

UNIVERSIDADE FEDERAL DE VIÇOSA
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

ELAINE SANTOS

**ESTUDO DO MOVIMENTO FACIAL COM A ACÚSTICA DA FALA
AVALIANDO O EFEITO LOMBARD SOBRE DIFERENTES RUÍDOS**

VIÇOSA
2010

ELAINE SANTOS

**ESTUDO DO MOVIMENTO FACIAL COM A ACÚSTICA DA FALA
AVALIANDO O EFEITO LOMBARD SOBRE DIFERENTES RUÍDOS**

Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 490 – Monografia e Seminário e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. Kétia Soares Moreira

VIÇOSA
2010

ELAINE SANTOS

**ESTUDO DO MOVIMENTO FACIAL COM A ACÚSTICA DA FALA
AVALIANDO O EFEITO LOMBARD SOBRE DIFERENTES RUÍDOS**

Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 490 – Monografia e Seminário e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.

Aprovada em de de 2010.

COMISSÃO EXAMINADORA

Prof. Dr. Kétia Soares Moreira- Orientador
Universidade Federal de Viçosa

Prof. Dr. Denílson Eduardo Rodrigues - Membro
Universidade Federal de Viçosa

Mestranda Paula Oliveira LaGatta - Membro
Universidade Federal de Juiz de Fora

“Ame muitas coisas, porque em amar está a verdadeira força. Quem ama muito conquistará muito, e o que for feito com amor estará bem feito.”

(Vicent Van Gogh)

Dedico esse trabalho, primeiramente a Deus, que é quem me dá forças nas horas difíceis e por realizar coisas maravilhosas na minha vida.

Dedico também aos meus tão queridos e amados, pais Maria das Graças e Maurício e minha não menos amada irmã Juliana pelo amor, dedicação, carinho, por sempre me incentivar nos estudos, por acreditar no meu potencial e por ser referência na minha vida. A vocês dedico esse trabalho com todo meu amor.

Agradecimentos

Tenho muitos agradecimentos a fazer, mas em primeiro lugar não poderia deixar de agradecer a Deus e Maria por terem me dado muita força e coragem em todos esses anos e terem me ajudado a chegar até aqui. A Eles devo tudo.

Minha mãe Maria, muito obrigada por todos esses sacrifícios para me manter esse tempo todo aqui. Obrigada pelo seu amor incondicional e pela sua simplicidade que me impulsiona a lutar cada vez mais e me inspira a me tornar sempre melhor.

A meu pai Maurício, muito obrigada por todos os conselhos e por essa educação maravilhosa que me ajudou a chegar onde estou com muita dignidade e honestidade. Obrigada também pelo seu amor e por todos os sacrifícios por mim.

A minha irmã Juliana, muito obrigada por todo seu exemplo que me fez enxergar que estudar sempre foi melhor opção. Obrigada por todos os “puxões de orelha” que me fizeram crescer e encarar a vida com mais maturidade. Obrigada também por todos os sacrifícios.

A minha madrinha Antonieta que sempre torceu por mim e sempre me apoiou e me ajudou durante todos esses anos.

Ao meu namorado Marcus Vinícius que se tornou minha base de sustentação em tudo o que eu faço na minha vida desde que nos conhecemos. Obrigada pelo seu amor, carinho e paciência.

A todos os meus amigos pela amizade e por todo o carinho dedicado. Laélia, Alexandra, Débora, Márcia, Rodrigo, Bárbara, Vanessa, Ana Paula a vocês um carinho especial!

À Ita por toda a amizade e companheirismo que sempre me ajudou muito desde que a conheci. Ao José Maria e Natália por todo o carinho.

Aos meus tios, avós, primos por toda a amizade.

A minha professora e orientadora desse trabalho Kétia que me estendeu a mão num momento muito difícil de minha vida e que sempre me orienta e me apóia.

A todos outros que porventura tenha esquecido, mas que com certeza foram de fundamental importância para que esse trabalho e a realização de um sonho acontecessem.
Muito obrigada!

Resumo

Esse trabalho é um estudo sobre relação entre o movimento facial com a acústica da fala, avaliando o efeito Lombard sobre variados níveis de ruído. A análise do efeito Lombard é de grande importância, pois visa o comportamento humano na produção e execução da fala, à melhoria de vida do cotidiano das pessoas com necessidades especiais, bem como uma melhor integração homem-máquina. Desta maneira, o estudo entre áudio e movimento, pode ser aplicado em diversas ferramentas áudios-visuais. Através de análise de parâmetros e da utilização de ferramentas matemáticas, como componentes principais, a locução feminina de uma sentença captada por um equipamento Optotrak, concedido pelo grupo de pesquisa CEFALA – UFMG será avaliada sob diferentes aspectos.

Abstract

This work is a study on relation enters the face movement with the acoustics of speaks, evaluating the effect Lombard on varied noise levels. The analysis of the effect Lombard is of great importance, therefore it aims at the human behavior in the production and execution of speaks, to the improvement of life of the daily one of the people with necessities special, as well as one better man-machine integration. In this way, the audio study between and movement, it can be applied in diverse tools audio-appearances. Through analysis of parameters and the use of mathematical tools, as component main, the feminine locution of a sentence caught for Optotrak equipment, granted for the group of research CEFALA - UFMG will be evaluated under different aspects.

Sumário

1	Introdução	14
1.1	Objetivo geral.....	15
2	Materiais e Métodos	16
2.1	Efeito Lombard	16
2.2	Aparelho de captação de dados: Optotrak	16
2.3	Análise em componente principal: PCA	18
2.4	Matriz de correlação cruzada	19
2.5	Variância acumulada.....	20
2.6	Parâmetros LPC e LSP.....	20
2.7	Base de Dados.....	21
3	Resultados e Discussões	32
4	Conclusões	36
	Referências Bibliográficas.....	37

Lista de Tabelas

Tabela 1 – Tabela de Variância Acumulada dos autovetores.....	33
Tabela 2 – Correlação cruzada do movimento para 6 autovetores.....	34
Tabela 3 – Correlação cruzada do movimento para 2 autovetores.....	34
Tabela 4 – Correlação cruzada da acústica para 6 autovetores	34
Tabela 5 – Correlação cruzada da acústica para 2 autovetores	34
Tabela 6 – Correlação Cruzada Movimento x Acústica.....	34

1 Introdução

Ao longo de muitos anos o desenvolvimento tecnológico tem beneficiado a vida humana em diversas áreas, contribuindo para uma vida melhor para a população. É por esse motivo que se torna relevante qualquer estudo para esse fim, pois qualquer que seja o problema (locomoção, comunicação, audição, visão, etc.) há sempre uma nova linha de estudo tentando solucionar tais indagações.

Acompanhando esse raciocínio grandes estudos sobre produção acústica humana já foram feitos e ainda continuam sendo executados. Tais estudos têm o intuito de entender e resolver cada vez mais os questionamentos sobre fala, leitura labial, movimentos faciais, movimentos visuais da fala, etc. e aplicá-los de forma a tornar mais rápida, fácil e sofisticada o dia-a-dia da população. Um sistema que faça a comunicação homem-máquina através de uma leitura labial que capta os movimentos labiais os transfere e codifica para o computador é um exemplo disso.

O processo de comunicação humana apesar de não ser tão simplório, resumidamente se dá por meio de vibrações das cordas vocais, de movimentos musculares dos lábios, da face e da cabeça. Idealmente essa transmissão ocorre sem nenhuma interferência, mas na realidade existem ruídos nos canais que prejudicam tal mecanismo.

Ao se fazer um estudo sobre sinais correlacionados como os movimentos de partes da face e a acústica, pode-se usar Análise em Componentes Principais (PCA). A PCA é utilizada para se obter as características mais importantes dos sinais para uma posterior análise (Draper, Baek, e Bartlett, 2003; Brunelli e Poggio, 1991; Chen, Liao, e Lin, 2001; Iwano, Tamura, e Furui, 2001; Arandjelovi e Cipolla, 2004; Lamar, Bhuiyan, e Iwata, 1999; Vatikiotis-Bateson, Yehia, e Kuratate, 2002; Barbosa e Yehia, 2001; Vatikiotis-Bateson e Yehia, 2000; Vatikiotis-Bateson, Kuratate, e Yehia, 1998b; Yehia, Rubin, e Vatikiotis-Bateson, 1998; Vatikiotis-Bateson e Yehia, 1997).

