JULIANO SANTIAGO DE MATTOS

UM ESTUDO COMPARATIVO ENTRE O SINAL ELETROGLOTOGRÁFICO E O SINAL DE VOZ

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Telecomunicações da Escola de Engenharia da Universidade Federal Fluminense como parte dos requisitos para obtenção do grau de Mestre em Ciências.

Professores Orientadores:

Edson Luiz Cataldo Ferreira, D. Sc. (GMA/UFF)
José Antonio Apolinário Junior, D. Sc. (SE/3/IME)

Niterói
2008

Resumo

Esta dissertação realiza um estudo comparativo entre o sinal obtido de um eletroglotógrafo, o sinal eletroglotográfico (EGG), e uma estimativa do sinal glotal (OFG) obtida pela filtragem inversa do sinal de voz.

Visando efetuar esta comparação, sinais de voz com seus sinais EGG associados foram gravados de um número de locutores do sexo masculino e feminino. Os sinais adquiridos na língua portuguesa falada no Rio de Janeiro formaram um *corpus* EGG/speech que será disponibilizado para os pesquisadores trabalhando nesta área.

Da razoavelmente grande massa de dados, incluindo os sinais EGG e OFG assim como suas primeira derivadas, algumas de suas características foram extraídas e comparadas estatisticamente em termos de suas médias e dispersão para a população considerada.

A comparação foi feita por meio de vogais sustentadas e concatenadas. Também, visando uma aplicação em perícias fonéticas, uma primeira discussão na capacidade discriminadora das características de ambos sinais EGG e OFG é abordada.

A análise realizada aqui, apesar de longe de ideal, objetivou contribuir com o início do estudo deste tópico na língua portuguesa falada no Brasil e o *corpus*, com seus resultados preliminares, poderia, devido ao seu caráter inovador, dar suporte a trabalhos futuros na tarefa de discriminar sinais de voz de maneira mais eficiente.

Palavras-chave: sinal glotal, sinal EGG, perícia forense.

Abstract

This dissertation performs a comparative study between the signal obtained from a eletroglotograph, the eletroglotographic signal or EGG, and an estimate of the glotal signal, OFG, obtained by inverse filtering of the speech signal.

In order to carry out this comparison, speech signals with their associated EGG signals were recorded from a number of male and female speakers. The signals acquired in Portuguese language spoken in Rio de Janeiro formed an EGG/speech *corpus* that will be available for those researchers working in this field.

From the reasonably large amount of data, comprising EGG and OFG signals as well as their first derivatives, some of their features were extracted and statistically compared in terms of their means and dispersion within the population taken into account.

The comparison was carried out by means of sustained and concatenated vowels. Also, aiming an application in forensic phonetics, a first discussion on the discriminative capability of both EGG and OFG features in speaker recognition is addressed.

The analysis carried out herein, although far from ideal, intended to contribute by starting the study of this topic in the Brazilian Portuguese language and the *corpus*, with its preliminary results, could, due to its innovative characteristic, support forthcoming works in the task of discriminating speech signals in a more efficient manner.

Key words: glotal signal, EGG signal, forensic phonetics.

Declaração de Originalidade

Esta dissertação foi produzida por mim e relaciona trabalho original de minha própria execução. A menos que de outra forma mencionado, os gráficos e tabelas exibidos foram produzidos a partir de dados obtidos durante a pesquisa. Sempre que materiais, idéias, ou algoritmos computacionais de outros pesquisadores tiveram sido usados ou adaptados, a fonte de informação foi claramente especificada. Esta dissertação não foi submetida para graduação ou qualificação profissional em nenhum outro lugar.

Juliano Santiago de Mattos

Agradecimentos

À minha família, que sempre me apoiou em todos os desafios que resolvi enfrentar, inclusive o início do mestrado.

Ao professor Edson Cataldo pela orientação, pela confiança demonstrada desde o início, inclusive durante o processo seletivo para ingresso no Mestrado.

Ao professor José Apolinário que me abriu as portas do IME, pela orientação e extrema disponibilidade.

Ao professor Dirceu Gonzaga pelo incondicional apoio, amizade e, principalmente, por ter abdicado do convívio familiar pra me ajudar a concluir este trabalho.

Ao curso de Pós-Graduação em Engenharia de Telecomunicações da Universidade Federal Fluminense que me concedeu esta grande oportunidade de aumentar meus conhecimentos.

Ao Instituto Militar de Engenharia (IME) pelo apoio em infraestrutura e equipamentos disponíveis no Laboratório de Voz.

Aos peritos do Instituto de Criminalística Carlos Éboli (ICCE-RJ) pelas informações disponibilizadas, que foram imprescindíveis para a conclusão deste trabalho.

À Jussara, funcionária da UFF, pelo incentivo em iniciar o mestrado nesta Universidade e pela sua extrema boa vontade.

Ao Universo por ter conspirado a meu favor, mais uma vez.

Dedicatória

Dedico este trabalho a:

Marcia, Zeno, Marcelo e Julia,

Patrícia, Jorge Henrique, Fabio,

Waldenir, José Malta,

Jeanne.

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

1.1 Introdução

O objetivo de um sistema de reconhecimento/verificação de locutor é identificar um locutor a partir da sua voz, tendo diversas aplicacões como, por exemplo, em segurança pública. Na identificação de criminosos a partir de gravações telefônicas, o processo de reconhecimento automático da identidade vocal se baseia na extração de parâmetros da voz, de um dado locutor, de forma a definir as suas características vocais que o tornem único.

Na atividade pericial, a verificação de locutor tem como principal objetivo atestar se determinada voz é de um locutor específico ou não, através da comparação entre falas distintas armazenadas em mídias de gravação.

A função do perito é avaliar a qualidade da mídia de gravação, extrair características das informações contidas na mídia, comparar padrões baseados em características extraídas dos sinais de voz, elaborar laudo técnico, entre outras, com o intuito de incriminar ou não um suspeito. Portanto, as técnicas de reconhecimento e verificação de locutor no contexto forense podem contribuir, significativamente,

com as investigações e julgamentos realizados pelas autoridades judiciárias.

Diferentes métodos podem ser aplicados, isoladamente ou em conjunto [1], para determinar (sugerir) se as vozes desconhecidas pertencem a um suspeito. Normalmente, os resultados obtidos são probabilidades que fornecem às autoridades uma indicação da força da evidência.

Apesar de todo o desenvolvimento, alcançado nos últimos anos, nos sistemas de reconhecimento/verificação de locutor, as típicas condições do contexto forense tais como diferenças nos equipamentos de gravação e canais de transmissão, a presença de ruído de fundo, entre outros, continuam sendo desafios a serem vencidos. Conseqüentemente, o impacto das tecnologias de reconhecimento automático de locutor, no contexto forense, ainda é modesto e extremamente dependente de uma grande variedade de procedimentos subjetivos [2]. Como conseqüência direta deste modesto impacto, nos últimos anos, os diferentes tipos de evidências forenses têm sido alvo de duras críticas que questionam seu *status* científico[3] [4] [5].

Atualmente, o que existe de concreto são cálculos de distâncias ou probabilidades que, quantitativamente, enumeram diferenças ou similaridades de certos parâmetros ou conjuntos de parâmetros. Ainda não há consenso científico na escolha dos parâmetros que permitam o cálculo das distâncias, ou na escolha da distribuição de probabilidade adequada para estabelecer um modelo estatístico [6].

A eletroglotografia é um método não invasivo descoberto por [7], que estima a variação da área de contato entre as cordas vocais (*vocal fold contact area - VFCA*) durante a produção da voz [8] [9] [10]. O eletroglotógrafo mede as variações da impedância elétrica causadas pela variação da área de contato entre as cordas vocais, sendo medida através de um par de eletrodos, presos ao pescoço do locutor.

Este eletrodos aplicam uma pequena corrente elétrica, incapaz de gerar desconforto, ao local. As aplicações práticas dos sinais EGG e DEGG (derivada do sinal EGG) que podem ser destacadas são: auxílio à detecção de patologias nas cordas vocais, modelagem do sinal de voz através dos parâmetros extraídos do sinal EGG, algoritmos para fins acadêmicos, determinação de parâmetros para reconhecimento de locutor.

A diferença de pressão entre o ar nos pulmões e o ar próximo à boca, causada pela expansão-contração dos pulmões, provoca um escoamento de ar que passa através das cordas vocais. Esta passagem de ar faz as cordas vocais vibrarem, tornando o fluxo de ar em um trem de pulso ou sinal glotal. O sinal glotal possui propriedades importantes ligadas às características anatômicas e fisiológicas da laringe e pode ser obtido experimentalmente, por filtragem inversa.

A despeito da forma do sinal glotal e do sinal eletroglotógrafico ser parecida, ainda não foi demonstrada uma relação física entre ambos, ou seja, não há uma técnica desenvolvida para se obter um através do outro. Este trabalho estuda, entre outros, a possibilidade de se obter parâmetros em comum entre os sinais, extraindo e comparando características de vogais sustentadas e de falas concatenadas.

1.2 Objetivos da dissertação

- Descrição da relação entre o sinal eletroglotográfico e o sinal de voz;
- Descrição da técnica de filtragem inversa PSIAIF e sua aplicação na obtenção do sinal glotal, a partir do sinal de voz;
- Obtenção e extração de parâmetros dos sinais glotal e eletroglotográfico de

vogais sustentadas e concatenadas, que possam auxiliar os peritos na identificação de locutores;

- Formação de um *corpus* EGG em português.

1.3 Estado da arte

Nos dias de hoje, ainda não é possível confirmar cientificamente que duas vozes gravadas, sob as melhores condições técnicas possíveis, são, sem sombra de dúvidas, oriundas do mesmo locutor [6].

Os reconhecimentos de voz e locutor têm se desenvolvido amplamente [11]. Porém, os resultados de alta precisão somente são obtidos com condições controladas, às quais, no contexto forense, estão longe de ocorrer [12].

O ponto vital dessa discussão está centrado no fato de que a corte, o júri ou os juízes, não se sentem confortáveis e, muitas vezes, incapazes de decidir por uma condenação, se não houver a certeza sobre a conduta ilícita do suspeito. Diante desse dilema, fica evidente que ainda há muito o que ser pesquisado sobre reconhecimento de locutor e que a perícia forense é um campo importante que necessita se desenvolver, principalmente, em função da demanda dos tribunais e de sua importância social.

Através da SENASP e do Departamento de Polícia Federal, hoje estão sendo treinados em Brasília peritos de todo o Brasil, como parte de um acordo de cooperação técnica que prevê, além de um curso de 580 horas, a construção de laboratórios completos com computadores e equipamentos para realização de exames nessa área. O Rio de Janeiro receberá dois conjuntos de equipamentos devido à enorme demanda desse tipo de exame [13].

A popularização dos equipamentos de gravação digital, a regulamentação da lei das interceptações telefônicas e a utilização de sistemas de monitoramento de linhas telefônicas pelas polícias são os fatores que explicam a grande quantidade de material remetido para a perícia. Estima-se que hoje cerca de 20.000 escutas autorizadas estão sendo realizadas pelas polícias, em todo país e, certamente, grande parte desse material será enviado para a perícia [13].

1.4 Contribuições desta dissertação

As contribuições desta dissertação são: a formação de um *corpus* EGG em português, a implementação do método de filtragem inversa PSIAIF, a explicação detalhada do funcionamento do eletroglotógrafo, o cálculo da freqüência fundamental da voz usando a derivada do sinal eletroglotográfico (EGG), a comparação de parâmetros obtidos do sinal EGG e do sinal glotal e a obtenção de parâmetros, que possuam bom poder de discriminação de locutores, do sinal eletroglotográfico.

Capítulo 2

Perícia forense

2.1 A importância da perícia forense

A Fonética forense, como o próprio nome sugere, pode ser definida como o estudo dos sons e das articulações próprias de uma língua para fins jurídicos e cíveis.

Com o propósito de condenar ou absolver um suspeito, os juízes de direito, freqüentemente, necessitam recorrer à memória de testemunhas para identificar o rosto ou a voz do acusado. A ausência de parâmetros concretos para a identificação, ocasionalmente, contribuem para a ocorrência de erros irreparáveis cometidos pela justiça.

Antigamente, a necessidade de determinar a identidade de parceiros comerciais, estimulou o desenvolvimento de uma técnica confiável, não invasiva e não traumática de identificação: a impressão digital.

Existem registros, encontrados na China, das primeiras marcas usadas em documentos como forma de autenticação dos originais.

Em 1686, o anatomista italiano Marcello Malpighi observou a diversidade de impressões digitais humanas. Entretanto, somente em 1823, J.E. Purkinje publicou

a primeira classificação das impressões digitais, organizadas em 19 tipos diferentes, mais conhecidas como *Os desenhos de Purkinje*, sem nenhuma aparente menção a aplicações de identificação [6]. Na prática, nem sempre as impressões digitais estão disponíveis. Há casos, cada vez mais freqüentes na mídia, que a única fonte de informação são vozes gravadas durante conversas telefônicas, como por exemplo, nos casos de crime de suborno, extorsão, seqüestro e chantagem. Diante deste fato, é possível afirmar que há uma forte demanda, perfeitamente justificável, existente por parte da comunidade policial e de magistrados no sentido de estabelecer formas legais e precisas de identificação de pessoas através da voz [14].

A partir desta premissa, questionamentos importantes imediatamente surgem: Qual a extensão do uso de gravações de vozes humanas como base de inquéritos policiais? Qual a extensão do uso de gravações de vozes para estabelecer a culpa ou inocência de um suspeito?

O considerável interesse em obter técnicas confiáveis para o reconhecimento de locutor, e utilizá-las com prova, é facilmente compreendido.

2.2 Histórico da perícia

O primeiro registro de uso forense da identificação de alguém pela voz foi em 1660 em um tribunal inglês. Ainda na Inglaterra, entre 1754 e 1780, o magistrado Sir John Fielding que era cego desde os 19 anos de idade, enquanto chefiava a primeira polícia profissional inglesa, chamada *Bow Street Runners*, identificou pelas vozes centenas de criminosos. Nos EUA, em 1861, um tribunal considerou possível a identificação de um cão pelo seu latido afirmando "...se uma pessoa pode ser identificada através de sua voz, um cachorro também pode ser através de seu latido". Porém,

o caso mais emblemático foi em 1935 quando o herói nacional americano Charles Lindberg, o primeiro homem que sobrevoou sozinho o Oceano Atlântico, teve seu filho seqüestrado e morto. O acusado, Bruno Hauptmann, teve sua voz reconhecida em juízo por Lindberg, que conversara com o seqüestrador pelo telefone [13].

A época do empirismo nessa área começou a findar com os experimentos de Alexander Melville Bell, pai de Alexander Graham Bell que criou o telefone, transformando o som em impulso elétrico, que em 1867 criou uma representação gráfica das palavras da forma como eram pronunciadas chamada *visible speech*. Os Laboratórios Bell, de New Jersey, foram referências nos estudos de identificação de locutores. Em 1941, seus engenheiros Ralph Potter, Kopp e Green produziram o primeiro Espectrógrafo Analógico de Som, que através de cálculos de transformadas de Fourier de amostras de voz, permitia a "visualização"do som. Esse equipamento veio substituir os espectrógrafos mecânicos de Henrici, concebidos dez anos antes [13]. Mas foi com a Segunda Guerra mundial que essa nova tecnologia ganhou importância estratégica. Com o objetivo de monitorar o deslocamento das tropas do eixo, o governo norte americano solicitou aos Laboratórios Bell um projeto que permitisse aos militares a identificação das vozes dos operadores de rádios alemães, para assim, acompanharem a movimentação das tropas. Em 1944, os doutores Gray e Koop entusiasmados com o projeto criaram o termo *voiceprint*, para equiparar a análise espectrográfica da voz à *fingerprint* (impressão digital), acreditando que a mesma objetividade da identificação dactilar pudesse ser aplicada à voz.

Na década de sessenta, no auge da guerra fria, a polícia dos Estados Unidos recebia inúmeras chamadas telefônicas com ameaças de bombas em companhias aéreas. A gravação em fitas magnéticas já era uma tecnologia amplamente difun-

dida, que permitia o armazenamento desses registros. Foi, então, que a polícia de Nova Iorque solicitou a ajuda dos Laboratórios Bell com o objetivo de identificar os indivíduos que faziam tais ameaças. O laboratório indicou o físico Lawrence G. Kersta que, em dois anos, desenvolveu uma metodologia, baseada na comparação espectrográfica das amostras que, segundo ele, era capaz de identificar uma voz com grau de 99,65% de certeza. Kersta e seu método foram tão bem aceitos que, em pouco tempo, ele abandonou os Laboratórios Bell e fundou sua própria empresa a *Voiceprint Laboratories, Inc.*, que oferecia vários serviços nessa área, trabalhando para a Agencia Federal de Aviação Americana e a Força Aérea, entre outras.

O físico foi o primeiro especialista em voz a depor em juízo, mas cometeu graves erros em sua metodologia, principalmente quando acreditou que a voz possuísse as mesmas características imutáveis das impressões digitais e superestimou a análise espectrográfica em detrimento da análise perceptual. Em 1968, em um processo judicial de grande repercussão onde Kersta atuava como perito, o ilustre foneticista Dr. Peter Ladefoged da Universidade da Califórnia, atacou veementemente a metodologia proposta pelo físico. Esse foi o início da decadência de seu método e de uma dissidência, que perdura até hoje, entre engenheiros e foneticistas que atuam nessa área [13].

Esse caso na esfera judicial levou as cortes americanas a reverem a admissibilidade das metodologias utilizadas na identificação de locutores. Em 1967, um juiz alegou a necessidade da aplicação do *Frye test* que diz: "...quando um novo princípio ou descoberta científica é utilizado nos tribunais para demonstrar alguma evidência, este deve contar com a aceitação geral da comunidade científica de seu entorno". Esse impasse produziu um importante avanço nessa área. Em 1968, o Departamento

de Justiça dos Estados Unidos convocou o Departamento de Ciências da Fala e Audiologia da Universidade de Michigan para elaborar um estudo sobre o assunto, o responsável, Oscar Tosi, Doutor em Física, realizou um grande trabalho em três anos o qual envolveu a análise espectrográfica de 34.992 casos. A conclusão do trabalho apontou para o aprimoramento da metodologia de Kersta, criando um modelo auditivo-espectrográfico que foi tão positivo que a polícia do Estado de Michigan decidiu criar a primeira unidade policial de identificação de voz sob o comando do Tenente Ernest Nash que havia trabalhado com Tosi. Em 1971, Tosi, Kersta e Nash fundam a I.A.V.I (*International Association of Voice Identification*).

Em 1986, o FBI publicou o resultado de um estudo que durou quinze anos, onde foram realizadas 2000 comparações de casos reais, que apontou para uma margem de erro inferior a 1%. Existem ainda, registros de estudos nessa área na antiga URSS, no Laboratório de Fonoscopia do Centro de Criminalística do Ministério do Interior; no Japão, em 1963, em um caso de seqüestro onde primeiro se utilizou esse exame. No início da década de sessenta, a Polícia Federal Alemã começou a pesquisar métodos automáticos de identificação da voz conhecido como AUROS (*Automatic Recognition of Speakers*). A Europa segue hoje uma linha metodológica que combina análise espectrográfica, fonético-lingüística e biométrica.

No Brasil, métodos de identificação da voz tiveram início, na perícia oficial, na década de 90, com a iniciativa isolada de alguns peritos dos Estados, da Polícia Federal e do Distrito Federal. O marco inicial foi o ano de 1994 quando ocorreu o primeiro Seminário de Fonética Forense patrocinado pela Associação Brasileira de Criminalística [13].

2.3 O conceito de identidade na criminalísitca

Em criminalística, o processo de identificação procura a individualização [15]. Identificar uma pessoa ou objeto significa poder distingui-los dos demais existentes. O processo de individualização forense pode ser considerado, inicialmente, como um processo de redução, com o intuito de minimizar a população até um único indivíduo ou objeto.

Quando estão envolvidas impressões digitais, marcas de calçados e ferramentas e armas de fogo, pode-se facilmente imaginar que o tamanho da população é gigantesco, tornando esse processo de redução indispensável para a elucidação do fato (um crime, por exemplo). A redução é alcançada a partir da obtenção de características específicas ou raras de cada rastro (i.e., impressões digitais na cena do crime) e o controle de material (i.e., impressões digitais colhidas do suspeito) [12].

2.4 Questões legais

Questões do direito civil, envolvendo disputas entre membros da sociedade, não serão consideradas neste trabalho e todos os procedimentos legais serão descritos de acordo com a legislação em vigor.

A legislação brasileira impõe uma série de limites, para que a repressão às práticas delituosas não sacrifique os direitos fundamentais a todos assegurados pela Constituição Federal. Se a lei determinar que uma pessoa é culpada por ter cometido certa transgressão, concomitantemente, lhe será garantido o direito de contraditório e ampla defesa, preservando seus direitos individuais, ou seja, o direito de se defender de toda e qualquer acusação durante o andamento do processo, até que seja

considerada culpada ou não por ter cometido um crime.

A Constituição brasileira, estabelece que toda pessoa acusada de um crime é presumidamente inocente até que se prove o contrário. Conseqüentemente, o acusado não será obrigado a provar sua inocência. A promotoria é quem será encarregada de provar a culpa do suspeito, dentro dos limites legais.

No Brasil, o início das investigações é realizado pela polícia judiciária, cujos poderes estão previamente definidos no texto infraconstitucional, mediante fatos que atribuam, ao mínimo, indícios de autoria e prova da materialidade do ato delituoso. Neste caso, são admitidas como evidências, impressões digitais, amostras de sangue e esperma, fios de cabelo ou interceptações telefônicas que deverão apresentar dois requisitos: ordem judicial para fins de investigação criminal ou instrução processual penal; nas hipóteses e na forma que a lei estabelecer (Inciso XII do art. 5.° da Constituição Federal).

Todas as evidências colhidas durante as investigações serão reunidas e apreciadas pela autoridade policial competente, que, de acordo com a coleta das provas, enviará o inquérito policial ao Ministério Público, que indiciará ou não o investigado. Diante da natureza técnica que a análise de algumas evidências requer, por exemplo as vozes gravadas durante ligações telefônicas, dois peritos, ou mais, deverão ser nomeados para confeccionar um laudo técnico sobre as evidências, mediante prazo pré-determinado. No exemplo citado, o laudo deverá comprovar a autenticidade do material, indícios de edição, com o intuito de modificar ou esconder o conteúdo da conversa, e até sugerir que a voz contida nas gravações pertence ou não à pessoa investigada.

Após a apresentação do laudo as partes (autor e réu) têm acesso ao trabalho do

perito, que poderá ser questionado por ambos. A credibilidade do perito, geralmente, não é posta em discussão, com excessão dos casos de erro grosseiro.

Em casos especiais, poderão ser convidados a participar da análise do laudo, especialistas em acústica, engenheiros e foneticistas.

Cabe ressaltar que a decisão final sempre caberá ao juiz, sendo a prova pericial técnica (laudo) um instrumento de auxílio à autoridade judiciária.

2.5 O termo *voiceprint*

Em 1962, um artigo publicado na revista *Nature* [68], intitulado *voiceprint identification* (uma alusão ao termo impressão digital, *fingerprint* em inglês), que pode ser traduzido como "Identificação pela Impressão da Voz", introduziu um termo que até hoje é utilizado em jornais, filmes policiais e de espionagem. Essa metáfora induziu várias pessoas a acreditarem que a representação gráfica da voz , assim como é feita no caso das impressões digitais, é suficiente para identificar uma pessoa (locutor). Atualmente, nenhum especialista em voz provou que a análise de espectrogramas é capaz de identificar locutores.

A gravação da voz não é um rastro deixado em uma superfície em contato com uma parte do corpo humano, nem é uma amostra direta: é, na verdade, uma gravação indireta de complexos movimentos articulatórios [6], conforme ilustra a Fig. 2.1-(a). É, pois, uma metáfora erroneamente associada às impressões digitais.

Os órgãos da fala induzem variações na pressão acústica instantânea, à qual pode ser recuperada por transdutores que convertem essas variações em tensão elétrica. Assim como os gestos humanos, a voz não pode ser reproduzida ao longo do tempo. Os parâmetros usados para descrever a voz mostram claramente sua

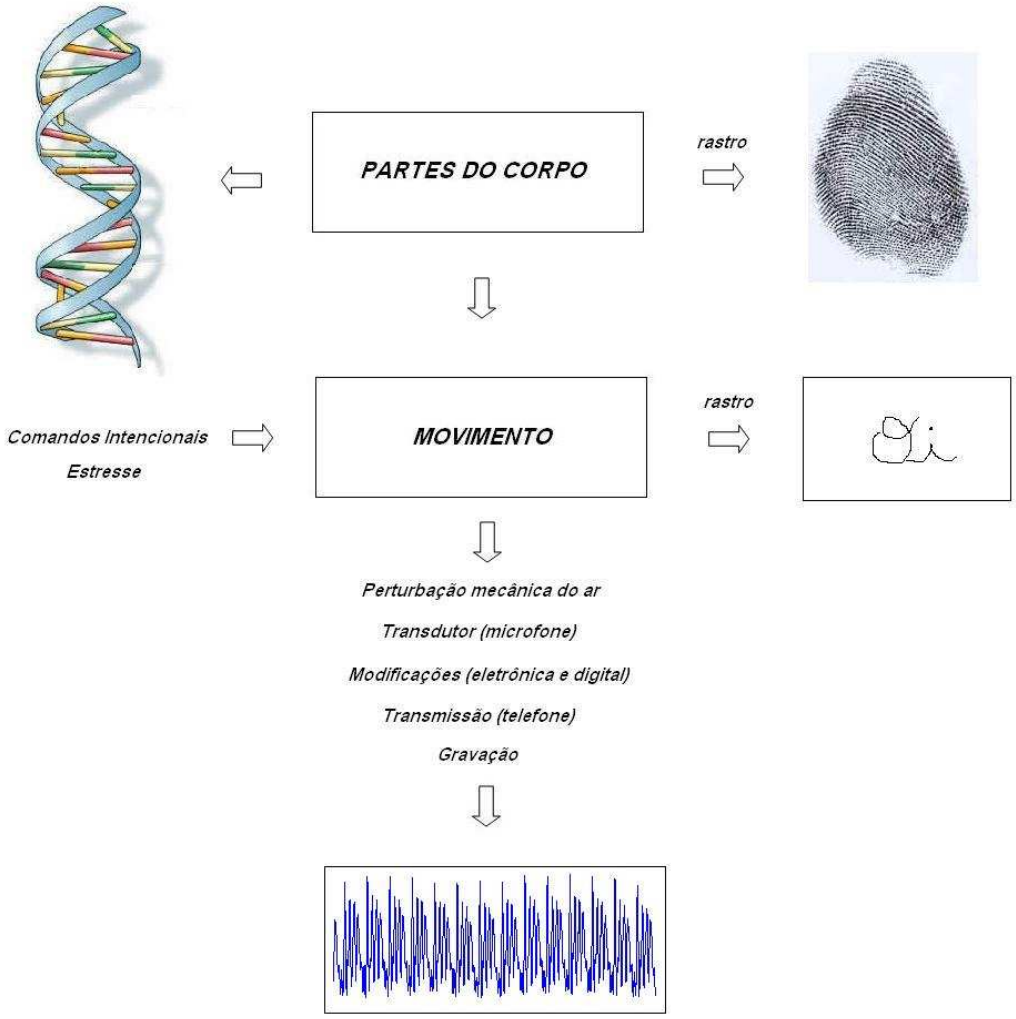

Figura 2.1: O termo *voiceprint* é uma metáfora erroneamente associada às impressões digitais.

dependência em relação à velocidade de articulação, volume da voz, o estado psicológico do locutor e estresse. O reconhecimento automático de locutor é diretamente confrontado com estas variações dependentes do locutor que estão intrinsicamente ligadas à produção da voz. Ademais, é extremamente importante considerar os parâmetros envolvidos na transmissão e gravação da voz e a possibilidade de outras vozes ou ruídos estarem presentes. No caso de gravações feitas em linhas telefônicas, as características do microfone, da linha telefônica, e do próprio pro-

cesso de gravação têm que ser analisadas, o que nem sempre é possível [6] .

Nos últimos anos, os diferentes tipos de evidências forenses têm sido alvo de duras críticas que questionam seu *status* científico [3] [4] [5].

Existem diversas formas de "individualizar a fonte", as quais incluem impressões digitais, voz, face, DNA, assinatura, marcas de ferramentas, tintas, vidros, fibras e armas de fogo. Geralmente, se concentram apenas em decisões de identificação ou exclusão - rejeição, mas se tornam problemáticas, pois estão relacionadas ao uso de limites subjetivos em técnicas que não fornecem uma identificação absoluta, apenas uma probabilidade. Então, o uso desses limites, em sua essência, qualificam o nível aceitável de dúvida adotado pelo perito [12]. Um outro problema encontrado, consiste no uso de escalas verbais de identificação de probabilidades tais como "muito provável", "provável"e "não conclusivo". Dessa forma, os mesmos erros são cometidos através do uso de limites subjetivos que ignoram as menores probabilidades relativas a cada caso, usurpando o direito do juiz ou júri de considerá-las [16].

