

Universidade Estadual de Campinas  
Faculdade de Engenharia Elétrica e de Computação  
Departamento de Comunicações

## **Jogos Computacionais Fonoarticulatórios para Crianças com Deficiência Auditiva**

**Antônio Marcos de Lima Araújo**

**Orientador: Prof. Dr. Fábio Violaro**  
DECOM – FEEC – UNICAMP

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da UNICAMP como requisito parcial para a obtenção do título de Doutor em Engenharia Elétrica.

Banca Examinadora:

Dr. Fábio Violaro – FEEC/UNICAMP – Presidente

Dra. Liliane Desgualdo Pereira – EPM/UNIFESP

Dra. Maria Cecília Marconi Pinheiro Lima – FCM/UNICAMP

Dr. Maurílio Nunes Vieira – ICEX/UFMG

Dr. João Marcos Travassos Romano – FEEC/UNICAMP

Dr. Luís Geraldo Pedroso Meloni – FEEC/UNICAMP

Campinas, 18 de setembro de 2000

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

Ar15j

Araújo, Antônio Marcos de Lima

Jogos computacionais fonoarticulatórios para crianças com deficiência auditiva / Antônio Marcos de Lima Araújo.--Campinas, SP: [s.n.], 2000.

Orientador: Fábio Violaro.

Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Processamento de sinais – Técnicas digitais. 2. Reconhecimento automático da voz. 3. Jogos por computador. 4. Deficientes auditivos - Reabilitação. I. Violaro, Fábio. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

## **RESUMO**

O desenvolvimento de jogos computacionais, para exercícios de aprimoramento da coordenação fonoarticulatória, em especial para crianças e adolescentes com deficiência auditiva, constituiu o objetivo principal deste trabalho. Os jogos desenvolvidos visam proporcionar um ambiente lúdico no qual o deficiente auditivo possa sentir prazer ao explorar as suas potencialidades, ao longo da prática que permitirá a coordenação articulatória. Esta prática é necessária para a obtenção do aprimoramento do controle sensorio-motor e da destreza exigida para a realização dos sons, com mínimo esforço articulatório. As aplicações exigem modelos matemáticos rápidos e precisos, tal que possam proporcionar uma retroalimentação conseqüente para a criança. A enorme variabilidade da voz, em especial de deficientes auditivos, requer algoritmos com grande adaptação. Conciliar rapidez, precisão e adaptabilidade foi o grande desafio deste projeto. Para superar este desafio foi realizado um minucioso estudo sobre as características físicas dos sons da língua. Partindo das propriedades físicas dos sons, foi possível desenvolver modelos matemáticos relativamente simples, precisos e com grande adaptabilidade.

## **ABSTRACT**

The development of computational games for exercising the articulatory coordination, specially for hearing-impaired children and adolescents, constitutes the main objective of this work. The developed games try to provide a playfull atmosphere in which hearing-impaired people can feel pleasure when exploring their potentialities, along with the practice that will allow the coordination articulatory. Practice this, that is necessary for the obtaining of the improvement of the sensory-motor control and of the necessary ability for the accomplishment of the sounds, with minimum I make an effort articulatory. The applications demand fast and necessary, such mathematical models that can provide a consequent feedback for the child. The enormous variability of the voice, especially of deficient auditory, it requires algorithms with great adaptation. To reconcile speed, precision and adaptability was the great challenge of this project. To overcome this challenge a meticulous study it was accomplished on the characteristics physics of the sounds of the language. Leaving of the physical properties of the sounds, it was possible to develop mathematical models relatively simple, necessary and with great adaptability.



## AGRADECIMENTOS

*Às crianças e adolescentes que, pacientemente, fizeram (e refizeram) as longas gravações que proporcionaram o material acústico essencial ao desenvolvimento deste trabalho.*

*Aos amigos e companheiros do Laboratório de Processamento Digital de Fala do Departamento de Comunicações da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, pelo constante apoio e incentivo, mas sobretudo pela amizade.*

*Às fonoaudiólogas e estagiárias que testaram a versão preliminar e que proporcionaram valiosas informações essenciais ao desenvolvimento deste trabalho.*

*Ao Centro de Estudos e Pesquisas em Reabilitação “Prof. Dr. Gabriel Porto” da Faculdade de Medicina da Universidade Estadual de Campinas, pelo apoio decisivo.*

*Ao Curso de Fonoaudiologia da Escola Paulista de Medicina da Universidade Federal de São Paulo, pela acolhida e incentivo.*

*Aos amigos que ganhamos em Campinas, personalizados no Roberto e na Ana Maria.*

*Aos meus irmãos e amigos José Marcos, João Marcos, Luiz Romano, Benedicto Oswaldo, Paulo Marcos, Ronaldo Marcos, Luiz Felipe e Angela Maria pelo apoio e incentivo.*

*Ao Conselho Nacional de Pesquisas que tornou possível a realização deste trabalho.*



*Aos amigos e à direção do Centro Federal de Educação Tecnológica do Pará, em especial ao Prof. Sérgio Cabeça Braz, pelo constante e decisivo apoio que tornaram possível a realização este trabalho.*

*Às professoras e amigas da Escola Paulista de Medicina, em especial as professoras Alda Borges, Brasília Chiari e Liliane Desgualdo, agradeço o constante incentivo e amizade. À fonoaudióloga Silvana Bommarito por ter acreditado e investido neste projeto.*

*A todos os meus professores, desde as professoras Denise Correa e Adelaíde Soares dos meus primeiros anos escolares, até os professores da pós-graduação, João Marcos Romano e Amauri Lopes, que muito contribuíram para minha formação.*

*Aos meus orientadores da graduação, Prof. Jorge Bechara, e do mestrado, Abraham Alcaim, meu muito obrigado.*

*Ao companheiro Carlos Marcel Goto, que dedicadamente deu vida a este trabalho. Ao meu filho Marcuns Leandro, que pacientemente fez e refez as figuras e objetos dos jogos. Valeu.*

*À professora Maria Cecília Marcondes Pinheiro Lima que, acreditando neste trabalho, com um espírito científico democrático na busca do conhecimento, proporcionou condições para seu desdobramento na sua forma final, meus mais sinceros agradecimentos.*

*Ao meu orientador, prof. Fábio Violaro, cuja grande competência, segurança, sensibilidade humana, elevado espírito científico e visão democrática, proporciona a cada um de seus orientados uma orientação serena e profícua. Obrigado pela oportunidade de ter podido compartilhar de seus conhecimentos, e principalmente pela amizade e carinho dispensados. A certeza de poder sempre contar com sua orientação e competência proporcionou um ambiente seguro e tranquilo para o desenvolvimento deste trabalho. Muito obrigado.*



# DEDICATÓRIA

*À memória de meus avôs Benedicto Oswaldo e Luiz Romano  
um que adquiriu a simplicidade do sábio e o outro que acumulou a cultura e conhecimento  
científico. Inesquecíveis.*

*À meus amados e queridos filhos, Marcuns Leandro e Aline Carolina,  
cuja amizade e paciência me deram força e tempo para realizar este trabalho.*

*À minha amada esposa Francisca Canindé  
que, com seu amor e companheirismo, tem sido meu constante suporte.*

*À minha mãe Laise Conceição  
pelo amor, amizade, carinho, apoio e dedicação.  
Acepção plena e superior da palavra mãe.*



# UMA ESPERANÇA

*Que os avanços da ciência proporcionem a cura da deficiência auditiva  
e que assim o resultado desse trabalho possa ser apenas um brinquedo.*



# ÍNDICE

<b>1. Introdução</b> .....	1
1.1. Objetivos e estrutura da tese .....	4
<b>2. A deficiência auditiva e a voz</b> .....	7
2.1 A educação de deficientes auditivos .....	10
2.2 A produção da fala e o sistemas de retroalimentação .....	19
2.3 A voz .....	21
2.4 Características das vozes de deficientes auditivos .....	23
2.4.1. Questões de fonte .....	24
2.4.1.1. A respiração .....	24
2.4.1.2. A freqüência de vibração das pregas vocais .....	26
2.4.2. Questões de filtro .....	30
<b>3. Os modelos e objetivos dos jogos</b> .....	33
3.1. Características fundamentais para produção da fala .....	34
3.1.1. As vogais .....	34
3.1.2. As Consoantes .....	35
3.2. Os jogos computadorizados .....	36
3.2.1. Jogos para exercícios de fonte .....	37
3.2.1.1. Aprimoramento do controle do fluxo respiratório .....	37
3.2.1.2. Aprimoramento do controle da energia .....	40
3.2.1.3. Aprimoramento do controle da freqüência fundamental .....	43
3.2.2. Exercícios de filtro .....	47
3.2.2.1. Exercícios com vogais .....	50
3.2.2.2. Exercícios com fricativas .....	54



<b>4. Parametrização</b> .....	57
4.1. Medidas da intensidade .....	59
4.1.1. Medida da energia do quadro .....	60
4.1.2. Medida do número de bits efetivos .....	61
4.1.3. Controle de ajuste de intensidade dos aplicativos .....	62
4.2. Modelamento para exercícios com as pregas vocais .....	63
4.2.1. Resultados obtidos com o medidor de frequência fundamental .....	72
4.3. Modelagem das vogais .....	73
4.3.1. Análise das vogais do português falado no Brasil .....	77
4.3.2. Resultados obtidos .....	82
4.3.3. Procedimentos para classificação de vogais .....	83
4.4. Modelagem de fricativas .....	85
4.4.1. As características das fricativas .....	86
<b>5. Aplicações e resultados</b> .....	99
5.1. Aplicações .....	100
5.2. Considerações sobre as aplicações .....	104
<b>6. Considerações finais</b> .....	107
<b>7. Bibliografia</b> .....	117
<b>Apêndices</b> .....	127
A. Procedimento para estimar energia e número de bits efetivos.....	127
B. Procedimento para estimar $F_0$ .....	129
C. Procedimento para estimar $F_1$ e $P_{70s}$ .....	131
D. Procedimento para estimar FPES e CEE .....	137
E. Relação das publicações .....	139

# *Capítulo 1*

## Introdução

Técnicas de processamento digital de sinais apresentam diversas aplicações em processamento de fala. Conversores texto-fala (ALLEN, HUNNICUTT & KLAT, 1987), sistemas para reconhecimento de fala (MORAES, 1997) e reconhecimento de locutor (LÖFQVIST, 1986; FIGUEIREDO, 1993), se transformam em aplicações cada vez mais próximas de nosso cotidiano.

O aperfeiçoamento das técnicas de processamento digital de sinais, concomitante ao avanço dos equipamentos de informática, propiciaram condições para o desenvolvimento de novas e importantes aplicações, entre as quais os sistemas para auxílio aos deficientes físicos. Sistemas como os conversores texto-fala podem ser muito úteis para deficientes visuais, proporcionando a versão acústica de um texto. Os sistemas de reconhecimento de fala (fala-texto) podem ser úteis para deficientes auditivos, proporcionando uma versão escrita da mensagem oral. Além destas, temos o desenvolvimento de sistemas computacionais para auxílio ao aprendizado da fala por deficientes auditivos (COSTA, 1993; CASTRO e col., 1995; CHEN & HUANG, 1999; LIMA e col., 1999).

Esse trabalho tem como objetivo ampliar e aprimorar um sistema anterior, desenvolvido para auxílio ao aprimoramento da voz de deficientes auditivos. Este sistema foi desenvolvido em trabalho conjunto realizado por professores e alunos da Universidade Federal do Pará e do Centro Federal de Educação Tecnológica do Pará, com apoio do

Instituto Felipe Smalldone. O desenvolvimento deste sistema, que será denominado protótipo base, se iniciou em 1994 (CASTRO e col., 1995), sendo concluído em 1996.

O protótipo base foi desenvolvido na forma de jogos de computador, operando sobre o sistema operacional DOS. Ele utiliza algoritmos desenvolvidos para codificação e transmissão de voz, e faz uso de parâmetros obtidos para outras línguas. O aplicativo para vogais, por exemplo, utiliza dados médios dessas vogais obtidos para a língua inglesa (CASTRO e col., 1995).

O desenvolvimento do protótipo base foi limitado devido à falta de pesquisas sobre a aplicabilidade dos jogos, a indefinição quanto aos tipos e formatos mais necessários e adequados dos jogos para crianças e adolescentes, além da falta de parâmetros que definissem e caracterizassem a voz de crianças e adolescentes, em especial na língua portuguesa.

A atualização do sistema base, com o desenvolvimento de algoritmos mais apropriados e precisos para a caracterização da voz de crianças e adolescentes, sua adequação aos sons da língua portuguesa e a ampliação do sistema com novas aplicações são alguns dos objetivos desse trabalho.

O carácter multidisciplinar do trabalho exigiu uma interação constante com fonoaudiólogos, para avaliar a aplicabilidade e limitações do protótipo base, com o intuito de determinar quais modificações precisavam ser realizadas e como deveria o sistema ser ampliado.

A receptividade e disponibilidade de pesquisadores do Centro de Estudos e Pesquisas em Reabilitação "Prof. Gabriel Porto" da Universidade Estadual de Campinas (CEPRE/UNICAMP) e do curso de Distúrbios da Comunicação Humana: Campo Fonoaudiológico da Universidade Federal de São Paulo - Escola Paulista de Medicina

(UNIFESP/EPM), possibilitou a instalação do protótipo base nessas instituições para pesquisa de sua aplicabilidade, viabilidade e implicações no aprimoramento do controle da voz dos deficientes auditivos. A interação com esses grupos de pesquisas proporcionou informações essenciais ao aprimoramento e ampliação do novo sistema, que será denominado *Jogos de Voz*.

O estudo para obtenção das características da voz, em especial de crianças e adolescentes, exigiu a montagem de um banco de dados com gravações de palavras isoladas e fonemas sustentados realizados por crianças, e fonemas sustentados realizados por adolescentes. O banco de dados serviu de base para o desenvolvimento e aprimoramento dos modelos matemáticos propostos para a caracterização das vozes de crianças e adolescentes, assim como para obtenção dos dados referenciais dos parâmetros resultantes.

Este trabalho, partindo das informações obtidas junto aos grupos de pesquisa que utilizaram (e avaliaram) o protótipo base, se concentra na concepção dos jogos computacionais que possibilitem a realização de exercícios fonoarticulatórios, no desenvolvimento e/ou aprimoramento dos algoritmos que forneçam características da voz e possibilitem a animação dos jogos, e na obtenção dos valores referenciais dos parâmetros que serão empregados, em especial para crianças e adolescentes.

Foram desenvolvidos vários algoritmos que possibilitaram a realização de aplicativos computacionais. Estes aplicativos foram desenvolvidos visando o aprimoramento do controle do fluxo respiratório, do controle da frequência de vibração das pregas vocais e do posicionamento articulatório. Além disso, esses aplicativos foram implementados sob a forma de jogos para motivar a criança com deficiência auditiva, usuária em potencial do presente trabalho.

Os aplicativos foram realizados na linguagem *C Builder*, sobre uma plataforma tipo IBM-PC. Esses aplicativos foram implementados por um graduando de Engenharia da Computação da UNICAMP, através de um projeto de Iniciação Científica. Os requisitos para operação dos aplicativos são um computador tipo PC, com relógio de frequência maior ou igual a 100 MHz, dotado de placa de som e microfone e operando sob o sistema operacional *Windows*. Os aplicativos podem ser executados com qualquer tipo de microfone.

### 1.1. Objetivo e estrutura da tese

O desenvolvimento de jogos computacionais, para exercícios de aprimoramento da coordenação fonoarticulatória, em especial para crianças e adolescentes com deficiência auditiva, é o objetivo principal deste trabalho. Os jogos visam proporcionar um ambiente lúdico no qual o deficiente auditivo possa sentir prazer no exercício exploratório de suas potencialidades, durante a realização da necessária prática que permitirá a coordenação articulatória. Esta prática é fundamental para a obtenção do aprimoramento do controle tátil-cinestésico e da destreza necessária para a realização dos sons, com mínimo esforço articulatório.

A consecução do objetivo principal exige:

1. Identificar quais as características da voz das crianças com deficiência auditiva que necessitam ser aprimoradas;
2. Obter dados sobre vozes de crianças ouvintes que possam servir como referência;

3. Determinar os modelos dos exercícios e o formato de sua implementação na forma de jogos de computador, que os estimulem no exercício exploratório de suas potencialidades.

A realização dos objetivos exigiu o desenvolvimento e ou o aprimoramento de modelos matemáticos sobre sinais de voz, que possibilitem a extração de parâmetros que:

1. Forneçam informações sobre a característica a ser trabalhada, com precisão suficiente;
2. Garantam uma retroalimentação conseqüente;
3. Permitam a animação dos aplicativos;
4. Possam ser ajustados a cada estágio de controle fonoarticulatório do indivíduo.

Esse trabalho está dividido em 6 capítulos. O capítulo 2 apresenta uma descrição sinóptica das metodologias e filosofias mais empregadas na educação de deficientes auditivos. Além disso são resumidas as principais características das vozes de deficientes auditivos. O capítulo 3 é dedicado à apresentação dos modelos dos jogos propostos com seus objetivos específicos, incluindo a descrição de alguns dos jogos implementados. A apresentação e discussão dos algoritmos matemáticos propostos para alimentação dos jogos, bem como a apresentação dos resultados obtidos com esses algoritmos são realizadas no capítulo 4. No capítulo 5, são apresentados alguns dos resultados obtidos com a aplicação do sistema base, e que proporcionaram informações essenciais para a consecução deste trabalho. O capítulo 6 apresenta as conclusões deste trabalho.



## *Capítulo 2*

# A deficiência auditiva e a voz

A deficiência auditiva é resultado da perda, maior ou menor, da capacidade de percepção normal do sons. Verifica-se a existência de vários tipos de portadores de deficiência auditiva, de acordo com a época em que a mesma ocorreu (se antes ou após a aquisição da língua oral), de acordo com os diferentes graus de perda da audição, de acordo com a causa, etc. (BRASIL, 1995).

As perdas auditivas podem ser classificadas em grupos separados, segundo o valor da média (em dB NA) dos limiares tonais de via aérea, nas frequências de 500, 1.000 e 2.000 Hz. Existem diferentes classificações referentes a perdas auditivas, cujas diferenças entre si, em termos de dB, são muito pequenas. REDONDO & LOPES FILHO (1997) recomendam o uso da seguinte classificação: normal para perdas até 25 dB; leve para perdas entre 26 e 40 dB; moderada para perdas entre 41 e 55 dB; moderadamente severa entre 56 e 70 dB; severa para perdas entre 71 e 90 dB; profunda para perdas acima de 91 dB. BESS & HUMES (1998) advertem que perdas entre 15 e 25 dB podem ter efeitos negativos sobre crianças do ponto de vista educacional.

Pesquisas mostram que um ou dois bebês entre mil nascem com problemas auditivos sérios e de vinte a trinta irão apresentar alguma perda auditiva até a idade de

cinco anos (BOOTHROYD, 1982; NORTHREN & DOWNS, 1989). Estudos mais recentes, referenciados pelo *Joint Committee on Infant Hearing* da *American Academy of Pediatrics* estimam a existência de 1,5 a 6,0 crianças com perdas auditivas em cada 1.000 nascimentos (JCIH, 1995)

A privação auditiva constitui grave distúrbio sensorial, que afeta a capacidade de comunicação oral e de aprendizagem.

Para SILVA (1996) a língua oral utiliza elementos simbólicos com capacidade quase infinita de construção e, sendo abstrata, tornou-se o instrumento fundamental de comunicação do pensamento humano. A língua oral é desenvolvida e organizada através das relações afetivas e sócio-culturais do indivíduo em interação com o ambiente.

A deficiência auditiva prejudica a recepção da fala e sua retroalimentação, essenciais ao aprendizado dos atos coordenados necessários à aquisição da língua oral. A perda auditiva é o impedimento maior para o desenvolvimento da língua oral, pois restringe o indivíduo na recepção desta e reduz sua habilidade em monitorar sua própria fala.

Deficientes auditivos severos e profundos são incapazes de aprender a falar sem técnicas especiais. A falta da audição reduz a monitoração primária necessária ao desenvolvimento normal da aquisição da língua oral.

Para Helen Keller, cega e surda, que aprendeu a comunicar-se pelo tato, os problemas da surdez são mais profundos, mais complexos e mais importantes que os da cegueira. Para Helen Keller, *“a surdez é o maior dos infortúnios, a perda do mais vital dos estímulos: o som da voz, que nos traz a linguagem, desencadeia os pensamentos e nos mantém na companhia intelectual dos homens”* (ALMEIDA & IORIO, 1996).

A linguagem é uma atividade mental que abrange os dois níveis de experiência: simbolização e conceituação. A linguagem surge quando há associação entre o significante

(sons e imagens) e o contexto que traz o significado (idéias ou experiências que se quer comunicar). O ser humano utiliza um código lingüístico, isto é, utiliza uma língua para se expressar (RINALDI, 1997).

A língua carrega consigo conceitos generalizados, que são fonte do conhecimento humano. Instrumentos culturais especiais, como a escrita e a aritmética, expandem enormemente os poderes do homem, tornando a sabedoria do passado analisável no presente e passível de aperfeiçoamento no futuro (VIGOTSKY, LURIA & LEONTIEV, 1988).

O objetivo central na educação dos deficientes auditivos é a aquisição da linguagem. Existem diversas teorias sobre como a criança a adquire. A teoria comportamentalista, organizada por SKINNER (1978), vê o desenvolvimento da linguagem como um processo de modelagem ambiental. A teoria inatista, proposta por CHOMSKY (1971), acredita que a criança deve estar preparada para processar a fala que ela ouve e formar as espécies de estruturas que são características da língua humana. A teoria sócio-interacionista, tendo VIGOTSKY (1993) como seu principal representante, afirma que a fala primitiva da criança é essencialmente social, dando lugar a uma teoria que privilegia o papel do outro na construção da linguagem da criança.

A aquisição da linguagem, conquanto seja o objetivo central da educação dos deficientes auditivos, extrapola os limites deste trabalho, que visa simplesmente proporcionar-lhes jogos computacionais para auxiliá-los no aprimoramento do controle articulatório, como substrato para a aquisição da língua oral.

Neste trabalho nos limitaremos à fala, partindo da concepção de ALBANO (1990), que discute sua importância como substrato indispensável para a construção da linguagem, construída na interação com o adulto. Para a autora, quatro condições são absolutamente

imprescindíveis para o desenvolvimento da linguagem: a presença de um interesse subjetivo por ela; a existência de pelo menos um sistema sensoriomotor que permita exorbitar da brincadeira; a inserção num meio onde a linguagem faça parte das rotinas significativas; a presença de uma língua minimamente referenciada. Na concepção da autora, aprender a falar é descobrir como gramaticalizar recursos já explorados ludicamente nos subsistemas da sensoriomotricidade que melhor se prestem à reinvenção de processos de auto-referência.

Para ALBANO (1990), o que o homem tem a mais que outros animais é simplesmente um excedente de substância neural eficientemente organizada de forma a permitir-lhe uma maior capacidade e vontade de brincar. Assim, o homem pode ser considerado o único animal capaz de tirar grande prazer do exercício exploratório das suas potencialidades.

## 2.1. A educação de deficientes auditivos

Até meados do século XVI, os deficientes auditivos, denominados surdos-mudos, eram considerados ineducáveis e, conseqüentemente, deixados à margem como inúteis à coletividade. O médico italiano Girolamo Cardano, ainda no século XVI, afirmou que os surdos-mudos podiam ser postos em condições de “ouvir lendo e de falar escrevendo“. Posteriormente, o frade espanhol Pedro Ponce de Leon conseguiu ensinar a linguagem articulada a surdos-mudos e, em 1620, outro espanhol, Juan Pablo Bonet, publicou o primeiro livro sobre o assunto, em que explica como exercitar o educando para a emissão dos sons. O reconhecimento de surdos filhos de famílias nobres como pessoas de lei (os

mudos não eram reconhecidos), para que pudessem herdar títulos e a fortuna da família, foi um fator para o desenvolvimento de métodos educacionais especiais para os deficientes auditivos (SACKS, 1998).

Até 1750, a situação das pessoas com surdez pré-lingüística era calamitosa, pois eram considerados incapazes de desenvolver a fala, e portanto “mudos”, incapazes de comunicar-se livremente até mesmo com pais e familiares, restritos a alguns sinais e gestos rudimentares. Isolados, exceto nas grandes cidades, até mesmo da comunidade de pessoas com o mesmo problema, privados de alfabetização e instrução, de todo o conhecimento do mundo, forçados a fazer os trabalhos mais desprezíveis, vivendo sozinhos, muitas vezes à beira da miséria, eram considerados pela lei e pela sociedade como pouco mais do que imbecis (SACKS, 1998).

O abade francês Charles Michel de L'Épée criou o chamado método silencioso no século XVIII. Sem desprezar a importância da palavra oral, L'Épée deu relevo especial ao emprego de sinais manuais, estabelecendo uma linguagem convencional, como meio de instrução dos deficientes auditivos. Esse sistema foi amplamente difundido em toda a Europa até o surgimento do método oral, que emprega a palavra e a leitura labial na transmissão das idéias. A controvérsia entre defensores do método oral e os partidários da linguagem de sinais que se estabeleceu em princípios do século XIX, persiste ainda hoje (ENCICLOPÉDIA BARSA, 1993).

A complexidade do problema e a enorme diversidade de tipos de deficiências auditivas resultam em uma multiplicidade de métodos e propostas. Existem três grandes correntes metodológicas na educação de pessoas com perdas auditivas: o Bilingüismo, a Comunicação Total e o Oralismo.

Para os bilingüistas, partidários da língua de sinais, o surdo não precisa almejar uma vida semelhante ao ouvinte, podendo aceitar e assumir sua surdez (GOLDFELD, 1997). Para o bilingüismo, os surdos formam uma comunidade, com cultura e língua próprias.

Existem duas maneiras da aplicação da posição bilingüista. A primeira acredita que a criança surda deve adquirir a língua de sinais e a modalidade oral da língua de seu país, sendo que posteriormente a criança deverá ser alfabetizada na língua oficial de seu país. Outros autores, como SÁNCHEZ (1993 apud GOLDFELD 1997), acreditam ser necessário para o surdo adquirir a língua de sinais e a língua oficial de seu país apenas na modalidade escrita e não na oral. Bilingüistas como PERLIN (1998) afirmam que a escrita do surdo não vai se aproximar da escrita ouvinte, e não há que se exigir do surdo uma construção simbólica tão natural como a do ouvinte.

Autores como KOLOWSKI (1998) preconizam, entre outros instrumentos, a leitura precoce, desenvolvida a partir de 18 meses, como parte de sua proposta pedagógica, visando garantir ao deficiente auditivo instrumentos para comunicação e estruturação do pensamento.

Outros autores defendem a abordagem da Comunicação Total, na qual a fala e a sinalização devem ser praticadas consistente e congruentemente. O termo genérico de comunicação simultânea parece não poder expressar, com clareza, esses aspectos. Porém, a simultaneidade numa comunicação bimodal, é também enfatizada como forma importante de se considerar aspectos educacionais valiosos (CICCONE, 1996). Para a autora, a melhor forma de linguagem a ser eleita deverá ser aquela que os familiares ouvintes possam aprender com a maior rapidez e a maior facilidade e da qual possam fazer um uso mais confortável, quando a praticarem com os filhos surdos.

Os defensores da Comunicação Total fomentam incentivos à comunicação aproveitando, quando válidas, as alternativas que viabilizem a comunicação quer por sinais acústicos, quer por sinais visuais. Um programa de Comunicação Total não exclui, entre outras técnicas e recursos, a estimulação auditiva, a adaptação de aparelhos de amplificação sonora individual, a leitura labial, a oralização, a leitura e a escrita (CICCONE, 1996).

Os defensores dos métodos orais (PERDONCINI & COUTO-LENZI, 1996; BEVILACQUA & FORMIGONI, 1997) preconizam uma intervenção que tem como proposta priorizar a alteração primária da criança: a audição. O objetivo é auxiliar as crianças a usarem sua audição residual e assim crescerem aprendendo a ouvir e a falar de tal forma que possam aumentar seus conhecimentos, suas experiências de vida e se tornarem pessoas integradas e participantes na sociedade em geral.

Abordagens orais visam o desenvolvimento da linguagem oral, para que a criança com deficiência auditiva possa alcançar melhor desempenho possível da língua falada. Para BEVILACQUA & FORMIGONI (op. cit.), 95% das crianças deficientes auditivas apresentam uma audição residual dos mais variados graus. Assim sendo, entendem que deva ser dada a estas crianças a oportunidade de desenvolverem os resíduos auditivos, por menor que sejam. As crianças com deficiência auditiva podem receber algum tipo de informação sobre o meio ambiente através da audição.

COUTO-LENZI (1997), avaliando audiometrias de 455 sujeitos portadores de surdez severa e profunda, mostrou que todos possuíam uma audição residual aproveitável. Muitos apresentavam audição residual numa faixa de frequências até 2.000 ou 4.000 Hz, que significa a possibilidade de atingir os sons da fala, desde que sejam devidamente aparelhados.

Relacionando o campo auditivo residual e a zona dos sons da fala, COUTO-LENZI (1997) mostrou que em todos os graus de surdez há uma variedade de situações, dependendo das frequências que tenham sido preservadas, não sendo possível pensar-se em um tipo de surdez severa nem profunda, quando se quiser estabelecer um programa de atendimento a um indivíduo com perdas auditivas. Para a educação auditiva e aquisição de linguagem, será da maior importância o conhecimento do campo frequencial do indivíduo.

