

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

**Desenvolvimento de Regras de Pronúncia
para a Síntese de Fala em Língua Portuguesa**

por

PAULO EDUARDO OSTERMANN FILHO

Dissertação submetida à avaliação,
como requisito parcial para a obtenção
do grau de Mestre em Ciência da Computação

Prof. Dante Augusto Couto Barone
Orientador

Porto Alegre, dezembro de 2002

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Ostermann Filho, Paulo Eduardo

Desenvolvimento de regras de pronúncia para a síntese de fala em língua portuguesa/por Paulo Eduardo Ostermann Filho. – Porto Alegre: PPGC da UFRGS, 2002.

70f.: il

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR – RS, 2002. Orientador: Barone, Dante Augusto Couto.

1. Síntese. 2. Lingüística Computacional. I. Barone, Dante Augusto Couto. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitora Prof^a Wrana Panizzi

Pró-Reitor de Ensino: Prof. José Carlos Ferraz Hennemann

Pró-Reitor Adjunto de Pós-Graduação: Prof. Jaime Evaldo Fensterseifer

Diretor do Instituto de Informática: Prof. Philippe Olivier Alexandre Navaux

Coordenador do PPGC: Prof. Carlos Alberto Heuser

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

Agradecimentos

Agradeço, primeiramente, a Deus pela oportunidade que ele me concedeu. Um agradecimento especial a minha família, principalmente, aos meus pais, minha irmã e namorada pela força e apoio. Depois, ao Prof. Dante pela sua orientação e compreensão. Por conseguinte, aos meus colegas do Projeto Spoltech pela colaboração no desenvolvimento deste trabalho. Não poderia deixar de agradecer também a minha prima pela sua ajuda na elaboração do Abstract, que agora está tinindo.

Por fim, também quero agradecer ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio financeiro e ao Instituto de Informática da Universidade Federal do Rio Grande do Sul pelo aporte tecnológico.

Sumário

<u>Lista de Abreviaturas</u>	6
<u>Resumo</u>	10
<u>Abstract</u>	11
<u>1 Introdução</u>	12
<u>2 Produção da Fala</u>	14
<u>2.1 O processo de leitura</u>	15
<u>2.2 Visão</u>	15
<u>2.3 Pensamento</u>	17
<u>2.4 Fala</u>	19
<u>2.5 Do pensamento para a fala</u>	21
<u>2.6 Audição</u>	24
<u>2.7 Fonêmica</u>	25
<u>2.7.1 Diferenças entre fonema e som [OLI 96]</u>	26
<u>2.7.2 Como se produz a voz</u>	26
<u>2.7.3 Classificação dos fonemas</u>	27
<u>2.7.4 As semivogais</u>	27
<u>2.7.5 Classificação das vogais</u>	27
<u>2.7.6 Classificação das consoantes</u>	28
<u>2.7.7 Correspondência entre fonema e letra [OLI 96]</u>	29
<u>2.7.8 Fonemas e alofones</u>	29
<u>2.7.9 As premissas da fonêmica</u>	32
<u>3 Síntese da fala</u>	36
<u>3.1 Introdução</u>	36
<u>3.2 Métodos de síntese</u>	37
<u>3.2.1 Síntese Articulatoria</u>	37
<u>3.2.2 Síntese de formantes</u>	38
<u>3.2.3 Síntese Código preditivo linear(LPC)</u>	39
<u>3.2.4 Concatenação da forma de onda</u>	40
<u>3.2.5 Síntese TPF</u>	40
<u>4 Processamento da Fala</u>	57
<u>4.1 As ferramentas utilizadas</u>	57
<u>4.1.1 CSLU Toolkit</u>	57
<u>4.2 O processo de reconhecimento</u>	65
<u>4.2.1 Treinamento dos reconhecedores de fala</u>	68
<u>4.2.2 Reconhecedores desenvolvidos</u>	70
<u>4.3 O processo de síntese da fala</u>	72
<u>5 Implementação</u>	75
<u>5.1 O problema</u>	75
<u>5.2 A solução</u>	75
<u>6 Conclusão</u>	91

<u>Anexo 1 Exemplo do dicionário eletrônico</u>	93
<u>Anexo 2 Passos básicos para criar uma definição léxica no Festival</u>	102
<u>Anexo 3 Forma básica de uma regra para o Festival</u>	103
<u>Anexo 4 Exemplo de construção de um conjunto de fones para o Festival</u>	104
<u>Anexo 5 Exemplo de regras letra para som</u>	106
<u>Anexo 6 Exemplo de uma Cart tree</u>	108
<u>Anexo 7 Resultado da análise estatística para a vogal “e”</u>	109
<u>Anexo 8 Resultado da análise estatística para a vogal “o”</u>	118
<u>Referências</u>	127
<u>Obras Consultadas</u>	129

Lista de Abreviaturas

CAA	Contraste em Ambiente Análogo
CAI	Contraste em Ambiente Idêntico
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CV	Consoante-Vogal
CVC	Consoante-Vogal-Consoante
DFT	Discrete Fourier Transform
F0	Frequência Fundamental
FSs	Structure Features
GP	Gerador de Prosódia
IPA	International Phonetic Association
LPC	Linear Predictive Coding
LTS	Letra para som
MFCC	Mel Frequency Cepstral Coefficients
MLDS	Multi-Level Data Structure
MLP	Multi-Layer Perceptron
NDA	Nos demais ambientes
OGI	Oregon Graduate Institute
PLN	Processamento da Linguagem Natural
PPGC	Pós-Graduação em Computação
PSD	Processamento de Sinal Digital
PSOLA	Pitch-Synchronous Overlap-and-Add
SFS	Sons foneticamente semelhantes
TPF	Texto para fala
UFRGS	Universidade Federal do Rio Grande do Sul
VODER	Voice Operating Demonstrator

Lista de Figuras

FIGURA 2.1 – Uma visão do trato vocal.....	14
FIGURA 2.2 – Um diagrama de fluxo de dados do processo de leitura oral.....	15
FIGURA 2.3 – Corte transversal de visão do olho humano.....	16
FIGURA 2.4 – Cada metade do campo de visão é processado por um hemisfério separado do cérebro	16
FIGURA 2.5 – Modelo de Wernicke-Geschwind para leitura em voz alta	18
FIGURA 2.6 – Uma visão cooperativa do reconhecimento de palavra	19
FIGURA 2.7 – Corte transversal do aparelho vocal	20
FIGURA 2.8 – Modelo de pensamento, linguagem e conversões de fala.....	22
FIGURA 2.9 – Conversões importantes no processamento da fala	23
FIGURA 2.10 – Aparelho auditivo	24
FIGURA 2.11 – Esquerda: Curvas isossônicas em um campo aberto. Direita: Máscara auditiva para um ruído de banda estreita	25
FIGURA 2.12 – Contextos mais freqüentes.....	32
FIGURA 3.1 – Modelo articulatório, com flechas rotuladas indicando graus de liberdade e linhas pontilhadas mostrando os relacionamentos entre os parâmetros.....	38
FIGURA 3.2 – Diagrama de bloco simplificado de um sintetizador de formantes	39
FIGURA 3.3 – Diagrama de bloco de um sintetizador de formantes em cascata-paralelo	39
FIGURA 3.4 – Diagrama funcional geral de um sistema TPF	41
FIGURA 3.5 – Esqueleto funcional de um módulo PLN.....	41
FIGURA 3.6 – Uma velha estratégia de processamento de linguagem natural para síntese TPF: estruturas de troca linear entre módulos de processamento organizados sequencialmente	43
FIGURA 3.7 – Um exemplo de estrutura característica quando o módulo de análise de sintaxe operou em um fragmento de sentença alemã <i>desto fester hüllte sich der Wanderer in seinen Mantel ein</i>	44
FIGURA 3.8 – MLDSs e FSs teoricamente permitem sincronização serial, hierárquica ou heterárquica	45
FIGURA 3.9 – O módulo PLN de um típico sistema de conversão TPF.....	45
FIGURA 3.10 – Típico analisador morfossintático	47
FIGURA 3.11 – Um típico sintetizador baseado em regras.....	53
FIGURA 3.12 – Um sintetizador geral baseado em concatenação	54
FIGURA 3.13 – Suavização linear paramétrica na borda de segmentos sucessivos	56
FIGURA 4.1 – CSLU Toolkit	58
FIGURA 4.2 – Níveis do CSLU Toolkit.....	58
FIGURA 4.3 – Ícones dos objetos base do RAD.....	59
FIGURA 4.4 – Ícones dos objetos Tucker Maxon.....	59
FIGURA 4.5 – Ícones dos objetos PSL.....	59
FIGURA 4.6 – Interface do BaldySinc	60
FIGURA 4.7 – Interface do Speech View	65
FIGURA 4.8 – Visão geral do processo de reconhecimento de fala.....	66
FIGURA 4.9 – Cálculo das propabilidades de cada categoria fonética pela rede neural	67

FIGURA 4.10 – Exemplo do uso do algoritmo de Viterbi	67
FIGURA 4.11 – Modelamento dependente de contexto	69
FIGURA 5.1 – Fluxograma da análise estatística	76
FIGURA 5.2 – Distribuição das palavras do dicionário	81
FIGURA 5.3 – Ocorrência do fonema “e” aberto e fechado.....	81
FIGURA 5.4 – Ocorrência do fonema “o” aberto e fechado	82

Lista de Tabelas

TABELA 2.1 – Fonemas consonantais	28
TABELA 2.2 – Correspondência entre fonema e letra	29
TABELA 2.3 – Fonemas distintos	29
TABELA 3.1 – Níveis de representação	48
TABELA 4.1 – Transcrições fonéticas	68
TABELA 4.2 – Exemplo de modelamento dependente de contexto.....	69
TABELA 4.3 – Lista de Fones	72
TABELA 5.1 – Fonemas respectivos das vogais “e” e “o”	76
TABELA 5.2 – Exemplo de palavras encontradas no dicionário	76
TABELA 5.3 – Palavras com “e” em seus radicais	77
TABELA 5.4 – Palavras com “o” em seus radicais	77
TABELA 5.5 – Exemplo da Análise estatística da ocorrência de ‘e’ fechado (/e/) e ‘E’ aberto (/E/) (onde ‘#’ significa nada)	78
TABELA 5.6 – Exemplo da Análise estatística da ocorrência de ‘o’ fechado (/o/) e ‘O’ aberto (/>/) (onde ‘#’ significa nada).....	79
TABELA 5.7 – Regras estabelecidas pela análise estatística.....	82
TABELA 5.8 – Palavras que geraram as regras.....	82
TABELA 5.9 – Lista de palavras usadas para o teste da regra 1 (AES)	83
TABELA 5.10 – Lista de palavras usadas para o teste da regra 2 (BEF)	83
TABELA 5.11 – Lista de palavras usadas para o teste da regra 3 (HEQ).....	84
TABELA 5.12 – Lista de palavras usadas para o teste da regra 4 (IEL).....	84
TABELA 5.13 – Lista de palavras usadas para o teste da regra 5 (JEC)	86
TABELA 5.14 – Lista de palavras usadas para o teste da regra 6 (JEL)	86
TABELA 5.15 – Lista de palavras usadas para o teste da regra 8 (LEL)	86
TABELA 5.16 – Lista de palavras usadas para o teste da regra 9 (XEL).....	87
TABELA 5.17 – Lista de palavras usadas para o teste da regra 10 (XEQ).....	87
TABELA 5.18 – Lista de palavras usadas para o teste da regra 11 (DOQ)	88
TABELA 5.19 – Lista de palavras usadas para o teste da regra 12 (DOX)	88
TABELA 5.20 – Número de acertos e respectivo percentual	89

Resumo

Sabe-se que a fala é a principal maneira de comunicação entre as pessoas. A Síntese de fala (geração automática da fala pelo computador) tem recebido atenção da comunidade acadêmica e profissional por várias décadas. Ela envolve a conversão de um texto de entrada em fala, usando algoritmos e algumas formas de fala codificada. O texto pode ser digitado pelo teclado ou obtido por reconhecimento de caracteres ou, ainda, obtido de um banco de dados. A síntese de fala pode ser usada em vários domínios de aplicação, tais como: auxílio para deficientes visuais, telecomunicações, multimídia, etc.

Este trabalho apresenta um estudo sobre a produção da fala e da área de síntese de fala visando servir de subsídio para dissertações e pesquisas futuras, bem como para o Projeto Spoltech, um projeto de cooperação entre os Estados Unidos e o Brasil para o avanço da tecnologia da língua falada no Brasil (Português Brasileiro).

Dentro deste estudo serão apresentadas as principais técnicas de síntese de fala, entre as quais destaca-se: Texto para Fala (TPF). Problemas de separação de sílabas, determinação da sílaba tônica, pronúncia das vogais “e” e “o” como um fonema aberto ou fechado, etc, são enfrentados dentro do contexto da área de síntese de fala para o português falado no Brasil.

Tendo conhecimento destes problemas, o principal objetivo deste trabalho será criar regras para resolver o problema de pronúncia das vogais “e” e “o” de forma automática, visando obter produção sonora mais inteligível, por intermédio da implementação de um analisador estatístico, o qual verificará a letra anterior e posterior ao “e” ou “o” de uma palavra e, com isso, determinar a pronúncia dos mesmos para aquela seqüência de letras. As mesmas poderão tornar-se regras válidas para a solução do problema se atingirem 80% dos casos de ocorrência no dicionário com fonema “e” ou “o” aberto (limiar), sendo que elas serão lidas por um interpretador Scheme utilizado pelo programa Festival - ferramenta para a construção de sistemas de síntese de fala desenvolvida pelo *Centre for Speech Technology Research* (University of Edinburgh, Reino Unido), a qual utiliza TPF como método de síntese. Sabendo-se que o Festival gera os fonemas “e” e “o” como fechados se não há uma regra para inferir o contrário, serão consideradas apenas as regras encontradas para os fonemas abertos.

Para possibilitar esta análise será utilizado um dicionário eletrônico de pronúncia (com 19.156 palavras), o qual possui a palavra e a sua respectiva pronúncia, conforme pode-se verificar no exemplo do Anexo 1.

Palavras-chave: Síntese de fala, Linguística Computacional.

TITLE: “DEVELOPMENTING OF PRONUNCIATION RULES FOR SPEECH SYNTHESIS IN BRAZILIAN PORTUGUESE LANGUAGE”

Abstract

Speech is the principal way people use to communicate. Speech Synthesis, automatic generation by Speech Computers, has received attention from the academic and professional community for many decades. That involves the conversion of plain text to speech, using algorithms and some forms of codified speech. The plain text can be typed through the keyboard, by letter recognition, or even exported from a database. The Speech Synthesis can be used in various applications such as: an auxiliary device for the visual handicap people, in telecommunications, in multimedia and so on.

This work introduces a study about the Speech Synthesis area aiming to help future research studies and the Spoltech Project, a project of co-operation between USA and Brazil for the advance of Spoken Language Technology in Brazil.

This work will present the Speech Generation and Speech Synthesis's principals techniques, among them: Text to speech. Problems of syllable separation, determination of tonic syllable, pronunciation of vowels “e” and “o” as open or closed phonemes, etc, are encountered in the context of Speech Synthesis in Portuguese Language Spoken in Brazil.