Atualmente, a abordagem bayesiana é que está sendo mais aceita pela comunidade científica quando o assunto é perícia forense. Existem diversos grupos de trabalho em diferentes áreas (DNA, fibras, impressões digitais, armas de fogo, grafologia, marcas de ferramentas, tintas e vidros, voz e audio) no ENFSI (*Europe Network of Forensic Science Institute*) lidando com a abordagem bayesiana, com individualização da fonte e buscando estabelecer padrões e procedimentos [17].

Capítulo 3

Fundamentos de produção da voz

3.1 A produção do sinal de voz

A compreensão dos processos físicos se faz indispensável para a descrição da geração e propagação do som no sistema vocal e, conseqüentemente, para a determinação de um modelo apropriado para a representacão dos sons da voz.

A produção da voz se inicia com uma expansão-contração dos pulmões, que gera uma diferença entre a pressão do ar nos pulmões e a pressão do ar próximo a boca, causando um escoamento de ar. O ar proveniente dos pulmões é forçado através do pequeno espaço existente entre as cordas vocais, causando o movimento das cordas em uma freqüência determinada pela tensão dos músculos associados [18].

Este movimento causa a modificação do fluxo de ar, resultando em pulsos de ar (conhecidos como trem de pulsos ou sinal glotal) que serão amplificados e modificados pelas cavidades oral e nasal até serem irradiados pela boca. Os pulsos de ar são modulados pela língua, pelos dentes e lábios; isto é, pela geometria destes órgãos, de forma a produzir o que conhecemos por voz.

Quando o locutor deseja gerar um determinado som, ele exerce diversos tipos de controles sobre o aparelho fonador, produzindo a configuração articulatória e a excitação apropriadas, resultando nos vários tipos de sons da voz (sons sonoros, sons surdos, etc.) [19]. A Fig. 3.1-(a) mostra um esquema do aparelho fonador.

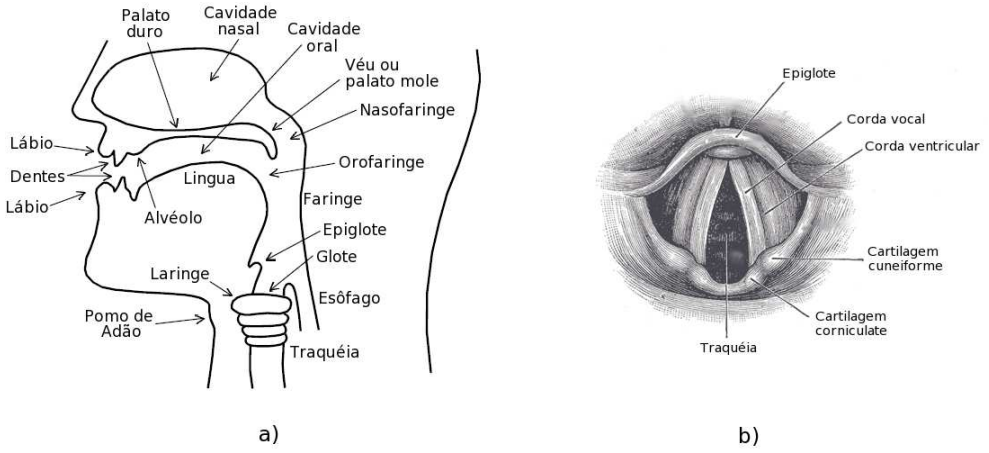

Figura 3.1: a) Aparelho Fonador b) Cordas Vocais. [70]

As cordas vocais, ilustradas na Fig. 3.2, principais elementos para a geração da voz, são duas membranas situadas na laringe (Fig. 3.1-(b)). Pela frente, as cordas vocais unem-se à cartilagem tiróide (o pomo de Adão) e, por trás, cada uma delas está presa a uma das cartilagens aritenóides, as quais podem se separar voluntariamente por meio de músculos.

Os dispositivos que fazem parte do aparelho fonador não são os únicos responsáveis pela produção da fala; existe uma interação entre o aparelho respiratório - pulmões, traquéia, laringe e pregas (ou cordas) vocais, faringe (parte comum aos aparelhos respiratório e digestivo, constituída por faringe oral e faringe nasal) e cavidade nasal, e a cavidade bucal, limitada pela mandíbula, pelos lábios, pelos dentes, pela língua, pelos palatos duro e mole (conhecidos popularmente como o "céu-da-boca") e pela faringe [20].

Fisiologicamente, estes tubos e cavidades compõem três subsistemas que atuam de modo sucessivo na produção da fala:

1. Respiratório: O subsistema respiratório é responsável pela passagem da corrente de ar dos pulmões pela traquéia e pela laringe;

2. Laringeal (ou laríngeo) : O subsistema laringeal (ou laríngeo), ou simplesmente laringe, situado na parte superior da traquéia, é o mais importante subsistema do aparelho fonador. Nele estão localizados a glote, a epiglote (válvula elástica que obstrui a glote durante a deglutição) e as cordas vocais. A parte mais importante deste subsistema é a glote, que consiste de uma pequena abertura de forma triangular situada próxima ao pomo-de-adão. Graças à chegada do fluxo de ar vindo dos pulmões, a glote pode abrir-se ou fechar-se, bastando que as bordas das pregas vocais se afastem ou se aproximem. Com a glote aberta, o ar passa livremente, sem fazer vibrar as cordas vocais, produzindo um fonema surdo ou não vozeado. Com o movimento cíclico de abertura da glote, causado pelo aumento da pressão do ar subglotal vindo dos pulmões, e fechamento, causado pela força de recuperação elástica e pelo efeito de Bernoulli (tendência de um orifício se fechar devido à redução da pressão quando da passagem de ar), as cordas vocais vibram numa freqüência fundamental e o fonema produzido, então, é dito sonoro ou vozeado. A taxa na qual a glote abre e fecha é controlada pela pressão de ar imposta pelos pulmões, pela tensão e rigidez das cordas vocais e pela área da abertura glotal, em condições de repouso. Resumindo, o subsistema laringeal é o responsável pela passagem da corrente de ar, que pode provocar ou não a vibração das cordas vocais.

3. Supralaringeal (ou supralaríngeo): Passagem dos pulsos ou da corrente contínua

de ar pela faringe, sujeitos a obstruções ou constrições em vários pontos de articulação (nas cavidades nasal e bucal). Este sistema completa o mecanismo da produção da fala.

O trato vocal compreende os subsistemas laringeal e supralaringeal, por ser a região situada desde as pregas vocais até as extremidades da cavidade nasal (as narinas) e da cavidade bucal (os lábios) [20].

Há estudos de modelos mecânicos para a produção da voz humana. Maiores detalhes podem ser encontrados em [21] e [22].

Figura 3.2: As cordas vocais [26].

3.1.1 Variações da voz

A voz é o produto resultante de uma seqüência complexa de transformações que ocorrem em diferentes níveis, quais sejam: 1) semântico; 2) lingüístico; 3) articulatório e 4) acústico [23].

As variações na voz relativas ao locutor são causadas pelas diferenças anatômicas no trato vocal e pelas diferenças nos hábitos de falar de diferentes indivíduos. As diferenças anatômicas estão relacionadas às diferenças de estruturas fixas, como a forma e o tamanho do trato vocal que podem variar consideravelmente de pessoa para pessoa. Por outro lado, as diferenças nos hábitos de falar resultam da maneira pela qual as pessoas aprenderam a usar o seu mecanismo de fala. Tais

diferenças aparecem nas variações temporais das características da voz de diferentes indivíduos. A forma de entonação de pessoas diferentes representa um bom exemplo dessas variações [23].

No reconhecimento de locutor deve-se considerar tanto as diferenças anatômicas quanto as diferenças nos hábitos de falar, para distinguir as vozes de diferentes locutores.

As variações intra-locutores também merecem atenção especial, pois uma mesma locução pode ser diferente, quando falada por um mesmo locutor em ocasiões diversas. Essas variações são causadas por fatores, tais como: diferenças na velocidade de emissão da fala, estado emocional do locutor, condições de saúde do locutor, etc.

É desejável selecionar para reconhecimento de locutor aqueles parâmetros acústicos de voz que apresentam pequenas variações intra-locutores e grande variação entre-locutores. Outro detalhe de suma importância é um estudo da constituição fonética do idioma o qual será objeto de estudo [23].

3.1.2 Os fonemas

A informação comunicada através da voz é intrinsecamente discreta, isto é, ela pode ser representada pela concatenação de elementos de um conjunto finito de símbolos. A fonologia é a parte da lingüística que estuda os sons da fala, do ponto de vista da função que possuem dentro de um sistema lingüístico particular. A fonologia tem por tarefa determinar, entre os sons que ocorrem em uma língua, quais são os símbolos básicos ou fonemas.

A fonética é a parte da lingüística que estuda os sons da fala enquanto realidade física e se interessa pelos mecanismos de produção e recepção dos sons pelo organismo

humano. A unidade da fonética é o fone ou som da fala, enquanto a unidade da fonologia é o fonema.

A maioria dos idiomas pode ser descrito em termos do conjunto de fonemas que possui. Este conjunto de símbolos básicos possui normalmente de 30 a 50 elementos que podem ser divididos basicamente em 4 classes: vogais, ditongos, semivogais e consoantes [23]. Neste trabalho, apenas será detalhada a classe das vogais.

3.1.3 Vogais

As vogais são produzidas pela excitação do trato vocal por pulsos de ar quase periódicos, causados pela vibração das cordas vocais. A Fig. 3.3 ilustra o sinal de voz obtido de uma vogal /a/ sustentada. A seção transversal ao longo do trato vocal determina as suas freqüências naturais, conhecidas como formantes. Portanto, cada vogal pode ser caracterizada pela configuração do trato vocal que é utilizada para a sua produção, em outras palavras, pelos formantes.

Em Português, as vogais são classificadas, de acordo com a Nomenclatura Gramatical Brasileira (NGB), considerando: a zona de articulação (conforme o posicionamento da língua), o timbre, o papel das cavidades bucal e nasal e a intensidade (átonas ou tônicas). Todas as vogais são sonoras; porém, quando sussuradas, são surdas.

Em resumo pode-se dizer que as vogais não encontram obstáculos ao serem emitidas, ou seja, a corrente de ar passa livremente; formam sílabas sozinhas ou são a base de uma sílaba; podem ser tônicas e receber o acento gráfico, quando escritas [23].

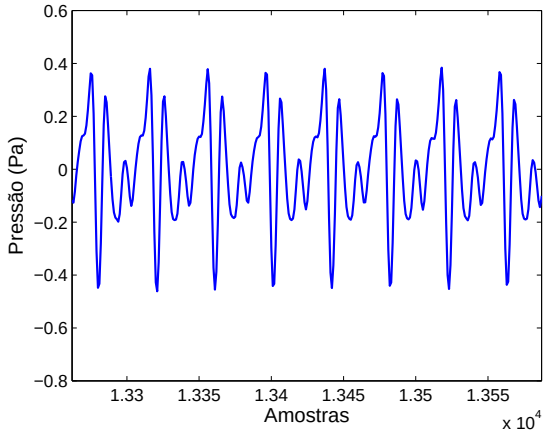

Figura 3.3: Exemplo de um sinal de voz (trecho da vogal sustentada */a/* obtida com uma freqüencia de amostragem fs=44.100 Hz).

3.2 Coarticulação

Os fonemas são conhecidos como os sons das palavras e suas características possibilitam, àqueles que os escutam, a identificar as palavras. Contudo, os fonemas que formam as palavras não são pronunciados isoladamente, pois todo o som utilizado na fala é afetado por aqueles que o antecedem e o sucedem. Portanto, os fonemas serão influenciados pelos fonemas anteriores e posteriores. A coarticulação refere-se aos efeitos de um fonema no outro, em um determinado contexto, ou seja, a articulação de cada som misturada com a articulação de um som vizinho (antes e depois). O efeito do contexto fonético pode afetar múltiplos fonemas, mas, freqüentemente, envolve os fonemas vizinhos mais próximos.

3.3 Modelo de produção sonoro/surdo da voz

Para um modelamento detalhado do processo de produção da voz, os seguintes efeitos devem ser considerados [24]: variação da configuração do trato vocal com o

tempo, perdas próprias por condução de calor e fricção nas paredes do trato vocal, a maciez das paredes do trato vocal, radiação do som pelos lábios, junção nasal, excitação do som no trato vocal, etc.

Um modelo detalhado para geração de sinais de voz, que leva em conta os efeitos da propagação e da radiação conjuntamente pode, em princípio, ser obtido através de valores adequados para excitação e parâmetros do trato vocal. A teoria acústica sugere uma técnica simplificada para modelar sinais de voz, a qual é bastante utilizada.

Essa técnica apresenta a excitação separada do trato vocal e da radiação. Os efeitos da radiação e do trato vocal são representados por um sistema linear variante com o tempo. O gerador de excitação gera um sinal similar a um trem de pulsos ou sinal aleatório (ruído). Os parâmetros da fonte e sistema são escolhidos de forma a se obter, na saída, o sinal de voz desejado [24]. Colocando-se todos os componentes necessários, obtém-se o modelo da Fig. 3.4, onde $u(n)$ é o sinal de excitação, $A_s(n)$ e $A_f(n)$ controlam a intensidade da excitação do sinal sonoro e do ruído, respectivamente, onde ocorre um chaveamento entre sonoro e surdo alterando o modo de excitação.

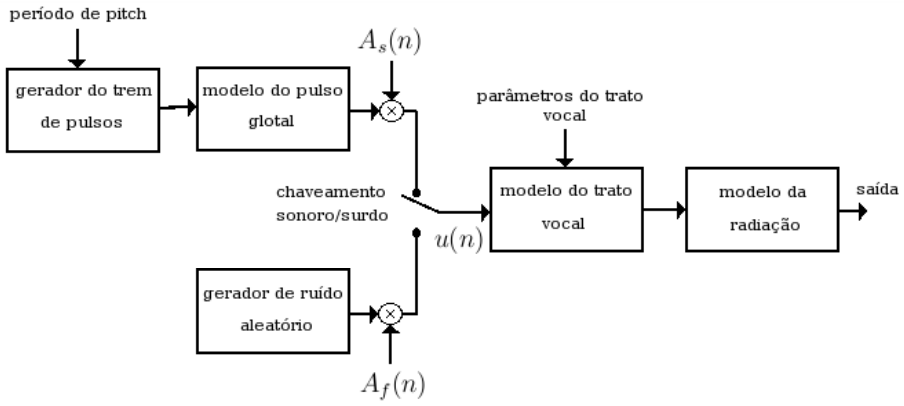

Figura 3.4: Modelo discreto da produção da voz [24].

3.4 A teoria fonte-filtro

Em 1960, Fant [25] introduziu a teoria fonte-filtro na produção da voz humana, que estabelecia que o mecanismo de produção da voz poderia ser modelado como uma fonte de excitação e um sistema de filtros conectados em série (a fonte e o filtro são considerados independentes entre si). A fonte da voz, em se tratando de sons sonoros, é proveniente do fluxo de ar que atravessa as cordas vocais. A função do trato vocal é filtrar esses sinais.

Na prática, existe uma interação entre a fonte e o trato vocal. Conseqüentemente, assumir que ambos são independentes entre si não é preciso, pois o fluxo glotal em determinado momento será influenciado pela configuração do trato vocal. Entretanto, a validade da teoria pode ser considerada suficiente para a maioria dos casos de interesse, sendo muito utilizada em processamento digital de sinais.

A análise acústica do mecanismo de produção da voz normalmente utiliza duas variáveis físicas: a pressão do som e o velocidade do volume de fluxo de ar. O fluxo glotal é freqüentemente expresso em termos de velocidade do volume [26]. A forma de onda da velocidade do volume na boca, determina a pressão do sinal, que se propagará no espaço livre. A radiação dos lábios, normalmente reduzida a uma operação de diferenciação [27], é considerada em detalhes por [28] e [25].

Capítulo 4

O sinal glotal

4.1 O sinal glotal

A expansão-contração dos pulmões pode ser considerada o ponto de partida para a geração do sinal glotal, pois gera a diferença de pressão entre o ar nos pulmões e o ar próximo à boca. O escoamento de ar provocado por essa diferença passa através das cordas vocais, que vibram em uma freqüência intimamente relacionada à tensão dos músculos associados [18]. Esta vibração altera o fluxo de ar, transformando-o em um trem de pulsos ou sinal glotal. O processo está esquematizado na Fig. 4.1. A Fig. 4.2 é um exemplo de sinal glotal obtido através do sinal de voz. Esse sinal foi obtido por filtragem inversa e será detalhado posteriormente.

O sinal glotal possui propriedades importantes de difícil reprodução que estão intimamente ligadas às características anatômicas e fisiológicas da laringe. Atualmente, a teoria mais aceita para a descrição do sinal glotal (isto é, o aparecimento do trem de pulsos) é a teoria chamada de aerodinâmica mioelástica [29] [30]. Esta teoria postulou que os movimentos de abrir e fechar as cordas vocais são regidos pelas propriedades mecânicas dos tecidos musculares que constituem, principalmente, as

cordas vocais e pelas forças aerodinâmicas que se distribuem ao longo da laringe durante a fonação. A ação neural consiste apenas em aproximar as cordas vocais de tal forma que a superfície destas vibrem [18].

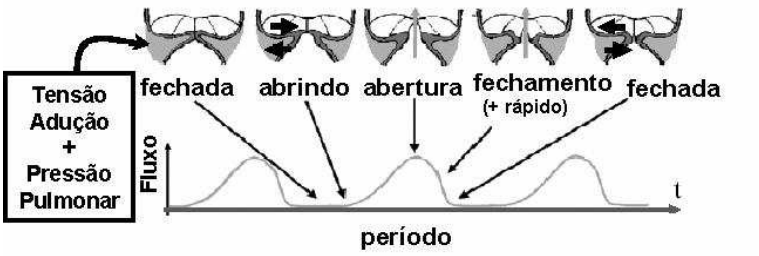

Figura 4.1: A formação do sinal glotal.

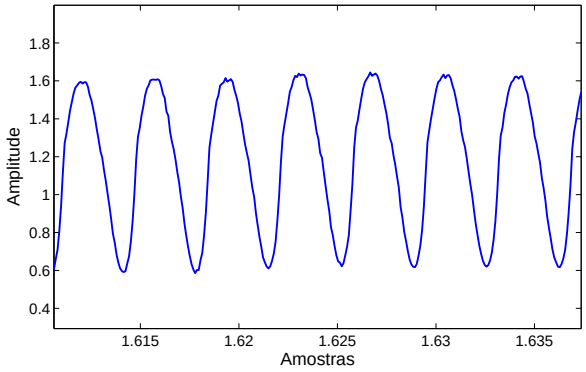

Figura 4.2: Sinal glotal da vogal sustentada representada na Fig. 3.3, obtido por filtragem inversa.

4.2 A derivada do sinal glotal

A derivada do sinal glotal auxilia na determinação dos diversos parâmetros da fonte glotal; um exemplo é o instante de início do fechamento da glote. Os picos de máximo da derivada do sinal glotal (DOFG) foram comparados aos picos de máximo da derivada do sinal eletroglotográfico (DEGG). Durante as simulações foi veri-

ficada a ocorrência de ambos em instantes próximos, sugerindo uma relação entre os sinais.

Existem modelos que representam o sinal glotal e sua derivada, porém o modelo da forma de onda da velocidade do volume glotal mais utilizado é o modelo Liljencrants-Fant (LF) [31], ilustrado na Fig. 4.3. Este modelo possui quatro parâmetros que, juntamente com o período do ciclo glotal, determinam a forma do pulso.

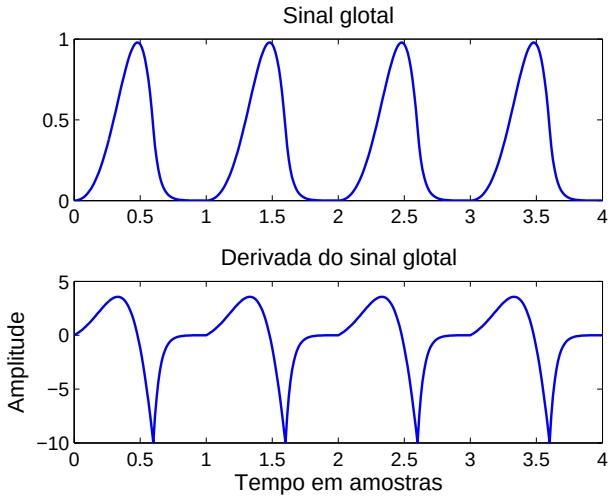

Figura 4.3: Modelo LF. Visualização do sinal glotal e sua derivada

4.3 Parâmetros

Os parâmetros que descrevem o fluxo glotal podem ser usados em uma variedade de aplicações, tais como: pesquisas sobre a produção da voz, codificação, síntese, reconhecimento automático de voz, uso clínico, verificação e identificação automática de locutor [32] e para quantificar a contribuição do pulso glotal na transmissão de sentimentos [33].

4.3.1 Instantes de máxima abertura e máximo fechamento glotal

O instante de máximo fechamento é definido como instante em que o fluxo glotal atinge seu valor mínimo. Fisiologicamente, corresponde ao instante que as cordas vocais começam a se separar. O instante de máxima abertura está associado ao máximo da excitação glotal, em outras palavras, corresponde ao instante que o fluxo glotal atinge seu valor máximo.

4.3.2 Diferença entre os instantes de máximo (Ko)

Após a obtenção dos instantes de máximo fechamento e máxima abertura, respectivamente, será calculada a diferença entre esses instantes, definida como (Ko). A Fig. 4.4 ilustra os instantes de máximo e sua diferença.

4.3.3 Amplitude de vozeamento

A amplitude de vozeamento (Av) é definida como a amplitude entre os valores mínimo e máximo do sinal glotal.

Existem outros parâmetros que descrevem a forma do pulso glotal, como por exemplo, o quociente de amplitude normalizada (NAQ) [34] que pode ser utilizado com a finalidade de parametrizar a fase de fechamento do pulso glotal. Existem, também, aqueles que foram utilizados para estudar os efeitos da carga vocal por filtragem inversa [35], entretanto, estes parâmetros não serão objeto de estudo deste trabalho.

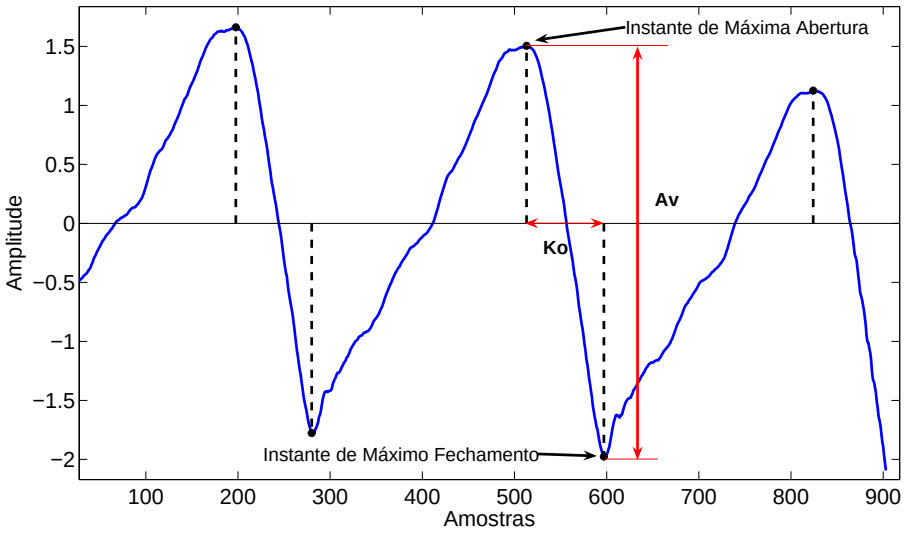

Figura 4.4: Sinal glotal e seus parâmetros obtidos de uma vogal sustentada /a/

Capítulo 5

O eletroglotógrafo

5.1 O sinal do eletroglotógrafo (EGG)

A eletroglotografia é um método não invasivo criado por [7], que estima a variação da área de contato entre as cordas vocais (*vocal fold contact area - VFCA*), durante a produção da voz [8] [9] [10], e vem sendo utilizado para fins clínicos e de pesquisa [7]. Resumidamente, o eletroglotógrafo mede as variações da impedância elétrica causadas pela variação da área de contato entre as cordas vocais, uma vez que estas vibram. A impedância é medida através de um par de eletrodos, presos ao pescoço do locutor, que aplicam uma pequena corrente elétrica ao local.

O princípio de funcionamento do eletroglotógrafo é baseado na medição da impedância entre dois eletrodos colocados no pescoço do locutor. Quando as cordas vocais estão fechadas, a corrente elétrica passa por elas, ou seja, há baixa impedância. Quando as cordas vocais estão separadas, devido ao fluxo de ar que as atravessa, a impedância da laringe é alta. Portanto, a impedância da laringe varia de acordo com a área de contato das cordas vocais [26] [10].

Diversos estudos comparativos foram realizados utilizando fotografia estro-

boscópica [36] [37] [38] [39], vídeo estroboscópico [40] [41], imagens de alta velocidade [42] [43] [44], fotoglotografia [45] [46] [47] [48] [49] [50], medidas de pressão subglotal [49], e filtragem inversa [39] [51] [52] [53]. Todos esses estudos confirmaram que o sinal do eletroglotógrafo está relacionado com a área de contato das cordas vocais: quanto maior o contato da superfície, maior a admitância medida [10].

Eletrodos são colocados por cima da pele, posicionados em cada lado da laringe e uma corrente alternada de alta freqüência circula entre eles, com o intuito de medir a impedância entre os eletrodos. A freqüência, normalmente, é na ordem de MHz e a corrente é limitada a alguns miliampères para garantir que será imperceptível, evitando desconforto [9]. A voltagem entre os eletrodos é em torno de 1 V rms [54]. A Fig. 5.1 ilustra o funcionamento do eletroglotógrafo.

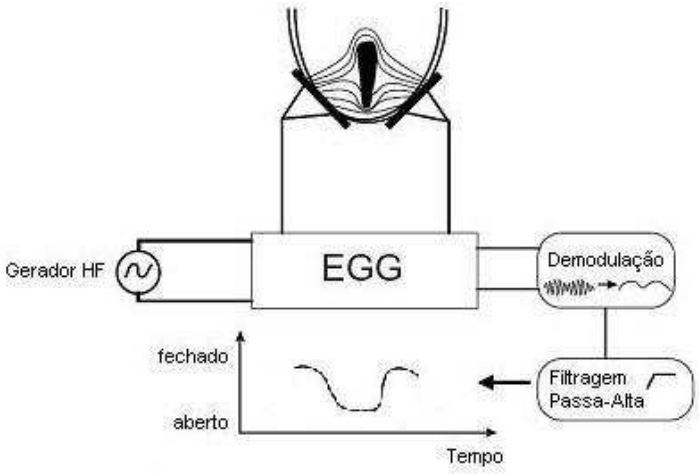

Figura 5.1: Eletroglotógrafo [10].

O sinal eletroglotográfico resultante, o eletroglotograma, mostra a variação, nas cordas vocais, da impedância em função do tempo. Essa variação é relativamente pequena, normalmente, apenas de 1 a 2% da total da impedância medida [9]. A impedância também varia consideravelmente com os tipos de pele e com os

movimentos verticais da laringe. Filtros passa-altas são usados para eliminar as interferências de baixa freqüência e extrair apenas as variações causadas pela vibração das cordas vocais. Ademais, um controle automático de ganho também foi construído, dentro do Eletroglotógrafo, para manter um nível de sinal apropriado a despeito de consideráveis variações de impedância entre pessoas e durante uma simples gravação. Essas técnicas causam distorções de fase e amplitude que podem influenciar a forma de onda do EGG [55]. Conseqüentemente, o sinal EGG não pode ser considerado uma medida absoluta do contato entre as cordas vocais, e certos cuidados devem ser tomados quando da interpretação do sinal. A despeito dessas limitações, o sinal EGG é uma informação útil sobre a situação das cordas vocais durante a fonação. A Fig. 5.2 ilustra um dos sinais EGG obtidos.