Uma educação auditiva correta e sistemática é, em todos os casos, o fator mais importante para o sucesso na aquisição da linguagem e no equilíbrio geral do indivíduo (COUTO-LENZI, 1997).

COUTO-LENZI (1999a, 1999b) verificou que mesmo deficientes auditivos com audição residual até 1.000 Hz apresentaram índices de 75% de acerto para detecção da melodia da fala e 52% para inteligibilidade da fala, após a devida educação auditiva. Estes resultados são consistentes com os índices obtidos por ouvintes em contexto da língua portuguesa falada no Brasil, filtrada com banda de zero a 1.000 Hz, e que resultaram índices de 79% para conservação do ritmo e 64% para inteligibilidade (ARAÚJO, VIOLARO & LIMA, 1999).

A educação auditiva (COUTO-LENZI, 1997; BEVILACQUA & FORMIGONI, 1997) e o uso de Aparelhos de Amplificação Sonora Individual (A.A.S.I.) permitem aprimorar as habilidades dos resíduos auditivos, proporcionando melhorias no auto-monitoramento, sendo portanto os mais importantes meios para o desenvolvimento da fala e linguagem em deficientes auditivos.

A aquisição da língua oral, embora reconhecidamente difícil e demorada, é recomendada por quase todos os estudiosos, de quase todas as metodologias propostas para a educação de deficientes auditivos.

Mesmo dentro de uma visão bilingüista, a necessidade da oralização é uma realidade para o surdo brasileiro quando se pensa em educação de nível superior, inserção social e profissional (KOLOWSKI, 1998). Autores como AGUIRRE, GOLDFELD e FERREIRA (1999) seguem a filosofia educacional do bilingüismo e pretendem criar um método de oralização coerente com suas bases teóricas. PIMENTA, RAMOS & SOARES (1998) atuam no modelo bilingüista de educação de surdos, e mesmo reconhecendo que nem todo surdo atinge a expressão oral com clareza, ainda assim acreditam que é imprescindível que todos possam interagir na cultura da maioria da população, buscando sua integração e cidadania.

A Secretaria de Educação Especial (BRASIL, 1995) preconiza a atuação dos serviços de educação especial para atendimento de deficientes auditivos em função do grau de perda auditiva. No caso de surdez leve, em que a criança apresenta dificuldades no período de alfabetização com relação à leitura e/ou à escrita, esta deverá receber um atendimento especializado, constituído das atividades de treinamento fonoarticulatório/fala e linguagem. As recomendações para portadores de surdez moderada são similares às anteriores com maiores atenções.

Para portadores de surdez severa, a Secretaria de Educação Especial recomenda que o aluno poderá freqüentar uma escola especial para adquirir a linguagem interior e receptiva (ou de compreensão) e expressiva (ou de expressão), uma vez que, em geral, não chega a adquirir a linguagem pelos processos naturais. Mesmo nos casos em que o aluno adquire alguma linguagem, esta não será suficiente para sua comunicação oral. Além de adquirir a linguagem oral, esse aluno poderá aprender a ler e escrever na escola especial. Tendo adquirido a linguagem e estando alfabetizado, o aluno poderá freqüentar a classe comum (BRASIL, 1995).

No caso de surdez profunda, o aluno precisará adquirir uma linguagem que permita sua comunicação com os ouvintes, além de possibilitar o desenvolvimento da escolaridade. A alfabetização do aluno com surdez profunda deverá ser realizada paralelamente à estruturação da linguagem oral. Poderão ser adotadas, para atendimento ao portador de deficiência auditiva, diferentes metodologias ou filosofias educacionais: o oralismo, a comunicação total, o bilingüismo e outras. A opção inicial por qualquer dos métodos deve ser feita pelos pais (BRASIL, 1995).

Os especialistas na aquisição da linguagem recomendam a intervenção para desenvolvimento da fala o mais cedo possível (MARCHESI, 1987; COUTO-LENZI, 1991). Para BEVILACQUA & FORMIGONI (1997), a intervenção precoce possibilitará a minimização das dificuldades que esse déficit impõe, não só daquelas que dizem respeito ao processo de comunicação, mas também das dificuldades encontradas durante todo o processo posterior. A precocidade é, portanto, um requisito fundamental nesta aquisição.

Para SHEETS (1971), o bebê surdo se engaja na produção de sons auto-reflexivos durante os primeiros meses de vida mas, devido à ausência da estimulação sonora, irá regredir e gradualmente se tornar silencioso, enquanto que o indivíduo ouvinte irá continuar com a produção dos sons, para o estágio do balbucio.

Nos primeiros meses de vida, mesmo sendo incapaz de ouvir sua própria voz, a criança com deficiência auditiva chora e vocaliza da mesma maneira que uma criança ouvinte. Entretanto, conforme vai ficando mais velha, seu repertório tende a manter-se restrito com respeito à qualidade e à quantidade de sons, e pode não desenvolver o balbucio silábico. No final do primeiro ano, sua vocalização reduz-se aos sons que ela pode sentir, sons guturais ou gritos agudos, similares àqueles produzidos por bebês de um mês de vida. Ela não desenvolve controle suficiente sobre sua voz e fala (SHEETS, 1971).

Entre as técnicas utilizadas para o aprendizado da fala por deficientes auditivos, estão aquelas baseadas nos sentidos preservados, em especial visão e tato (CALVERT & SILVERMAN, 1978).

A importância do sentido cinestésico é também encontrado na literatura (RINALDI, 1997). A palavra cinestesia resulta da fusão de duas palavras gregas: kinein (mover) e aistesis (sensação). O termo sentido cinestésico indica que os movimentos são experimentados através deste sentido. Outra designação para este sentido é sentido proprioceptivo. Este termo indica que os estímulos são recebidos pela células sensoriais no próprio músculo, em seus tendões e nas cápsulas articulares (HOLLE, 1990)

Para RINALDI (1997), a criança deve ser preparada para a emissão dos sons, desenvolvendo o controle da tensão e relaxamento, sua respiração, sensibilidade tátil, mobilidade e propriocepção da região fonoarticulatória.

O relaxamento tem como objetivo geral a aquisição de controle de tensão e relaxamento da musculatura do corpo em geral, com ênfase especial no tórax, pescoço e face. A regra geral na estimulação da sensibilidade e da mobilidade orofacial é o uso e o abuso do tato, brincando com a criança para estimular o uso da voz para produzir vogais e combinações entre vogais e consoantes (RINALDI, 1997).

O objetivo geral do trabalho com a respiração é a aquisição de hábitos corretos, propiciando melhor ritmo de fala e melhor emissão dos fonemas e de uma voz mais natural. Os melhores resultados serão obtidos através de situações lúdicas motivadoras. Entre as atividades sugeridas temos: apagar velas; soprar tiras de papel, penas ou bolinhas de sabão; soprar através de canudos; brincar com apitos e gaitas; cheirar perfumes; falar, expirando, soltando e sustentando a voz; conscientizar-se do ato respiratório, colocando uma mão sobre seu próprio peito e a outra sobre o do professor (RINALDI, 1997).

O tato é utilizado, por exemplo, em exercícios para controle das pregas vocais. Em uma das práticas utilizadas, a criança deve colocar um dedo de uma mão sobre o pescoço do professor e um dedo da outra mão sobre o próprio pescoço, e verificar se as pregas vocais estão vibrando ou não, procurando imitar o professor.

O tato é limitado para determinar variações de frequência. A referência, em termos de frequência, será a do profissional que a estiver orientando (adulto, sexo masculino ou feminino). A frequência de referência poderá ser impossível de ser realizada pela criança, em função das diferenças de idade e/ou sexo, restando-lhe realizar o ato vibratório, sem o ajuste do valor da frequência. As limitações do sentido associadas a referências com padrões inadequados ensejam atividades repetitivas e demoradas, tanto para o fonoaudiólogo quanto para o deficiente, no necessário aprimoramento do controle das pregas vocais.

O desenvolvimento da mobilidade e propriocepção da região fonoarticulatória no educando visam: dar consciência para a existência e o uso da respiração oral; dar informações sobre as vias nasais como instrumento de respiração e de sua correta utilização; dar consciência sobre como a inspiração e expiração devem ser realizadas em movimentos distintos, alternando as duas vias respiratórias; ampliar e coordenar a capacidade fonoarticulatória; dar ciência das partes do corpo e dos diferentes graus de tonicidade muscular; adquirir movimentação e controle articulatorios necessários aos padrões desejáveis de emissão e adquirir condições de articular fonemas, estabelecendo a correta retroalimentação acústica, proprioceptiva, visual e tátil (RINALDI, 1997).

A combinação de crianças deficientes com tenra idade, utilizando sentidos limitados, padrões de referência impróprios, em atividades repetitivas e demoradas, constitui algumas das limitações no desenvolvimento da fala por deficientes auditivos.

## 2.2. A produção da fala e o sistemas de retroalimentação

A produção da fala é constituída por cinco fases: respiração, fonação, articulação, ressonância e retroalimentação (ZEMLIN, 2000).

O processo de produção da fala começa no nível cortical. O processo de pensamento conduz a uma seqüência de impulsos neurais, que são transmitidos para a musculatura do mecanismo respiratório, para a laringe e para a estruturas articuladoras (ZEMLIN, 2000).

Receptores especializados em nossas articulações, tendões e músculos, fornecem informações ao cérebro, sobre o modo como as coisas estão ocorrendo – a retroalimentação. Algumas informações nunca atingem o nível da consciência.

Para ZEMLIN (op. cit.), a retroalimentação auditiva é uma via principal pela qual monitoramos nossa produção sonora. Para o autor, quase todas as interrupções da retroalimentação auditiva resultam em degradação da produção da fala. Isso fica mais evidente na fala das crianças que perderam a audição bem cedo. Uma vez que a fala foi bem estabelecida, o papel da retroalimentação auditiva pode ser diminuído, como demonstram os indivíduos que tiveram perdas de audição graves com mais idade, mas que conseguem manter a articulação adequada, principalmente através da retroalimentação cinestésica.

Existe ainda a interação entre a modalidade motora e outras modalidades sensoriais, que, embora ainda pertençam mais ao inconsciente, controlam todo o nosso mecanismo de produção da fala. Os músculos, tendões e mucosas possuem receptores de estiramento, pressão, táteis e outros, sensíveis e elaborados, que comunicam as informações sobre a

extensão de movimentos, grau de tensão muscular, velocidade do movimento, entre outras. Essas informações voltam para o cérebro e para a medula espinhal, onde são integradas a comandos neurais ordenados em série para os mecanismos dos músculos da fala. Esses receptores apresentam, em sua maior parte, adaptação muito rápida, enviando informações enquanto o movimento está ocorrendo (ZEMLIN, 2000).

Perdas da sensibilidade sensorial, como as que ocorrem quando se aplica um anestésico na cavidade oral (como em tratamento odontológico, por exemplo), reduzem a retroalimentação proprioceptiva, prejudicando a destreza articulatória e a sincronização, embora a fala possa permanecer inteligível.

MENN e STOEL-GAMMON (1997, p.280) afirmam que *“a fala possui um componente de destreza e, como ocorre com qualquer atividade que exige uma destreza, a prática aumenta o controle e a precisão com que um movimento é realizado. Assim, quanto maior a frequência com que um bebê produz os movimentos que moldam o trato vocal para produção de determinados sons e seqüências de sons, mais automáticos se tornam estes movimentos e, por fim, torna-se mais fácil executá-los na produção da fala com significado”*.

Os sistemas de controle biológicos (BORDEN & HARRIS apud ISSLER, 1996) podem ser classificados quanto à rapidez da resposta em:

1. Sistemas internos do sistema nervoso central, que são teoricamente velocíssimos;
2. Propriocepção, bastante rápida no sistema nervoso periférico, capaz de dar conta, na hora, por retroalimentação do movimento e posição dos músculos no controle do ato motor;
3. Audição, o menos rápido de todos, pelas variações de pressão e tato.

De acordo com MYSAK (1988), o sensor auditivo, em circunstâncias normais, monitora o pensamento verbal e controla e monitora a efetiva produção de código verbo-acústico, tanto quanto suas dimensões de tempo e sonorização. Dependendo da idade do indivíduo, o sensor auditivo também faz uma importante contribuição ao controle e monitoração da articulação dos sons da fala. Teoricamente, quanto mais a idade do indivíduo se aproxima e ultrapassa os sete ou oito anos, menos é utilizado o sensor auditivo no controle e monitoria da articulação. A propriocepção coopera com o controle e monitoração da articulação efetiva da fala. Este mecanismo torna-se aparentemente mais importante nesta atividade, logo após a maturação.

### 2.3. A voz

A voz é produzida pelo trato vocal, a partir de um som básico gerado na laringe, o chamado ‘buzz’ laríngeo. Este som ainda não representa a voz que ouvimos. A fonação na laringe gera um som de fraca intensidade, não sendo parecido com nenhuma vogal ou consoante de nossa língua. Contudo, esse som básico, assim que produzido, vai percorrendo um caminho pelo trato vocal, dentro de nosso corpo, passando por determinadas estruturas até atingir a saída pela boca e/ou nariz. Neste trajeto o som é modificado através de um processo chamado ressonância. As cavidades de ressonância, portanto, constituem um amplificador natural de fonação e são formadas pela própria laringe, faringe, boca, nariz e seios paranasais (BEHLAU & PONTES, 1999).

Este conceito de voz explicita a teoria fonte-filtro (FANT, 1960). Por esta teoria, a energia acústica produzida é resultado de um processo de amplificação seletiva realizada

pelo ressonador (ou filtro) sobre a excitação proveniente da fonte. A fonte deve produzir um padrão aproximadamente periódico para a realização de sons sonoros ou um ruído para produção de sons surdos. O padrão periódico é realizado por vibração na laringe pela ação da pressão de ar dos pulmões sobre as pregas vocais e pelos ajustes musculares a que estas estão submetidas. O padrão surdo é obtido com as pregas abertas permitindo a passagem do ar livremente pelo espaço glótico.

Na realização do padrão sonoro, a vibração das pregas vocais produz um espectro sonoro composto de raias com frequências múltiplas da frequência de vibração das pregas. Este espectro decresce em intensidade a uma razão de 12 dB por oitava. A frequência de vibração das pregas vocais é denominada de frequência fundamental.

O filtro, realizado em função da posição assumida pelas estruturas dos tratos vocal e/ou nasal (os articuladores), proporciona uma amplificação seletiva de frequências, as frequências de ressonância ou formantes. Uma formante é um modo natural de vibração (ressonância) dos tratos, responsável pelo timbre característico de cada som. Embora exista um grande número de formantes, para propósitos práticos, somente as primeiras três ou quatro são de interesse para a caracterização dos sons. Cada formante pode ser descrito por duas características: frequência central (chamada de frequência formante) e largura de banda (KENT & READ, 1992).

As mudanças na frequência de vibração para um locutor, ocorrem somente na fonte e não tem necessariamente qualquer efeito no filtro. A relativa independência da fonte e filtro torna possível produzir voz inteligível com uma grande variabilidade na fonte (KENT & READ, 1992).

Para BEHLAU & PONTES (1999), a produção da voz e da fala exige uma série de atos coordenados pelo cérebro, que ocorrem na seqüência: o cérebro dispara um comando

para a laringe e articuladores; inicialmente precisamos inspirar o ar, o que requer um afastamento das pregas vocais; na seqüência as pregas vocais devem aproximar-se entre si, com tensão adequada, controlando e bloqueando a saída de ar dos pulmões; a pressão do ar coloca em vibração as pregas vocais; as caixas de ressonância devem estar ajustadas para amplificar a saída do som pela boca; os articuladores devem ser posicionados adequadamente propiciando as necessárias freqüências de ressonância .

## 2.4. Características das vozes de deficientes auditivos

A voz de deficientes auditivos tem sido objeto de diversos estudos. CALVERT & SILVERMAN (1978) verificaram problemas comuns tais como: omissão de consoantes iniciais, ensurdecimento de consoantes, substituição de nasal por sua oral cognata (ex.: /m/ por /b/), substituições de vogais, hipernasalidade, tendência para prolongar excessivamente as vogais, ditongação de vogais, estridência, elevados valores de freqüência fundamental, irregularidade no ritmo. Os problemas se agravam com o grau da perda auditiva.

Estudos mostram que os deficientes auditivos acham mais inteligíveis vozes com valores baixos de freqüência fundamental, provavelmente devido à estrutura com mais harmônicas (LING, 1978). Para o autor, sons de fricativos como /f/ de **f**aca, /s/ de sapo e /ʃ/ de **ch**uva são geralmente omitidos, distorcidos ou substituídos em decorrência das crianças deficientes auditivas apresentarem reduzida sensibilidade na faixa de freqüências ocupada por esses sons, notadamente acima de 2.000 Hz.

Problemas com a nasalidade, resultante da falta de precisão no grau de fechamento ou velocidade da válvula palatofaríngea, é uma característica freqüentemente incluída como

característica da voz de deficientes auditivos. Outra característica comum entre os deficientes é a utilização de frequência fundamental elevada e com excessiva variação (BORGES, 1998). Para a autora, há indícios de que algumas variações da frequência fundamental, que ocorrem na fala do deficiente auditivo, sejam resultantes de sua tentativa de aumentar a retroalimentação proprioceptiva na atividade da produção de fala.

BORGES (1998) verificou que a qualidade vocal tensa está muito presente. Esta qualidade é decorrente de uma constrição laríngea e, como consequência, os sons das vogais, que são importantes para a inteligibilidade da fala, acabam sendo indistintos. A autora, revendo bibliografias antigas e recentes, chegou à conclusão que há uma semelhança quanto às características das vozes dos deficientes auditivos.

A produção da voz exige a ação coordenada de diversos órgãos adaptados, atuando em sincronismo e com precisão. A fonte de energia precisa gerar um sinal acústico básico com energia e padrão harmônico suficiente para a geração do som. Os articuladores devem assumir uma posição que possibilite a ressonância nas frequências adequadas.

## 2.4.1. Questões de fonte

### 2.4.1.1. A respiração

Para respirar e falar temos que vencer a pressão atmosférica. A respiração é a força-motriz da fonação, realizada durante a expiração. Existem diferenças entre a respiração vital, utilizada para manutenção da vida, e a fônica, utilizada para a comunicação. Em repouso a inspiração corresponde a 40% do ciclo e a expiração corresponde a 60% do ciclo.

Na fala, a duração da expiração se amplia para 90% do ciclo e a inspiração se reduz a 10% do ciclo (ISSLER, 1996). Para BLOCH (1963), na respiração em repouso a relação entre a duração da inspiração e a expiração é quase igual a 1, ao passo que na respiração fônica esta relação pode alcançar de 1 para 10 até 1 para 40.

O tempo máximo de fonação (TMF) é uma das mais simples e importantes medidas fonatórias. É obtido pela medida do tempo máximo que um indivíduo consegue sustentar a emissão de um som, numa só expiração, e permite uma investigação quantitativa e qualitativa da fonação. Em testes com vogais sustentadas, produzidas por falantes da cidade de São Paulo, BEHLAU & PONTES (1995) encontraram uma duração média de 20 s para homens e 14 s para mulheres. Para crianças até a puberdade, os valores de tempo máximo de fonação acompanham em média o valor do número de anos. A importância dos tempos máximos de fonação reside no fato do falante, na fala encadeada, realizar recargas aéreas a cada um terço de seu tempo máximo de fonação. Para BLOCH (1963), o maior segredo da boa voz está na coordenação da respiração com a emissão sonora.

BORGES (1998) relacionou a dificuldade do deficiente auditivo na coordenação respiratória, como caracterizada pela menor produção de sílabas por ciclo respiratório. O deficiente auditivo inspira em intervalos de somente uma, duas ou três palavras, e expira gastando mais sobre cada sílaba, quando comparado com a respiração do ouvinte. Para a autora, o fluxo de ar insuficiente e a pressão subglótica perturbam a aerodinâmica da vibração, requerendo maior esforço muscular do que o normal e/ou aumentando a tensão vocal para a fonação. A autora concluiu que a fala do deficiente auditivo é fisicamente ineficiente e, por essa razão, mais cansativa. Há um desperdício de esforço que acaba resultando numa péssima inteligibilidade de fala.

CARNEY, HIGGINS & SCHULTE (1994, apud FRANCO 1998) observaram que, apesar de alguns indivíduos deficientes auditivos terem boa inteligibilidade de fala, apresentam anormalidades na voz (soprosa, tensa, frequência fundamental elevada, ressonância *cul-de-sac*). Os autores salientaram ainda que o pouco controle de expiração do ar leva o deficiente auditivo a cansar-se rapidamente ao falar, apresentando, conseqüentemente, um ritmo alterado da fala.

FORNER & HIXON (1977, apud BORGES 1998) verificaram que os deficientes auditivos apresentam as funções respiratórias, durante o repouso, dentro dos limites dos ouvintes, mas que ao falarem apresentam, associadas às dificuldades lingüísticas, alterações no ajuste da função respiratória, nas funções da laringe e das vias aéreas.

PERELLO & TORTOSA (1978, apud FRANCO 1998) afirmaram que não há coordenação da inspiração e expiração quando o deficiente auditivo fala.

ARAÚJO & LIMA (1999) mediram o tempo máximo de fonação de deficientes auditivos pré-adolescentes e verificaram que o TMF médio estava 42,5% abaixo do padrão apresentado por locutores ouvintes na mesma faixa etária.

#### 2.4.1.2. A frequência de vibração das pregas vocais

BOONE (1996, apud FRANCO, 1998) observou que os deficientes auditivos apresentam nível mais alto de frequência fundamental, especialmente os do sexo masculino. Em relação às adolescentes deficientes auditivas, observou uma variação entre elas, pois algumas apresentavam registro de frequência fundamental normal e outras muito

alto. Concluiu que o uso impróprio de frequência fundamental alta ocorre mais no adolescente do que na criança com deficiência auditiva.

MONSEN (1979, apud FRANCO 1998) observou que o controle da frequência fundamental parece ser o mais importante fator para diferenciar o melhor do pior falante surdo. Para que a qualidade de voz seja considerada boa, parece ser necessário não somente que o registro da frequência fundamental esteja dentro de uma faixa apropriada, mas que o controle seja realizado de uma maneira adequada.

FRANCO (1998), avaliando a voz de 89 sujeitos de ambos os sexos, com deficiência auditiva, verificou que 34,8% apresentavam qualidade tensa, 23,6% soprosa, 23,6% faríngea, 16,9% nasal e 1,1% normal. Observou que sujeitos do sexo masculino apresentaram piores alterações do controle da frequência fundamental, duração, prosódia e menores índices de inteligibilidade. No controle de expiração do ar, 62,9% apresentaram problemas severos. Observou ainda que, embora a maioria dos sujeitos não apresentassem o comportamento de evitar falar, os sujeitos que assim procediam apresentaram os piores níveis de controle de frequência fundamental, enquanto que os que não evitavam falar tinham melhores níveis de controle de frequência fundamental, o mesmo ocorrendo com a inteligibilidade.

ROBB & SAXMAN (1985), na síntese de vários trabalhos com ouvintes, mostraram que, ao nascimento, a frequência fundamental média é de 440 Hz, aos 10 meses está em torno de 400 Hz e, aos 2 anos, atinge um valor próximo a 300 Hz. Para COLTON & CASPER (1990), a frequência fundamental mantém-se entre 300 e 270 Hz até os 10 anos, reduzindo-se para o valor médio de 224 Hz para mulheres ouvintes adultas e 132 Hz para homens ouvintes adultos. No Brasil, BEHLAU & PONTES (1995) obtiveram para homens,

mulheres e crianças ouvintes, valores para frequência fundamental de 113 Hz, 205 Hz e 236 Hz respectivamente.

A produção da fala requer da fonte um padrão harmônico adequado. Na realização da vogal /u/ por um locutor masculino, com a primeira formante próxima de 300 Hz (PETERSON & BARNEY, 1952), é necessário que o padrão harmônico produzido pelas pregas vocais contemple frequências fundamentais menores ou iguais a 300 Hz, tal que o padrão produzido na laringe possa ser efetivamente amplificado pelas cavidades ressonantes. Caso a frequência fundamental seja da ordem de 400 Hz, por exemplo, a componente amplificada será de 400 Hz, produzindo um som mais próximo ao da vogal /o/ (com formante próximo a 400 Hz) do que o da vogal /u/. Quanto menor for uma frequência formante, menor deverá ser a frequência fundamental, tal que o perfil harmônico seja capaz de representar adequadamente a formante, facilitando a inteligibilidade.

A vogal /a/ apresenta a primeira frequência formante mais elevada de todas as vogais, sendo, portanto, a menos exigente em termos de frequência fundamental. É, aos 18 meses, a única vogal usada corretamente em 80% ou mais na fala (BOONE & PLAUTE, 1994). A simplicidade da posição articulatória aliada à pouca exigência em termos de frequência fundamental, facilitam a produção, monitoração e, conseqüentemente, o domínio desta vogal.

Aos 2 anos, praticamente todas as vogais são utilizadas corretamente em 80% ou mais da produção de fala de crianças normais (BOONE & PLAUTE, 1994).

ARAÚJO & LIMA (1999) mediram a frequência fundamental de deficientes auditivos pré-adolescentes e obtiveram valores médios de 368 Hz, portanto 56,6% acima da média apresentada por ouvintes na mesma faixa etária. Este valor de frequência

fundamental, inadequado para a realização de vogais fechadas como /i/ e /u/, resulta um padrão harmônico restrito, que dificulta a inteligibilidade.

LING (1989, apud BORGES, 1998) explicou que a causa básica e óbvia sobre a falha na fala do deficiente auditivo é a percepção inadequada da retroalimentação e a qualidade insuficiente de padrões de fala e de como eles são produzidos.

Paradoxalmente os deficientes auditivos utilizam frequência fundamental alta na sua produção oral embora considerem mais inteligíveis vozes de outras pessoas com frequências fundamentais mais baixas. Este paradoxo talvez possa ser entendido avaliando-se o efeito psicofísico da variação da frequência fundamental. As curvas de igual audibilidade obtidas por Fletcher & Munson (NEPOMUCENO, 1977) mostram que há um ganho de sensação de até 40 dB aumentando-se a frequência de 100 Hz para 1.000 Hz. O deficiente auditivo, aumentando a frequência fundamental, aproxima a primeira harmônica da voz, em geral com energia significativa, de uma região de maior sensibilidade auditiva. Portanto, aumentar a frequência fundamental implica em ampliar a sensação auditiva e, em consequência, a retroalimentação auditiva, facilitando para o deficiente auditivo detectar a presença (ou ausência) de sua voz, ainda que o padrão gerado seja pouco inteligível.

É notável como as crianças ouvintes realizam, entre 10 e 24 meses, um significativo ajuste da frequência fundamental, aproximando-o do padrão necessário para a produção dos sons que são realizados com baixas frequências formantes. Neste período, a criança ouvinte sai do balbúcio para a fala.

As crianças, dispendo de tratos vocais menores e, conseqüentemente, frequências formantes mais elevadas, podem realizar a fala com frequências fundamentais mais elevadas. Mulheres adultas com tratos de dimensões médias, que estão associados a frequências formantes intermediárias, realizam fala inteligível com frequências

fundamentais médias. Homens adultos, com tratos de maiores dimensões freqüências formantes mais baixas, precisam utilizar freqüências fundamentais menores para produzir fala mais inteligível. É o natural ajuste da fonte ao filtro, pelo natural crescimento e modificações hormonais.

Deficientes auditivos adolescentes do sexo masculino, com tratos oral e nasal em crescimento acelerado, mais adequados para produção de ressonâncias em freqüências cada vez mais baixas, e produzindo freqüências fundamentais elevadas, haverão de ter mais dificuldades de produzir fala inteligível, tal como foi observado por FRANCO (op. cit.).

O controle da fonte, tanto no que se refere à coordenação respiratória, quanto no ajuste da produção harmônica, é essencial para a produção de sons inteligíveis que se adequem aos propósitos da comunicação oral efetiva.

#### 2.4.2. Questões de filtro

Do ponto de vista da fonética, a vogal /a/, por exemplo, é central e produzida com os lábios distensos, a boca aberta e a língua abaixada (CÂMARA JR.,1991). A vogal /a/ pode, na verdade, ser realizada com infinitos posicionamentos. Movimentos compensatórios permitem realizar essa vogal, por exemplo, com a ponta da língua para fora da boca ou no palato. A posição definida pela fonética corresponde a uma produção média, de mínimo esforço articulatório, que requer menor utilização de músculos e movimentação de massa. Essa posição média é obtida no processo de aprendizado, através do monitoramento auditivo e reajuste fonoarticulatório.

O filtro, representado pelos articuladores, deve assumir uma posição que possibilite as ressonâncias nas frequências necessárias para a produção de cada som, desde que a fonte gere um sinal acústico com energia e padrão harmônico compatível.

Encontrar a posição de mínimo custo articulatorio para cada som, com a coordenação explicitada por BEHLAU & PONTES (1999), aprimorando a destreza preconizada por MENN & STOEL-GAMMON (1997), propiciando o aprendizado e cooperação efetiva do sensor tátil-cinestésico conforme MYSAK (1988), exige uma longa e necessária prática, essenciais à efetiva comunicação oral. Para isso, são necessárias condições para que a criança obtenha o prazer do exercício exploratório de suas potencialidades de acordo com a concepção de ALBANO (1990).

Valendo-se da brincadeira e do jogo exploratório, a criança pode encontrar as posições articulatorias mínimas, aprimorar a destreza, coordenação e cooperação necessárias à produção de cada som e suas conexões, propiciando condições para a fala articulada.