O movimento facial é fortemente ligado à acústica da fala. A componente linear dessa ligação pode ser modelada por meio dos autovetores da matriz de correlação

cruzada entre a posição de pontos da face do locutor e parâmetros extraídos da acústica da fala. De forma similar, a ligação entre regiões diferentes da própria face pode ser modelada pelos autovetores da matriz de covariância das posições dos pontos da face. Observa-se, entretanto que, tanto os padrões de movimento da face, quanto ligação entre o movimento facial e a acústica variam com o tempo. (Moreira, 2008).

Nesse trabalho em particular, utilizando-se de um aparelho Optotrak obteve-se o banco de dados para um estudo sobre a análise do movimento facial no momento da produção da fala avaliando o efeito Lombard.

1.1 Objetivo geral

O objetivo desse trabalho é fazer um estudo do movimento facial e da acústica da fala considerando uma sentença gravada através de uma locutora num aparelho Optotrak sobre diferentes ruídos avaliando a influência do “Efeito Lombard”.

Como objetivo específico tem-se:

Avaliar as frequências em Hz (frequência fundamental), RMS, LSP;

Fazer um estudo sobre a variância acumulada nos diferentes níveis de ruídos de 38 dB, 55 dB, 65 dB, 75 dB e 85 dB;

Fazer um estudo da correlação entre os diferentes níveis de ruídos de 38 dB, 55 dB, 65 dB, 75 dB e 85 dB usando análise em componente principal (PCA);

2 Materiais e Métodos

2.1 Efeito Lombard

O Efeito Lombard é um fenômeno observado entre os seres humanos e está relacionado com o nível de ruído que por convenções é medido em Decibel.

A partir do momento em que o ruído ambiente ultrapassa 30 dB o nível sonoro médio da fala tende a aumentar. Tal fenômeno foi definido por Lombard em 1911 e batizado de efeito Lombard. Esse efeito pode ser observado tanto em ouvintes normais quanto em indivíduos com perdas auditivas (Quedas, 2007).

Tal fenômeno será observado na sentença que foi proferida por um locutor feminino e avaliado se existe alguma alteração que modifica a correlação entre a acústica da fala e movimento facial.

2.2 Aparelho de captação de dados: Optotrak

Para a captação de dados de um determinado estudo de análise de movimento facial e acústica da fala, necessita-se de um aparelho chamado Optotrak. Tal aparelho faz a gravação dos dados da acústica bem como a captação dos dados do movimento dos marcadores em três dimensões: x, y, z. Ele é normalmente usado em indústrias, universidades e instituições de pesquisa para aquisição de dados em várias linhas de estudo. A dimensão x mostra a posição do marcador no eixo x. A dimensão y mostra a posição do marcador no eixo y. Na dimensão z tem-se o movimento da cabeça.

Nesse trabalho o uso do Optotrak para aquisição dos dados foi concedido pelo grupo de pesquisa CEFALA – UFMG. Utilizando-se de doze marcadores sendo que oito deles contornavam a boca da locutora e os quatro restantes ficavam posicionados num capacete. As sentenças foram gravadas e transferidas ao computador no formato Excel. Posteriormente optou-se por trabalhar com apenas uma sentença: “Leila tem um lindo jardim.”. Os marcadores ao redor da boca objetivam captar os movimentos da mesma enquanto era proferida a sentença. Os marcadores posicionados no capacete têm o objetivo de captar o movimento da cabeça que involuntariamente acompanha a fala de

uma pessoa enquanto ela conversa, para posteriormente ser possível retirar esse movimento para não atrapalhar a análise.



Figura 1 – Locutora usando Optotrak para aquisição dos dados

O aparelho gravou os movimentos dos marcadores numa frequência de 100 Hz e os dados da acústica numa frequência de 22050 Hz. Esses dados foram convertidos em dados numéricos e tabelados no formato Microsoft Excel.

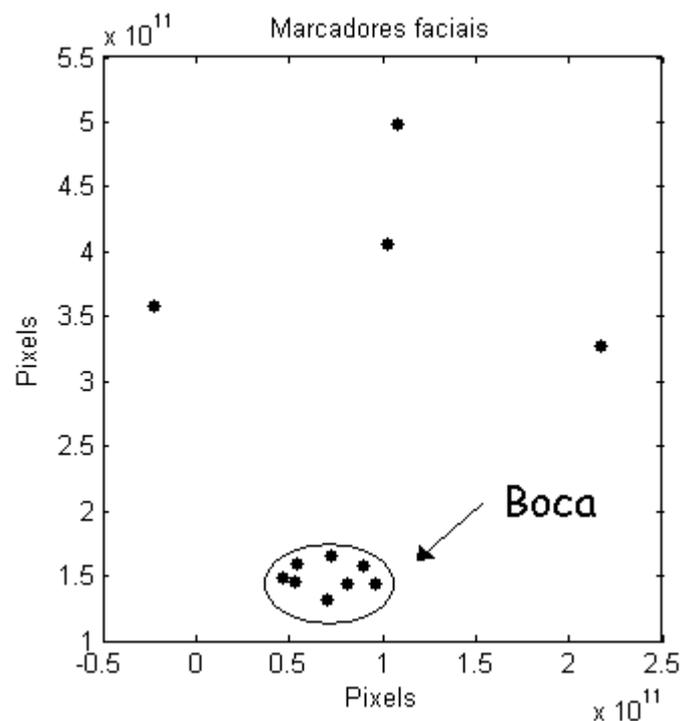


Figura 2 – Posição dos marcadores na face e na cabeça

Os oito primeiros marcadores se localizam ao redor da boca, já os quatro últimos ficam nas pontas do capacete (Fig. 1) em volta da cabeça.

2.3 **Análise em componente principal: PCA**

A análise de componentes principais (PCA) é um processo que permite projetar o espaço original de variáveis num espaço de dimensão mais reduzida. As variáveis derivadas das originais são designadas por componentes principais. O procedimento de PCA faz uma partição da variância nas componentes principais sendo que cada componente principal é calculada por forma a reter a maior quantidade de variância presente nas variáveis originais (Durães, 2009). A utilização da análise de componentes principais tem por finalidade determinar novas variáveis, que para a metodologia proposta seja capaz de ter confiabilidade, ou seja, através de uma combinação linear das variáveis originais determinam-se novas variáveis de análise de componentes principais (Lopes, Samohyl, 2003).

Matematicamente cada componente principal é uma combinação linear das variáveis originais que irá gerar os autovetores. Esses autovetores representam uma direção que é preservada por essa combinação. A Eq. 1 ilustra como se obtém a componente principal $T_{i,j}$ a partir de um vetor amostra de dados $(X(i,:))$ e da matriz de cargas P.

$$T_{i,j} = X(i,:)P(:,j) = X_{i,1}P_{1,j} + X_{i,2}P_{2,j} + \dots + X_{i,m}P_{m,j} \quad (1)$$

Na Eq. 1, $T_{i,j}$ é a componente principal, j (contagem) obtida a partir da amostra $X(i,:)$ e do coeficiente (carga) P correspondendo a componente j. As cargas constituem uma base ortonormal para o subespaço da matriz X (Lopes, Samohyl, 2003). Geometricamente as componentes principais podem ser vistas como projeções dos dados originais sobre eixos ortogonais (cargas) que cobrem o espaço das variáveis. O primeiro eixo é determinado por forma a capturar o máximo de variância passando através das zonas mais densas dos dados. A segunda componente principal é determinada através do mesmo critério com a restrição de que deve ser ortogonal em relação ao primeiro. O processo repete-se para todos os eixos (Durães, 2009).

A principal vantagem em se usar PCA é que além de reduzir a dimensionalidade de um grupo de dados ela retém aquelas características do conjunto de dados que

contribuem mais a sua variância, mantendo uma ordem de baixo nível das componentes principais e ignorando as de alto nível.

2.4 Matriz de correlação cruzada

A correlação indica a força e a direção de uma relação linear entre duas variáveis aleatórias. Considera-se que duas variáveis quantitativas estão correlacionadas quando os valores de uma delas variam sistematicamente com respeito aos valores homônimos da outra: se temos duas variáveis (A e B) existe correlação se ao aumentar os valores de A fazem-no também os de B e vice-versa (Aguirre, 2007).

Existem diversos coeficientes que medem o grau de correlação, adaptados à natureza dos dados. O mais conhecido é o coeficiente de correlação de Pearson, como pode ser visto na Eq.2, que se obtém dividindo a covariância de duas variáveis pelo produto de seus desvios regulares (Yehia, Rubin, Vatikiotis-Bateson, 1998).

$$p = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2)$$

Onde, $x_1, x_2 \dots x_n$ e $y_1, y_2 \dots y_n$ são os valores medidos de ambas as variáveis; $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ e $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Usando essa ferramenta foi feita a correlação cruzada do movimento e da acústica a qual podemos chamar de “Matriz de Correlação”. Tal artifício foi utilizado para o quanto um sinal está correlacionado com outro e vice-versa.

