)$ fornece a função densidade de probabilidade de que um determinado conjunto de dados desconhecidos \mathbf{X} seja pertencente ao modelo λ_s , $p(\mathbf{X})$ é a função densidade de probabilidade dos dados e $P_r(\lambda_s)$ probabilidade de um dado locutor. Para o nosso caso, $P_r(\lambda_s|\mathbf{X}) \propto p(\mathbf{X}|\lambda_s)$, pois todos os dados recebidos possuem igual probabilidade de pertencerem a qualquer um dos locutores modelados, bem como todos os modelos possuem iguais probabilidades *a priori* de que os dados de teste pertençam a eles. Estima-se então um modelo para o locutor à partir dos dados de treinamento \mathbf{x}_t , com esses, é possível construir um modelo ótimo que represente um locutor \hat{s} . Suponha que no processo de reconhecimento (testes) tem-se o seguinte conjunto de dados independentes $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t, \dots, \mathbf{x}_T$ como parâmetros de um locutor. A probabilidade de que esse conjunto de dados pertença a um locutor é

$$p(\mathbf{X}|\lambda_s) = p(\mathbf{x}_1|\lambda_s) \times p(\mathbf{x}_2|\lambda_s) \times \dots p(\mathbf{x}_T|\lambda_s). \quad (12)$$

O modelo ótimo citado anteriormente é obtido aplicando-se o método conhecido como estimação da máxima verossimilhança, conforme segue-se

$$\hat{s} = \left(\underset{1 \leq s \leq S}{arg\ max} \right) \prod_{t=1}^T p(\mathbf{x}_t|\lambda_s), \quad (13)$$

sendo \hat{s} o provável locutor dentre um grupo de locutores S . Maximizar a verossimilhança é equivalente a maximizar o logaritmo da verossimilhança

$$\hat{s} = \log \left(\left(\underset{1 \leq s \leq S}{arg\ max} \right) \prod_{t=1}^T p(\mathbf{x}_t|\lambda_s) \right), \quad (14)$$

transformando o produtório em um somatório

$$\hat{s} = \left(\underset{1 \leq s \leq S}{arg\ max} \right) \sum_{t=1}^T \log p(\mathbf{x}_t|\lambda_s). \quad (15)$$

Para estimar os parâmetros dos modelos de cada locutor (fase de treinamento ou modelagem), aplica-se o logaritmo na equação (8), obtém-se

$$\log(p(\mathbf{X}|\lambda)) = \sum_{t=1}^T \log(p(\mathbf{x}_t|\lambda)) = \sum_{t=1}^T \log\left(\sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_t|\mu_k, \Sigma_k)\right). \quad (16)$$

A seguir, para obter cada um dos parâmetros do modelo λ de maneira independente um do outro, considere a equação (16) derivada parcialmente em relação ao vetor de médias μ_k ,

$$\frac{\partial}{\partial \mu_k} \left\{ \sum_{t=1}^T \log\left(\sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_t|\mu_k)\right) \right\} = 0, \quad (17)$$

segue-se,

$$\sum_{t=1}^T \frac{\pi_k \mathcal{N}(\mathbf{x}_t|\mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_t|\mu_j, \Sigma_j)} \Sigma_k (\mathbf{x}_t - \mu_k) = 0, \quad (18)$$

definindo

$$\frac{\pi_k \mathcal{N}(\mathbf{x}_t|\mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_t|\mu_j, \Sigma_j)} = \gamma_{tk}. \quad (19)$$

O termo γ_{tk} representa-se a normalização da função densidade de probabilidade multi-dimensional. Substituindo a equação (19) na equação (18), obtém-se

$$\sum_{t=1}^T \gamma_{tk} \Sigma_k (\mathbf{x}_t - \mu_k) = 0. \quad (20)$$

A partir da equação (20), define-se o vetor de médias

$$\mu_k = \frac{1}{N_k} \sum_{t=1}^T \gamma_{tk} \mathbf{x}_t \quad (21)$$

e

$$N_k = \sum_{t=1}^T \gamma_{tk}, \quad (22)$$

sendo que N_k representa a soma dos pesos de todas as gaussianas.

Agora, considere a equação (16) derivada parcialmente em relação a matriz de covariâncias Σ_k e igualada a zero,

$$\frac{\partial}{\partial \Sigma_k} \left\{ \sum_{t=1}^T \log\left(\sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_t|\mu_k)\right) \right\} = 0. \quad (23)$$

Com isso, obtém-se

$$\Sigma_k = \frac{1}{N_k} \sum_{t=1}^T \gamma_{tk} (\mathbf{x}_t - \mu_k)(\mathbf{x}_t - \mu_k)^T, \quad (24)$$

que fornece a matriz de covariâncias para cada uma das gaussianas que compõem a mistura. Assim sendo, o vetor de pesos correspondentes a cada gaussiana é

$$\pi_{\mathbf{k}} = \frac{N_{\mathbf{k}}}{N}, \quad (25)$$

de forma que,

$$N = \sum_{k=1}^K N_k. \quad (26)$$

Para obtenção dos parâmetros, segundo (CARDOSO, 2009), “a resolução direta não é factível dada a não-linearidade dos parâmetros do modelo λ . Assim, um processo alternativo na determinação do modelo λ é ajustá-lo a cada iteração de forma que o conjunto \mathbf{X} se torne mais verossímil. Isto é possível com o emprego do algoritmo EM”.

4.7.1 ALGORITMO EXPECTATION MAXIMIZATION

Os dados obtidos nas fases de processamento do sinal de voz são utilizados para compor o modelo de cada locutor. Porém, é necessário um tratamento computacional para que cada locutor tenha um modelo ótimo. Este é conseguido por meio do algoritmo EM. Busca-se uma estimativa maximizada dos seguintes parâmetros: vetor média, matriz de covariância e os pesos de cada componente gaussiana. O algoritmo EM inicializa os valores dos parâmetros para cada gaussiana, as médias $\mu_{\mathbf{k}}$ assumem valores aleatórios do seu respectivo conjunto de dados, as matrizes de covariância $\Sigma_{\mathbf{k}}$ assumem um valor calculado a partir do vetor de características e os pesos $\pi_{\mathbf{k}}$ são tomados como iguais para cada agrupamento de dados. Os parâmetros são recalculados de maneira iterativa até que atinjam o critério de convergência estabelecido, como será mostrado a seguir. Com os parâmetros $\mu_{\mathbf{k}}$, $\Sigma_{\mathbf{k}}$ e $\pi_{\mathbf{k}}$ inicializados, calcula-se γ_{tk} por

$$\gamma_{tk} = \frac{\pi_{\mathbf{k}} \mathcal{N}(\mathbf{x}_t | \mu_{\mathbf{k}}, \Sigma_{\mathbf{k}})}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_t | \mu_j, \Sigma_j)}. \quad (27)$$

Com o valor encontrado em γ_{tk} , obtém-se os novos valores dos parâmetros $\mu_{\mathbf{k}}$, $\Sigma_{\mathbf{k}}$ e $\pi_{\mathbf{k}}$

$$\mu_{\mathbf{k}}^{new} = \frac{1}{N_{\mathbf{k}}} \sum_{t=1}^T \gamma_{tk} \mathbf{x}_t \quad (28)$$

$$\Sigma_{\mathbf{k}}^{new} = \frac{1}{N_{\mathbf{k}}} \sum_{t=1}^T \gamma_{tk} (\mathbf{x}_t - \mu_{\mathbf{k}}^{new})(\mathbf{x}_t - \mu_{\mathbf{k}}^{new})^T \quad (29)$$

$$\pi_{\mathbf{k}}^{new} = \frac{N_{\mathbf{k}}}{N} \quad (30)$$

$$N_{\mathbf{k}} = \sum_{t=1}^T \gamma_{tk}. \quad (31)$$

Após cada iteração, verifica-se se o critério de convergência foi atingido por

$$\log(p(\mathbf{X}|\lambda)) = \sum_{t=1}^T \log\left(\sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_t | \mu_{\mathbf{k}}, \Sigma_{\mathbf{k}})\right). \quad (32)$$

Do exposto acima, o algoritmo EM termina quando a variação dos valores de médias e de matrizes de covariâncias atingem um dado valor limiar.

5 MATERIAIS E MÉTODOS

O presente trabalho de conclusão de curso intitulado Reconhecimento de Locutor Utilizando Modelo de Misturas Gaussianas Treinado pelo Algoritmo de Maximização da Expectativa, cujo pesquisador responsável é o professor ALBERTO YOSHIHIRO NAKANO e os acadêmicos Juliano Rodrigues Dourado e Helio Rodrigues da Silva, foi submetido a apreciação do Comitê de Ética (CAAE), cujo certificado é 83708018.5.0000.5547, em 25/02/2018 e aprovado em 08/03/2018 pela instituição proponente UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANA - UTFPR. Logo, por se tratar de uma pesquisa que envolve seres humanos, os envolvidos na pesquisa seguiram todas as orientações exigidas pela UTFPR. Assim sendo, os seguintes documentos submetidos e aprovados são:

- Termo de compromisso, de confidencialidade de dados e envio do relatório final - apêndice A;
- Termo de consentimento e de livre esclarecimento e termo de consentimento para uso de imagem e som de voz - apêndice B;
- Projeto detalhado - apêndice C;
- Uso das instalações - apêndice D.