With these concepts in mind, the project also aims at creating rules to solve the problem of pronunciation for vowels “e” and “o” by automatic way, obtaining sound production more intelligible for Portuguese Language Spoken in Brazil, by implementation of an Statistics Analyzer, that will verify the previous and next letter to “e” and “o” in different words and will determine the pronunciation of these vowels for some sequence of letters in study. These sequence of letters including “e” and “o” would be transformed in automatic pronunciation rules if they reach 80% of entries in dictionariy (threshold), which will be read by scheme interpretator used by Festival Software – an engine for construction of Speech Synthesis Systems developed by the Centre for Speech Technology Research (University of Edinburgh, UK), that use TTS like Speech Synthesis method. Knowing that Festival generates phonemes “e” and “o” as closed ones if there isn't a rule determining the opposite in this work, only the rules matched for open phonemes will be considered.

For the establishing of rules, a pronunciation dictionary was analyzed (19,156 entries), that possesses words and its pronounciations, what can be verified in annex 1.

Keywords: Speech Synthesis, Computational Linguistic.

1 Introdução

A linguagem, seja escrita ou falada, é o meio de comunicação entre o homem e a máquina. Até os anos 70, tinha-se um "mosteiro" de especialistas em informática para manipular o computador; homens que dominavam a utilização de linguagens especializadas/específicas no uso do computador. Na década de 80, com o advento e expansão dos microcomputadores e suas redes, o usuário final foi colocado frente-a-frente com a máquina e seus dados. Desta forma, linguagens mais naturais, bem como meios de acesso mais naturais eram prementes[HUG 95].

Os substratos físicos da comunicação da fala humana são sinais audio-visuais. Estes sinais audíveis e visíveis são produzidos pelos movimentos de diferentes órgãos tais como o diafragma, a laringe, a mandíbula, o palato, a língua e os lábios. Estes movimentos devem ser coordenados de tal maneira que os sinais audio-visuais resultantes sejam suficientemente distinguíveis com respeito à estrutura da linguagem e mais geralmente, a situação da comunicação. Enquanto as expectativas dos ouvintes podem compensar a raridade de informações físicas, o locutor deveria ser capaz de manter um grande conjunto de contrastes fonológicos dependendo, em particular, da preparação do fonema da linguagem[BAI 90].

Síntese de fala é um processo que artificialmente produz fala para várias aplicações, diminuindo a dependência em usar uma voz gravada de uma pessoa. Os métodos de síntese capacitam uma máquina a passar informações ou instruções para o usuário através da fala[FUR 89]. Os principais beneficiados com estas aplicações de síntese de fala serão os deficientes visuais, pois poderão interagir com o sistema computacional mais facilmente.

O que levou a estudar este assunto foi o fato de ser uma área bastante promissora na área da informática, a qual oferece grandes ganhos para os usuários de computador, principalmente deficientes visuais, por ser um assunto pouco abordado nas dissertações de mestrado do Instituto de Informática e o interesse no desenvolvimento científico próprio e por consequência do Instituto de Informática na referida área. Ou seja, desenvolver um trabalho que sirva de base para novas investigações e progressos dentro do âmbito da Pós-Graduação em Computação para o assunto referido.

Este trabalho situa-se dentro do Projeto Spoltech, o qual é um projeto de cooperação entre os Estados Unidos e o Brasil para o avanço da tecnologia da linguagem falada no Brasil (Português), dentro da área de tecnologias da fala. Nele apresenta-se uma visão da forma cognitiva de produção da fala até a sua representação fonêmica, descrição do que é síntese de fala, quais métodos existem, uma descrição geral do Projeto Spoltech, as ferramentas utilizadas, o processo de reconhecimento e síntese de fala e, por fim, o problema a ser solucionado por este trabalho e a implementação para o mesmo.

O objetivo principal deste trabalho é apresentar melhorias de desempenho nas saídas produzidas pelo sistema sintetizador de fala Festival, adaptado para o Português, já que o mesmo foi inicialmente desenvolvido para Inglês. Várias melhorias de pronúncia e entonação poderiam ser realizadas. No entanto, devido à complexidade da tarefa, optou-se por desenvolver regras de pronúncia para o Festival com o intuito de

resolver o problema do “e” e “o”, já que o sintetizador sem uma regra gera os fonemas fechados para qualquer palavra com “e” e “o”, ou seja, corrigir a pronúncia daquelas palavras que possuam um fonema aberto. Utilizou-se para isso, um dicionário eletrônico de pronúncia e a implementação de um analisador estatístico, o qual analisará as palavras do respectivo dicionário para determinar a letra anterior e posterior ao ‘e’ ou ‘o’ aberto ou fechado. Após este levantamento, determina-se qual o conjunto de letras anterior e posterior que atingem 80% de ocorrências de ‘e’ ou ‘o’ aberto (limiar) e, por consequência, torna-se uma regra válida para a solução do problema em questão. Por exemplo, se no dicionário consta a palavra “bola” e o fonema “o” entre as letras “b” (anterior) e “l” (posterior) fosse aberto em 80% dos casos de ocorrência dessas letras entre o fonema, esta seqüência de letras seria considerada uma regra de pronúncia (**bol**) para todas as palavras que a possuírem, ou seja, gerando a pronúncia do fonema “o” como aberto.

Por conseguinte, ver-se-á uma descrição sumária do que cada capítulo do trabalho aborda.

No capítulo 2 apresenta-se a forma cognitiva de produção da fala até a representação fonêmica. Ou seja, dá-se uma breve descrição de como é formada a fala, a partir do pensamento até a mesma. Mais adiante fala-se sobre a fonêmica, apresentam-se os fonemas da língua portuguesa, as diferenças entre fonema e som, a classificação dos fonemas, a correspondência entre fonema e letra e suas premissas.

Já no capítulo 3 apresentam-se alguns métodos de síntese de fala (articulatória, formantes, LPC, concatenação da forma de onda e Texto para Fala), centrando o enfoque deste capítulo no método Texto para Fala. Neste enfoque, apresentam-se os módulos e procedimentos deste meio de síntese.

No capítulo 4 explana-se sobre o projeto em que este trabalho está situado(**Spoltech**), seus objetivos, as ferramentas utilizadas, o processo de reconhecimento e, por fim, o processo de síntese da fala.

No capítulo 5 faz-se referência à implementação desenvolvida. Situa-se o leitor no problema a ser desenvolvido neste trabalho (o problema do “e” e “o” aberto ou fechado). A seguir, mostra-se a solução desenvolvida por este trabalho (análise estatística) sobre as palavras de um dicionário. Concomitantemente, apresentam-se os testes realizados e os resultados alcançados por este desenvolvimento.

Seguem-se a conclusão, anexos, referências e obras consultadas.

2 Produção da Fala

Fala serve para comunicar informação de um locutor para um ou mais ouvintes. A produção da fala inicia quando o locutor formula a mensagem que deseja transmitir através da voz. O locutor produz um sinal de fala na forma de ondas de pressão que viajam da cabeça do locutor para os ouvidos dos ouvintes. As variações de amplitude correspondem a desvios de pressão atmosférica causados pela viagem das ondas. Fala pode ser dividida em segmentos de som, que compartilham algumas propriedades acústicas e articulatórias comuns umas com as outras, por um curto intervalo de tempo. Para cada som, há um posicionamento dos articuladores do trato vocal, que são: cordas vocais, língua, lábios, dentes, palato e maxilar (figura 2.1). Os sons são tipicamente divididos em duas classes: (a) vogais (conhecidas como sons sonoros ou vozeados), que permitem fluxo de ar irrestrito no trato vocal, e (b) consoantes, que restringem o fluxo de ar no mesmo ponto e têm uma intensidade mais fraca do que as vogais.

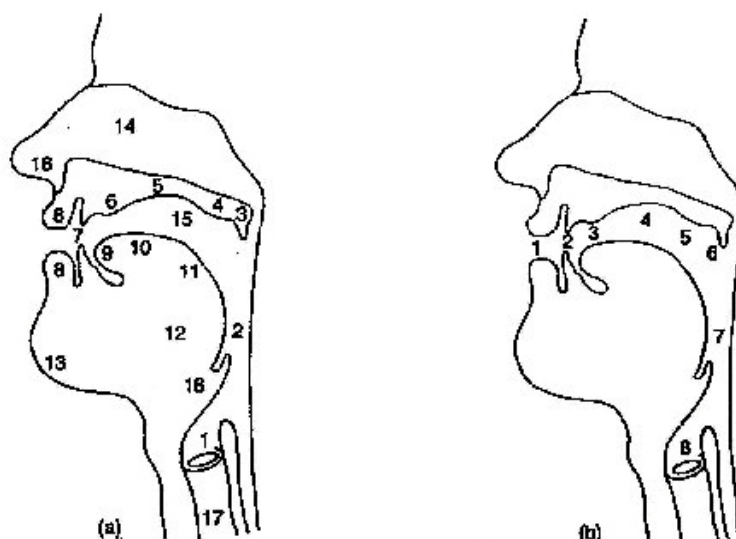


FIGURA 2.1 - Uma visão do trato vocal. (a) Articuladores da fala: (1) cordas vocais, (2) faringe, (3) palato, (4) palato mole, (5) palato duro, (6) cadeia alveolar, (7) dente, (8) lábios, (9) extremidade da língua, (10) lâmina, (11) dorso, (12) raiz, (13) mandíbula, (14) cavidade nasal, (15) cavidade oral, (16) narinas, (17) traquéia, (18) epiglote. (b) locais de articulação: (1) labial, (2) dental, (3) alveolar, (4) palatal, (5) velar, (6) uvular, (7) faríngeo, (8) glótico [OSH 2000]

Depois de uma inalação preparatória do ar nos pulmões, a fala é produzida quando o ar é exalado. Mudanças nas posições articulatórias influenciam a saída de ar dos pulmões. A produção de fala pode ser vista como uma operação de filtragem em que uma fonte de som excita o filtro do trato vocal; a fonte pode ser periódica, resultando em uma fala vozeada, ou não-periódica, causando fala não-vozeada. A fonte de fala ocorre na laringe, onde o fluxo de ar pode ser interrompido periodicamente pela vibração das cordas vocais. Os pulsos de ar produzidos pela abdução e adução (abertura e fechamento, respectivamente) das cordas vocais geram uma excitação periódica para o trato vocal.

Fala não vozeada é ruidosa devido à natureza randômica do sinal gerado em uma constrição estreita do trato vocal para tais sons. Para a excitação vozeada e não-vozeada, o trato vocal, agindo como um filtro, amplifica certas frequências de sons enquanto atenua outras.

Como um sinal periódico, fala vozeada tem um espectro consistindo de harmônicas da frequência fundamental da vibração do trato vocal; esta frequência, freqüentemente abreviada como F0, é o aspecto físico da fala correspondendo a *pitch* (frequência fundamental) percebida. As harmônicas são as concentrações de energia nos múltiplos de F0. Um sinal periódico tem um espectro discreto, mas desde que o trato vocal muda a forma quase continuamente, sons vozeados são de fato apenas localmente quase periódicos[OSH 2000].

2.1 O processo de leitura

Uma descrição razoável do processo de leitura então chamada de modo claro para aspectos funcionais tanto quanto fisiológicos, como na figura 2.2, que introduz um simples diagrama do fluxo do ato da leitura. A imagem é inserida nos neurônios sensoriais dos nossos olhos e transmitida na forma de estímulos elétricos para a nossa massa cinzenta, onde é processada para comandar os neurônios motores responsáveis pela correta ativação de nossos pulmões, cordas vocais e músculos articulatórios. Isto resulta na produção da fala, permanentemente monitorada pelo cérebro, principalmente através dos órgãos da audição, para ajustar em tempo real a configuração do trato vocal, junto com a energia do sinal expresso e a entonação aplicada a este[DUT 97].

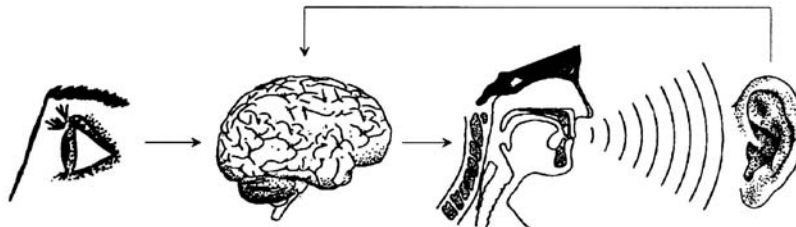


FIGURA 2.2 - Um diagrama de fluxo de dados do processo de leitura oral [DUT 97]

2.2 Visão

A figura 2.3 apresenta um corte transversal do olho. A imagem é capturada pela pupila, através da abertura ajustada pela íris, invertida pelas lentes, e jogada na retina. A retina contém dois tipos de fotorreceptores – chamados de cones e bastonetes. Cones são as células que detectam forma e cor e são responsáveis pela visão diurna; os bastonetes capacitam a visão noturna. A informação sensorial coletada pelos cones é agrupada no disco óptico, ou ponto cego, uma região da retina que não tem fotorreceptores, e é transmitida para o córtex visual através do nervo óptico.

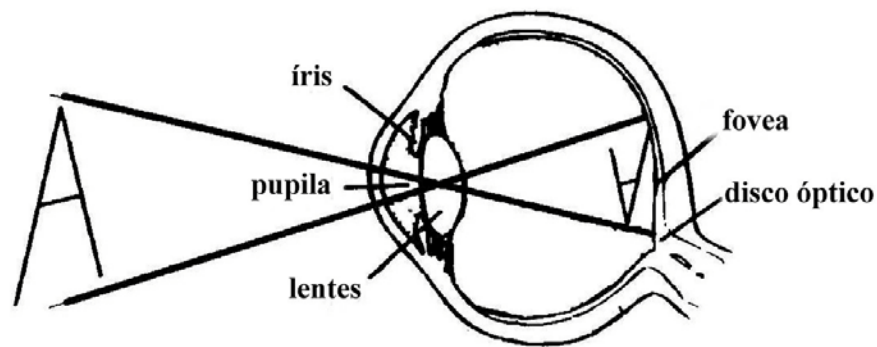


FIGURA 2.3 - Corte transversal de visão do olho humano [DUT 97]

É frequentemente errado acreditar que a informação sensorial capturada por cada olho é processada pelo hemisfério contralateral do cérebro. O que realmente acontece é que os neurônios em cada retina são separados em dois grupos, que respectivamente constituem as hemiretinas temporal e nasal. Ambas são projetadas a um hemisfério diferente, em uma região chamada de núcleo articulado lateral (figura 2.4). Também, é mais adequado dizer que cada metade do cérebro é preocupada com o campo semi-contralateral de visão [DUT 97].

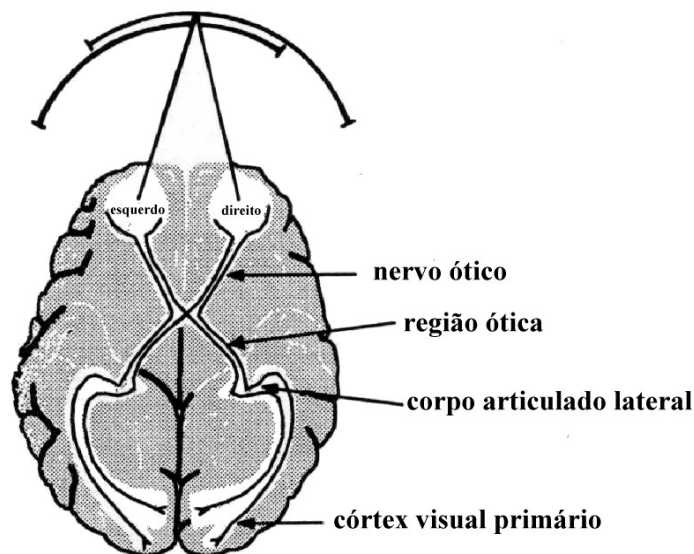


FIGURA 2.4 - Cada metade do campo de visão é processada por um hemisfério separado do cérebro [DUT 97]

O núcleo articulado envia os seus resultados para a massa cinzenta, em uma região chamada de córtex visual primário na região occipital do cérebro, onde eles são extensivamente analisados para decompor a parte visual em segmentos curtos de várias orientações (discriminação de forma e movimento) e combinam a saída dos dois olhos (percepção de profundidade).