## *Capítulo 3*

# Os modelos e objetivos dos jogos

O desenvolvimento de jogos computacionais para exercícios de aprimoramento da coordenação fonoarticulatória, visando o auxílio a deficientes auditivos, exige a definição dos modelos dos exercícios a serem realizados, as aplicações, a adequação a cada estágio de controle articulatório e o modelagem sobre o sinal de voz que propicie medidas adequadas e precisas para a consecução dos jogos.

A definição dos exercícios constitui um dos principais objetivos desse trabalho. Procurou-se estabelecer situações lúdicas em que produções sonoras inadequadas possam ser corrigidas, aproximando-as de padrões acústicos adequados a uma comunicação efetiva. É fundamental que os aplicativos possam adequar-se ao nível de controle articulatório da criança e possam ser ajustados conforme esse controle se aprimore.

A modelagem sobre o sinal de voz deve fornecer parâmetros adequados para o controle do jogo. Esses parâmetros devem ser suficientemente precisos, de modo a garantir para a criança uma retroalimentação conseqüente.

### 3.1. Características fundamentais para produção da fala

A voz é resultante da ação dos articuladores exercendo uma filtragem seletiva sobre a excitação produzida pela fonte.

A fonte deve ser capaz de produzir uma excitação adequada tanto em conteúdo espectral quanto em duração. Os articuladores precisam adquirir a velocidade e precisão suficientes para a produção de sons inteligíveis.

A excitação pode sofrer ação vibratória das pregas vocais, gerando os fones sonoros, caso contrário a excitação (e o sinal) será surda. Os diversos movimentos e posicionamentos articulatorios produzem amplificação e/ou atenuação seletiva (ressonância e/ou anti-ressonância) das frequências da excitação.

Os sons da fala são constituídos de vogais e consoantes. As vogais se caracterizam pelo livre fluxo do ar no trato vocal, enquanto que as consoantes se caracterizam pela presença de obstruções ou constrictões no trato vocal.

#### 3.1.1. As vogais

Uma vogal é realizada por um fluxo de ar contínuo proveniente dos pulmões até a laringe, onde as pregas vocais vibram produzindo um trem de pulsos de ar quase periódico. As várias vogais são resultados de posicionamentos diversos dos tratos oral e nasal. As vogais podem ser realizadas com o ponto de articulação próximo aos lábios (anteriores), no centro da boca (medial) ou distante dos lábios (posteriores). Podem ser realizadas com a língua mais elevada (alta) ou mais abaixada no interior da boca (baixa). As vogais podem

ser nasais quando o véu se abaixa e o ar escapa pelas cavidades oral e nasal. No português falado no Brasil (PB) existem 7 vogais orais e 5 vogais nasais.

A tabela 3.1 apresenta as 7 vogais orais tônicas do PB em função do ponto de articulação e altura da língua

ALTURA DA LÍNGUA	PONTO DE REALIZAÇÃO		
	Anterior	Medial	Posterior
Alta	/i/ de <i>ivo</i>		/u/ de <i>uva</i>
Variante de 2º grau	/e/ de <i>ele</i>		/o/ de <i>ovo</i>
Variante de 1º grau	/ɛ/ de <i>ela</i>		/ɔ/ de <i>ova</i>
Baixa		/a/ de <i>ata</i>	

Tabela 3.1: Vogais tônicas do português falado no Brasil

### 3.1.2. As Consoantes

As consoantes são realizadas quando o fluxo de ar que sai dos pulmões encontra obstáculos na altura da boca. Nas oclusivas, o obstáculo é total, enquanto nas constrictivas a obstrução é parcial. A realização pode ser com fricção (fricativas) ou sem fricção. Nas sonoras temos a vibração das pregas vocais, caso contrário temos as surdas. O abaixamento do véu permite a realização das nasais enquanto as orais são produzidas com o véu levantado. A obstrução pode ser de ordem labial, como as bilabiais e labiodentais, de ordem anterior, como as linguodentais e alveolares, e de ordem posterior, como as palatais e velares. Essas características são combinadas permitindo a realização dos diversos tipos de consoantes. No PB, utilizamos basicamente 19 consoantes.

A Tabela 3.2 apresenta as consoantes da língua portuguesa em função do modo de articulação, do papel das pregas vocais, do papel das cavidades e do ponto de articulação.

Modo de Articulação		Oclusivas			Constritivas			
		Plosivas		Nasais	Fricativas		Laterais	Vibrantes
Papel	das pregas	Surdas	Sonoras		Surdas	Sonoras		
	das cavidades	Oral		Nasal	Oral			
Ponto de articulação	Bilabiais	/p/ <i>pato</i>	/b/ <i>bola</i>	/m/ <i>amor</i>				
	Lábilodentais				/f/ <i>fofo</i>	/v/ <i>vovó</i>		
	Linguodentais	/t/ <i>tatu</i>	/d/ <i>dado</i>	/n/ <i>ano</i>				
	Alveolares				/s/ <i>selo</i>	/z/ <i>zelo</i>	/l/ <i>mala</i>	/r/ <i>caro</i>
	Palatais			/ɲ/ <i>manha</i>	/ʃ/ <i>chato</i>	/ʒ/ <i>jeito</i>	/ʎ/ <i>malha</i>	
	Velares	/k/ <i>casa</i>	/g/ <i>gata</i>					/R/ <i>carro</i>

Tabela 3.2: Consoantes da língua portuguesa falada no Brasil.

### 3.2. Os jogos computadorizados

Na consecução de jogos, os modelos matemáticos precisam ser adequados aos objetivos dos exercícios a serem realizados e devem ser suficientemente precisos de forma a garantir uma adequada retroalimentação para a criança.

Os jogos estão divididos em dois grupos, os jogos para exercícios de aprimoramento de fonte e os jogos para aprimoramento do posicionamento articulatorio ou jogos para exercícios do filtro.

### 3.2.1. Jogos para exercícios de fonte

Nesta classe de exercícios, três características são fundamentais, o controle do fluxo expiratório – tempo máximo de fonação, o controle da intensidade e o controle da frequência de vibração das pregas vocais.

#### 3.2.1.1. Aprimoramento do controle do fluxo respiratório

O exercício para aprimoramento do controle do fluxo inspiratório-expiratório visa proporcionar ao deficiente auditivo maior domínio na utilização da fonte de energia do som – o ar armazenado nos pulmões.

Quando falamos, liberamos o ar lentamente do pulmão, excitando o trato e produzindo a fala. A cada pausa realizamos uma inspiração para reabastecer o pulmão de ar, propiciando ao organismo fonte vital de oxigênio e mais energia para produção da fala.

Em repouso, a duração da inspiração e da expiração são quase iguais. Na fala, a duração da expiração se amplia e a duração da inspiração se reduz

Ao longo da primeira infância aprendemos que, durante a fala, devemos inspirar rapidamente o ar e que este deve ser expirado lentamente durante a produção da fala. Este é um processo que exige um certo tempo para ser adquirido, mesmo para crianças ouvintes, que inicialmente falam recortando as palavras, como pa-pa, precisando inspirar no meio da palavra. Somente após o amadurecimento, a criança será capaz de produzir uma palavra ou frase de forma contínua, sem reabastecimento da fonte – a inspiração.

Os exercícios para aprimoramento do controle respiratório para produção da fala visam estimular o aumento do tempo máximo de fonação, ou seja, aumentar o tempo máximo que a criança pode manter o fluxo expiratório, respeitando o ar de reserva.

Algumas das técnicas utilizadas para o controle de inspiração-expiração consistem em soprar uma vela e exercícios de mergulho em piscina com expiração sob a água. Em qualquer caso, as crianças são estimuladas a aumentar a duração do tempo em que sopram a vela e/ou permanecem expirando sob a água na piscina.

Em computador, os exercícios para aprimoramento do controle de respiração, para a produção da fala, devem também estimular o aumento do tempo máximo de fonação e a conseqüente redução do tempo de inspiração.

Os exercícios propostos, por exemplo, permitem que a criança mova um objeto durante a expiração e realize movimento contrário durante a inspiração (por exemplo, subir e descer). Na expiração, com voz ou sopro, o objeto sobe, aumentando a pontuação. Na inspiração o objeto cai, reduzindo a pontuação. Para aumentar a pontuação, a criança deve aumentar a duração do tempo de expiração e reduzir o tempo de inspiração.

O jogo proposto neste trabalho apresenta níveis de dificuldade que estão associados à taxa com que diminuem os pontos durante a inspiração. Níveis de dificuldades menores reduzem mais lentamente a pontuação durante a inspiração (queda do objeto), ao passo que níveis maiores de dificuldade reduzem mais rapidamente a pontuação durante a inspiração.

O parâmetro utilizado para controle deste jogo é o valor da energia do sinal, calculado sobre quadros de 44 ms e convertido em decibéis, sendo o valor da energia por quadro submetido a um filtro passa baixas para suavizar sua variação. No capítulo 4, serão fornecidos os detalhes sobre os procedimentos para o cálculo deste parâmetro.



As nuvens e o sol são apenas animações que movem-se em sentido contrário ao pássaro, não interferindo sobre sua movimentação.

### 3.2.1.2. Aprimoramento do controle da energia

As limitações dos deficientes auditivos reduzem o aprendizado do controle da energia de sua produção sonora. O nível médio da intensidade dos sons da fala a 1 m dos lábios do locutor ouvinte é de 65 dB NPS. Vozes com baixa intensidade apresentam intensidade de 45 dB NPS e vozes muito intensas apresentam intensidade de 85 dB NPS. Os níveis médios de intensidade de vozes produzidas pelas mulheres estão 3 dB abaixo dos níveis médios produzidos por locutores do sexo masculino (CALVERT & SILVERMAN, 1975).

Os deficientes auditivos precisam adquirir, via pista tátil-cinestésica e/ou via audição residual, o controle sobre o nível de energia de sua produção sonora.

Os exercícios em computador para aprimoramento do controle do nível da energia para produção de fala visam aprimorar os sistemas sensoriais, possibilitando que o sistema tátil-cinestésico e/ou a audição residual possam reconhecer se o nível de energia da produção está (ou não) adequado.

Assim, como os jogos para aprimoramento do controle respiratório, os jogos para controle de energia são baseados em medidas de energia em dB tomadas sobre quadros de 44 ms do sinal de voz.

Os jogos em computador propostos para aprimoramento do controle do nível da energia da produção sonora são exercícios em que um estímulo positivo ocorre quando a

intensidade está na faixa adequada, e um estímulo negativo aparece quando a produção está fora da faixa adequada. Além disso, a criança deve aprender a modular a voz, aumentando ou reduzindo a energia, dependendo do contexto e/ou da ênfase necessária à sílaba (tônica ou átona) ou à curva melódica da frase (afirmativa, negativa ou interrogativa).

São propostos, inicialmente, dois jogos, um no qual o nível filtrado da energia em decibéis é utilizado para mostrar uma imagem positiva (por exemplo, uma moça bonita) quando a energia está na faixa especificada, e uma imagem negativa (por exemplo, uma bruxa) quando o nível de energia está acima (ou abaixo) da energia especificada.

A Figura 3.2 apresenta uma seqüência de telas do jogo para aprimoramento do controle do nível de energia proposto. No quadro 2.a, o nível de energia está abaixo do nível mínimo especificado. No quadro 2.b, o nível de energia esta dentro da faixa de valores especificados. No quadro 2.c, o nível de energia está acima do nível máximo fixado.



Figura 3.2: Telas do jogo para controle do nível de energia.

Os níveis mínimos e máximos precisam ser ajustados em função do ganho da placa de aquisição e da sensibilidade do microfone.

Neste jogo, o principal objetivo é que a criança, mantendo a energia no nível adequado, ou seja, mantendo a figura mais bonita na tela, tenha tempo para adquirir as informações táteis-cinestésicas e/ou auditivas (residuais) necessárias ao aprendizado do controle do nível de energia por uma dessas vias ou por ambas as vias.

Além de aprender a manter o nível de energia em valores adequados, o deficiente auditivo precisa aprender a modular sua voz, ampliando (ou reduzindo) a energia, conforme a necessidade. A Figura 3.3 mostra uma tela do jogo proposto para aprimoramento do controle de energia visando a modulação da voz.



Figura 3.3 : Tela do jogo para controle de energia visando a modulação da voz

Neste jogo, o nível de energia controla a altura da nave espacial na tela, que move-se da esquerda para a direita, enquanto os planetas e astronautas movem-se da direita para a esquerda. A nave deve recolher os astronautas, aumentando a pontuação, e deve desviar-se dos planetas, evitando diminuir a pontuação. Para isso é necessário aumentar ou diminuir o nível de energia da voz.

### 3.2.1.3. Aprimoramento do controle da freqüência fundamental

Os sons da língua portuguesa são, em sua grande maioria, constituídos de sons sonoros, ou seja, são sons realizados com vibração das pregas vocais. Estudos sobre a estatística da língua mostram que a vibração das pregas vocais ocorre em mais de 83% dos fonemas utilizados na língua portuguesa falada no Brasil (ALCAIM, SOLEWICZ & MORAES, 1992).

A função vital das pregas vocais é auxiliar na proteção de entrada dos pulmões, fechando-se durante a ingestão de alimentos e evitando que os mesmos possam ir para o interior dos pulmões que, não dispondo de meios para eliminá-los, os reteriam, e seriam gravemente afetados pela presença de restos alimentares no seu interior. A utilização da vibração das pregas para produção de fala é um mecanismo adaptado há poucos milhares de anos pelo homem.

No Capítulo 2, discutiu-se a importância do aprimoramento do controle da freqüência de vibração das pregas vocais, a freqüência fundamental. As técnicas mais utilizadas para controle da freqüência de vibração das pregas vocais são baseadas no tato e/ou na audição residual.

O sentido do tato é trabalhado colocando-se um dedo de uma mão da criança no pescoço do instrutor e um dedo da outra mão da criança no seu próprio pescoço. O instrutor deve realizar uma sonorização e a criança deverá sentir a vibração do instrutor e tentar imitá-lo. O tato possibilita verificar a existência (ou não) da vibração das pregas, mas é limitado para realizar um ajuste adequado da freqüência – tanto pela característica do sensor quanto pelo modelo fornecido (normalmente de locutor adulto).

A adequação da frequência fundamental é, mesmo para ouvintes, um processo gradual. Um instrutor, utilizando o tato para ajuste da frequência fundamental, precisaria de grande habilidade para produzir uma frequência modelo com valor entre a frequência alvo e a frequência de produção da criança, sendo ainda capaz de inferir os valores da sua própria frequência e da criança, procedimento este extremamente difícil. Em geral, as crianças com deficiência auditiva aprendem a utilizar a vibração das pregas vocais sem o adequado ajuste do valor da frequência.

A utilização das pregas vocais na fala é um mecanismo adaptado, não essencial à produção da fala, como no caso da produção de fala sussurrada. As pregas vocais, ao vibrarem, produzem um espectro composto de harmônicas múltiplas da frequência fundamental, concentrando a energia proveniente dos pulmões sobre as harmônicas, propiciando uma amplificação do sinal. No caso da voz sussurrada (sem utilização da vibração das pregas vocais), a energia da excitação e da fala se distribui sobre todo o espectro, enquanto que na fala normal, que utiliza a vibração das pregas vocais, a energia se concentra nas frequências múltiplas da frequência fundamental.

A intensidade de um sussuro forte é 20 dB menor que a intensidade da fala de conversação, sendo também um modo muito perdulário de utilizar o suprimento de ar. Para um indivíduo que possa fonar por até 30 s durante a vocalização, a duração do sussuro pode ser sustentada por apenas 10 s antes da ocorrência de outra inspiração (ZEMLIN, 2000).

Os jogos com exercícios para aprimoramento do controle das pregas vocais visam fazer com que os sistemas proprioceptivo e/ou auditivo do deficiente auditivo aprendam a reconhecer se a frequência utilizada está ou não adequada aos padrões da fala.

Devem ser consideradas duas situações distintas, uma na qual a frequência produzida está muito acima (ou muito abaixo – vocal *fry*) dos valores médios para sua

idade e sexo, e outra na qual a frequência está mais próxima aos valores médios utilizados pelos ouvintes de mesma idade e sexo. ARAÚJO & LIMA (1999) chegaram a medir a frequência fundamental de 680 Hz em deficiente auditivo pré-adolescente, quando o padrão médio ouvinte correspondente é de 236 Hz (BEHLAU & PONTES, 1995). BOMMARITO (2000) mediu a frequência fundamental de 140 Hz em deficientes auditivos do sexo feminino.

Para aprimoramento do controle da frequência fundamental, são propostos jogos nos quais a pontuação aumenta quando a frequência de vibração das pregas vocais medida sobre uma locução sonora sustentada estiver na faixa especificada, e a pontuação diminui quando a frequência estiver fora da faixa de frequências alvo.

A Figura 3.4 apresenta a tela de um jogo para iniciar o aprimoramento do controle da frequência fundamental. Neste jogo, o objeto (por exemplo, um helicóptero) subirá quando a frequência da locução estiver na faixa de frequências especificada, e descerá quando a frequência da locução estiver acima (ou abaixo) da frequência máxima ou mínima especificada.



Figura 3.4: Tela do jogo para iniciar o aprimoramento do controle das pregas vocais

Caso a frequência máxima da criança esteja muito acima (ou muito abaixo) da frequência padrão médio do ouvinte de mesmo sexo e idade, deve-se fixar valores mínimos e máximos compatíveis da frequência que permitirão à criança o controle do objeto. Gradualmente, de acordo com o progresso da criança, deve-se alterar os valores máximos e mínimos na direção dos padrões ouvintes médios para mesma idade e sexo (LIMA, ARAÚJO & ROMANO, 1999).

Após a devida aproximação da frequência fundamental dos padrões ouvintes médios para a mesma idade e sexo, o jogo de alvo pode ser utilizado para fixar a frequência fundamental em uma faixa adequada.

A Figura 3.5 apresenta a tela de um jogo para aprimoramento do controle da frequência fundamental.



Figura 3.5: Telas do jogo de aprimoramento do controle da frequência fundamental

No quadro (a) da Figura 3.5, a frequência da fonação está na faixa especificada. Nesse caso, os objetos (frutas) lançados pela árvore atingem o centro do alvo, aumentando a pontuação. No quadro (b) da mesma Figura 3.5, a frequência está acima da frequência fixada e os objetos não atingem ao alvo, diminuindo a pontuação. Neste jogo o alvo

corresponde a uma pequena faixa de frequências em torno da frequência fixada na opção de configuração.

No caso em que a frequência está muito acima (ou muito abaixo) dos padrões ouvintes de mesmo sexo e idade, certamente o tiro ao alvo poderá se constituir em um exercício com exagerado grau de dificuldade, praticamente impossível de ser realizado, havendo necessidade de um jogo de controle mais fácil que poderá ser realizado na fase inicial, o jogo do helicóptero.

### 3.2.2. Exercícios de filtro

Os exercícios de filtro são destinados ao aprimoramento dos posicionamentos articulatórios. Estes posicionamentos devem ser tais que possibilitem as ressonâncias nas frequências necessárias para a produção de cada som.

Os ouvintes utilizam a retroalimentação auditiva para monitorar sua produção e, gradualmente, vão ajustando-a ao padrão de sua língua, brincando com a voz desde o balbúcio, estimuladas pela família, até alcançar a fala. Na deficiência auditiva, a retroalimentação é limitada, prejudicando o aprendizado (ARAÚJO, 1999).

Os exercícios de filtro visam propiciar condições para que a criança: 1) realize o exercício exploratório de suas potencialidades; 2) adquira a destreza necessária para realização da fala; 3) adquira a coordenação fonoarticulatória; 4) municie o sistema tátil-cinestésico e/ou auditivo residual de informações necessários para seu aprendizado.

O objetivo dos exercícios para controle do filtro é o aprimoramento do controle do posicionamento dos articuladores, permitindo que a criança encontre a sua posição de mínimo custo articulatório para cada fonema e aprimore seus sistemas sensoriais.

As possibilidades do trato oral são virtualmente infinitas. Cada língua seleciona um conjunto de possibilidades aos quais associa um fonema. Além disso, um mesmo fonema pode ser realizado com posições articulatórias distintas por movimentações compensatórias em outras posições do trato.

O auto-monitoramento é a base para a determinação do posicionamento articulatório adequado. Realizações baseadas apenas em descrições de posicionamento haverão de resultar em fala com movimentos compensatórios exagerados, pois utilizam posições fixadas arbitrariamente por instrução e não obtidas na busca de uma posição articulatória confortável.

O auto-monitoramento articulatório deve, sempre que possível, ser realizado com a utilização de Aparelhos de Amplificação Sonora Individual (A.A.S.I) e a adequada educação auditiva (COUTO-LENZI, 1997; BEVILACQUA & FORMIGONI, 1997).

Conquanto a audição residual deva ser a principal via para o auto-monitoramento articulatório, existem fonemas cujas frequências mais relevantes estão fora da faixa de frequências de que dispõe o deficiente auditivo.

ARAÚJO, VIOLARO & LIMA (1999), estudando a inteligibilidade da língua portuguesa falada no Brasil, limitada a frequências abaixo de 1.000 Hz, verificaram que as vogais posteriores podem efetivamente ser discriminadas nessa faixa, enquanto as vogais anteriores se projetam acusticamente sobre as posteriores com mesmo grau de abertura. Por exemplo, nessa faixa de frequências, a vogal /i/ é discriminada como sendo a vogal /u/.

A discriminação dos fonemas por deficientes auditivos é efetiva, embora possa existir confusões entre sons cognatos (/p/, /t/ e /k/; /b/, /d/ e /g/; /m/ e /n/) (ERBER, 1981)

Jogos computacionais para aprimoramento do controle articulatório são basicamente instrumentos para auto-monitoramento da voz. Neste caso, a retroalimentação é realizada na tela do computador. A criança pode aprimorar a fonação e obter naturalmente uma posição articulatória mais confortável, a qual permitirá maior controle do jogo com menor esforço articulatório (ARAÚJO, 1999).

A importância da língua como articulador da fala é consubstanciada pela utilização da palavra língua para referir-se ao órgão e ao idioma. A língua tem mais fibras sensoriais de tato que qualquer outra parte do corpo (ISSLER, 1996). Para SÁNCHEZ (1973) não existe órgão tão móvel quanto a língua, a qual atua na mastigação, na deglutição, na sucção, na articulação dos fonemas, na produção da voz, etc.

HARDCASTLE (1976, apud ZEMLIN 2000) relaciona apenas sete parâmetros articulatorios que podem ser responsáveis pela grande variedade de posições e configurações da língua durante a fala, a saber: 1) movimento horizontal para frente e para trás do corpo da língua; 2) movimento vertical para cima e para baixo do corpo da língua; 3) movimento para frente e para trás do ápice-corpo da língua; 4) movimento vertical para cima e para baixo de ápice-corpo da língua; 5) configuração perpendicular transversal do corpo da língua, convexo-côncava com relação ao palato; 6) configuração perpendicular transversal estendendo-se por todo o comprimento da língua, em especial no ápice e no corpo – grau de sulcamento central; 7) plano superficial do dorso da língua.

As vogais são menos complexas e utilizam basicamente os dois primeiros parâmetros. Já as fricativas, como o /s/, exigem a participação máxima de todos os parâmetros articulatorios (ZEMLIN, 2000).

### 3.2.2.1. Exercícios com vogais

As vogais correspondem a 54,66% dos sons produzidos na língua portuguesa falada no Brasil (ALCAIM, SOLEWICZ & MORAES, 1992). Os sistemas vocálicos variam de uma língua para outra, exigindo um modelagem próprio de acordo com suas características específicas.

Todos os sistemas vocálicos do mundo estão construídos sobre uma dupla oposição entre, por um lado, o agudo (energia significativa em alta frequência) e grave (energia significativa em baixa frequência) (i-u), e, por outro lado, o difuso e o compacto (i-a). Muitas línguas utilizam apenas uma série de vogais agudas (por exemplo, o português, o italiano, o espanhol, o inglês) (MALMBERG, 1954).

É usual simbolizar esquematicamente a zona de articulação das vogais na boca por uma figura geométrica que, em português, para as tônicas, de acordo com CÂMARA JR (1970), terá a seguinte forma.

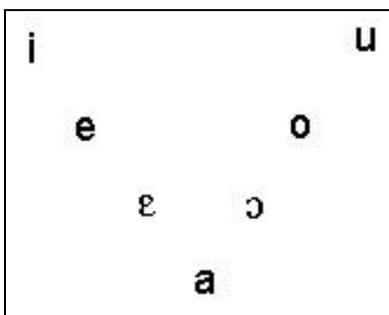


Figura 3.6. Triângulo fonético das vogais tônicas para o PB

O triângulo fonético apresenta relação inequívoca com a realidade fonética, pois é traduzível em termos tanto articulatórios como acústicos. A versão articulatória é obtida medindo-se na ordenada a altura do ponto de constrição máxima e na abscissa a distância da constrição máxima aos lábios. Analogamente, a versão acústica é obtida medindo-se na

ordenada o inverso da frequência da primeira formante e na abscissa o inverso da frequência da segunda formante (MAIA, 1991)

A língua francesa se serve de quatro graus (quatro níveis, como na figura 3.6) na série vertical e há outras línguas que utilizam mais ou menos graus. Algumas línguas utilizam também séries intermediárias entre o agudo e o grave (por exemplo, o inglês e o sueco). Outras línguas, enfim, possuem ainda duas séries de vogais graves. Há poucas línguas que, como o francês e o português, utilizam uma série particular de vogais nasais (MALMBERG, 1954).

A multiplicidade de sistemas vocálicos enseja que o tipo de desenvolvimento proposto nesse trabalho seja efetivamente desenvolvido para cada língua.

Explorando as peculiaridades do sistema vocálico que compõe uma determinada língua, é possível realizar sua caracterização com maior precisão, possibilitando uma retroalimentação mais efetiva.

Os jogos com vogais são basicamente jogos de busca, manutenção e alternância do alvo, proporcionando ao deficiente uma retroalimentação sobre o som que está sendo produzido e um tempo de exercício que propicie o aprendizado de seu sistema tátil-cinestésico e/ou auditivo residual (ARAÚJO, GOTO & VIOLARO, 2000).

Os alvos podem ser agrupados em função da altura da língua na realização das vogais (alta, variante de 2º grau, variante de 1º grau e baixa) e/ou em função do seu ponto de articulação (anterior, medial e posterior). Por exemplo, pode-se agrupar os alvos na forma de um triângulo, similar ao triângulo fonético.

Os jogos com as vogais podem ser concebidos apresentando todos os alvos possíveis (as sete vogais tônicas) ou combinando-as em subgrupos menores, permitindo o aprimoramento por vogal ou subgrupo de vogais. Os jogos devem contemplar as duas

possibilidades. A divisão em subgrupos exige a definição de critérios de como estes serão organizados. Na divisão das vogais em subgrupos, há que se considerar a inteligibilidade dessas vogais e os resíduos auditivos presentes na maioria dos deficientes auditivos.

Os sistemas vocálicos do mundo, de acordo com MALMBERG (op. cit.), estão construídos sobre o triângulo formado pelas vogais /i/ - /a/ - /u/, no qual temos máxima diferenciação acústica (aguda - média - grave) para os ouvintes. Para deficientes auditivos, por outro lado, parece mais adequado que o sistema de aprendizado utilize um padrão /a/-/ɔ/-/o/-/u/, no qual maximizamos a diferenciação acústica (média - grave - mais grave - muito grave), otimizando a utilização de seus resíduos auditivos.

A vogal /a/, pela sua simplicidade articulatória e mínimos requisitos em termos de frequência fundamental, deve ser a base dos jogos.

Os diversos jogos devem contemplar subgrupos de alvos nos quais seja enfatizado o aprimoramento do controle de altura da língua (por exemplo, /a/ x /ɔ/, /a/ x /ɔ/ x /o/ x /u/, etc.) e outros nos quais seja enfatizado o aprimoramento do controle da posição articulatória (por exemplo, /ɛ/ x /ɔ/, /e/ x /o/, /i/ x /u/, etc.), além da possibilidade de realizar ambas as funções (/e/ x /ɛ/ x /ɔ/ x /o/, etc.), até a o sistema completo (/i/ x /e/ x /ɛ/ x /a/ x /ɔ/ x /o/ x /u/).

A base proposta para os jogos com vogais é fixa, com sete colunas verticais, cada uma associada a uma vogal, dispostas na seqüência /i/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/. Ao realizar uma vogal, um objeto (por exemplo, um macaco) move-se verticalmente na coluna correspondente à vogal realizada. Um objeto prêmio (por exemplo, uma fruta) é colocado aleatoriamente na coluna da vogal alvo, na porção oposta (acima ou abaixo). Quando o

objeto prêmio é alcançado, a pontuação é aumentada e o objeto prêmio é colocado aleatoriamente sobre a coluna correspondente a outra vogal.

No caso de se trabalhar com um subgrupo de vogais, o objeto prêmio deve se posicionar unicamente sobre as colunas correspondentes às vogais selecionadas. A velocidade do objeto móvel é função do grau de dificuldade escolhida para o jogo.

Os parâmetros empregados para animação dos jogos com vogais são a primeira formante e a frequência abaixo da qual está contida parcela significativa da energia do quadro medida sobre o espectro LPC (P70s), medidas sobre quadros de 44 ms do sinal de voz. No capítulo 4 são descritos os procedimentos para obtenção destes dois parâmetros.

A Figura 3.8 apresenta um quadro do jogo de vogais desenvolvido.