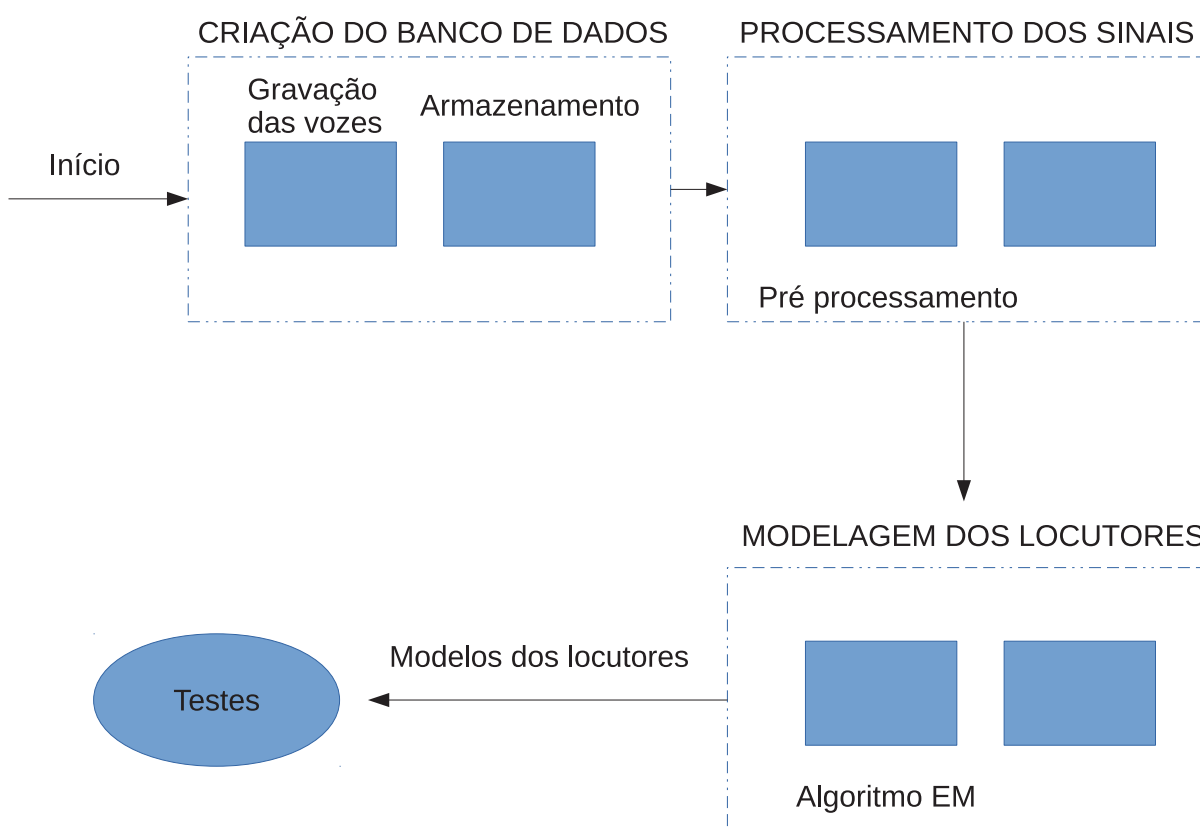
Portanto, todos os locutores voluntários foram esclarecidos sobre os objetivos do projeto, assim como, de sua participação na pesquisa. Estando cientes de todos os aspectos legais, os mesmos receberam uma via do documento que consta no B. Para mais detalhes, consultar os respectivos apêndices.

5.1 PROCEDIMENTOS GERAIS

Após seguidos todos os trâmites legais, foram coletadas as vozes dos locutores para a criação do banco de dados. O conjunto de locutores escolhidos é constituído de 17 homens e 23 mulheres na faixa etária de 18 a 60 anos. As gravações das vozes ocorreram nas dependências da Universidade Tecnológica Federal do Paraná - campus Toledo, em um dos laboratórios que constituem a coordenação do curso de Engenharia Eletrônica conforme apêndice D. No momento da gravação o local esteve o mais silencioso possível para que não ocorresse adição de ruídos nos sinais de vozes armazenados. As 200 frases balanceadas gravadas foram obtidas de (ABRAHAM Alcaim, 1992) e constam no anexo A. Durante o processo de gravação procurou-se deixar um intervalo de, aproximadamente 1 segundo de silêncio, com isso, obter um melhor resultado no processo de segmentação e de obtenção dos dados de cada locutor. O processo de gravação foi feito utilizando um gravador com resolução de 16 bits e frequência de amostragem de 44,1 kHz. O tempo médio das gravações foi 15 minutos por locutor. Assim, as etapas de desenvolvimento deste trabalho estão mostradas no diagrama de blocos da Figura 8 são: a criação do banco de dados, o processamento dos sinais gravados, a criação dos respectivos modelos e testes realizados

para cada um dos locutores.

Figura 8 – Diagrama de blocos representando a sequência de atividades do projeto.



Fonte: **Fonte:** Autoria própria.

Após a obtenção do banco de dados com os sinais de vozes gravados e armazenados, foi realizado o processamento de tais sinais e extração dos parâmetros característicos MFCCs. O processamento dos sinais foi realizado através de software interativo de alta performance para cálculo numérico. O processamento realizado com os sinais de áudio dividiu-se em duas partes: uma primeira, chamada de pré-processamento consistindo nas etapas de filtragem de pré-ênfase, segmentação em quadros e janelamento; e uma segunda realizada em frequência, nas etapas de extração da magnitude do sinal, aplicação do logaritmo da magnitude, aplicação do banco de filtros na escala MEL, aplicação da DCT e extração dos MFCCs. Para extração dos parâmetros característicos foi aplicado o processo de *Liftering*, pois há a separação em frequência das funções do trato vocal e da voz. Logo, conseguiu-se remover a função da voz por esse processo. Adicionalmente, foi estimado os parâmetros estáticos de log-energia e os respectivos parâmetros dinâmicos delta e delta-delta.

A modelagem dos locutores criada a partir dos MFCCs obtidos na etapa de processamento foi realizada também com uso de software de alta performance para cálculo numérico. Os modelos foram gerados utilizando uma mistura de 16 gaussianas e matrizes de covariância completas para cada um dos 40 modelos de locutores deste estudo conforme Equação 10, este último visando a criação de modelos ótimos. O método para tal modelagem consiste na aplicação

do modelo de misturas gaussianas e do algoritmo EM. Com os modelos criados, foi possível realizar os testes com os dados de cada locutor comparando com os modelos criados e armazenados. Assim, foram obtidos os valores de máxima verossimilhança para cada locutor e com isso, obtido os resultados mostrado em uma matriz de confusão (confusion matrix).

6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Os resultados experimentais obtidos neste trabalho utilizam como métrica os próprios modelos e dados de testes dos locutores. Os sinais de voz empregados no treinamento e testes foram obtidos em um ambiente praticamente livre de ruídos, ou seja, os resultados retratam o reconhecimento de locutores em um ambiente praticamente ideal. Para isso, foram construídos quatro sistemas para reconhecimento cada um com os seguintes parâmetros listados a seguir:

- Sistema 1 - MFCCs;
- Sistema 2 - MFCCs e log Energia;
- Sistema 3 - MFCCs, log Energia e Delta;
- Sistema 4 - MFCCs, log Energia, Delta e Delta-Delta.

Portanto, pode-se comparar os resultados obtidos para o reconhecimento de locutor em cada um dos sistemas listados acima. Os resultados esperados devem, teoricamente, se apresentarem segundo a ordem crescente de desempenho de reconhecimento: Sistema 1, Sistema 2, Sistema 3 e Sistema 4. Para apresentar os resultados, utilizou-se de duas matrizes de classificação denominadas, matriz *n-best* e matriz de confusão. As linhas e as colunas das matrizes representam, respectivamente, os dados de teste e os modelos para cada um dos locutores. Para exemplificar a aplicação destas matrizes, considere os exemplos das Tabelas 1 e 2 abaixo.