A informação em cada área do córtex visual é claramente processada em um caminho distribuído. De fato, o córtex é composto de seis camadas organizadas em

colunas verticais. A organização completa do processamento é uma combinação de computação hierárquica e distribuída. A hierarquia é realçada quando é observada a ativação das células de área à área como uma função do sinal visual apresentado no olho. Células complexas, por exemplo, pesquisam a atividade de grupos de células simples, que pesquisam a atividade de grupos de células articuladas. Por outro lado, o paralelismo origina-se do fato que diferentes tipos de informação são enviadas para áreas diferentes do cérebro, onde elas são processadas simultaneamente.

Isto levou muitos pesquisadores a contextualizar a percepção como uma característica distribuída. Células em muitas áreas codificariam para um aspecto particular do estímulo: forma, profundidade, movimento, e cor.

Entre as observações úteis em favor da hipótese do processamento paralelo, há uma evidência clínica de uma série de perdas parciais da percepção visual devido a um dano cerebral denominado de **visual agnosias**. Pacientes, por exemplo, perdem independentemente a habilidade para perceber cores, ou formas e padrões, ou movimentos e profundidades[DUT 97].

2.3 Pensamento

Uma resposta parcial para questões como: “do que é feito o pensamento?” e “O que acontece à informação visual quando ela é processada pelo córtex visual?”, pode ser encontrada no estudo das desordens da linguagem e da leitura, respectivamente denominadas de afasias e alexias, e da justificação neurobiológica do modelo de Wernicke-Geschwind, apresentado na figura 2.5. De acordo com este modelo, pronunciar um texto envolve transferir informação visual do córtex visual primário para o córtex visual de ordem mais alta, e de lá para uma região específica da junção do córtex denominada de giro angular. Esta é então projetada para a chamada área de Wernicke, situada no hemisfério esquerdo, onde a percepção das palavras são formadas e onde as imagens de memória relacionadas são armazenadas[DUT 97]. As imagens são produzidas por meio da *arcuate fasciculus*, a área do Broca, também no hemisfério esquerdo, onde a percepção de palavras isoladas é combinada em uma estrutura gramatical e semântica e, por conseguinte, a memória para articulação de palavras é armazenada. Finalmente, o padrão de som da sentença é transmitida para o córtex motor da área facial, que está encarregado da articulação.

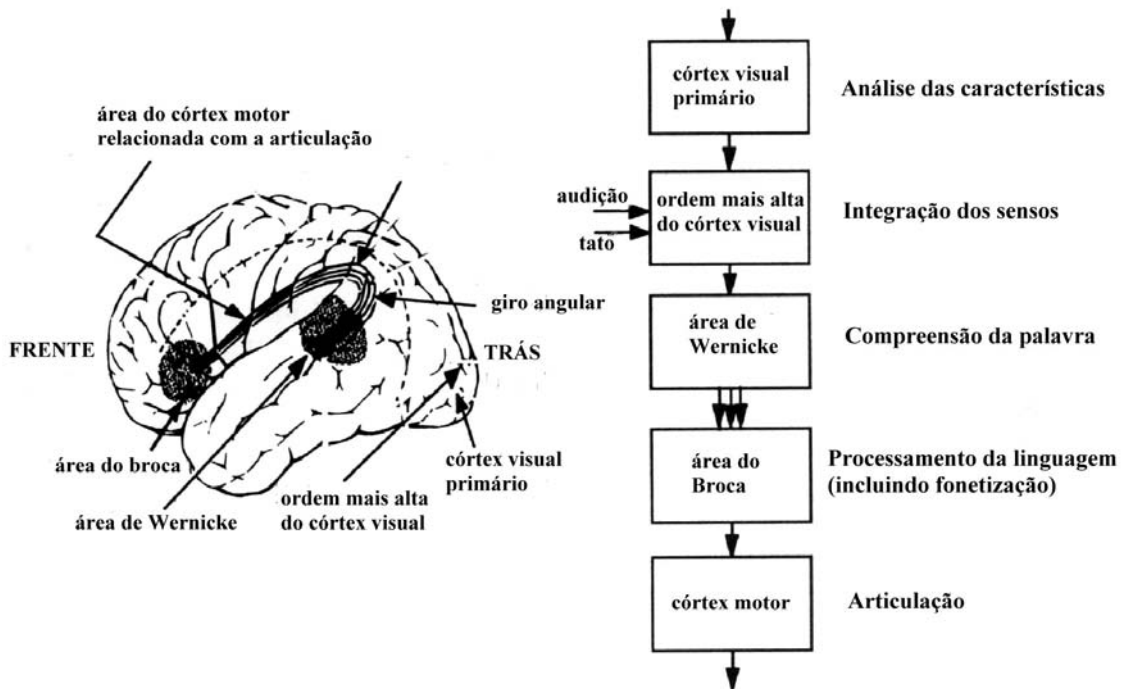


FIGURA 2.5 - Modelo de Wernicke-Geschwind para leitura em voz alta [DUT 97]

A justificação do modelo de Wernicke-Geschwind reside principalmente no número de previsões que é permitido sobre *afasias*. Primeiro, este prevê o efeito de uma lesão na área de Wernicke (que resulta na chamada afasia de Wernicke): pacientes são incapazes de entender o texto que eles lêem, enquanto a fala espontânea deles permanece fluente (ritmo e melodia). Eles também falham em expressar as idéias que eles têm em mente. Segundo, a afasia de Broca, causada pela lesão na área de Broca. A compreensão é mantida, mas a leitura e a fala espontânea são fortemente prejudicadas. Pacientes que sofrem dessa doença, tipicamente usam apenas palavras-chave e principalmente as bases da palavra (substantivos singulares e infinitivos). Terceiro, prever uma lesão que poderia deixar a área de Broca e de Wernicke intactas, mas desconectadas uma da outra: afasia condutiva. Compreensão é boa e a saída verbal permanece fluente, mas desde que a área de Broca não recebe entrada da área de Wernicke, a leitura em voz alta é prejudicada [DUT 97].

O que realmente acontece na área de Broca e de Wernicke é ainda desconhecido. O que é bastante certo, entretanto, é que a percepção da palavra e compreensão precedem uma profunda análise lingüística, tais como a codificação fonética ou análise sintática. Isto leva-se a pensar que, quando lê-se, primeiro associa-se significados às palavras e deixa-se estes significados abastecerem-se com a maneira que eles são expressos.

Tal transformação conceito para fonema provavelmente acontece para muitas das palavras que nós usamos correntemente; algumas nós reconhecemos num piscar de olhos, com uma organização do processamento paralelo distribuído como aquela que é mostrada no ramo direito da figura 2.6. Nosso conhecimento lingüístico nos designa a encontrar rapidamente um conceito associado, que induz a contraparte fonética, da mesma maneira que nós colocamos um nome em uma face familiar sem ter de analisar

cada um de seus traços faciais. Além disso, a existência de tal associação conceito para fonema, como oposto a um analítico, baseado em regras, conversão grafema para fonema, está em perfeito entendimento com o fato que, durante um diálogo, nós dificilmente nos referimos a ortografia das palavras que nós usamos: conceitos são diretamente colocados em correspondência com as pronúncias.

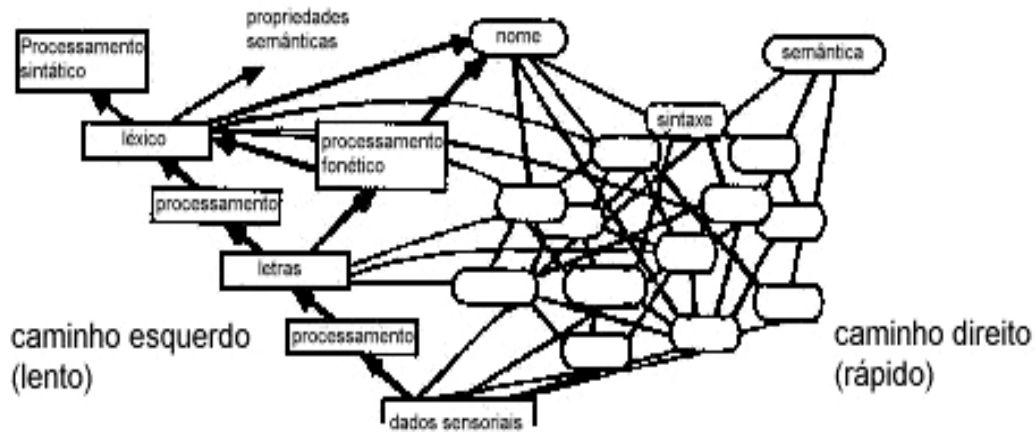


FIGURA 2.6 - Uma visão cooperativa do reconhecimento de palavra [DUT 97]

Na pronúncia de uma sentença é recomendável associá-la com uma melodia coerente, ritmo e intensidade – prosódia. Os únicos dados fisiológicos que nós temos neste aspecto da fala originam-se de recentes descobertas dos distúrbios dos componentes afetivos da linguagem: *aprosodias*. Foi descoberto, de fato, que a prosódia depende de certas áreas especializadas no hemisfério direito. A organização anatômica do hemisfério direito parece espelhar a organização da contraparte esquerda a este respeito: pacientes com lesões na região que reflete a área de Broca são capazes de organizar a prosódia fluentemente, enquanto lesões na região que corresponde a área de Wernicke impedem a compreensão [DUT 97].

2.4 Fala

Ler um texto não apenas implica reconhecer uma seqüência de caracteres e entender as palavras resultantes, mas também, grupos de palavras e sentenças. Eventualmente requer muitos músculos articulatórios do aparelho articulatório (figura 2.7) para ser comandado de uma maneira coordenada [DUT 97].

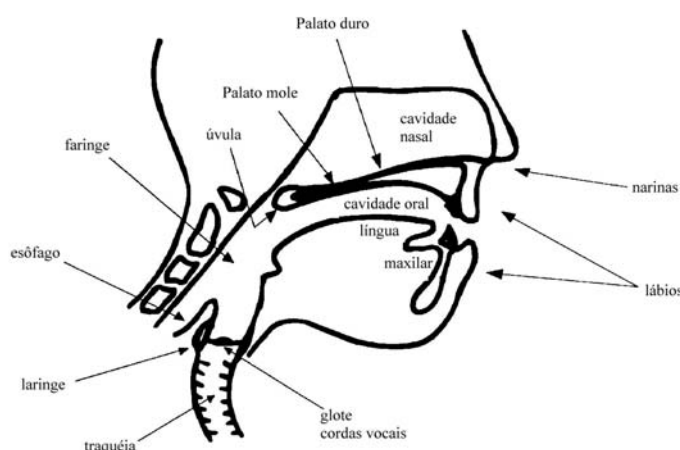


FIGURA 2.7 - Corte transversal do aparelho vocal [DUT 97]

Muitos estudos mostram que esta tarefa é carregada por algumas áreas corticais especializadas: as partes faciais dos córtices pré-motor e motor, que existem em ambos os hemisférios cerebrais. Além disso, estas áreas estão incluídas no modelo de Wernicke-Geschwind, o que pode explicar a base anatômica para alguns tipos de desordens no entendimento ou na execução de movimentos aprendidos, embora nenhuma paralisia possa ser encontrada. Tais perdas nos componentes cognitivos dos movimentos aprendidos são denominados como apraxias ideomotoras e são freqüentemente acompanhados por afasia para algumas extensões. Isto é facilmente entendido no caso da afasia condutiva. Pacientes com afasia condutiva (*conduction aphasia*), que se origina do rompimento do *arcuate fasciculus* do lobo parietal esquerdo, são ainda capazes de entender o que eles lêem, mas a informação relacionada não pode ser transmitida para a área de Broca ou para o córtice pré-motor. Como resultado, eles não podem articular as sentenças que eles lêem, enquanto eles podem articular fala espontânea. Outros tipos de *apraxia* podem resultar de destruição cirúrgica de algumas partes do corpo caloso (que conecta o hemisfério direito ao esquerdo). Os pacientes são incapazes de realizar comandos articulatórios com os músculos faciais direitos, enquanto a parte esquerda funciona perfeitamente[DUT 97].

Obviamente, estas áreas são a fonte de um problema clássico em pesquisa fonética e fala: a confusa diferença entre descrições físicas e lingüísticas da fala – entre o que é dito e o que realmente significa o que foi dito. Como mencionou Lindblom(1989): apesar de várias décadas de pesquisa fonética em muitas linguagens, ainda encontram-se sérias dificuldades quando especificam-se unidades fonológicas de tal maneira que a descrição fonética delas permanecerá invariante em uma grande escala de contextos que comunicativamente os atos da vida real apresentam. A questão é: que tipo de fonte faz o córtice pré-motor referir-se a transformar fonemas simbólicos em gestos articulatórios verdadeiros: articulatório, acústico ou auditivo[DUT 97].

Talvez nenhuma dessas. A busca por invariância tem conduzido as teorias alvo, que afirmam que os fonemas podem ser descritos em termos dos objetivos articulatórios que não seriam alcançados dada à preguiça dos articuladores. O resultante *undershoot* seria uma função da duração e contexto dos gestos articulatórios, enquanto os objetivos permaneceriam invariantes. Estas teorias têm se mostrado imperfeitas com os anos, embora elas tenham produzido consideráveis soluções para o problema. Este também

tem sugerido que sinais de fala poderiam conter invariâncias físicas absolutas na forma de padrões característicos correspondendo aos segmentos fonéticos. As muitas dificuldades que ainda surgem quando tenta-se automaticamente dividir fala contínua em fones provêm de algumas indicações da robustez de tal suposição. Eventualmente, *auditory gestures* pensava-se que exibiam invariância; embora gestos articulatórios pudessem mudar, a percepção deles permanecia constante. Novamente, descobertas experimentais mostraram a fraqueza destas idéias. Na prática, nós somos perfeitamente incapazes de expressar duas vezes uma sentença da mesma maneira, de um ponto de vista articulatório, acústico ou auditivo. Isto sugere que a invariância lingüística foi definida apenas no nível da compreensão do ouvinte: a variabilidade que os sinais de fala exibem é meramente uma função dos obstáculos que surgem nas interações locutor-ouvinte[DUT 97].

2.5 Do pensamento para a fala

Suponha-se a seguinte situação: duas jovens no Museu de arte da Filadélfia param diante da pintura de Henri Matisse intitulada de “Odalisque Jaune”. Uma delas diz a outra: “Veja esta pintura. Há algo sobre a face e as formas que me lembra de algumas pinturas que eu vi em um museu em Nova Iorque.” Nós não podemos presumir conhecer como esta locução foi derivada do conhecimento lingüístico da jovem e dos processos do pensamento dela, mas nós devemos supor que alguma referência foi feita de experiências visuais armazenadas sobre pinturas japonesas, e que associações foram feitas entre as formas da pintura japonesa e a justaposição de formas da pintura de Matisse[BOR 94].