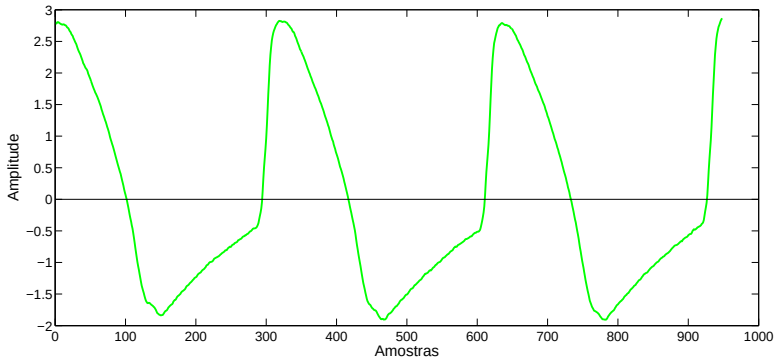

Figura 5.2: Sinal EGG

A referência [52] apresentou um modelo de diferentes fases do período do sinal EGG e sua relação com os eventos fisiológicos que ocorrem na laringe. Existem outros modelos similares [51], que são, contudo, simplificações idealizadas, que não devem ser interpretadas literalmente. Diversos autores apontam que o sinal EGG não permite determinar, exatamente, o instante de início de fechamento glotal e que localizar o instante de início de abertura glotal a partir do sinal EGG é ainda mais

impreciso [8] [9].

O estudo do sinal do eletroglotógrafo (sinal EGG) é importante, pois sua derivada primeira fornece os instantes de início de abertura e início de fechamento glotal e a estimação precisa da freqüência fundamental do sinal [10].

5.1.1 A derivada do sinal do eletroglotógrafo (DEGG)

Comparando a forma de onda do sinal EGG com imagens de alta velocidade, a referência [51] relatou que o fechamento glotal ocorre no instante o qual a derivada primeira do sinal EGG (DEGG) possui seu pico de máximo. Estes picos do sinal DEGG são claramente identificados, conforme Fig. 5.3 - Sinal DEGG.

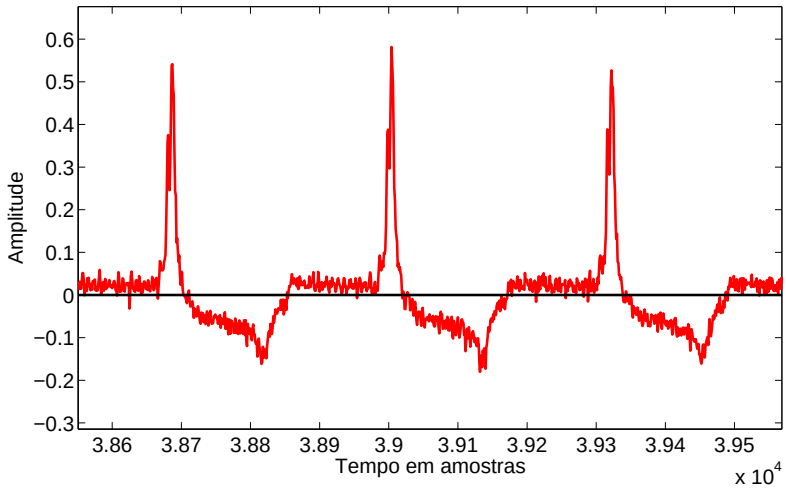

Figura 5.3: Sinal DEGG

Esta abordagem foi corroborada por [10], que considerou os picos do DEGG como indicadores de início de abertura e início de fechamento glotal. As Figs. 5.4 e 5.5 ilustram esta relação. Entretanto, em alguns casos, podem ocorrer picos imprecisos, duplos e até a ausência de picos.

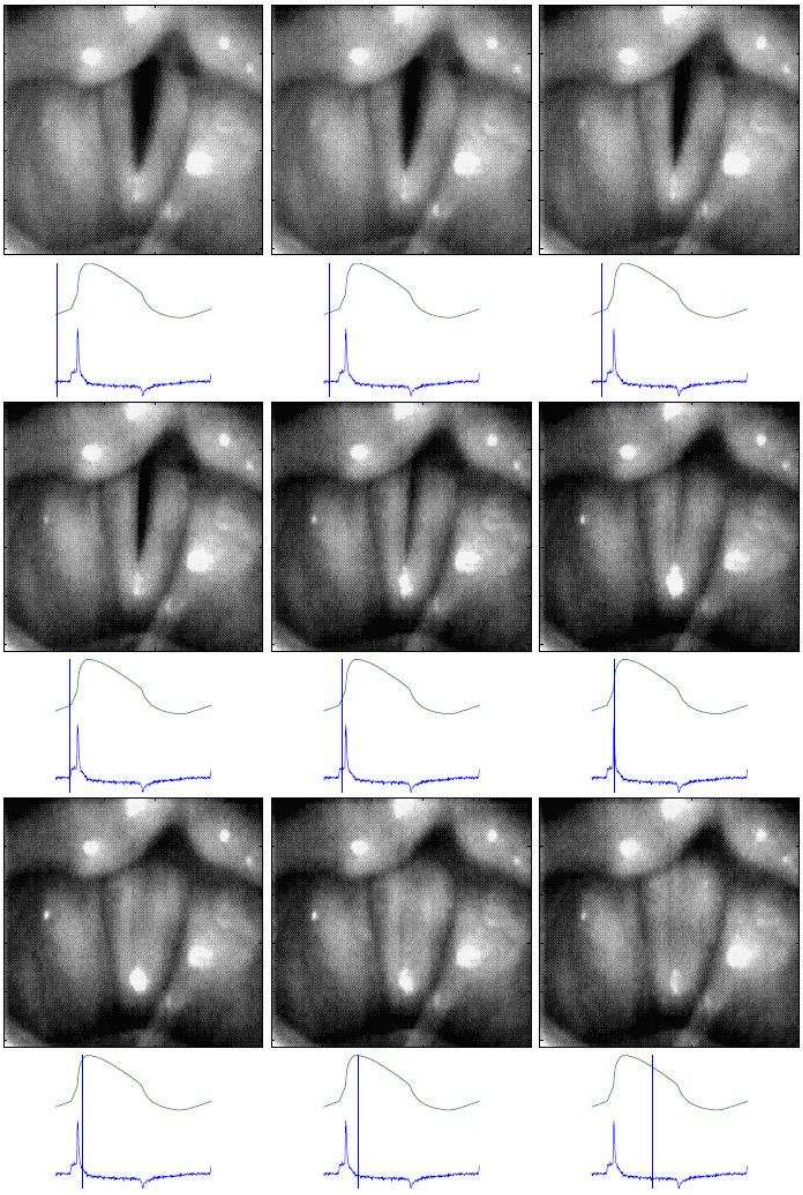

Figura 5.4: Visualização do fechamento por cinematografia ultra-rápida e eletroglotografia simultâneas (locutor em fonação normal e freqüência fundamental igual a 110 Hz - sinais EGG e DEGG) [10].

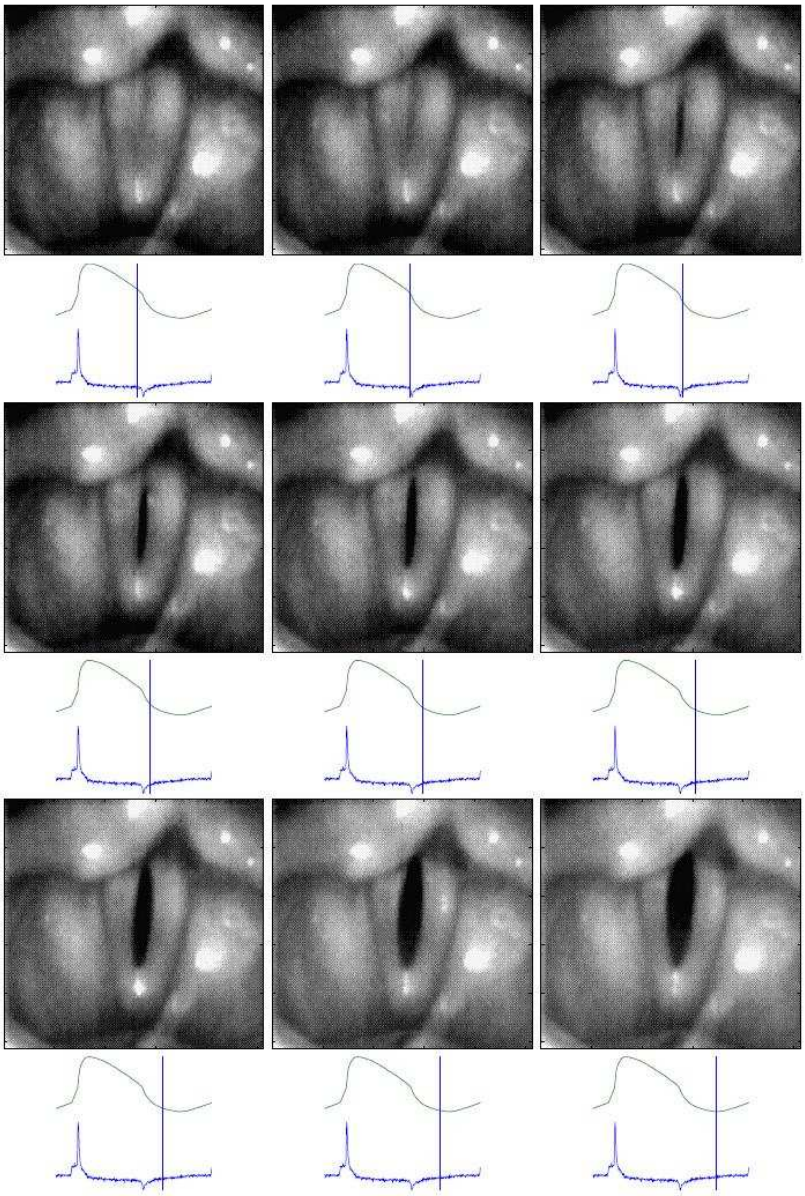

Figura 5.5: Visualização da abertura por cinematografia ultra-rápida e eletroglotografia simultâneas (locutor em fonação normal e freqüência fundamental igual a 110 Hz - sinais EGG e DEGG) - sinais EGG e DEGG [10].

Em adição à resistência em torno do pescoço, a medida da impedância também é influenciada pela reatância (capacitiva ou indutiva) da carga examinada. Uma capacitância variável pode hipoteticamente existir na glote quando as duas cordas vocais estão separadas por uma fina camada de ar, como suposto por [52]. Esta hipótese pode ser verificada alterando a freqüência da corrente alternada usada para medição da impedância: a corrente permanecerá a mesma apenas se a carga for puramente resistiva. De acordo com [55], a impedância é essencialmente resistiva em uma ampla faixa de freqüência [26].

5.2 Parâmetros

Os parâmetros que podem ser obtidos do sinal do eletroglotógrafo são os instantes de início de fechamento e início de abertura glotal, encontrados a partir do sinal DEGG, e os instantes de máximo fechamento e máxima abertura, encontrados a partir do sinal EGG. As Figs. 5.6 (a),(b) e (c) apresentam um trecho de um sinal de voz de uma vogal /a/, o sinal EGG e o sinal DEGG, respectivamente.

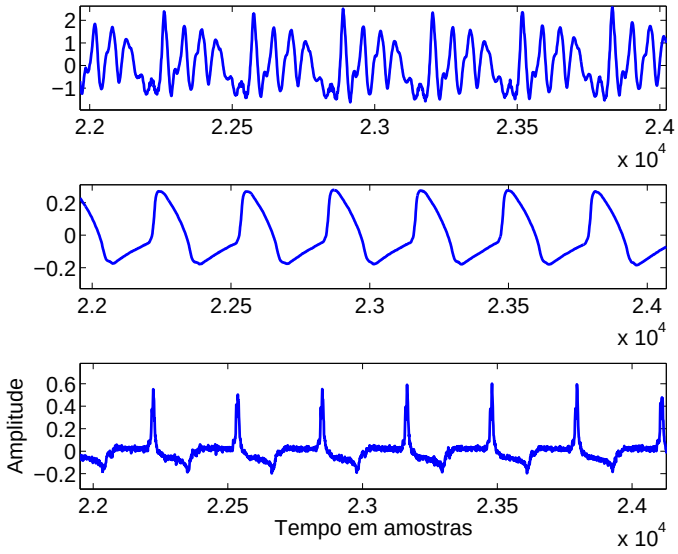

Figura 5.6: (a) sinal de voz, (b) sinal EGG e (c) sinal DEGG.

5.2.1 Instante de início de abertura

Fisiologicamente, esse instante corresponde ao instante no qual as cordas vocais iniciam sua separação (início da redução da área de contato entre as cordas vocais). O instante de início de abertura (*iabert*) é definido como o instante em que o sinal DEGG (derivada do sinal EGG) atinge o pico de mínimo, conforme Figs. 5.5 e 5.7.

5.2.2 Instante de início de fechamento

Fisiologicamente, corresponde ao instante no qual as cordas vocais iniciam seu fechamento (aumento da área de contato entre as cordas vocais). O instante de início de fechamento (*ifec*) ocorre quando o sinal DEGG atinge seu pico de máximo, conforme Figs. 5.4 e 5.7.

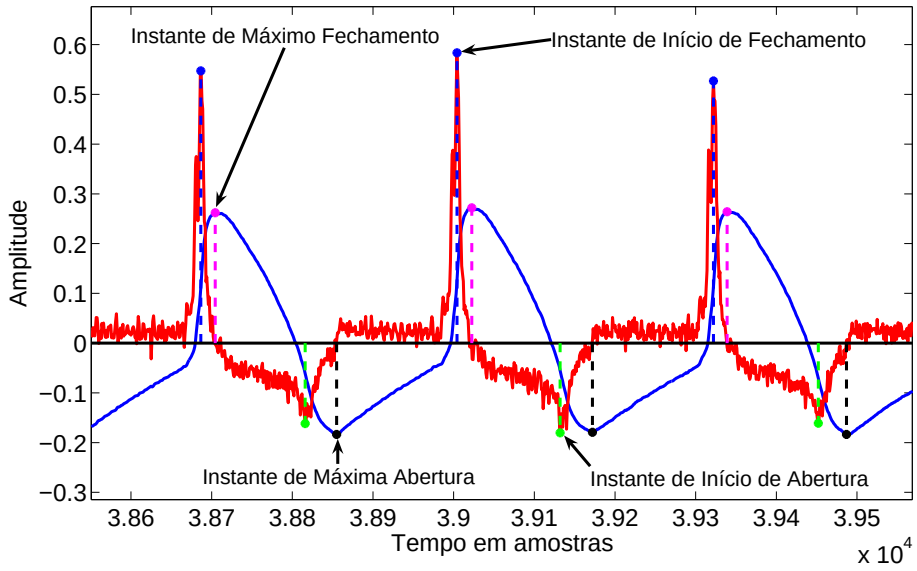

Figura 5.7: Sinais EGG e DEGG com seus instantes de início de fechamento, início de abertura e os instantes de máximo fechamento e abertura.

5.2.3 Instantes de máximo fechamento e máxima abertura

Define-se o instante de máximo fechamento como o momento de maior área de contato entre as cordas vocais, que ocorre após o instante de início de fechamento, quando o sinal EGG atinge seu pico de máximo, como mostrado na Fig. 5.7. Analogamente, o instante de máxima abertura é definido como o momento de menor área, que ocorre após o instante de início de abertura, quando o sinal EGG atinge seu pico de mínimo.

5.2.4 Diferença entre os instantes de máximo (Ke) e Amplitude EGG

Após a obtenção dos instantes de máxima abertura e máximo fechamento, será calculada a diferença entre esses instantes, definida como (Ke). A amplitude EGG

(*Avegg*) é definida como amplitude entre os valores mínimo e máximo do sinal EGG. A Fig. 5.8 ilustra os instantes de máximo, a diferença entre eles e a amplitude do sinal EGG.

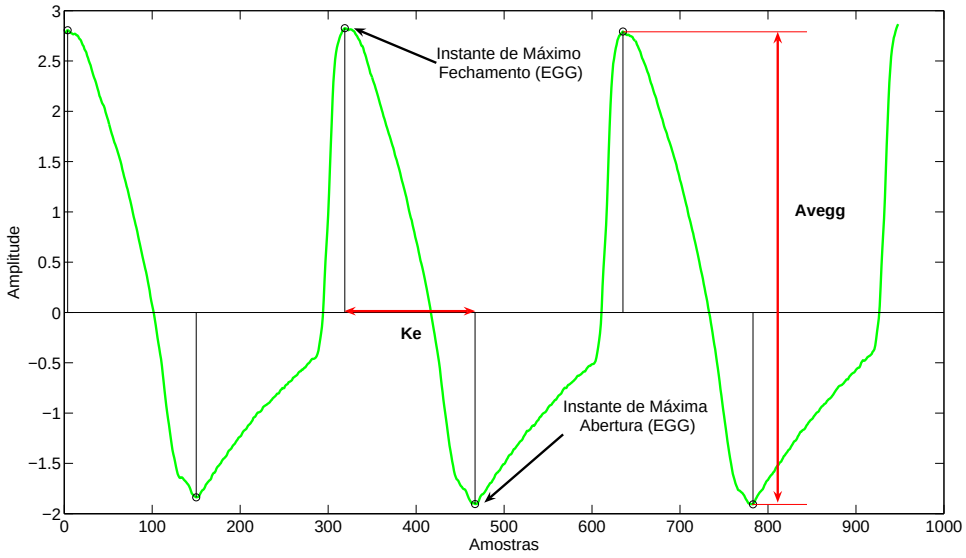

Figura 5.8: Sinal EGG e os instantes de máximo fechamento, abertura e amplitude EGG.

5.2.5 Comparação entre os instantes de máximo fechamento e máxima abertura dos sinais OFG e EGG

Inicialmente, foram calculados os instantes de máximo fechamento e máxima abertura do sinal glotal e do EGG. Os parâmetros (*Da*) e (*Df*) visam quantificar a diferença entre os instantes encontrados em cada sinal, em outras palavras, fazem uma comparação entre os instantes de máxima abertura (*Da*) e de máximo fechamento (*Df*) encontrados nos dois sinais. A Fig. 5.9 ilustra essas duas diferenças.

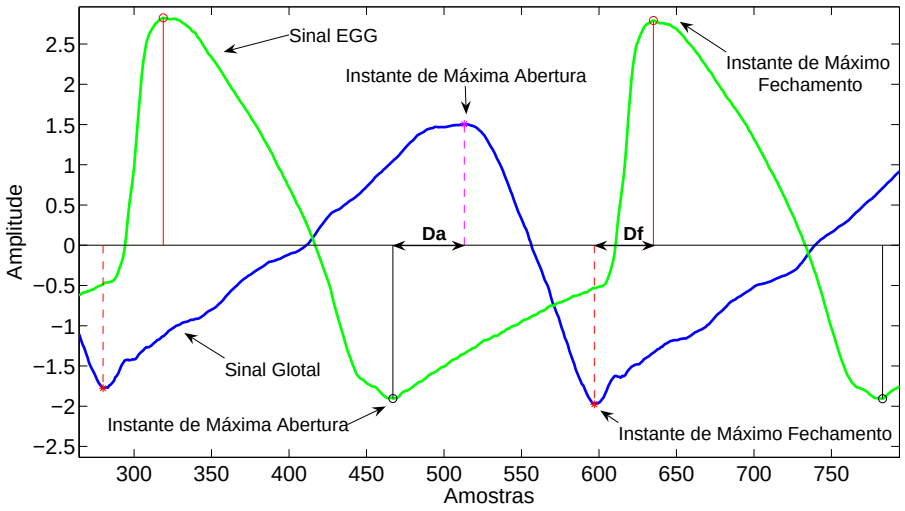

Figura 5.9: Comparação entre os instantes de máximo fechamento e máxima abertura dos sinais glotal e EGG.

5.2.6 Diferença entre instantes de início de abertura e início de fechamento

A diferença entre os instantes de início de abertura e início de fechamento, definida como *Kd1* e a diferença entre os instantes de início de fechamento e início de abertura, definida com *Kd2*, ambas encontrados a partir do sinal DEGG estão representadas na Fig. 5.10.

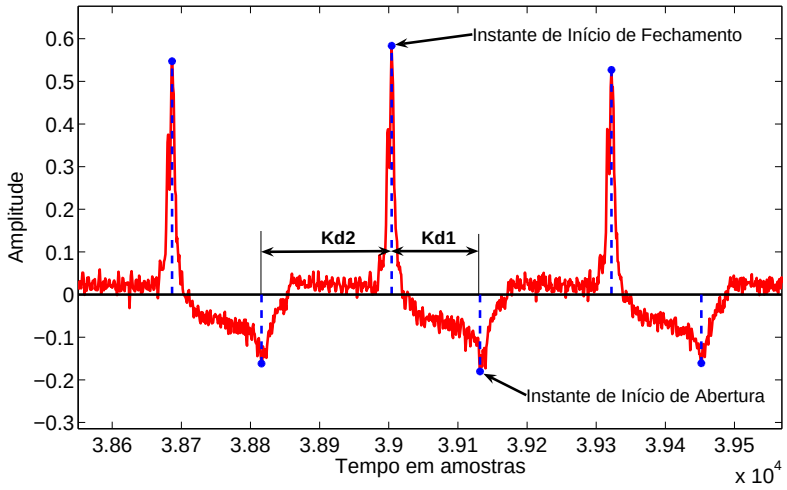

Figura 5.10: Diferença entre instantes de início de fechamento e início de abertura.

5.2.7 Comparação entre os picos de máximo da derivada do sinal glotal (DOFG) e da derivada do sinal EGG (DEGG) e a variação Koe.

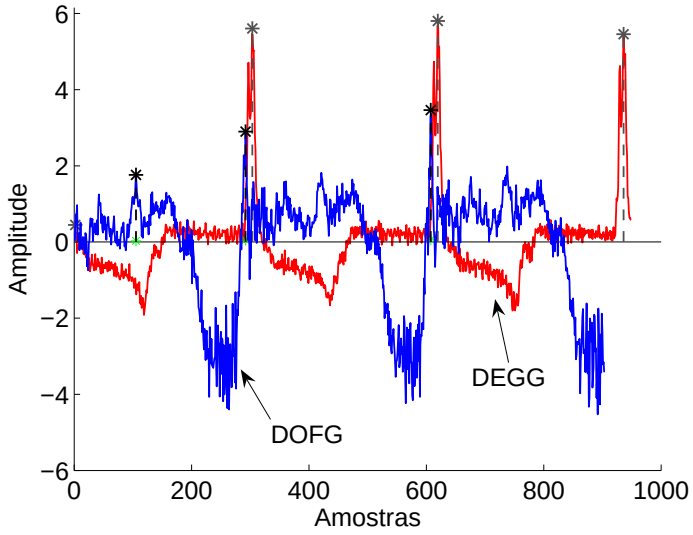

Figura 5.11: Picos de máximo dos sinais DOFG e DEGG.

Durante as simulações foi observado que os picos de máximo do sinal DEGG,

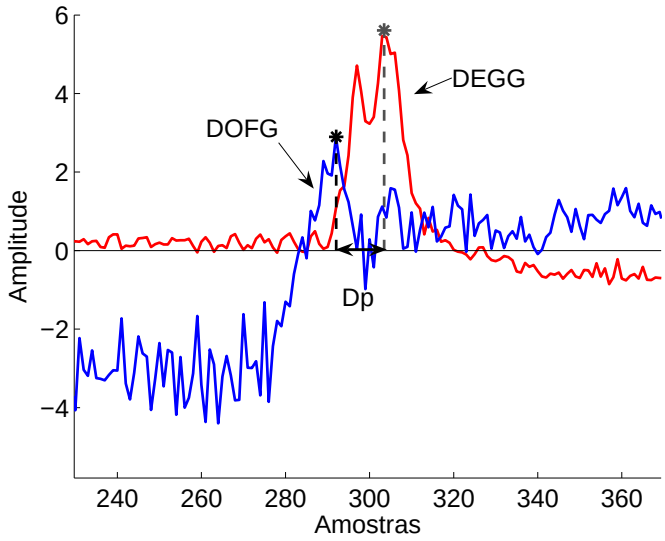

Figura 5.12: Diferença entre os picos de máximo (parâmetro *Dp*).

representando o instante de início de fechamento, ocorriam próximos dos picos do sinal DOFG (Fig. 5.11). A partir desta observação foram calculados os instantes de ocorrência de ambos os picos e seus valores comparados entre si, sendo o resultado definido como *Dp*, ilustrado na Fig. 5.12. Outro parâmetro de comparação encontrado foi a variação *Koe*, definida com a diferença entre os parâmetros *Ko* e *Ke*, definidos anteriormente.

5.2.8 Resumo dos parâmetros

A Tabela 5.1 contêm um resumo de todos os parâmetros contidos, neste trabalho.

Tabela 5.1: Resumo dos parâmetros.

	RESUMO DOS PARÂMETROS
	Sinal EGG
Ke	Diferença entre os instantes de máxima abertura e de máximo fechamento
Avegg	Amplitude EGG
	Sinal DEGG
iabert	Instante de início de abertura
ifec	Instante de início de fechamento
Kd1	Diferença entre o instante de início de abertura e o instante de início de fechamento
Kd2	Diferença entre o instante de início de fechamento e o instante de início de abertura
	Sinal OFG
Ko	Diferença entre os instantes de máximo fechamento e de máxima abertura
Av	Amplitude de vozeamento
	Comparação entre EGG e OFG
Df	Diferença entre os instantes de máximo fechamento
Da	Diferença entre os instantes de máxima abertura
Keo	Ke-Ko
Dp	Diferença entre os picos de máximo do DOFG e do DEGG

5.3 O eletroglotógrafo EG2-PCX

O modelo EG2-PCX, fabricado pela *Glottal Enterprises*, foi o eletroglotógrafo cedido pelo Instituto Militar de Engenharia (IME) para ser utilizado neste trabalho. As Fig. 5.13 e 5.14 ilustram o equipamento.

Figura 5.13: Parte frontal do eletroglotógrafo.

Figura 5.14: Parte traseira do eletroglotógrafo.

Conforme explicado no início deste capítulo, o sinal EGG é captado por dois eletrodos posicionados, apropriadamente, próximos às cordas vocais. O eletrodo que deverá ser posicionado no lado esquerdo do pescoço do locutor possui dois fios na cor vermelha. A Fig. 5.15 revela a existência de duas placas douradas em cada eletrodo, separadas por uma tarja de material isolante que, em contato com o pescoço, deverá permanecer na posição horizontal (tarja paralela ao velcro que prende os eletrodos ao pescoço). O gel que acompanha o EG2-PCX deve ser colocado nas placas douradas para facilitar a passagem da corrente elétrica.

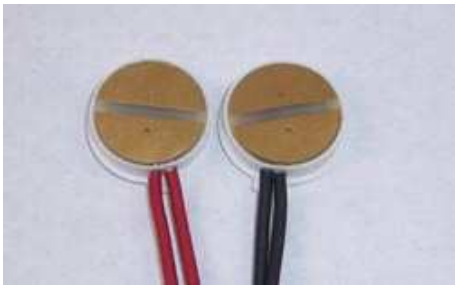

Figura 5.15: Eletrodos do eletroglotógrafo - diâmetro 34mm.

O indicador no painel frontal do eletroglotógrafo que auxilia no posicionamento dos eletrodos e provê uma indicação quantitativa do movimento vertical da laringe está representado na Fig. 5.16. O ideal é que, durante a fala, um ou mais *leds* verdes estejam acessos; caso contrário, os eletrodos deverão ser reposicionados. Este teste deverá ser realizado antes do início das gravações sob pena de comprometer a qualidade do sinal obtido.

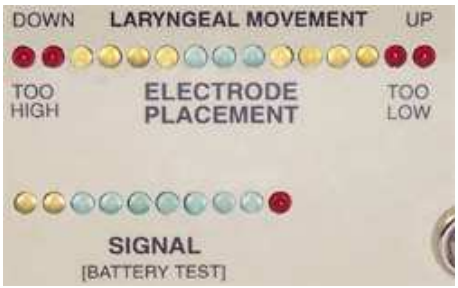

Figura 5.16: Indicação quantitativa dos movimentos verticais da laringe e auxílio visual ao posicionamento dos eletrodos.

Após o correto posicionamento dos eletrodos, o microfone (Fig. 5.17) que captará o sinal de voz também deverá ser conectado ao eletroglotógrafo (parte traseira), que apresentará, em sua saída, o sinal EGG e o sinal da voz sincronizados, porém separados. A saída do eletroglotógrafo deverá ser conectada a um computador, via porta *USB* ou cabo estéreo que, através de um *software* de voz, efetuará as

gravações dos sinais no disco rígido ou em outra unidade previamente selecionada. Caso a conexão seja efetuada com o cabo estéreo, o canal esquerdo fornecerá o sinal de áudio (microfone) e o direito o sinal EGG.

Neste trabalho, o *software* de voz escolhido foi o *Audacity* e a conexão entre o eletroglotógrafo e o computador foi realizada através da porta *USB*.

Figura 5.17: Microfone usado com o eletroglotógrafo para captação do sinal de voz, compensador de fase (C-1) e simulador de laringe (LS-1), respectivamente.