Tabela 1 – Matriz *n-best*

	Modelos			
	M1	M2	M3	M4
Locutor 1	1	2	4	3
Locutor 2	4	1	2	3
Locutor 3	2	4	1	3
Locutor 4	3	2	4	1

Tabela 2 – Matriz de confusão (%)

	Modelos			
	M1	M2	M3	M4
Locutor 1	70	3	7	10
Locutor 2	19	40	20	1
Locutor 3	10	6	82	2
Locutor 4	6	12	6	76

A matriz de classificação *n-best* mostrada na Tabela 1 mostra as posições dos modelos dos locutores mais prováveis de terem gerados os dados de teste, com suas respectivas classificações. Essas posições são obtidas de acordo com a soma das probabilidades de cada um dos dados do conjunto de teste dos locutores, aplicados modelo a modelo. Como pode ser visto, os dados de teste do Locutor 2, por exemplo, resultaram no modelo 2 como mais provável, no modelo 3

como segundo mais provável, no modelo 4 como terceiro mais provável e no modelo 1 como quarto mais provável.

Em relação a matriz de confusão mostrada na Tabela 2, tem-se o percentual da quantidade de dados de teste de cada um dos locutores terem sido gerados por cada um dos modelos. Como pode ser visto, 70% dos dados de teste do Locutor 1 indicaram o modelo 1 como o mais provável, 10% para o modelo 4, 7% para o modelo 3, e 3% para o modelo 2.

Com o intuito de deixar claro os dados obtidos, referentes aos 40 locutores, como mencionado anteriormente, as linhas se referem aos dados de testes de cada locutor e as colunas se referem aos modelos. Assim, na vertical têm-se **L** que significa **Locutor** e **M** ou **F** que significam o gênero **Masculino** e **Feminino**. Do mesmo modo, na horizontal têm-se **M** para representar a palavra **Modelo**, seguida da letra **M** ou **F** que representam o gênero **Masculino** e **Feminino**, respectivamente. Isto posto, os resultados obtidos para cada um dos respectivos sistemas de reconhecimento implementados (Sistema 1, Sistema 2, Sistema 3 e Sistema 4) estão mostrados nos apêndices E, F, G e H utilizando matrizes de classificação *n-best*. Nos quatro sistemas, as matrizes obtidas mostram que o modelo mais provável aponta para seu próprio locutor. Isso pode ser visto, analisando a diagonal principal das matrizes, sendo que todos os seus valores são iguais a um.

Os resultados obtidos na forma de matriz de confusão para cada um dos sistemas estão mostrados nos apêndices I, J, K e L. Nos quatro casos, as matrizes obtidas mostram que o modelo mais provável também aponta para seu próprio locutor. Nota-se que os maiores valores percentuais estão posicionados na diagonal principal.

No apêndice H percebe-se, tomando alguns locutores como exemplo que, com exceção do próprio locutor, as posições dos demais que mais se aproximam tendem a ser do mesmo gênero. Por exemplo, o locutor 1 que é do sexo masculino possui seu próprio modelo como mais provável, e o dos locutores 33, 4 e 11 como segundo, terceiro e quarto mais prováveis respectivamente, todos estes também do sexo masculino. O locutor 8, que é do sexo feminino, possui seu próprio modelo como mais provável, e o dos locutores 23, 28 e 13 como segundo, terceiro e quarto mais prováveis respectivamente, sendo todos estes também do sexo feminino.

Para análise de desempenho, toma-se como referência o Sistema 1. Os quatro locutores que obtiveram os maiores e os menores percentuais de reconhecimento para os quatro sistemas estão respectivamente, listados a seguir:

- Locutores com o maior percentual de reconhecimento:
 - Sistema 1 no apêndice I
 - $L_{07} \rightarrow 72,98\%$;
 - $L_{29} \rightarrow 63,43\%$;
 - $L_{15} \rightarrow 61,89\%$;
 - $L_{38} \rightarrow 60,37\%$;
 - Sistema 2 no apêndice J
 - $L_{07} \rightarrow 75,08\%$;

Analisando as locuções dos locutores 7 e 13, mencionados anteriormente, observou-se que as mesmas possuem características discrepantes dos demais. Essas características representam a percepção do som pelo ouvido humano de forma subjetiva, ou seja, são características de difícil percepção.

Em termos de desempenho, percebe-se que há um aumento do mesmo em função dos parâmetros dinâmicos utilizados, o que pode ser avaliado em termos de suas médias de reconhecimento dadas a seguir:

- Sistema 1 no apêndice I sua média é 46,56%;
- Sistema 2 no apêndice J sua média é 47,86%;
- Sistema 3 no apêndice K sua média é 55,37%;
- Sistema 4 no apêndice L sua média é 65,32%,

estes valores que, foram obtidos calculando a média aritmética da diagonal principal das matrizes de confusão dos respectivos sistemas. Com isso, cada sistema fica caracterizado quantitativamente, sendo possível compará-los entre si.

Em relação aos percentuais de reconhecimento dos locutores, considere o apêndice L. Observando os dados de entrada e o modelo do locutor 7, nota-se que o percentual de reconhecimento é de 85,96%, e 14,04% ficam distribuídos entre os 39 locutores restantes.

6.1 APLICAÇÕES

Considere um Sistema Hipotético para abertura de portas, com acesso restrito para 40 usuários. Assim, pode-se definir limiares de percentuais de classificação para os dados de teste, do seguinte modo:

- Limiar A - 50%;
- Limiar B - 55%;
- Limiar C - 60%.

De acordo com o limiar estabelecido para o sistema, o mesmo permitirá o acesso apenas aqueles que obtiverem percentuais de classificação acima dos estabelecidos. Portanto, utilizando o Sistema 4 que possui o melhor desempenho, e os resultados dos testes descritos no apêndice H, os acessos para essa aplicação hipotética, seriam:

- Limiar A - Acesso: Todos os locutores;
- Limiar B - Acesso: Todos os locutores, exceto L12, L13 e L31;
- Limiar C - Acesso: Todos os locutores, exceto L4, L9, L12, L13, L20, L26, L28, L31, L32, L34.

Nos casos em que o acesso é negado, solicita-se que o locutor repita a locução devido à alguma adversidade que possa ter ocorrido durante a identificação. Assim, a técnica de reconhecimento de locutor pode ser aplicada em outras situações tais como controle de frequência em escolas, acesso de contas bancárias, entre outras.

7 CONCLUSÃO

Neste trabalho de conclusão de curso, pôde-se identificar uma pessoa utilizando a voz como característica biométrica. Para isso, utilizando os coeficientes mel-cepstrais foi possível modelar o trato vocal de 40 locutores. Esta modelagem foi realizada empregando o GMM treinado pelo algoritmo EM.

Assim, foi implementado quatro sistemas diferentes, utilizando os coeficientes estáticos MFCCs e log-energia e os coeficientes dinâmicos delta log-energia, delta-delta log-energia, delta e delta-delta. Com a adição dos dinâmicos, verificou-se uma melhora no desempenho dos sistemas, tornando-os mais otimizados, ou seja, com maiores percentuais de reconhecimento. Concluímos que, o número de componentes mel-cepstrais adotadas, afeta diretamente no desempenho do sistema de identificação. Portanto, o aparelho fonador de um locutor fica melhor modelado com o emprego dos coeficientes estáticos e dinâmicos.

Devido ao tempo e aos recursos computacionais disponíveis, os modelos gerados utilizaram parâmetros não otimizados no processo de treinamento, como por exemplo, a mistura de apenas 16 gaussianas. O aumento do valor desse parâmetro para algo próximo do que é utilizado em sistemas reais como 256 gaussianas, resultaria em sistemas com melhores desempenhos. Sendo assim, todos os resultados obtidos nos testes foram satisfatórios, haja vista que, todos os locutores testados obtiveram o seu próprio modelo como o locutor mais provável.

Enfim, foi possível definir limiares que tornam o sistema mais seletivo no processo de identificação em uma aplicação hipotética, de acordo com o grau de segurança exigido na aplicação.

7.1 TRABALHOS FUTUROS

Finalmente, este Trabalho de Conclusão de Curso permite que trabalhos futuros possam ser realizados, com o intuito de aperfeiçoar o sistema de reconhecimento de locutor e poder aplicá-lo, por exemplo, em algumas atividades do dia-dia, conforme os itens descritos a seguir:

- Os sistemas de reconhecimento de locutor apresentados neste Trabalho de Conclusão de Curso foram implementados em um ambiente controlado, com pouquíssimas interferências de ruídos. Considerando isto, o processo de reconhecimento ficou facilitado. Então, uma proposta futura é realizar o reconhecimento de locutores em ambientes ruidosos;
- Em se tratando de aplicações no cotidiano, pode ser construído um sistema de reconhecimento de locutor para acesso de um local restrito em tempo real. Como este tipo de sistema fica sujeita aos diversos tipos de ruídos, um tratamento de sinal deverá ser realizado de modo a torná-lo mais robusto;
- Pode-se, a partir deste trabalho, elaborar um sistema de reconhecimento de locutor em um mesmo dispositivo que permita acesso à diferentes funções de acordo com o locutor.