Um modelo de pensamento, linguagem e conversões de fala é apresentado na figura 2.8. Os círculos sobrepostos sugerem a inter-relação envolvida entre eles e a simultaneidade. As experiências visuais e estéticas da jovem, ambas no presente e no passado, relatam as idéias que ela tem sobre as similaridades e os sentimentos delas sobre as pinturas[BOR 94].

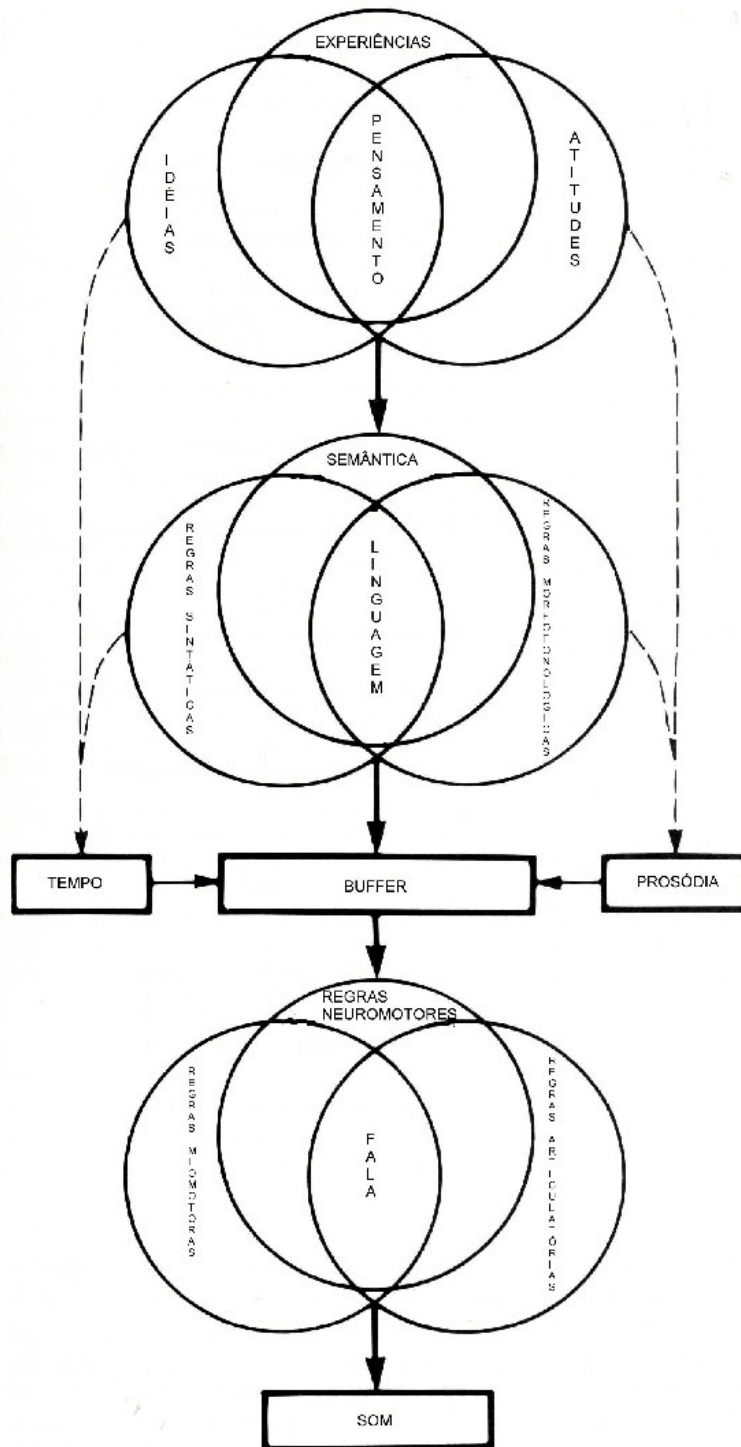


FIGURA 2.8 – Modelo de pensamento, linguagem e conversões de fala [BOR 94]

Há um número de maneiras em que a mulher poderia ter formado as suas idéias e sentimentos, mas baseado em certa semântica, sintaxe e decisões morfológicas. Ela foi compelida pelas regras da sua linguagem e pelas regras do seu mecanismo de produção da fala[BOR 94].

Tempo e aspectos prosódicos da expressão são vistos em nosso modelo como acrescentados sob a mensagem onde esta é convertida em fala. Também, a expressão pode ser dita em uma variedade de taxas temporais, da mais rápida para a mais lenta, indicando comandos de tempo um tanto distintos. O tempo dentro e através dos segmentos que distinguem fonemas, entretanto, é intrínseco às especificações deles armazenadas no buffer.

Pode haver uma transformação no nível da fala de uma representação abstrata da fala para a atual atividade neuromotora que controla a atividade muscular, mudanças da cavidade, e modificações da pressão do ar ouvidas como fala. Nas conversões de linguagem, as regras lingüísticas são usadas; nas conversões de fala, neuromotor, *myomotor* e regras articulatórias são usadas, omitindo pelo objetivo de simplicidade qualquer menção dos mecanismos de controle sensoriais. A figura 2.9 esquematiza as conversões vistas como importantes no processamento da fala.

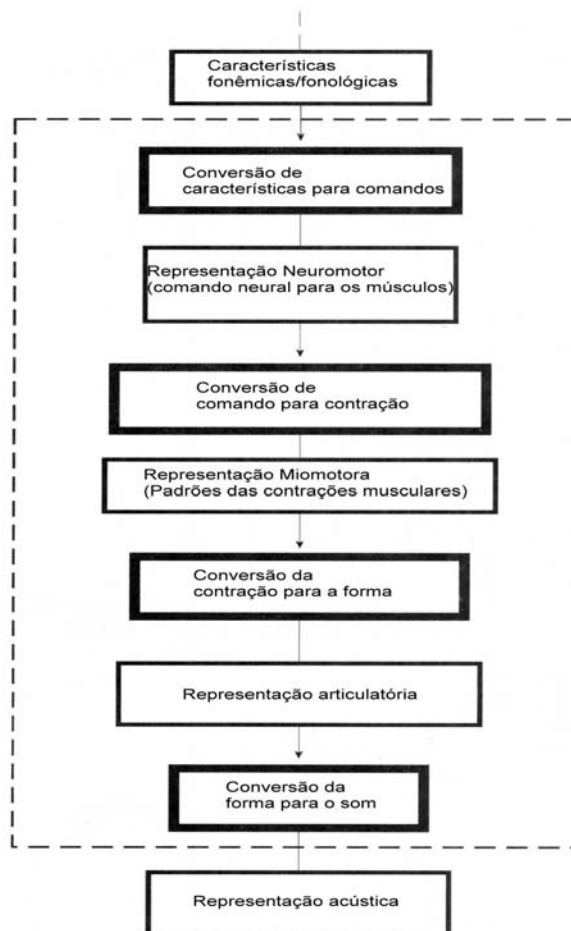


FIGURA 2.9 – Conversões importantes no processamento da fala [BOR 94]

No modelo (figura 2.8) as mesmas regras da fala são aplicadas. Elas ocorrem simultaneamente e relacionam-se umas com as outras.

No buffer, a expressão existe como uma representação interna da meta acústica do locutor e como uma representação interna da fisiologia da produção de fala em termos de coordenadas espaciais tridimensionais e tempo relativo. O locutor sabe os sons pretendidos e o que tem de ser feito para produzi-los[BOR 94].

2.6 Audição

A fala é coletada pelo aparelho auditivo, o qual é subdividido em ouvido externo, ouvido médio e ouvido interno (Figura 2.10). A vibração do tímpano é transmitida através de ossículos para a cóclea (ouvido interno), onde a transdução mecânica-neural toma lugar[DUT 97]. Cada som sinusoidal inicia a viagem de uma onda ao longo da membrana basilar, da qual a amplitude pico ocorre em diferentes posições como uma função da frequência do estímulo. Dada a distribuição delas ao longo da cóclea, cada película capilar é mais sensível a uma diferente banda de frequência, tal que uma pode de modo geral aproximar-se do processamento realizado pelo ouvido interno como análise espectral[DUT 97].

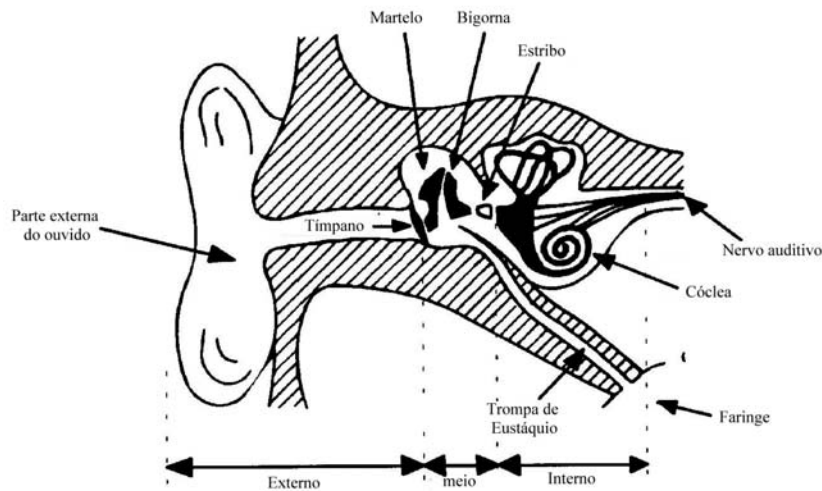


FIGURA 2.10 - Aparelho auditivo [DUT 97]

Contribuições de ambos os ouvidos são enviadas para ambas as metades do cérebro, de quem o processamento hierárquico paralelo resulta em uma percepção distribuída de sons.

Fundamentalmente, o ouvido responde diferentemente a todas as frequências. A figura 2.11 exhibe as curvas isossônicas – as curvas de intensidade perceptual igual, como uma função da frequência. Elas mostram que a máxima sensibilidade é alcançada na banda [500 Hz, 10 kHz][DUT 97]. Além disso, os sons podem ser mascarados por outros. Esta propriedade psicoacústica do ouvido humano é exemplificada na figura 2.11, que mostra como um ruído de banda estreita centrado em 1kHz localmente eleva o primeiro limiar de audibilidade. Não apenas qualquer som afeta a audição, mas o efeito é relacionado a esta força de uma maneira não linear (como é mostrado no painel esquerdo da figura 2.11).

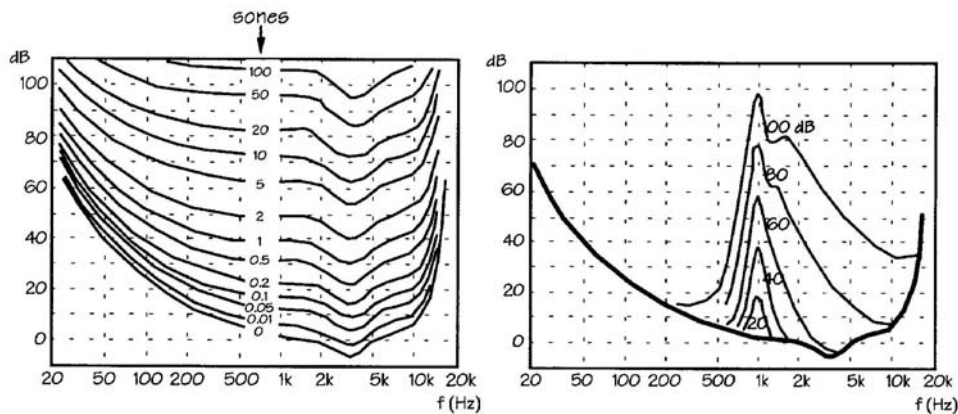


FIGURA 2.11 - Esquerda: Curvas isossônicas em um campo aberto. Direita: Máscara auditiva para um ruído de banda estreita [DUT 97]

Conhecimento lingüístico interfere nas propriedades psicoacústicas quando ouvindo, exatamente como este faz quando lêem-se textos. Conhecimento léxico, sintático e semântico é extensivamente empregado nesta tarefa. Fones são expressos sozinhos. O contexto fonético em que eles são incorporados é obviamente usado pelo cérebro. Alguns fonemas produzem mais informação do que outros, na medida em que eles são menos prováveis de aparecer em uma palavra para uma dada linguagem, tal que o reconhecimento exato deles de modo vantajoso restringe-se ao espaço de busca para os vizinhos plausíveis. Similarmente, as seqüências de fonemas são organizadas em palavras, que têm de obedecer a uma sintaxe e constituir uma sentença com significado. Por conseguinte, o conhecimento de uma determinada linguagem gera uma tremenda ajuda para o reconhecimento da fala.

2.7 Fonêmica

A organização da cadeia sonora da fala é orientada por certos princípios. Tais princípios agrupam segmentos consonantais e vocálicos em cadeia e determinam a organização das seqüências sonoras possíveis de uma determinada língua. Considere um exemplo concreto do português. Mesmo sem saber o significado da palavra, como por exemplo, “sali”, sabemos que a cadeia de segmentos é possível de ocorrer em uma palavra do português. Portanto, falantes do português interpretam “sali” como sendo uma palavra possível do português. Entretanto, uma palavra como “spali” não tem a mesma interpretação – uma vez que os falantes da língua portuguesa sabem que a seqüência “sp” não ocorre em início de palavra em português. Certamente, a palavra “spali” é interpretada como uma palavra estrangeira para os falantes do português.

Os segmentos consonantais e vocálicos organizam-se em estruturas silábicas formando palavras possíveis em uma determinada língua. Línguas variam quanto aos seus sons e quanto à organização da estrutura silábica.

Outro aspecto importante na organização da fala é a maneira como os segmentos consonantais e vocálicos afetam os seus segmentos adjacentes (que os precedem ou que os seguem). Sendo a fala contínua, observa-se que um segmento pode ser alterado por um segmento que o precede ou que o segue. A alteração de um segmento a partir de segmentos adjacentes se dá pelo fato dos segmentos em questão compartilharem de certas propriedades fonéticas. Um exemplo do português é a palatização de consoantes

velares – [k,g] – quando estas são seguidas da vogal **i**: “quilo” e “guia”. A propriedade de ser anterior da vogal **i** é compartilhada pela consoante precedente [k,g].

O objetivo da análise fonêmica é analisar a organização da cadeia sonora da fala do português a partir de pressupostos teóricos de tendência estruturalista. O termo fonologia passa a ser utilizado por modelos pós-estruturalistas que analisam a organização da cadeia sonora – ou componente fonológico. Portanto, ambos os termos fonêmica e fonologia referem-se a modelos que tratam do estudo da cadeia sonora da fala.

A relação entre uma representação fonética – entre colchetes – e uma representação fonêmica – entre barras transversais – não será necessariamente idêntica como o exemplo da palavra “babá”:[ba’ba] e /ba’ba/. Pode ter, por exemplo, a representação fonética [p̃l’toh] “pintor” que relaciona-se com a representação fonêmica /piN’toR/. Observe que no exemplo da palavra “pintor” a representação fonética – [p̃l’toh] é diferente da representação fonêmica - /piN’toR/[SIL 99].