O modelo EG2-PCX ainda possui um simulador de laringe (LS-1) e um compensador de fase (C-1), externos ao equipamento. O simulador de laringe, como o próprio nome sugere, é usado no lugar dos eletrodos para simular as pequenas variações de resistência do pescoço medidas pelo eletroglotógrafo durante a produção da voz. O LS-1 é um auxílio à medição e à verificação do desempenho dos equipamentos, uma vez que pode ser conectado em qualquer unidade do fabricante, possibilitando um comparação objetiva entre eletroglotógrafos. O C-1 foi desenvolvido para minimizar as distorções de fase, inseridas pela filtragem passa alta, e pode ser utilizado em conjunto com o LS-1, quando a aquisição de dados for realizada pela placa de som de um computador. A Fig. 5.17 ilustra os referidos simulador de laringe e compensador de fase.

Capítulo 6

Filtragem inversa

Como já discutimos, o fluxo de ar, proveniente dos pulmões, é alterado pela vibração das cordas vocais gerando o sinal glotal, que serve de excitação do trato vocal e gerando, finalmente, a voz. Portanto, o estudo do sinal glotal é de suma importância na compreensão da produção da voz.

A estimação do sinal glotal tem sido alvo de várias pesquisas e, ao longo dos anos, várias técnicas têm sido desenvolvidas visando obter informações a respeito da formação e modelagem deste sinal. Porém, um método de análise do sinal glotal que seja fácil de usar e obtenha bons resultados com uma grande variedade de sinais de voz ainda não foi desenvolvido [34].

Uma técnica amplamente empregada na estimação do sinal glotal é a filtragem inversa. Entretanto, mesmo entre as diversas versões desta técnica, todas são baseadas na mesma idéia: o pulso glotal é obtido cancelando os efeitos dos formantes na voz. O trato vocal deve ser modelado e, então, os efeitos dos formantes são cancelados filtrando o sinal de voz através do inverso do trato vocal [34].

O PSIAIF (*Pitch Synchronous Iterative Adaptive Inverse Filtering*) é um método

de filtragem inversa, semi-automático, desenvolvido por [34], que utiliza o sinal de voz como entrada e apresenta na saída uma estimação do fluxo glotal correspondente.

Neste capítulo, será apresentada a teoria relacionada ao método PSIAIF, assim como os algoritmos usados para sua implementação.

6.1 Filtro de pré-ênfase

A filtragem de pré-ênfase serve para atenuar as componentes de baixa freqüência e incrementar as componentes de alta freqüência do sinal de voz, prevenindo contra instabilidade numérica e, também, minimizando o efeito dos lábios e da glote [56].

A função de transferência mais usada para um filtro de pré-ênfase é dada por [24]:

$$H(z) = 1 - az^{-1}, \qquad 0,9 \leq a \leq 1,0. \tag{6.1}$$

Neste caso, a saída do sistema de pré-ênfase $\tilde{s}(n)$ está relacionada à entrada $s(n)$ pela equação de diferenças:

$$\tilde{s}(n) = s(n) - as(n-1) \tag{6.2}$$

onde o valor de a usualmente empregado é 0,95 [24].

6.2 Janelamento

Após a pré-ênfase, passa-se à etapa de "janelamento"do sinal de voz. Nesta etapa, são extraídos quadros de N amostras a partir do sinal $\tilde{s}(n)$, tendo uma superposição de M amostras (ver Fig. 6.1). Tal divisão é extremamente importante devido ao fato de um sinal de fala não estacionário. O janelamento de pequenos segmentos, que variam de 10 a 30 ms, possibilita minimizar as descontinuidades do sinal no

começo e no final de cada janela (*frame*) e admitir que ele seja aproximadamente estacionário nesses intervalos [24] permitindo, assim, o uso de métodos tradicionais de análise espectral. Geralmente, para separar cada segmento do sinal de voz, usa-se uma janela de Hamming [24] [56], definida por

$$h(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), & 0 \leq n \leq N-1 \\ 0, & c.c. \end{cases} \tag{6.3}$$

onde n é o índice da amostra e N é o número total de amostras da janela.

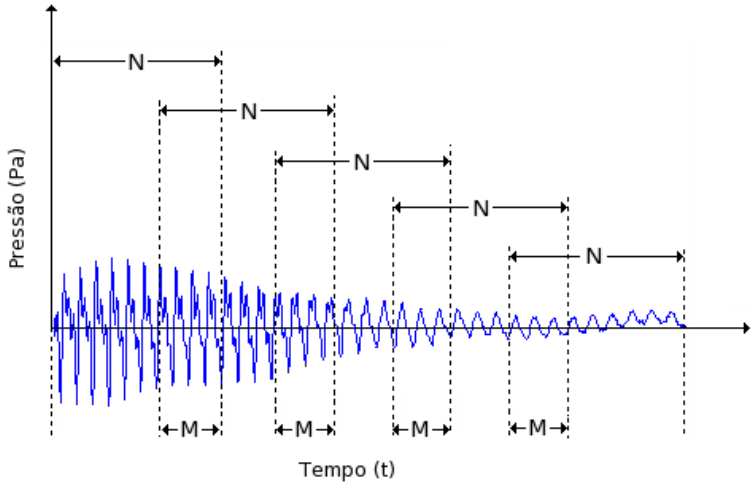

Figura 6.1: Divisão em quadros do sinal de voz.

6.3 Algoritmo de filtragem inversa

A teoria fonte-filtro da produção da voz provê o embasamento teórico necessário para a técnica de filtragem inversa. Se a função de transferência do filtro do trato vocal é conhecida, uma filtragem inversa poderá ser realizada. Em princípio, o sinal da excitação glotal pode ser reconstruído passando o sinal de voz pelo inverso do filtro do trato vocal. Na prática, a função de transferência do filtro do trato vocal pode

ser aproximada, baseando-se no sinal de voz e no mecanismo de produção da voz. Aplicando a técnica de filtragem inversa ao sinal de voz obteremos uma estimação da excitação glotal, a forma de onda do fluxo glotal, que também é conhecida como FGG (*flow glottogram*) [57] [58]. Atualmente, a maioria das técnicas de filtragem inversa são digitais devido à flexibilidade e facilidade de implementação quando comparadas aos filtros analógicos.

Os métodos de filtragem inversa digital podem ser divididos em técnicas manuais, semi-automáticas e automáticas. Os métodos manuais requerem o ajuste dos filtros para determinar os formantes do sinal de voz, diferentemente das ténicas automáticas que constroem um modelo do filtro do trato vocal e encontram os parâmetros dos filtros, normalmente por análise LPC [58]. Os métodos semi-automáticos encontram-se entre os dois extremos. O método proposto por [34] é um bom exemplo de método semi-automático, pois, basicamente, o filtro do trato vocal é encontrado automaticamente, mas o usuário pode controlar certos parâmetros que afetarão o resultado final do fluxo glotal. A referência [59] comparou um método de filtragem inversa automático com um manual e concluiu que há extrema semelhança entre os resultados obtidos em cada método.

A filtragem inversa, basicamente, envolve a extração de dois sinais, o sinal glotal e o efeito do filtro do trato vocal, de uma única fonte de sinal. Entretanto, a técnica adota diversas aproximações a respeito do fluxo glotal e da função de transferência do trato vocal. Conseqüentemente, o resultado da filtragem inversa deve ser considerado uma estimação do sinal glotal. O fluxo glotal em si ainda não é conhecido exatamente [26].

A precisão da filtragem inversa se deteriora caso a freqüência fundamental da

voz seja alta, pois a estrutura espaçada dos harmônicos do espectro da excitação interfere com os formantes, que são ressonâncias locais no espectro [26].

A despeito dessas limitações a filtragem inversa provou ser uma ferramente valiosa para a pesquisa do mecanismo de produção da voz [60]. É uma técnica não evasiva e que não requer equipamentos caros.

6.4 Análise LPC

A predição linear é uma técnica muito utilizada em processamento digital de sinais e consiste em usar amostras anteriores do sinal para estimar a amostra atual [24]. A Fig. 6.2 ilustra o modelo de predição linear da voz.

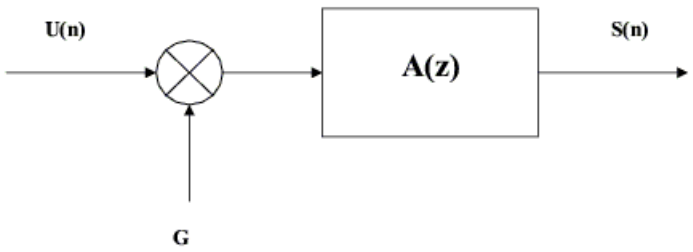

Figura 6.2: Modelo de predição linear da voz [24].

A formulação geral considera inicialmente $s_p(n)$ como uma seqüência de amostras obtidaspor um preditor linear de M-ésima ordem, conforme Eq. (6.4):

$$\mathrm{s(n)} = \sum_{\mathrm{k=1}}^{\mathrm{M}} \mathrm{a_k s(n-k)} + \mathrm{Gu(n)} \tag{6.4}$$

$$\mathrm{s_p(n)} = \sum_{\mathrm{k=1}}^{\mathrm{M}} \mathrm{a_k s(n-k)} \tag{6.5}$$

onde a_k, com k=1,2,3,...,M, representam os coeficientes do preditor linear, $u(n)$ é a excitação normalizada e G o ganho [24].

O erro de predição entre a amostra atual s(n) e a amostra predita $s_p(n)$ é dado pela Eq.6.6.

$$\mathrm{e(n) = s(n) - s_p(n)\,.} \tag{6.6}$$

Substituindo-se (6.5) em (6.6), e calculando a respectiva transformada-Z, tem-se:

$$E(z) = S(z) - \sum_{\mathrm{k=1}}^{\mathrm{M}} \mathrm{a_k z^{-k}} S(z) \tag{6.7}$$

sendo

$$S(z) = \sum_{\mathrm{i=1}}^{\mathrm{M}} \mathrm{a_i z^{-i}} S(z) + GU(z) \tag{6.8}$$

e

$$\frac{\mathrm{E(z)}}{\mathrm{S(z)}} = \mathrm{A(z)} = 1 - \sum_{\mathrm{k=1}}^{\mathrm{M}} \mathrm{a_k z^{-k}}\,. \tag{6.9}$$

6.4.1 IAIF

O método de filtragem inversa, semi-automático e conhecido como IAIF, foi desenvolvido por [34] e utiliza o sinal de voz como entrada a fim de obter, na saída, uma estimação do fluxo glotal correspondente. O modelo de produção da voz, o qual o IAIF é baseado está representado na Fig. 6.3 .

O IAIF é composto de três blocos fundamentais. São eles: análise LPC, filtragem inversa e integração. A análise LPC é responsável pela filtragem de pré-

ênfase, pela estimação do trato vocal e da contribuição glotal, definidas através da ordem de seus coeficientes discutida, ainda, neste capítulo. A filtragem inversa é responsável pela eliminação do trato vocal e da contribuição glotal no sinal da voz e a integração pela eliminação da radiação dos lábios.

Como visto na Fig. 6.5, o sinal de entrada é passado por um filtro passa alta com intuito de eliminar asfreqüências baixas, que provocam flutuações na saída. O sinal filtrado é usado como entrada para os blocos subseqüentes (blocos 1, 2, 4, 7 e 9). A freqüência de corte deve ser ajustada de modo que não seja maior que a freqüência fundamental do sinal de voz,caso contrário perderá informações relevantes.

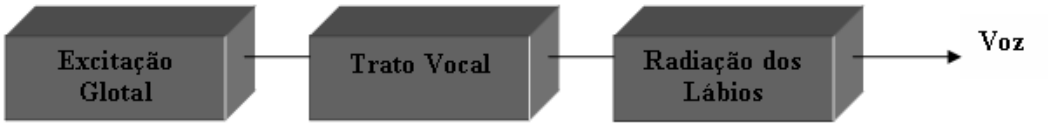

Figura 6.3: Modelo de produção da voz utilizado no método IAIF [34].

O método IAIF é baseado no prévio conhecimento da função de transferência do trato vocal. Logo, se todo o efeito da fonte glotal é eliminado do espectro da voz, o trato vocal pode ser estimado, mais precisamente, por análise LPC ou outro método de predição linear. A estimação da contribuição glotal e a função de transferência do trato vocal é computada pelo algoritmo IAIF em uma estrutura que se repete duas vezes. Inicialmente, a primeira estimativa da contribuição glotal é obtida do sinal de voz por análise LPC de ordem um e, posteriormente, eliminada por filtragem inversa. A ordem da análise LPC, neste caso, se for maior que um pode modelar os formantes, efeito indesejável por enquanto [34].

Um modelo preliminar do trato vocal é obtido aplicando análise LPC, de ordem

elevada (no nosso caso, a ordem usada foi quarenta e cinco), ao sinal do qual o efeito da contribuição glotal inicial foi eliminado. A primeira estimativa da excitação glotal é obtida cancelando o efeito do trato vocal e da radiação dos lábios, por filtragem inversa e integração, respectivamente.

O resultado desta primeira estrutura é o sinal glotal (excitação glotal ou contribuição glotal) que é usado como entrada da segunda estrutura a fim de estimá-lo de forma mais precisa. A Fig. 6.4 ilustra a diferença entre o sinal glotal obtido na primeira e na segunda estruturas.

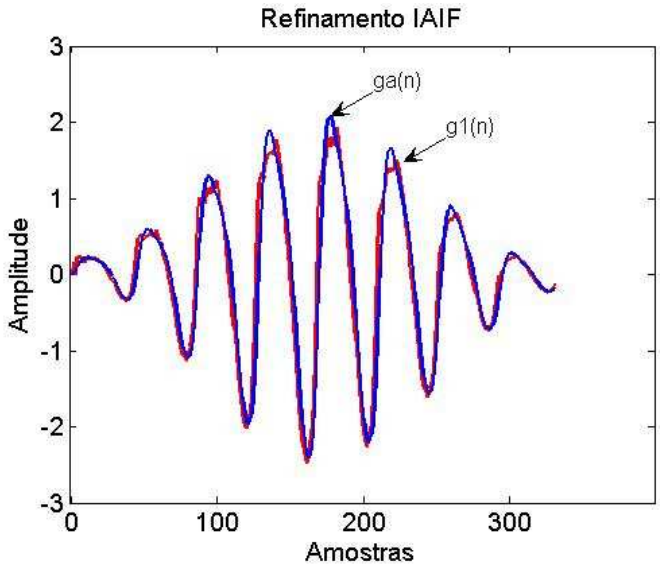

Figura 6.4: Refinamento na estimação do sinal glotal. O sinal glotal estimado pela primeira estrutura está representado por $g_1(n)$, obtido na saída do bloco 6. O sinal glotal estimado pela segunda estrutura está representado por $g_a(n)$, obtido na saída do bloco 10.

O espectro da excitação glotal é estimado no início da segunda estrutura usando análise LPC de ordem igual a dois ou quatro. Após cancelar a contribuição glotal, o modelo do trato vocal é encontrado, novamente usando análise LPC de

ordem elevada. O resultado final é obtido pela fitragem inversa do efeito do trato vocal e da radiação dos lábios do sinal original da voz [34].

A primeira estrutura do algoritmo contém os blocos de 1 a 5 e a segunda os blocos de 6 a 10.

O processamento pelo IAIF é feito em janelas de 30ms com 75 por cento de superposição. A Fig. 6.5 ilustra o diagrama do processamento do IAIF.

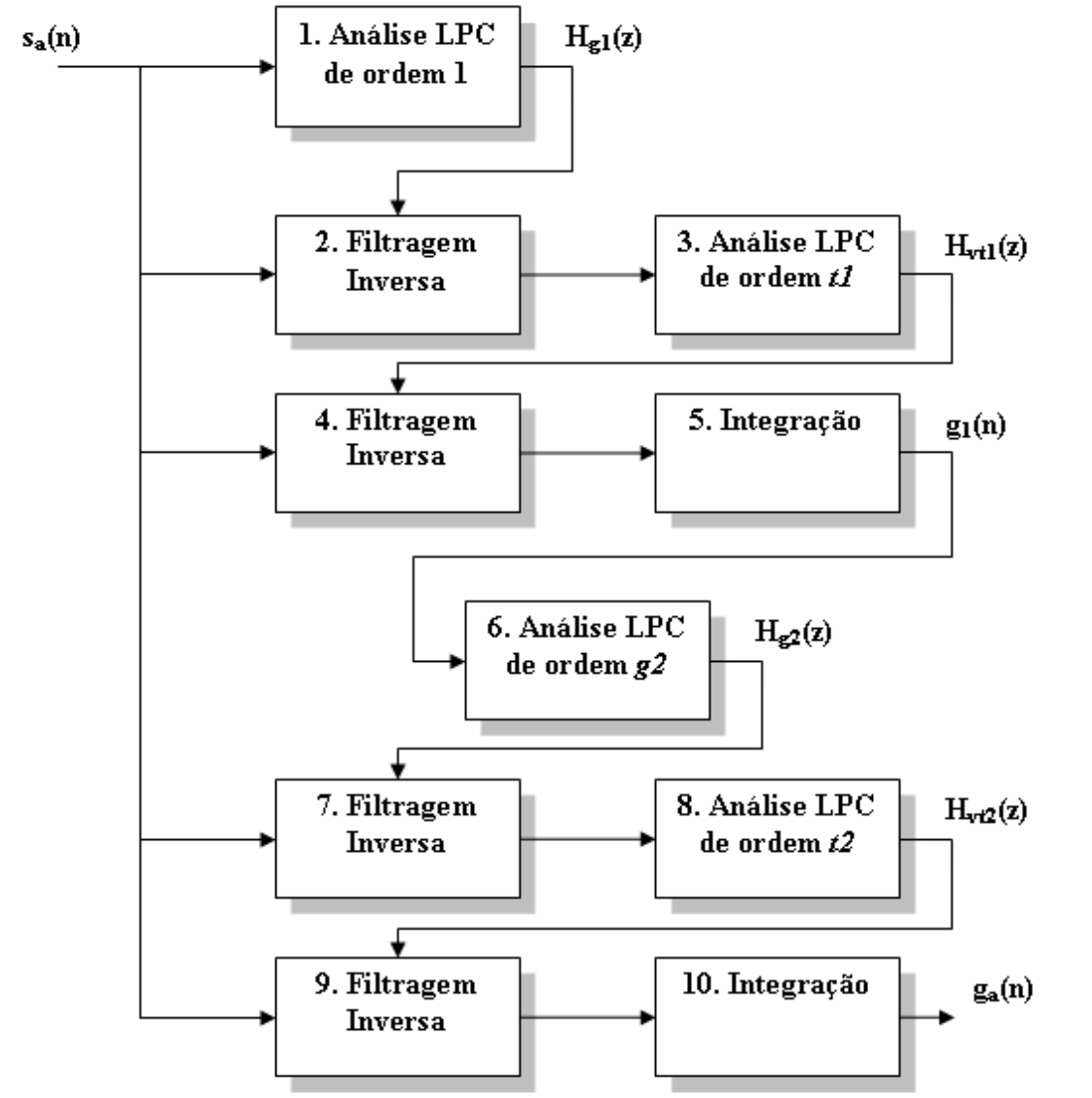

Figura 6.5: IAIF (*Iterative Adaptive Inverse Filtering*) [34].

Bloco 1. Análise LPC de primeira ordem - O efeito da contribuição glotal no espectro da voz é preliminarmente estimado pela análise LPC de ordem 1. A saída deste bloco é representada pela Eq. 6.10.

$$H(z) = 1 - az^{-1}, \qquad 0,9 \leq a \leq 1,0. \tag{6.10}$$

onde o valor de a é 0,98.

Bloco 2. Filtragem Inversa - A contribuição glotal é eliminada passando $s_a(n)$ por $H_{g1}(z)$.

Bloco 3. Análise LPC de ordem t_1- a primeira estimativa do trato vocal é obtida, aplicando análise LPC à saída do bloco anterior. A saída deste bloco é dada pela Eq. (6.11) (no caso, $t_1 = 45$).

$$H_{vt1}(z) = 1 + \sum_{k=0}^{t1} b(k)z^{-k}. \tag{6.11}$$

Bloco 4. Filtragem Inversa - o efeito do trato vocal é eliminado passando $s_a(n)$ por $H_{vt1}(z)$.

Bloco 5. Integração - a primeira estimativa para a excitação glotal, $g_1(n)$, é obtida pelo cancelamento do efeito da radiação dos lábios através da integração. Este bloco marca o final da primeira estrutura usada no IAIF. Sua saída servirá de entrada para o bloco seguinte, diferentemente dos blocos 1, 2, 4, 7 e 9, que possuem o sinal de voz como entrada.

Bloco 6. Análise LPC de ordem g_2- a segunda estrutura se inicia pela nova estimação do efeito da fonte no espectro da voz, porém a análise LPC tem sua ordem alterada para dois ou quatro. O sinal no qual a contribuição glotal é estimada é $g_1(n)$. A saída deste bloco é representada pela Eq. (6.12) (no caso, $g_2 = 4$).

$$H_{g2}(z) = 1 + \sum_{k=0}^{g2} c(k) z^{-k} \,. \tag{6.12}$$

Bloco 7. Filtragem Inversa- o efeito da contribuição glotal é eliminado, passando $s_a(n)$ através de $H_{g2}(z)$. Eliminando a contribuição glotal, no espectro do sinal de voz, é possível estimar o trato vocal de forma mais precisa no próximo bloco.

Bloco 8. Análise LPC de ordem t_2 - o modelo final do trato vocal é obtido, aplicando análise LPC de ordem t_2 à saída do bloco 7. O bloco 8 tem saída representada pela Eq. (6.13). ($t_2 = 45$)

$$H_{vt2}(z) = 1 + \sum_{k=0}^{t2} d(k) z^{-k} \,. \tag{6.13}$$

Bloco 9. Filtragem Inversa - o efeito do trato vocal é eliminado da voz, passando $s_a(n)$ através de $H_{vt2}(z)$.

Bloco 10. Integração - o resultado final do algoritmo ou **sinal glotal**, $g_a(n)$, é obtido pelo cancelamento do efeito da radiação dos lábios, integrando a saída do bloco 9.

6.4.2 PSIAIF

No método IAIF, a contribuição glotal no espectro da voz é inicialmente estimada por uma estrutura iterativa. A função de transferência do trato vocal é modelada após eliminar a contribuição glotal média. A excitação glotal é obtida cancelando os efeitos do trato vocal e da radiação dos lábios, por filtragem inversa e integração, respectivamente. No método PSIAIF (*Pitch Synchronous Iterative Adaptive inverse Filtering*), a forma do pulso glotal é obtida aplicando-se o algoritmo IAIF duas

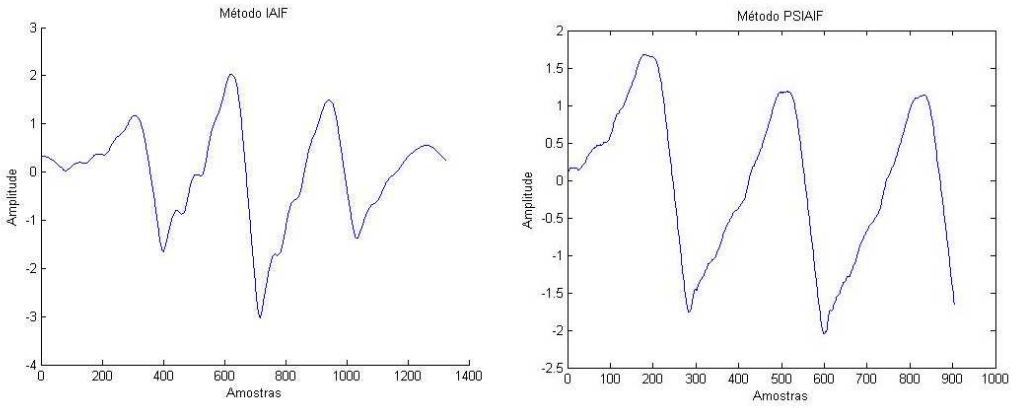

Figura 6.6: Vantagem na adoção do PSIAIF. Estimação mais precisa do sinal glotal, quando comparado ao método IAIF.

vezes, ao mesmo sinal, sendo o resultado da primeira aplicação servindo apenas para identificar o período fundamental que será a base para o cálculo do novo janelamento, antes da segunda aplicação do IAIF. Isto é ilustrado na Fig. 6.7

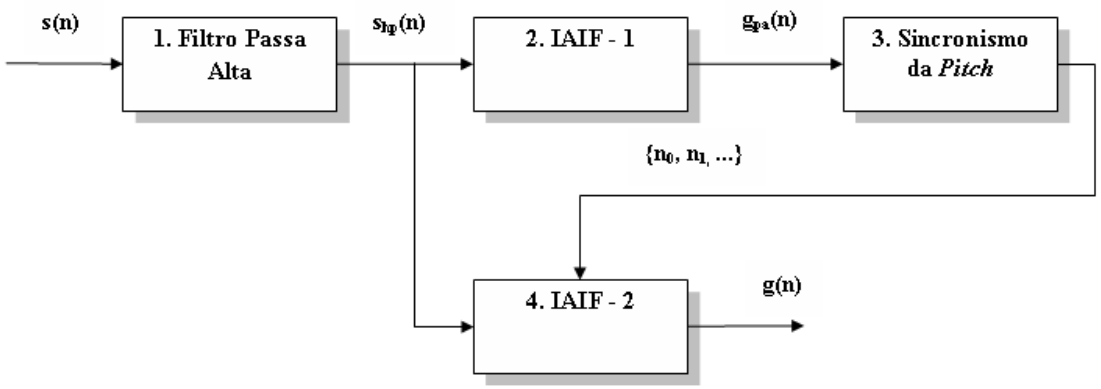

Figura 6.7: PSIAIF [34]

A primeira análise realizada pelo IAIF fornece o resultado da excitação glotal que ocorre entre vários períodos da *pitch* ($g_{pa}(n)$), que tem como entrada o sinal de voz previamente filtrado ($s_{hp}(n)$ - bloco 1 da Fig. 6.7). Este pulso é usado para determinar posições e larguras de janelas para uma análise síncrona da *pitch*. O

resultado final será obtido analisando o sinal de voz original com o algoritmo IAIF em um período por vez, ou seja, a estimativa final da forma do pulso glotal será obtida aplicando o método IAIF ao sinal de voz, usando o intervalo de tempo entre dois máximos de abertura glotal consecutivos (n_0, n_1, ...) [34]. Outros tamanhos de janela podem ser utilizados, mas sempre tendo como referência o período fundamental. Neste trabalho foram usados três períodos fundamentais consecutivos. A principal vantagem na utilização do método PSIAIF está na obtenção do sinal glotal de forma mais precisa. A Fig. 6.6 ilustra os resultados obtidos com os dois métodos, a partir do mesmo sinal de voz.

Capítulo 7

Resultados experimentais

7.1 Introdução

Apesar da similaridade da forma entre o sinal glotal (OFG) e o sinal eletroglotográfico (EGG), até o momento não foi demonstrada uma relação física entre ambos: o primeiro representa o fluxo de ar que atravessa a glote ao longo de tempo e o segundo a influência do movimento glotal durante a passagem de uma corrente elétrica pelo pescoço [10]. Isto é, ainda não foi desenvolvida uma técnica de se obter um através do outro.

Este capítulo visa detalhar as possíveis similaridades entre parâmetros de ambos os sinais,a fase experimental indispensável para a conclusão deste trabalho, o processo de gravação dos sinais EGG e de voz, e, finalmente, discutir a relevância de alguns parâmetros encontrados na identificação semi-automática de locutor.

A formação de um *corpus*, composto de sinais gravados em um eletroglotógrafo sincronizado com os sinais de voz correspondentes, gravados no laboratório do IME,

constitui um importante resultado desta dissertação que, sem dúvida, poderá servir de ponto inicial de outros estudos dedicados a aprimorar o conhecimento da voz, através da eletroglotografia.

7.2 Obtenção e processamento dos dados experimentais

Nesta dissertação, a parte experimental, pode ser, basicamente, dividida em: gravação das vozes e dos sinais EGG de doze locutores; obtenção dos sinais OFG, pelo método PSIAIF, a partir dos sinais de voz; seleção manual de 10 vogais de falas concatenadas e 05 vogais sustentadas de cada locutor, para comparação entre os parâmetros extraídos de sustentadas e concatenadas; extração de parâmetros dos sinais OFG, DOFG (primeira derivada do OFG), EGG e DEGG (primeira derivada do OFG) de vogais sustentadas e concatenadas; comparação dos sinais OFG e ODG e das suas respectivas derivadas.

7.2.1 A gravação dos sinais de voz e EGG

A gravação das vozes e dos sinais EGG foi realizada em uma câmara acústica, no Laboratório de Voz do Instituto Militar de Engenharia (IME). As frases e as vogais retiradas de falas concatenadas, foram escolhidas com o auxílio dos peritos criminais do Instituto de Criminalística Carlos Éboli (ICCE-RJ). Os sinais foram gravados com um taxa de amostragem de 44.100 Hz, com 16 bits de resolução e no formato PCM. Um dos primeiros obstáculos para a gravação dos sinais foi a obtenção de voluntários; isto se deu particularmente à necessidade de levá-los ao laboratório.