A Função de Correlação Cruzada (FCC) entre dois sinais $u(t)$ e $y(t)$ é definida como (Aguirre, 2007):

$$r_{uy}(\tau, t) = E[u(t)y(t + \tau)] = \lim_T \frac{1}{2T} \int_{-T}^T u(t)y(t + \tau) \quad (3)$$

No caso discreto, a FCC torna-se:

$$r_{uy}(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{i=-N}^N u(i)y(i + k) \quad (4)$$

Onde $E[.]$ é a esperança matemática (Aguirre, 2007).

2.5 **Variância acumulada**

Em teoria de probabilidade e estatística, a variância de uma variável aleatória é uma medida da sua dispersão estatística, indicando quão longe em geral os seus valores se encontram do valor esperado (Moraes, 2007).

Seja $\mu = E(X)$ o valor esperado (média) da variável aleatória X . A variância pode ser dada por:

$$\text{var}(x) = E((X - \mu)^2) \quad (5)$$

Isto é, o valor esperado do quadrado do desvio de X da sua própria média. Em linguagem comum isto pode ser expresso como “a média do quadrado da distância de cada ponto até a média”. É assim a “média do quadrado dos desvios” (Moraes, 2007). A variância da variável aleatória “ X ” é geralmente designada por $\text{var}(x), \sigma_x^2$ (Moraes, 2007).

No referido trabalho a variância acumulada foi adquirida para todos os níveis de ruídos avaliando os autovetores que cada marcador carrega. Muitos dos algoritmos clássicos de processamento de sinal recorrem, de uma ou de outra forma, a métodos de decorrelação de dados. As vantagens de tais técnicas baseiam-se num princípio relativamente simples: ao decorrelacionar os dados, estamos a eliminar parte da informação redundante em cada dimensão. Certos sinais, como é o caso de sinais fala ou de imagens, que apresentam uma substancial correlação entre amostras ou “pixels” consecutivos, são os candidatos ideais para os algoritmos de decorrelação. Os benefícios resultantes são vários: os dados podem ser descritos de uma forma mais concisa, certas características escondidas dos dados podem vir à luz depois de transformadas, a distribuição dos dados pode ser representada (aproximadamente) pelas

densidades individuais de cada dimensão (Marques, 2005/2006). Por esse motivo é que se utilizou de variância acumulada e de componentes principais.

2.6 Parâmetros LPC e LSP

No final dos anos 60 foi desenvolvido um sintetizador de voz chamado LPC (Linear Prediction Coding). Constituído por um filtro de dez ou mais pólos, um gerador de impulsos e ruído, e um amplificador variável, sendo hoje em dia empregados em sistemas de aparelhos celulares (Alencar; Alcaim, 2007).

Os esquemas de codificação de voz usados em sistemas de comunicações móveis e redes IP operam a baixas taxas de bits e utilizam, em geral, codificação preditiva linear ou LPC (*Linear Predictive Coding*), com base em um modelo de produção da fala. Nesse modelo, um sinal de excitação é aplicado a um filtro tudo-pólo (caracterizado por parâmetros LPC), que representa a informação da envoltória espectral do sinal de voz (Alencar; Alcaim, 2007).

É uma das técnicas de análise de discurso mais poderoso, e um dos métodos mais úteis para a codificação de voz de boa qualidade a uma baixa taxa de bits e fornece estimativas precisas de parâmetros extremamente expostos (Gray, 2006).

Os parâmetros LSP são um mapeamento dos parâmetros ou coeficientes de representações LPC, e relacionam-se com as frequências formantes do trato vocal (Werneck, Malta, Carneiro, Yehia). São de grande popularidade uma vez que garantem a estabilidade do indicador, e os erros espectrais são locais para os pequenos desvios do coeficiente (Gray, 2006).

Em linguagem gráfica usando o MATLAB foi desenvolvido funções para os parâmetros LPC. Transmitindo o preditor linear tal que a soma dos quadrados dos erros foi minimizada pode-se fazer o mapeamento dos parâmetros LPC definindo os coeficientes LSP.

Os coeficientes LSP estão ligados a frequências de ressonância do trato vocal sendo que o mesmo sofre influência do formato da face (Moreira, 2008).

2.7 Base de Dados

Inicialmente os dados utilizados nesse trabalho foram coletados no laboratório do grupo de pesquisa CEFALA-UFMG, usando um aparelho Optotrak. Foram gravadas 10 pequenas sentenças, ou seja, 10 frases curtas de um locutor feminino. Posteriormente optou-se por usar nesse trabalho apenas uma sentença: “Leila tem um lindo jardim.”

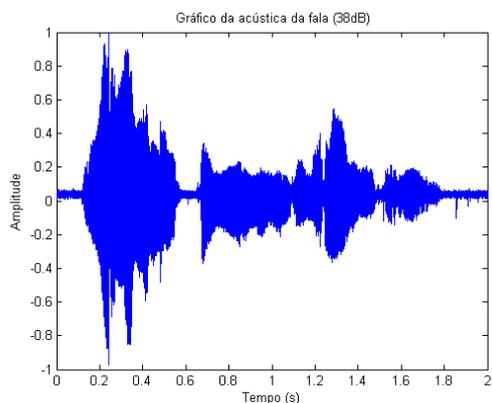


Figura 3 – “Leila tem um lindo jardim” - 38dB

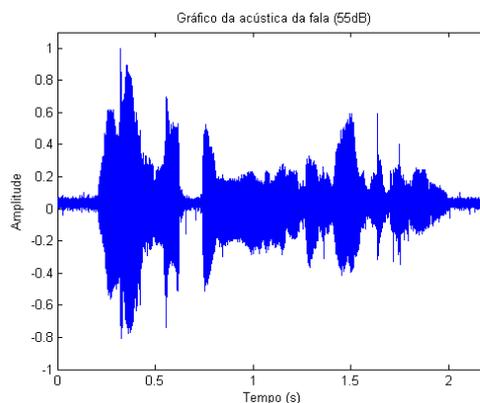


Figura 4 – “Leila tem um lindo jardim” - 55dB

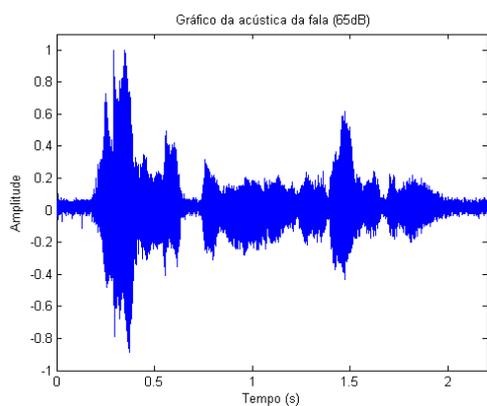


Figura 5 – “Leila tem um lindo jardim” - 65dB

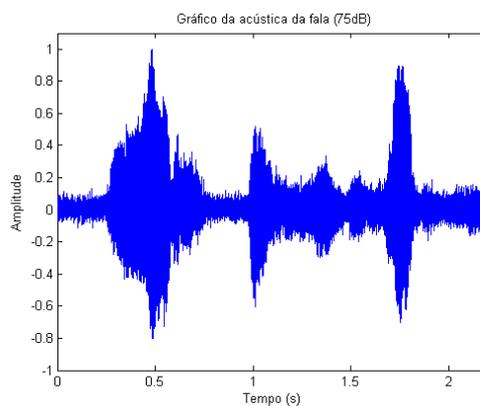


Figura 6 – “Leila tem um lindo jardim” - 75dB

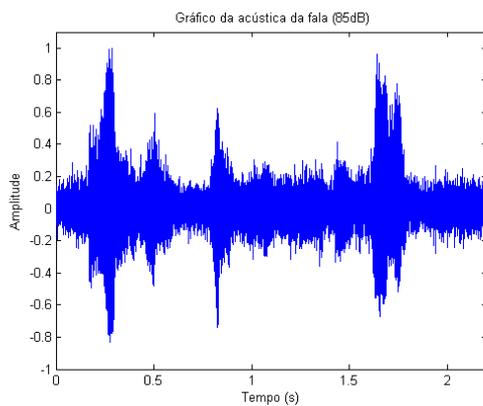


Figura 7 – “Leila tem um lindo jardim” - 85dB

A faixa de ruído utilizada na elocução varia de 38 dB a 85 dB, sendo caracterizada por:

Ruído a 38 dB → Frequência audível normal, sem barulho.