REFERÊNCIAS

- ABRAHAM Alcaim. Frequência de ocorrência dos fones e listas de frases foneticamente balanceadas no português falado no rio de janeiro. **Revista da Sociedade Brasileira de Telecomunicações**, v. 7, p. 40–47, 1992. Citado na página 17.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. [S.l.]: Springer, 2006. Citado na página 12.
- CARDOSO, D. P. **Identificação de locutor usando modelos de mistura de gaussianas**. maio 2009. 88 f. Dissertação (Mestrado em Engenharia – Sistemas Eletrônicos) — Escola Politécnica da Universidade de São Paulo, São Paulo, 2009. Citado 3 vezes nas páginas 1, 3 e 15.
- CUADROS, C. D. R. **RECONHECIMENTO DE VOZ E DE LOCUTOR EM AMBIENTES RUIDOSOS: COMPARAÇÃO DAS TÉCNICAS MFCC E ZCPA**. maio 2007. 121 f. Dissertação (Pós-Graduação em Engenharia de Telecomunicações) — Escola de Engenharia da Universidade Federal Fluminense, Niterói, 2007. Citado 3 vezes nas páginas 1, 3 e 6.
- GORDILLO, C. A. **Reconhecimento de Voz Contínua Combinando os Atributos MFCC e PNCC com Métodos de Robustez SS, WD, MAP e FRN**. março 2013. 99 f. Dissertação (Mestrado em Engenharia Elétrica) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013. Citado 2 vezes nas páginas 10 e 11.
- MACHADO, A. F. **Conversão de Voz Inter-Linguística**. Maio 2013. 145 f. Tese (Doutorado em Ciências) — Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, 2013. Citado na página 10.
- MOLAU, S. et al. Computing mel-frequency cepstral coefficients on the power spectrum. In: IEEE. **Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on**. [S.l.], 2001. v. 1, p. 73–76. Citado na página 7.
- OPPENHEIM, A. V.; SCHAFER, R. W. **Processamento em tempo discreto de sinais**. [S.l.]: Pearson, 2013. v. 2013. Citado 6 vezes nas páginas 5, 6, 7, 8, 9 e 10.
- SIQUEIRA, J. K. **Reconhecimento de Voz Contínua com Atributos MFCC, SSCH e PNCC, Wavelet Denoising e Redes Neurais**. setembro 2011. 15 f. Dissertação (Mestrado em Engenharia Elétrica) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2011. Citado na página 5.
- YOUNG, S. J. et al. **The HTK Book Version 3.4**. [S.l.]: Cambridge University Press, 2006. Citado na página 11.

Apêndices

APÊNDICE A – TERMO DE CONFIDENCIALIDADE

MODELO 1:

TERMO DE COMPROMISSO, DE CONFIDENCIALIDADE DE DADOS E ENVIO DO RELATÓRIO FINAL

Eu, **Prof. Dr. Alberto Yoshihiro Nakano**, docente do Magistério Superior do curso de Engenharia Eletrônica da Universidade Tecnológica Federal do Paraná - campus Toledo, pesquisador responsável pelo projeto de pesquisa (Trabalho de conclusão de curso - TCC) intitulado **RECONHECIMENTO DE LOCUTOR UTILIZANDO MODELO DE MISTURAS GAUSSIANAS TREINADO PELO ALGORITMO DE MAXIMIZAÇÃO DA EXPECTATIVA**, comprometo-me a dar início a este estudo somente após apreciação e aprovação pelo Comitê de Ética em Pesquisa em Seres Humanos da Universidade Tecnológica Federal do Paraná e registro de aprovado na Plataforma Brasil.

Com relação à coleta de dados da pesquisa, nós pesquisadores, abaixo firmados, asseguramos que o caráter anônimo dos dados coletados nesta pesquisa será mantido e que suas identidades serão protegidas. Bem como as fichas clínicas e/ outros documentos não serão identificados pelo nome, mas por um código.

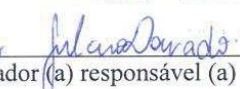
Nós pesquisadores, manteremos um registro de inclusão dos participantes de maneira sigilosa, contendo códigos, nomes e endereços para uso próprio. Os formulários: **Termo de Consentimento Livre e Esclarecido, Termo de Assentimento Livre e Esclarecido e /ou Termo de Consentimento de Uso de Voz e Imagem**, assinados pelos participantes serão mantidos pelo pesquisador em confidência estrita, juntos em um único arquivo.


Asseguramos que os participantes desta pesquisa receberão uma cópia do Termo de Consentimento Livre e Esclarecido; Termo de Assentimento Livre e Esclarecido; e/ou Termo de Consentimento de Uso de Voz e Imagem, que poderá ser solicitada de volta no caso deste não mais desejar participar da pesquisa.


Eu, como professor (a) orientador (a), declaro que este projeto de pesquisa, sob minha responsabilidade, será desenvolvido pelos alunos **Helio Rodrigues da Silva e Juliano Rodrigues Dourado** do curso de Engenharia Eletrônica da Universidade Tecnológica Federal do Paraná - campus Toledo.

Declaro, também, que li e entendi a Resolução 466/2012 (CNS) responsabilizando-me pelo andamento, realização e conclusão deste projeto e comprometendo-me a enviar ao CEP/UTFPR, relatório do projeto em tela quando da sua conclusão, ou a qualquer momento, se o estudo for interrompido.

Toledo, 16 de Fevereiro de 2018

Juliano Rodrigues Dourado 
Nome completo e assinatura do pesquisador (a) responsável (a)

Helio Rodrigues da Silva 
Nome completo e assinatura do pesquisador (a) responsável (a)

ALBERTO YOSHIHIRO NAKANO 
Nome completo e assinatura do prof. orientador (a), se houver

APÊNDICE B – MODELO TCLE TCUISV

TERMO DE CONSENTIMENTO E LIVRE ESCLARECIMENTO (TCLE) E TERMO DE CONSENTIMENTO PARA USO DE IMAGEM E SOM DE VOZ (TCUISV)

Título da pesquisa: Reconhecimento de locutor utilizando modelo de misturas gaussianas treinado pelo algoritmo de maximização da expectativa

Pesquisador (es/as) com Endereços e Telefones:

Hélio Rodrigues da Silva residente na Rua Vitória régia 1034, Bairro São Francisco, Toledo/PR. Tel: (45) 99848-2967

Juliano Rodrigues Dourado residente na Rua Caramuru 56, Distrito de Pérola Independente, Maripá/PR. Tel: (44) 99948-6830

Responsável pela pesquisa, com Endereços e Telefones:

Prof. Dr. Alberto Yoshihiro Nakano, Rua Cristo Rei, 19, Vila Becker, Toledo - Paraná; Tel: +55-(45)-3379-6847, e-mail: nakano@utfpr.edu.br

Local de realização da pesquisa: Universidade Tecnológica Federal do Paraná (UTFPR), campus Toledo.

Endereço, telefone do local: Rua Cristo Rei, 19, Vila Becker, Toledo - Paraná; Tel: +55-(45)-3379-6847.

A) INFORMAÇÕES AO PARTICIPANTE

Prezado (a) gostaria de convidá-lo (a) para participar da criação de um banco de dados de vozes para se empregado no Trabalho de Conclusão de Curso “Reconhecimento de locutor utilizando modelo de misturas gaussianas treinado pelo algoritmo de maximização da expectativa” sob responsabilidade dos alunos Hélio Rodrigues da Silva e Juliano Rodrigues Dourado orientado pelo Prof. Dr. Alberto Yoshihiro Nakano.

1. Apresentação da pesquisa.

Este estudo propõe desenvolver um sistema para reconhecimento automático de locutor. Para o desenvolvimento deste projeto são necessárias 5 (cinco) etapas, criação de um banco de dados de vozes, criação de modelos estatísticos, realização de testes, análise dos resultados e, por último, conclusões. Você estará participando da primeira etapa, a criação de um banco de dados de vozes.

Nesta pesquisa você será orientado a ler uma sequência de frases pré-definidas que será gravada em áudio. As frases são extraídas do trabalho: **“Frequência de ocorrência dos fones e listas de frases foneticamente balanceadas no português falado no Rio de Janeiro”, Alcaim, A.; Solewicz, A.; Moraes, J. A., Revista da Sociedade Brasileira de Telecomunicações, Vol. 7, nº.1 Dez. 1992.**

Estas frases serão posteriormente processadas para compor o banco de dados de vozes. Durante a realização do projeto o material será acessível apenas aos pesquisadores para a devida análise do material garantindo-se o sigilo a sua identidade e a sua não associação com os áudios gravados.

2. Objetivos da pesquisa.

Aplicar os conhecimentos adquiridos durante a graduação em engenharia eletrônica no desenvolvimento de um sistema de reconhecimento de locutor; Compreender como funciona sistemas biométricos, no caso, controlados pela voz; Levantamento de banco de dados de vozes para desenvolvimento de pesquisas científicas.