Por ser tratar de uma Organização Militar, predominantemente lotada por homens, houve também uma maior dificuldade na aquisição de dados de locutores do sexo feminino. O uso imperativo da câmara acústica, a fim de padronizar o ambiente de aquisição de dados e evitar o surgimento de ruídos, também inviabilizou a realização das gravações em outros locais. Porém, conseguimos, no total, a gravação de vozes de 11 pessoas, sendo 06 do sexo masculino e 05 do sexo feminino.

O eletroglotógrafo modelo EG2-PCX, fabricado pela *Glottal Enterprises*, foi o equipamento utilizado nas gravações.

Conforme o andamento da pesquisa avançava, outras informações foram colhidas, principalmente, no manual do equipamento e no site do fabricante, entre elas, a existência de outros modelos de eletroglotógrafo que estimam, além do VFCA (*Vocal Fold Contact Crea*) ou área de contato entre as cordas vocais, o qual o modelo do IME estima, o IVFCA (*Inverse Vocal Fold Contact Area*), a derivada do sinal EGG na polarização VFCA (conhecida como DEGG - *Differentiated EGG*) e o EXT LF (*Extended Low Frequency Response*), que permite a observação das componentes de baixa freqüência dos movimentos da laringe. Esta variedade de informações, provenientes do mesmo equipamento, tornou-se, no início, uma razoável dificuldade, pois os diversos artigos e teses lidos abordam cada uma dessas informações de forma variada.

A conexão entre o eletroglotógrafo e o computador foi realizada através de uma porta *USB*. O microfone que capta o sinal da voz é ligado diretamente ao eletroglotógrafo. O sinal EGG foi captado por dois eletrodos posicionados próximos às cordas vocais, sendo que o eletrodo posicionado no lado esquerdo do pescoço do locutor possuía fios na cor vermelha. O indicador no painel frontal do eletroglotógrafo

auxiliou no correto posicionamento dos eletrodos, parte fundamental no processo de gravação. Com todos os equipamentos ligados, foi iniciado o processo de gravação tendo os locutores lido as frases com breves pausas.

Após o encerramento das gravações o próximo passo foi a seleção, corte e exportação manual das vogais de interesse, ocorrida, simultaneamente, à seleção dos sinais EGG. Nesta etapa, foi fundamental o uso de um *software* de voz que permitisse a visualização de ambos os sinais e que possuísse um ambiente amigável que facilitasse a correta realização desta etapa. O *software* escolhido foi o *Audacity* [61]. Ao todo foram selecionados, cortados e exportados (ou salvos) manualmente, no formato "*.wav*", 380 (trezentos e oitenta) arquivos, divididos igualmente entre sinais de voz e sinais EGG.

7.3 A formação de uma nova base de dados

Durante a realização deste trabalho, foi identificada a escassez de base de dados de sinais EGG em português, resultando na formação de uma nova base de dados, complementar àquela inicialmente obtida, cujos detalhes podem ser obtidos na referência [62]. A artigo contendo as informações sobre esta nova base EGG intitulado "Incursionando pelos domínios da eletroglotografia: proposta de um *corpus* EGG", foi submetido e aceito para apresentação no XXVI Simpósio Brasileiro de Telecomunicações (SBrT2008), promovido pela Sociedade Brasileira de Telecomunicações (SBrT) [62].

A base de dados é constituída de sinais de voz e sinais do eletroglotógrafo (EGG) [62]. Os sinais EGG e de voz obtidos foram gravados com uma taxa de amostragem de 44.100 Hz, com 16 bits de resolução e no formato PCM. As gravações foram realizadas em uma câmara acústica no Laboratório de Voz do Instituto Militar de Engenharia (IME). A Figura 7.1 ilustra o processo de gravação.

Figura 7.1: Foto que simula o processo de gravação do sinal EGG (VFCA), realizado no Laboratório de Voz do IME.

A base de dados é formada por frases, sendo 10 balanceadas foneticamente

para o português falado no Rio de Janeiro [63], 5 palavras e 28 frases de interesse para a perícia forense, elaboradas com a colaboração dos peritos do Instituto de Criminalística Carlos Éboli (ICCE, RJ) e 8 vogais sustentadas [62]. Todas as frases, palavras e vogais sustentadas foram gravadas por 5 locutores do sexo masculino e 5 locutores do sexo feminino, com idades relacionadas na Tabela 7.1, com exceção das frases de interesse para perícia forense que foram gravadas por 12 locutores (6 homens e 6 mulheres). A referência [64] também disponibilizou uma base de sinais gravados com eletroglotógrafo, contendo 15 locutores, que pode ser encontrada no servidor ftp *ftp.cs.keele.dc.uk*.

A Tabela 7.1 representa a faixa etária dos locutores usados na formação da base de dados e a duração total aproximada das gravações.

Tabela 7.1: Faixa etária dos grupos de locutores e a duração total aproximada da gravação.

	Idade	Duração
Homens	20-38	75-105s
Mulheres	18-28	70-95s

Tabelas mostrando as vogais, palavras e frases de interesse, como indicados pelo Instituto de Criminalística Carlos Éboli (ICCE, RJ), estão no apêndice.

7.3.1 Obtenção do sinal OFG

Os sinais OFG foram obtidos, experimentalmente, dos 380 (trezentos e oitenta) sinais de voz pelo método de filtragem inversa conhecido como PSIAIF (*Pitch Synchronous Iterative Adaptive Inverse Filtering*), com o auxílio do MATLAB. Antes da aplicação do método, porém, os sinais de voz passaram por uma filtragem de pré-ênfase, para

evitar a instabilidade numérica e reduzir o efeito dos lábios. Após a pré-ênfase, uma etapa de janelamento é efetuada devido a necessidade de estacionariedade. Ademais, o janelamento (*hamming*) de pequenos segmentos minimiza os efeitos de borda.

O método PSIAIF se inicia com a estimação do período fundamental da voz com o intuito de analisar o sinal glotal de forma síncrona. Esta estimação foi possível aplicando o método IAIF (*Iterative Adaptive Inverse Filtering*), utilizando janelas de 30 ms com 75 por cento de superposição, que proporciona na saída o sinal OFG. A junção das janelas foi realizada pela técnica *Overlap and Add* [67]. Em seguida, o período fundamental, desta primeira estimação do sinal OFG, foi encontrado através dos picos de máximo do sinal e este resultado, usado como base para o novo tamanho de janela para uma segunda estimação do método IAIF, mais precisa. Os picos do sinal OFG e dos demais sinais foram encontrados pela rotina *findpeaks* [65]. Resumidamente, o método PSIAIF se baseia na aplicação do método IAIF duas vezes, sendo a primeira apenas para encontrar o período fundamental, que servirá como referência para o janelamento na segunda aplicação do método. Neste trabalho, o tamanho da janela escolhido foi de três períodos fundamentais.

Apesar de ser de rápida implementação, a análise LPC no MATLAB gerou alguns problemas, em função da ordem dos coeficientes adotada inicialmente. A referência [52] sugere que a ordem dos coeficientes LPC dos blocos do IAIF deve ser 10 (dez). Entretanto, as simulações com vozes de diversos locutores não apresentaram sinais glotais satisfatórios, ou seja, com muita distorção e forte influência do trato vocal, que, teoricamente, deveriam ser eliminados pela filtragem inversa que se sucede a cada bloco de análise LPC do algoritmo IAIF. Com o intuito de solucionar o problema foi a implementado um modelo de estimação de ordem automático con-

hecido como AIC [66], que calculou a melhor ordem dos coeficientes LPC para cada janela analisada, aumentado significativamente a precisão da estimação do trato vocal. Como conseqüência direta da adoção do AIC [66] o algoritmo se tornou consideralvelmente mais lento. A ordem de cada LPC das janelas processadas foi armazenada em um vetor, para possilitar sua posterior visualização, a fim de encontrar um valor fixo que atendesse satisfatoriamente o algoritmo. Dessa forma, foi possível concluir que 45 coeficientes eram suficientes para a obtenção de uma boa estimação do trato vocal, culminando com um sinal glotal mais próximo do esperado na saída do algoritmo. A Fig. 7.2 ilustra o sinal glotal com forte influência do trato vocal, obtido com 10 coeficientes LPC e o sinal glotal do mesmo locutor, porém obtido com 45 coeficientes. Portanto, a estimação do trato vocal realizada com apenas 10 coeficientes não cumpriu seu objetivo, afetando o resultado final (o sinal glotal).

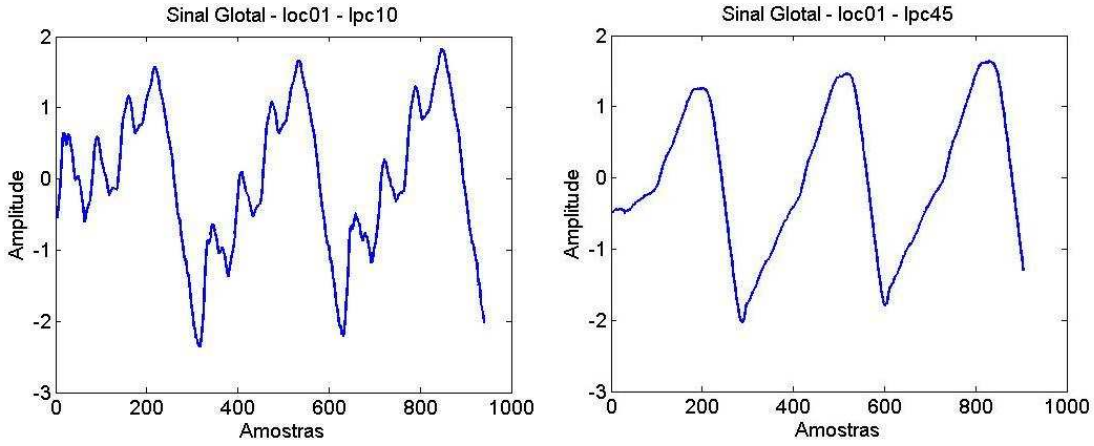

Figura 7.2: Diferença entre a estimação do sinal glotal de uma vogal sustentada */a/* com 10 e 45 coeficientes LPC.

7.3.2 Cálculo da freqüência fundamental usando a base de dados

Os picos de máximo do sinal DEGG foram utilizados para o cálculo da freqüência fundamental (*pitch*) do sinal de voz. Os resultados foram comparados com a técnica de extração da *pitch* amplamente conhecida como *fxrapt* [74]. O sinal de voz usado na simulação foi uma vogal sustentada */a/*. Os resultados das simulações mostraram que não houve diferenças significativas entre os valores de pitch calculados pelo *fxrapt* e pelos instantes de início de fechamento do sinal EGG (picos de máximo do sinal DEGG). As Figs. 7.3, 7.4, 7.5 e ilustram osresultados encontrados.

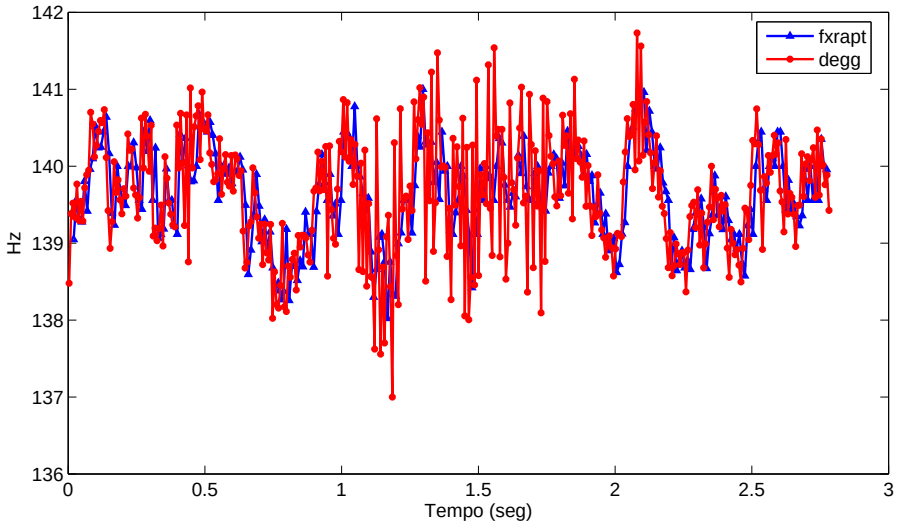

Figura 7.3: Freqüência fundamental extraída pelo DEGG e pelo algoritmo do *fxrapt*.

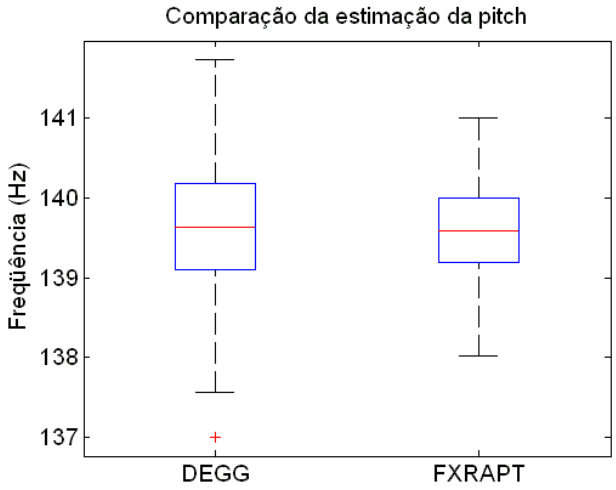

Figura 7.4: Boxplot dos resultados encontrados com DEGG e com o algoritmo *fxrapt*.

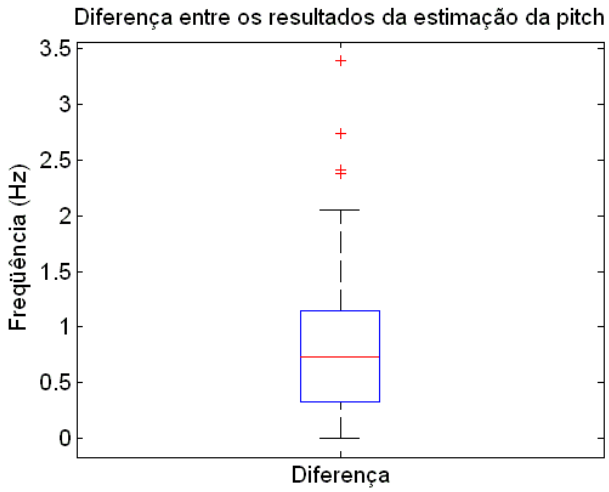

Figura 7.5: Diferença entre os valores estimados para a *pitch*.

Os valores de freqüência fundamental encontrados usando o sinal DEGG, foram similares aos calculados pelo *fxrapt*, ratificando o valor da base e suas aplicações.

7.3.3 Parâmetros obtidos e organização dos resultados

Após a obtenção dos sinais OFG e EGG, foram calculadas suas respectivas derivadas (DOFG e DEGG), pois, principalmente, a derivada do sinal EGG contém importantes informações sobre o estado físico das cordas vocais (instante de início de abertura e instante de início de fechamento). De posse destes sinais, doze (12) parâmetros de cada locutor foram encontrados, de acordo com o objetivo desta dissertação. A Tabela 7.2 contêm os parâmetros obtidos.

Tabela 7.2: Parâmetros obtidos.

	Descrição dos Parâmetros
Ke	Diferença entre os instantes de máxima abertura e de máximo fechamento
Avegg	Amplitude EGG
Kd1	Diferença entre o instante de início de abertura e o instante de início de fechamento
Kd2	Diferença entre o instante de início de fechamento e o instante de início de abertura
iabert	Instante de início de abertura
ifec	Instante de início de fechamento
Ko	Diferença entre os instantes de máximo fechamento e de máxima abertura
Avegg	Amplitude de vozeamento
Df	Diferença entre os instantes de máximo fechamento
Da	Diferença entre os instantes de máxima abertura
Keo	(Ke-Ko)
Dp	Diferença entre os picos de máximo do DOFG e do DEGG

Os doze (12) parâmetros foram obtidos das cinco (5) vogais sustentadas e das dez (10) vogais de falas concatenadas, extraídas dos locutores, totalizando 2160 (dois mil cento e sessenta) vetores com diferentes resultados armazenados. Para viabilizar a apresentação e a análise desses resultados foram efetuados mais de quinhentos (500) gráficos do tipo *boxplot* (rotina do MATLAB) contendo comparações dos resultados de cada parâmetro das vogais sustentadas em relação aos obtidos

com as vogais concatenadas de um mesmo locutor, comparação dos resultados dos parâmetros apenas entre as vogais sustentadas de um mesmo locutor e comparação dos resultados dos parâmetros de cada vogal sustentada entre locutores.

Os resultados foram salvos em estruturas do MATLAB separadas por locutor e parâmetro, conforme o exemplo da Tabela 7.3. As vogais de falas concatenadas foram analisadas de acordo com seu contexto fonético. A vogal */o/* da palavra "digo"foi considerada como */u/*.

Tabela 7.3: Exemplo de organização da estrutura *Keoloc12result.mat* que concentra os resultados encontrados do parâmetro *Keo* para o locutor 12.

	Vogal	**Parâmetro**
1	Vogal sustentada */a/*	Keo (*Keo* da vogal sustentada */a/*)
2	*Aaa* tá, amanhã eu ligo.	Keo a2 (*Keo* da vogal concatenada */a/*)
3	Alô, *a*lô, alô!	Keo a1 (*Keo* da vogal concatenada */a/*)
4	Vogal sustentada */e/*	Keo e (*Keo* da vogal sustentada */e/*)
5	Caf*é* ou açúcar?	Keo e1 (*Keo* da vogal concatenada */e/*)
6	Eu quero dez quilos de açúcar e dois quilos de caf*é*.	Keo e2 (*Keo* da vogal concatenada */e/*)
7	Vogal sustentada */i/*	Keo e (*Keo* da vogal sustentada */i/*)
8	Eu d*i*go alô baixinho.	Keo i1 (*Keo* da vogal concatenada */i/*)
9	Eu d*i*go parada baixinho.	Keo i2 (*Keo* da vogal concatenada */i/*)
10	Vogal sustentada */u/*	Keo u (*Keo* da vogal sustentada */u/*)
11	Eu dig*o* alô baixinho.	Keo u1 (*Keo* da vogal concatenada */u/*)
12	Eu dig*o* parada baixinho.	Keo u2 (*Keo* da vogal concatenada */u/*)
13	Café ou aç*ú*car?	Keo u3 (*Keo* da vogal concatenada */u/*)
14	Eu quero dez quilos de aç*ú*car e dois quilos de café.	Keo u4 (*Keo* da vogal concatenada */u/*)

7.3.4 Sincronismo

Para que o estudo da relação entre os sinais OFG e EGG fosse possível, a necessidade de sincronismo entre eles foi primordial. O método PSIAF, além de proporcionar uma estimação mais precisa do que o método IAIF, viabiliza o sincronismo entre os sinais, caracterizando mais uma vantagem para a sua adoção. A obtenção do sincronismo se inicia com o cálculo do período fundamental da voz, através do sinal OFG. Este período também pode ser estimado a partir do sinal EGG ou DEGG, uma vez que este possui os picos de máximo melhor definidos.

A partir do período fundamental foi escolhido o novo tamanho de janela que a repetição do IAIF usaria. Durante as simulações, foi observado que com três períodos fundamentais o algoritmo respondia de forma satisfatória sendo, então, adotado para todos os locutores. A Fig.7.6 ilustra os sinais de voz e OFG com três períodos fundamentais marcados.

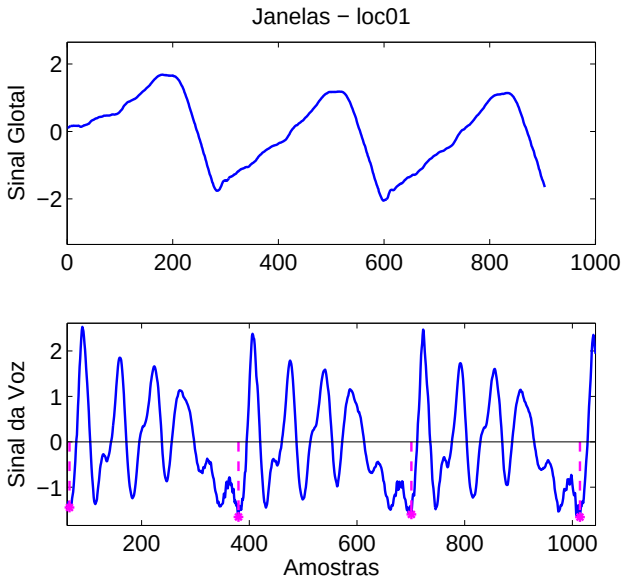

Figura 7.6: Janelamento efetuado com três períodos fundamentais.

Cabe ressaltar que os sinais EGG, DEGG e DOFG devem ser analisados com

este mesmo tamanho de janela. Os parâmetros escolhidos para mensurar o erro de sincronismo entre os sinais OFG e EGG foram *Df* e *Da*.

O parâmetro *Df* se apresentou praticamente estável entre todos os locutores sugerindo que os instantes de máximo fechamento podem ser obtidos tanto do sinal OFG quanto do sinal EGG. O erro de sincronismo entre os sinais foi estimado em torno de 1ms para todos os locutores. As Figs. 7.7 e 7.8 ilustram este resultado.

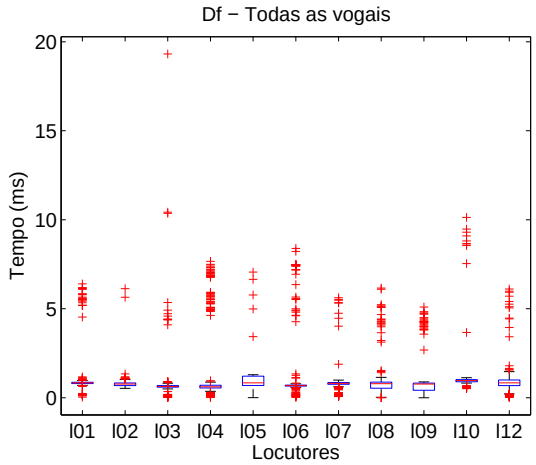

Figura 7.7: Exemplo de sincronismo, usando o parâmetro *Df*

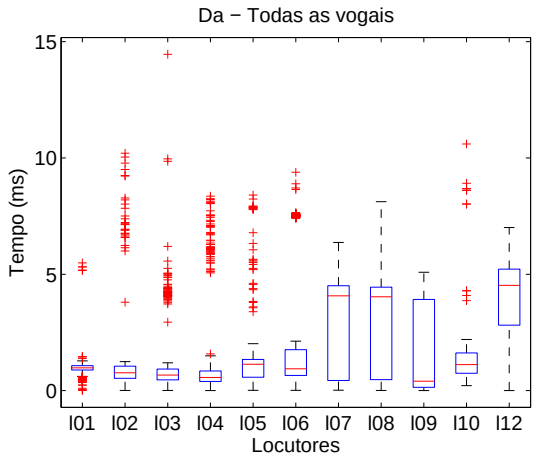

Figura 7.8: Exemplo de sincronismo, usando o parâmetro *Da*

O parâmetro *Df* teve sua estimação favorecida, quando comparada à estimação do *Da*, devido à facilidade de obtenção do picos de máximo no sinal EGG, menos ruidoso que o sinal OFG.

A Fig. 7.9 ilustra a comparação entre os parâmetros *Df* de cada locutor obtidos de vogais sustentadas.

Observamos que alguns dados, localizados em pontos isolados, são discrepantes, em relação aos demais. Como é o caso, por exemplo, do locutor 04.

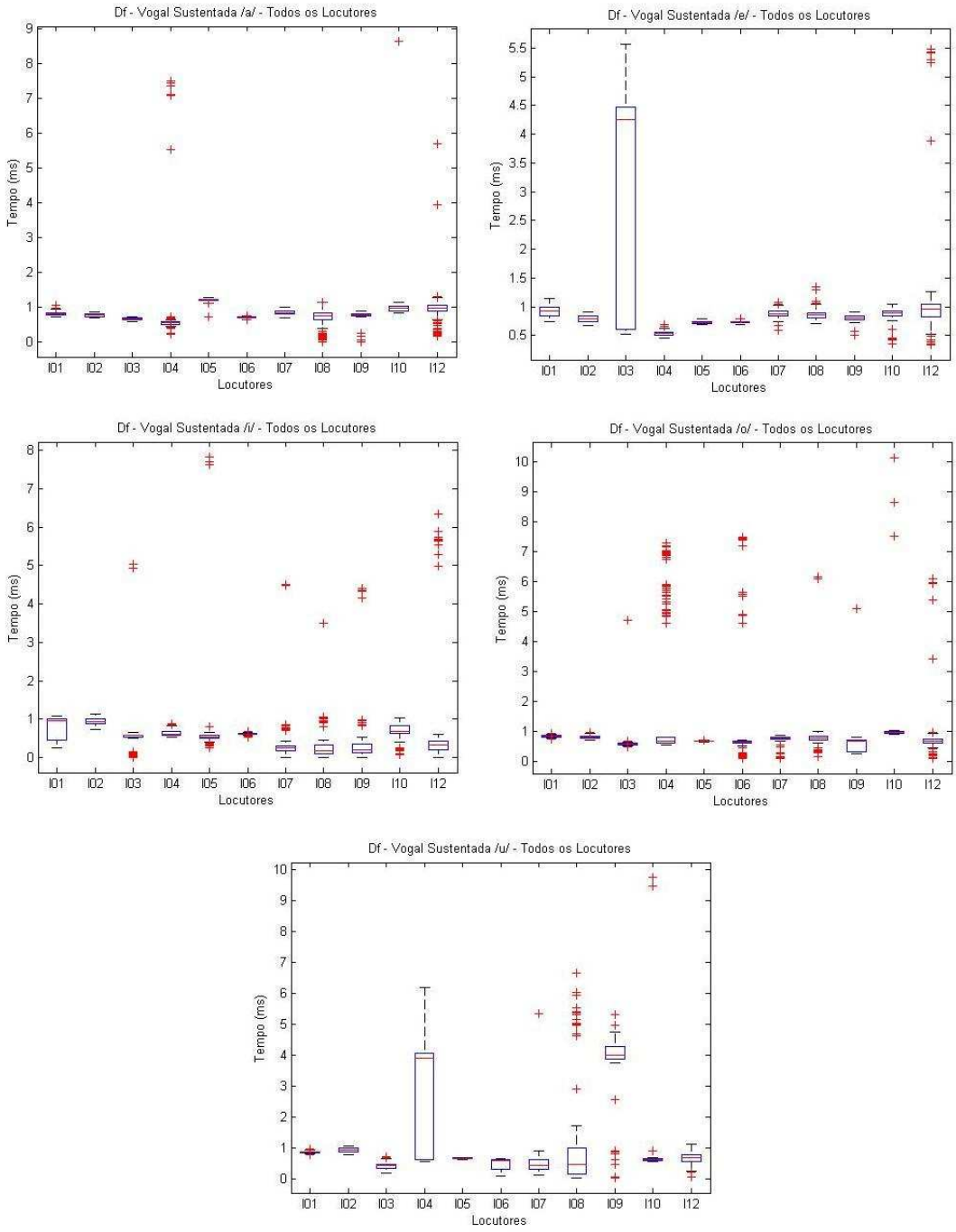

Figura 7.9: Comparação entre os parâmetros *Df* de cada locutor obtidos das vogais sustentadas */a/*, */e/*, */i/*, */o/* e */u/*.

7.4 Considerações importantes sobre a base de dados e para a perícia forense

Foram estudados doze locutores, seis do sexo masculino e seis do sexo feminino, sendo que o locutor 11, do sexo feminino, apresentou sinais EGG visivelmente diferentes quando comparados com os sinais dos demais locutores. Essas fortes diferenças, apesar de inviabilizarem o cálculo dos parâmetros com o algoritmo implementado, permitem segregar o locutor 11 dos demais analisados, uma vez que as distorções encontradas estão presentes em todos os sinais EGG. Se houvesse um banco de dados contendo sinais EGG previamente gravados, estas distorções constituiriam um importante indício para sistemas de identificação semi-automáticos, principalmente usados em perícias.

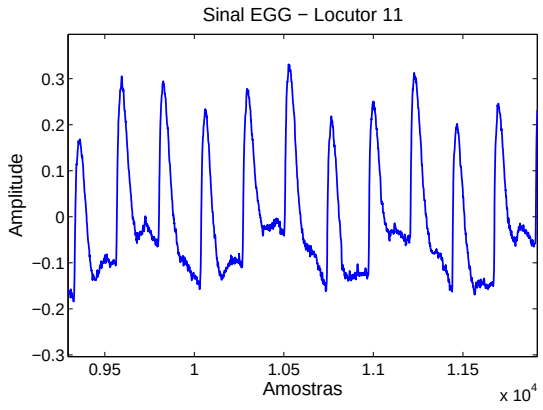

Figura 7.10: Sinal EGG distorcido de uma vogal sustentada */a/* do locutor 11

As distorções do sinal EGG também podem indicar a presença de patologia nas cordas vocais, logo, as gravações do locutor 11 foram descartadas para o cálculo dos parâmetros estudados neste trabalho. Um exemplo de sinal EGG distorcido do locutor 11 está ilustrado na Fig. 7.10.