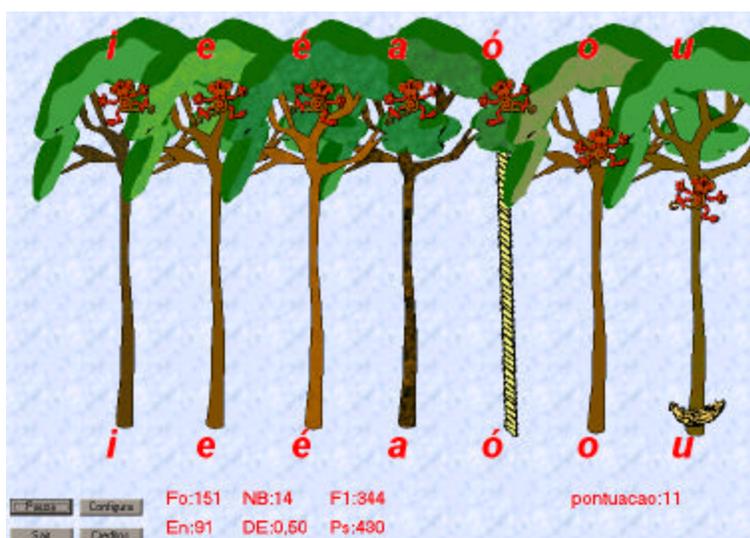


Figura 3.8: Tela de um quadro do jogo de vogais

Na figura 3.8 o objeto (o macaco) move-se em direção ao prêmio (a banana) durante a locução da vogal /u/. Os mostradores indicam uma frequência fundamental (Fo) igual a 151 Hz, energia do quadro (En) igual a 91 dB, número efetivo de bits igual a 14, primeira formante (F1) igual a 344 Hz e o parâmetro P70s (Ps) igual a 430 Hz.

### 3.2.2.2. Exercícios com fricativas

As fricativas constituem um importante grupo no sistema desenvolvido, pois são sons compostos de componentes de altas frequências que, em grande parte dos casos, estão acima da faixa de frequências dos resíduos auditivos dos deficientes auditivos. Além disso, exigem, como no caso do /s/, a participação máxima de todos os parâmetros articulatorios, possibilitando o exercício de uma ampla gama de músculos necessários à produção da fala, como também ao aprendizado do sistema tátil-cinestésico.

Em uma das técnicas mais utilizadas, o sentido do tato é trabalhado colocando-se a palma de uma das mãos da criança à frente da boca do instrutor e a palma da outra mão à frente da sua boca. As diferentes sensações produzidas, por cada uma das fricativas, são utilizadas como instrumento de aprendizado da criança. Em geral, o fluxo de ar é exagerado para aumentar a sensação na palma da mão, resultando em uma produção com excesso de fluxo de ar, o que é fisicamente ineficiente, pois provoca gasto excessivo de ar.

Os jogos desenvolvidos contemplam exercícios com fricativas sonoras e surdas, labiais, lábio-dentais e alveolares. São divididos em grupos propiciando o exercício de sonoridade (/f/ x /v/, /s/ x /z/ ou /ʃ/ x /ʒ/), ou exercício de posicionamento (/f/ x /s/ x /ʃ/ ou /v/ x /z/ x /ʒ/).

Os parâmetros empregados para animação dos jogos com fricativas são a frequência de pico do espectro suavizado (FPES), o coeficiente de espalhamento espectral e a frequência fundamental, medidas sobre quadros de 44 ms do sinal de voz. No capítulo 4 são

descritos os procedimentos para obtenção desses parâmetros. Os parâmetros são obtidos sobre quadros adquiridos com frequência de amostragem de 44.100 Hz.

Os jogos com fricativas são também jogos de busca, manutenção e alternância do alvo, proporcionando ao deficiente uma retroalimentação sobre o som que está sendo produzido e um tempo de exercício suficiente de modo que propicie condições para o aprendizado de seu sistema tátil-cinestésico e/ou auditivo residual.

Nos jogos com fricativas, cada fricativa é associada a uma coluna horizontal. Um objeto móvel (por exemplo, um pássaro) desloca-se horizontalmente na direção de uma barreira. Ao produzir a fricativa, a barreira é rompida na direção da coluna correspondente, permitindo que o objeto móvel passe através da barreira. Caso a produção seja insuficiente, o objeto móvel se choca com a barreira e retorna, permanecendo preso.

A Figura 3.9 apresenta um quadro do jogos de fricativas desenvolvido.

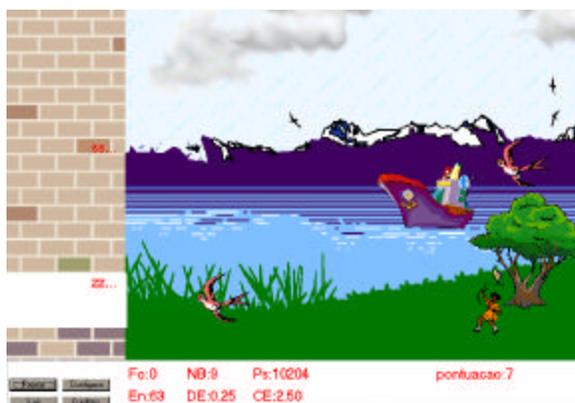


Figura 3.9: Tela de um quadro do jogo de fricativas



## *Capítulo 4*

# Parametrização

A modelagem matemática para o desenvolvimento dos jogos requer sua adequação aos objetivos dos exercícios, a possibilidade de adaptar-se a cada estágio de controle articulatorio da criança, além de propiciar medidas suficientemente precisas para a consecução dos jogos.

O problema geral é de reconhecimento de voz. Particularmente, reconhecimento das características sobre fonemas. Conquanto as técnicas mais utilizadas atualmente para reconhecimento de voz utilizem redes neurais artificiais e/ou cadeias escondidas de Markov, que realizam o reconhecimento após treinamento com bancos de dados de teste, neste trabalho optou-se por uma modelagem paramétrica obtida a partir da determinação das características físicas do sinal de voz.

A modelagem paramétrica fornece, de forma direta, um parâmetro que indica quão próximo do modelo se encontra a produção. A variação do parâmetro obtido permite alimentar um jogo, no qual seja possível uma retroalimentação que indique para a criança quando sua produção está se aproximando ou se afastando do alvo acústico desejado.

A dificuldade inicial para modelagem paramétrica é a falta de dados sobre vozes de crianças ouvintes que possam ser utilizados como padrão para definição dos alvos acústicos. Visando superar esta dificuldade, foi montado um banco de dados com vozes de

10 crianças ouvintes com idade variando entre 7 e 10 anos. Cada criança realizou gravação de vogais sustentadas (sete tônicas), consoantes fricativas sustentadas e de 100 palavras isoladas, com 3 locuções de cada palavra. As palavras utilizadas foram obtidas de listas de palavras utilizadas em Logaudiometria (PUPO, 1981 apud RUSSO & SANTOS, 1994). Foram selecionadas para cada palavra as duas melhores locuções de cada criança, gerando um banco de dados com 2.000 locuções. O banco de dados foi ampliado com a gravação de vogais tônicas e fricativas sustentadas produzidas por pré-adolescentes ouvintes com idade variando entre 11 e 14 anos.

A gravação do banco de dados foi feita diretamente em microcomputador, sendo realizada em ambiente tipo escritório, similar aos ambientes nos quais devem ser utilizados os jogos. Os sinais foram adquiridos com 16 bits por amostra, garantindo uma elevada relação sinal ruído de quantização. A frequência de amostragem utilizada foi de 44.100 Hz, por ser um padrão multimídia encontrado na maioria dos sistemas disponíveis para aquisição de sinais de áudio em computador e que assegura uma qualidade de áudio em padrão CD.

O banco de dados foi utilizado para caracterização acústica e modelagem matemática dos fonemas. Quando possível, foram utilizados dados constantes da literatura, em especial os dados referentes à frequência de vibração das pregas vocais. Características específicas da língua portuguesa, como no caso da composição e distribuição espectral das vogais, foram utilizadas para definição dos parâmetros.

Fonemas como as fricativas exigiram a determinação de mais características físicas que as usualmente utilizadas, ensejando o desenvolvimento de novos modelos paramétricos para esses fonemas.

As características da voz de crianças com deficiência auditiva foram também levadas em consideração na definição dos parâmetros a serem utilizados.

Os parâmetros foram obtidos sobre quadros de 2048 amostras (46,4 ms), para o sinal adquirido a uma frequência de amostragem de 44.100 Hz. Alguns parâmetros, como os propostos para caracterização das fricativas, foram obtidos diretamente do sinal amostrado à taxa de 44.100 Hz. Outros parâmetros, como os propostos para caracterização de vogais, foram obtidos com o sinal amostrado à frequência de 11.025 Hz. Para tal, submeteu-se a seqüência original a uma filtragem passa-baixas com corte em 5.000 Hz (filtro Chebyshev tipo II, obtido via transformação bilinear) e em seguida tomou-se uma em cada 4 amostras.

Durante a execução dos jogos são adquiridos quadros com 2048 amostras do sinal de voz. Quando o quadro é preenchido, as amostras são transferidas para um vetor. Após a transferência das amostras é dado um comando para leitura de um novo quadro. Concomitante à aquisição de um novo quadro, são realizadas as medidas dos parâmetros e a animação dos jogos. O sinal, assim, é adquirido com um mínimo de amostras perdidas entre quadros.

#### 4.1. Medidas da intensidade

As informações relativas à intensidade do sinal produzido são importantes, pois proporcionam medidas da energia e do número de bits efetivamente utilizados na aquisição do quadro.

#### 4.1.1. Medida da energia do quadro

A medida da energia do quadro é base para animação dos jogos para exercícios de aprimoramento do controle do fluxo respiratório e do controle de energia. Esta medida é simples e requer apenas a determinação do valor quadrático médio dos quadros do sinal em decibéis. Os sinais são adquiridos em quadros contíguos de 2.048 amostras, com um conversor de 16 bits e utilizando uma frequência de amostragem de 44.100 Hz. O processamento é realizado sem superposição entre os quadros.

Antes de calcular a energia, a seqüência foi submetida a um pré-processamento que consistiu em eliminar, quadro a quadro, a componente DC do sinal, subtraindo-se de cada amostra o valor médio do quadro. Este procedimento visa eliminar eventuais componentes DC introduzidas pelo sistema de aquisição.

Para a seqüência resultante  $x(n)$ , com nível DC nulo e correspondente a um quadro com  $N$  amostras, a medida do valor quadrático médio do quadro em consideração ( $E_q$ ) foi obtida através de:

$$E_q = 10 \cdot \log_{10} \left( \frac{1}{N} \sum_{n=1}^N x^2(n) \right), \quad n = 1 \cdots N, \quad N = 2048 \quad (4.1)$$

A medida da energia está em decibéis, mas não está relacionado a uma unidade específica (referência). Os diversos tipos de microfones e placas de aquisição utilizados em multimídia, com controle de ganho bastante amplo para o sinal de entrada, tornam virtualmente impossível uma normalização do valor da energia.

Na animação dos jogos o valor da energia de cada quadro em decibéis, foi passado por um filtro passa-baixas. Esta filtragem visa reduzir a velocidade da movimentação das

animações decorrentes de eventuais variações abruptas da energia entre um quadro e o seguinte. O procedimento para estimação da energia filtrada para controle do jogos é descrito no apêndice A.

#### 4.1.2. Medida do número de bits efetivos

A medida do número de bits efetivos informa se o nível de intensidade do quadro está (ou não) adequado, mostrando se os picos do sinal estão saturando (ou próximos de saturar) a placa de aquisição ou se o nível do sinal está muito baixo. Um sinal muito intenso que sature a placa de aquisição provocará distorção no sinal. Outro sinal com pouca intensidade estará efetivamente utilizando tão poucos bits na aquisição que reduzirá a relação sinal ruído de quantização. Em ambas as situações extremas o desempenho dos algoritmos para parametrização será prejudicado.

O número de bits efetivamente utilizado para representação do quadro pode ser calculado por:

$$N_{be} = 1 + \text{int}(\log_2(\max(\text{abs}(x(n))) + 0,5)), \quad n = 1 \cdots N \quad (4.2)$$

onde  $x(n)$  é a seqüência resultante da aquisição a 44.100 Hz, antes da remoção do nível DC para cálculo da energia.

O número de bits efetivos ( $N_{be}$ ) varia entre 0 e o número total de bits do conversor, no caso 16. O valor nulo indica ausência de sinal e o valor máximo indica que o sinal está saturado. Valores de  $N_{be}$  baixos ( $N_{be} < 8$ ) indicam que nível do sinal está muito baixo, enquanto que valores de número de bits efetivos altos ( $N_{be} > 14$ ) indicam que o nível está saturando, ou muito próximo de saturar o sistema de aquisição.

Cabe ao usuário verificar os valores apresentados na tela para esse parâmetro, de acordo com o microfone e a placa de aquisição utilizados, e ajustar o ganho da placa de aquisição. Recomenda-se que o ganho da placa de aquisição seja ajustado de forma que o número de bits efetivos fique entre 11 e 12, para a locução de uma vogal sustentada.

O procedimento para obtenção do número de bits efetivos é descrito no apêndice A.

#### 4.1.3. Controle de ajuste de intensidade dos aplicativos

O sistema apresenta para todos os jogos uma opção de configuração, no qual o ruído de fundo é gravado. O valor da energia por quadro desse ruído de fundo é utilizado como limitante para o funcionamento do jogo. Quando o valor da energia estiver menos de 5 dB acima do ruído de fundo, não haverá cálculo de parâmetros, permanecendo ativos apenas as animações automáticas do jogo. Para que o jogo se desenrole, o nível do sinal deverá estar pelo menos 5 dB acima do nível do ruído de fundo.

A calibração do sistema antes da utilização do jogo é muito importante. A calibração é simples e pode ser feita de forma instrumental (medidor de nível de pressão sonora) e/ou biológica.

Na calibração biológica um locutor ouvinte deverá executar o jogo e verificar o valor da energia obtida. Deverá realizar locuções de baixa, média e alta intensidade, anotar os valores mostrados e utilizar esses valores para fixar os níveis mínimos e máximos do valor da energia para controle do jogo no painel de configuração. Na calibração instrumental um medidor de nível de pressão sonora pode ser utilizado para fazer correspondência entre o valor em decibéis do instrumento e o valor apresentado na tela do programa.

A calibração, em qualquer caso, será válida para o microfone utilizado e para o ganho fixado na placa de aquisição, devendo ser reavaliada para qualquer mudança na configuração do sistema.

## 4.2. Modelagem para exercícios com as pregas vocais

A laringe é utilizada para transformar o fluxo de ar pulmonar numa série de pulsos de fluxo aéreo com intervalos regulares. Inicialmente as pregas vocais estão aduzidas, criando um obstáculo à passagem do ar. A força da pressão que a massa de ar contida nas cavidades subglotais exerce sobre as pregas vocais tende, então, a afastá-las e um jato de ar de alta velocidade passa através da glote. Quando o ar escapa através da glote, a alta velocidade, tem lugar o efeito de depressão de Bernoulli que, em conjunto com as forças elásticas dos tecidos, tende a fechar novamente a glote. Uma vez aduzidas as pregas, a pressão subglotal aumenta e o ciclo se repete. Conquanto outros fatores concorram para a vibração das pregas vocais, esses fatores estão fora do escopo deste trabalho.

Os pulsos de ar ocorrem em intervalos regulares, sendo a frequência fundamental da voz ( $F_0$ ) o número de ciclos realizados em um segundo.

O sinal proveniente das pregas vocais apresenta-se sob a forma de um trem de pulsos periódicos com padrão espectral descendente à razão de 12 dB por oitava. Os pulsos são produzidos periodicamente de acordo com a frequência de vibração das pregas.

A frequência fundamental pode ser mensurada determinado-se o inverso do intervalo de tempo transcorrido entre dois pulsos glóticos sucessivos, ou tomando-se a frequência correspondente à primeira harmônica do espectro de frequências.

A Figura 4.1 apresenta a aproximação de um trem de pulsos glóticos, com uma frequência fundamental de 100 Hz. Na Figura 4.2 é apresentado o espectro de frequências correspondente.

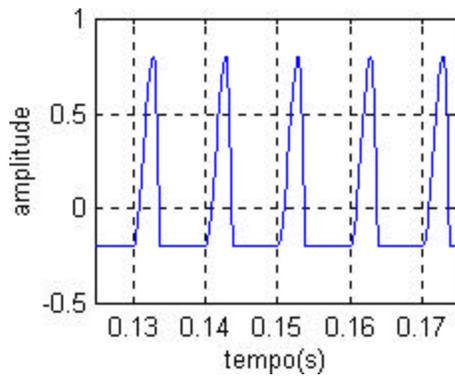


Figura 4.1: Trem de pulsos glóticos

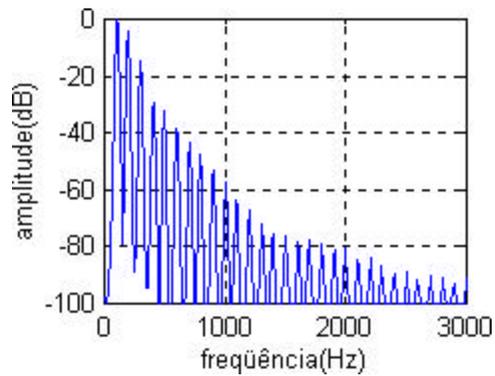


Figura 4.2:Espectro do trem de pulsos

As harmônicas, ao atingirem o trato, são seletivamente amplificadas (ou atenuadas) por ação das ressonâncias (ou antirressonâncias), em função das diversas posições assumidas pelo trato, produzindo os diversos tipos de sons sonoros. A Figura 4.3 apresenta um padrão sintético com frequências de ressonância de 500 Hz e 2.200 Hz.

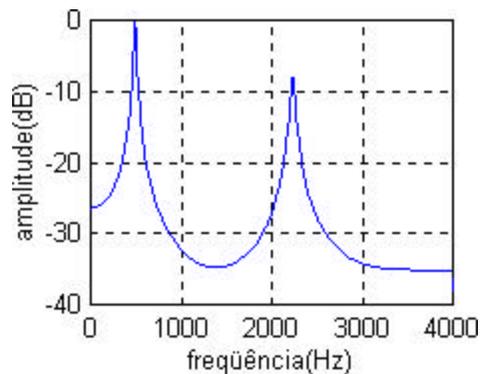


Figura 4.3: Padrão espectral com ressonâncias de 500 Hz e 2.200 Hz

A Figura 4.4 apresenta a forma de onda sintética resultante da filtragem do trem de pulsos apresentado na Figura 4.2, passado pelo filtro cuja resposta em frequência é mostrada na Figura 4.3. Na Figura 4.5, é apresentado o espectro correspondente.

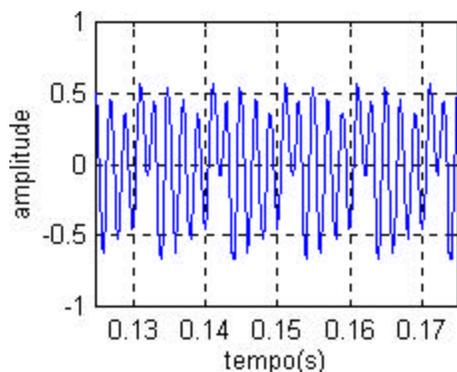


Figura 4.4: Forma de onda do sinal sintético,  
à saída do filtro que simula o trato vocal.

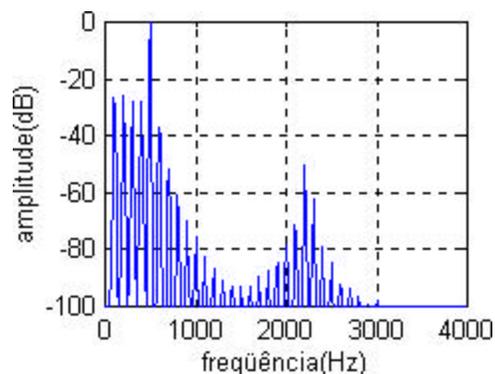


Figura 4.5: Espectro do sinal sintético.

Na Figura 4.4, observa-se a presença de oscilações amortecidas intra-ciclos resultante da ação dos ressonadores atuando sobre o trem de pulsos (Figura 4.1).

A frequência fundamental pode ser facilmente determinada a partir da forma de onda do sinal mostrada na Figura 4.1 ou do seu espectro apresentado na Figura 4.2. Entretanto, para aplicações sobre o sinal acústico, o valor de  $F_0$  deve ser obtida a partir do sinal de voz como mostrado na Figura 4.4, ou a partir do seu espectro (Figura 4.5).

A determinação automática de  $F_0$  para sinais reais apresenta algumas dificuldades. Os picos máximos são desiguais, tanto em decorrência de mudanças no trato vocal quanto devido à falta de sincronismo entre os instantes de amostragem e os picos do sinal. Assim, medidas automáticas no domínio do tempo são susceptíveis a essas variações. As harmônicas do espectro de frequências de sinais reais apresentam-se contaminadas por

ruído introduzido pela fonte glótica e pelo ruído existente no ambiente de aquisição. Essas são algumas limitações na medida automática de  $F_0$  diretamente a partir do espectro de frequências.

Para PICONE (1993), a medida de  $F_0$  tem sido negligenciada em sistemas de reconhecimento de voz por diversas razões, incluindo o grande esforço computacional, e a dificuldade na caracterização da interação entre  $F_0$  e fenômenos suprasegmentais.

As dificuldades ensejam uma grande variedade de tipos de algoritmos para extração do valor de  $F_0$ . Existem algoritmos que realizam a medida diretamente no domínio do tempo (GOLD & RABINER, 1969; SUKKAR e col., 1988). Outros utilizam a função AMDF (Average Magnitude Difference Function) como RABINER & SCHAEFER (1978). Existem algoritmos que realizam a medida utilizando análise cepstral (NOLL, 1967). Outros algoritmos realizam a medida a partir do resíduo da análise LPC (MARKEL & GRAY, 1976), ou utilizam a função de correlação (SONDHI, 1968).

Para PICONE (op. cit.), os algoritmos baseados na função de autocorrelação apresentam alto desempenho sobre uma grande faixa de aplicações, inclusive na presença de ruído. Nesse caso, os ruídos que acompanham o sinal são minimizados (baixa correlação) e as características periódicas do sinal são maximizadas (alta correlação). Existe um custo computacional associado, que é superado pelo crescente avanço tecnológico.

A frequência  $F_0$  pode ser obtida tomando-se o inverso do tempo no qual ocorre o máximo da função de autocorrelação. Este máximo deve ser buscado entre dois valores limites de tempo ( $T_1$  e  $T_2$ ). O menor valor limite ( $T_1$ ) corresponde ao inverso da frequência fundamental máxima admitida e o maior valor limite ( $T_2$ ) corresponde ao inverso da frequência fundamental mínima aceita pelo algoritmo.

A Figura 4.6 apresenta a forma de onda do sinal de voz de uma criança do sexo masculino, de 8 anos, produzindo a vogal /o/. A figura 4.7 apresenta a função de autocorrelação correspondente.

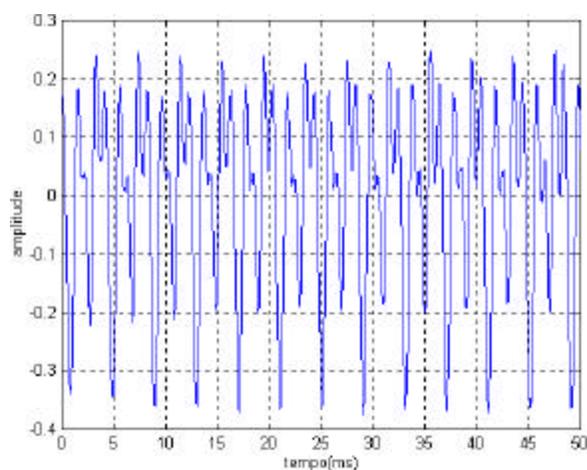


Figura 4.6: Exemplo de forma de onda da vogal /o/.

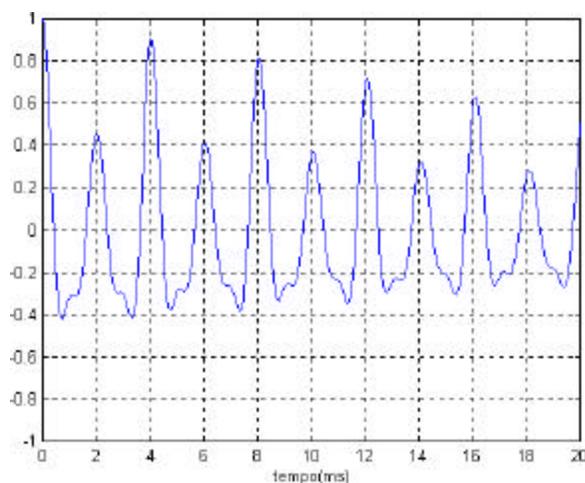


Figura 4.7: Função de autocorrelação obtida para a vogal /o/

Analisando a Figura 4.7, observa-se que o pico máximo ocorre em 4 ms, o que corresponde a uma frequência fundamental de 250 Hz. O pico em 2 ms, na função de autocorrelação, resulta das oscilações amortecidas intra-ciclos, introduzidas pela ação do trato vocal. O valor de pico da função de autocorrelação normalizada é comumente utilizado para definir se o quadro de sinal sob análise é ou não sonoro. Na Figura 4.6, pode-se observar ainda que os valores máximos dos picos de cada ciclo não são iguais entre si, nem para uma produção sustentada.

A utilização da função de autocorrelação sobre sinais reais apresenta algumas dificuldades. O principal problema decorre da presença de eventuais picos da função de

autocorrelação maiores que o relativo à frequência fundamental, devido à ação do trato vocal. Diversos métodos tem sido utilizados para reduzir a ação do trato na estimativa automática de  $F_0$ . MARKEL & GRAY (1976) propõem realizar a medida de  $F_0$  a partir do resíduo da análise LPC. RABINER & SCHAEFER (1978) afirmam que este procedimento é ineficiente para vozes com valores elevados de  $F_0$ , sendo portanto inadequado para vozes de crianças.

Em 1968, SONDHY propôs uma operação não linear sobre o sinal, buscando reduzir o efeito do trato sobre a função de autocorrelação. Um dos efeitos da ação do trato sobre o pulso glótico é a introdução de oscilações amortecidas entre pulsos glóticos. O algoritmo de SONDHY consiste, basicamente, em reduzir (ou eliminar) estas oscilações tornando nulas as amostras da sequência menores que uma determinada magnitude do sinal (Cl) e subtraindo as demais amostras do valor de Cl. O autor sugere para Cl o valor de 30% da amplitude máxima do quadro. A interferência do trato pode ser minimizada aumentando-se o valor de Cl, entretanto existe o risco de que pulsos com menor magnitude possam ser desprezados, acarretando erros na medida final. Quando o pico máximo da função de autocorrelação normalizada, entre  $T_1$  e  $T_2$ , obtida sobre o sinal após a transformação não linear, é superior a 0,3, o sinal é considerado sonoro e é efetuada a medida de  $F_0$ .

Sinais de voz apresentam diferenças de intensidade a cada ciclo, mesmo durante fonação sustentada. Estas diferenças decorrem da variação da pressão da fonte glótica, da dificuldade de manter constante os ajustes musculares das pregas vocais e das próprias modificações do ganho introduzido pelo trato, que não se mantém absolutamente imóvel, nem para uma locução sustentada. Estas variações na intensidade do sinal diminuem o valor máximo da função de autocorrelação, reduzindo a confiabilidade do algoritmo.

Na transformação proposta por SONDHI (op. cit.), as amostras são subtraídas por uma constante (CI). Diferenças absolutas de amplitude existentes entre os picos dos ciclos sucessivos do sinal, após o processamento não linear, são mantidas fixas sobre um sinal resultante com menor intensidade, acentuando portanto a diferença relativa entre ciclos sucessivos do sinal. Este procedimento, entretanto, reduz o valor máximo da função de autocorrelação normalizada entre  $T_1$  e  $T_2$ .

É importante no algoritmo que o valor de CI seja, o mais elevado possível, sem desprezar picos efetivos. RABINER & SCHAEFER (1978) sugerem uma variante para este algoritmo, na qual o quadro é dividido em 3 sub-quadros, sendo determinado o valor de pico do primeiro e último sub-quadro. Em seguida, o valor de CI é fixado entre 60 e 80% do menor valor de pico obtido anteriormente.

Neste trabalho propõe-se uma modificação no algoritmo de autocorrelação, para obtenção da frequência fundamental, tendo sido obtido após uma série de testes visando reduzir a influência do trato.

Conforme foi descrito anteriormente, os deficientes auditivos apresentam, em geral, anormalidades na voz: elevada soprosidade indicando presença significativa de ruído glótico, dificuldades na ressonância e expiração que produzem variações de intensidade, e frequências fundamentais tão elevadas quanto 680 Hz. Estas são algumas das características que devem ser consideradas na determinação de  $F_0$  que sejam adequados em aplicativos para exercitar deficientes auditivos. O algoritmo deve ser eficiente mesmo na presença de ruído, deve ser capaz de reduzir os efeitos das variações de intensidade e, ainda, deve ser adequado em faixa de frequências.

A utilização da função de autocorrelação possibilita reduzir o efeito da presença de ruído, sendo a adotada neste trabalho para determinação de  $F_0$ .