2.7.1 Diferenças entre fonema e som[OLI 96]

Fonema:

- Unidade sonora mínima da língua
- Constitui a imagem mental do som
- É representado entre barras: /p/
- Seu número é limitado. Em português existem 33 fonemas
- Não tem significado por si mesmo, e sim um valor diferencial; serve para diferenciar uma palavra da outra

Fonologia:

- Pertence ao plano abstrato, isto é, à língua
- Sua unidade é o fonema

Som:

- Unidade sonora mínima da fala
- É a realização oral, física e individual do fonema
- É representado entre colchetes: [p]
- O número de sons é limitado

Fonética:

- Pertence ao plano concreto, ao da fala
- Sua unidade é o som

2.7.2 Como se produz a voz

Os animais emitem sons: miam, ladram, piam, relincham. Só os seres humanos articulam os sons. Para isso se utilizam dos órgãos da fala, chamados, em seu conjunto, de aparelho fonador. O ar sai dos pulmões, passa pelos brônquios, pela traquéia e chega à laringe. É esse ar que faz vibrar as cordas vocais. Diversos órgãos transformam a voz em sons articulados. O ar expirado passa à boca através da faringe. As cavidades supralaríngeas (faringe, boca e fossas nasais) atuam como caixas de ressonância. De

acordo com a posição que adotam os diferentes órgãos, são produzidos efeitos acústicos distintos[OLI 96].

2.7.3 Classificação dos fonemas

Os fonemas são as unidades sonoras mais simples da língua portuguesa. Dividem-se em vogais, semivogais e consoantes.

As vogais são sons que resultam das vibrações das cordas vocais. Na pronúncia das vogais, a boca sempre permanece aberta ou entreaberta. Já as consoantes são ruídos ocasionados pela ação dos órgãos da boca (língua, dentes, lábios, etc.), ao criarem obstáculo ou resistência à corrente de ar emitida dos pulmões[AND 82].

2.7.4 As semivogais

Entre as vogais e as consoantes estão as semivogais, que, diferentemente das vogais, nunca são base de sílaba, nem são tônicas. Elas acompanham as vogais, com as quais formam sílabas. São representadas fonologicamente pelos símbolos /j/ e /w/, que correspondem aproximadamente às vogais /i/ e /u/: caixa, lousa, pão[OLI 96].

2.7.5 Classificação das vogais

As vogais são classificadas, de acordo com quatro critérios [AND 82]:

A. quanto a zona de articulação:

- a. média: a – som produzido quando a língua se conserva quase em repouso, baixa, a boca aberta e o véu palatino (palato mole) se levanta impedindo a passagem da corrente de ar para as fossas nasais;
- b. anteriores: ê, é, i – sons produzidos quando a ponta da língua se eleva gradualmente em direção ao palato duro, e ao mesmo tempo vai diminuindo a abertura da boca; assim, ao proferirmos é, a língua principia aquela elevação e estreita-se um pouco a abertura da boca; ao pronunciarmos ê os dois movimentos se acentuam; quando dizemos i, a elevação da língua em direção ao palato duro é máxima e o estreitamento da boca é bem acentuado.
- c. posteriores: ó, ô, u – sons produzidos quando recuamos e elevamos progressivamente a parte posterior da língua em direção ao véu palatino, arredondando e estreitando gradativamente os lábios.

B. quanto ao timbre:

- a. abertas: a (*casa*), é (*café*), ó (*ótimo*) – sons proferidos quando a boca está mais aberta e é grande a distância entre a língua e o céu da boca; aquela abertura e esta distância são mais amplas para o a, e menos amplas para o é e ó;
- b. fechadas: ê (*cedo*), ô (*povo*), i (*missa*), u (*uva*) – sons proferidos quando a boca estreita a abertura e é pequena a distância entre a língua e o céu da boca; aquele estreitamento e esta aproximação são menos acentuados no ê e no ô, e mais acentuados no i e no u;
- c. reduzidas: a (*casa*), e (*leve*), o (*pálido*) – sons intermediários entre as vogais abertas e as fechadas, proferidos fracamente.

- C. quanto ao papel das cavidades bucal e nasal:
- orais: a, é, ê, i, ó, ô, u – quando a corrente de ar ressoa apenas na boca;
 - nasais: ã, ~e, ~i, õ, ~u
- D. quanto à intensidade:
- tônicas – aquelas vogais proferidas com maior intensidade: cipó, livre, pálido, pêndulo, pura;
 - átonas – aquelas vogais pronunciadas com intensidade fraca: cipó, livre, pálido, pura.

2.7.6 Classificação das consoantes

As consoantes são classificadas de acordo com quatro critérios[AND 82]:

- quanto ao modo de articulação: se o ar encontra um obstáculo total (oclusivas) ou parcial (constritivas). Se é parcial, o ar pode sofrer fricção (fricativas), passar pelos lados da cavidade bucal (laterais) ou fazer vibrar a língua ou o véu palatino (vibrante).
- quanto ao ponto de articulação: é o ponto em que se localiza o obstáculo encontrado pelo ar (lábios, palato duro, palato mole).
- quanto ao papel das cordas vocais: se vibram (sonoras), ou se não vibram (surdas).
- quanto ao papel das cavidades bucal e nasal: se o ar passa apenas pela cavidade bucal (orais), ou se ressoa na cavidade nasal (nasais).

Na tabela a seguir podemos verificar os fonemas consonantais de acordo com as suas classificações.

TABELA 2.1 – Fonemas consonantais

Função das cavidades bucal e nasal	Orais						Nasais	
	Oclusivas		Constritivas					
Modo de articulação			fricativas		vibrantes	laterais		
Função das cordas vocais	surdas	sonoras	surdas	sonoras	sonoras	surdas	sonoras	
Zona de Articulação	Bilabiais	p	b				m	
	Labiodentais			F	v			
	Linguodentais	t	d					
	Alveolares			S C Ç	s z	R Rr	l	n
	Palatais			X Ch	g j		lh	nh
	Velares	c(k) q	g(guê)					

Fonte: [AND82]

2.7.7 Correspondência entre fonema e letra [OLI 96]

Nem sempre a cada fonema corresponde uma só letra:

- Um mesmo fonema pode ser representado por várias letras:

TABELA 2.2 – Correspondência entre fonema e letra

Fonema	Letra
/s/	S (sapo); ss (passo) ç (maço); x (máximo)
/k/	C (caixa); q (queijo) k (karaokê)

Fonte: [OLI96]

- A mesma letra pode representar fonemas distintos

TABELA 2.3 – Fonemas distintos

Letra	Fonema
S	/s/ sopa; /z/ mesa
R	/r/ cara; /R/ roupa
I	/i/ saci; /j/ peixe
U	/u/ uva; /w/ lousa

Fonte: [OLI96]

2.7.8 Fonemas e alofones

Um dos objetivos de uma análise fonêmica é definir quais são os sons de uma língua que têm valor distintivo. Sons que estejam em oposição – por exemplo [f] e [v] em “faca” e “vaca” – são caracterizados como unidades fonêmicas distintas e são denominados fonemas.

O procedimento habitual de identificação de fonemas é buscar duas palavras com significados diferentes, cuja cadeia sonora seja idêntica. Assim, em português, definimos /f/ e /v/ como fonemas distintos uma vez que o par mínimo “faca” e “vaca” demonstra a oposição fonêmica. Diz-se que o par mínimo “faca/vaca” caracteriza os fonemas /f,v/ por contraste em ambiente idêntico (CAI).

Quando pares mínimos não são encontrados para um grupo de sons em uma determinada língua, podemos caracterizar os dois segmentos em questão como fonemas distintos pelo contraste em ambiente análogo (CAA). Assim, duas palavras que ocorram em ambientes similares podem caracterizar o contraste em ambiente análogo, desde que as diferenças entre os sons não sejam atribuídas aos sons. Ilustra-se o contraste em ambiente análogo com os sons [s] e [z] em português. Sabemos que em posição intervocálica os segmentos [s] e [z] são fonemas distintos, pois têm-se os pares mínimos que demonstram o contraste em ambiente idêntico entre estes dois sons.

Consideremos, contudo, o contraste entre [s] e [z] em início de palavra. Suponha que não encontre-se um par mínimo que demonstre o contraste em ambiente idêntico entre [s] e [z] em início de palavra. Para prosseguir-se à análise fonêmica, pode-se buscar um par de palavras bastante semelhante que caracterize a oposição fonêmica em início de palavra entre [s] e [z] por contraste em ambiente análogo. Um par de palavras

que demonstre o contraste fonêmico em ambiente análogo apresenta diferença segmental em relação à mais de um segmento. Um exemplo para demonstrar o contraste fonêmico em ambiente análogo entre [s] e [z] em posição inicial é o par de palavras “sumir/zunir”. Note que em “sumir/zunir” além da diferença segmental de [s] e [z] temos a diferença entre [m] e [n] precedendo a vogal tônica. Portanto, o par de palavras “sumir/zunir” demonstra o contraste em ambiente análogo entre [s] e [z] em posição inicial.

Trabalhar com uma língua que se conheça bem, certamente contribui para que os dados necessários para a análise sejam encontrados e quase que certamente pares mínimos são identificados para todos os fonemas da língua.

Do ponto de vista de representação tem-se aqui dois níveis: o fonético e o fonêmico. No plano fonético temos fones que transcrevemos entre colchetes, por exemplo [a]. São fones todos aqueles segmentos consonantais e vocálicos identificados na transcrição fonética do corpus. Em outras palavras, fones são os segmentos encontrados no quadro fonético. No plano fonêmico temos fonemas que transcrevemos entre barras transversais, por exemplo, /a/. A determinação de fonemas se dá a partir da identificação de pares mínimos para um grupo de dois segmentos. Certamente quanto mais se conhece a língua, mais dispõe-se de dados para identificar pares mínimos para quaisquer segmentos desta língua. Entretanto, há grandes chances de que segmentos como l e k sejam fonemas distintos em qualquer língua. Assim, mesmo que não tenhamos encontrado ainda pares mínimos para eles, podemos postular que l e k são fonemas distintos. Isto se dá, porque l e k não têm nenhuma similaridade fonética a não ser pelo fato de serem ambas consoantes.

(4) Sons foneticamente semelhantes

- a. *Um som vozeado e seu correspondente desvozeado.*
- b. *Uma oclusiva e as fricativas e africadas com ponto de articulação idêntico ou muito próximo.*
- c. *As fricativas com ponto de articulação muito próximo.*
- d. *As nasais entre si.*
- e. *As laterais entre si.*
- f. *As vibrantes entre si.*
- g. *As laterais, vibrantes e o tepe.*
- h. *Sons com propriedades articulatórias muito próximas.*
- i. *As vogais que se distinguem por apenas uma propriedade articulatória. Assim, [e, ε] constituem um par suspeito porque estas vogais diferem apenas quanto a uma propriedade articulatória (referente à altura). Por outro lado, [i,u] não representam pares suspeitos uma vez que estes segmentos diferem quanto à anteriorização/posteriorização e arredondamento/não-arredondamento.*

Viu-se então que na busca de identificar os fonemas de uma língua listam-se os pares suspeitos (sons foneticamente semelhantes) de segmentos consonantais e vocálicos. Passa-se então a buscar um par de palavras que venha a constituir um par mínimo para determinar os fonemas em questão. Quando não encontram-se pares mínimos (ou análogos) para dois segmentos suspeitos, concluímos que os segmentos em questão não são fonemas. Se não conseguir caracterizar dois segmentos suspeitos como

fonemas distintos deve-se buscar evidência para caracterizá-los como alofones de um mesmo fonema. Alofones (ou variantes) de um fonema são identificados por meio do método de distribuição complementar. Quando dois segmentos estão em distribuição complementar, eles ocorrem em ambientes exclusivos. Em outras palavras, onde uma das variantes ou alofone ocorre, a outra variante não ocorrerá. Esta distribuição deve ser válida para todas as palavras da língua em questão. O procedimento de identificação de alofones a partir do método de distribuição complementar é ilustrado abaixo[SIL 99].

(5) Considere os dados:

a. tatu [ταἔτυ]	e. tipo [τΣιπΥ]	ι. pátio [ἔπατΣιΥ]
b. tudo [τυδΥ]	φ. χαντιγα [κᾱ'τΣιγ↔]	φ. teto [□τΕτΥ]
χ. tinge [ἔτΣιΖΙ]	g. g. tingido [τΣ ι~~~ἔΖιδΥ]	κ. ética
[‘ΕτΣΙκ↔]		
d. trevo [τΡεϖΥ]	h. kátia [ἔκατΣΙ↔]	l. atlas [ἔατλ↔σ]

Observe que os segmentos [t] e [tΣ] correspondem, respectivamente, a uma oclusiva e a uma africada com pontos de articulação próximos. De acordo com os principais grupos de sons foneticamente semelhantes (SFS) listados em (4), uma oclusiva e uma africada com pontos de articulação próximos constituem um par suspeito. Para um par suspeito de sons devemos encontrar um par mínimo (ou análogo) que caracterize os segmentos em questão como fonemas distintos. Se não encontrar um par mínimo (ou análogo) devemos constatar a distribuição complementar identificando então a distribuição dos alofones. Uma análise preliminar dos dados acima, nos mostra que [t] ocorre seguido de [a,u, Y, P, E,I] e que [tΣ] ocorre seguido de [i,ĩ,I]. Podemos então formular uma hipótese de investigação. Tal hipótese tem por objetivo definir os ambientes em que [t] e [tΣ] ocorrem[SIL 99].

(6) Hipótese: O segmento [tΣ] ocorre seguido de [i] e suas variantes [ĩ,I] e o segmento [t] ocorre nos demais ambientes (NDA).

Poderia-se questionar a afirmação de que [t] e [tΣ] são alofones – ao apresentar pares mínimos como “tal-tchau” ou “tê(letra)-tchê(sulista)”. Em princípio, estes pares mínimos demonstram o status de fonemas distintos de [t] e [tΣ]. Contudo, o fato de pares mínimos como “tal-tchau” ou “tê(letra)-tchê(sulista)” ocorrerem em português, não invalida a análise de distribuição complementar. Isto ocorre porque em todos os dados de pares mínimos para [t] e [tΣ], as palavras que ilustram o exemplo com o [tΣ] devem ter foneticamente um [tΣ] em todo e qualquer dialeto do português (“tchau, tchê”, por exemplo). As palavras que apresentam [tΣ] em qualquer dialeto do português – tchau, tchê, tcheco-eslováquia, tcheco, tchurma – constituem um grupo restrito e são justificáveis como empréstimos.

Conclui-se então que a análise de distribuição complementar proposta – que define [t] e [tΣ] como alofones – é adequada. O próximo passo é definir um fonema que represente os alofones envolvidos na distribuição complementar dos segmentos [t] e [tΣ]. Tanto [t] quanto [tΣ] são considerados alofones e devemos selecionar um destes segmentos para representar o fonema. Optou-se por representar os alofones [t] e [tΣ] pelo fonema /t/ na distribuição complementar discutida acima. A escolha do fonema,

geralmente se dá por aquele alofone que tenha uma ocorrência mais abrangente ou mais geral em termos de distribuição. O outro alofone – geralmente com ocorrência mais restrita ou específica – representará um dos alofones daquele fonema. Escolhemos /t/ para representar o fonema dos alofones [t] e [t̺] porque o alofone [t] ocorre de maneira mais abrangente [NDA].

A seguir serão apresentadas as premissas da fonêmica, segundo [SIL 99].

2.7.9 As premissas da fonêmica

Apresentam-se nesta seção as quatro premissas básicas da fonêmica.

2.7.9.1 Premissa 1

Os sons tendem a ser modificados pelo ambiente em que se encontram.

Interpretando-se a fala continuamente, observa-se que os sons sofrem alterações dependendo do ambiente em que se encontram. A seguir, alguns ambientes ou contextos que mais frequentemente causam alteração na fala:

1) **Ambientes ou contextos propícios à modificação de segmentos**

- (a) *Sons vizinhos (precedentes ou seguintes)*
- (b) *Fronteiras de sílabas, morfemas, palavras e sentenças*
- (c) *A posição do som em relação ao acento*

Alguns símbolos são formalmente utilizados para caracterizar os contextos mais freqüentes, conforme ilustrado na figura 2.12. Observe que na caracterização dos contextos listados na figura, o espaço sublinhado indica o local em que se encontra o segmento, cujo contexto deseja-se descrever.