O vocabulário de interesse para perícia forense, apesar de restrito, é extrema-

mente dinâmico. Com o passar do tempo, a polícia passa a conhecê-lo, compelindo os usuários a mudá-lo constantemente, a fim de dificultar o monitoramento de suas atividades ilícitas, principalmente através de escutas telefônicas, por exemplo. As frases estudas neste trabalho, portanto, e aquelas que compõem a base de dados, não esgotam o assunto, sendo apenasum ponto de partida para os que desejam concentrar esforços na área de perícia forense. As frases e as vogais concatenadas utilizadas (destacadas em *itálico*) estão relacionadas na Tabela 7.4.

Tabela 7.4: Frases e as respectivas vogais concatenadas utilizadas na obtenção dos parâmetros dos sinais OFG e EGG.

	Frases - Perícia Forense
1	*Aaa* tá, amanhã eu ligo.
2	Alô, *a*lô, alô!
3	Eu d*i*g*o* alô baixinho.
4	Eu d*i*g*o* parada baixinho.
5	Caf*é* ou aç*ú*car?
6	Eu quero dez quilos de aç*ú*car e dois quilos de caf*é*.

O contorno de *pitch*, denominado contorno melódico, é considerado como uma das características dos sinais de voz mais importantes para análise em Fonética Forense, por ser extremamente dependente do locutor [20]. O sinal DEGG permite o cálculo da *pitch*, conforme Fig. 7.3, a qual ilustra os valores de freqüência fundamental, encontrados usando o sinal DEGG, similares aos calculados pelo *fxrapt*, uma técnica de extração da *pitch* amplamente conhecida [74] .

7.4.1 Picos duplos no sinal DEGG

Durante as simulações realizadas, foram encontrados picos duplos no sinal DEGG de um locutor, dificultando a estimação precisa do instante de início de fechamento e do instante de início de abertura. Segundo a refêrencia [10], os picos duplos ocorrem devido à forma com que as cordas vocais se fecham e se abrem. Se as cordas vocais começarem a abrir pela parte posterior, haverá picos duplos de abertura (picos para baixo no DEGG). Caso as cordas vocais terminem de se fechar pela parte posterior, como se fosse um *zipper*, haverá picos duplos de fechamento (picos para cima no DEGG). Os picos duplos de fechamento e abertura estão ilustrados nas Figs. 7.11 e 7.12. É importante ressaltar que a forma como as cordas vocais se fecham e se abrem depende de cada locutor, portanto, assim como as distorções encontradas nos sinais EGG do locutor 11, este resultado constitui um importante indício para sistemas de identificação semi-automáticos usados em perícias de voz, com aplicação condicionada à existência de uma base de sinais EGG, gravadas previamente.

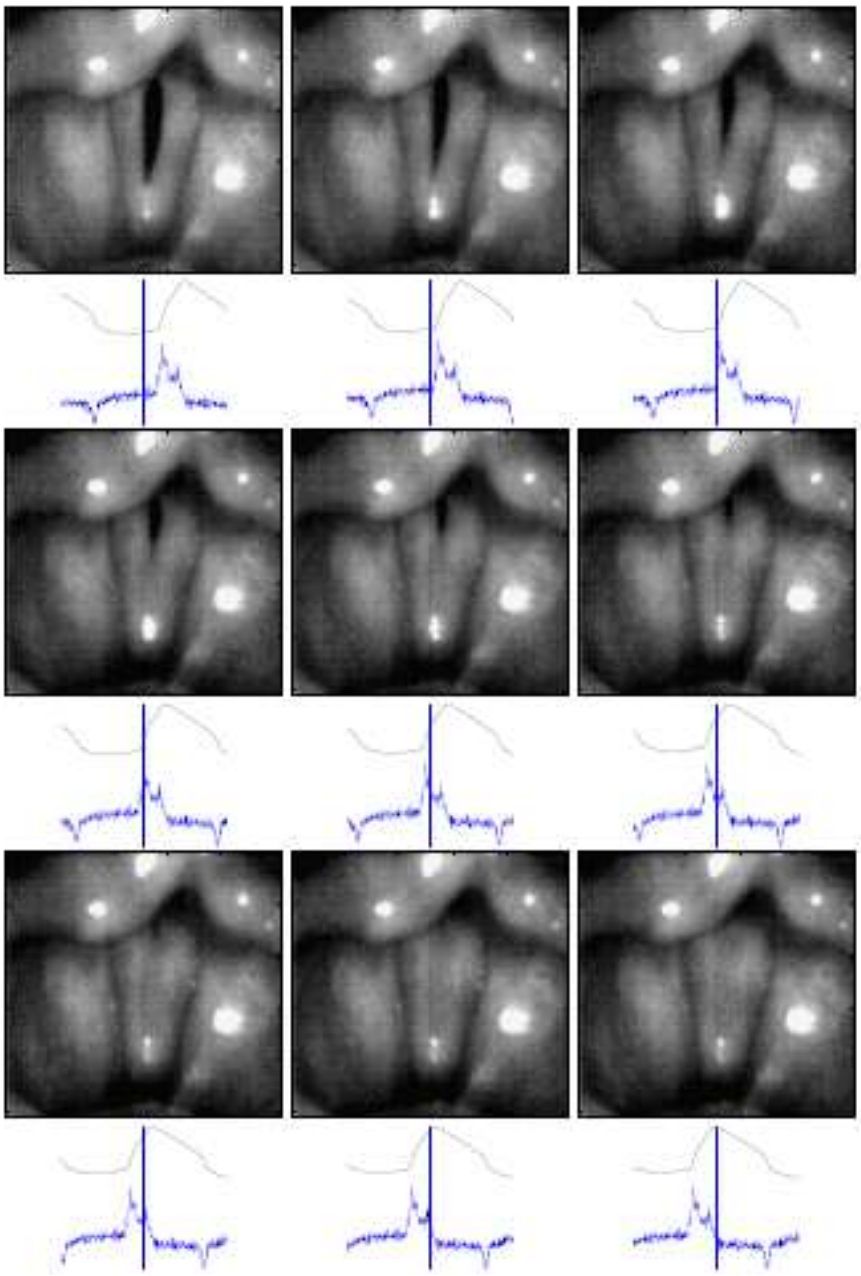

Figura 7.11: Visualização de picos duplos de fechamento por cinematografia ultrarápida e eletroglotografia simultâneas (locutor em fonação normal e freqüência fundamental igual a 110 Hz - sinais EGG e DEGG) [10].

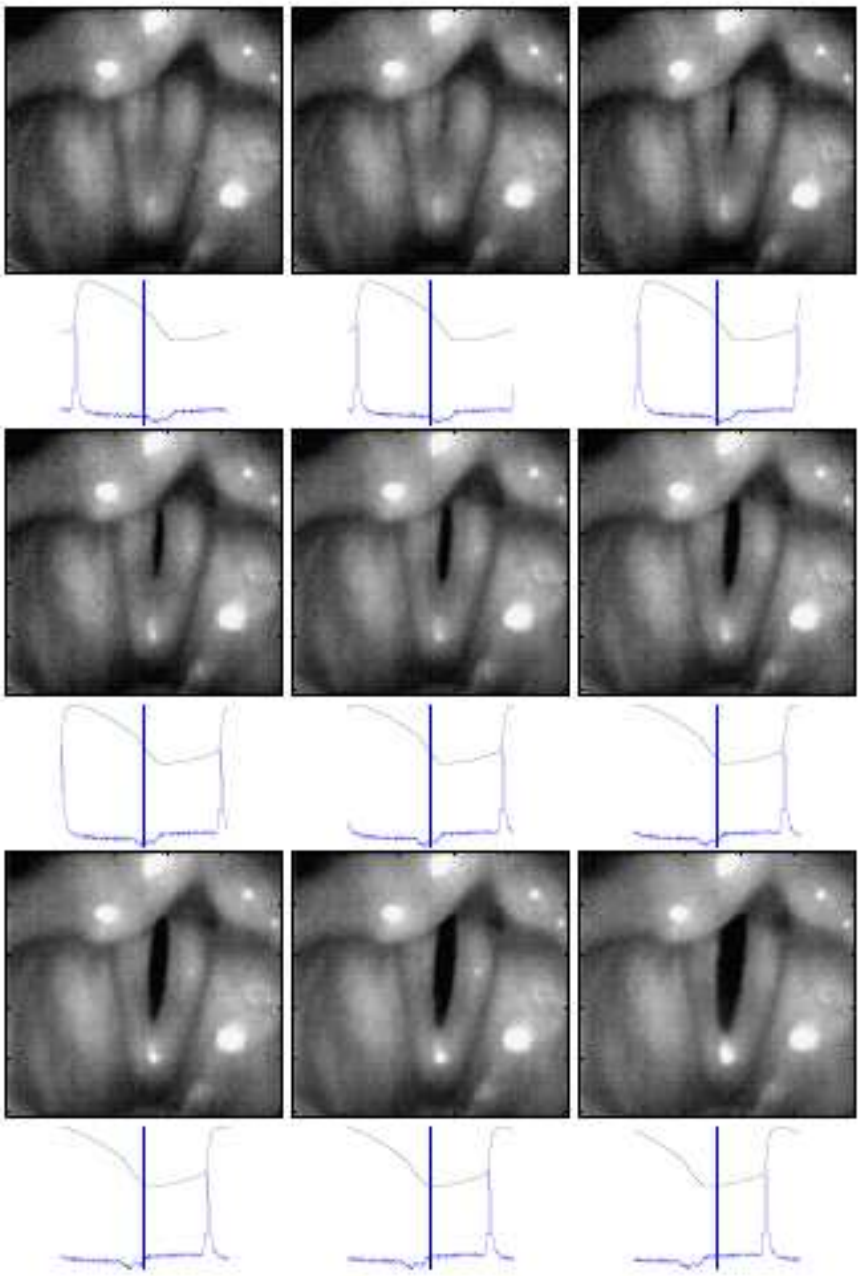

Figura 7.12: Visualização de picos duplos de abertura por cinematografia ultrarápida e eletroglotografia simultâneas (locutor em fonação normal e freqüência fundamental igual a 110 Hz - sinais EGG e DEGG) [10].

7.5 Comportamento dos parâmetros com vogais sustentadas e concatenadas

As vogais sustentadas e vogais retiradas de falas concatenadas foram estudadas, de acordo com seu contexto fonético, através de medidas realizadas nos respectivos sinais glotais e eletroglotográficos obtidos, verificando a existência de relação entre esses sinais, pela comparação entre as medidas. O objetivo é encontrar uma aplicação na identificação e na verificação de locutores, principalmente voltada para perícia forense.

A vogal */o/* da palavra "digo"foneticamente corresponde a vogal */u/* para o português falado no Rio de Janeiro, sendo assim considerada. No total foram duas vogais */a/*, */e/*, */i/* e quatro vogais */u/*. A vogal sustentada */o/* (pronunciada como */ó/*) não foi considerada no estudo, pois as frases no contexto da perícia, frequentemente, apresentam vogais foneticamente equivalentes a */o/*.

As Tabelas 7.5, 7.6, 7.7 e 7.8 ilustram alguns dos resultados obtidos comparando vogais sustentadas e concatenadas. As médias e variâncias encontradas indicam que não é possível usar vogais concatenadas em substituição às vogais sustentadas, pois os valores dos parâmetros variam muito entre elas, principalmente, devido à influência dos fonemas próximos às vogais retiradas de falas concatenadas. Todas as Tabelas estudadas podem ser encontradas no Apêndice.

Tabela 7.5: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 03 - vogal /a/.

	Vogal /a/			
Locutor 03	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.1879	0.7426	2.6952	0.5824
Avegg	1.5870	0.0107	1.8674	0.0515
Ko	2.0141	0.1190	5.5568	83.4870
Ke	1.9928	0.0045	3.6082	6.1441
Kd1	1.2425	0.0798	7.1458	120.0100
Kd2	4.0058	0.1479	2.3394	6.3945
Da	0.6183	0.1006	3.0591	14.8920
Dp	0.3274	0.4639	3.4951	52.6860
Df	0.6612	0.0008	2.4285	20.0760
Keo	0.3041	0.0381	5.3483	130.8100

Tabela 7.6: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 04 - vogal /a/.

	Vogal /a/			
Locutor 04	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.7177	0.8209	1.9196	0.4231
Avegg	13.8130	0.0369	13.3170	0.3150
Ko	1.9401	0.1138	NaN	NaN
Ke	4.0526	0.0232	4.6423	0.1581
Kd1	4.0626	0.0248	4.9044	0.0472
Kd2	4.0286	0.0175	3.8300	0.0684
Da	0.9388	0.4026	1.4358	0.9711
Dp	0.5258	1.3785	0.7396	4.9077
Df	1.2039	0.0037	NaN	NaN
Keo	2.1125	0.1585	4.8593	28.6400

Tabela 7.7: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 06 - vogal /a/.

	Vogal /a/			
Locutor 06	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.5453	0.2400	3.1409	0.4047
Avegg	6.9642	0.0107	5.7777	0.2017
Ko	1.8594	0.1116	1.7843	0.1200
Ke	4.9310	0.0032	4.4248	0.0692
Kd1	3.0394	0.1792	2.7827	0.7157
Kd2	4.5859	0.1806	4.8770	1.2627
Da	5.0464	11.2030	0.6080	0.0851
Dp	0.9161	4.7219	2.1071	11.5820
Df	0.6993	0.0004	0.6952	0.0004
Keo	3.0716	0.1236	2.6406	0.0431

Tabela 7.8: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 03 - vogal /e/.

	Vogal /e/			
Locutor 03	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.2924	0.3698	2.8153	0.3568
Avegg	2.2424	0.0079	1.7222	0.4314
Ko	2.3782	0.7203	1.7395	0.3977
Ke	2.2811	0.0154	2.3451	0.0285
Kd1	1.9185	0.0109	1.7911	0.0666
Kd2	3.4831	0.0093	3.0887	0.5739
Da	1.0019	0.0296	1.8614	4.4614
Dp	0.8430	1.0031	1.4487	1.6480
Df	3.0606	3.9439	0.7055	0.0021
Keo	0.8275	0.0337	0.8332	0.1393

A partir da análise das tabelas de média e variância das vogais sustentadas e concatenadas e do *boxplot* dos parâmetros *Ko* e *Ke*, obtidos do sinal OFG e EGG, respectivamente, foi observado, uma tendência de *Ke* variar consideravelmente entre locutores, se comparado a *Ko*. As Figs. 7.13 e 7.14 ilustram os resultados obtidos. A estimação do parâmetro *Ke* foi obtida com precisão, pois a forma de onda do sinal EGG tem comportamento bem definido e livre de ruídos que dificultem sua estimação, logo, as variações nos resultados de *Ke* não podem ser atribuídas a deficiências do algoritmo e sim, a características de fechamento e abertura das cordas vocais de cada locutor. Lembrando que este parâmetro estima o intervalo entre os instantes de máxima abertura e máximo fechamento do sinal EGG e o parâmetro *Ko* estima o intervalo entre os instantes de máximo fechamento e máxima abertura no sinal OFG. Portanto, o parâmetro *Ke* possui características relevantes para a

discriminação de locutores, como ilustrado na Fig. 7.13.

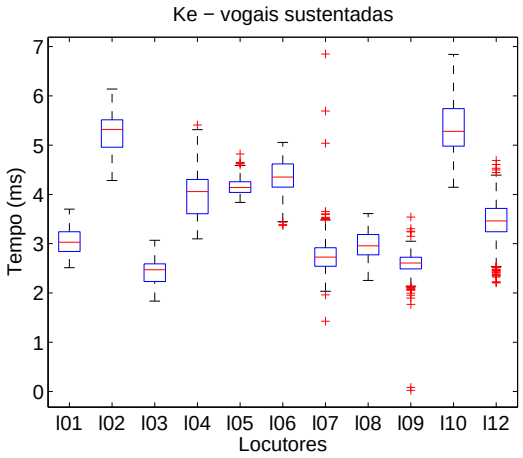

Figura 7.13: Visualização do *boxplot* do parâmetro *Ke* para todos os locutores. Este gráfico foi obtido unindo os resultados das vogais sustentadas.

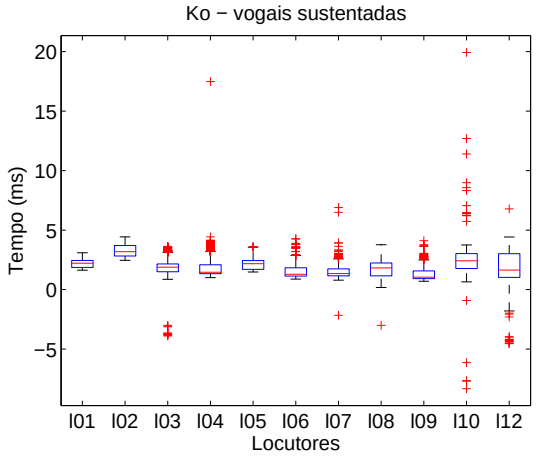

Figura 7.14: Visualização do *boxplot* do parâmetro *Ko* para todos os locutores. Este gráfico foi obtido unindo os resultados das vogais sustentadas.

7.6 Perspectivas para reconhecimento de locutor

Com os resultados obtidos, foram confeccionados os diversos gráficos dos parâmetros, tais como: *Ke versus Avegg*, *Ke versus Kd1*, *Ke versus Kd2*, *Ko versus Av*, *Kd1 versus Kd2*, *Avegg versus Kd1* e *Avegg versus Kd2*. O objetivo era detectar uma possível discriminação visual entre locutores, sendo os melhores resultados obtidos com os gráficos *Ke versus Avegg* e *Ke versus Kd1*. As Figs. 7.15, 7.16, 7.17, ilustram alguns dos resultados obtidos.

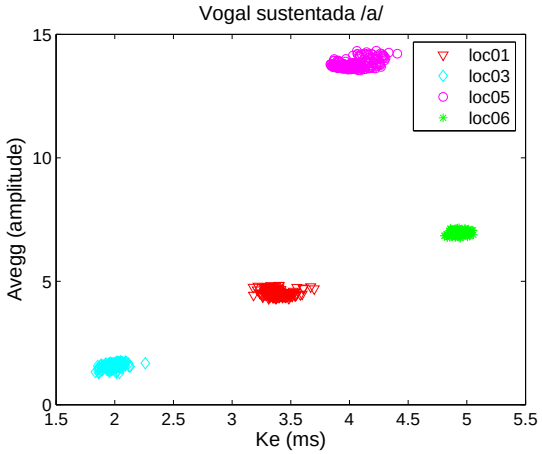

Figura 7.15: Discriminação visual entre os locutores 01, 03, 05 e 06, utilizando a vogal sustentada */a/*

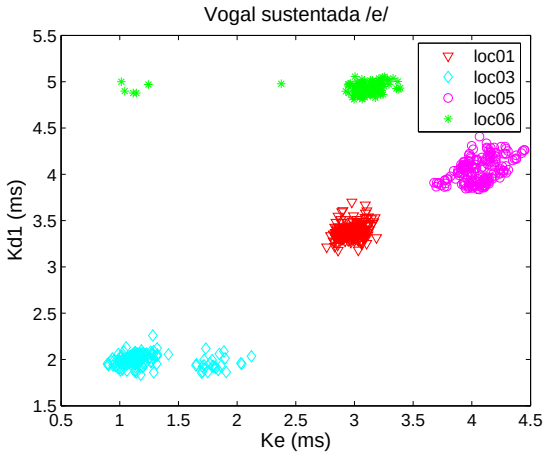

Figura 7.16: Discriminação visual entre os locutores 01, 03, 05 e 06, utilizando a vogal sustentada */e/*

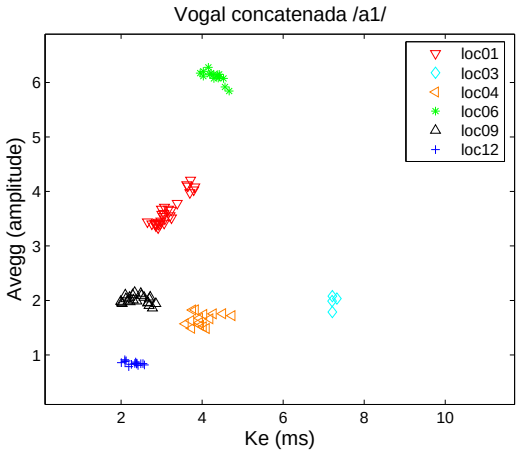

Figura 7.17: Discriminação visual entre os locutores 01, 03, 04, 06, 09 e 12, utilizando a vogal concatenada */a1/*

Apesar da boa discriminação exibida, esses gráficos não podem ser considerados como um resultado final. Estes parâmetros devem ser testados para um grande número de locutores e seu resultado, validado por métodos estatísticos.

7.7 Comparação dos sinais OFG e EGG

O principal objetivo dessa dissertação é a comparação entre os sinais OFG e EGG. Para tal, foram escolhidos os seguintes parâmetros: *Df, Da, Dp e Keo.* Estes parâmetos indicam que, se os resultados obtidos em ambos forem similares, haverá uma grande possibilidade de dispensar o eletroglotógrafo para a obtenção dos parâmetros citados. É evidente que, para se afirmar isto, um estudo mais abrangente deve ser realizado, entretanto, este trabalho pode ser considerado um marco inicial para aqueles que desejarem estudar as similaridades entre estes sinais. A vantagem obtida em dispensar o uso de eletroglotógrafo é clara, dada a natureza e forma de aquisição dos sinais. O sinal eletroglotográfico, conforme detalhado anteriormente, é um método extremamente dependente do locutor, pois a aquisição do sinal EGG e do sinal da voz ocorre mediante eletrodos presos ao pescoço e um microfone, devidamente posicionados. O sinal OFG é obtido, experimentalmente, por um método de filtragem inversa que necessita exclusivamente do sinal de voz do locutor, logo, não há a necessidade da presença ou autorização direta do locutor para a aquisição do sinal. Dessa forma, a obtenção de parâmetros do sinal OFG se torna interessante, principalmente para perícia forense, cujos métodos são baseados na análise do sinal da voz dos envolvidos, lembrando que o sinal EGG provê informações importantes a respeito do estado físico das cordas vocais; abertas, fechadas, fechando ou abrindo e até mesmo a forma como o fechamento ou abertura ocorrem (picos duplos).

O parâmetro *Df* teve sua estimação favorecida, quando comparada à estimação do *Da*, devido à facilidade de obtenção do picos de máximo no sinal EGG, menos ruidoso que o sinal OFG.

As Fig. 7.18 e 7.19 ilustram a comparação entre os parâmetros *Df* e *Da* de

cada locutor, obtidos de todas as vogais sustentadas e concatenadas.

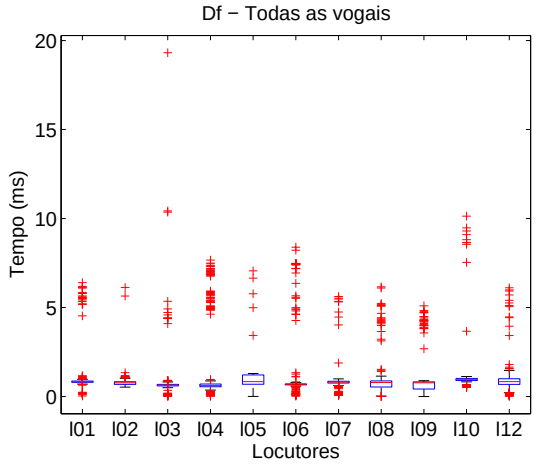

Figura 7.18: Comparação entre os parâmetros *Df* de cada locutor obtidos de todas as vogais sustentadas e concatenadas.

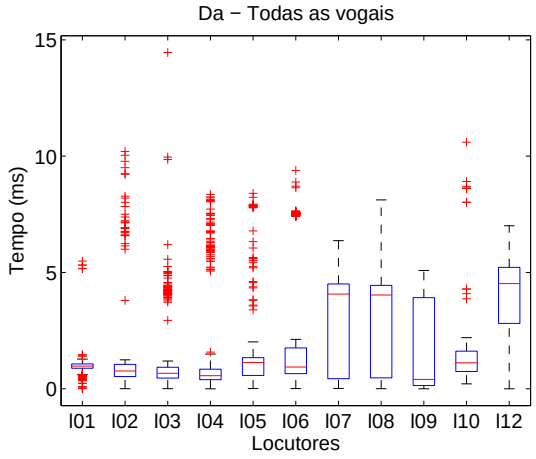

Figura 7.19: Comparação entre os parâmetros *Da* de cada locutor obtidos de todas as vogais sustentadas e concatenadas.

A vogal sustentada */o/* de todos os locutores foi inserida no cálculo destes quatro parâmetros de comparação, mesmo não havendo vogais retiradas de falas concatenadas, que dentro do contexto fonético analisado, correspondessem a essa vogal. O parâmetro *Df* apresentou erro de sincronismo, entre os sinais, em torno de *1ms* para todos os locutores, podendo ser considerado praticamente estável. Este resultado indica que os instantes de máximo fechamento podem ser obtidos tanto do sinal OFG quanto do sinal EGG. O parâmetro *Da* apresentou resultado similar ao *Df*, exceto para os locutores 7, 8, 9 e 12, pois a forma obtida do sinal glotal destes não favoreceu a estimação dos picos de máximo.

Analisando os resultados do parâmetros *Keo*, foi constatada uma tendência similar à ocorrida com o parâmetro *Df*, de permanecer próximo a um determinado valor (aproximadamente *1,5ms*) independente do locutor, conforme Fig.7.20.

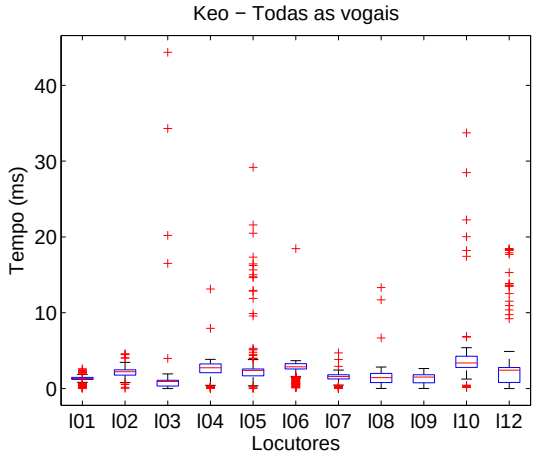

Figura 7.20: Comparação entre os parâmetros *Keo* de cada locutor obtidos de todas as vogais sustentadas e concatenadas

Este parâmetro representa a diferença entre os intervalos de ocorrência dos instantes de máximo entre o sinal EGG e o sinal OFG. O resultado obtido graficamente indica que os sinais têm comportamento similar ao longo do tempo. É mais

um resultado que ratifica a hipótese de que os instantes de máximo de fechamento e máxima abertura podem ser igualmente obtidos em ambos os sinais. As Figs. 7.21 e 7.22 ilustram os parâmetros *Keo* e *Df*, obtidos de vogais sustentadas.