Ruído a 55 dB → Frequência audível normal, com barulho.

Ruído a 65 dB → Frequência audível normal, com barulho incomodando.

Ruído a 75 dB → Frequência audível no limiar, com barulho incomodando.

Ruído a 85 dB → Frequência audível excedente (dor no Tímpano), com muito barulho.

Observe que nas figuras 6 e 7 o ruído é alto, com amplitude média variando entre 0,12 a 0,23. Os áudios foram normalizados, sendo que a amplitude não ultrapassa de uma unidade.

As funções foram definidas no programa gráfico MATLAB e a Eq. 6 é a função referente aos gráficos das figuras 3, 4, 5, 6 e 7. Observe:

$X(n)$ → Matriz de áudio com n frames

F_s → Frequência do áudio

$V = \left[\frac{X(1), X(2), X(3), \dots, X(n)}{F_s} \right]$ → Vetor contendo o tamanho de X

$x = \frac{X}{\text{Máx}(X(n))}$ → Normalizando

$$y = x(n)V(n) \tag{6}$$

O estudo proposto foi avaliar a frequência fundamental da voz da elocução para diferentes níveis de ruído. A frequência fundamental é o que conhecemos como voz aguda (fina) ou grave (grossa) e relaciona-se, principalmente, com as características referentes ao tamanho e espessura das pregas vocais, e seu valor é extraído a partir do número de vibrações que a prega vocal realiza em um segundo (Mariano, 2005). No homem adulto espera-se que a frequência fundamental esteja ao redor de 80 a 150 Hz, pois o mesmo tem as pregas vocais maiores e espessas. Já a mulher adulta espera-se que

a frequência fundamental esteja ao redor de 150 a 250 Hz, pois a mesma tem a prega vocal maior e menos espessa (Mariano, 2005).

A frequência fundamental foi adquirida porque indica diretamente a vocalização, ou seja, as cordas vocais estão vibrando. Essas figuras demonstram que a vibração pode ser alterada de acordo com o ruído.

Assim foram retiradas as frequências fundamentais mostradas nas figuras de 8 a 12. Observe:

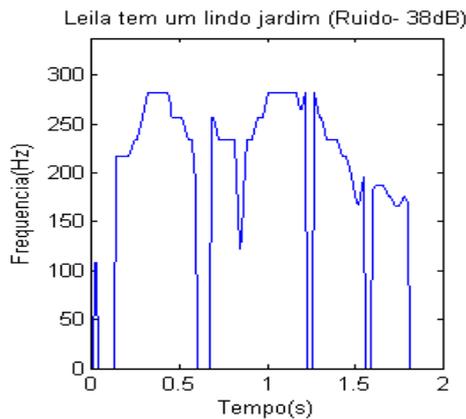


Figura 8 – Frequência em Hz a 38dB

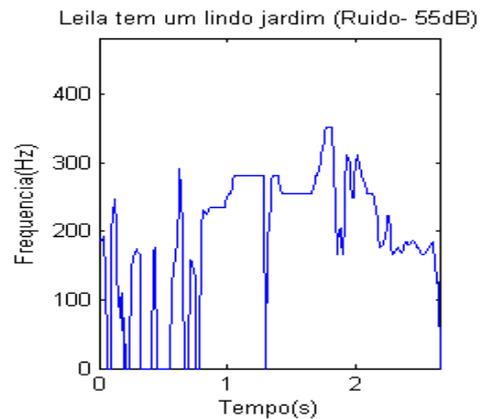


Figura 9 – Frequência em Hz a 55dB

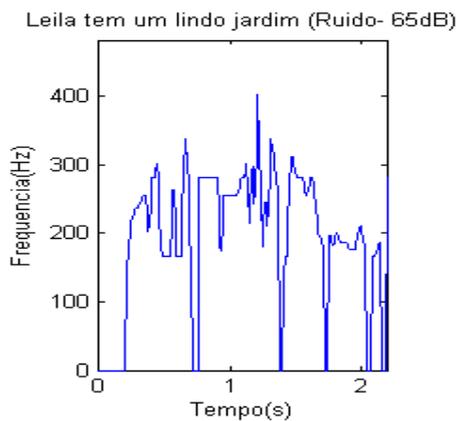


Figura 10 – Frequência em Hz a 65dB

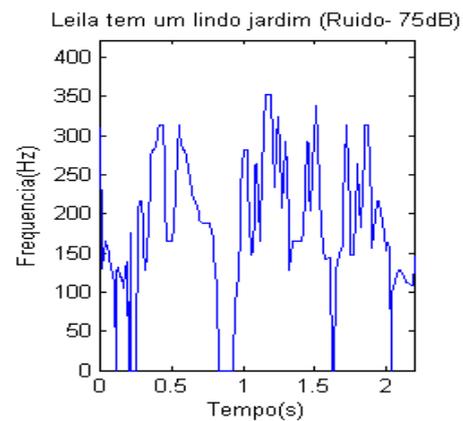


Figura 11 – Frequência em Hz a 75dB

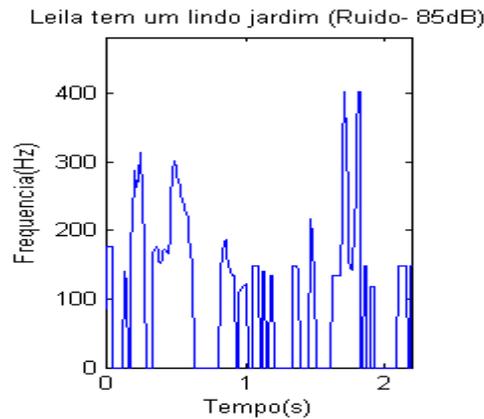


Figura 12 – Frequência em Hz a 85dB

Como mencionado anteriormente, o trato vocal é um tubo flexível limitado entre as pregas Vocais (cordas Vocais) até os lábios e as narinas. Nele, ocorre a passagem de ar, vindo dos pulmões, criando uma pressão entre suas paredes (estruturas flexíveis) e vibrando as pregas vocais. Essa onda sonora inicialmente produzida pelas pregas vocais é modificada pela geometria (estrutura) do trato vocal, ou seja, pela ressonância do trato vocal. Ligados a freqüências de ressonância do trato vocal (formantes) estão os coeficientes LSP. Vale salientar que, o formato do trato vocal tem forte influência sobre o formato da face. Assim, para verificar alguma alteração influenciada pelo “Efeito Lombard” uma análise dos parâmetros LSP (freqüências LSP) em Hz foi feita. Nas figuras 13-17 temos o resultado dessa análise.

Na representação do sinal acústico por meio da freqüência em LSP, definiram-se inicialmente os parâmetros LPC.