V <u> </u> V	Representa o contexto intervocálico(entre vogais)
# <u> </u>	Representa o início de palavra
<u> </u> #	Representa o final de palavra
<u> </u> + <u> </u>	Representa um limite de morfema
<u> </u> \$ <u> </u>	Representa um limite de sílaba

FIGURA 2.12 Contextos mais freqüentes [SIL 99]

Finalizando a discussão da primeira premissa – que estabelece que os sons tendem a ser modificados pelo ambiente em que se encontram – se discutirá alguns aspectos relacionados à nasalidade no português brasileiro. A nasalidade no português brasileiro relaciona-se ao fato de uma vogal ser nasalizada quando seguida de consoante nasal. Há contudo, grande variação quanto à nasalidade no português brasileiro dependendo do dialeto em questão. Em vários dialetos da região Sudeste, uma vogal tônica é obrigatoriamente nasalizada quando seguida de consoante nasal – “c[ã]ma”. Contudo, se a vogal seguida de consoante nasal ocorre em posição pretônica, a nasalidade é opcional: “c[a]mareira” ou “c[ã]mareira”. Já em certos dialetos do estado de São Paulo, nenhuma vogal seguida de consoante nasal é obrigatoriamente nasalizada: “c[a]ma” e “c[a]mareira”. Em vários dialetos do Nordeste do Brasil toda vogal (tônica ou pretônica) seguida de consoante nasal é obrigatoriamente nasalizada: “c[ã]ma” e “c[ã]mareira”. Quando a consoante nasal é palatal (ou glide nasal correspondente) as

vogais tônicas e pretônicas são geralmente nasalizadas na grande maioria dos dialetos do português brasileiro: “b[ã]nho” e “b[ã]nheiro” [SIL 99].

2) Nasalidade

a. cama [ɛkãμ↔]	*[ɛkαμ↔]	d. camareira [kãμãpɛIP↔]	
	~[kαμαɔPɛIP↔]		
b. sono [ɛsõvY]	*[‘s□vY]	e. soneira [sõɔvɛIP↔]	~[soɔvɛIP↔]
c. cana [ɛkãv↔]	*[‘kαv↔]	φ. canavial	
	[kãvαπiɔαω]		~[kavαπiɔαω]

Os dados em (2) mostram que uma vogal tônica deve ser obrigatoriamente nasalizada quando seguida de consoante nasal (cf. 2a – c). Quando a vogal seguida de consoante nasal ocorre em posição pré-tônica (cf. 2d – f) a nasalidade é opcional. Portanto, os exemplos em (2) mostram que a nasalidade de uma vogal seguida por consoante nasal ocorre obrigatoriamente em posição tônica e, opcionalmente em posição pré-tônica. Note que não apenas a presença da consoante nasal, mas também a posição da vogal em relação ao acento tônico influencia a modificação da vogal – que passa a ser nasalizada. A nasalidade de vogais seguidas de consoantes nasais ilustrada nos exemplos em (2) reflete um outro caso de assimilação, em que uma vogal assimila a nasalidade da consoante seguinte dependendo da posição do acento tônico da palavra [SIL 99].

Os processos de alteração segmental discutidos – vozeamento e nasalidade – ocorrem por assimilação ou ajuste fonético.

2.7.9.2 Premissa 2

Os sistemas sonoros tendem a ser foneticamente simétricos.

Assume-se que os sistemas sonoros tendem a ser simétricos. Por simetria espera-se que para cada som de uma língua seja encontrado um outro som correspondente. Assim, encontra-se um segmento “oclusivo bilabial desvozeado” [p] espera-se encontrar o seu correspondente vozeado [b]. No caso de vogais deve-se, portanto, buscar sons correspondentes que sejam anterior/posterior e arredondado/não-arredondado. Contudo, a simetria não é obrigatória, mas reflete apenas uma tendência das línguas naturais.

O sistema vocálico do português é bastante simétrico, apresentando sete vogais. Observe que para cada vogal anterior – [i, e, ɛ] – há uma vogal posterior correspondente – [u, o, □]. As vogais anteriores são não arredondadas e as vogais posteriores são arredondadas, refletindo a tendência dos sistemas vocálicos das línguas naturais.

2.7.9.3 Premissa 3

Os sons tendem a flutuar.

Para ilustrar a premissa discutem-se aspectos relacionados à articulação das consoantes oclusivas vozeadas e desvozeadas na língua krenak (falada em MG: nação Krenak) e os comparamos ao português. A discussão que se segue considera o parâmetro vozeamento/desvozeamento em termos fonêmicos. Pretende-se demonstrar que em Krenak o vozeamento de oclusivas é previsível por contexto. Assim, segmentos

oclusivos em Krenak podem variar a pronúncia entre vozeados/desvozeados sem causar prejuízo para a compreensão da língua. Já em português, o vozeamento deve ser marcado em categorias distintas: vozeado e desvozeado[SIL 99].

Em Krenak temos os segmentos oclusivos desvozeados [p,t,k] e os segmentos oclusivos vozeados [b,d,g]. Os segmentos oclusivos desvozeados [p,t,k] ocorrem em início de palavra (como em [p□k] “fechar”, [t□n] “feio” e [kr□t] “mamão”); ocorrem em final de palavra (como em [w↔p] “chorar”, [kurit] “folha” e [krak] “faca”), e ocorrem entre vogais (como em [kuparak] “onça”, [Xataran] “arara” e [Xakukan] “coruja”). Os segmentos oclusivos vozeados [b,d,g] ocorrem sempre precedidos de consoante nasal homorgânica (como em [mb□k] “peixe”, [nd↔ŋ] “torto” e [ŋgr□t] “grosso”). Observe que o vozeamento de oclusivas em Krenak é previsível pelo contexto: os segmentos oclusivos vozeados [b,d,g] ocorrem precedidos de consoante nasal homorgânica e os segmentos oclusivos desvozeados [p,t,k] ocorrem nos demais contextos[SIL 99].

O que é interessante é que falantes de Krenak podem variar o grau de vozeamento das oclusivas sem prejuízo para o sistema sonoro da língua. Quer dizer-se com isto que independente do grau de vozeamento utilizado na pronúncia de uma oclusiva – se vozeado ou desvozeado – o falante de Krenak identifica o segmento como vozeado ou desvozeado em termos fonêmicos. Para ilustrar este fato tomar-se-á como exemplo a pronúncia de uma palavra como [nd↔ŋ] “torto” que pode variar de uma forma em que a oclusiva seja completamente vozeada – [nd↔ŋ] – ou o vozeamento da oclusiva pode ser parcial – [nḍ↔ŋ] – ou o vozeamento pode não ocorrer durante a produção da oclusiva – [nt↔ŋ]. O mesmo pode ocorrer com uma forma como “feio” [t□n] em que uma oclusiva desvozeada ocorre no início da palavra. Nesta forma a oclusiva pode alternativamente ocorrer com vozeamento parcial – [ḍ□n] – ou pode ocorrer completamente vozeada – [d□n].

Em português, o vozeamento é foneticamente relevante. Temos [t] em “tato” e [d] em “dado”, que não podem ser confundidos em termos de vozeamento. O segmento [t] é desvozeado e o segmento [d] é vozeado. Isto implica que o vozeamento é distintivo em português (cf. “tato” e “dado”). Já na língua Krenak, o vozeamento é previsível por contexto, portanto não tem caráter distintivo. Uma vez que a distinção de vozeamento é foneticamente relevante em português, os falantes têm facilidade em agrupar segmentos vozeados e desvozeados em línguas em que o vozeamento não é distintivo. Em outras palavras, falantes do português são capazes de identificar os segmentos oclusivos vozeados e desvozeados em uma palavra do Krenak como “feio” [t□n]~[d□n] (pode haver dificuldade na interpretação de segmentos parcialmente vozeados em Krenak como na pronúncia – [ḍ□n]).

Enquanto na articulação de consoantes oclusivas, falantes de Krenak variam o grau de vozeamento de um modo mais vozeado até a ausência de vozeamento, os falantes do português separam as oclusivas vozeadas e desvozeadas em grupos distintos em palavras do Krenak.

Uma conseqüência da terceira premissa é que, em português, devemos empregar símbolos distintos no sistema escrito para caracterizarmos [t,d] que ocorrem foneticamente.

2.7.9.4 Premissa 4

Seqüências características de sons exercem pressão estrutural na interpretação fonêmica de segmentos suspeitos ou seqüências de segmentos suspeitos[SIL 99].

A noção de segmentos suspeitos ou seqüências de segmentos suspeitos decorre das possíveis interpretações silábicas que podem ser dadas a um segmento ou a uma seqüência de segmentos. Interpretação silábica é a análise de um segmento como consonantal ou vocálico em relação à estrutura silábica ou estrutura fonotática da língua. Ilustra-se o aspecto prático desta premissa (língua hipotética):

- | | |
|--------------------|-----------------|
| (3) a. [ma] “gato” | d. [sa] “folha” |
| b. [bo] “correr” | e. [ia] “lua” |
| c. [su] “céu” | f. [tsa] “dez” |

A primeira questão que se coloca é quanto à interpretação fonêmica da forma (3e): [ia] “lua”. O segmento inicial [i] deve ser interpretado como vogal ou consoante? Foneticamente não há dúvidas de que o segmento [i] é uma vogal, uma vez que este é um segmento produzido sem obstrução na parte central do trato vocal. A questão que se coloca aqui é quanto ao comportamento fonêmico deste segmento em relação ao sistema sonoro da língua. O segmento [i] pode ser foneticamente interpretado como vogal ou consoante. Tendo mais de uma interpretação possível, o segmento [i] passa então a ser um segmento suspeito. Vogal ou consoante? Sua interpretação na língua é dada pelo comportamento do sistema sonoro como um todo. Especificamente, a interpretação fonêmica do segmento [i] faz-se a partir da análise da estrutura silábica da língua hipotética, cujos dados são apresentados na premissa 4. Em tal língua não se observa a ocorrência de segmentos vocálicos em início de sílaba. Todas as sílabas são constituídas de seqüências de consoante-vogal (chamadas línguas CV). Portanto, para a língua hipotética ilustrada nesta última premissa a interpretação fonêmica do segmento [i] deve ser assumida como uma consoante (e não como vogal, pois esta língua não permite vogal sem uma consoante que a preceda). Temos então que a forma fonética [ia] “lua” é transcrita fonemicamente como /ya/. O símbolo /y/ indica que o segmento [i] é fonemicamente interpretado como uma consoante. Note que a interpretação de [i] como consoante segue o padrão silábico da língua.

Neste capítulo viram-se os passos necessários para chegar a geração da fala por um indivíduo. Também viram-se os princípios para a cadeia sonora da fala. E, por fim, as diferenças entre fonema e som e, por conseguinte, a classificação dos fonemas.

3 Síntese da fala

3.1 Introdução

O primeiro trabalho em síntese da fala foi feito por volta de 1779 por Christian Gottlieb Krazenstein da Academia Imperial de St. Petersburg. Krazenstein inventou um instrumento que usou uma palheta vibrante e um constante fluxo de ar, como o mecanismo de um órgão. Mais tarde em 1791, Wolfgang von Kempelen de Viena criou uma “máquina falante”. Esta máquina produziu mais que sons de vogais, chegando a produzir sentenças completas.

Trabalhos em outras “máquinas falantes” não ocorreram até 1939, quando o VODER ou Voice Operating Demonstrator foi desenvolvido para a Feira Mundial na cidade de Nova Iorque. Foi construído como um teclado de órgão, no qual ao pressionar-se as teclas geravam-se sons. Este tinha um pedal de controle da frequência operado pelo pé e 10 teclas para operar o sistema de controle das ressonâncias.

Hoje, a tecnologia de voz tem muitas aplicações, das máquinas automáticas à aeronaves automatizadas. Um sistema de fala digital toma muitas amostras da fala por segundo e digitaliza-as em séries de uns e zeros para uso do computador. A frequência de amostragem varia. Certamente, o número de vezes por segundo que esta é amostrada melhora a qualidade da mesma[MOS 90].

Os progressos recentes em síntese de fala têm produzido sintetizadores com alta inteligibilidade, mas a qualidade do som e a naturalidade ainda continuam sendo o grande problema. Entretanto, a qualidade dos produtos tem alcançado um nível adequado para várias aplicações, tais como multimídia e telecomunicações. Com alguma informação audiovisual, ou animação facial é possível aumentar a inteligibilidade da voz, consideravelmente.

Um tipo de síntese de fala é o texto para fala(TPF). A síntese texto para fala consiste de duas fases principais. A primeira é a análise do texto, onde o texto de entrada é transcrito em fonema ou outra representação lingüística, e uma segunda é a geração de formas de onda da voz, onde a saída acústica é produzida deste fonema e da informação prosódica. Estas duas fases são usualmente chamadas de síntese de alto e baixo nível, respectivamente[LEM 90].

A maneira mais simples de produzir voz sintética é reproduzir amostras pré-gravadas da voz natural, tais como sentenças ou palavras simples. Esta concatenação gera alta qualidade e naturalidade, mas tem um vocabulário limitado. O método é muito adaptável para alguns sistemas de informação. Entretanto, é bastante claro que não podemos criar um banco de dados de todas as palavras e nomes comuns no mundo. Para uma síntese de fala irrestrita nós temos de usar pedaços mais curtos do sinal de voz, tais como sílabas, fonemas, ou mesmo segmentos mais curtos.

Outro método usado largamente para produzir síntese de fala é a síntese de formantes. O método é algumas vezes chamado de analogia terminal, porque modela a fonte de som e as frequências formantes e não qualquer característica física do trato vocal. O sinal de excitação poderia ser vozeado com frequência fundamental ou ruído

não vozeado. A excitação misturada destes dois pode também ser usada para consoantes vozeadas e alguns sons de aspiração.

Há ainda um método para gerar voz artificial que modela diretamente o sistema de produção da fala humana, chamado de síntese articulatória, o qual tipicamente envolve modelar os articuladores humanos e as cordas vocais. Os articuladores são usualmente modelados com um conjunto de funções de área de seções pequenas. O modelo de corda vocal é usado para gerar um sinal de excitação apropriado.

Todos os métodos de síntese tem alguns benefícios e problemas e é bastante difícil dizer qual método é o melhor. Com os métodos concatenativo e por formantes, muitos resultados promissores foram alcançados, mas também a síntese articulatória pode surgir como um método potencial no futuro[LEM 90].

3.2 Métodos de síntese

3.2.1 Síntese Articulatória

A modelagem do movimento articulatório evita algumas das complexidades da fonética-acústica: a acústica para vários sons é uma função complicada da informação fonêmica, taxa de fala, prosódia, etc. Efeitos de coarticulação devem ser especificados por regras nos sistemas *terminal-analog*, mas segue mais diretamente em síntese articulatória. A síntese articulatória pode ser dividida em três classes, distinguidas pelo nível em que elas recebem a entrada no processo de produção da fala: comando neuromotor, articulador e a forma do trato vocal.