3. Participação na pesquisa.

Durante esta pesquisa, você estará sentado e um microfone será posicionado a sua frente sobre uma mesa ou afixado em sua vestimenta próximo a sua boca. Após você estar devidamente posicionado, uma lista de contendo 200 frases lhe será entregue. Você irá ler cada frase pronunciando-as naturalmente inserindo uma pausa entre as frases. Caso haja problemas na pronúncia de uma frase, a mesma deverá ser repetida. A gravação total terá duração entre 10 a 15 minutos, considerando as falas e as pausas.

4. Confidencialidade.

A privacidade será respeitada, ou seja, nome ou qualquer outro dado ou elemento que possa, de qualquer forma, identificar os participantes da pesquisa, será mantido em sigilo, garantindo a não utilização das informações em prejuízo das pessoas, inclusive em termos de auto estima, conforme previsto na Resolução 466/2012.

5. Riscos e Benefícios.

5a) Riscos:

A participação na pesquisa não traz risco físico, no entanto, pode ocorrer algum tipo de constrangimento, desconforto, por exemplo, cansaço e garganta seca, durante o decorrer das atividades. Neste caso os pesquisadores estão aptos a intervir, caso o desconforto não cessar o convidado (a) pode desistir da pesquisa a qualquer momento. Para os participantes na pesquisa a gravação (áudio) pode ser considerado de risco mínimo.

5b) Benefícios:

Após a pesquisa você poderá saber como a sua voz compõe o banco de dados de sistemas biométricos em função da voz. Adicionalmente, o banco de dados poderá ser usado na pesquisa e desenvolvimento de diversos trabalhos acadêmicos, sempre mantendo o sigilo dos dados dos convidados.

6. Critérios de inclusão e exclusão.

6a) Inclusão: Os critérios de inclusão dos participantes são: Ter no mínimo 18 anos; saber ler.

6b) Exclusão: Não se aplica.

7. Direito de sair da pesquisa e a esclarecimentos durante o processo.

Durante a sua participação na pesquisa, você tem o direito de deixar o estudo a qualquer momento, sem nenhum prejuízo ou coação. Da mesma forma, em qualquer momento, você tem o direito de receber esclarecimentos sobre quaisquer dúvidas que possam surgir. Em todo e qualquer momento você terá a liberdade de recusar ou retirar o seu consentimento sobre a sua participação na pesquisa, não ocorrendo nenhuma forma de penalização diante disso.

Você pode assinalar o campo a seguir, para receber o resultado desta pesquisa, caso seja de seu interesse :

quero receber os resultados da pesquisa (email para envio : _____)

não quero receber os resultados da pesquisa

8. Ressarcimento e indenização.

A pesquisa não traz nenhum custo aos participantes, desta forma não será necessário o ressarcimento aos participantes ou aos seus acompanhantes. Em caso de algum dano material ao participante da pesquisa, este terá direito a indenização. Caso haja algum dano ao participante, o pesquisador responsável entrará em contato para que seja escolhida a melhor alternativa para que ocorra a indenização.

B) CONSENTIMENTO (do participante de pesquisa ou do responsável legal – neste caso anexar documento que comprove parentesco/tutela/curatela)

Eu declaro ter conhecimento das informações contidas neste documento e ter recebido respostas claras às minhas questões a propósito da minha participação direta (ou indireta) na pesquisa e, adicionalmente, declaro ter compreendido o objetivo, a natureza, os riscos, benefícios, ressarcimento e indenização relacionados a este estudo.

Após reflexão e um tempo razoável, eu decidi, livre e voluntariamente, participar deste estudo, permitindo que os pesquisadores relacionados neste documento obtenham gravação de voz de minha pessoa para fins de pesquisa científica/ educacional. As gravações ficarão sob a propriedade do grupo de pesquisadores pertinentes ao estudo e sob sua guarda.

Concordo que o material e as informações obtidas relacionadas a minha pessoa possam ser publicados em aulas, congressos, eventos científicos, palestras ou periódicos científicos. Porém, não devo ser identificado por nome ou qualquer outra forma.

Estou consciente que posso deixar o projeto a qualquer momento, sem nenhum prejuízo. Após reflexão e um tempo razoável, eu decidi, livre e voluntariamente, participar deste estudo.

Nome Completo: _____

RG: _____ Data de Nascimento: ___/___/___ Telefone: _____

Endereço: _____

CEP: _____

Cidade: _____

Estado: _____

Assinatura: _____ Data: ___/___/___

Eu declaro ter apresentado o estudo, explicado seus objetivos, natureza, riscos e benefícios e ter respondido da melhor forma possível às questões formuladas.

Nome completo: _____

Assinatura pesquisador (a): _____ Data: ___/___/___
(ou seu representante)

Para todas as questões relativas ao estudo ou para se retirar do mesmo, poderão se comunicar com _____,
via e-mail: _____
ou telefone: _____.

ESCLARECIMENTOS SOBRE O COMITÊ DE ÉTICA EM PESQUISA:

O Comitê de Ética em Pesquisa envolvendo Seres Humanos (CEP) é constituído por uma equipe de profissionais com formação multidisciplinar que está trabalhando para assegurar o respeito aos seus direitos como participante de pesquisa. Ele tem por objetivo

Rubrica do Pesquisador

Rubrica do participante de pesquisa

avaliar se a pesquisa foi planejada e se será executada de forma ética. Se você considerar que a pesquisa não está sendo realizada da forma como você foi informado ou que você está sendo prejudicado de alguma forma, entre em contato com o Comitê de Ética em Pesquisa envolvendo Seres Humanos da Universidade Tecnológica Federal do Paraná (CEP/UTFPR). **Endereço:** Av. Sete de Setembro, 3165, Bloco N, Térreo, Bairro Rebouças, CEP 80230-901, Curitiba-PR, **Telefone:** (41) 3310-4494, **e-mail:** coep@utfpr.edu.br.

Contato do Comitê de Ética em Pesquisa que envolve seres humanos para denúncia, recurso ou reclamações do participante pesquisado:

Comitê de Ética em Pesquisa que envolve seres humanos da Universidade Tecnológica Federal do Paraná (CEP/UTFPR)

Endereço: Av. Sete de Setembro, 3165, Bloco N, Térreo, Rebouças, CEP 80230-901, Curitiba-PR, **Telefone:** 3310-4494, **E-mail:** coep@utfpr.edu.br

OBS: este documento deve conter 2 (duas) vias iguais, sendo uma pertencente ao pesquisador e outra ao participante da pesquisa.

APÊNDICE C – PROJETO DETALHADO

PESQUISA PARA O TRABALHO DE CONCLUSÃO DE CURSO EM ENGENHARIA ELETRÔNICA DA UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ (UTFPR) CAMPUS TOLEDO

Título:

Reconhecimento automático de locutor utilizando modelo de misturas gaussianas treinado pelo algoritmo de maximização da expectativa

Grupo de Pesquisadores:

Orientador: **Prof. Dr. Alberto Yoshihiro Nakano**

Pesquisadores (alunos de graduação):

Hélio Rodrigues da Silva

Juliano Rodrigues Dourado

Financiamento:

Financiamento próprio.

Desenho:

Gravação de áudio de 20 pessoas do sexo masculino e 20 do sexo feminino; Idade igual ou superior a 18 anos.

Resumo

Desenvolver sistemas que possam reconhecer ou identificar indivíduos vem se tornando uma necessidade cada vez maior em aplicações que exigem a verificação e a garantia da identidade humana. Há vários sistemas que utilizam a biometria para reconhecer um determinado indivíduo, dentre estes, o reconhecimento automático de locutor que utiliza a fala como dado de reconhecimento. Neste trabalho, os parâmetros acústicos mel-cepstrais serão extraídos para modelagem de locutores através de ferramentas estatísticas. Para realizar a modelagem do trato vocal de um indivíduo, utilizar-se-á o modelo de misturas gaussianas (GMM). Os parâmetros do modelo GMM são adaptados ou treinados pelo algoritmo de maximização da expectativa. O reconhecimento de locutor se mostra promissor para diversas aplicações práticas, compreendendo tarefas que possam vir a facilitar, agilizar e melhorar processos de verificação de identidade.