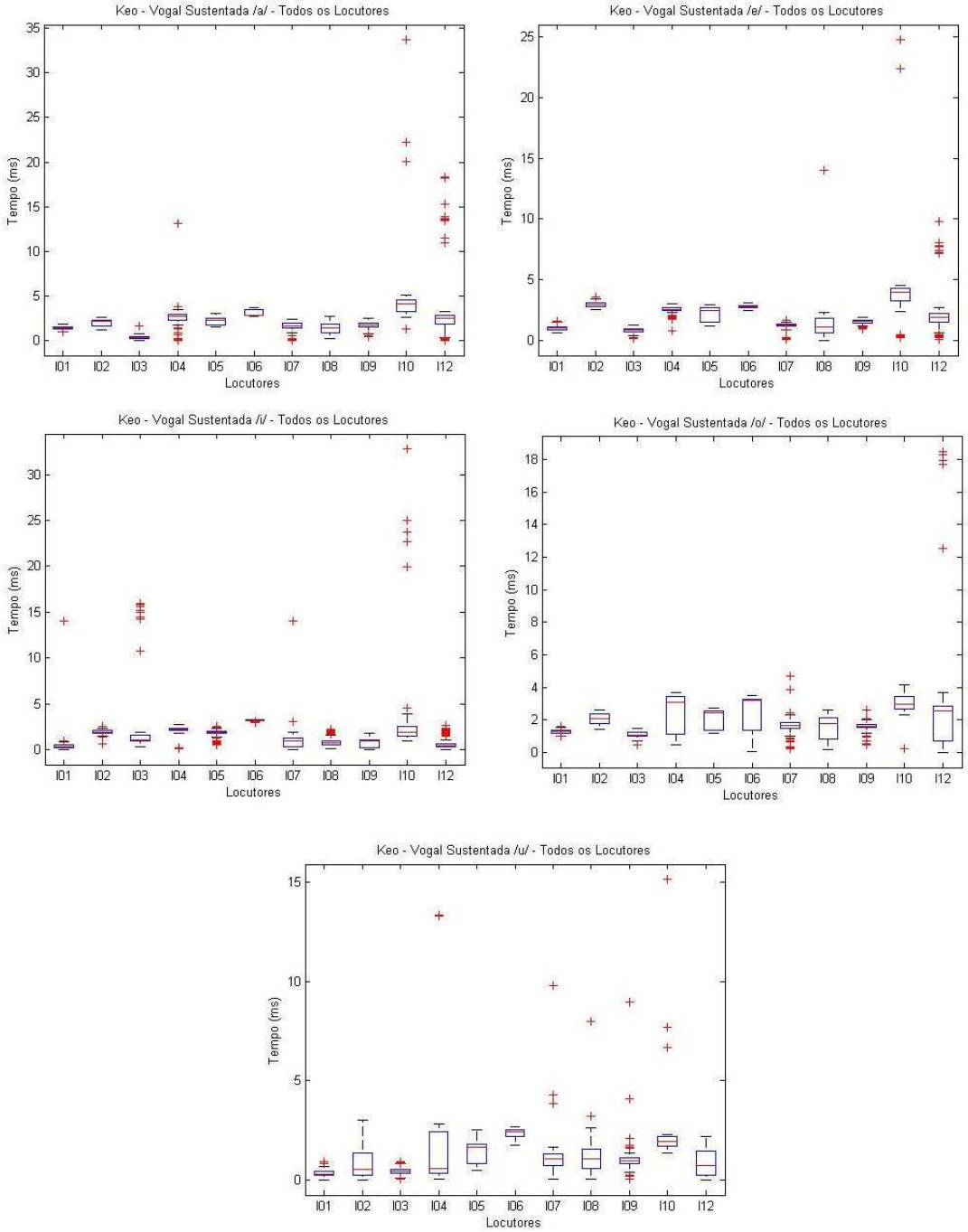

Figura 7.21: Comparação entre os parâmetros *Keo* de cada locutor, obtidos das vogais sustentadas */a/*, */e/*, */i/*, */o/* e */u/*.

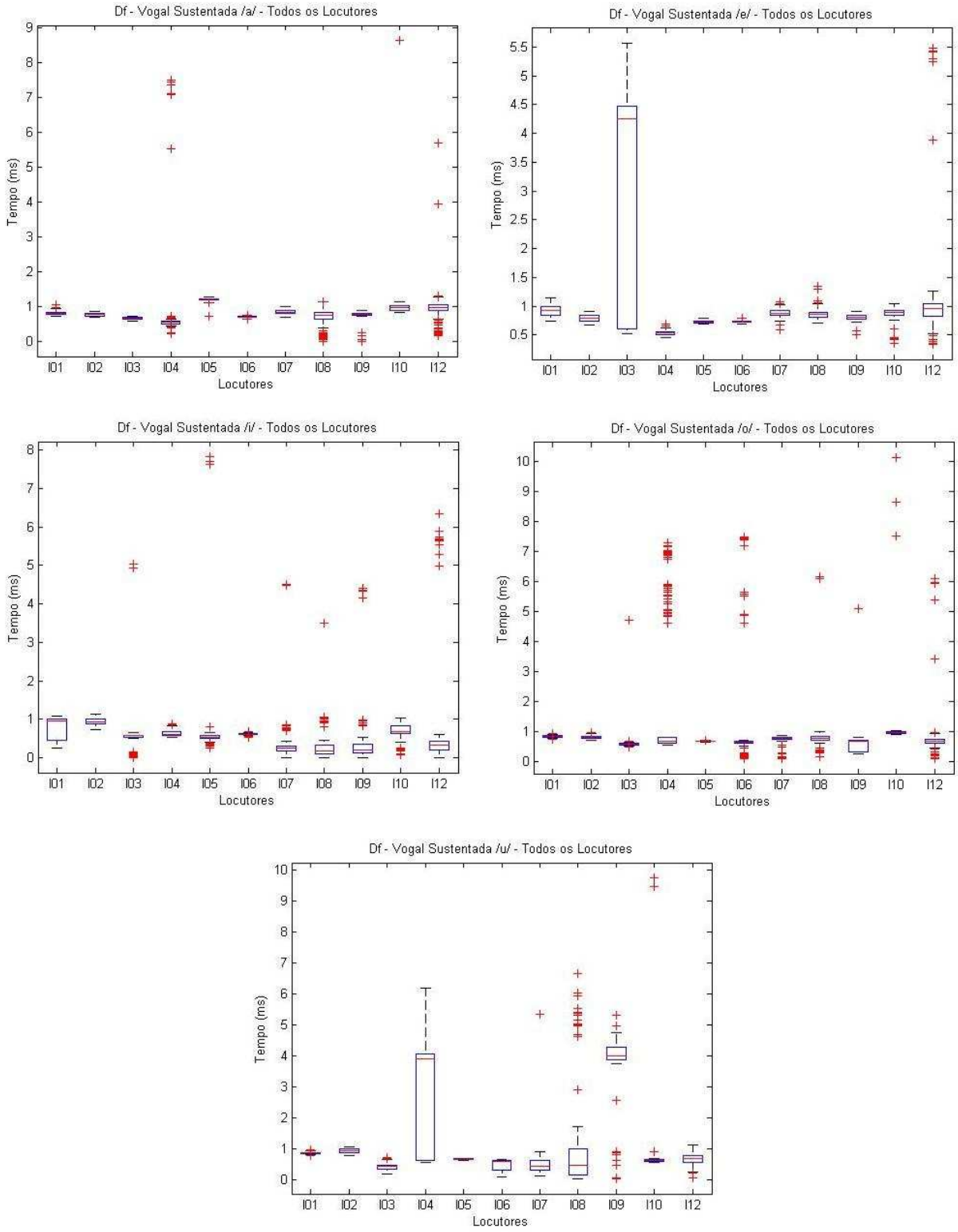

Figura 7.22: Comparação entre os parâmetros *Df* de cada locutor, obtidos das vogais sustentadas */a/*, */e/*, */i/*, */o/* e */u/*.

A derivada do sinal glotal (DOFG) não apresentou uma forma de onda que favorecesse a comparação dos picos de máximo deste sinal com os picos de máximo do sinal DEGG (instante de início de fechamento), principalmente nos sinais OFG dos locutores 1, 2, 4 e 10. A Fig. 7.23 ilustra o resultado obtido.

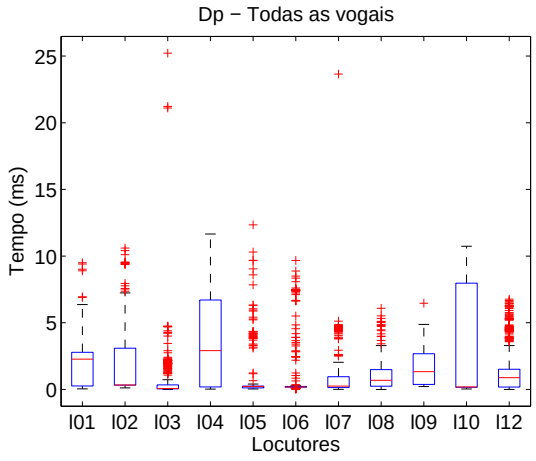

Figura 7.23: Comparação entre os parâmetros *Dp* de cada locutor obtidos de todas as vogais sustentadas e concatenadas

Apesar de apresentarem similaridades, uma filtragem prévia do sinal DOFG se faz necessária para amortecer sua forma de onda, possibilitando uma estimação mais precisa dos seus picos, conseqüentemente, melhorando a estimação do parâmetro *Dp*. Este parâmetro indica que o instante de início de fechamento, encontrado a partir do sinal DEGG, pode ser obtido do sinal DOFG, indicando que é possível abdicar de um método extremamente dependente da presença e autorização do locutor para aquisição de seus sinais, por outro que necessita apenas do sinal da voz, obtido através de uma escuta telefônica, por exemplo. Esta característica também é muito importante para perícia forense, pois independe de um banco de dados EGG. Ademais, caso o parâmetro *Dp* seja usado, com sucesso, em sistemas de reconhecimento de locutor, a sua importância se torna ainda maior. Embora tenha sido observado

o surgimento de picos duplos no sinal DEGG o mesmo não pode ser observado no sinal DOFG. O parâmetro *Dp* foi encontrado em torno de 0,8 e 1,5 ms.

Capítulo 8

Conclusões e trabalhos futuros

8.1 Conclusões

1. Neste trabalho, foi proposto um estudo sobre o sinal de voz e o sinal do eletroglotógrafo, visando mensurar suas semelhanças e estimar determinados parâmetros que, hipoteticamente, poderiam ser obtidos de um sinal ou do outro. A vantagem de se obter os mesmos parâmetros do sinal EGG, a partir do sinal de voz está concentrada na facilidade de se obter gravações de vozes de diversos locutores, se comparadas às gravações efetuadas com o eletroglotógrafo que, além de ser um equipamento caro e de difícil aquisição, requer o uso de eletrodos presos ao pescoço durante o processo de obtenção do sinal. O sinal de voz foi filtrado inversamente através da técnica conhecida como PSIAIF, com o intuito de estimar o sinal glotal, ponto inicial do estudo. O método PSIAIF apresentou resultados compatíveis com o esperado, mostrando-se uma ferramenta eficaz para a estimação do sinal glotal.

2. A estimação do sinal glotal apresenta resultados mais precisos quando imple-

mentado um modelo de estimação de ordem automático, para as etapas de análise LPC, contidas no método IAIF. Dessa forma, os janelas são analisadas individualmente, sendo escolhida a ordem dos coeficientes LPC que proprorciona o melhor resultado. Entretanto, a perda da velocidade de processamento é inevitável.

3. O estudo comparativo entre o sinal EGG e o sinal de voz evidenciaram que parâmetros como o instante de máximo fechamento, máxima abertura e instante de início de fechamento podem ser obtidos tanto do sinal glotal quanto do sinal EGG com um pequeno erro de sincronismo, próximo a *1ms*, que, dependendo da aplicação, pode ser minimizado. O mesmo não pode ser dito, entretanto, do instante de início de abertura estimado a partir do sinal DEGG; isto se deve ao fato da derivada do sinal glotal (DOFG) ter se apresentado extremamente ruidosa, dificultando a estimação dos picos que identificam os referidos instantes.

4. O surgimento de picos duplos no sinal DEGG de um locutor caracterizam a forma como as cordas vocais do locutor se fecham e se abrem. Nesse caso, constitui um forte indício na identificação semi-automática de locutores, pois a forma de fechamento e abertura das cordas vocais está intimamente ligada a cada locutor. No contexto da perícia forense, o número de locutores analisados é, em geral, em torno de dois ou três, pois as gravações de vozes são obtidas de escutas telefônicas. Considerando este pequeno espaço amostral e caso o acesso aos investigados fosse possível, a detecção de picos duplos no sinal DEGG poderia constituir mais um parâmetro de segregação ou identificação de locutores. As distorções do locutor 11 possuem igual valor para segregação

ou identificação de locutores.

5. A base de dados em português, formada por sinais gravados com o eletroglotógrafo (sinais EGG), representa uma importante ferramenta inicial para aqueles que desejam estudar o sinal da voz. A derivada do sinal EGG se mostrou uma importante informação sobre o estado físico das cordas vocais através dos instan-

   tes de início de fechamento e início de abertura, facilmente identificados pelos picos que o sinal DEGG possui.

6. Os picos de máximo do sinal DEGG, bem definidos na forma de onda e que representam os instantes de início de fechamento das cordas vocais, podem ser utilizados para o cálculo da freqüência fundamental da voz (*pitch*), como ilustrado no Capítulo 7. Os resultados foram comparados com a técnica de extração da *pitch* conhecida como *fxrapt* [74]. Os resultados das simulações mostraram que não houve diferenças significativas entre os valores de pitch calculados a partir de uma vogal sustentada */a/*, pelo *fxrapt* e pelos picos de máximo do sinal DEGG.

7. Os gráficos *Ke versus Avegg* e *Ke versus Kd1* possuem boa discriminação visual entre locutores o que pode auxiliar os peritos na elaboração de laudos. Entretanto, estes gráficos precisam ser testados para um grande número de locutores e seu resultado validado por métodos estatísticos. A quantidade de dados e características é grande, sendo esta dissertação uma primeira tentativa de apresentá-los.

8.2 Trabalhos futuros

Apresentamos, a seguir, algumas posibilidades de continuação da presente pesquisa:

- Implementar filtros que suavizem os sinais OFG, DOFG e DEGG.
- Ampliar o *corpus* EGG.
- Estudar o problema da estimação de picos duplos no sinal DOFG.
- Implementar um sistema de verificação de locutor independente do texto, inserindo os parâmetros encontrados nessa dissertação para avaliar as taxas de acerto.
- Pesquisar um aperfeiçoamento do método PSIAIF.
- Ampliar o conjunto de parâmetros para comparação.

Apêndice A

Frases, palavras e vogais da base de dados

As Tabelas A.1, A.2, A.3 e A.4, contêm as frases, palavras e vogais sustentadas que compõem a base de dados disponível em *www.ime.eb.br/~labvoz/*.

A Tabela A.1 contém as frases foneticamente balanceadas para o portugês falado no Rio de Janeiro.

Tabela A.1: Frases foneticamente balanceadas para o português falado no Rio de Janeiro.

	Frases Balanceadas
1	Eu vi logo a Iôiô 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 coberto 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 teencontro.
9	Esses são nossos times.
10	Procurei Maria na copa.

Tabela A.2: Frases de interesse para perícia forense.

	Frases - Perícia Forense
1	Aaa tá, amanhã euligo.
2	Aaa tá, tamo junto irmão.
3	Alô, quem fala?
4	Alô, alô, alô!
5	Alô, quer falar com quem?
6	Eu digo alô baixinho.
7	Cadê a parada?
8	Eu digo parada baixinho.
9	Esse bagulho é bom.
10	Cadê o bagulho?
11	Eu digo bagulho baixinho
12	Cara, cadê você?
13	Cara, amanhã faço a entrega.
14	Eu digo cara baixinho.
15	Café ou açúcar?
16	Eu quero dois gramas de açúcar.
17	Eu quero dez quilos de açúcar e dois quilos de café.
18	Eu digo café baixinho.
19	Eu digo açúcar baixinho.
20	Fala praele que tô bolado com essa parada.
21	Tô bolado.
22	Eu digo bolado baixinho.
23	Eu digo copa baixinho.
24	Eu digo tudo baixinho.
25	Iiiiiii, isso não vai dar certo.
26	Tá tudo dominado no Turano
27	Eeeeeeeee, acho que sim.
28	Pede pra sair.

Tabela A.3: Palavras de interesse para perícia forense.

	Palavras - Perícia Forense
1	[pataká]
2	[peteké]
3	[pitikí]
4	[potokó]
5	[putukú]

Tabela A.4: Vogais sustentadas.

	Vogais Sustentadas
1	a
2	ã
3	é
4	ê
5	i
6	ó
7	ô
8	u

Apêndice B

Comparação entre vogais sustentadas e concatenadas

Tabela B.1: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 01 - vogal /a/

	Vogal /a/			
Locutor 01	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.2489	0.0620	3.1848	0.1312
Avegg	4.5206	0.0143	3.7656	0.1199
Ko	1.9808	0.0279	1.9901	0.0690
Ke	3.3838	0.0060	3.2297	0.1098
Kd1	2.9800	0.0078	2.8791	0.2452
Kd2	4.1817	0.0060	4.0878	0.0880
Da	0.9874	0.0326	0.9380	0.0649
Dp	0.3975	0.5560	0.8724	1.2782
Df	0.8099	0.0026	0.9355	0.5113
Keo	1.4029	0.0364	1.2395	0.1067

Tabela B.2: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 01 - vogal /e/

	Vogal /e/			
Locutor 01	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.4090	0.0340	3.1787	0.1777
Avegg	4.4948	0.0528	4.1942	0.1619
Ko	2.1513	0.0108	1.9503	0.2658
Ke	3.1736	0.0184	3.6241	0.9389
Kd1	2.8710	0.0201	2.9655	0.8839
Kd2	4.3187	0.0206	4.0598	1.6170
Da	0.9369	0.0141	1.0059	1.6690
Dp	0.3551	0.2489	1.9497	11.1990
Df	0.9234	0.0077	0.8775	0.0054
Keo	1.0222	0.0385	1.6738	0.3126

Tabela B.3: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 01 - vogal /i/

	Vogal /i/			
Locutor 01	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.8619	0.0355	2.7990	0.0528
Avegg	4.0995	0.0030	4.1601	0.0207
Ko	NaN	NaN	2.5461	0.0380
Ke	2.8670	0.0087	2.8902	0.0244
Kd1	2.5211	0.0056	2.7200	0.0303
Kd2	4.1814	0.0061	3.5566	0.2339
Da	0.5221	0.6100	0.7872	0.0139
Dp	2.1982	0.7307	2.5304	0.2354
Df	0.8252	0.0661	3.6609	7.2297
Keo	0.4072	0.9463	0.3500	0.0253

Tabela B.4: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 01 - vogal /u/

	Vogal /u/			
Locutor 01	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.2707	0.0289	2.8696	0.22997
Avegg	4.371	0.039978	4.3673	0.41732
Ko	2.4587	0.005022	2.2021	0.031619
Ke	2.7902	0.017234	3.1321	0.31944
Kd1	2.2622	0.0039681	2.8291	0.82304
Kd2	4.1717	0.0044151	3.7003	0.076136
Da	0.40496	0.50812	1.3656	3.2623
Dp	1.4198	1.2042	1.3811	1.8373
Df	0.86169	0.001078	NaN	NaN
Keo	0.3316	0.0256	0.9300	0.3756

Tabela B.5: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 02 - vogal /a/

	Vogal /a/			
Locutor 02	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.5396	0.0192	3.6803	0.0750
Avegg	2.0248	0.3799	2.6690	0.0077
Ko	3.1359	0.1197	1.6073	0.0579
Ke	5.1173	0.0206	3.9620	0.1312
Kd1	3.5015	1.6180	2.2995	0.4059
Kd2	6.3348	1.6583	4.6875	0.4484
Da	0.8257	0.0963	0.7035	0.0557
Dp	1.2496	6.0036	2.4156	9.5628
Df	0.7647	0.0022	0.6723	0.0011
Keo	1.9814	0.1554	2.3547	0.1393

Tabela B.6: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 02 - vogal /e/

	Vogal /e/			
Locutor 02	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.7660	0.0126	3.5391	0.1524
Avegg	1.8860	0.0061	2.7284	0.0847
Ko	2.5689	0.0024	2.0116	0.1395
Ke	5.5250	0.0330	4.4642	1.9787
Kd1	4.3337	1.5534	3.1422	3.2039
Kd2	5.5470	1.5827	4.6287	0.2630
Da	0.9942	0.0202	1.3001	3.4280
Dp	3.3143	19.1560	2.5339	12.2110
Df	0.7917	0.0037	0.6270	0.0023
Keo	2.9561	0.0352	2.4527	1.4559

Tabela B.7: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 02 - vogal /i/

	Vogal /i/			
Locutor 02	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.7018	0.2869	3.2686	0.1387
Avegg	2.3045	0.0357	3.1292	0.0182
Ko	3.6550	0.0172	1.9312	0.3687
Ke	5.5680	0.0292	4.6011	0.0478
Kd1	4.7622	1.1100	3.3560	0.3013
Kd2	4.8259	1.3220	4.0707	0.1202
Da	8.9965	0.0432	2.9115	12.1040
Dp	2.2937	11.4390	2.2879	4.2523
Df	0.9474	0.0071	0.7214	0.0288
Keo	1.9131	0.0592	2.6699	0.3319

Tabela B.8: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 02 - vogal /u/

	Vogal /u/			
Locutor 02	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.6634	0.1428	2.9459	0.5419
Avegg	2.7608	0.0107	2.8320	0.0961
Ko	3.9191	0.3451	2.5794	1.0889
Ke	4.8445	0.2054	4.6285	0.7660
Kd1	4.2923	0.1138	3.9123	1.0715
Kd2	5.1908	0.1134	3.9890	0.4410
Da	7.3371	11.7890	5.1564	13.9540
Dp	2.1258	4.6858	2.9903	7.3539
Df	0.9281	0.0041	1.2605	1.8267
Keo	0.9269	0.7380	2.1486	0.9534

Tabela B.9: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 03 - vogal /a/

	Vogal /a/			
Locutor 03	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.1879	0.7426	2.6952	0.5824
Avegg	1.5870	0.0107	1.8674	0.0515
Ko	2.0141	0.1190	5.5568	83.4870
Ke	1.9928	0.0045	3.6082	6.1441
Kd1	1.2425	0.0798	7.1458	120.0100
Kd2	4.0058	0.1479	2.3394	6.3945
Da	0.6183	0.1006	3.0591	14.8920
Dp	0.3274	0.4639	3.4951	52.6860
Df	0.6612	0.0008	2.4285	20.0760
Keo	0.3041	0.0381	5.3483	130.8100

Tabela B.10: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 03 - vogal /e/

	Vogal /e/			
Locutor 03	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.2924	0.3698	2.8153	0.3568
Avegg	2.2424	0.0079	1.7222	0.4314
Ko	2.3782	0.7203	1.7395	0.3977
Ke	2.2811	0.0154	2.3451	0.0285
Kd1	1.9185	0.0109	1.7911	0.0666
Kd2	3.4831	0.0093	3.0887	0.5739
Da	1.0019	0.0296	1.8614	4.4614
Dp	0.8430	1.0031	1.4487	1.6480
Df	3.0606	3.9439	0.7055	0.0021
Keo	0.8275	0.0337	0.8332	0.1393

Tabela B.11: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 03 - vogal /i/

	Vogal /i/			
Locutor 03	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.5999	0.8435	2.5646	0.4902
Avegg	2.7899	0.0011	2.2110	0.0351
Ko	1.4575	2.0875	1.8568	0.1031
Ke	2.8370	0.0061	2.3876	0.0033
Kd1	1.9931	0.0096	1.6507	0.0046
Kd2	2.9229	0.0091	2.7503	0.0042
Da	3.1058	3.8756	3.5116	2.4451
Dp	0.2874	0.0823	1.9412	1.6748
Df	0.5732	0.2890	0.8891	2.0087
Keo	2.0111	12.1010	0.5307	0.1162

Tabela B.12: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 03 - vogal /u/

	Vogal /u/			
Locutor 03	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.8544	0.1507	2.6700	0.3584
Avegg	2.6368	0.0035	2.3069	0.0147
Ko	2.0095	0.0222	1.6789	0.1249
Ke	2.4449	0.0084	2.3386	0.0118
Kd1	1.4398	0.0168	1.6000	0.0222
Kd2	3.3527	0.0146	2.8989	0.0489
Da	3.6976	3.5656	2.1150	4.1038
Dp	0.8536	0.2504	1.5099	0.9124
Df	0.4240	0.0109	1.0517	1.6063
Keo	0.4410	0.0241	0.6645	0.1183

Tabela B.13: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 04 - vogal /a/

	Vogal /a/			
Locutor 04	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.4889	0.4288	2.4934	0.5746
Avegg	1.9624	0.0623	1.5584	0.0230
Ko	1.9998	2.0212	1.8226	0.6545
Ke	4.3616	0.0952	4.3427	0.2617
Kd1	3.2766	0.2720	2.9355	1.2035
Kd2	3.8936	0.2601	4.1992	2.1809
Da	1.0738	2.6393	1.2513	3.5788
Dp	2.9532	6.9841	0.4226	1.7165
Df	0.7736	1.5063	1.7185	5.3178
Keo	2.5239	1.2838	2.5308	0.9273

Tabela B.14: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 04 - vogal /e/

	Vogal /e/			
Locutor 04	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.8346	0.2915	2.9013	0.3765
Avegg	2.1216	0.0177	2.7983	0.2422
Ko	1.5734	0.0500	1.9077	0.4219
Ke	4.1609	0.0086	5.0670	0.2445
Kd1	3.4231	0.7861	2.4689	2.5202
Kd2	3.7947	0.7761	5.0749	1.8474
Da	0.9546	0.0401	5.9502	6.0382
Dp	3.6902	11.1950	0.7070	3.3055
Df	0.5386	0.0011	0.9836	1.4165
Keo	2.5875	0.0606	3.1593	0.2030

Tabela B.15: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 04 - vogal /i/

	Vogal /i/			
Locutor 04	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.1279	0.1922	3.0461	0.2123
Avegg	2.9177	0.0731	2.4440	0.1938
Ko	1.4012	0.0451	1.4742	0.2032
Ke	3.6153	0.0311	3.7861	0.0803
Kd1	3.0251	0.1317	3.1572	0.7330
Kd2	2.8432	0.1309	3.3210	0.5450
Da	1.4436	4.9080	0.4876	0.0511
Dp	1.8423	6.1434	1.8289	3.5564
Df	0.6275	0.0051	0.5535	0.0352
Keo	2.2146	0.0724	2.4897	1.1598

Tabela B.16: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 04 - vogal /u/

	Vogal /u/			
Locutor 04	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.6672	0.4076	2.4922	0.4098
Avegg	3.1287	0.0063	2.4970	0.5922
Ko	2.3202	1.0348	1.9840	0.2356
Ke	3.5728	0.0387	3.7726	0.2599
Kd1	2.7721	0.0187	2.7968	0.0280
Kd2	2.6772	0.0188	3.5530	0.3869
Da	1.5711	5.2698	2.8327	6.8569
Dp	2.1024	3.3185	2.6464	0.1338
Df	2.8304	3.8350	NaN	NaN
Keo	1.3775	2.5304	1.7945	0.6213

Tabela B.17: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 04 - vogal /a/

	Vogal /a/			
Locutor 05	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.7177	0.8209	1.9196	0.4231
Avegg	13.8130	0.0369	13.3170	0.3150
Ko	1.9401	0.1138	NaN	NaN
Ke	4.0526	0.0232	4.6423	0.1581
Kd1	4.0626	0.0248	4.9044	0.0472
Kd2	4.0286	0.0175	3.8300	0.0684
Da	0.9388	0.4026	1.4358	0.9711
Dp	0.5258	1.3785	0.7396	4.9077
Df	1.2039	0.0037	NaN	NaN
Keo	2.1125	0.1585	4.8593	28.6400

Tabela B.18: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 05 - vogal /e/

	Vogal /e/			
Locutor 05	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.8333	0.5642	1.5072	0.6148
Avegg	13.1010	0.2232	14.9160	4.9168
Ko	2.0991	0.3198	2.0168	8.9038
Ke	4.2856	0.0210	4.1562	1.5688
Kd1	3.7669	0.0192	3.6476	1.2999
Kd2	4.2359	0.0209	3.4217	1.8217
Da	1.2404	2.3591	1.3654	2.1498
Dp	1.3987	7.1723	1.9885	9.9565
Df	0.7247	0.0005	NaN	NaN
Keo	2.1865	0.3539	3.9268	31.6580

Tabela B.19: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 05 - vogal /i/

	Vogal /i/			
Locutor 05	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.6422	0.7208	2.6863	0.0940
Avegg	15.2370	0.0458	15.6720	0.1416
Ko	2.3319	0.1330	1.5302	0.0279
Ke	4.1253	0.0088	4.1955	0.0249
Kd1	3.5994	0.0705	4.0303	0.0390
Kd2	4.4370	0.0713	3.8961	0.0607
Da	1.2188	0.6343	1.4454	0.0889
Dp	1.7596	1.8578	2.4953	16.3630
Df	0.6606	1.0254	0.6697	0.0017
Keo	1.7934	0.1493	5.7441	39.2770

Tabela B.20: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 05 - vogal /u/

	Vogal /u/			
Locutor 05	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.1682	0.8964	1.4551	0.9054
Avegg	14.7110	0.0223	14.8580	0.7094
Ko	2.5795	0.3357	3.0816	13.1510
Ke	4.0119	0.0047	3.9808	0.6554
Kd1	3.6128	0.0085	3.7921	0.5324
Kd2	4.2528	0.0093	3.3560	0.2520
Da	2.9822	9.3499	3.0573	5.5681
Dp	0.8771	4.6458	1.9581	6.1180
Df	0.6569	0.0002	NaN	NaN
Keo	1.4324	0.3073	5.4815	29.7830

Tabela B.21: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 06 - vogal /a/

	Vogal /a/			
Locutor 06	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.5453	0.2400	3.1409	0.4047
Avegg	6.9642	0.0107	5.7777	0.2017
Ko	1.8594	0.1116	1.7843	0.1200
Ke	4.9310	0.0032	4.4248	0.0692
Kd1	3.0394	0.1792	2.7827	0.7157
Kd2	4.5859	0.1806	4.8770	1.2627
Da	5.0464	11.2030	0.6080	0.0851
Dp	0.9161	4.7219	2.1071	11.5820
Df	0.6993	0.0004	0.6952	0.0004
Keo	3.0716	0.1236	2.6406	0.0431