Para falar, uma pessoa pensa uma mensagem e envia comandos para os músculos do trato vocal, que faz os articuladores moverem-se, mudando a forma do trato vocal. A mais básica síntese articulatória transforma uma sequência de fonemas de entrada em um conjunto de comandos musculares neuromotores, seguindo um modelo baseado nos dados de estudos eletromiográficos. Este também necessita de um modelo relatando os comandos musculares para o movimento do articulador e um algoritmo para converter o conjunto de posições do articulador dentro das formas do trato vocal ou funções de área. Uma sequência de funções de área no tempo $A_m(n)$ poderia especificar a variação do tempo no filtro de rede de quem os coeficientes de reflexão são determinados pelos raios das áreas adjacentes.

Enquanto a síntese neuromotora é mais agradável teoricamente, na qual os fonemas relacionam-se diretamente com os comandos dos músculos, é também mais complexa de ser implementada. Comandos para o gesto da fala tendem a ser similares aos diferentes fonemas que usam o mesmo gesto, mas os relacionamentos entre comandos e gestos são complexos. Um sintetizador que mapeia diretamente os fonemas para posições articulatórias e movimentos evita este passo, mas deve explicitamente considerar o comportamento do trato vocal que pode mais facilmente ser descrito no nível muscular. Desde que os dados experimentais são mais vastamente disponíveis para o movimento articulatório do que para os comandos musculares, tem havido mais tentativas de síntese articulatória do que síntese neuromotora. Sistemas típicos assumem que 7 a 11 parâmetros podem adequadamente descrever o comportamento articulatório(figura 3.1).

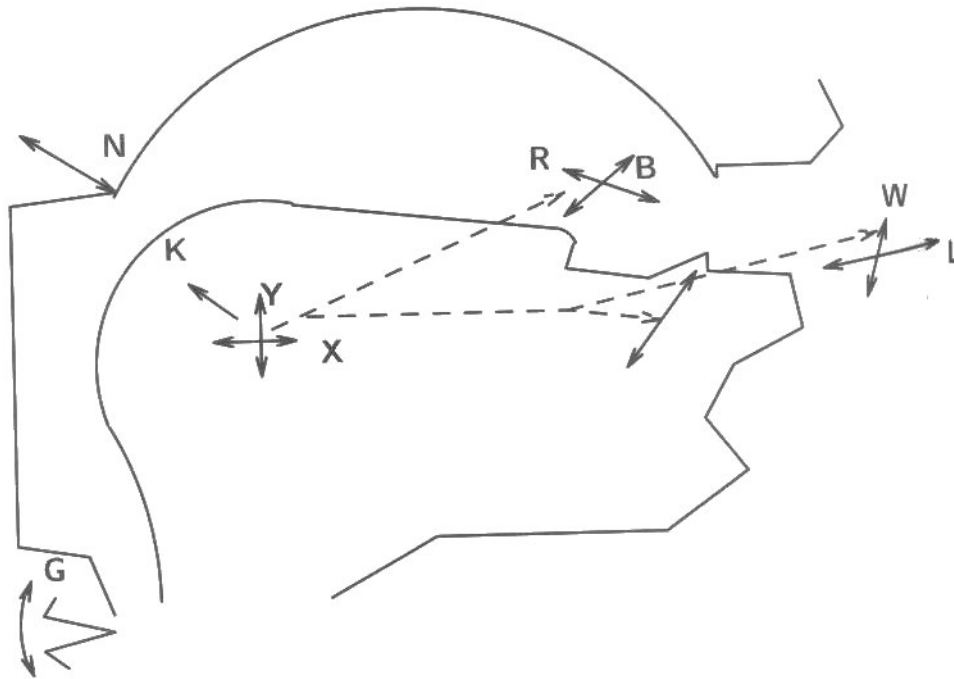


FIGURA 3.1 – Modelo articulatório, com flechas rotuladas indicando grau de liberdade e linhas pontilhadas mostrando os relacionamentos entre os parâmetros.[OSH 2000]

Valores destinados para estes parâmetros são armazenados para cada fonema, e a sequência de tempo dos valores dos parâmetros é obtida pela interpolação entre os destinos, usando limitações de coarticulação e diferentes constantes de tempo para os vários articuladores.

O modelo para o movimento articulador usualmente é derivado do raio X do trato vocal durante locuções simples. E a forma do trato vocal é apenas bruscamente aproximada pelo modelo simplificado usando 7 a 11 parâmetros. Fala sintética da síntese articulatória, baseada em generalizações de dados limitados, produz fala de mais baixa qualidade do que a síntese *terminal-analog*.

Uma vantagem da síntese articulatória é que os modelos da área do trato vocal permitem exata modelagem dos transientes devido as mudanças repentinas da área, tão bem quanto automaticamente gerando excitação turbulenta para as constrições estreitas[OSH 2000].

3.2.2 Síntese de formantes

Muitos sintetizadores empregam síntese de formantes, usando um modelo similar ao que é mostrado na figura 3.2. O sistema da figura 3.3 mostra uma implementação específica que forma a base de dois sintetizadores avançados, estrutura cascata e paralela. A excitação é geralmente um trem periódico de impulsos (para fala vozeada), ruído pseudo-randômico (para fala não vozeada), ou ruído periódico (para fricativos vozeados). O trato vocal é usualmente modelado como uma cascata de ressonadores digitais de segunda ordem, cada um representando ou uma formante ou

uma forma espectral da fonte de excitação. A estrutura do filtro cascata aproxima-se do espectro da fala para vogais e pode ser controlada com um parâmetro amplitude[OSH 2000].

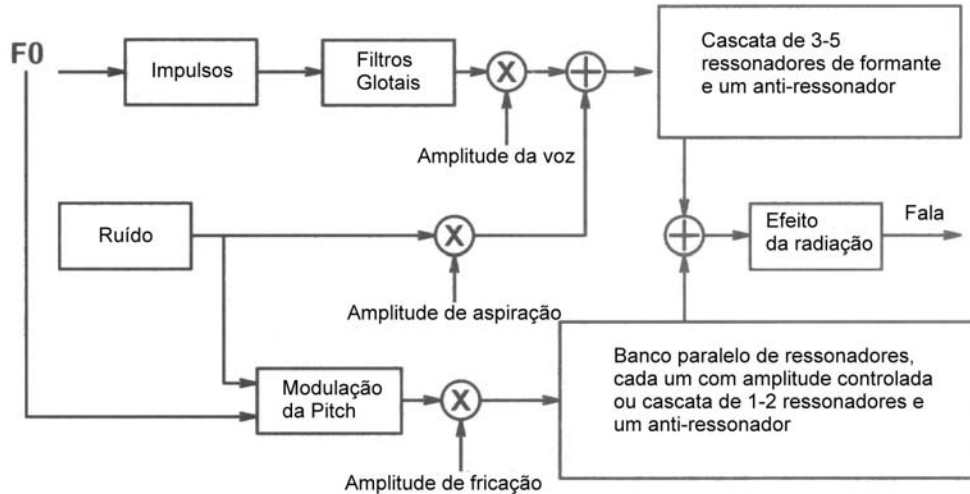


FIGURA 3.2 - Diagrama de bloco simplificado de um sintetizador de formantes [OSH 2000]

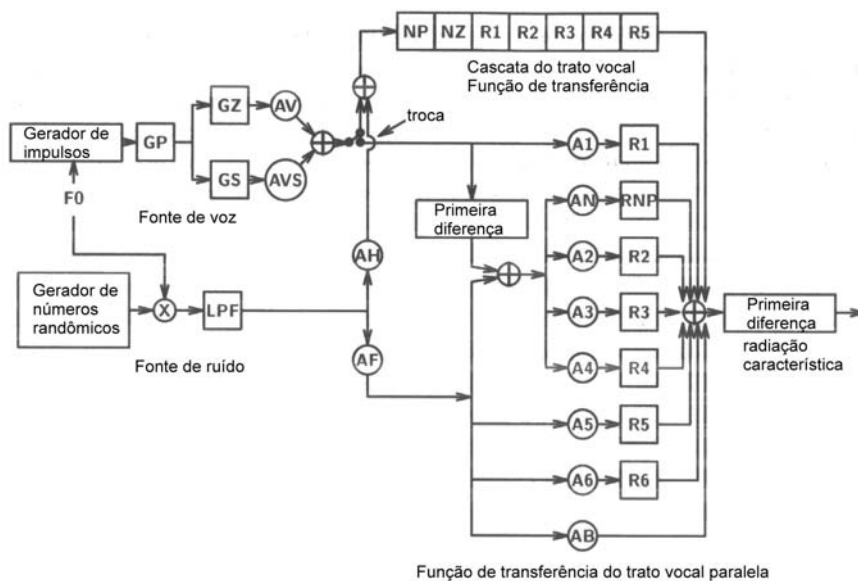


FIGURA 3.3 - Diagrama de bloco de um sintetizador de formantes em cascata-paralelo [OSH 2000]

3.2.3 Síntese Código preditivo linear(LPC)

Para evitar a necessidade de manualmente desenvolver o conjunto complexo de regras de coarticulação para a síntese de formantes, o método automático de síntese LPC tornou-se popular. Este tem uma estrutura mais simples do que a síntese de formantes, porque todas as propriedades espectrais da fala (exceto intensidade e periodicidade) estão incluídas nos coeficientes LPC, que são calculados

automaticamente na análise da fala natural. O Filtro de rede é freqüentemente usado para síntese porque os coeficientes do filtro podem ser linearmente interpolados entre amostras sem produzir um filtro instável. A excitação é especificada pelo parâmetro ganho, um bit vozeado-não vozeado, e (se vozeado) um valor F0.

LPC tem um mais simples e completo procedimento de análise automática e uma estrutura de síntese mais simples do que a de formantes. Por outro lado, o entendimento do comportamento da formante facilita a interpolação do parâmetro nos segmentos limites da síntese de formantes, enquanto que a interpolação dos coeficientes de reflexão LPC é usualmente limitado, porque cada coeficiente afeta um vasto intervalo de freqüências da fala em uma forma complexa. Em acréscimo, síntese de formantes é mais flexível em permitindo transformações simples para simular várias vozes. Porque os vales espectrais são modelados mais pobremente do que os picos na análise LPC, as larguras de banda são freqüentemente imprópriamente estimadas, enquanto que determinação da largura de banda direta do espectro DFT na aproximação formante pode produzir melhores estimativas[OSH 2000].

3.2.4 Concatenação da forma de onda

Muitos sintetizadores têm adotado a aproximação espectral, onde as unidades são descritas em termos de formantes, amplitudes Fourier ou coeficientes LPC, devido ao mais eficiente código de fala e a flexibilidade para modificar as unidades armazenadas abaixo de mudanças na prosódia ou locutor. Como uma alternativa para a síntese baseada em espectro, a síntese da forma de onda pode produzir muito boa fala sintética, ao custo do aumento da memória e a inflexibilidade do locutor.

Muito boa qualidade de síntese tem sido obtida com o método de concatenação chamado de PSOLA(*pitch-synchronous overlap-and-add*): este sobrepõe e soma pequenas unidades de forma de onda, usando tipicamente unidades de dois períodos de *pitch*(freqüência fundamental) adjacentes. Como outra forma de onda ou método de síntese difone, PSOLA não pode ser facilmente modificado para simular outras vozes além da do locutor treinado.

3.2.5 Síntese TPF

A última meta de um sintetizador texto para fala é ler qualquer texto, introduzido por um operador ou escaneado e submetido a um sistema de reconhecimento de caracteres. A leitura deveria ser inteligível e natural. Sistemas de resposta por voz produzem fala artificial, simplesmente concatenando palavras isoladas ou partes das sentenças. Eles são, entretanto, aplicáveis apenas quando um vocabulário limitado é requerido (não mais do que 100 palavras, em geral) e quando as sentenças a serem pronunciadas compartilham uma estrutura muito restrita. No contexto da síntese texto para fala, é impossível gravar e armazenar todas as palavras da linguagem em uso. É mais conveniente definir texto para fala como a produção de fala pelas máquinas, pela maneira da fonetização automática das sentenças. A figura 3.5 introduz o diagrama de um sintetizador TPF geral.



FIGURA 3.4 - Diagrama funcional geral de um sistema TPF [DUT 97]

Como em uma leitura humana, texto para fala compreende um módulo de processamento de linguagem natural (PLN), capaz de produzir uma transcrição fonética do texto a ser lido, junto com a entonação e ritmo desejado, e um módulo de processamento digital de sinais, que transforma a informação simbólica recebida desse em fala de som natural. O formalismo particular escolhido para intermediar a informação varia de um sintetizador para outro.

É interessante fazer uma distinção entre características segmentais e suprasegmentais de um sistema TPF. Denominar-se-á como qualidade segmental a eficiência da máquina para produzir sons de fala natural, assumindo que o módulo PLN atualmente distribui informação de alta qualidade. Por outro lado, nós referiremos qualidade suprasegmental como a riqueza do contorno prosódico que este é capaz de explorar. Este é principalmente a responsabilidade do módulo PLN. Naturalidade e inteligibilidade, propriedades fundamentais em síntese TPF[DUT 97].

3.2.5.1 Do texto para a transcrição fonética

Nós introduzimos na figura 3.4 o diagrama geral de um sistema de síntese TPF. Entrando em um nível mais profundo na descrição ver-se-á o módulo PLN. De fato, desde que o sintetizador necessita conhecer a transcrição fonética das palavras expressas nós podemos facilmente imaginar que o módulo PLN conterá dois sub-blocos: o módulo letra para som (*LTS*), e um gerador de prosódia (*GP*), como mostrado na figura 3.5.

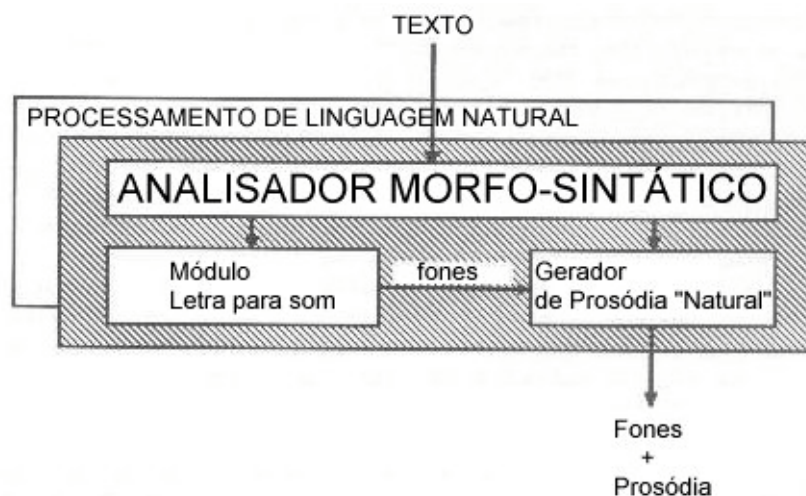


FIGURA 3.5 - Esqueleto funcional de um módulo PLN [DUT 97]

O que é menos óbvio em uma primeira visão, é a necessidade de um poderoso analisador morfo-sintático. De fato, quando questionado na habilidade de leitura deles, muitos não-especialistas acreditam que os sons que eles expressam um a um relacionam-se com as letras que eles originam e as entonações que eles produzem não tem outra origem do que a pontuação da sentença que eles lêem. Na prática, nenhum módulo *LTS* nem *PG* tem uma chance real de sucesso se eles não estão adequadamente equipados com informação de alto nível, tal como a parte da fala de palavras isoladas e a dependência de relacionamento entre as palavras conectadas para o bloco *LTS*, e algumas formas de descrição hierárquica das sentenças no caso do *PG*[DUT 97].