Em geral, os algoritmos são projetados para determinar  $F_0$  entre 50 Hz e 500 Hz (MARKEL & GRAY, 1976). Para a aplicação proposta a faixa de frequência foi fixada entre 80 Hz e 800 Hz.

O algoritmo proposto neste trabalho consiste em: reduzir por filtragem as componentes de frequência fora da faixa de 80 Hz a 800 Hz, reduzir a frequência de amostragem para aumentar a velocidade do cálculo, realizar uma operação de truncamento sobre a seqüência filtrada, tomar a porção unipolar mais significativa da seqüência truncada e, finalmente, obter  $F_0$  a partir do tempo associado ao pico da função de autocorrelação normalizada sobre a seqüência unipolar, filtrada e truncada, entre  $T_1$  e  $T_2$ .

A seqüência  $x(n)$ , amostrada a 44.100 Hz, com 2048 amostras de cada quadro é passada inicialmente por dois filtros digitais, um passa-altas e um passa-baixas. O filtro passa-altas é um filtro digital IIR Chebyshev tipo II de 4ª ordem, obtido por transformação bilinear, com frequência de corte de 65 Hz. Este filtro visa reduzir o efeito das componentes de 60 Hz da rede elétrica que eventualmente se associam ao sinal, na placa de aquisição. O filtro passa-baixas é também um filtro digital IIR Chebyshev tipo II, obtido por transformação bilinear, com frequência de corte de 800 Hz. O filtro passa-baixas visa reduzir a ação das formantes do sinal de voz acima da faixa especificada.

O valor da constante para truncamento (CI) é obtido quadro a quadro, dividindo-se cada quadro em 2 sub-quadros de igual comprimento e determinando-se o valor de pico de cada um dos sub-quadros. O valor para o truncamento do sinal (CI) foi fixado em 0,6 do menor valor de pico obtido.

A seqüência é em seguida sub-amostrada tomando-se uma em cada 4 amostras, reduzindo-se a freqüência de amostragem para 11.025 Hz e o número de amostras para 512.

Em seqüência, a operação de truncamento é realizada através de uma operação que torna nulas as amostras entre  $-Cl$  e  $+Cl$ , mantendo inalteradas as demais amostras. A função proposta para truncamento do sinal é apresentada na Figura 4.8.

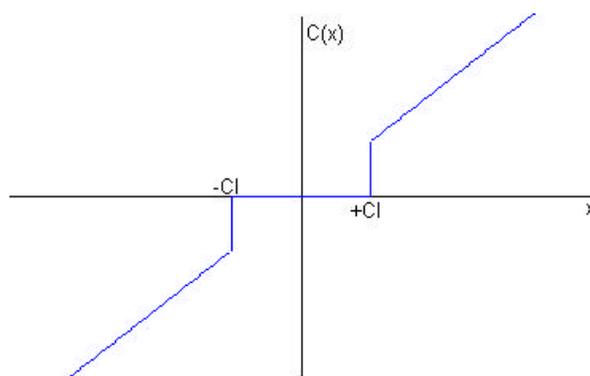


Figura 4.8: Função de truncamento do sinal

Essa função de truncamento reduz a ação das oscilações intra-ciclo sem reduzir as diferenças relativas entre os picos do sinal.

Sinais de fala podem apresentar o pico máximo positivo ou negativo. Sendo a oscilação do sinal uma característica introduzida pelo trato, pode-se eliminar a porção menos significativa, seja ela positiva ou negativa, tornando nulas as amostras com polaridade inversa ao pico máximo.

A função de autocorrelação da seqüência filtrada, truncada e unipolar, é calculada e normalizada. O quadro é considerado sonoro se o valor de pico da função de autocorrelação, entre  $T_1$  e  $T_2$ , é igual ou superior a 0,4, sendo os valores de  $T_1$  e  $T_2$  iguais a 1,25ms e 12,5ms, que correspondem a 800 Hz e 80 Hz respectivamente. A freqüência

fundamental é então determinada, tomando-se o inverso do tempo no qual ocorre o valor de pico da função de autocorrelação no intervalo acima.

O procedimento detalhado para estimação da frequência fundamental é descrito no apêndice B.

#### 4.2.1. Resultados obtidos com o estimador de frequência fundamental.

A seguir são apresentados alguns resultados obtidos com o medidor de frequência fundamental implementado .

A Figura 4.9 apresenta as medidas de  $F_0$  traçadas sobre um espectrograma, para locução sonora (vogal /o/) com duração de 314 ms e  $F_0$  média de 249 Hz, realizada por um menino de 8 anos. A Figura 4.10 apresenta o espectrograma e as medidas da frequência fundamental para uma locução sonora (vogal /a/) com duração de 11 s e  $F_0$  média de 638 Hz, realizada por pré-adolescente com deficiência auditiva profunda.

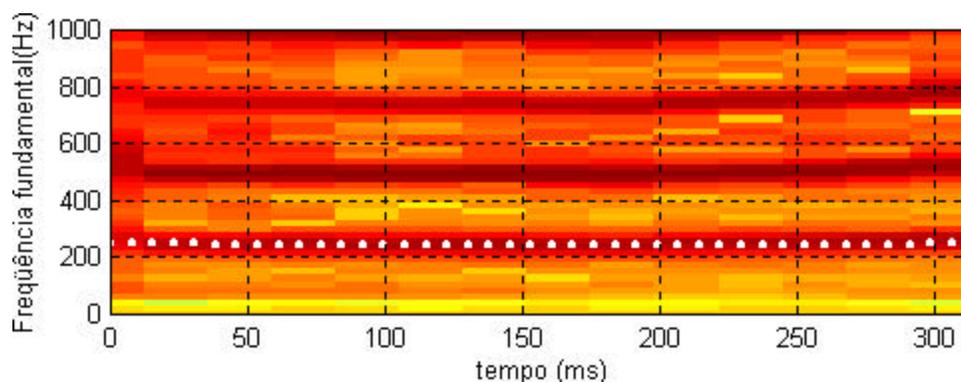


Figura 4.9: Espectrograma da vogal /o/ e rastreamento da frequência fundamental

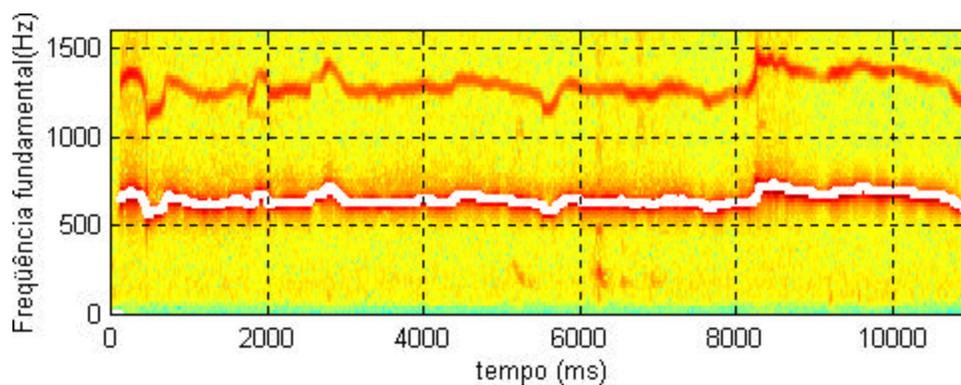


Figura 4.10: Espectrograma e rastreo da frequência fundamental da vogal /a/, de locutor pré-adolescente com deficiência auditiva profunda.

Nas Figuras 4.9 e 4.10, é possível verificar que o algoritmo proposto foi capaz de acompanhar a primeira harmônica do sinal – a frequência fundamental ( $F_0$ ). Na Figura 4.10 pode-se verificar a existência de valores de frequência fundamental de até 740 Hz.

O algoritmo implementado respondeu satisfatoriamente quando avaliado com vozes masculinas com baixos valores de  $F_0$  (100 Hz-140 Hz), com vozes de mulheres com valores médios de  $F_0$  (170 Hz-230 Hz), com vozes de crianças com valores de  $F_0$  altos (220 Hz-300 Hz) e com vozes de crianças portadoras de deficiência auditiva e produzindo frequências fundamentais extremamente altas.

### 4.3. Modelagem das vogais

As vogais são realizadas por posicionamentos diversos no interior do trato vocal. As diferentes configurações assumidas pelo trato introduzem ressonâncias sobre o sinal

proveniente da glote. As ressonâncias acentuam algumas frequências, propiciando a distinção entre as diversas vogais. Assim, para cada vogal observa-se um “posicionamento característico” do trato, cada posicionamento correspondendo a um conjunto de frequências de ressonância, as formantes das vogais. As formantes das vogais são um importante instrumento para análise de sinais de voz.

Diversas técnicas são empregadas para extração das formantes. Utiliza-se desde técnicas por inspeção visual do espectro do sinal (PETERSON & BARNEY, 1952; BEHLAU, 1984), até técnicas automáticas de análise numérico-computacional (HILLENBRAND e col., 1995). Em qualquer caso, as frequências das formantes correspondem aos picos do espectro de frequências.

Na análise de sinais de voz, as duas primeiras formantes,  $F_1$  e  $F_2$ , são da maior relevância (PETERSON & BARNEY, 1952). Essas duas primeiras formantes possibilitam a caracterização do plano de vogais (alta x baixa, anterior x posterior), e suas trajetórias são a base para a teoria do locus, amplamente utilizada para análise dos sons de diversas consoantes - { (/p/, /b/), (/t/, /d/), (/k/, /g/) } (DELATTRE, LIBERMAN & COOPER, 1955; FRUCHTER, 1994; NITTROUER, STUDDERT-KENNEDY & NEELY, 1996; SUSSMAN e col., 1996 )

Diversos fatores limitam as técnicas para extração das formantes, entre as quais se destacam as diferenças físicas entre locutores, a diferença entre diversas locuções de um mesmo locutor, e as diferenças de idade e sexo. Em princípio, observa-se que uma dada vogal nunca é realizada exatamente da mesma maneira. A cada realização são modificadas as frequências formantes, suas amplitudes relativas e a frequência fundamental, isto para um mesmo locutor. Além disto, a análise por fonemas indica que estes são afetados por fonemas vizinhos, em maior ou menor grau, dependendo do fonema, da prosódia, do

locutor e da locução em questão. Estas são algumas razões que configuram o principal fator complicador para a análise de voz, a falta de um único padrão específico para cada som. Dispomos para cada som apenas padrões com características de similaridade – os sons da fala não são determinísticos, mas se caracterizam como eminentemente aleatórios.

A caracterização estatística do sinal da fala permite realizar a análise do sinal de um determinado fonema esperando encontrar estas ou aquelas propriedades, mas esta será sempre uma esperança matemática.

Múltiplos fatores influenciam as medidas de formantes para cada técnica utilizada:

- Na *técnica de inspeção visual direta dos espectros*, o espectro de Fourier apresenta diversos picos, correspondentes aos harmônicos da frequência fundamental, nem sempre com definição clara da envoltória e impossibilitando, portanto, a definição exata das frequências formantes. Nesta técnica é necessário, além de admitir um certo erro na medida, um forte domínio teórico sobre os sinais e uma boa experiência prática. Neste tipo de análise, frequências fundamentais elevadas, em geral resultantes de sons produzidos por crianças e mulheres, aumentam as dificuldades.
- *Técnicas numérico-computacionais* são afetadas por outros fatores, entre os quais se destacam as dificuldades dos algoritmos com sinais com frequências de ressonâncias muito próximas (por exemplo as duas primeiras formantes da vogal /u/), ou muito distantes entre si (por exemplo, as duas primeiras formantes da vogal /i/). Na análise LPC (MARKEL & GRAY, 1976), formantes próximas podem ser modeladas por um único par de pólos,

provocando um erro de fusão, quando a ordem do sistema de análise é muito baixa. Erros de fusão ocorrem, em geral, na análise de fonemas com formantes muito próximas (ex.: a vogal /u/). Harmônicas da frequência fundamental podem ser caracterizadas como formantes, produzindo erro de inserção, quando a ordem do modelo de análise é muito alta. Erros de inserção, em geral, ocorrem na análise de fonemas com formantes muito afastadas (ex.: a vogal /i/). A ordem adequada do modelo depende do sexo e idade do locutor.

Em qualquer caso, erros são inerentes às diferenças entre os modelos assumidos na análise e o efetivamente utilizado na realização. A infinidade de modelos utilizados na realização, torna virtualmente impossível a utilização de um único ou mesmo de um conjunto finito de modelos para efetuar uma análise isenta de erros.

A análise da voz sobre o modelo fonte-filtro consiste em determinar a frequência fundamental da excitação e as frequências de ressonâncias do filtro que modela o trato vocal.

A determinação das formantes é dificultada pela própria dinâmica da voz. Estas dificuldades ensejam um conjunto substancial de técnicas matemáticas em sua análise, entre as quais podemos citar: transformada de Fourier, análise cepstral, análise linear preditiva (LPC). Os bancos de filtros, como os utilizados na análise espectrográfica, propiciam uma avaliação da envoltória espectral, eximindo-se de determinar a posição exata das formantes. A análise linear preditiva (LPC), que procura determinar as frequências de ressonância do trato, é extremamente sensível ao número e à posição relativa

das formantes, e além disso depende da ordem do preditor utilizado e também é bastante sensível as harmônicas da frequência fundamental.

O grande problema de compromisso é que com preditores de baixa ordem, somente a envoltória espectral é capturada e, com preditores de ordem alta, detalhes da estrutura fina associados à excitação podem ser representados no modelo (MARKEL & GRAY, 1976).

Medidas das duas primeiras frequências formantes, com modelos de ordens mais independentes do locutor e/ou locução, podem ser obtidas através da análise LPC-MEL (ARAÚJO & VIOLARO, 1998). Esta análise é realizada calculando-se uma função de autocorrelação modificada, obtida a partir da DFT computada sobre frequências uniformemente espaçadas no plano MEL. O plano MEL impõe uma quantização não linear no domínio da frequência, impossibilitando o uso de métodos rápidos para cálculo da DFT (FFT). Estas medidas exigem, em consequência, um elevado esforço computacional (ARAÚJO & VIOLARO, 1998). O elevado esforço computacional foi um dos fatores que levaram à não utilização desta técnica na elaboração dos jogos de voz.

#### 4.3.1. Análise das vogais do português falado no Brasil.

As vogais tônicas do português falado no Brasil foram estudadas por BEHLAU (1984), sendo caracterizados a duração, intensidade e valores médios das formantes das vogais. Os resultados obtidos por BEHLAU constituem o ponto de partida para o desenvolvimento do algoritmo de parametrização que permita a classificação das vogais.

BEHLAU (1984) obteve, para o português falado no Brasil, as frequências formantes de adultos e crianças de ambos os sexos. A Tabela 4.1 apresenta uma síntese dos resultados obtidos pela pesquisadora, para as duas primeiras formantes, em Hz.

Grupos	Formantes	Anteriores			Medial	Posteriores		
		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/
Homens	$F_1$	398	563	699	807	715	558	400
	$F_2$	2456	2339	2045	1440	1201	1122	1182
Mulheres	$F_1$	425	628	769	956	803	595	462
	$F_2$	2984	2712	2480	1634	1317	1250	1290
Crianças	$F_1$	465	698	902	1086	913	682	505
	$F_2$	3176	2825	2606	1721	1371	1295	1350

Tabela 4.1.  $F_1$  e  $F_2$  das vogais tônicas do português falado no Brasil (BEHLAU, 1984).

Na Tabela 4.1 pode-se observar que os valores de  $F_1$  aumentam com o grau da altura da língua (por exemplo,  $F_{1/a/} > F_{1/i/}$ ) e os valores de  $F_2$  aumentam com a posição articulatória (por exemplo,  $F_{2/i/} > F_{2/a/}$ ).

As distâncias absolutas e relativas entre os formantes das vogais são importantes características que devem ser consideradas na sua parametrização e classificação. Uma medida da distância absoluta ( $DF_\alpha$ ) entre as vogais pode ser obtida tomando-se o valor absoluto da diferença entre os valores das formantes das vogais consideradas, onde  $\alpha = 1$  corresponde a  $F_1$  e  $\alpha = 2$  corresponde a  $F_2$ . A distância relativa ( $\Delta F_\alpha$ ) pode ser obtida por:

$$\Delta F_\alpha = \frac{\text{abs}(F_{\alpha/v1/} - F_{\alpha/v2/})}{\min \begin{bmatrix} F_{\alpha/v1/} & F_{\alpha/v2/} \end{bmatrix}} \cdot 100\%, \quad \alpha = 1,2 \quad (4.3)$$

onde  $v_1$  e  $v_2$  correspondem as vogais.

Considerando-se duas vogais com pares de formantes próximas, como por exemplo as vogais /i/ e /e/ produzidas pelas crianças, pode-se verificar que, conquanto o valor de  $DF_1$  (233 Hz) seja menor que o de  $DF_2$  (351 Hz), o valor de  $\Delta F_1$  (50,1%) é muito maior que o valor de  $\Delta F_2$  (12,4%). Para o grupo de crianças, por exemplo, a Tabela 4.2 apresenta a diferença relativa, tomando-se como base o menor valor, entre os valores médios das formantes, em percentagem.

	/i/ - /e/	/e/ - /ɛ/	/ɛ/ - /a/	/a/ - /ɔ/	/ɔ/ - /o/	/o/ - /u/
$\Delta F_1$	50,1 %	29,2 %	20,7 %	18,9 %	33,9 %	35,0 %
$\Delta F_2$	12,4 %	8,4 %	51,4 %	25,5 %	5,9 %	4,2 %

Tabela 4.2: Diferença relativa entre os valores médios das formantes, em percentagem

Na Tabela 4.2, pode-se observar que os valores de  $\Delta F_1$ , tomados sobre vogais com pares de formantes próximos, mantêm-se acima de 18,9 %, enquanto os valores de  $\Delta F_2$  podem assumir valores tão pequenos quanto 4,2 %. Pequenos valores de  $\Delta F_2$  mostram que os valores de  $F_2$  são praticamente indistintos de uma vogal para outra.

Para o grupo de crianças, por exemplo, BEHLAU (1984) obteve medidas de  $F_2$  entre 2.950 Hz e 3.500 Hz para a vogal /i/, de 2.200 Hz a 3.200 Hz para a vogal /e/ e entre 2.350 Hz e 3.020 Hz para a vogal /ɛ/. É interessante verificar que, embora o valor médio de  $F_2$  obtido para a vogal /ɛ/ seja maior que o obtido para a vogal /e/, a faixa contendo as medidas individuais de  $F_2$  para a vogal /ɛ/ está totalmente contida na faixa de frequências relativas à vogal /e/. Portanto, embora com médias de  $F_2$  diferentes, há muita superposição

entre as medidas individuais, superposição fortemente associada à posição da articulação e ao comprimento do trato vocal.

A utilização de  $F_2$  para controle de jogos pode produzir confusões inerentes aos valores do próprio parâmetro. A primeira formante, por outro lado, está intimamente relacionada à altura da língua, mas é insuficiente para propiciar a separação entre as sete vogais orais da língua portuguesa.

Os valores absolutos de  $F_2$ , embora não pareçam adequados para elaboração de jogos para aprimoramento do controle articulatório por deficientes auditivos, não são características que devam ser desconsideradas. Os valores elevados de  $F_2$  estão associados a ressonâncias em frequências altas, sendo característicos de vogais anteriores, enquanto baixos valores de  $F_2$ , que estão associados a ressonâncias em baixas frequências, constituem uma característica das vogais posteriores.

A presença de ressonâncias em frequências baixas, na vogais posteriores, faz com que a energia do sinal se concentre em baixas frequências, enquanto que para as vogais anteriores, com ressonâncias em frequências elevadas, o espectro de frequências se mostra distribuído, com energia significativa sobre uma larga faixa de frequências. Uma medida da distribuição de energia no espectro de frequências possibilita portanto avaliar se a vogal é posterior ou anterior.

O algoritmo proposto neste trabalho, consiste em determinar a altura da vogal pela determinação da primeira formante por análise LPC, e obter a posição articulatória a partir da distribuição espectral:

- O *valor da primeira formante* ( $F_1$ ) obtida por análise LPC, permite classificar a vogal em **alta**, **variante de 1º grau**, **variante de 2º grau** ou **baixa**. A utilização da análise LPC para obtenção unicamente da primeira formante

reduz a possibilidade de erros de inserção, aumentando a robustez do algoritmo, tornando-o menos susceptível à ordem do modelo utilizado. Nesse caso uma ordem mínima deve ser fixada para evitar erros de fusão.

- A *posição articulatória* obtida através da análise da distribuição espectral do sinal, permite classificar a vogal em **anterior** ou **posterior**. Propõe-se, nas vogais produzidas por crianças e adolescentes, separar as vogais anteriores das posteriores. Para tal será determinada a frequência abaixo da qual está contida parcela significativa da energia do quadro medida sobre o espectro LPC.

A proposta para separar as vogais anteriores das posteriores requer a definição do valor da frequência limiar ( $f_l$ ) que proporcionará a separação, assim como exige um procedimento para obter quanto da parcela da energia do quadro deve ser considerada.

Perfis espectrais de energia ( $P_\beta$ ) podem ser estabelecidos determinando-se, a partir do espectro de frequências ( $X(k)$ ,  $k = 1 \dots M$ ), a frequência abaixo da qual estão contidas uma determinada percentagem ( $\beta$ ) da energia total calculada sobre o espectro de frequências, desde a frequência nula até metade da frequência de amostragem. O valor de  $P_\beta$  é obtido por:

$$P_\beta = k_\beta \frac{f_a}{M} \quad (4.4)$$

onde  $f_a$  é a frequência de amostragem e  $k_\beta$  é tal que

$$\frac{1}{M} \left[ |X(0)|^2 + 2 \cdot \sum_{k=1}^{k_\beta} |X(k)|^2 \right] = \left( \frac{\beta}{100} \right) \cdot \frac{1}{M} \left[ |X(0)|^2 + \left| X\left(\frac{M}{2}\right) \right|^2 + 2 \cdot \sum_{k=1}^{\frac{M}{2}-1} |X(k)|^2 \right] \quad (4.5)$$

A determinação dos valores de  $f_1$  e  $\beta$  foi obtida por minimização da contagem de erros na classificação das vogais do banco de dados, variando-se  $\beta$  entre 25 e 95, em passo de 5, e  $f_1$  entre 1.000 Hz e 3.000 Hz, em passo de 100 Hz. Os valores de  $\beta$  e  $f_1$  obtidos foram de 70 e 2.000 Hz respectivamente. O parâmetro energia foi obtido sobre o espectro LPC, já obtido anteriormente para cálculo da primeira formante, sendo referido como  $P_{70s}$ , onde  $s$  é de suavizado para diferenciar o cálculo da energia através da DFT da seqüência original.

A utilização do parâmetro  $P_{70s}$  e a frequência limiar de 2.000 Hz para separação entre as anteriores e as posteriores propiciou uma classificação com taxa de erro de 1,3% para locuções produzidas por crianças com idades entre 7 e 10 anos e uma taxa de erro de 4,5% na locução produzida por pré-adolescentes com idade entre 11 e 14 anos. A vogal medial (/a/) apresenta  $P_{70s}$  acima ou abaixo de 2.000 Hz, devendo ser classificada somente a partir da altura da língua.

O procedimento detalhado para estimação da primeira formante e do parâmetro  $P_{70s}$  é descrito no apêndice C.

#### 4.3.2. Resultados obtidos

A Tabela 4.3 mostra os valores de  $F_1$  obtidos por análise LPC, e o parâmetro  $P_{70s}$  para crianças com idade entre 7 e 10 anos e pré-adolescentes com idade entre 11 e 14 anos.

Grupos	Parâmetros	Anteriores			Medial	Posteriores		
		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/
Crianças	F1 (Hz)	322	465	636	1122	759	526	411
	P70s (Hz)	4086	3627	3354	1942	1053	1059	808
Pré-adolescentes	F1 (Hz)	337	463	677	973	677	509	423
	P70s (Hz)	3455	3089	2712	1821	1462	1308	1002

Tabela 4.3:  $F_1$  e  $P_{70s}$  das vogais do PB, para crianças e pré-adolescentes.

Estes valores serão utilizados para classificar uma determinada locução sustentada como sendo uma determinada vogal.

#### 4.3.3. Procedimentos para classificação de vogais

O modelo proposto para classificar uma locução como sendo uma dada vogal utiliza duas medidas, a primeira formante ( $F_1$ ) para determinação do grau de abertura da língua e a frequência que contém 70% da energia do espectro suavizado ( $P_{70s}$ ) para determinação da posição articulatória.

A decisão sobre a posição articulatória é realizada pela medida de  $P_{70s}$ . Valores acima de 2.000 Hz indicam uma vogal anterior ou medial, enquanto que medidas de  $P_{70s}$  abaixo deste limiar indicam uma vogal posterior ou medial.

Após a decisão sobre a posição articulatória, é realizada a decisão sobre a altura da língua. Esta classificação é realizada pela medida de  $F_1$ . A decisão recai sobre a vogal que

apresenta valor médio de  $F_1$  com menor distância absoluta ( $DF_1$ ) do valor de  $F_1$  obtido da locução. A distância absoluta é dada por:

$$DF_1 = \text{abs}(F_{1/v1/} - F_{1/v2/}) \quad (4.6)$$

A classificação das vogais, para crianças e pré-adolescentes, é apresentada de forma sumária na Tabela 4.4.

	Crianças	Pré-adolescentes	Vogal
$P70_s < 2000$	$F_1 \leq 468,5$ Hz	$F_1 \leq 466,0$ Hz	/u/
	$468,5$ Hz $< F_1 \leq 642,5$ Hz	$466,0$ Hz $< F_1 \leq 593,0$ Hz	/o/
	$642,5$ Hz $< F_1 \leq 940,0$ Hz	$593,0$ Hz $< F_1 \leq 825,0$ Hz	/ɔ/
	$F_1 > 940,5$	$F_1 > 825,0$	/a/
$P70_s \geq 2000$	$F_1 \leq 393,5$ Hz	$F_1 \leq 400,0$ Hz	/i/
	$393,5$ Hz $< F_1 \leq 550,5$ Hz	$400,0$ Hz $< F_1 \leq 570,0$ Hz	/e/
	$550,5$ Hz $< F_1 \leq 879,0$ Hz	$570,0$ Hz $< F_1 \leq 825,0$ Hz	/ɛ/
	$F_1 > 879,0$	$F_1 > 825,0$	/a/

Tabela 4.4: Classificação das vogais em função de  $F_1$  e  $P70_s$ , para crianças e pré-adolescentes

A modelagem utilizada é apropriada para línguas como a portuguesa e/ou espanhola, que não utilizam séries intermediárias de vogais tônicas. Outras línguas, como

por exemplo a língua francesa, com séries intermediárias de vogais, exigiriam outro tipo de modelagem.

#### 4.4. Modelagem de fricativas

Sons fricativos são caracterizados pela formação de uma constricção no trato vocal, produzindo aceleração e turbulência no fluxo de ar proveniente dos pulmões, resultando em um ruído de fricção. O ruído primário excita os formantes associados à cavidade à frente da constricção, produzindo o som característico (FANT, 1960). As consoantes fricativas são contínuas, podendo ser prolongadas, em princípio, tanto quanto permita o ar pulmonar.

As consoantes fricativas da língua portuguesa são classificadas em labiodentais ou anteriores (/f/ surda e /v/ sonora), alveolares ou médias (/s/ surda e /z/ sonora), palatais ou posteriores (/ʃ/ surda e /ʒ/ sonora). Esses fonemas apresentam características espectrais relativamente fixas, embora sejam observadas variações que dependem dos locutores e/ou do contexto fonético (KLATT, 1987). Na língua portuguesa falada no Brasil, as consoantes fricativas correspondem a 12,12% dos sons (ALCAIM, SOLEWICZ & MORAES, 1992).

As fricativas surdas requerem audição em uma faixa de frequências acima de 2.000 Hz para sua detecção, faixa esta de frequências para as quais as crianças com deficiência auditiva usualmente apresentam reduzida sensibilidade auditiva (LING, 1978). Para CALVERT & SILVERMAN (1978), as fricativas estão no grupo de fonemas com aquisição mais difícil por parte dos deficientes auditivos.

A obtenção de um modelamento paramétrico sobre os sinais de voz que possibilite a discriminação das (e entre) fricativas e que permita a sua utilização no desenvolvimento de sistemas para o auxílio de deficientes auditivos é o objetivo desta sessão.

#### 4.4.1. As características das fricativas

LAUNAY & BOREL-MAISONNY (1989) observaram que as faixas de frequências dos espectros dos fonemas /ʃ, ʒ/ se estendem de 3.000 a 6.000 Hz, as dos fonemas /s, z/ se estendem de 4.500 a 10.000 Hz e as dos fonemas /f, v/ se espalham entre 1.000 Hz e 10.000 Hz, com o sinal laríngeo abaixo de 1.000 Hz para as sonoras.

RUSSO & BEHLAU (1993) obtiveram, para as consoantes fricativas, uma faixa de frequências entre 1.200 e 7.000 Hz para as anteriores, frequências variando de 4.500 a 8.000 Hz para as médias e frequências entre 2.500 Hz e 6.000 Hz para as posteriores.

Para RUSSO & BEHLAU (1993), as anteriores /f, v/ apresentam a intensidade mais reduzida de todas as consoantes do português. Para SANTOS (1987), essas consoantes são tão débeis, que alguns indivíduos praticamente não apresentam nenhuma área de intensidade visível na análise espectrográfica. FANT (1960) sugere que talvez devêssemos descrever esses sons como um fraco ruído de faixa larga, sem ressonâncias observáveis.