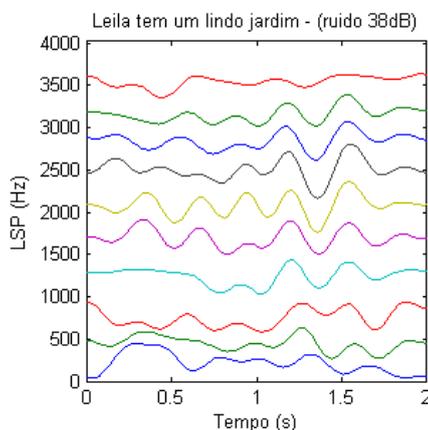


Figura 13 – Frequência em LSP a 38dB

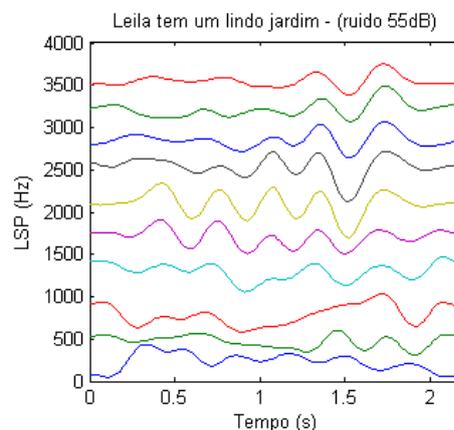


Figura 14 – Frequência em LSP a 55dB

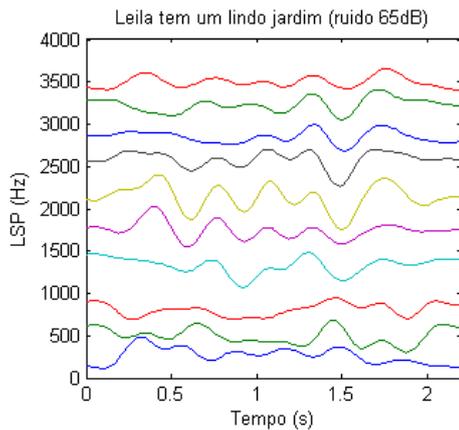


Figura 15 – Frequência em LSP a 65dB

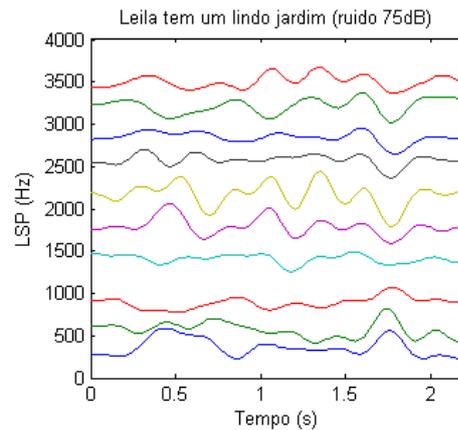


Figura 16 – Frequência em LSP a 75dB

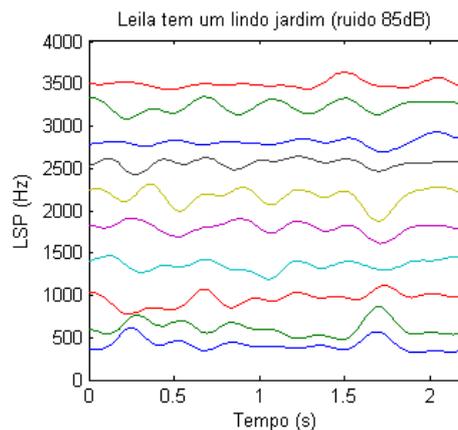


Figura 17 – Frequência em LSP a 85dB

Outra característica analisada no sinal de voz é a energia RMS. As energias em RMS de 55 dB, 65 dB, 75 dB e 85 dB, foram normalizadas utilizando os valores de 38 dB. Ou seja, os gráficos das figuras 19, 20, 21 e 22 mostram o quanto energia da sentença proferida com seu ruído respectivo é maior que a energia da sentença proferida a um ruído de 38 dB.

Basicamente para se obter a RMS dos dados, extraiu-se a raiz quadrada da média dos dados elevados ao quadrado. Como resultado, obteve-se que à medida que o ruído aumenta, o locutor profere a sentença com uma maior energia, o que confirma a existência do “Efeito Lombard”. Chegando a aumentar cerca de cinco vezes para um ruído de 85 DB.

Analisando as frequências em RMS temos as seguintes figuras:

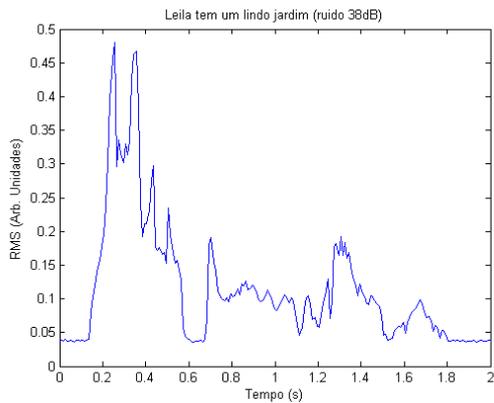


Figura 18 – Energia em RMS a 38dB

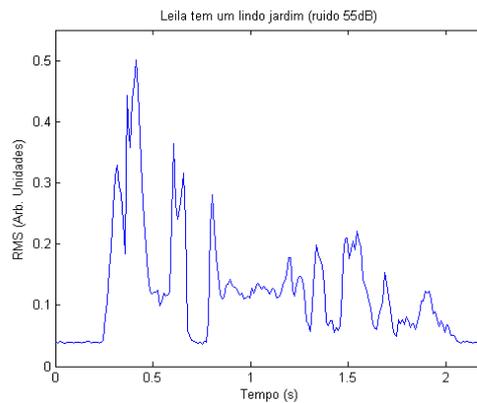


Figura 19 – Energia em RMS a 55dB

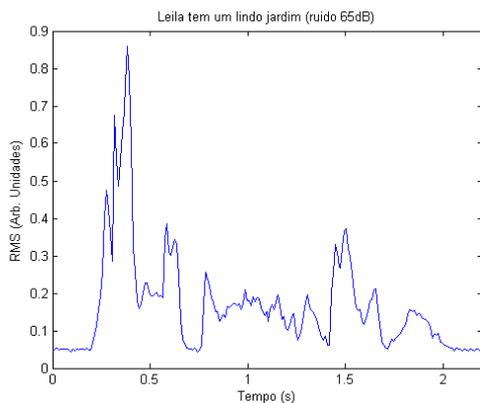


Figura 20 – Energia em RMS a 65dB

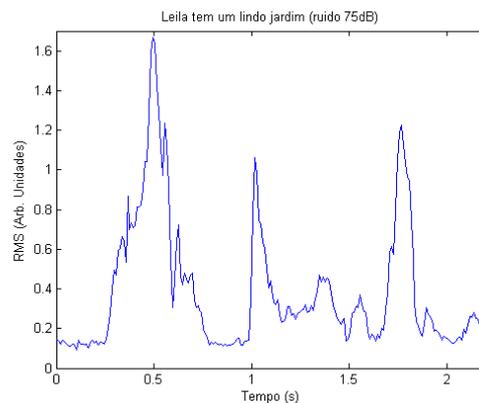


Figura 21 – Energia em RMS a 75dB

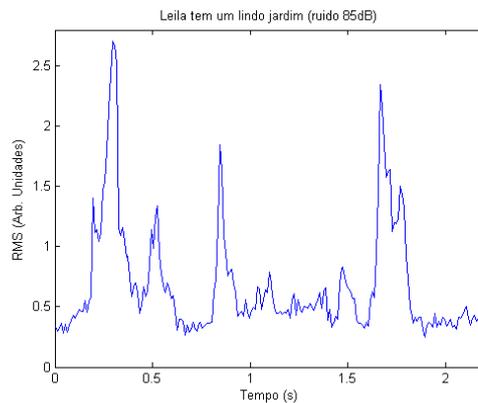


Figura 22 – Energia em RMS a 85dB

Até o momento toda análise voltou-se a acústica da fala. Entretanto, outro foco desse trabalho relaciona ao movimento facial. Para a análise da correlação entre os movimentos faciais e a acústica da fala, utilizou-se das componentes principais.

Inicialmente, calcula-se o autovetor de uma matriz que contém os movimentos dos sensores no eixo X e Y. Sendo que cada coluna da matriz X corresponde a um frame. Obtendo como resultado uma matriz onde as colunas da matriz U são os

autovetores da matriz de covariância ($X0*X0'$) e os elementos ao longo da diagonal de S são os respectivos autovalores. A matriz $X0$ corresponde a matriz com a posição dos sensores subtraída do movimento médio.

Para variância acumulada foi retirada a porcentagem de informação em componentes principais dos autovetores. Essa variância foi obtida com o arquivo de dados do movimento e através dela é possível perceber qual é o mínimo de informação que cada autovetor carrega. A Fig. 28 faz um comparativo da variância acumulada dos ruídos analisados.

Analisando a variância acumulada temos os seguintes gráficos:

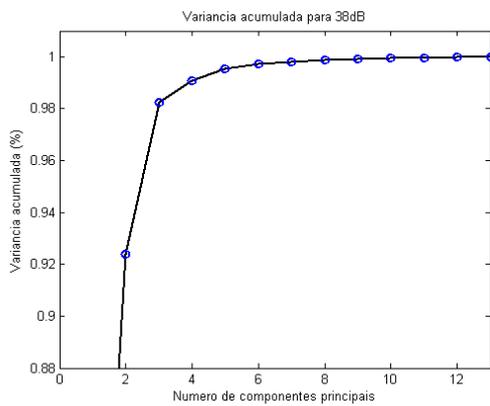


Figura 23 – Variância acumulada 38dB

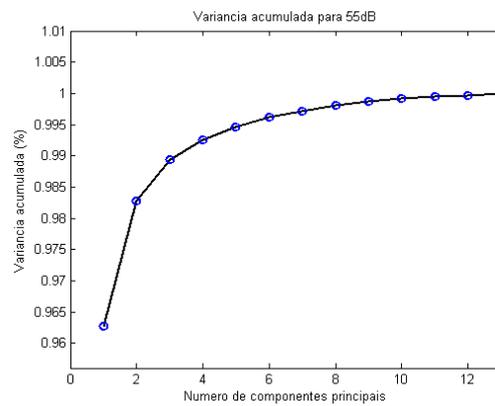


Figura 24 – Variância acumulada 55dB

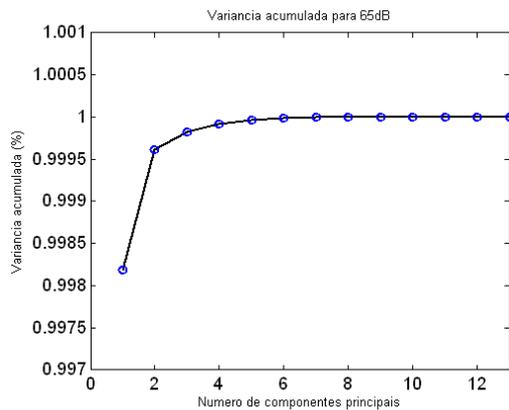


Figura 25 – Variância acumulada 65dB

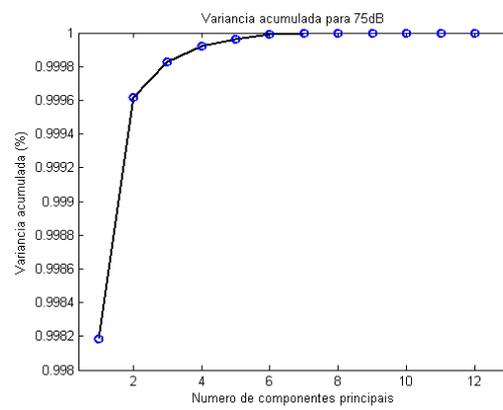


Figura 26 – Variância acumulada 75dB

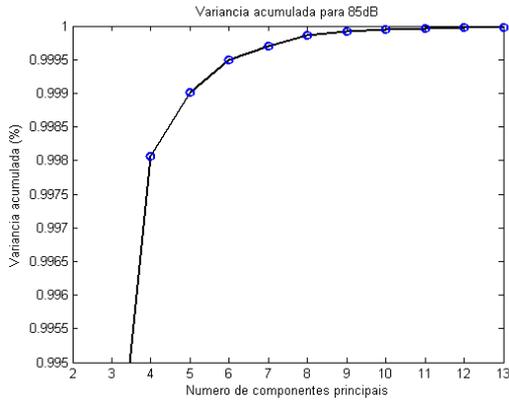


Figura 27 – Variância acumulada 85dB

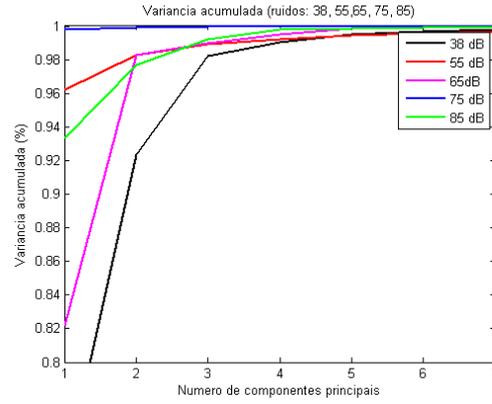


Figura 28 – Variância acumulada do conjunto

Outra comparação foi feita analisando a distância Euclidiana ou distância Métrica entre autovetor de uma sentença total e o autovetor de uma janela de 2 segundos deslocada de 0,3 segundos. Essa distância mostra a distância entre dois pontos, que pode ser provada pela aplicação repetida do Teorema de Pitágoras.

A distância Euclidiana da Fig. 29 foi calculada da seguinte forma:

$$d_{jk} = \sqrt{\sum_{i=1}^{2N} (\bar{u}_{ik} - u_{ijk})^2} \quad (7)$$

Onde: $\bar{u}_{ik} \rightarrow$ i-ésimo marcador do k-ésimo autovetor de referência sendo calculado com base no trecho completo; $u_{ijk} \rightarrow$ representa o i-ésimo marcador do k-ésimo autovetor calculado com base na j-ésima janela analisada; e d_{jk} é distância do k-ésimo autovetor calculado com base na j-ésima janela analisada e o autovetor de referência correspondente (Moreira, 2008).

A Eq. 7 está representada no gráfico da Fig. 29 e mostra o quanto cada autovetor analisado se distancia da informação original de acordo com o ruído exposto. Ela informa a variabilidade dos autovetores ao longo do tempo. Estudos mostram que o primeiro autovetor permanece mais estável ao longo do tempo.

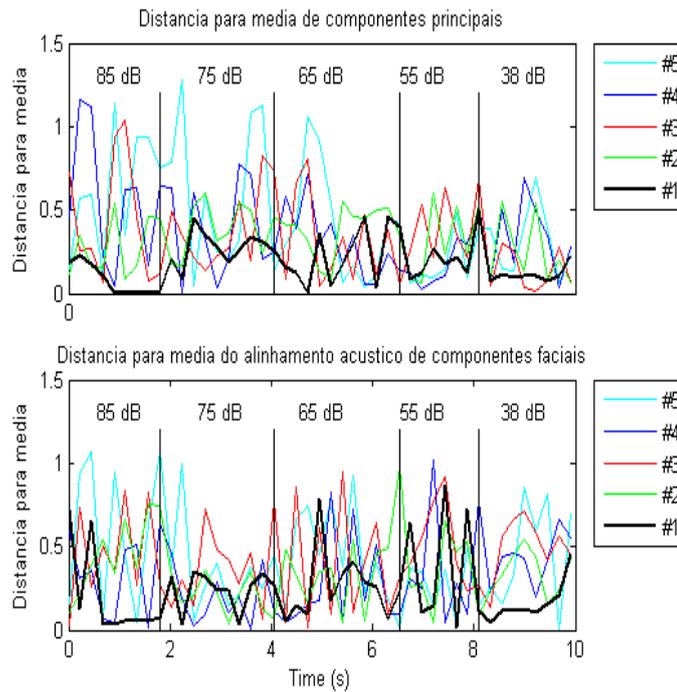


Figura 29 – Distância euclidiana de componentes principais e do alinhamento acústico

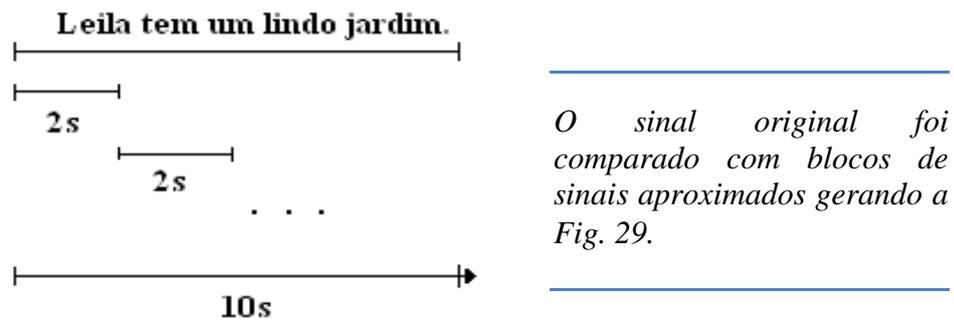


Figura 30 – Esquema da divisão dos blocos para distância Euclidiana

Analisando o autovetor que representa a componente principal (PCA) ao longo do tempo, temos os seguintes gráficos:

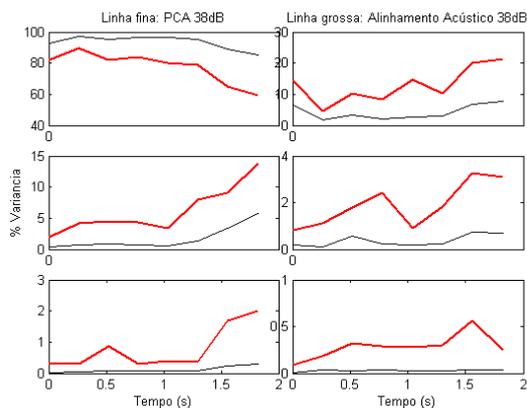


Figura 31 – PCA 38 dB

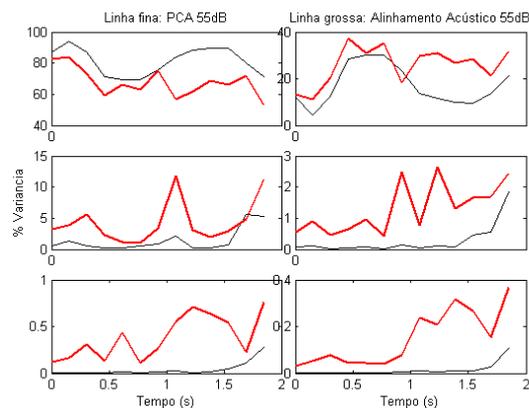


Figura 32 – PCA 55 dB

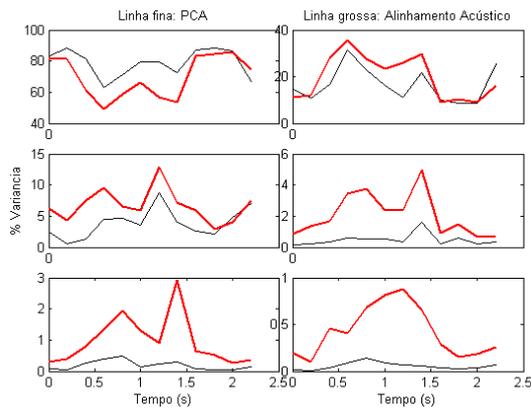


Figura 33 – PCA 65 dB

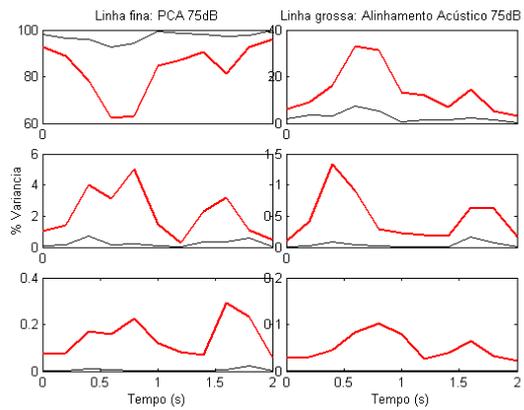


Figura 34 – PCA 75 dB

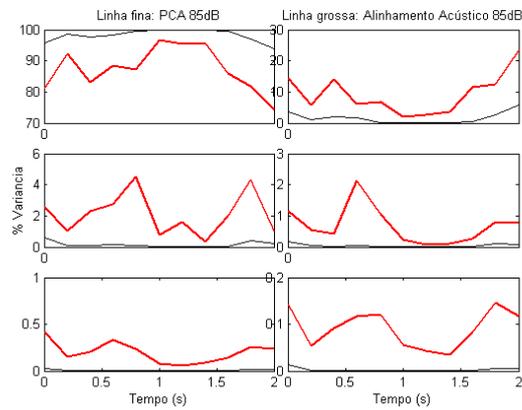


Figura 35 – PCA 85 dB

A PCA foi utilizada para se extrair as características mais importantes da imagem. A linha em vermelho representa o alinhamento acústico e a linha em preto representa a componente principal do movimento facial. Nos gráficos está representado o quanto cada autovetor representa, no movimento facial e no alinhamento acústico.

3 Resultados e Discussões

Com relação às figuras mostradas até o presente momento podemos fazer algumas inferências.

A frequência fundamental das figuras 8, 9, 10, 11 e 12 mostra que quanto mais o ruído aumenta há também um aumento da distorção da mesma, ou seja, uma maior vibração é percebida.

Na Fig. 8 de 38 dB a frequência fundamental em Hz não ultrapassa de 300 Hz sendo que o ruído é quase imperceptível.

A Fig. 9 de 55 dB a frequência fundamental em Hz chega até 350 Hz sendo que o ruído apesar de ser perceptível ainda é baixo.

Na Fig. 10 de 65 dB a frequência fundamental em Hz chega até 400 Hz sendo mais constante no pico de 300 Hz. O ruído apesar de ser perceptível ainda é baixo.

Na Fig. 11 de 75 dB a frequência fundamental em Hz chega até 350 Hz sendo que o ruído é alto e interfere na locução da frase. Pode-se notar que a vibração das pregas vocais são mais variáveis em relação a frequência fundamental.

Na Fig. 12 de 85 dB a frequência fundamental em Hz chega até 400 Hz sendo que a interferência do ruído é alta chega a alterar intermitentemente a frequência fundamental da voz da locutora.

Com base nas frequências em Hz de parâmetros LSP das figuras 13, 14, 15, 16 e 17 não se percebe muitas alterações com o Efeito Lombard. Isso pode ser justificado por serem esses coeficientes fortemente ligados aos formantes (frequência de ressonância do trato vocal), os quais são determinados pela geometria do trato vocal.

Para a energia em RMS mostradas nas figuras 18, 19, 20, 21 e 22 temos uma confirmação do Efeito Lombard, ou seja, à medida que o ruído foi elevado, a locutora instintivamente eleva seu tom de voz consideravelmente. Observe que na Fig. 22 temos

um pico acima de 2.5 por volta de 5.5 vezes a mais que o pico de 0.45 em 38 dB mostrado na Fig. 18.

Ao analisar a Fig. 29 que mostra a distância Euclidiana verifica-se a estabilidade foi alterada quando se utilizou as elocuições com ruído.

Para a Variância Acumulada observe a Tabela 1.

Tabela 1 – Tabela de Variância Acumulada dos autovetores

Autovetores	Variância Acumulada %				
	38dB	55dB	65dB	75dB	85dB
1	0.92	0.96	0.99	0.99	0.94
2	0.98	0.98	0.99	0.99	0.97
3	0.98	0.98	0.99	0.99	0.98
4	0.99	0.99	0.99	0.99	0.99
5	0.99	0.99	0.99	0.99	0.99
6	0.99	0.99	1	1	0.99
7	1	0.99	1	1	0.99
8	1	0.99	1	1	0.99
9	1	1	1	1	1
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1

Observe que pela Tabela 1 observa-se uma variância muito pequena, indicando a conservação do sinal independente do ruído. Aumentando-se o ruído leva-se a crer que essa variância aumenta. Considerando as figuras 23, 24, 25, 26 e 27 as primeiras seis componentes principais representam cerca de mais de 95% das características principais do movimento facial.

A Fig. 29 que mostra a distância Euclidiana de cinco autovetores comparados aos sinais originais da sentença, vem destacando o primeiro autovetor. Tal motivo é devido porque como se pode observar que o primeiro autovetor é o que menos se distancia quando comparado aos outros autovetores independente do sinal-ruído. Observa-se também que para os ruídos de 38 dB e 85 dB o mesmo autovetor se mantém mais estável que os outros. Isso infere que o referido assume uma representação maior do movimento.

Uma comparação por meio da correlação foi feita entre o movimento captado pelos sensores e o movimento estimado através das principais componentes principais. Este resultado pode ser visto nas Tabelas 2, 3, 4 e 5:

Tabela 2 – Correlação cruzada do movimento para 6 autovetores

Correlação Cruzada Movimento										
6 Autovetores										
	X_{aprox}					X_{real}				
	38dB	55dB	65dB	75dB	85dB	38dB	55dB	65dB	75dB	85dB
38dB	0,9973	0,9583	0,9584	0,9583	0,9583	0,9885	0,9583	0,9583	0,9583	0,9583
55dB	0,9583	0,9972	0,9583	0,9584	0,9582	0,9583	0,9886	0,9583	0,9583	0,9583
65dB	0,9584	0,9583	0,9970	0,9583	0,9583	0,9583	0,9583	0,9883	0,9583	0,9583
75dB	0,9583	0,9584	0,9583	0,9971	0,9583	0,9583	0,9583	0,9583	0,9883	0,9583
85dB	0,9583	0,9582	0,9583	0,9583	0,9970	0,9583	0,9583	0,9583	0,9583	0,9883

Tabela 3 – Correlação cruzada do movimento para 2 autovetores

Correlação Cruzada Movimento										
2 Autovetores										
	X_{aprox}					X_{real}				
	38dB	55dB	65dB	75dB	85dB	38dB	55dB	65dB	75dB	85dB
38dB	0,9883	0,9580	0,9581	0,9583	0,9584	0,9884	0,9583	0,9583	0,9583	0,9583
55dB	0,9580	0,9884	0,9583	0,9584	0,9582	0,9583	0,9883	0,9583	0,9583	0,9583
65dB	0,9581	0,9583	0,9884	0,9583	0,9583	0,9583	0,9583	0,9884	0,9583	0,9583
75dB	0,9583	0,9584	0,9583	0,9883	0,9582	0,9583	0,9583	0,9583	0,9884	0,9583
85dB	0,9584	0,9582	0,9583	0,9582	0,9884	0,9583	0,9583	0,9583	0,9583	0,9884