Tabela B.22: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 06 - vogal /e/

	Vogal /e/			
Locutor 06	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.4101	0.1475	2.4443	0.8471
Avegg	7.9079	0.0094	3.3409	0.1771
Ko	1.8134	0.0086	2.6489	0.7190
Ke	4.6063	0.0051	4.2800	0.3814
Kd1	3.1031	0.1105	3.8770	12.4690
Kd2	4.3170	0.1049	4.3519	8.3691
Da	0.2662	0.0132	2.1771	10.2250
Dp	2.2764	10.3480	2.2319	8.4868
Df	0.7345	0.0004	1.5405	6.1008
Keo	2.7929	0.0142	1.6416	0.3644

Tabela B.23: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 06 - vogal /i/

	Vogal /i/			
Locutor 06	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.2903	0.2416	2.4667	0.5835
Avegg	8.5870	0.0525	7.8178	0.4247
Ko	1.1625	0.0032	1.6126	0.0870
Ke	4.3501	0.0020	3.3672	0.1923
Kd1	2.6272	0.0020	2.5891	0.0261
Kd2	3.9488	0.0023	3.7904	0.3445
Da	0.4533	0.0028	1.3314	0.0076
Dp	0.8385	2.5564	1.4820	1.0188
Df	0.6083	0.0010	1.9956	8.5697
Keo	3.1877	0.0047	1.7546	0.2258

Tabela B.24: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 06 - vogal /u/

	Vogal /u/			
Locutor 06	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.0586	0.2451	2.0815	0.9152
Avegg	8.7422	0.0156	5.3467	2.1689
Ko	1.2616	0.0296	3.5792	29.1890
Ke	3.6221	0.0070	4.2207	0.1400
Kd1	2.5405	0.0585	2.7382	0.2432
Kd2	3.4955	0.0604	4.5618	0.5283
Da	0.6547	0.0098	0.7924	0.0909
Dp	0.7763	0.2673	2.2522	4.2258
Df	0.4875	0.0274	NaN	NaN
Keo	2.3605	0.0402	3.2774	20.0780

Tabela B.25: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 07 - vogal /a/

	Vogal /a/			
Locutor 07	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.3778	1.0725	2.3407	0.6010
Avegg	1.6846	0.0195	1.9921	0.0380
Ko	1.2690	0.1152	1.5835	1.2275
Ke	2.8914	0.0861	2.6161	0.1286
Kd1	2.0850	0.4445	3.2579	12.5410
Kd2	2.5161	0.5307	2.6831	0.2923
Da	3.5649	2.4445	3.3619	3.9135
Dp	1.1512	2.6879	2.8995	49.2000
Df	0.8385	0.0040	1.2367	1.3576
Keo	1.6234	0.2069	1.5647	0.2879

Tabela B.26: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 07 - vogal /e/

	Vogal /e/			
Locutor 07	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.3613	0.9769	2.0393	0.9916
Avegg	1.9471	0.0290	2.2599	0.1289
Ko	1.3771	0.1691	1.3168	0.4890
Ke	2.5369	0.0194	2.5250	0.2152
Kd1	1.7951	0.3047	1.6159	0.1788
Kd2	2.8500	0.3196	2.6908	0.4083
Da	2.6667	4.7249	1.7967	3.6765
Dp	1.1994	3.0559	0.5130	1.4082
Df	0.8801	0.0039	0.6496	0.0171
Keo	1.1997	0.0837	1.3417	0.1680

Tabela B.27: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 07 - vogal /i/

	Vogal /i/			
Locutor 07	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.2438	0.6240	1.6725	0.1080
Avegg	2.3362	0.0531	2.5278	0.0118
Ko	1.9510	0.3060	2.2093	0.0294
Ke	2.7832	0.1055	2.3601	0.0800
Kd1	2.0555	0.0897	1.8497	0.0039
Kd2	2.2109	0.1139	2.6030	0.0224
Da	3.2498	1.3253	3.9997	0.0828
Dp	0.8092	0.1706	1.1624	0.0771
Df	0.3175	0.2356	0.1549	0.0026
Keo	0.9444	1.3067	0.2676	0.0626

Tabela B.28: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 07 - vogal /u/

	Vogal /u/			
Locutor 07	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	1.7635	0.7659	1.6495	0.8275
Avegg	2.8943	0.0618	2.7810	0.2206
Ko	2.0729	1.1230	1.9018	0.6024
Ke	2.7705	0.0261	2.3569	0.1166
Kd1	1.9403	0.1073	1.7485	0.1105
Kd2	2.6591	0.0502	2.5871	0.0192
Da	4.0935	0.6134	2.8578	2.5084
Dp	1.2220	0.6568	0.9527	0.6128
Df	0.5529	0.4898	1.6986	3.8715
Keo	1.2367	1.9134	0.6256	0.2196

Tabela B.29: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 08 - vogal /a/

	Vogal /a/			
Locutor 08	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.0468	1.2342	1.9523	0.8339
Avegg	2.0749	0.1515	2.3367	0.0092
Ko	1.7123	0.2590	1.1427	0.1495
Ke	3.0862	0.0560	3.0596	0.1464
Kd1	1.4530	0.1475	1.8379	1.3612
Kd2	3.4883	0.1480	3.4733	0.6038
Da	3.6849	2.3944	1.1274	3.1266
Dp	0.9793	1.0653	1.8924	2.3143
Df	0.6674	0.0769	0.6353	0.4485
Keo	1.3814	0.3614	1.9169	0.4899

Tabela B.30: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 08 - vogal /e/

	Vogal /e/			
Locutor 08	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.3291	1.0197	2.2758	1.5255
Avegg	2.1462	0.1956	2.1266	0.0741
Ko	1.5547	0.5319	1.2701	0.4226
Ke	2.7805	0.0375	3.0615	0.0318
Kd1	1.5061	0.1642	1.5362	1.2055
Kd2	3.6147	0.1647	3.2763	1.4382
Da	2.4951	3.8892	2.7915	7.4390
Dp	1.4017	3.5283	0.7858	2.2450
Df	0.8682	0.0093	1.2262	1.1313
Keo	1.3195	1.8712	2.1583	3.3952

Tabela B.31: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 08 - vogal /i/

	Vogal /i/			
Locutor 08	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.1836	0.9032	2.0736	0.5695
Avegg	2.0565	0.0976	2.6710	0.0161
Ko	2.2477	0.3291	1.6429	0.5619
Ke	3.0411	0.0951	2.6248	0.1383
Kd1	1.8198	0.3084	1.5763	0.0715
Kd2	3.0992	0.1368	2.7532	0.0517
Da	3.5082	1.7167	3.9853	0.4160
Dp	0.7050	0.1527	1.0543	0.5155
Df	0.3190	0.1767	0.8251	1.9158
Keo	0.8667	0.2743	1.5851	10.8470

Tabela B.32: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 08 - vogal /u/

	Vogal /u/			
Locutor 08	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.0256	0.8398	1.8581	0.9748
Avegg	2.7051	0.1082	2.5409	0.0145
Ko	1.8925	0.3733	1.9530	0.3467
Ke	3.0093	0.0606	2.6820	0.2273
Kd1	2.2695	0.8108	1.8252	1.3387
Kd2	2.6803	0.9645	2.6083	1.3451
Da	3.6141	3.8022	4.0776	0.1819
Dp	1.6064	1.1134	1.6214	0.7310
Df	1.2465	3.2557	1.4183	3.0165
Keo	1.1715	0.9351	0.8462	0.4125

Tabela B.33: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 09 - vogal /a/

	Vogal /a/			
Locutor 09	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.4115	0.7128	1.9200	1.4298
Avegg	2.0004	0.0139	1.9808	0.0207
Ko	0.9857	0.1864	0.8942	0.1215
Ke	2.6239	0.0792	2.5254	0.0915
Kd1	2.2659	0.1020	2.0808	0.1387
Kd2	2.1271	0.0378	2.0549	0.4951
Da	1.3494	3.8641	2.2590	3.6468
Dp	1.6719	2.1872	1.2383	1.2853
Df	0.7584	0.0214	0.8146	0.1257
Keo	1.6685	0.1532	1.6312	0.2220

Tabela B.34: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 09 - vogal /e/

	Vogal /e/			
Locutor 09	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	3.1045	0.6385	2.4027	0.6491
Avegg	2.1660	0.0199	2.4316	0.0089
Ko	0.9924	0.0113	1.1906	0.2397
Ke	2.5296	0.0292	2.7210	0.2405
Kd1	2.4000	0.0239	2.8437	0.2128
Kd2	2.0834	0.0480	1.5980	0.1006
Da	1.0506	2.7612	3.4037	2.5932
Dp	1.2818	1.9534	2.1289	1.9304
Df	0.8021	0.0044	0.7931	0.3870
Keo	1.5372	0.0507	1.5372	0.2963

Tabela B.35: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 09 - vogal /i/

	Vogal /i/			
Locutor 09	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.6244	0.2207	2.4750	0.1887
Avegg	2.6482	0.0454	3.0397	0.0063
Ko	1.9279	0.3322	1.9497	0.2477
Ke	2.5413	0.0366	2.2769	0.0271
Kd1	2.3976	0.0451	2.2005	0.0153
Kd2	1.7350	0.0235	1.7241	0.0417
Da	3.0689	1.6465	1.9889	1.9465
Dp	0.9024	0.1118	0.9780	0.0088
Df	0.4784	0.9171	0.6701	1.3827
Keo	0.7253	0.2217	0.4746	0.1749

Tabela B.36: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 09 - vogal /u/

	Vogal /u/			
Locutor 09	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.1945	0.6385	2.2728	0.3887
Avegg	3.3265	0.0196	3.1247	0.0270
Ko	1.7599	0.2622	1.9439	0.1150
Ke	2.5559	0.2493	2.3889	0.0543
Kd1	2.7249	1.2006	2.4705	0.0754
Kd2	1.5206	1.2935	1.5903	0.0171
Da	1.4104	3.6462	2.6371	2.9900
Dp	1.3847	0.6342	1.5111	0.7774
Df	3.6529	1.7006	NaN	NaN
Keo	1.1675	1.3924	0.4573	0.0761

Tabela B.37: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 10 - vogal /a/

	Vogal /a/			
Locutor 10	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.3420	1.0845	2.9542	0.2213
Avegg	12.6620	0.2798	11.9660	2.0287
Ko	1.6032	3.8818	1.8622	0.0534
Ke	5.7825	0.2605	5.2677	0.0915
Kd1	2.0318	2.8203	2.0894	1.4959
Kd2	8.2475	1.0519	6.7239	1.4839
Da	1.4161	0.1229	1.6450	7.4280
Dp	2.5503	19.0950	5.9358	9.0988
Df	1.1249	1.2624	0.9304	0.0014
Keo	5.2654	31.0080	3.4055	0.0395

Tabela B.38: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 10 - vogal /e/

	Vogal /e/			
Locutor 10	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.1260	1.1352	2.6294	0.9315
Avegg	14.4360	0.1150	10.9180	1.3461
Ko	NaN	NaN	3.2257	1.8870
Ke	5.9008	0.0997	6.5637	3.0042
Kd1	3.5054	3.0846	2.2854	0.1341
Kd2	6.6447	4.0182	7.6142	8.5429
Da	1.3594	0.1588	0.9661	0.1613
Dp	4.1184	23.1100	4.8546	12.2190
Df	0.8279	0.0309	NaN	NaN
Keo	5.1449	34.0890	3.3380	1.0597

Tabela B.39: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 10 - vogal /i/

	Vogal /i/			
Locutor 10	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.1717	0.8038	3.4443	0.0037
Avegg	14.5370	3.0250	17.2820	0.0216
Ko	2.7431	6.6068	1.9415	0.0075
Ke	4.9567	0.0852	4.6907	0.0066
Kd1	2.8503	2.6573	4.5322	0.0068
Kd2	6.4847	3.9873	3.6926	0.0044
Da	2.3482	9.5833	0.9189	0.0146
Dp	3.5917	14.2810	4.7780	7.1620
Df	NaN	NaN	0.5992	0.0023
Keo	4.0568	46.1760	2.7492	0.0009

Tabela B.40: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 10 - vogal /u/

	Vogal /u/			
Locutor 10	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.4724	0.7661	1.7940	0.7284
Avegg	13.4920	0.3096	11.8570	3.8924
Ko	4.4038	16.0790	4.4034	38.6170
Ke	4.9471	0.0096	5.3339	0.5668
Kd1	3.7302	0.7210	2.8837	3.6173
Kd2	5.6453	0.9719	6.2129	2.9645
Da	0.6913	0.0799	4.5476	12.7230
Dp	1.4647	11.1370	2.8891	14.4190
Df	1.3219	5.9586	NaN	NaN
Keo	2.8121	8.4698	6.0360	55.7400

Tabela B.41: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 12 - vogal /a/

	Vogal /a/			
Locutor 12	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.4694	0.9297	2.4419	0.8205
Avegg	0.5373	0.0031	0.8038	0.0083
Ko	1.4354	1.9169	1.1709	0.1132
Ke	3.6530	0.0513	2.6774	0.5784
Kd1	2.3765	2.2615	1.5898	0.7578
Kd2	3.4565	0.7334	2.9263	0.5569
Da	3.8305	3.8591	2.6326	4.1811
Dp	1.7339	3.2873	1.1761	0.6124
Df	0.9492	0.1318	0.6906	0.0413
Keo	2.4618	6.0240	1.5064	0.6406

Tabela B.42: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 12 - vogal /e/

	Vogal /e/			
Locutor 12	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.1096	0.9935	2.1473	1.3457
Avegg	0.7093	0.0077	0.5151	0.0279
Ko	1.2808	0.5159	2.1379	3.2953
Ke	3.0301	0.1261	2.8891	0.4299
Kd1	2.3218	0.9754	2.0073	2.1016
Kd2	3.2272	0.6652	3.3945	2.2758
Da	1.0141	2.2710	2.8136	4.2221
Dp	2.1028	3.8049	1.1709	2.2242
Df	1.0349	0.5758	1.1004	0.6307
Keo	1.9901	1.3551	1.8203	2.7788

Tabela B.43: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 12 - vogal /i/

	Vogal /i/			
Locutor 12	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.0555	0.8335	2.2635	0.9892
Avegg	0.9395	0.0083	0.8533	0.0089
Ko	2.8538	0.2679	1.9333	0.2214
Ke	3.4359	0.0394	2.7394	0.0535
Kd1	2.8687	1.6441	2.0458	0.1765
Kd2	2.7323	0.7983	2.4546	0.1581
Da	3.5727	1.2094	3.2661	2.5628
Dp	1.5202	0.5818	1.0760	0.3537
Df	NaN	NaN	1.3929	3.8212
Keo	0.5895	0.2872	0.8061	0.2986

Tabela B.44: Resultado da comparação entre vogais sustentadas e concatenadas para o locutor 12 - vogal /u/

	Vogal /u/			
Locutor 12	**Sustentada**		**Concatenada**	
Parâmetros	**Média**	**Variância**	**Média**	**Variância**
Av	2.1096	0.9498	2.3123	1.3638
Avegg	0.9767	0.0018	0.7487	0.0045
Ko	2.4113	0.3955	1.7421	0.3842
Ke	3.2698	0.0290	2.8472	0.3367
Kd1	2.3299	0.7800	2.3395	0.6152
Kd2	2.8099	0.5814	2.4678	0.4667
Da	3.8306	0.4198	2.4833	3.6672
Dp	1.3529	0.1710	1.3295	1.7759
Df	0.6579	0.0276	NaN	NaN
Keo	0.8712	0.4043	1.8293	4.1455

Bibliografia

[1] Majewski, W., Basztura, C.,"Integrated approach to speaker recognition in forensic applications", Forensic Linguistics 3 (1), pp.50-64, 1996.

[2] Broeders,"Forensic Speech and Audio Analysis Forensic Linguistics 1998 to 2001", A Review, Dept. of Handwriting, Speech and Document Examination Forensic Institute, Ministry of Justice, 13th INTERPOL Forensic Science Symposium, Lyon, France, October 16-19, 2001.

[3] Robertson, B. et al., "Interpreting Evidence Evahating Forensic Science in the Courtroom", 1997.

[4] Wiley, U.K., Foster, K.R.e Huber P.W.,"Judging Science: Scientific Knowledge and the Federal Courts", MIT Press, 1997.

[5] Taroni, F., Aitkeu, C.G.C., "Forensic Science at Trial", Jurimem Journal 37, pp. 327-337, 1997.

[6] Louis-Jean Boe,"Forensic voice identification in France", Institut de la Communication Parlée, 2000.

[7] Fabre, P., "Sphygmographie par simple contact d électrodes cutanées, introduisant dans l arterè de faibles courants de haute fréquence détecteurs de ses variations volumétriques", Comptes Rendus Soc. Biol., vol. 133, pp. 639-641, 1940.

[8] Colton, R. H. e Conture, E. G., "Problems and pitfalls of electroglottography", Journal of Voice, vol. 4, no. 1, pp. 10-24, 1990.

[9] Baken, R. J., "Electroglottography", Journal of Voice, vol. 6, no. 2, pp. 98-110, 1992.

[10] Henrich, N., "Etude de la source glottique en voix parlée et chantée: modélisation et estimation, mesures acoustiques et électroglottographiques, perception", Thèse de doctorat de l Université Paris 6, PhD thesis, pp. 87-96, 2001.

[11] Campbell, Jr. J. P., "Speaker Recognition: A tutorial", Proceedings of the IEEE, vol. 85, no. 9, pp. 1437-1462, 1997.

[12] Christophe Champod, Didier Meuwly,"The inference of identity in forensic speaker recognition", Institut de Police Scientifique et de Criminologie, University of Lausanne, pp. 193-203, 2000.

[13] R. B. de Sousa, "Identificação Humana pela Voz", Perito Criminal, Instituto de Criminalística Carlos Éboli (ICCE), Perícias de Áudio e Vídeo, 2008.

[14] Fombonne, "La Criminalistique", Que Sais-je, PUF, Paris, 1996.

[15] Tuthill, H., "Individualization: Principles and Procedures in Criminalistics", Lightning Powder, Salem, 1994.

[16] J. Gonzalez-Rodriguez, J. Fiewez-Aguilar and J. Ortega-Garcia, "Forensic identification Reporting using Automatic Speaker Recognition Systems", Dpt. Audio-visual and Communication Engineering Universidad Politecnica de Madrid (SPAIN), 1962.

[17] Meuwly, D., "Current Discussions of the, ENFSI-WG About the Use of the Bayesian Approach for the Interpretation of Evidence", ENFSl Speech and Audio Group Meeting, Pans (France), 2001.

[18] Cataldo E., Rubens S., Nicolato L., "Uma Discussão sobre Modelos Mecânicos de Laringe para Síntese de Vogais", ENGEVISTA, v. 6, n. 1, p. 47-57, abr. 2004

[19] Fechine J. M., "Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística", Tese de Doutorado em Engenharia Elétrica da Universidade Federal da Paraíba, 2002.

[20] Ten. J. F. V. C. Flores, "Novas contribuições à verificação automática de locutor para fins forenses", Dissertação de Mestrado, Instituto Militar de Engenharia (IME), 2008.

[21] Cataldo E., Sampaio R., Ludero J., Soize C.,"Modeling random uncertainties in voice production using a parametric approach", Mechanics Research Communications, v. 35, p. 429-490, 2008.

[22] Cataldo E., Sampaio R.,Soize C., Desceliers C., "Probabilistic modelling of a nonlinear dynamical system used for producing voice", Computational Mechanics, v. 1, p. Disp on line, 2008.

[23] Silva, D. G., Lima, C. B., "Tutorial sobre caracterisitcas de voz"Ministério da Defesa, Exército Brasileiro, Secretaria de Ciência e Tecnologia, Instituto Militar de Engenharia (IME), 2006.

[24] Rabiner, L. R., Juang, B., "Fundamentals of Speech Recognition", Prentice Hall, p. 493, 1993.

[25] Fant, G., Acoustic Theory of Speech Production, Mouton, The Hague, 1960.

[26] Pulakka Hannu," Analysis of Human Voice Production Using Inverse Filtering, High-Speed Imaging, and Electroglottography", Helsinki University of Technology, Dept. of Computer Science and Engineering, 2005.

[27] Flanagan, 1972; Javkin et al., 1987; Veldhuis,1998

[28] Flanagan, J. L., "Speech Analysis Synthesis and Perception", Springer-Verlag, secondedn, 1972.

[29] Van den Berg, J.,"Myoelasticaerodynamic theory of voice production", Journal of Speech and Hearing Research, vol.1, pp. 227- 244, 1958.

[30] Titze, I. R., "Comments on the myoelastic-aerodynamic theory of phonation", The Journal of the Acoustical Society of America, vol. 23, pp. 495-510, 1980.

[31] Fant, G., Liljencrants, J. e Lin, Q., "A four-parameter model of glottal flow", Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR), Royal Institute of Technology, Stockholm, vol. 4, pp. 1-13, 1985.

[32] Strik, H., "Automatic parametrization of differentiated glottal flow: Comparing methods by means of synthetic flow pulses", The Journal of the Acoustical Society of America, vol. 103, no. 5, pp. 2659-2669, 1998.

[33] Gobl, C. e Chasaide, A. N., "The role of voice quality in communicating emotion,mood and attitude", Speech Communication, vol. 40, no. 1-2, pp. 189-212, 2003.

[34] Alku, P., "Glottal wave analysis with Pitch Synchronous Adaptive Inverse Filtering", Speech Communication, vol. 11, pp. 109-118, 1992.

[35] Vilkman, E., Lauri, E.-R., Alku, P., Sala, E. e Sihvo, M., "Loading changes in timebased parameters of glottal flow waveforms in different ergonomic conditions", Folia Phoniatrica et Logopaedica, vol. 49, pp. 247-263, 1997.

[36] Lecluse, F. L. E., Brocaar, M. P., and Verschuure, J., "The electroglottographyand its relation to glottal activity", Folia Phoniatr. 27, 215- 224, 1975.

[37] Pedersen, M. F., "Electroglottography compared with synchronized stroboscopy in normal persons", Folia Phoniatr. 29, 191-199, 1977.

[38] Teaney, D., and Fourcin, A. J., "The electrolaryngography as a clinicaltool for the observation and analysis of vocal fold vibration", The Voice Foundation, 1980.

[39] Fourcin, A. J., "Laryngographic examination of vocal fold vibration, in Ventilatory and Phonatory Function", edited by B. Wyke Oxford University Press, London, pp. 315-326, 1974.

[40] Anastaplo, S., and Karnell, M. P., "Synchronized videostroboscopic and electroglottographic examination of glottal opening", J. Acoust. Soc. Am. 83, 1883-1890, 1988.

[41] Karnell, M. P., "Synchronized videostroboscopy and electroglottography", J. Voice 3, 68-75, 1989.

[42] Childers, D. G., Hicks, D. M., Moore, G. P., Eskenazi, L., and Lalwani, A.L., "Electroglottography and vocal fold physiology", J. Speech,Hear. Res, 33, 245-254, 1990.

[43] Childers, D. G., and Krishnamurthy, A. K., "A critical review ofelectroglottography", CRC Crit. Rev. Biomed. Eng. 12, 131-161, 1985.

[44] Childers, D. G., and Larar, J. N., "Electroglottography for laryngealfunction assessment and speech analysis", IEEE Trans. Biomed. Eng. BME-31, 807-817, 1984.

[45] Baer, T., Lofqvist, A., and McGarr, N. S., "Laryngeal vibrations: Acomparison between high-speed filming and glottographic techniques", J. Acoust. Soc. Am. 73, 1304-1308, 1983.

[46] Berke, G. S., Moore, D. M., Hantke, D. R., Hanson, D. G., Gerratt, B. R., and Burstein, F., "Laryngeal modeling: Theoretical, in vitro, in vivo", Laryngoscope 97, 871-881, 1987.

[47] Dejonckere, P., "Comparison of two methods of photoglottographyin relation to electroglottoraphy", Folia Phoniatr. 33, 338, 1981.

[48] Gerrat, B. R., Hanson, D. G., and Berke, G. S., "Laryngeal configurationassociated with glottography", Am. J. Otolaryngol. 9, 173-17, 1988.

[49] Kitzing, P., "Photo- and electroglottographical recording of the laryngealvibratory pattern during different registers", Folia Phoniatr. 34, 234-241, 1982.

[50] Titze, I. R., Baer, T., Cooper, D., and Scherer, R., "Automatic extraction of glottographic waveform parameters and regression to acoustic and physiologic variables", in Vocal Fold Physiology: Contemporary Research Clinical Issues, edited by A. J. Bless DM College Hill, San Diego, pp. 146-154, 1984.

[51] Childers, D. G., Naik, J. M., Larar, J. N., Krishnamurthy, A. K., and Moore,G. P., "Electroglottography, speech and ultra-high speed cinematography", in Vocal Fold Physiology and Biophysics of Voice, edited by I. Titze and R. Scherer Denver Center for the Performing Arts, Denver, pp. 202-220, 1983.

[52] Rothenberg, M., "Some relations between glottal air flow and vocalfold contact area", ASHA Rep. 11, 88-96, 1981.

[53] Rothenberg, M., and Mahshie, J. J., "Monitoring vocal fold abductionthrough vocal fold contact area", J. Speech Hear. Res. 31, 338-351, 1988.

[54] Marasek, K., "EGG and voice quality", web page. Referenced 20 December 2004.URL http://www.ims.uni-stuttgart.de/phonetik/EGG/, 1997.

[55] Scherer, R. C., Druker, D. G. and Titze, I. R., "Electroglottography and direct measurement of vocal fold contact area", in O. Fujimora, ed., "Vocal Physiology: Voice Production,Mechanisms and Functions", Raven Press, New York, pp. 279-291, 1988.

[56] Deller J. R., Hansen J. H., Proakis J. G., "Discrete-Time Processing of Speech Signals", IEEE Press, p. 936, 2000.

[57] Gauffin, Hertegard, A. Lindestad, "A comparison of subglottal and intraoral pressure measurements during phonation", Journal of Voice, vol. 9, pp. 149-155, 1995.

[58] Hertegard, S., Gauffin, J. e Karlsson, I., "Physiological correlates of the inverse filtered flow waveform", Journal of Voice, vol. 6, no. 3, pp. 224-234, 1992.

[59] Sodersten, M., Hakansson, A. e Hammarberg, B., "Comparison between automatic and manual inverse filtering procedures for healty female voices", Logopedics Phoniatrics Vocology, vol. 24, pp. 26-38, 1999.

[60] Fritzell, "Inverse filtering"Journal of Voice, vol. 6, no. 2, pp. 111-114, 1992.

[61] http://audacity.sourceforge.net/

[62] Mattos, J. S., Silva, D. G., Cataldo, E., Apolinário, J.A., "Incursionando pelos domínios da eletroglotografia: proposta de um corpus EGG", XXVI Simpósio Brasileiro de Telecomunicações, 2008.

[63] Alcaim, A., Solewicz, J. A. e Moraes, J. A. "Frequência de ocorrência dos fones e lista de frases foneticamente balanceadas no português falado no Rio de Janeiro". Revista da Sociedade Brasileira de Telecomunicações, vol. 7, nr. 1, dez 1992.

[64] F. Plante, "A Pitch Extraction Reference Database", ESCA EUROSPEECH 95, 4th European Conf. on Speech Communication and Technology, 1995.

[65] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[66] Petre, S. , Yngve S.,"Model-order selection: a review of information criterion rules", IEEE Signal Processing Magazine, 1053-5888/04, 2004.

[67] Rabiner, Lawrence R.; Gold, Bernard (1975). Theory and application of digital signal processing. Englewood Cliffs, N.J.: Prentice-Hall, pp 63-67

[68] Kersta, L.G.,"Voiceprint identification", Nature 196, 1253-1257, 1962.

[69] Henrich, N., d' Alessandro, C., Doval, B., Castellengo, M., "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation", The Journal of the Acoustical Society of America, vol. 115, no. 3, pp. 1321-1332, 2004.

[70] Riquelme, C.,"Reconhecimento de voz e de locutor em ambientes ruidosos: Comparação das técnicas MFCC e ZCPA", Dissertação de Mestrado, Universidade Federal Fluminense, 2007.

[71] Rosenberg, A. E., "Effect of glottal pulse shape on the quality of natural vowels", The Journal of the Acoustical Society of America, vol. 49, no. 2, pp. 583-590, 1971.

[72] Holmberg, E., Hillman, R., e Perkell, J., "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice", The Journal of the Acoustical Society of America, vol. 84, no. 2, pp. 511-529, 1988.

[73] Fourcin, A. J., "Laryngographic assessment of phonatory function", ASHA Rep. 11, 116-124, 1981.

[74] Talkin, "A Robust Algorithm for Pitch Tracking (RAPT) in Speech Coding and Synthesis", W B Kleijn, K K Paliwal eds, Elsevier ISBN 0444821694, 1995.