As Figuras 4.11, 4.13 e 4.15 apresentam espectrogramas de palavras com fricativas anteriores, médias e posteriores surdas, respectivamente. As Figuras 4.12, 4.14 e 4.16 mostram os espectrogramas para palavras com as fricativas sonoras correspondentes.

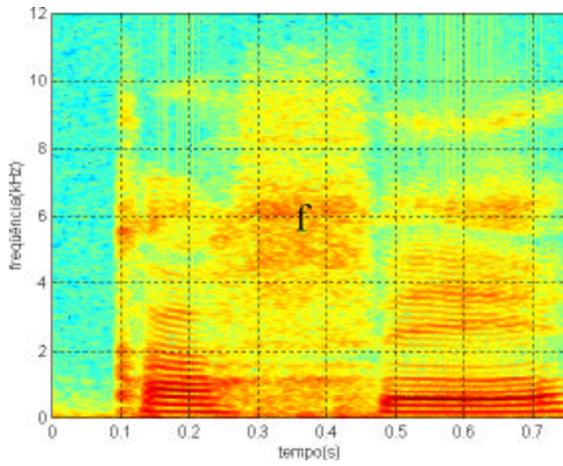


Figura 4.11: Espectrograma da palavra *café*

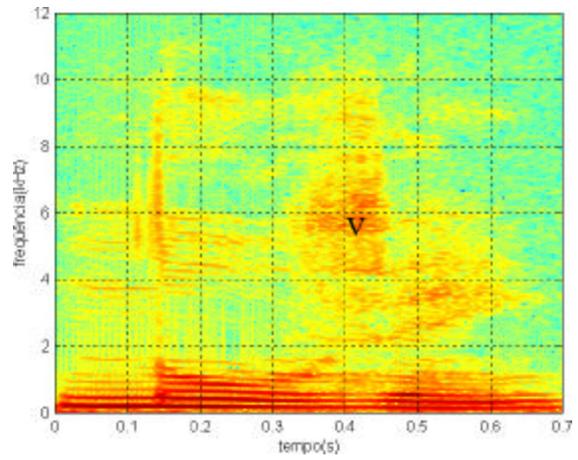


Figura 4.12: Espectrograma da palavra *nuvem*

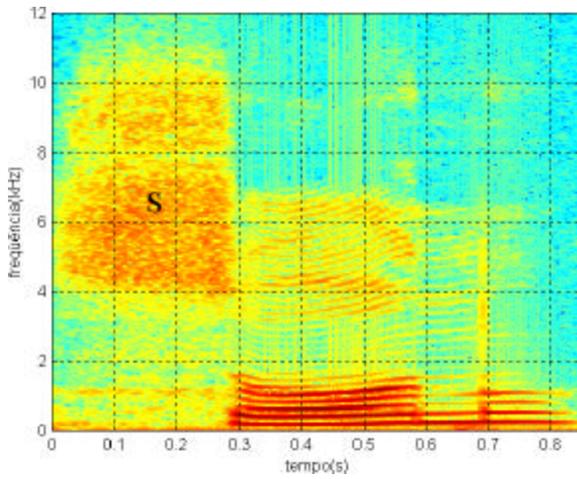


Figura 4.13: Espectrograma da palavra *sono*

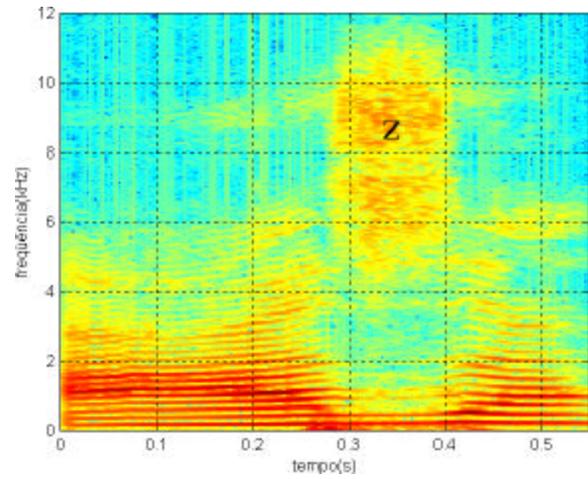


Figura 4.14: Espectrograma da palavra *asa*

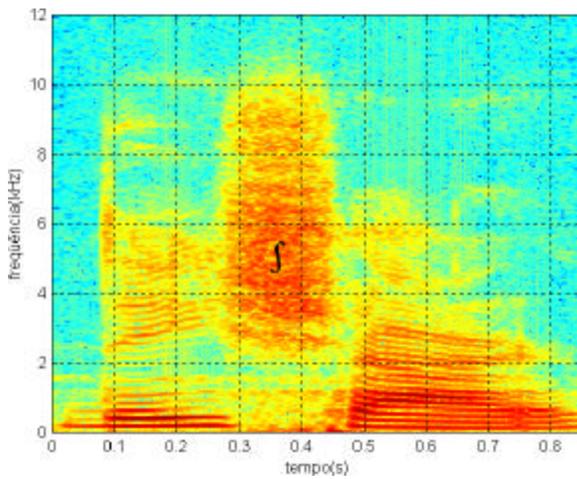


Figura 4.15: Espectrograma da palavra *deixar*

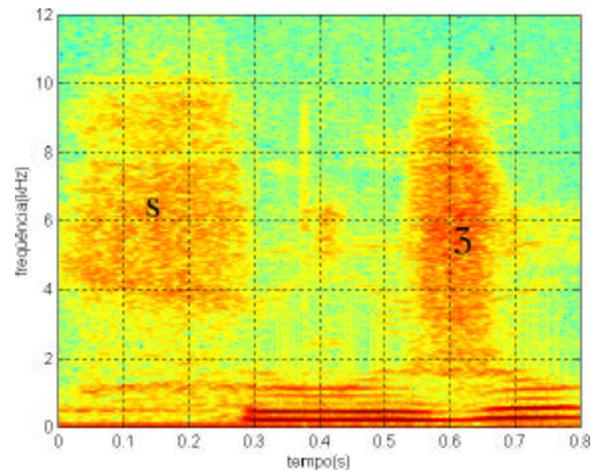


Figura 4.16: Espectrograma da palavra *sujo*

Nas figuras anteriores, pode-se observar que os fonemas fricativos apresentam componentes significativas de elevadas frequências com duração superior a 50 ms. Os espectrogramas mostram que as fricativas sonoras diferem das surdas pela presença (ou ausência) do traço de sonoridade nas frequências abaixo de 2.000 Hz. As fricativas posteriores são mais graves que as demais. As anteriores e médias são mais agudas. As anteriores apresentam um espectro mais disperso.

A existência de componentes significativas em altas frequências, sugere a utilização da taxa de cruzamentos por zero para modelar e localizar quadros contendo fricativas, em especial as surdas (RABINER & SCHAEFER, 1978; VIEIRA, 1989). O centro de gravidade espectral definido como sendo a frequência abaixo (ou acima) da qual estão concentradas 50% da energia também pode ser utilizado para caracterização das fricativas. KENT & READ (1992) sugerem a frequência para a qual a amplitude do espectro é máxima, a frequência de pico, para diferenciar /s/ de /ʃ/.

As Figuras 4.17, 4.18 , 4.19, 4.20, 4.21 e 4.22 mostram a trajetória obtida sobre as medidas, tomadas quadro a quadro, da taxa de cruzamentos por zero, do centro de gravidade (CG) e a da frequência de pico ( $F_{\text{pico}}$ ) para as locuções cujos espectros são apresentados nas figuras 4.11 a 4.16, respectivamente. A taxa de cruzamentos por zero foi normalizada, sendo apresentada em número de cruzamentos por zero por segundo (NCZ/s). Os sinais foram adquiridos com 16 bits por amostra, frequência de amostragem de 44.100 Hz, divididos em quadros de 1024 amostras (23,2 ms), sem superposição entre quadros.

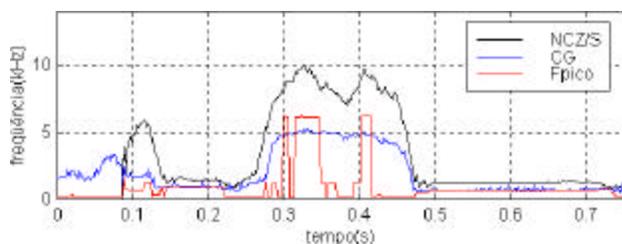


Figura 4.17: NCZ/s, CG e  $F_{pico}$  para *café*

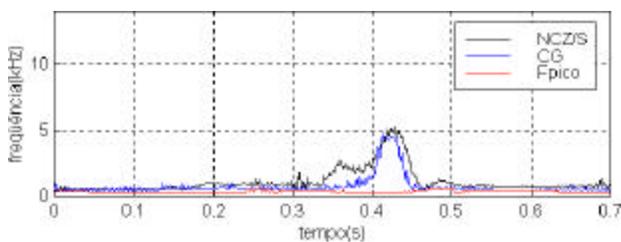


Figura 4.18: NCZ/s, CG e  $F_{pico}$  para *nuvem*

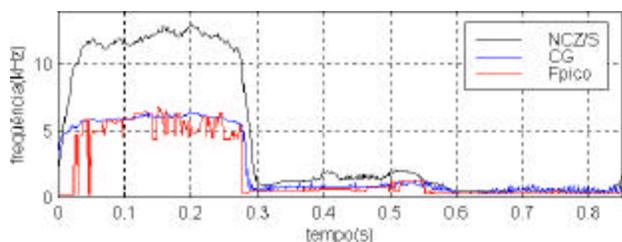


Figura 4.19: NCZ/s, CG e  $F_{pico}$  para *sono*

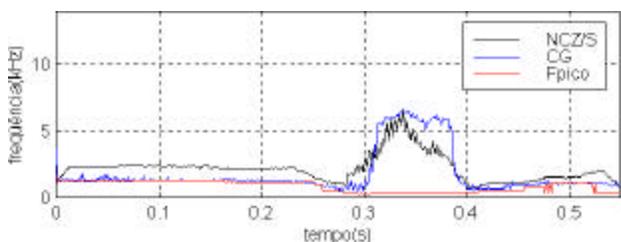


Figura 4.20: NCZ/s, CG e  $F_{pico}$  para *asa*

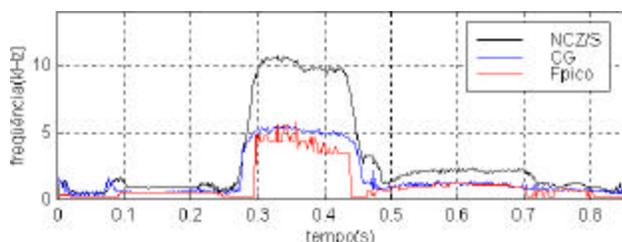


Figura 4.21: NCZ/s, CG e  $F_{pico}$  para *deixar*

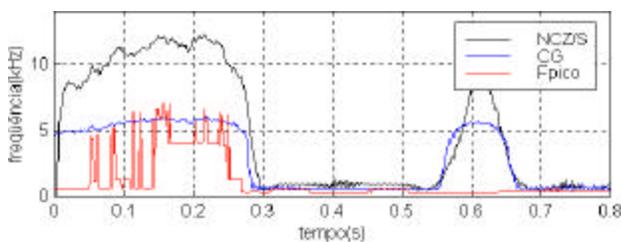


Figura 4.22: NCZ/s, CG e  $F_{pico}$  para *sujo*

Os resultados apresentados nas Figuras 4.17 a 4.22 indicam que a taxa de cruzamentos por zeros e o centro de gravidade apresentam menores variações intra fonema que a frequência de pico. Os resultados mostram que, para a mesma posição articulatória, o centro de gravidade é o parâmetro com menor variação, quando é modificada apenas a natureza da fonte (surda/sonora).

Nas locuções surdas, o NCZ/s mostra relação com a maior frequência com energia significativa do espectro, enquanto o CG e a  $F_{pico}$  mostram maior relação com as frequências onde concentram-se as componentes mais significativas do espectro. Estes resultados estão coerentes com os obtidos com sinais tonais puros, quando o NCZ/s deve

ser igual a duas vezes a frequência do tom e o centro de gravidade e a frequência de pico devem ser da ordem da frequência do tom.

Os valores obtidos para a taxa de cruzamentos por zero e o centro de gravidade espectral são maiores para fonemas fricativos médios surdos que os obtidos para os fonemas fricativos posteriores surdos.

Os parâmetros avaliados, em especial NCZ/s e CG, mostraram diferenças nas medidas entre quadros fricativos /s/ e /ʃ/.

As fricativas apresentam características espectrais definidas (FANT, 1960; SANTOS, 1987; LAUNAY & BOREL-MAISONNY, 1989), portanto a frequência de pico do espectro e/ou o centro de gravidade podem ajudar a discriminação das (e entre as) fricativas. Deve-se buscar parâmetros que reduzam a influência de picos locais do espectro e/ou da presença (ou ausência) da sonoridade.

Neste trabalho, propõe-se um novo parâmetro, a medida da frequência de pico do espectro suavizado (FPES), que consiste na medida da frequência associada ao par de pólos complexos conjugados resultante da análise LPC de ordem 3 sobre o sinal de voz (ARAÚJO & VIOLARO, 2000).

Para sinais de voz de fricativas adquiridos com altas frequências de amostragem (ex. 44.100 Hz), as harmônicas da frequência fundamental estão próximas à frequência zero e podem ser modeladas com a imposição de um pólo real, possibilitando ao par de pólos complexos conjugados remanescentes acompanharem a frequência de pico do espectro.

As Figuras 4.23 a 4.28 apresentam, para as palavras exemplificadas anteriormente o espectrograma e a trajetória da frequência de pico do espectro suavizado (FPES). Os sinais foram divididos nos mesmos quadros utilizados para a obtenção de NCZ/s, CG e Fpico.

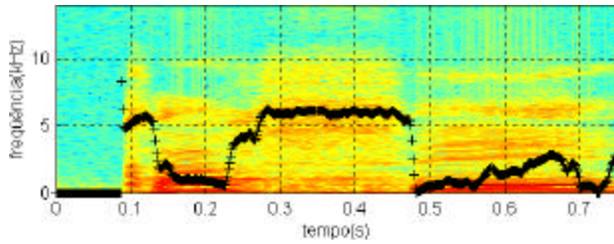


Figura 4.23: FPES da palavra *café*

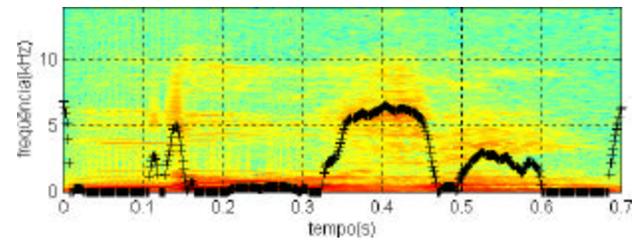


Figura 4.24: FPES da palavra *nuvem*

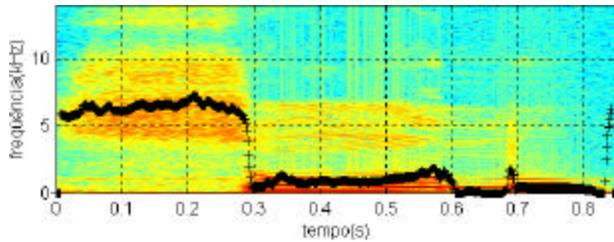


Figura 4.25: FPES da palavra *sono*

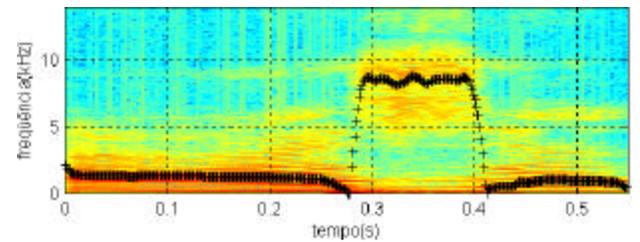


Figura 4.26: FPES da palavra *asa*

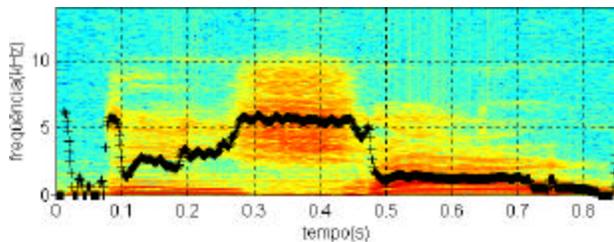


Figura 4.27: FPES da palavra *deixar*

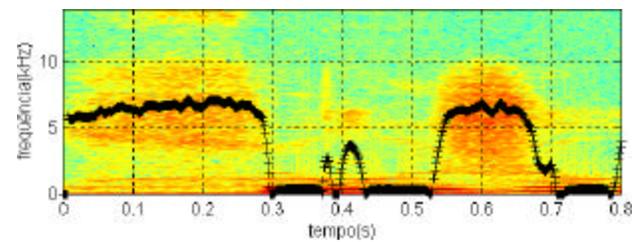


Figura 4.28: FPES da palavra *sujo*

Analisando-se os resultados obtidos para a trajetória da frequência de pico do espectro suavizado sobre as palavras utilizadas, pode-se verificar que a sonoridade produz pouco efeito sobre a FPES, pois é modelada pelo pólo real, permitindo ao par de pólos complexos conjugados restante modelar o ruído de fricção. A suavização proposta foi suficiente para reduzir efeitos de picos espúrios, sem eliminar a informação sobre a região de maior concentração da energia espectral.

A medida de FPES sobre as consoantes oclusivas pode atingir valores elevados nos momentos da oclusão com curta duração, enquanto nas fricativas o valor de FPES permanece com valores elevados por mais de 50 ms.

A Figura 4.29 apresenta uma estimativa da função densidade de probabilidade (fdp) das medidas da frequência de pico do espectro suavizado de fricativas. Essas medidas foram obtidas por fonação sustentada por 8 crianças com idade entre 7 e 10 anos, obtidas em 40 quadros de 46 ms.

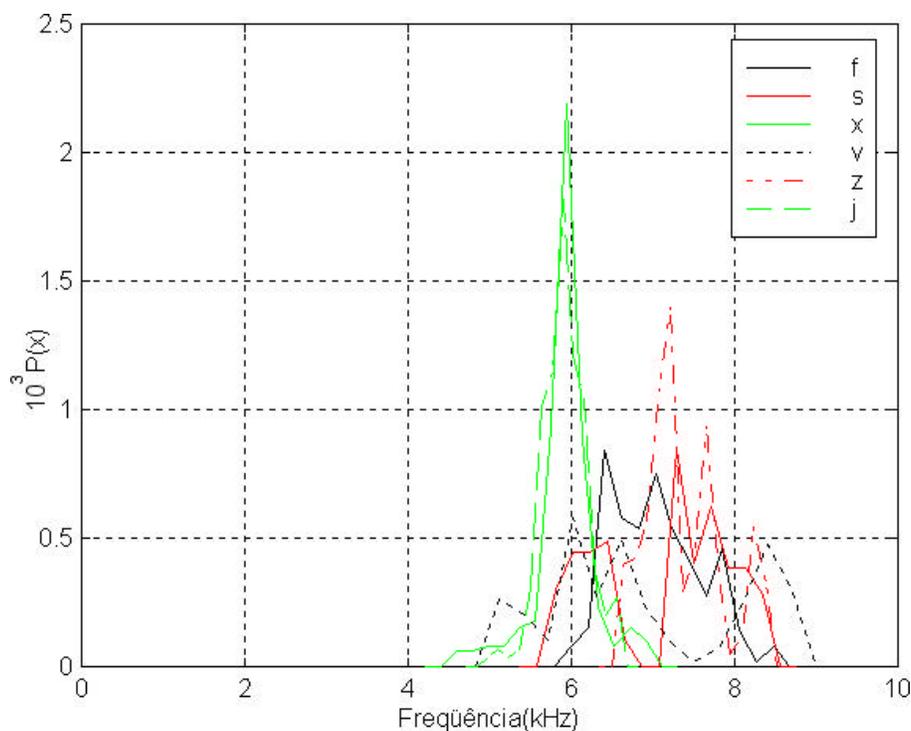


Figura 4.29: fdp das medidas da FPES de fricativas sustentadas.

As medidas da frequência de pico do espectro suavizado estão entre 4.500 Hz e 8.850 Hz, sendo que 99,3% das medidas estão acima de 5.000 Hz.

A Tabela 4.4 apresenta o valor médio e o desvio padrão (em Hz) da medida da frequência de pico do espectro suavizado em função do ponto de articulação e da sonoridade.

<i>Zona de Articulação</i>	<i>Surda</i> (/f/, /s/, /ʃ/)	<i>Sonora</i> (/v/, /z/, /ʒ/)	<i>Sur+Son</i>
Lábiodentais (La)	7031 (548)	6883 (1148)	6957 (903)
Alveolares (Al)	7130 (809)	7378 (481)	7246 (686)
Palatais (Pa)	5928 (366)	5921 (276)	5925 (328)
La+Al+Pa	6739 (808)	6788 (967)	67634 (888)

Tabela 4.4: Valor médio (desvio padrão) , em Hz, da medida de FPES das fricativas

As fricativas palatais (mais graves) apresentam menores valores de FPES (média de 5.925 Hz) e as fricativas anteriores e mediais apresentam valores médios de FPES mais elevados (6.957 Hz e 7.246 Hz, respectivamente). Os valores de pico das fdp's, assim como valores médios de FPES das anteriores são menores que os correspondentes das mediais, resultados concordantes com as análises apresentadas na literatura (SANTOS, 1987; LAUNAY & BOREL-MAISONNY, 1989; KENT & READ, 1992; RUSSO & BEHLAU, 1993). As medidas da FPES apresentaram valores médios de 6.739 Hz para as surdas e 6.788 Hz para as sonoras, com uma variação inferior a 0,73%, mostrando que o parâmetro FPES é bastante imune à natureza da fonte (sonora ou surda), conforme foi especificado.

As medidas de FPES das fricativas palatais são menores e com menor desvio padrão que as demais, apresentando-se bastante concentradas pouco abaixo de 6.000 Hz, permitindo assim realizar a sua classificação.

A Figura 4.30 apresenta uma estimativa da função densidade de probabilidade agrupando-se as medidas de FPES sobre as palatais e as medidas sobre as labiodentais e alveolares, independente da sonoridade.

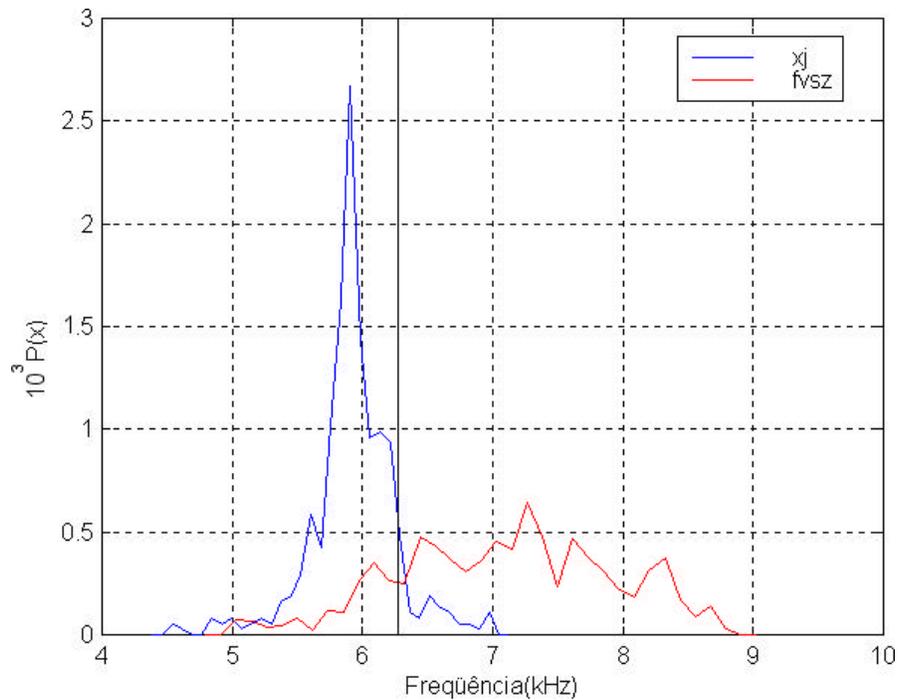


Figura 4.30: fdp das medidas de FPES das fricativas palatais x labiodentais e alveolares.

Na figura 4.30, está traçado um limiar de decisão obtido por contagem, minimizando a probabilidade de erro. A frequência limiar para decisão palatal x labiodental e alveolar é de 6.270 Hz (média), resultando em 14,3% de erros.

A regra resultante para classificação de uma fricativa palatal é FPES entre 5.000 Hz e 6.270 Hz. A decisão sonora/surda deve ser realizada independentemente. Em consequência, FPES superiores a 6.270 Hz indicam que o fonema deve ser labiodental ou alveolar. A FPES é insuficiente para realizar uma decisão labiodental x alveolar entre as fricativas.

A literatura (FANT, 1960, SANTOS, 1987; KENT & READ, 1992; RUSSO & BEHLAU, 1993) e a observação dos registros espectrográficos de palavras contendo fricativas labiodentais e alveolares mostram que as alveolares apresentam-se mais

concentrados espectralmente, enquanto as labiodentais apresentam espectros mais espalhados.

A realização da decisão entre labiodentais e alveolares pode ser realizada com uma medida do espalhamento espectral (ARAÚJO & VIOLARO, 2000). O coeficiente de espalhamento espectral (CEE) proposto neste trabalho é dado por:

$$CEE = \frac{P_{80} - P_{20}}{P_{20}} \quad (4.7)$$

onde  $P_{\beta}$  é a frequência abaixo do qual esta concentrada  $\beta\%$  da energia do espectro do sinal acima de 2.000 Hz. A utilização das componentes acima de 2.000 Hz visa extrair as componentes relativas ao traço de sonoridade.

A Figura 4.31 apresenta as funções densidade de probabilidade das medidas do coeficiente de espalhamento espectral para as fricativas médias e anteriores.

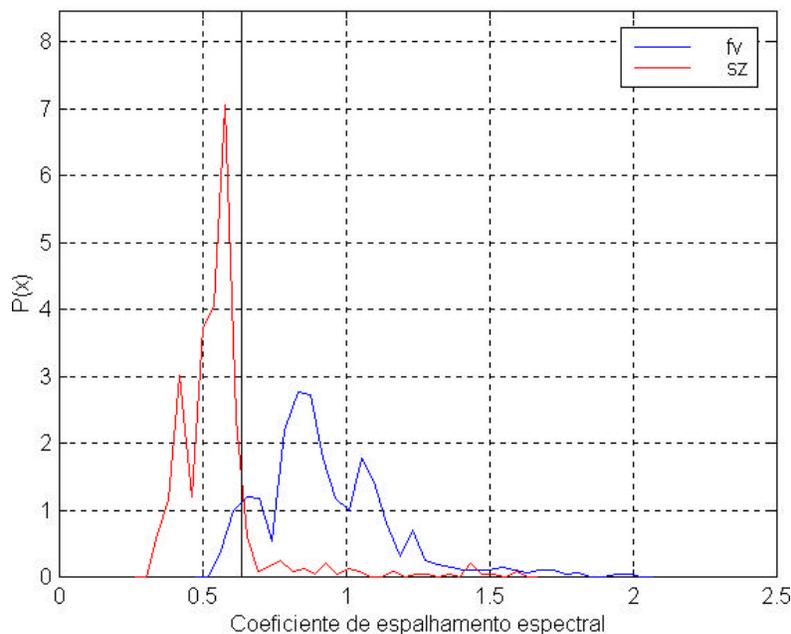


Figura 4.31: fdp das medidas de CEE das fricativas médias x anteriores

As fdp's mostram que o espalhamento espectral das médias é menor que o espalhamento espectral das anteriores. A Figura 4.31 apresenta, ainda, o limiar de decisão obtido minimizando a probabilidade de erro, por contagem. O limiar obtido sobre o coeficiente de espalhamento espectral para decisão anterior x médias foi de 0,64, resultando em 7,79% de erros. O limiar foi obtido para crianças entre 7 e 10 anos em fonação sustentada, sendo portanto válidos somente para esta faixa etária e na condição especificada. Para outras idades e para fala fluente o limiar deverá ser alterado.

Em síntese, as principais medidas utilizadas para a caracterização de fricativas (NCZ e CG) são fortemente afetadas pela presença (ou ausência) de sonoridade. Essas medidas podem ficar mais imunes à sonoridade eliminando-se, por filtragem, as componentes abaixo de 2.000 Hz, mas uma filtragem pode fazer com que medidas sobre outros fonemas, como as vogais /i, e/, possam ser confundidas com as fricativas.

Para sinais adquiridos com frequências de amostragem altas, a utilização de um filtro preditor com 3 pólos para modelar o pico do espectro permitiu que um pólo real modelasse a sonoridade e possibilitou que o par de pólos complexos conjugados remanescentes modelasse a frequência de pico do espectro. O modelo foi suficiente para reduzir efeitos de picos espúrios, mantendo as informações sobre a faixa de frequências com maior concentração do espectro de energia.

A avaliação de sinais fricativos com crianças entre 7 e 10 anos permitiu verificar que as fricativas apresentam FPES superior a 5.000 Hz. Para crianças na faixa etária indicada e para fonação sustentada, a FPES média obtida foi de 6.957 Hz para labiodentais, 7.254 Hz para alveolares e de 5.609 Hz para palatais.