Tabela 4 – Correlação cruzada da acústica para 6 autovetores

Correlação Cruzada Acústica										
6 Autovetores										
	X_{aprox}					X_{real}				
	38dB	55dB	65dB	75dB	85dB	38dB	55dB	65dB	75dB	85dB
38dB	0,9880	0,9583	0,9583	0,9583	0,9580	0,9882	0,9583	0,9582	0,9583	0,9583
55dB	0,9583	0,9882	0,9583	0,9583	0,9582	0,9583	0,9883	0,9583	0,9580	0,9581
65dB	0,9583	0,9583	0,9882	0,9583	0,9583	0,9582	0,9583	0,9882	0,9583	0,9585
75dB	0,9583	0,9583	0,9583	0,9883	0,9583	0,9583	0,9580	0,9583	0,9882	0,9583
85dB	0,9580	0,9582	0,9583	0,9583	0,9880	0,9583	0,9581	0,9585	0,9583	0,9883

Tabela 5 – Correlação cruzada da acústica para 2 autovetores

Correlação Cruzada Acústica										
2 Autovetores										
	X_{aprox}					X_{real}				
	38dB	55dB	65dB	75dB	85dB	38dB	55dB	65dB	75dB	85dB
38dB	0,9884	0,9581	0,9582	0,9583	0,9581	0,9882	0,9583	0,9582	0,9583	0,9580
55dB	0,9581	0,9880	0,9583	0,9583	0,9582	0,9583	0,9882	0,9583	0,9580	0,9584
65dB	0,9582	0,9583	0,9880	0,9583	0,9584	0,9582	0,9583	0,9883	0,9583	0,9583
75dB	0,9583	0,9583	0,9583	0,9880	0,9583	0,9583	0,9580	0,9583	0,9883	0,9583
85dB	0,9581	0,9582	0,9584	0,9583	0,9880	0,9580	0,9584	0,9583	0,9583	0,9880

Tabela 6 – Correlação Cruzada Movimento x Acústica

Correlação Cruzada Movimento x Acústica										
6 Autovetores										
	X_{aprox}					X_{real}				
	38dB	55dB	65dB	75dB	85dB	38dB	55dB	65dB	75dB	85dB
38dB	0,7790	0,7300	0,7400	0,7300	0,7200	0,7580	0,7200	0,7106	0,6999	0,6897
55dB	0,7300	0,7542	0,7230	0,7400	0,7502	0,7209	0,7760	0,7002	0,6999	0,7575
65dB	0,7400	0,7230	0,7997	0,7502	0,7400	0,7106	0,7002	0,7830	0,7575	0,7575
75dB	0,7300	0,7400	0,7500	0,7971	0,7501	0,6999	0,6999	0,7575	0,7731	0,7490
85dB	0,7200	0,7502	0,7400	0,7501	0,7970	0,6897	0,7575	0,7575	0,7490	0,7723

A avaliação da matriz X_{aprox} com relação aos valores aproximados e com relação à matriz real X_{real} do movimento para seis autovetores visto na Tabela 2 mostra que a correlação é alta independente da relação sinal-ruído. E mesmo para a

Tabela 3 com apenas dois autovetores essa correlação se mantém alta não sendo diferenciada pelo Efeito Lombard.

Para a acústica as mesmas observações podem ser percebidas tanto para a avaliação de seis autovetores na Tabela 4 como para dois autovetores na Tabela 5 levando a crer que o Efeito Lombard não interfere no alinhamento acústico.

Na Tabela 6 temos a confirmação de estudos anteriores que diz que a acústica tem forte ligação com o movimento facial.

4 *Conclusões*

Apesar de haver necessidade de mais testes utilizando sentenças maiores ou mesmo considerando um locutor masculino para explicar as sentenças, temos um fechamento diante do estudo feito.

O movimento facial percebido pelos ouvintes acontece em baixa resolução. Apenas com dois autovetores foi possível conseguir uma correlação de no mínimo 96 % no movimento dos sensores, não importando qual o nível de ruído. Fica clara a correlação do movimento facial e da acústica do sinal comprovando estudos feitos já anteriormente que também denotam tal afirmação.

A influência do Efeito Lombard não caracteriza grande ou nenhuma alteração no sinal analisado, apesar de ter havido uma vibração observada na frequência fundamental à medida que o ruído se eleva. Mas as características principais do sinal são preservadas, não sendo influenciadas pelo fenômeno.

A acústica da fala e o movimento facial, se apresentam bem correlacionados e carregam as informações principais do sinal em todas as avaliações feitas.

Referências Bibliográficas

Adriano Vilela Barbosa e Hani Camille Yehia. Measuring the relation between speech acoustics and 2d facial motion. 26th International Conference on Acoustics, Speech, and Signal Processing - ICASSP'2001, 1:181{184, 2001.

Luis Antonio Aguirre. Introdução à identificação de sistemas: técnicas lineares e não-lineares aplicadas a sistemas reais. 3. Ed. Ver e ampl. Belo Horizonte: Editora UFMG, 2007.

Bruce A. Draper, Kyungim Baek, e Marian Stewart Bartlett. Recognizing faces with pca and ica. Computer Vision and Image Understanding, 91:115{137, 2003.

Eric Vatikiotis-Bateson, Hani Yehia, e Takaaki Kuratate. Speaking mode variability in multimodal speech production. IEEE Transaction on Neural Networks, 13(4):894{899, July 2002.

Eric Vatikiotis-Bateson e Hani Camille Yehia. Estimation and generalization of multimodal speech production. IEEE Signal Processing Society Workshop, 1:23{32, December 2000.

Eric Vatikiotis-Bateson, Takaaki Kuratate, e Hani Camille Yehia. Kinematics-based synthesis of realistic talking faces. International Conference on Auditory-Visual Speech Processing (AVSP'98), pages 185{190, 1998b.

Eric Vatikiotis-Bateson e Hani Camille Yehia. Uni_ed physiological model of audible-visible speech production. V EUROSPEECH, pages 22{25, 1997.

Hani Camille Yehia, Philip Rubin, e Eric Vatikiotis-Bateson. Quantitative association of vocal-tract and facial behavior. Speech Communication, 26:23{43, October 1998.

INSTITUTO DE EDUCAÇÃO AVANÇADO DA AUDIÇÃO – IEAA.
<<http://www.ieaa.com.br/clinica/index.php?page=articles&op=readArticle&id=8&title=Audicao-e-Ruido-em-escolares>>. Acesso em março de 2010.

Ketia Soares Moreira. Um estudo sobre as relações de padrões do movimento facial com a acústica da fala e com a identidade do locutor. Belo Horizonte. Maio de 2008.

Koji Iwano, Satoshi Tamura, e Sadaoki Furui. Bimodal speech recognition using lip movement measured by optical-ow analysis. International Workshop on Hands-Free Speech Communication (HSC 2001), pages 187{190, 2001.

Li-Fen Chen, Hong-Yuan Mark Liao, e Ja-Chen Lin. Person identi_cation using facial motion. IEEE - International Conference on Image Processing, 02:677{680, 2001.

Luis Felipe Dias Lopes, Robert Wayne Samohyl. Análise em components principais: técnica alternativa de análise de confiabilidade em sistemas complexos multivariados. Junho de 2003.

Marcus V. Lamar, Md. Shoaib Bhuiyan, e Akira Iwata. Hand alphabet recognition using morphological pca and neural networks. Proc. of International Joint Conference on Neural Networks, 4:2839{2844, July 1999a.

Nicolau Werneck; Lucas Malta; Leonardo Carneiro; Hani Yehia. Um sistema de reprodução multimodal da fala. <<http://cefala.org/~nwerneck/projeto/cartaz/cartaz.pdf>>. Acesso em agosto de 2010.

Ognjen Arandjelovi e Roberto Cipolla. Face recognition from face motion manifolds using robust kernel resistor-average distance. Conference on Computer Vision and Pattern Recognition Workshop, 5, 2004.

Rodrigo Leite Durães. Validação de modelos baseados em RNA utilizando análise estatística de dados e lógica fuzzy. Outubro de 2009.

Robert M. Gray. IEEE signal processing society distinguished lecturer program. <<http://www-ee.stanford.edu/~gray/dl.html>>. Acesso em novembro 2010.

Roberto Brunelli e Tomaso Poggio. Face recognition: features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 1042{1052, October 1991.

Sirlei Aparecida Mariano. Análise acústica da voz. <
http://www.acesa.com/viver/arquivo/vida_saudavel/2005/01/14-cal/>. Acesso em
novembro de 2010.

Valdeci A. Moraes. Estatística básica. Junho 2007.

Vladimir F. S. de Alencar; Abraham Alcaim. Atributos eficientes em reconhecimentos
de voz distribuído.
<<http://www.deetc.isel.ipl.pt/comunicacoesep/disciplinas/pes/pca.pdf>>. Rio de Janeiro,
2007. Acesso em novembro de 2010.