Palavras Chaves

Gaussian Mixture Model (GMM) - Modelo de Mistura de Gaussianas, *Expectation Maximization (EM)* - Maximização da Expectativa, Reconhecimento de locutor

Introdução

Utilizar a voz humana como uma característica biométrica vem se tornando uma técnica simples e segura a ser empregada nas mais diversas situações de controle e investigação como, por exemplo, no acesso a dispositivos pessoais e acesso a locais restritos. Esta segurança e simplicidade advêm da extração de características acústicas do sinal de fala de maneira não invasiva e seu emprego na modelagem de locutores por ferramentas estatísticas. A biometria é a parte da ciência que busca analisar e quantificar dados biológicos. Existem diversas formas, tais como impressões digitais, retinas e íris, reconhecimento de locutor (padrões de voz), padrões faciais e medições de mão. Para o reconhecimento de um indivíduo pela sua voz é necessário captar o som por um microfone e extrair dados do sinal captado. A partir destes sinais criar modelos estatísticos que irão representar determinado locutor. Posteriormente, o processo de reconhecimento consiste em verificar qual o modelo mais provável que gerou a sequência de teste. Existem diversas formas para se trabalhar com o problema de reconhecimento de um locutor. Neste trabalho, utilizar-se-á de coeficientes cepstrais de frequência Mel (MFCC), como parâmetros extraídos. A modelagem de cada locutor será feita por modelo de misturas de gaussianas (GMM). Será utilizado o algoritmo de Maximização da Expectativa (EM), que estima parâmetros para modelos estatísticos com base na máxima verossimilhança. O critério para reconhecimento de locutor é baseado também no critério de máxima verossimilhança empregando-se o conjunto de modelos criados. Para dar prosseguimento ao trabalho de pesquisa, há necessidade de se criar um banco de dados de áudio contendo amostras de locuções vocais. Para permitir isso, submete-se este projeto para a avaliação do Comitê de Ética em Pesquisa Envolvendo Seres Humanos para coleta e armazenamento de áudio para fins de pesquisa.

Hipótese

Este projeto não emprega o conceito de hipótese, mas apenas de verificação de técnica já existente em reconhecimento de locutor, sendo que, possuir banco de dados de áudio é fundamental para o desenvolvimento do projeto.

Objetivo Primário

O objetivo deste trabalho é estudar o problema de reconhecimento de locutor independente de texto e para isso há a necessidade de criar um banco de dados de vozes.

Metodologia Proposta

O (A) voluntário (a) será instruído (a) dos procedimentos empregados para aquisição de amostras de áudio por um dos pesquisadores do projeto. Após as dúvidas serem esclarecidas o (a) voluntário (a) será conduzido (a) ao local onde será realizada a aquisição de áudio. O (A)

voluntário (a) deverá estar sentado (a) e um microfone será posicionado à sua frente sobre uma mesa ou afixado em sua vestimenta próximo à boca. Após o (a) voluntário (a) estar devidamente posicionado (a), uma lista de conteúdo com 200 frases lidas será entregue. O (A) voluntário (a) deverá ler cada frase pronunciando-as naturalmente inserindo uma pausa entre as frases. Caso haja problemas na pronúncia de uma frase, a mesma deverá ser repetida. A gravação total terá duração entre 10 a 15 minutos, considerando as falas e as pausas. Terminada a leitura das frases, o (a) voluntário (a) estará dispensado (a).

Critério de Inclusão

Os critérios de inclusão dos participantes são: Ter no mínimo 18 anos; saber ler.

Critério de Exclusão

Não se aplica

Riscos

A participação na pesquisa não traz risco físico, no entanto, pode ocorrer algum tipo de constrangimento, desconforto, por exemplo, cansaço e garganta seca, durante o decorrer das atividades. Neste caso os pesquisadores estão aptos a intervir, caso o desconforto não cessar o convidado (a) pode desistir da pesquisa a qualquer momento. Para os participantes na pesquisa a gravação (áudio) pode ser considerado de risco mínimo.

Benefícios

Após a pesquisa você poderá saber como a sua voz compõe o banco de dados de sistemas biométricos em função da voz. Adicionalmente, o banco de dados poderá ser usado na pesquisa e desenvolvimento de diversos trabalhos acadêmicos, sempre mantendo o sigilo dos dados dos convidados.

Metodologia de análise de dados

Os dados de áudio serão separados em conjunto de treinamento e conjunto de teste. O conjunto de treinamento será empregado para a criação de modelos estatísticos representando cada locutor (voluntário). A etapa de teste emprega os modelos anteriormente criados e o conjunto de dados de teste para verificar a validade do modelo com o critério de máxima verossimilhança.

Desfecho Primário

Gravação de áudio de 20 pessoas do sexo masculino e 20 do sexo feminino; Criação do sistema de reconhecimento de locutor; e realização de teste de verificação do sistema.

Tamanho da Amostra

40 indivíduos voluntários para coleta e amostras de voz.

Cronograma de Execução:

Atividade	Mar. 2018	Abr. 2018	Mai. 2018	Jun. 2018
Criação do Banco de dados	X	X		
Modelagem dos locutores	X	X	X	
Testes com o sistema			X	X
Elaboração do relatório final	X	X	X	X
Defesa do Trabalho de Conclusão de curso				X

- Criação do banco de dados: gravação das vozes dos locutores, e armazenamento do conteúdo;
- Modelagem dos locutores: utilizar as ferramentas computacionais para criar os modelos ótimos para cada locutor;
- Testes com o sistema: efetuar testes a partir dos modelos criados, obtendo os resultados da implementação do projeto, e aplicação em um sistema de acesso a local restrito (abertura de uma porta por voz);
- Elaboração do relatório final: descrever como se comportou o sistema implementado, quais foram os resultados obtidos e se os mesmos satisfazem o esperado;
- Defesa do TCC: apresentar os resultados do trabalho conforme os objetivos definidos.

Orçamento

Identificação do Orçamento	Tipo	Valor (R\$)
Água	Custeio	50,00

Referências

CARDOSO, Denis Pirttiho. Identificação de locutor usando modelos de mistura de gaussianas. maio 2009. 88 f. Dissertação (Mestrado em Engenharia - Sistemas Eletrônicos) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2009.

CUADROS, Carlos Daniel Riquelme. RECONHECIMENTO DE VOZ E DE LOCUTOR EM AMBIENTES RUIDOSOS: COMPARAÇÃO DAS TÉCNICAS MFCC E ZCPA. maio 2007. 121 f. Dissertação (Pós-Graduação em Engenharia de Telecomunicações) - Escola de Engenharia da Universidade Federal Fluminense, Niterói, 2007.

MOLAU, Sirko; PITZ, Michael; SCHLUTER, Ralf; NEY, Hermann. Computing melfrequency cepstral coefficients on the power spectrum. In: IEEE. Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on., 2001. v. 1, p. 73–76.

SIQUEIRA, Jan Krueger. Reconhecimento de Voz Contínua com Atributos MFCC, SSCH e PNCC, Wavelet Denoising e Redes Neurais. setembro 2011. 15 f. Dissertação (Mestrado em Engenharia Elétrica) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2011.

ALCAIM, A.; SOLEWICZ, A.; MORAES, J. A., “Frequência de ocorrência dos fones e listas de frases foneticamente balanceadas no português falado no Rio de Janeiro,” Revista da Sociedade Brasileira de Telecomunicações, Vol. 7, nº.1 Dez. 1992.

APÊNDICE D – USO DAS INSTALAÇÕES



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Toledo
Coordenação do Curso de Engenharia Eletrônica
COELE



DECLARAÇÃO PARA USO DE LABORATÓRIOS

A respeito do pedido do **Prof. Dr. Alberto Yoshihiro Nakano**, lotado na Coordenação de Engenharia Eletrônica (COELE), sobre a utilização do espaço físico de laboratórios da COELE para gravação de áudio que será empregado no Trabalho de Conclusão de Curso (TCC2) dos alunos **Hélio Rodrigues da Silva** e **Juliano Rodrigues Dourado** intitulado: “RECONHECIMENTO DE LOCUTOR UTILIZANDO MODELO DE MISTURAS GAUSSIANAS TREINADO PELO ALGORITMO DE MAXIMIZAÇÃO DA EXPECTATIVA”, respeitando as condições de que:

1. as gravações não serão realizadas em horário de aula dos alunos orientandos;
2. as gravações não virão a atrapalhar as atividades desenvolvidas por outros docentes/discentes/estagiários/técnicos e outros usuários do laboratório;

o pedido está deferido.