A FPES média é de 6.486 Hz sobre quadros surdos e de 6.517 Hz sobre os quadros sonoros, o que corresponde a uma variação inferior a 0,73%, indicando que este parâmetro é relativamente insensível à sonoridade, como desejado.

O procedimento detalhado para estimação de FPES e CEE é descrito no apêndice D.

A utilização da FPES permite caracterizar uma fricativa e ainda possibilita caracterizar as palatais. O coeficiente de espalhamento espectral obtido sobre quadros de sinais labiodentais é maior que o coeficiente de espalhamento espectral sobre quadros de alveolares, permitindo realizar a classificação entre estas duas zonas de articulação. A sonoridade é calculada de acordo com o algoritmo descrito na seção 4.2.1.

A classificação das fricativas para locução sustentada por crianças é apresentada de forma sucinta na Tabela 4.5.

FPES	$CEE < 0,64$		$CEE \geq 0,64$	
	Son	Sur	Son	Sur
$FPES < 5000 \text{ Hz}$	Não é fricativa			
$5.000 \text{ Hz} \leq FPES < 6.270 \text{ Hz}$	/ʒ/	/ʃ/	/ʒ/	/ʃ/
$FPES \geq 6.270 \text{ Hz}$	/z/	/s/	/v/	/f/

Tabela 4.5: Classificação das fricativas em função de FPES, CEE e sonoridade para crianças.



## *Capítulo 5*

# Aplicações e resultados

A natureza multidisciplinar do trabalho exigiu, ao longo do seu aprimoramento, o apoio de outros profissionais, em especial de fonoaudiólogos. As aplicações descritas neste capítulo foram realizadas com o protótipo base, desenvolvido anteriormente em trabalho conjunto realizado por professores e alunos da Universidade Federal do Pará e do Centro Federal de Educação Tecnológica do Pará (CASTRO e col., 1995).

O protótipo base dos jogos para auxílio no aprimoramento do controle fonoarticulatório de deficientes auditivos foi instalado no Centro de Estudos e Pesquisas em Reabilitação "Prof. Gabriel Porto" - CEPRE da Universidade Estadual de Campinas e na Escola Paulista de Medicina - UNIFESP, para fins de avaliação de sua aplicabilidade, viabilidade e implicações no aprimoramento do controle da voz dos deficientes auditivos. Neste capítulo serão apresentados alguns dos resultados obtidos nos estudos conduzidos por fonoaudiólogas nas duas instituições.

As aplicações utilizaram dois jogos, um para aprimoramento do controle respiratório e energia, e o outro para aprimoramento do controle da frequência de vibração das pregas vocais.

Estas aplicações com o protótipo base foram importantes para nortear o novo sistema desenvolvido no presente trabalho.

## 5.1. Aplicações

O trabalho inicial de aplicação dos jogos foi realizado no CEPRE com um pré-adolescente de sexo masculino com idade de 10,5 anos, perda auditiva neurosensorial profunda, cursando o primeiro grau em escola regular. O estudo ocorreu ao longo do ano de 1998. Em abril, a frequência fundamental medida foi de 680 Hz e, após 7 meses em terapia de 15 minutos por semana, a frequência fundamental foi reduzida para 280 Hz. A criança, que não apresentava nenhuma produção vocal no início da terapia, durante a terapia apresentou um incremento considerável na sua produção oral (LIMA, ARAÚJO & ROMANO, 1999).

O segundo trabalho, realizado ainda no CEPRE, com um grupo de 7 crianças com surdez moderada, severa e profunda, participantes do Programa de Apoio à Escolaridade de Crianças, Adolescentes e Adultos, utilizou os programas uma vez por semana, durante 15 minutos.

Foram realizadas 3 avaliações ao longo do ano: uma no início do programa, uma após 3 meses e uma ao final do ano. Considerou-se tanto avaliações de parâmetros da voz quanto da produção sonora. Foram medidos parâmetros como tempo máximo de fonação e frequência fundamental. Para comparação, os resultados são apresentados tanto no seu valor absoluto quanto relativo a crianças ouvintes com mesma faixa etária.

Na avaliação prévia, a média do tempo máximo de fonação das crianças era de 5,76 s, que corresponde a 42,4% da média de crianças ouvintes de mesma faixa etária, que é de 10,0 s (COLTON & CASPER, 1996). A média da frequência fundamental era de 368 Hz, 56,6 % acima do padrão ouvinte de mesma faixa etária, que é de 235 Hz (BEHLAU & PONTES, 1995). Após 3 meses utilizando os jogos, a média do tempo máximo de fonação foi aumentada para 6,88 s ficando em 32,2% abaixo do padrão ouvinte. A média da frequência fundamental foi reduzida para 320 Hz, ficando em 36,2% acima do padrão considerado. Ao final do ano, a média do tempo máximo de fonação foi ampliada para 7,75 s, ou seja, 22,6% menor que a média do padrão considerado. A média da frequência fundamental ficou reduzida a 256 Hz, ou seja, 8,93% maior que o padrão ouvinte de mesma faixa etária.

As avaliações sobre a quantidade e qualidade da produção da fala foram realizadas em interação com as crianças, sendo possível observar que os alunos passaram a usar mais a emissão oral como forma de estabelecer contato com as outras pessoas (ARAÚJO & LIMA, 1999).

Outro trabalho de avaliação do protótipo inicial foi feito na Universidade Federal de São Paulo. Este trabalho foi realizado com 20 indivíduos (18 do sexo feminino e 2 do sexo masculino), na faixa etária dos oito aos vinte e oito anos, com perda auditiva neurosensorial, bilateral, de grau severo a profundo. As perdas auditivas eram congênitas ou adquiridas até o segundo ano de vida. Dos vinte indivíduos da amostra, apenas 7 faziam uso de Próteses Auditivas. Destes 7 indivíduos, seis apresentavam limiares em torno de 60 dB NA (nível de audição) e um em torno de 90 dB NA. Todos os indivíduos da amostra tiveram pelo menos cinco anos de tratamento fonoaudiológico e apresentavam domínio básico da língua oral, para entender e se fazer entender verbalmente, na comunicação

interpessoal. Dos 20 indivíduos, quatro haviam estudado em escolas para ouvintes e os outros 16 em escolas especiais para surdos. Estas escolas utilizavam métodos de ensino Comunicação Total e Verbo-Tonal. Nenhum dos participantes tinha sido submetido anteriormente a qualquer tipo de terapia vocal específica (BOMMARITO, 2000).

O programa de terapia idealizado e conduzido por BOMMARITO (2000) foi composto de 16 sessões de terapia vocal, com duração de 1 hora cada sessão, duas vezes por semana. Os primeiros 30 minutos foram destinados a exercícios de voz para adequar a qualidade vocal e os outros 30 minutos a atividades com os jogos. Os componentes da amostra foram separados em cinco grupos de quatro indivíduos, com idades aproximadas.

Cada grupo de 4 indivíduos realizou 1 hora/semana de exercícios com os jogos. Portanto, cada membro do grupo utilizou os jogos por um tempo médio de 15 minutos/semana.

Foram realizadas medidas sobre: frequência fundamental modal, frequência fundamental máxima, frequência fundamental mínima, *jitter* (medida da oscilação da frequência fundamental em locução sustentada), *shimmer* (medida da variação da intensidade do sinal em locução sustentada) e ERG (energia de ruído glótico). As medidas foram realizadas com auxílio do aplicativo VOICE ASSESSMENT, parte do programa *Dr. Speech Sciences*, versão 3.0, da TIGER DSR.

A análise acústica da voz foi realizada na primeira sessão e após cada 15 dias de terapia, perfazendo um total de cinco avaliações acústicas durante os 60 dias. Foi realizada, ainda, uma avaliação perceptivo-auditiva da voz, sobre registros da emissão prolongada da vogal /a/, da contagem dos números de um a quinze e da repetição de cinco frases. As amostras foram apresentadas de maneira aleatória, sendo considerados os aspectos: a) qualidade vocal; b) ressonância; c) inteligibilidade de fala. A avaliação foi realizada por

quatro especialistas em voz do Centro de Estudos da Voz – CEV, os quais deviam fornecer um grau entre zero e três pontos para cada um dos itens considerados. Após o procedimento terapêutico realizado por BOMMARITO, a pesquisadora constatou valores estatisticamente significantes quanto a:

- a) Redução da frequência fundamental modal e da frequência fundamental máxima.
- b) Redução do *jitter* e do *shimmer*.
- c) Redução da energia do ruído glótico.
- d) Aumento dos tempos máximos de fonação das vogais sustentadas e dos sons fricativos /s/ e z/.
- e) Diminuição do grau geral de alteração vocal, do grau de ressonância hipernasal, e do grau de ressonância faríngea.

Não houve variações estatisticamente significantes no que se refere ao grau de rouquidão da voz, soproidade, astenia, tensão vocal, instabilidade vocal e hiponasalidade.

É interessante destacar que entre os 20 indivíduos, dois do sexo feminino apresentavam produção oral tipo vocal *fry*, com frequências fundamentais de 162,9 Hz e 166,7 Hz, abaixo do padrão para o sexo feminino. Nestes casos os jogos foram utilizados para aumentar a frequência fundamental. Ao final da terapia, a frequência fundamental modal destes dois indivíduos estava em 193,3 Hz e 209,2 Hz, respectivamente.

Os dois únicos sujeitos do sexo masculino apresentavam frequências fundamentais modais de 294,8 Hz e 158,6 Hz e, após a terapia, apresentaram frequências fundamentais modais de 191,2 Hz e 127,9 Hz.

A percentagem de vozes melhoradas foi de 80%. Os demais 20% não apresentaram melhorias na voz, não havendo em nenhum caso piora da voz. Para a autora, uma das causas da melhora observada no estudo pode ser justificada pelo fato de ter sido realizado um trabalho em nível da fonte glótica, o que proporcionou uma maior coaptação glótica, melhor produção desses sons e, conseqüentemente, melhora na inteligibilidade da fala.

## 5.2. Considerações sobre as aplicações

Os estudos realizados no Centro de Estudos e Pesquisas em Reabilitação "Prof. Gabriel Porto" da Universidade Estadual de Campinas e no curso de Fonoaudiologia da Escola Paulista de Medicina - UNIFESP, mostraram que aplicativos lúdicos controlados por parâmetros de voz auxiliam no aprimoramento do controle da frequência fundamental, no aumento do tempo máximo de fonação e inteligibilidade da fala, fatores que contribuíram para o aumento da quantidade e qualidade da produção oral de pessoas com deficiência auditiva.

No CEPRE, o atendimento foi individual, de 15 minutos por semana, por um período de 7 meses, com interrupção de 45 dias. Na UNIFESP o atendimento foi em grupo de 4, em duas sessões semanais de 30 minutos, resultando em tempo médio de 15 minutos por semana por sujeito, em um período de 2 meses sem interrupção.

No CEPRE foram atendidos basicamente pré-adolescentes, com idades variando entre 10 e 14 anos. Na UNIFESP foram atendidos desde crianças com 8 anos de idade até adultos com 28 anos.

Os aplicativos proporcionaram condições para realizar reduções drásticas da frequência fundamental (680 Hz para 280 Hz), assim como possibilitaram, quando necessário, o aumento da frequência fundamental. Em especial, o aprimoramento do controle fonoarticulatório proporcionou aumento da quantidade e qualidade da produção oral.

A interação com os pesquisadores que realizaram as aplicações serviu para a definição dos novos jogos resultantes deste trabalho.



## *Capítulo 6*

# Considerações finais

O principal objetivo deste trabalho foi o desenvolvimento de jogos computacionais para auxiliar deficientes auditivos no aprimoramento do controle fonoarticulatório. As crianças e adolescentes com perdas auditivas constituem o público alvo prioritário.

As dificuldades específicas dos deficientes auditivos no controle articulatório nortearam o desenvolvimento deste trabalho. As dificuldades com o controle da respiração, intensidade da voz e frequência de vibração das pregas vocais, assim como problemas com a produção de vogais e fricativas foram especialmente considerados.

As vogais são os fonemas mais utilizados na língua falada no Brasil, correspondem a mais de 54% dos sons de nossa língua. Além disso, as vogais utilizadas na língua portuguesa formam um conjunto próprio, que difere dos conjuntos de vogais de outras línguas, o que exige o desenvolvimento de um algoritmo adequado para discriminação das vogais da língua portuguesa falada no Brasil.

As consoantes fricativas, que contribuem com mais de 12% dos sons da nossa língua, mereceram uma atenção especial neste trabalho, pois de acordo com a literatura são os sons mais frequentemente omitidos na produção dos deficientes auditivos. Além disso, as consoantes fricativas requerem a utilização de todos os movimentos que a língua permite

realizar, possibilitando o exercício para aprimoramento do controle do principal órgão utilizado na articulação, a língua.

Os estudos realizados com o protótipo base, desenvolvido antes do início deste trabalho, mostraram que jogos computadorizados controlados por voz são efetivos para aprimoramento do controle da frequência fundamental, para o aumento do tempo máximo de fonação e da inteligibilidade da fala, fatores que contribuíram para o aumento da quantidade e qualidade da produção oral de pessoas com deficiência auditiva.

Os jogos, desenvolvidos anteriormente a este trabalho, mostraram-se adequados para utilização com crianças, pré adolescentes e adultos. Foram utilizados para reduzir drasticamente a frequência fundamental (680 Hz para 280 Hz), assim como possibilitaram, quando necessário, aumentar a frequência fundamental.

Em geral, jogos computadorizados controlados por voz proporcionaram:

- a) Ajuste da frequência fundamental.
- b) Redução do *jitter*, do *shimmer* e da energia do ruído glótico.
- c) Aumento dos tempos máximos de fonação das vogais sustentadas e dos sons fricativos /s/ e z/.
- d) Diminuição do grau geral de alteração vocal, do grau de ressonância hipernasal, do grau de ressonância faríngea.
- e) Importante percentagem de vozes melhoradas.

Os estudos mostraram que aplicativos computacionais lúdicos controlados por voz são eficientes para aprimoramento do controle de voz de deficientes auditivos, em procedimentos de curta duração (2 e 7 meses) e utilizados por curtos períodos de tempo (15 minutos por semana). Em especial, proporcionam aumento da quantidade e qualidade da produção oral.

O desenvolvimento do sistema base foi prejudicado por diversos fatores, entre os quais: a falta de informações sobre as características da vozes de crianças e adolescentes que pudessem servir de referência para os jogos; a inexistência de estudos sobre a aplicabilidade dos jogos para crianças e adolescentes com deficiência auditiva; a indefinição sobre quais eram as características mais importantes da voz dos deficientes auditivos que deveriam ser consideradas no desenvolvimento dos jogos; a falta da definição dos parâmetros mais adequados para o controle dos jogos.

A montagem de um banco de dados com locuções realizadas por crianças e adolescentes possibilitou o desenvolvimento de novos algoritmos e permitiu a obtenção dos valores médios dos parâmetros propostos e utilizados, em especial de vozes de crianças e adolescentes.

Os requisitos fixados para os algoritmos foram respostas que pudessem ajustar-se aos mais diversos padrões de voz, precisão tão alta quanto possível e tempo de processamento em intervalo inferior a um quadro de aquisição, tal que não prejudicasse a animação. A base para tempo de processamento foi tomado sobre microcomputadores tipo Pentium com relógio de 100 MHz, permitindo que os jogos possam ser executados em qualquer computador tipo PC, com relógio de frequência maior ou igual a 100 MHz.

A velocidade dos algoritmos, requerida para as aplicações, exige modelos matemáticos rápidos. A precisão necessária para propiciar uma retroalimentação conseqüente exige modelos matemáticos precisos. A enorme variabilidade da voz, em especial da voz de deficientes auditivos, exige algoritmos com grande adaptação. Conciliar rapidez, precisão e adaptabilidade foi o grande desafio deste projeto. Para superar este desafio foi realizado um minucioso estudo sobre as características físicas dos sons da língua. Partindo das propriedades físicas dos sons, foi possível desenvolver modelos

matemáticos relativamente simples, precisos e com boa adaptabilidade.

No presente trabalho, que culminou com o desenvolvimento de um sistema denominado *Jogos de Voz*, foram propostos vários algoritmos para extração de parâmetros da voz e discriminação de fonemas, tais como algoritmos para: medida da energia; medida da frequência fundamental; medidas para discriminação de vogais como a medida de formante de vogais e medida da posição articulatória ( $P_{70s}$ ); medidas para discriminação de fricativas como a medida da frequência de pico do espectro suavizado e a medida do coeficiente de espalhamento espectral.

A definição dos objetivos dos aplicativos, a concepção dos formatos dos exercícios a serem realizados e sua adequação a cada estágio de controle articulatório foram outra preocupação ao longo desse trabalho.

Estudos sobre a característica da voz de deficientes auditivos registram que estes apresentam, em geral, locução com frequências fundamentais elevadas. Neste trabalho foi mostrado a necessidade da adequação da fonte ao filtro, sendo verificado a necessidade do ajuste da frequência de vibração das pregas vocais às exigências das frequências de ressonância do trato vocal. Particularmente, foi mostrado que uma frequência fundamental muito elevada é imprópria para a produção dos sons da língua.

O algoritmo proposto para medir a frequência fundamental, foi capaz de extraí-la de vozes produzidas por homens, mulheres, adolescentes e crianças ouvintes e, principalmente, de crianças e adolescentes portadores de deficiência auditiva. O algoritmo se mostrou efetivo para determinação da frequência fundamental com valores variando desde 90 Hz até 740 Hz.

O algoritmo para discriminação de vogais utiliza dois parâmetros, a primeira formante e a frequência que contém 70% da energia do espectro LPC ( $P_{70s}$ ). A primeira

formante, associada à altura da língua, é um parâmetro já consagrado na literatura. A segunda formante, outro parâmetro consagrado, cujo valor é associado à posição articulatória, mostrou-se efetivo para medidas de um único locutor, mas seus valores tendem a uma grande superposição para as várias vogais produzidas por diversos locutores. A utilização da frequência que contém 70% da energia do espectro LPC, parâmetro proposto neste trabalho, permitiu separar as vogais anteriores das posteriores com probabilidade de erro inferior a 1,7% para crianças até 10 anos e inferior a 4,5% para pré-adolescentes com idades entre 11 e 14 anos, valores otimizados na base de dados de treinamento.

Uma aplicação óbvia para o aprimoramento do controle da produção das vogais é o traçado das medidas das duas primeiras frequências formantes ( $F_1$  e  $F_2$ ) sobre um plano. Neste caso, cada vogal é associada a uma região no plano  $F_1 \times F_2$ . Esta aplicação é prejudicada pela dificuldade da medida simultânea de  $F_1$  e  $F_2$  que seja suficientemente rápida, precisa, e que possa adequar-se às diferentes características dos locutores, tais como sexo e idade. Além disso, essa aplicação também é prejudicada pela grande variabilidade dos valores da segunda formante. Nesse caso, a criança com deficiência auditiva deveria alcançar simultaneamente dois alvos relativamente rígidos, o que pode resultar em produção com compensações articulatórias inadequadas. A medida do parâmetro  $P_{70s}$  é menos crítica que a medida de  $F_2$  e a decisão tomada com o parâmetro  $P_{70s}$  (anterior x posterior) aumenta o grau de liberdade na produção da vogal.

O algoritmo para discriminação de fricativas é baseado em dois novos parâmetros, a medida da frequência de pico do espectro suavizado (FPES) e o coeficiente de espalhamento espectral (CEE). Parâmetros clássicos para detecção de fricativas como número de cruzamentos por zero, centro de gravidade espectral e frequência de pico foram

também avaliados mas sofrem influência de outros fatores como pico locais e presença ou ausência de sonoridade.

Para crianças com idades entre 7 e 10 anos, as medidas da frequência de pico do espectro suavizado sobre consoantes fricativas estão entre 4.500 e 8.850 Hz, sendo que 99,3% das medidas estão acima de 5.000 Hz. As fricativas palatais (mais graves) apresentam menores valores de FPES (média de 5.925 Hz) e as fricativas anteriores e médias apresentam valores médios de FPES mais elevados (6.957 Hz e 7.246 Hz, respectivamente). As medidas de FPES apresentaram valores médios de 6.739 Hz para as surdas e 6.788 Hz para as sonoras, variação inferior a 0,73%. As medidas de FPES das fricativas palatais são menores e com menor desvio padrão que as demais, apresentando-se bastante concentradas pouco abaixo de 6.000 Hz, permitindo assim realizar a sua classificação.

O coeficiente de espalhamento espectral (CEE) foi proposto para realizar a separação entre fricativas anteriores e médias. As anteriores apresentam o espectro de energia quase plano, enquanto as médias apresentam uma ressonância mais definida, resultando em espectro mais concentrado. As medidas do CEE são relativamente independentes da sonoridade. Os sinais de fones fricativos anteriores apresentam maiores valores do CEE que os sinais de fricativos médios.

Uma característica importante do algoritmo proposto para discriminação de consoantes fricativas é a possibilidade da discriminação das fricativas anteriores (/f/ e /v/). Os parâmetros especificados na literatura pesquisada foram propostos para discriminação entre fricativas médias e posteriores.

A FPES possibilitou separar as palatais das demais fricativas com taxa de erro de 14,3% e o CEE permitiu separar as labiodentais das alveolares com taxa de erro de 7,79%.

O sistema decorrente desse trabalho incorpora algoritmos mais precisos e que abrangem uma variedade maior de sons da língua.

O sistema possibilita uma ampla variedade de ajustes, incluindo-se definição de: valores máximos e mínimos de energia, valor alvo para energia, valores mínimos e máximos da frequência fundamental, valor alvo para frequência fundamental, valor da primeira formante de vogais, valor da frequência abaixo da qual está contida 70% da energia do espectro LPC, valor da frequência de pico do espectro suavizado, valor do coeficiente do espalhamento espectral, parâmetros este utilizados para controle das animações.

Uma janela fornece o valor dos parâmetros de voz utilizados no controle das animações e a medida do número de bits efetivos utilizados na aquisição do sinal. A medida do número de bits efetivo possibilita ao usuário ajustar o controle de ganho do sistema de aquisição, adequando-o ao microfone utilizado. As indicações com o valor dos parâmetros que controlam cada jogo permitem reconfigurá-lo, adequando-o ao estágio de controle articulatorio do indivíduo.

O desenvolvimento de aplicativos para aprimoramento do controle respiratório e da energia são complementados com jogos para o aprimoramento da energia para modulação da voz.

A importância do controle de vibração das pregas vocais ensejou o desdobramento dos aplicativos para aprimoramento dessa função em dois estágios. No estágio inicial, o sistema possibilita o controle sobre uma ampla faixa de valores da frequência fundamental, permitindo fixar limiares máximos e mínimos que possam ser adequados ao grau de controle proporcionado pelo indivíduo e que possa ser gradualmente modificado para uma faixa de frequências mais compatível com sua idade e sexo. No segundo estágio, o sistema

dispõe de jogos com alvo para consolidar a faixa de frequência fundamental obtida.

Os novos parâmetros propostos para discriminação de vogais possibilitaram o desenvolvimento de um aplicativo mais preciso e abrangente. A parametrização e concepção adotada possibilitam o exercício de aprimoramento do controle articulatório de vogais com maior grau de liberdade, exigindo menos compensações articatórias. Os aplicativos com vogais possibilitam definir quais serão as vogais alvo, possibilitando que o exercício possa ser gradualmente adequado às possibilidades articatórias da criança. Em especial, espera-se que as posições articatórias para produção das vogais, que venham ser adquiridas com os jogos, exijam o mínimo de compensações.

Os parâmetros propostos para discriminação de fricativas possibilitaram o desenvolvimento de jogos que contemplam todas as fricativas. Nesse caso, espera-se que as produções das fricativas possam vir a ser realizadas com mínimo consumo de ar.

Os aplicativos resultantes desse trabalho incorporam os exercícios que se mostraram conseqüentes na versão base. O sistema atual, *Jogos de Voz*, utiliza algoritmos mais precisos e específicos, proporcionando uma retroalimentação mais efetiva. Ele inclui exercícios com uma nova concepção para controle de energia e aprimoramento da frequência fundamental e, especialmente, introduz novos jogos para aprimoramento da produção de vogais e fricativas. Com o novo sistema, *Jogos de Voz*, tem-se a firme convicção que resultados bem mais promissores podem ser obtidos.

Os aplicativos que compõem os *Jogos de Voz* deverão futuramente ser avaliados pelos fonoaudiólogos para terapia de voz com deficientes auditivos. Os *Jogos de Voz* serão disponibilizados gratuitamente através da Internet.

Os requisitos para operação dos *Jogos de Voz* são um computador tipo IBM-PC com relógio igual a superior a 100 MHz, placa de som e microfone operando sob o sistema

operacional *Windows*. Não havendo restrições quanto a placa e/ou microfone a ser utilizado.

Estudos futuros para determinação dos valores médios dos parâmetros utilizados no jogos podem e devem ser realizadas para outras faixas etárias. Os próprios jogos garantem essa possibilidade através dos mostradores e das suas amplas possibilidades de configuração.

Estudos futuros também devem ser elaborados para o desenvolvimento de aplicativos com fonemas oclusivos, palavras isoladas e frases para auxílio a deficientes auditivos. Nesse caso, a utilização de redes neurais e HMM devem constituir a base para sua consecução.

Os parâmetros propostos neste trabalho exigem relativamente baixo esforço computacional e proporcionam boa discriminação sobre razoável parcela dos sons da fala. Os parâmetros propostos podem (e devem) ser avaliados, e quiçá utilizados, em outras aplicações. Em particular, espera-se que os parâmetros propostos possam corroborar com os parâmetros já consagrados, como os coeficientes MEL-CEPSTRAIS, em sistemas para reconhecimento de fala.

Espera-se, principalmente, que os jogos computadorizados para exercícios de aprimoramento da coordenação fonoarticulatória possam efetivamente proporcionar um ambiente lúdico, no qual as crianças e os adolescentes deficientes auditivos possam sentir prazer no exercício exploratório de suas potencialidades. Os jogos poderão, assim, auxiliar a realização da necessária prática que permitirá a coordenação articulatória e o aprimoramento do controle tátil-cinestésico, propiciando desta forma a destreza necessária à realização dos sons da fala.



# Bibliografia

- AGUIRRE, C., GOLDFELD, M. & FERREIRA, A . *Uma proposta bilingüe de atendimento à criança surda baseada no interacionismo. Anais do I Simpósio de Fonoaudiologia*, Campinas, **1999**.
- ALBANO, E. C. *Da fala à linguagem: tocando de ouvido*. São Paulo: Martins Fontes. **1990**.
- ALCAIM, A., SOLEWICZ, J. A. & MORAES, J. A. *Frequência de ocorrência dos fonemas e listas de frases foneticamente balanceadas no português falado no Rio de Janeiro. Revista da Sociedade Brasileira de Telecomunicações*, 7(1):23-42, Dez. **1992**.
- ALLEN, J., HUNNICUTT, S. & KLATT, D. *From text to speech: the MITalk system*. Cambridge: Cambridge University Press, **1987**.
- ALMEIDA, K. & IORIO, M.C.M. *Próteses auditivas – fundamentos teóricos & aplicações clínicas*. São Paulo: Lovise, **1996**.
- ARAÚJO, A.M.L. *Jogos computacionais fonoarticulatórios. Revista Espaço*, nº 12: 86-90, Dez. **1999**.
- ARAÚJO, A.M.L., GOTO, C.M. & VIOLARO, F. *Jogos computadorizados para auxílio ao aprendizado da fala em crianças com deficiência auditiva. Actas del Iberdiscap 2000, 3º Congreso de Comunicación Alternativa y Aumentativa y 1º de Tecnologías de Apoyo Para la Discapacidad*. 37-40, Madrid-ESP, Out. **2000**.

- ARAÚJO, A. M. L. & LIMA, M. C. P. M. *Trabalhando a voz de deficientes auditivos com auxílio de computador*. In: QUEVEDO, A. A. F., OLIVEIRA, J. R. & MANTOAN, M. T. E. (org.) ***Mobilidade e comunicação: Desafios à tecnologia e à inclusão social***. Campinas: UNICAMP, **1999**.
- ARAÚJO, A. M. L. & VIOLARO, F. *Formant frequency estimation using a MEL scale LPC algorithm*. ***Proceedings of ITS/IEEE International Telecommunications Symposium***, vol. I, 207-212, São Paulo- SP, **1998**.
- ARAÚJO, A. M. L., VIOLARO, F. & LIMA, M. C. P. M. *Inteligibilidade do português falado no Brasil limitado às frequências abaixo de 1.000 Hz*. ***Revista Pró-Fono***, 11(2):15-21, Set. **1999**.
- ARAÚJO, A. M. L. & VIOLARO, F. *Análise e parametrização de fricativas*. Aceito para apresentação e publicação nos **anais do XVIII Simpósio Brasileiro de Telecomunicações**, Gramado –RS, **2000**.
- BEHLAU, M. *Uma análise das vogais do português falado em São Paulo: Perceptual, espectrográfica, de formantes e computadorizada de frequência fundamental*. Tese de Mestrado – Escola Paulista de Medicina. São Paulo, **1984**.
- BEHLAU, M. & PONTES, P. *Avaliação e tratamento das disfonias*. São Paulo: Lovise, **1995**.
- BEHLAU, M. & PONTES, P. *Higiene vocal – cuidando da voz*. 2ª ed., Rio de Janeiro: Revinter, **1999**.
- BEVILACQUA, M. C. & FORMIGONI, G. M. P. *Audiologia Educacional: uma opção terapêutica para a criança deficiente auditiva*. Carapicuíba: Pró-Fono, **1997**.
- BESS, F. H. & HUMES, L. E. *Fundamentos de Audiologia*. 2ª ed., Porto Alegre: Artmed, **1998**.

- BLOCH, P. *Problemas da voz e da fala*. Rio de Janeiro: Letras e Artes, **1963**.
- BOMMARITO, S. *O efeito de um método de terapia de voz na qualidade vocal e na inteligibilidade da fala de indivíduos surdos*. Tese de Doutorado – Universidade Federal de São Paulo – Escola Paulista de Medicina. São Paulo, **2000**.
- BOONE, D. & PLAUTE, E. *Comunicação humana e seus distúrbios*. Porto Alegre: Artes Médicas, **1994**.
- BOOTHROYD, A. *Hearing impairments in children*. New York: Prentice Hall, **1982**.
- BORGES, R. H. M. *Avaliação de voz em deficientes auditivos: uma análise crítica*. In: FERREIRA, L. P. (org.) *Dissertando sobre voz*. Carapicuíba: Pró-Fono, **1998**.
- BRASIL, Ministério da Educação e do Desporto, Secretaria de Educação Especial, *Subsídios para organização e funcionamento de serviço de educação especial: área de deficiência auditiva*. Brasília: MEC, Série Diretrizes, nº 6, **1995**.
- CALVERT, D. R. & SILVERMAN, S. R. *Speech and deafness*. 2<sup>nd</sup> ed., Washington: The Alexander Graham Bell Association for the Deaf, **1978**.
- CÂMARA JR., J. M. *História e estrutura da Língua Portuguesa*. Rio de Janeiro: Padrão, **1970**.
- CÂMARA JR., J. M. *Estrutura da Língua Portuguesa*. 20<sup>a</sup> ed., Petrópolis: Vozes, **1991**.
- CASTRO, D. B., ANJOS, J. C. A., ARAÚJO, A. M. L. & KLAUTAU, A. *Sistema áudio gráfico visual computadorizado para auxílio no treinamento de deficientes auditivos*. *Anais XVIII Congresso Nacional de Matemática Aplicada e Computacional*, vol. II, 725-729, Curitiba – PR, **1995**.
- CHEN, Y. Y. & HUANG, D. *The application of multimedia technique to speech training for deaf children*. *Proceedings of 2<sup>o</sup> World Voice Congress and 5<sup>o</sup> International Symposium on Phonosurgery*. São Paulo, Fev. **1999**.

- CHOMSKY, N. *Linguagem e pensamento*. 2<sup>a</sup> ed. Petrópolis: Vozes, **1971**.
- CICCONI, M. *Comunicação total – introdução – estratégias – a pessoa surda*. 2<sup>a</sup> ed., Rio de Janeiro: Cultura Médica, **1996**.
- COLTON, H. R. & CASPER, J. K. *Compreendendo os problemas de voz*. Porto Alegre: Artes Médicas, **1996**.
- COUTO-LENZI, A. *Como posso falar*. Rio de Janeiro: AIPEDA, **1991**.
- COUTO-LENZI, A. *Percepção da fala – teste*. Rio de Janeiro: AIPEDA, **1997**.
- COUTO-LENZI, A. *Surdos severos e profundos: acessos à zona dos sons da fala*. *Revista Espaço*, 11:24-28, Jun. **1999(a)**.
- COUTO-LENZI, A. *Os mitos da surdez profunda*. *Revista Integração*, 21:41-48, Jun. **1999(b)**.
- COSTA, F. *Um ambiente para o treinamento da fala*, Relatório de Graduação, Ciências da Computação, UFRN, **1993**.
- DELATTRE, P., LIBERMAN, A.M. & COOPER, F. S., *Acoustic loci and transitional cues for consonants*. *Journal of the Acoustical Society of America*, 27:769-774, **1955**.
- ENCICLOPÉDIA BARSA. Rio de Janeiro: Encyclopaedia Britannica do Brasil Publicações, **1993**.
- ERBER, N. P. *Speech perception by hearing-impaired children*. In: BESS, F. H., FREEMAN, B. A. & SINCLAIR, J. S. E. (org.) *Amplification in education*. Washington: The Alexander Graham Bell Association for the Deaf, **1981**.
- FANT, G. *Acoustic theory of speech production*. The Hague: Mouton, **1960**.
- FIGUEIREDO, R. M. *A eficácia de medidas extraídas do espectro de longo termo para a identificação de falantes*. *Caderno de Estudos Lingüísticos*, (25):129-160, Jul./Dez **1993**.

- FRANCO, E. *Surdos de uma escola especial: caracterização da voz e seus fatores determinantes*. In: FERREIRA, L. P. (org.) *Dissertando sobre voz*. Carapicuíba: Pró-Fono, **1998**.
- FRUCHTER, D. *Perceptual Significance of locus equations*. *Journal of the Acoustical Society of America*, (95):2977-2980, **1994**.
- GOLD, B. & RABINER, L.R. *Parallel processing techniques for estimating pitch periods of speech in the time domain*. *Journal of the Acoustical Society of America*, 34(7):916-921, **1969**.
- GOLDFELD, M. *A criança surda – linguagem e cognição numa perspectiva sócio-interacionista*. São Paulo: Plexus, **1997**.
- HAYKIN, S. *Adaptive filter theory*. 3<sup>a</sup> ed., New Jersey: Prentice-Hall, **1996**.
- HILLENBRAND, J., GETTY, L. A., CLARK, M. J. & WHEELER, K., *Acoustic characteristics of american english vowels*. *Journal of the Acoustical Society of America*, 97(5):3099-3111, May **1995**.
- HOLLE, B. *Desenvolvimento motor na criança normal e retardada*. São Paulo: Manole, **1990**.
- ISSLER, S. *Articulação e Linguagem*. 3<sup>a</sup> ed., São Paulo: Lovise, **1996**.
- JCIH – JOINT COMMITTEE ON INFANT HEARING, *1994 Position Statement*. *Pediatrics*. 95(1):152-156, Jan. **1995**.
- KENT, R. D. & READ, C. *The acoustic analysis of speech*. San Diego: Singular Publishing, **1992**.
- KOSLOWSKI, L. *A proposta bilíngüe de educação do surdo*. *Revista Espaço*, 10:47-53, Dez. **1998**.

- LAUNAY, C. & BOREL-MAISONNY, S. *Distúrbios da linguagem da fala e da voz na infância*. São Paulo: Roca, **1989**.
- LIMA, M. C. M. P., ARAÚJO, A. M. L. & ROMANO, A. M. C. *Utilization of a computational program to assist voice production of a deaf adolescent. Proceedings of 2<sup>o</sup> word voice congress and 5<sup>o</sup> international symposium on phonosurgery*. São Paulo, Fev. **1999**.
- LING, D. *Amplification to speech*. In CALVERT, D. R. & SILVERMAN, S. R. *Speech and deafness*. 2<sup>a</sup> ed. Washington: The Alexander Graham Bell Association for the Deaf, **1978**.
- LÖFQVIST, A. *The long-time average spectrum as a tool in voice research*. *J. Phon.* (14):471-476, **1986**.
- MAIA, E. M. *No reino da fala: a linguagem e seus sons*. São Paulo: Ática, **1991**.
- MALMBERG, B. *A fonética*. Lisboa: Livros do Brasil, **1954**.
- MARCHESI, A. *El desarrollo cognitivo y linguístico de los niños sordos: perspectivas educacionais*. Madrid: Alianza Editorial, **1987**.
- MARKEL, J. D. *Digital inverse filtering – a new tool for formant trajectory estimation*. *IEEE Transactions on Audio and Electroacustics*, AU-20(2):129-137, Jun. **1972**.
- MARKEL, J. D. & GRAY, A. H. *Linear prediction of speech*. New York: Springer-Verlag, **1976**.
- MENN, L., STOEL-GAMMON, C. *Desenvolvimento fonológico*. In: FLETCHER, P. & MACWHINNEY, B. (org.) *Compêndio da linguagem da criança*. Porto Alegre: Artes Médicas, **1997**.

- MORAES, E. *Reconhecimento automático de fala contínua empregando modelos híbridos ANN+HMM*. Tese de Mestrado – Faculdade de Engenharia Elétrica e de Computação – Universidade Estadual de Campinas. Campinas, **1997**.
- MYSAK, E. D. *Patologia dos sistemas da fala*. 2ª ed. Rio de Janeiro: Atheneu, **1988**.
- NEPOMUCENO, L. X. *Acústica*. São Paulo: Edgard Blücher, **1977**.
- NITTROUER, S., STUDDERT-KENNEDY, M. & NEELY, S. T. *How children learn to organize their speech gestures: further evidence from fricative-vowel syllables*. ***Journal of the Speech and Hearing Research***, 39:379-389, Apr. **1996**.
- NORTHERN, J. & DOWNS, M. P. *Audição em crianças*. 3ª ed. São Paulo: Manole, **1989**.
- NOLL, A. M. *Cepstrum pitch determination*. ***Journal of the Acoustical Society of America***, (41):293-309, Feb. **1967**.
- PERDONCINI, G. & COUTO-LENZI, A. *Audição – é o futuro da criança surda*. Rio de Janeiro: AIPEDA, **1996**.
- PERLIN, G. T. T. *Identidades surdas*. In: SLKIAR, C. (org.) *A Surdez, um olhar sobre as diferenças*. Porto Alegre: Mediação, **1998**.
- PETERSON, G. E. & BARNEY, H. L. *Control methods used in a study of the vowels*. ***Journal of the Acoustical Society of America***, 24(2):175-185, Mar., **1952**.
- PICONE, J. W. *Signal modeling techniques in speech recognition*. ***Proceedings of the IEEE***, 81(9):1215-1247, Sep. **1993**.
- PIMENTA, E. M., RAMOS, M. I. B. & SOARES, R. C. A. *A fonoaudiologia numa proposta bilingüe*. ***Revista Espaço***, 10:74-75, Dez. **1998**.
- RABINER, L. R. & SCHAFER, R. W. *Digital processing of speech signals*. New Jersey: Prentice-Hall, **1978**.

- REDONDO, M. C. & LOPES FILHO, O. C. *Testes básicos de avaliação auditiva*. In: LOPES FILHO, O. C. (editor) *Tratado de fonoaudiologia*. São Paulo: Roca, **1997**.
- RINALDI, G. *Deficiência auditiva*. In: *Série Atualidades Pedagógicas*, nº 4, vol. I e II. Brasília: MEC/SEEP, **1997**.
- ROBB, M. P. & SAXMAM, J. H. *Developmental trends in vocal fundamental frequency of young children*. *Journal of Speech and Hearing Research*, (28):421-427, Sept. **1985**.
- RUSSO, I. C. P. & BEHLAU, M. *Percepção da fala: análise acústica do português brasileiro*. São Paulo: Lovise, **1993**.
- RUSSO, I. C. P. & SANTOS, T. M. M. *Audiologia infantil*. São Paulo: Cortez, **1994**.
- SACKS, O. *Vendo vozes*. São Paulo: Companhia das Letras, **1998**.
- SADOSKY, M. *Cálculo numérico e gráfico*. Rio de Janeiro: Interciência, **1980**.
- SÁNCHEZ, T. C. *Defectos en la diction infantil: procedimientos para su correccion*. Buenos Aires: Kapelusz, **1973**.
- SANTOS, M. T. *Uma análise espectrográfica dos sons fricativos surdos e sonoros do português brasileiro*. Monografia – Especialização – Escola Paulista de Medicina. São Paulo, **1987**.
- SHEETS, B. V. *The development of speech*. In: CONNOR, L. E. (ed.) *Speech for the deaf child*. Washington: The Alexander G. B. Association for the Deaf, **1971**.
- SONDHI, M. M. *New methods of pitch extraction*. *IEEE Transactions on Audio and Electroacoustics*, AU-16(2):262-266, Jun. **1968**.
- SILVA, A. A. *A surdez e a pessoa surda: revisão e tópicos básicos*. In: CICCONE, M. *Comunicação Total*. 2ª ed. Rio de Janeiro: Cultura Médica, **1996**.

- SKINNER, B. F. *O comportamento verbal*. São Paulo: Universidade de São Paulo, **1978**.
- SUKKAR, R. S., LOCICERO, J. L. & PICONE, J. *Design. and implementation of a parallel processing based pitch detector. IEEE Journal of Selected Areas on Communications*, 6(2):441-451, Feb. **1988**.
- SUSSMAN, H. M., MINIFIE, F. D., BUDER, E. H., STOEL-GAMMON, C. & SMITH, J., *Consonant-Vowel Interdependencies in Babbling and Early Words: Preliminary Examination of a Locus Equation Approach. Journal of the Speech and Hearing Research*, (39):424-433, Apr. **1996**.
- VIEIRA, M. N. *Módulo frontal para um sistema de reconhecimento automático de voz*. Tese de Mestrado – Faculdade de Engenharia Elétrica e de Computação – Universidade Estadual de Campinas. Campinas, **1989**.
- VIGOTSKY, L. S. *Pensamento e linguagem*. 3<sup>a</sup> ed. São Paulo: Martins Fontes, **1993**.
- VIGOTSKY, L. S., LURIA, A. R. & LEONTIEV, A. N. *Linguagem, desenvolvimento e aprendizagem*. São Paulo: Icone, **1988**.
- WOODWORTH, R. S. & SCHLOSBERG, H. *Psicologia Experimental – Tomo I*, 3<sup>a</sup> ed. Buenos Aires: Editoria Universitária de Buenos Aires, **1964**.
- ZEMLIN, W. R. *Princípios de anatomia e fisiologia em fonoaudiologia*. 4<sup>a</sup> ed., Porto Alegre: Artmed, **2000**.



## Apêndice A

### Procedimento para estimar energia e número de bits efetivos

O processo de estimação de parâmetros para alimentação dos jogos controlados pela energia e para os mostradores do número de bits efetivos é realizado sobre quadros de  $N$  amostras da seqüência  $x(n)$ , amostrada a  $f_a = 44.100$  Hz, para  $n=1\dots N$ , é descrito a seguir:

1. Determinar o valor médio do quadro,  $\bar{X} = \frac{1}{N} \sum_{n=1}^N x(n)$ , onde  $N = 2048$ .

2. Eliminar a componente DC do quadro,  $x_1(n) = x(n) - \bar{X}$ ,  $n = 1 \dots N$ .

3. Obter o valor quadrático médio do k-ésimo quadro em decibéis, de acordo com

$$E(k) = 10 \cdot \log_{10} \left( \frac{1}{N} \sum_{n=1}^N x_1^2(n) \right).$$

4. Filtrar o parâmetro energia do quadro em decibéis através de um filtro passa-baixas de 1ª ordem,  $E_f(k) = \alpha \cdot E_f(k-1) + (1-\alpha) \cdot E(k)$ , onde o valor da constante  $\alpha$  foi fixado em 0,5 e o valor inicial da energia ( $E_f(0)$ ) foi fixado em  $-80$ .

5. Obter o número efetivo de bits,  $N_{be} = 1 + \text{int}(\log_2(\max(\text{abs}(x(n)))) + 0,5)$ ,  $n = 1..N$



## Apêndice B

### Procedimento para estimar $F_0$

A estimativa da frequência fundamental é realizada sobre quadros contíguos do sinal contendo 2.048 amostras, adquiridos com um conversor A/D de 16 bits, utilizando uma frequência de amostragem de 44.100 Hz e processamento sem superposição entre quadros.

O processo de estimação da frequência fundamental pode ser sistematizado para cada quadro de acordo com os seguintes passos:

1. Obter uma seqüência correspondente a um quadro com  $N$  amostras do sinal,  $x(n)$ ,  $n = 1 \cdots N$ , onde  $N = 2048$  amostras e  $f_a = 44.100$  Hz.
2. Filtrar a seqüência  $x(n)$  através de um filtro IIR passa-altas Chebyshev tipo II de 4ª ordem, com frequência de corte (-3 dB) de 65 Hz, obtendo a seqüência  $x_1(n)$ .
3. Filtrar a seqüência  $x_1(n)$  através de um filtro IIR passa-baixas Chebyshev tipo II de 6ª ordem, com frequência de corte (-3 dB) de 800 Hz, obtendo a seqüência  $x_2(n)$ .
4. Obter de  $x_2(n)$  as seqüências  $x_{21}(n)$  e  $x_{22}(n)$ , onde  $x_{21}(n) = x_2(n)$ ,  $n = 1 \cdots 1024$  e  $x_{22}(n) = x_2(n + 1024)$ ,  $n = 1 \cdots 1024$ . Determinar o valor máximo absoluto das

seqüências  $x_{21}(n)$  e  $x_{22}(n)$ ,  $X_{21\max}$  e  $X_{22\max}$  respectivamente. Tomar o menor dos

$$\text{máximos, } X_{2\max} = \min \left[ X_{21\max} \quad X_{22\max} \right].$$

5. Calcular o limiar de truncamento,  $Cl = 0,6 \cdot X_{2\max}$ .

6. Obter a polaridade da amostra de  $x_2(n)$  associada a  $X_{2\max}$ ,

$$p_{ol} = \begin{cases} +1, & \text{se } x_2(n) \text{ associado a } X_{2\max} \text{ é positivo} \\ -1, & \text{caso contrário} \end{cases}$$

7. Inverter a polaridade de toda a seqüência se  $p_{ol}$  é negativa,

$$x_3(n) = \begin{cases} +x_2(n), & \text{se } p_{ol} = +1 \\ -x_2(n), & \text{se } p_{ol} = -1 \end{cases}.$$

8. Reduzir a freqüência de amostragem da seqüência  $x_3(n)$ , obtendo a seqüência

$$x_4(n) = x_3(4 \cdot n), \quad n = 1 \cdots N_2 \text{ onde } N_2 = 512 \text{ e } f_a = 11.025 \text{ Hz.}$$

9. Tornar nulas todas as mostras de  $x_4(n)$  menores que  $Cl$ ,

$$x_5(n) = \begin{cases} x_4(n), & \text{se } x_4(n) \geq Cl \\ 0, & \text{caso contrário} \end{cases}$$

10. Determinar a posição das amostras associadas às freqüências fixadas como limites do

algoritmo. No caso os valores mínimo e máximo para a freqüência fundamental foram

fixados em  $F_{o\min} = 80 \text{ Hz}$  e  $F_{o\max} = 800 \text{ Hz}$ .

$$K = \frac{T}{\Delta T} = \frac{1/F}{1/f_a} = \frac{f_a}{F}, \text{ resultando } \begin{cases} K_{o\min} = \frac{11.025}{800} = 13,78 \approx 13 \\ K_{o\max} = \frac{11.025}{80} = 137,8 \approx 138 \end{cases}$$

11. Calcular a função de autocorrelação,

$$r(k) = \frac{1}{N-k} \sum_{m=1}^{N-k} x_4(m) \cdot x_4(m+k), \quad k = 0 \dots K_{0max}$$

12. Normalizar a função de autocorrelação,  $r_n(k) = \frac{r(k)}{r(0)}$ ,  $k = 0 \dots K_{0max}$ .

13. Determinar o valor de  $k$  associado ao máximo de  $r_n(k)$  entre  $K_{0min}$  e  $K_{0max}$ ,

$$K_{max} = k, \text{ tal que } r_n(K_{max}) = \max(r_n(k), k = K_{0min} \dots K_{0max}).$$

14. Se  $r_n(K_{max}) > 0,5$  o quadro é considerado como sonoro e a frequência fundamental é obtida pela razão entre a frequência de amostragem da sequência  $x_4(n)$  e  $K_{max}$ ,

$$F_o = \frac{f_a}{K_{max}}. \text{ No presente caso } F_o = \frac{11.025}{K_{max}}.$$

Devido à quantização temporal imposta pela frequência de amostragem  $f_a = 11.025$  Hz, a precisão da estimação da frequência fundamental fica limitada a 0,36% para  $F_o = 80$  Hz e 3,63% para  $F_o = 800$  Hz



## Apêndice C

### Procedimentos para estimar $F_1$ e $P_{70s}$

O algoritmo para determinação de  $F_1$  e  $P_{70s}$  utiliza quadros de sinal com 2.048 amostras, adquiridos com 16 bits por amostra e frequência de amostragem de 44.100 Hz. O processamento é realizado sem superposição de quadros.

O processo de estimação de  $F_1$  e  $P_{70s}$  está sistematizado para cada quadro de acordo com os seguintes passos:

1. Obter uma seqüência correspondente a um quadro com  $N$  amostras do sinal,

$$x(n), \quad n = 1 \cdots N, \text{ onde } N = 2048 \text{ amostras e } f_a = 44.100 \text{ Hz.}$$

2. Filtrar a seqüência  $x(n)$  através de um filtro IIR passa-altas Chebyshev tipo II de 4ª ordem, com frequência de corte (-3 dB) de 65 Hz, obtendo a seqüência  $x_1(n)$ .

3. Filtrar a seqüência  $x_1(n)$  através de um filtro IIR passa-baixas Chebyshev tipo II de 6ª ordem, com frequência de corte (-3 dB) de 5.000 Hz, obtendo a seqüência  $x_2(n)$ .

4. Reduzir a frequência de amostragem da seqüência  $x_2(n)$ , obtendo a seqüência

$$x_3(n) = x_2(4 \cdot n), \quad n = 1 \cdots N_2, \text{ onde } N_2 = 512 \text{ e } f_a = 11.025 \text{ Hz.}$$

5. Efetuar uma pré-ênfase sobre  $x_3(n)$ , obtendo a seqüência

$$x_4(n) = x_3(n) - 0,95x_3(n-1), \quad n = 2 \cdots N_2.$$

6. Ponderar a seqüência  $x_4(n)$  com uma janela de Hamming, obtendo a seqüência

$$x_5(n) = x_4(n) \cdot h(n), \quad h(n) = 0,54 - 0,46 \cos\left(\frac{2\pi(n-1)}{(N-1)}\right), \quad n = 1 \cdots N, \quad N = N_2 = 512.$$

7. Calcular a função de autocorrelação,

$$r(k) = \sum_{m=1}^{N-k} x_5(m) \cdot x_5(m+k), \quad k = 0 \cdots K_o$$

onde  $K_o$  é a ordem do modelo LPC utilizado.

8. Obter o modelo LPC pelo método da autocorrelação, através do algoritmo de Levinson-Durbin (HAYKIN, 1996). Da análise LPC resulta o filtro de ordem  $K_o$  que simula o

trato vocal,  $H(z) = \frac{1}{A(z)}$ , onde  $A(z) = 1 - \sum_{i=1}^{K_o} a_i z^{-i}$ . Variando a ordem do modelo entre

$K_o = 9$  e  $K_o = 13$ , obteve-se medidas similares de  $F_1$ . A ordem adotada no trabalho foi  $K_o = 11$ .

9. Calcular a FFT de  $a(n) = \{1 - a_1 - a_2 \cdots - a_{k_o} 0 \cdots 0\}$  onde  $a(n)$  corresponde à transformada inversa de  $A(z)$ , de comprimento  $K_o + 1$ , completada com  $M - (K_o + 1)$  zeros, onde  $M$  é o número de pontos da FFT. O valor do espectro LPC nas freqüências

$\omega_k = k \cdot \frac{2\pi}{M}$ ,  $k = 0 \cdots (M - 1)$  é dado por

$$|H(k)| = \frac{1}{\left[ (\text{Re}(A(k)))^2 + (\text{Im}(A(k)))^2 \right]^{1/2}}$$

onde  $\text{Re}(\cdot)$  e  $\text{Im}(\cdot)$  são a parte real e imaginária. O valor de  $M$  foi fixado em 512.

10. A primeira formante ( $F_1$ ) é determinada tomando-se a freqüência correspondente ao

primeiro máximo do espectro LPC,  $F_1 = k_{\max} \cdot \frac{fa}{M}$  onde  $k_{\max}$  é tal que

$$|H(k_{\max})| = \max\{|H(k)|, k = 0 \dots (M-1)\}.$$

11. Computar  $k_{\beta}$  para  $\beta = 70$  através da equação 4.5,

$$\frac{1}{M} \left[ |H(0)|^2 + 2 \cdot \sum_{k=1}^{k_{\beta}} |H(k)|^2 \right] = \left( \frac{\beta}{100} \right) \cdot \frac{1}{M} \left[ |H(0)|^2 + \left| H\left(\frac{M}{2}\right) \right|^2 + 2 \cdot \sum_{k=1}^{\frac{M}{2}-1} |H(k)|^2 \right]$$

12. Calcular o parâmetro  $P_{70s}$ , com a equação 4.4,  $P_{\beta} = k_{\beta} \frac{f_a}{M}$ , onde

$$P_{70s} = k_{70} \frac{11.025}{512}$$

O cálculo do parâmetro  $P_{70}$  sobre o espectro LPC proporcionou resultados mais precisos que os obtidos diretamente sobre o espectro do quadro do sinal (via FFT).



## Apêndice D

### Procedimentos para estimar FPES e CEE

A discriminação das fricativas é realizada através de dois parâmetros, a frequência de pico do espectro suavizado (FPES) e o coeficiente de espalhamento espectral (CEE). Estes parâmetros são obtidos de quadros de sinal com 2.048 amostras, adquiridos com 16 bits por amostra e frequência de amostragem ( $f_a$ ) de 44.100 Hz. O processamento é realizado sem superposição de quadros.

O processo de estimação da FPES e do CEE está sistematizado para cada quadro de acordo com os seguintes passos:

1. Obter uma seqüência correspondente a um quadro com  $N$  amostras do sinal,

$$x(n), \quad n = 1 \cdots N, \quad \text{onde } N = 2048 \text{ amostras e } f_a = 44.100 \text{ Hz.}$$

2. Filtrar a seqüência  $x(n)$  através de um filtro IIR passa-altas Chebyshev tipo II de 4ª ordem, com frequência de corte (-3 dB) de 65 Hz, obtendo a seqüência  $x_1(n)$ .

3. Ponderar a seqüência  $x_1(n)$  com uma janela de Hamming, obtendo a seqüência

$$x_2(n) = x_1(n) \cdot h(n), \quad h(n) = 0,54 - 0,46 \cos\left(\frac{2\pi(n-1)}{(N-1)}\right), \quad n = 1 \cdots N, \quad N = 2048.$$

4. Calcular a função de autocorrelação,

$$r(k) = \sum_{m=1}^{N-k} x_2(m) \cdot x_2(m+k), \quad k = 0, \dots, 3$$

5. Obter o modelo LPC de ordem 3, através do método da autocorrelação, empregando o algoritmo de Levinson-Durbin (HAYKIN, 1996).
6. Calcular as raízes do polinômio resultante da modelagem LPC, usando a fórmula de Cardano-Tartaglia (SADOSKY, 1980).
7. Determinar a FPES obtendo a frequência associada às raízes complexas obtidas no passo anterior.
8. Calcular o espectro de frequências,

$$X(k) = DFT(x_1(n)), \quad k = 0 \dots 2047$$

9. Zerar as componentes correspondentes às frequências de 0 a 2000 Hz obtendo  $X_h(k)$ ,  $k = 0 \dots 2047$ .
10. Determinar o valores de  $P_{20}$  e  $P_{80}$  a partir de  $X_h(k)$
11. Calcular CEE com a equação 4.7.

## Apêndice E

### Relação das publicações

#### Capítulo de Livro

<b>Título</b>	Trabalhando a voz de deficientes auditivos com auxílio de computador.
<b>Livro</b>	Mobilidade e comunicação: Desafios à tecnologia e à inclusão social.
<b>Edição</b>	UNICAMP, Campinas 1999.
<b>Autores</b>	ARAÚJO, A. M. L. & LIMA, M. C. M. P.
<b>Org.</b>	QUEVEDO, A. A. F., OLIVEIRA, J. R. & MANTOAN, M. T. E.

#### Artigos em revista

<b>Artigo</b>	Jogos Computacionais fonoarticulatórios.
<b>Periódico</b>	Revista Espaço, (12): 86-90, Dezembro 1999.
<b>Autor</b>	ARAÚJO, A. M. L.

**Artigo** Inteligibilidade do português falado no Brasil limitado às frequências abaixo de 1.000 Hz.

**Periódico** Revista Pró-Fono, 11(2):15-21, Setembro 1999.

**Autores** ARAÚJO, A. M. L., VIOLARO, F. & LIMA, M. C. M. P.

#### Artigos em anais de congressos

**Artigo** Formant frequency estimation using a MEL scale LPC algorithm.

**Congresso** ITS/IEEE International Telecommunications Symposium.

**Local e data** São Paulo - SP, Setembro 1998.

**Autores** ARAÚJO, A. M. L. & VIOLARO, F.

**Artigo** Utilization of a computational program to assist voice production of a deaf adolescent.

**Congresso** 2º word voice congress and 5º international symposium on phonosurgery.

**Local e data** São Paulo, Fevereiro 1999.

**Autores** LIMA, M. C. M. P., ARAÚJO, A. M. L. & ROMANO, A. M. C.

- Artigo** Análise e parametrização de fricativas.
- Congresso** XVIII Simpósio Brasileiro de Telecomunicações.  
Gramado – RS, Setembro 2000.
- Autores** ARAÚJO, A. M. L. & VIOLARO, F.
- Artigo** Jogos computadorizados para auxílio ao aprendizado da fala em crianças com deficiência auditiva.
- Congresso** Iberdiscap 2000, Congreso Iberoamericano, 3º de comunicación alternativa y aumentativa y 1º de tecnologías de apoyo para la discapacidad.
- Local e data** Madri - Espanha, Out. 2000.
- Autores** ARAÚJO, A. M. L., GOTO, C. M. & VIOLARO, F.