Prof. Ms. Jorge Augusto Vasconcelos Alves
Coordenador do Curso de Engenharia Eletrônica
UTFPR - campus Toledo

APÊNDICE E – CLASSIFICAÇÃO NBEST - SISTEMA 1

Quadro 1 – Classificação nBest - Sistema 1

Modelos dos Locutores

	MM1	MM2	MF3	MM4	MM5	MM6	MM7	MF8	MF9	MF10	MM11	MF12	MF13	MF14	MF15	MM16	MF17	MM18	MM19	M20	MF21	MF22	MF23	MF24	MF25	MM26	MF27	MF28	MM29	MF30	MF31	MM32	MM33	MF34	MF35	MF36	MM37	MM38	MM39	MM40
L1	1	36	29	2	8	35	39	28	24	17	5	10	7	25	40	12	13	27	32	34	31	11	22	19	26	23	38	20	37	33	15	4	3	18	30	14	6	21	16	9
L2	17	1	7	5	10	6	39	31	21	30	4	27	16	33	40	24	25	8	13	14	36	18	23	19	35	32	38	20	37	28	26	2	12	34	15	22	3	9	29	11
L3	30	33	1	29	34	35	36	12	6	8	26	7	20	21	40	31	5	37	24	18	22	23	17	19	3	14	10	4	39	11	13	15	27	9	16	2	25	38	32	28
L4	2	35	28	1	10	29	39	31	14	19	8	12	3	24	40	17	18	32	27	34	25	5	30	20	26	22	38	15	37	36	21	6	4	11	33	16	9	23	13	7
L5	4	34	30	3	1	35	39	24	11	16	10	17	6	31	40	2	14	29	27	33	21	13	20	22	23	28	38	26	37	36	19	9	5	15	32	18	7	25	12	8
L6	11	15	30	6	16	1	39	33	18	20	8	21	17	22	40	32	27	2	24	31	36	14	29	23	35	28	38	9	37	34	19	12	5	25	10	26	3	7	13	4
L7	20	22	12	16	31	32	1	33	11	14	17	10	21	35	7	19	26	40	3	18	23	28	38	36	9	37	5	6	29	4	34	30	8	27	39	24	2	25	13	15
L8	28	36	16	25	29	38	40	1	15	8	31	10	9	12	39	22	3	35	26	5	18	20	2	4	19	13	14	7	37	32	11	21	23	6	33	17	24	34	30	27
L9	27	37	12	18	26	32	39	13	1	2	25	3	7	19	40	21	10	36	33	28	9	14	23	15	4	20	34	8	38	31	11	16	17	5	30	6	22	35	29	24
L10	27	37	16	25	31	34	39	11	5	1	24	4	13	18	40	30	6	35	33	28	10	22	15	17	8	14	32	9	38	26	7	12	19	3	20	2	23	36	29	21
L11	4	28	29	5	11	31	39	35	25	16	1	13	10	27	40	12	9	32	22	36	34	14	17	19	26	33	38	20	37	30	18	6	3	21	15	23	2	24	7	8
L12	26	38	22	24	32	36	40	15	3	8	27	1	7	11	39	31	10	35	29	16	18	17	14	20	13	12	28	2	37	19	6	9	23	5	30	4	25	33	34	21
L13	16	37	29	10	22	38	39	14	8	19	25	6	1	4	40	24	3	33	28	23	21	7	17	2	27	11	35	9	36	31	13	12	15	5	32	18	26	34	30	20
L14	22	37	16	27	32	38	40	13	21	19	23	5	3	1	39	33	8	34	20	11	25	18	9	6	30	2	12	4	24	17	7	14	26	10	29	15	31	35	36	28
L15	35	38	10	24	36	40	11	17	7	8	29	9	12	23	1	31	15	39	6	4	27	19	32	26	13	30	2	5	16	3	22	20	25	14	37	18	21	33	34	28
L16	15	34	25	4	2	36	39	16	3	12	24	13	6	30	40	1	8	31	23	27	10	20	17	21	9	32	38	26	37	33	29	14	11	19	28	22	5	35	18	7
L17	28	35	20	22	29	38	39	7	11	10	21	9	3	16	40	23	1	34	31	25	12	19	2	4	15	13	33	8	37	32	6	17	24	5	18	14	26	36	30	27
L18	20	17	27	11	8	3	39	15	19	29	9	30	16	26	40	18	10	1	31	13	34	23	4	12	35	22	38	14	37	36	7	25	21	28	2	32	5	33	24	6
L19	15	30	25	2	12	40	36	31	21	28	11	18	3	24	37	17	19	38	1	8	32	4	34	22	27	33	29	13	14	10	35	6	7	20	39	26	5	16	23	9
L20	31	33	12	21	25	38	40	2	20	22	28	15	3	6	39	23	8	24	4	1	35	17	13	9	29	7	10	5	32	14	18	11	26	16	34	19	27	37	36	30
L21	27	38	28	23	25	34	39	15	4	7	26	9	6	19	40	18	2	36	30	31	1	17	12	10	5	22	32	8	35	33	11	24	20	3	29	13	16	37	21	14
L22	16	35	24	3	29	36	39	26	10	22	13	8	2	20	40	31	9	33	19	21	23	1	27	14	25	18	37	4	38	34	11	6	17	5	32	12	15	30	28	7
L23	24	35	23	28	27	36	39	7	22	15	18	8	4	13	40	21	2	31	33	26	17	30	1	6	19	5	34	12	38	32	3	14	25	10	11	9	20	37	29	16
L24	23	35	13	22	31	38	40	8	16	18	26	9	2	6	39	24	3	33	25	12	17	19	4	1	27	5	28	11	37	30	7	10	20	15	21	14	29	36	34	32
L25	30	39	10	27	34	38	37	12	3	5	29	7	15	20	33	31	8	40	25	21	4	18	19	16	1	23	13	6	36	11	17	14	22	9	28	2	26	35	32	24
L26	16	36	19	23	18	37	39	12	22	15	20	8	3	2	40	28	7	29	31	14	24	17	6	5	30	1	34	9	38	33	4	13	21	11	26	10	27	35	32	25
L27	32	39	6	30	34	38	35	3	12	13	28	9	17	11	22	33	4	40	19	5	18	25	21	15	14	20	1	2	24	8	7	23	27	10	31	16	26	37	36	29
L28	28	38	17	24	32	36	39	22	8	9	23	2	7	12	40	35	15	37	18	13	27	6	26	29	19	16	21	1	34	10	5	11	25	3	33	4	14	31	30	20
L29	21	39	30	12	27	38	40	20	11	22	29	8	3	10	26	15	16	36	4	6	25	17	34	28	35	31	7	2	1	13	24	18	19	5	37	23	14	32	33	9
L30	29	37	9	31	36	39	35	16	8	6	30	2	14	18	21	34	11	38	20	10	28	27	19	22	12	17	7	5	33	1	15	4	23	13	25	3	24	32	40	26
L31	21	36	27	25	24	32	39	16	10	12	22	2	6	13	40	29	4	31	34	30	14	18	7	11	19	3	35	9	38	33	1	15	17	5	26	8	20	37	28	23
L32	5	34	25	4	18	36	39	27	22	19	9	8	3	23	40	21	11	35	32	30	33	16	12	13	26	17	38	20	37	29	15	1	6	14	24	2	10	31	28	7
L33	2	36	32	4	9	31	39	27	17	12	5	13	8	25	40	11	15	34	28	35	19	16	26	22	20	24	38	21	37	33	23	6	1	10	30	18	3	29	14	7
L34	23	38	29	19	26	32	39	14	11	4	24	3	2	15	40	25	7	34	31	30	10	21	13	17	20	8	36	9	37	33	6	12	18	1	27	5	22	35	28	16
L35	15	34	20	19	14	35	39	27	26	16	5	25	11	29	40	23	6	8	28	33	30	24	3	22	31	18	38	13	37	32	7	9	21	10	1	12	2	36	17	4
L36	26	37	15	24	27	33	39	16	13	2	25	6	12	18	40	31	7	34	32	29	14	23	10	17	9	8	35	11	38	28	4	3	21	5	20	1	22	36	30	19
L37	4	30	23	7	10	24	38	33	25	13	3	21	18	35	40	11	16	32	15	36	28	17	22	31	26	34	39	12	37	27	20	8	6	19	9	14	1	29	5	2
L38	2	13	31	6	9	14	39	32	28	20	4	12	11	34	40	29	26	15	18	27	36	10	33	24	35	30	38	17	37	19	23	5	3	21	25	16	7	1	22	8
L39	6	34	27	5	9	31	39	28	19	15	2	22	10	24	40	8	13	32	25	35	16	11	21	26	20	29	38	17	37	36	18	12	4	14	30	23	3	33	1	7
L40	5	36	32	4	16	25	39	30	23	10	6	12	11	26	40	19	18	28	31	35	22	20	21	27	29	24	38	15	37	34	17	3	7	9	13	8	2	33	14	1

Anexos

ANEXO A – FRASES BALANCEADAS

Lista 01

1. A questão foi retomada no congresso.
2. Leila tem um lindo jardim.
3. O analfabetismo é a vergonha do país.
4. A casa foi vendida sem pressa.
5. Trabalhando com união rende muito mais.
6. Recebi nosso amigo para almoçar.
7. A justiça é a única vencedora.
8. Isso se resolvera de forma tranquila.
9. Os pesquisadores acreditam nessa teoria.
10. Sei que atingiremos o objetivo.

Lista 02

- Nosso telefone quebrou.
2. Desculpe se magoei o velho.
 3. Queremos discutir o orçamento.
 4. Ela tem muita fome.
 5. Uma índia andava na mata.
 6. Zé, vá mais rápido!
 7. Hoje dormirei bem.
 8. Joao deu pouco dinheiro.
 9. Ainda são seis horas.
 10. Ela saia discretamente.

Lista 03

1. Eu vi logo a Iôô e o Léo.
2. Um homem não caminha sem um fim.
3. Vi Zé fazer essas viagens seis vezes.
4. O atabaque do Tito é encoberto com pele de gato.
5. Ele lê no leito de palha.
6. Paira um ar de arara rara no Rio Real.
7. Foi muito difícil entender a canção.
8. Depois do almoço te encontro.
9. Esses são nossos times.
10. Procurei Maria na copa.

Lista 04

1. A pesca é proibida nesse lago.
2. Espero te achar bem quando voltar.
3. Temos muito orgulho da nossa gente.
4. O inspetor fez a vistoria completa.
5. Ainda não se sabe o dia da maratona.
6. Será muito difícil conseguir que eu venha.
7. A paixão dele e a natureza.
8. Você quer me dizer a data?
9. Desculpe, mas me atrasei no casamento.
10. Faz um desvio em direção ao mar!

Lista 05

1. A velha leoa ainda aceita combater.
2. É hora do homem se humanizar mais.
3. Ela ficou na fazenda por uma hora.

4. Seu crime foi totalmente encoberto.
5. A escuridão da garagem assustou a criança.
6. Ontem não pude fazer minha ginastica.
7. Comer quindim e sempre uma boa pedida.
8. Hoje eu irei precisar de você.
9. Sem ele o tempo flui num ritmo suave.
10. A sujeira lançada no rio contamina os peixes.

Lista 06

1. O jogo será transmitido bem tarde.
2. E possível que ele já esteja fora de perigo.
3. A explicação pode ser encontrada na tese.
4. Meu voo tinha sido marcado para as cinco.
5. Daqui a pouco a gente ira pousar.
6. Estou certo que mereço a atenção dela.
7. Era um belo enfeite todo de palha.
8. O comércio daqui tem funcionado bem.
9. É a minha chance de esclarecer a noticia
10. A visita transformou-se numa reunião intima.

Lista 07

1. O cenário da história é um subúrbio do Rio.
2. Eu tenho ótima razão para festejar.
3. A pequena nave medira o campo magnético.
4. O premio será entregue sem sessão solene.
5. A ação se passa numa cidade calma.
6. Ela e a namorado vão a Portugal de navio.
7. O adiamento surpreendeu a mim e a todos.
8. A gente sempre colhe o que plantou.
9. Aqui é onde existem as flores mais interessantes.
10. A corrida de inverno aconteceu com vibração.

Lista 08

1. Esse empreendimento será de enorme sucesso.
2. As feiras livres não funcionam amanhã.
3. Fumar é muito prejudicial a saúde.
4. Entre com seu código e o número da conta.
5. Reflita antes e discuta depois.
6. As aulas dele são bastante agradáveis.
7. Usar aditivos pode ser desastroso.
8. O clima não e mau em Calcutá.
9. A locomotiva vem sem muita carga.
10. Ainda é uma boa temporada para o cinema.

Lista 09

1. Os maiores picos da Terra ficam debaixo d' agua.
2. A inauguração da vila e quarta-feira.
3. S6 vota quem tiver o titulo de eleitor.
4. É fundamental buscar a razão da existência.
5. A temperatura só é boa mais cedo.
6. Em muitas regiões a população esta diminuindo.
7. Nunca se pode ficar em cima do muro.

8. Pra quem vê de fora o panorama é desolador.
9. É bom te ver colhendo flores.
10. Eu me banho no lago ao amanhecer.

Lista 10

1. É fundamental chegar a uma solução comum.
2. Ha previsão de muito nevoeiro no Rio.
3. Muitos móveis virão às cinco da tarde.
4. A casa pode desabar em algumas horas.
5. O candidato falou como se estivesse eleito.
6. A ideia é falha, mas interessa.
7. O dia esta bom para passear no quintal.
8. Minhas correspondências não estão em casa.
9. A saída para a crise dele é o dialogo.
10. Finalmente o mau tempo deixou o continente.

Lista 11

1. Um casal de gatos come no telhado.
2. A cantora foi apresentar seu grande sucesso.
3. Lá é um lugar ótimo para tomar uns chopinhos.
4. O musical consumiu sete meses de ensaio.
5. Nosso baile inicia após as nove.
6. Apesar desses resultados, tomarei uma decisão.
7. A verdade não poupa nem as celebridades.
8. As queimadas devem diminuir este ano.
9. O Vão entre o trem e a plataforma e muito grande.
10. Infelizmente não compareci ao encontro.

Lista 12

1. As crianças conheceram o filhote de ema.
2. A bolsa de valores ficou em baixa.
3. O congresso volta arras em sua palavra.
4. A médica receitou que eles mudassem de clima.
5. Não é permitido fumar no interior do ônibus.
6. A apresentação foi cancelada por causa do som.
7. Uma garota foi presa ontem anoite.
8. O prato do dia e couve com atum.
9. Eu viajarei ao Canada amanhã.
10. A balsa e o meio de transporte daqui.

Lista 13

1. O grêmio ganhou a quadra de esportes.
2. Hoje irei a vila sem meu filho.
3. Essa magia não acontece todo dia.
4. Será bom que você estude esse assunto.
5. O menu incluía pratos bem saborosos.
6. Podia dizer as horas, por favor?
7. A casa é ornamentada com flores do campo.
8. A Terra e farta, mas não infinita.
9. O sinal emitido é captado por receptores.
10. A mensalidade aumentou mais que a inflação.

Lista 14

1. O tele-jornal termina às sete da noite.
2. A cabine telefônica fica na próxima rua.
3. Defender a ecologia e manter a vida.
4. Nesse verão o calor está insuportável.
5. Um jardim exige muito trabalho.
6. O mamão que eu comprei estava ótimo.
7. Meu primo falara com a gerência amanhã.
8. De dia apague a luz sempre.
9. A sociedade uruguaia tem que se mobilizar.
10. Suas atitudes são bem calmas.

Lista 15

1. Dezenas de cabos eleitorais buscavam apoio.
2. A vitória foi paga com muito sangue.
3. Nossa filha tem amor por animais.
4. Esse peixe é mais fatal que certas cobras.
5. O time continua lutando pelo sucesso.
6. Essa medida foi devidamente alterada.
7. O estilete é uma arma perigosa.
8. Aguarde quinta eu venho jantar em casa.
9. A mudança é lenta, porém duradoura.
10. O clima não é mais seco no interior.

Lista 16

1. A sensibilidade indicara a escolha.
2. A Amazônia é a reserva ecológica do globo.
3. O ministério mudou demais com a eleição.
4. Novos rumos se abrem para a informática.
5. O capital de uma empresa depende da produção.
6. Se não fosse ela, tudo teria sido contido.
7. A principal personagem no filme é uma gueixa.
8. Receba seu jornal em sua casa.
9. A juventude tinha que revolucionar a escola.
10. A atriz terá quatro meses para ensaiar seu canto.

Lista 17

1. Muito prazer em conhecê-lo.
2. Eles estavam sem um bom equipamento.
3. O sol ilumina a fachada de tarde.
4. A correção do exame está coerente.
5. As portas são antigas.
6. Sobrevoamos Natal acima das nuvens.
7. Trabalhei mais do que podia.
8. Hoje eu acordei muito calma.
9. Esse canal é pouco informativo.
10. Parece que nascemos ontem.

Lista 18

1. Receba meus parabéns pela apresentação.
2. Eu planejo uma viagem no feriado.
3. Nalado de cá do rio ha uma boa sombra.

4. A maioria dos visitantes gosta deste monumento.
5. Minha filha é especialista em musica sacra.
6. A casa só tem um quarto.
7. A duração do simpósio é de cinco dias.
8. Ao contrário de nossa expectativa, correu tranquilo.
9. A intenção é obter apoio do governante.
10. A fila aumentou ao longo do dia.

Lista 19

1. À noite a temperatura deve ir á zero.
2. A proposta foi inspecionada pela gerência.
3. O quadro mostra uma face do cotidiano.
4. Já era bem tarde quando ele me abordou.
5. O canário canta ao amanhecer.
6. A lojinha fica bem na esquina de casa.
7. Meu time se consagrou como o melhor.
8. Um instituto deve servir a sua meta.
9. Ele entende quando se fala pausadamente.
10. Seu saldo bancário esta baixo.

Lista 20

1. O termômetro marcava um grau.
2. O discurso de abertura é bem longo.
3. Eu precisei de microfone na conferência.
4. Joyce esticou sua temporada ate quinta.
5. Nada como um almoço ao ar livre.
6. Nossa filha e a primeira aluna da classe.
7. Gostaria de deitar um pouco.
8. Não fizemos uma viagem muito cansativa.
9. Ainda tenho cinco telefonemas para dar.
10. Os hotéis do sudoeste são fantásticos.