3.2.5.1.1 Arquiteturas PLN para síntese TPF

A análise do texto e, melhor ainda, entendimento do texto são alguns dos maiores desafios encontrados na inteligência artificial. Nos últimos 30 anos, os especialistas em lingüística computacional propuseram um número impressionante de formalismos lingüísticos e métodos de inferência para cuidar destes problemas.

Os sistemas *TPF* a priori têm de lidar com qualquer texto, de simples sentenças isoladas para parágrafos complexos, com possivelmente palavras incorporadas, algumas delas com discursiva, se não estruturas não-gramaticais: qualquer que sejam os dados fornecidos na entrada de um sistema *TPF*, a fala deve ser gerada! Agora qualquer um que tenha estado alguma vez envolvido em *PLN* sabe que o tempo requerido para extrair algumas informações lingüisticamente relevantes de um texto, rapidamente vai ao infinito como a exatidão da informação esperada aumenta. Pior ainda, obter informações elementares torna-se questionável quando o texto a ser examinado torna-se mais elaborado. Isto origina-se tipicamente do fato que a linguagem natural constantemente desrespeita regras, em qualquer nível do processo de análise, tal que levemente refinando uma dada análise ou aumentando a cobertura de uma dada regra freqüentemente implica aumentar a complexidade do sistema de análise em uma maneira que é difícil justificar pelo refinamento que este fornece.

Nesta consideração, algumas análises heurísticas têm sucesso em abastecer informação com uma razoável cobertura, com baixo custo computacional e com um tamanho de léxico reduzido. Todavia, em muitos casos tais como aproximações orientadas para otimização para problemas lingüísticos têm também a descrever regras em termos procedurais (especificando ações a serem realizadas nos dados para a regra a ser aplicada) mais que nas declarativas (especificando condições nos dados, a verificação de quais são deixados para uma máquina de inferência geral). Isto tem resultado em *parsers hard-coded*, onde a reusabilidade e a extensibilidade são intrinsicamente limitadas. Em uma perspectiva a longo período, uma clara separação entre as regras de uma gramática e a maneira que elas são tratadas parece sem escapatória. Da mesma forma, os formalismos dos dados deveriam ser mantidos independente das especificações das regras[DUT 97].

3.2.5.1.2 Formalismos dos dados

Muitos sistemas *TPF* anteriores faziam uso do intercâmbio linear de estruturas entre módulos de processamento organizados sequencialmente. A informação processada por cada unidade foi unida com a cadeia de caracteres de entrada, e fornecida como uma cadeia de entrada da próxima unidade(figura 3.6). Como resultado, representações intermediárias tornaram-se mais complexas.

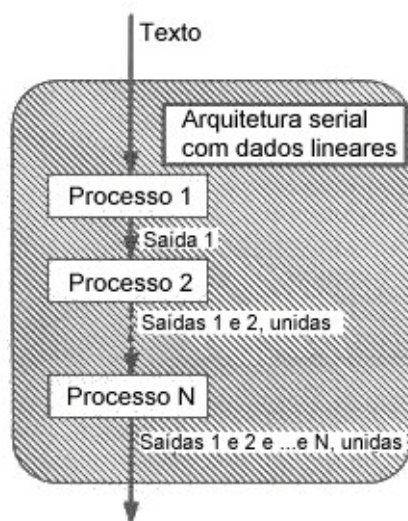


FIGURA 3.6 - Uma velha estratégia de processamento de linguagem natural para síntese TPF: estruturas de troca linear entre módulos de processamento organizados seqüencialmente [DUT 97]

Em contraste, dadas as várias conexões entre os muitos níveis de descrição de uma linguagem (fonética, fonologia, morfologia, sintaxe, semântica, e pragmática), e visto que o caminho deles pode ser relacionado um ao outro é raramente conhecido em avanço, é comum na prática organizar os dados mantidos por um sistema de processamento de linguagem natural em um sistema de análise que permite relacionamentos descritos em muitas maneiras - isto é, com muitos tipos de gramáticas quanto possível - e mais tarde ser refinado por especificando dependências adicionais e possivelmente níveis adicionais. Quando tentando catalogar recentes sistemas *TPF* com respeito a organização dos dados internos deles, duas tendências emergem. Alguns beneficiam-se das poderosas estruturas de descrição de características mais e mais usadas pelos lingüistas computacionais - *feature structures (FSs)* (figura 3.7), em que cada nível é associado a uma categoria *PATR*.

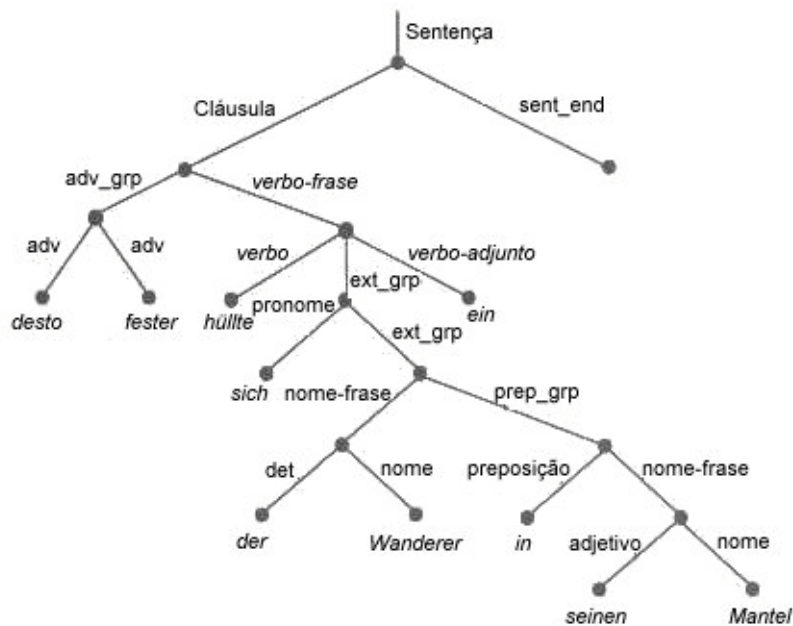


FIGURA 3.7 - Um exemplo de estrutura característica quando o módulo de análise de sintaxe operou em um fragmento de sentença alemã *desto fester hüllte sich der Wanderer in seinen Mantel ein* [DUT 97]

As vantagens dos *MLDSs* e *FSs* sobre uma organização linear da informação são numerosas:

- naturalmente aumentam a legibilidade, relativamente ambos os dados e regras.
- A extensibilidade é muito melhorada.
- O debug é mais fácil.
- Desde que dados são independentes de formalismos de regras, portabilidade interlinguagem é muito melhor assegurada.
 - É possível trocar uma dada implementação de um bloco de processamento por outra implementação da mesma funcionalidade, mas usando outra regra, outro algoritmo, outra linguagem de programação, e assim por diante.

Na realidade, a estratégia completa de análise do texto pode ainda ser escolhida como serial, hierárquica, ou heterárquica (figura 3.8) [DUT 97].

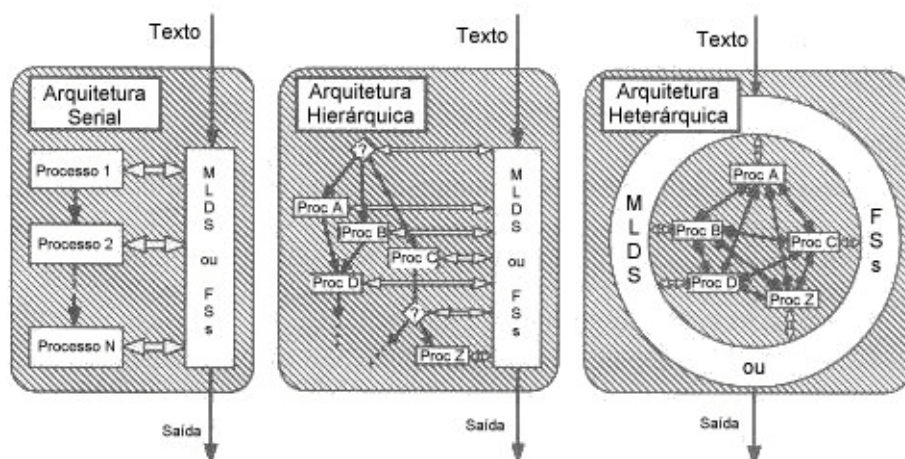


FIGURA 3.8 - MLDSs e FSSs teoricamente permitem sincronização serial, hierárquica ou heterárquica [DUT 97]

3.2.5.1.3 Formalismos das regras

A figura 3.9 mostra um típico módulo de processamento de linguagem dos sistemas *TPF* atuais. Este engloba um analisador de texto, dirigido em abastecer informação sintática útil para os subseqüentes módulos letra para som e geração de prosódia.

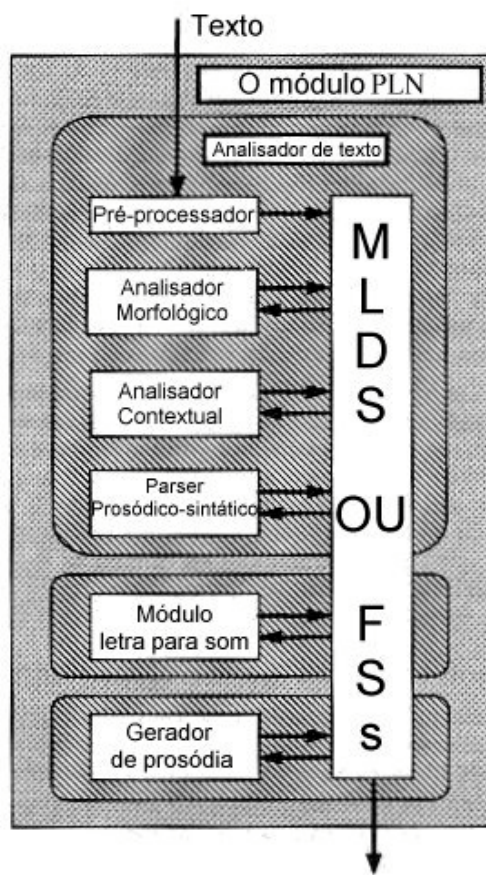


FIGURA 3.9 - O módulo PLN de um típico sistema de conversão TPF [DUT 97]

A questão dos formalismos das regras é central no contexto do processamento de linguagem para síntese *TPF*. De maneira rigorosa, eles irremediavelmente confinam o conjunto de eventos lingüísticos que eles podem descrever, ou ao menos eles constroem os escritores das regras a usar um estratagema para alcançar o final deles. Um formalismo deveria ser adaptado a descrever um conjunto de fatos.

Agora, freqüentemente acontece que uma dada tarefa de processamento da linguagem natural pode ser considerada para um número limitado de regras simples, que aplicam muito do tempo, mas não todo o tempo: uma quantia considerável ou regras altamente confusas são necessárias, a aplicabilidade das quais é infelizmente muito mais restrita. Para que o conjunto de regras heterogêneas possa ser mantido de maneira que as regras mais simples não sejam penalizadas pelas regras mais complexas devem-se adaptar os formalismos das regras para a complexidade de cada sub-problema. Uma vantagem adicional dos *MLDSs* e dos *FSs* é que eles são suficientemente gerais para aceitar muitos tipos de processamento da informação[DUT 97].

3.2.5.2 Análise morfossintática

Ser capaz de reduzir uma dada sentença em algo como uma seqüência de partes da fala, e além disso, descrevê-la na forma de uma árvore de sintaxe, que descobre a estrutura interna dela, é requerida por ao menos duas razões:

- Transcrição fonética precisa pode ser alcançada apenas se parte da categoria da fala de algumas palavras está disponível e pode depender do conhecimento do relacionamento de dependência entre palavras sucessivas.
- Prosódia natural depende pesadamente da sintaxe. Isto também, obviamente tem muito a fazer com a semântica e a pragmática, mas desde que muito poucos dados estão correntemente disponíveis nos aspectos produtivos desta dependência, sistemas *TPF* meramente concentram-se na sintaxe.

No contexto da síntese *TPF*, sintaxe deveria ser entendida como a organização de sentenças na forma de relações de dependência entre palavras, examinadas através da parte da fala. Isto depende de quais categorias da parte da fala serão consideradas, assim como o que significa palavra. De fato, no caso extremo de uma parte da fala por palavra, nossa definição de sintaxe permanece sem sentido. Observe que tal partição está longe de ser irreal.

O analisador morfossintático de vários sistemas *TPF* recentes é mostrado na figura 3.10. Este compreende:

- Um módulo de pré-processamento, cujo principal propósito é servir como uma interface entre a entrada de texto e a estrutura de dados interna do analisador, tipicamente uma estrutura de dados multinível (*MLDS*) ou estruturas de características (*FSs*). Todas as palavras na sentença corrente, incluindo números, abreviações, acrônimos, expressões idiomáticas, etc, são identificadas e possivelmente transformadas em texto.
- Um módulo de análise morfológica, cuja tarefa é propor todas as possíveis partes das categorias da fala para cada palavra individualmente.
- Um módulo de análise contextual, que considera uma palavra neste contexto e então torna possível reduzir a lista destas possíveis partes das categorias da fala para

um número restrito de hipóteses possíveis, dadas as correspondentes partes da fala de palavras vizinhas.

- Finalmente, um *parser* sintático-prosódico, que examina o espaço de busca restante e descobre a estrutura do texto que mais relaciona-se com a esperada realização prosódica.

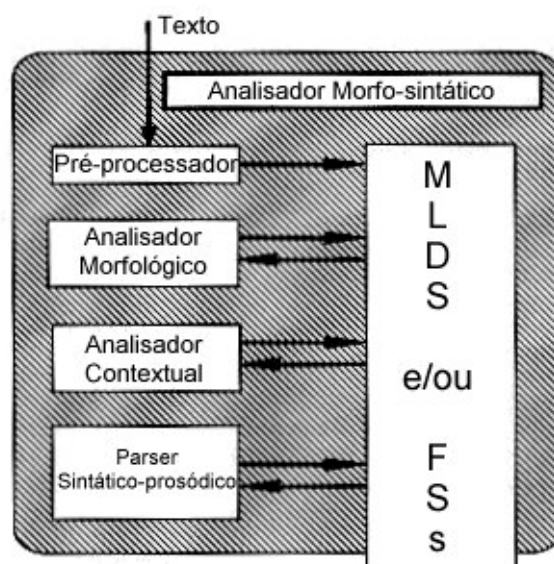


FIGURA 3.10 - Típico analisador morfossintático [DUT 97]

3.2.5.3 Geração automática da prosódia

A impressão de naturalidade que emana de um sistema *TPF* é uma função da riqueza do contorno melódico e da qualidade dos padrões rítmicos que este é capaz de aplicar a fala que produz. O gerador de prosódia é responsável por estes dois aspectos[DUT 97].

3.2.5.3.1 Prosódia

O termo prosódia refere-se a certas propriedades do sinal de fala, tais como mudanças audíveis na *pitch* (frequência fundamental), ruído(loudness) e comprimento da sílaba.

Pelo fato de os eventos prosódicos mostrarem-se alinhados no tempo com as sílabas ou grupo de sílabas, mais do que com segmentos (sons, fonemas), eles são também referidos como suprasegmentais.

O termo entonação é usado por alguns como sinônimo para prosódia. Este é restrito aos aspectos melódicos da prosódia.

3.2.5.3.2 Níveis de representação do fenômeno prosódico

Como para outras propriedades do sinal de fala, eventos prosódicos podem ser estudados em vários níveis de representação (ver tabela 3.1):

- Primeiro, o nível acústico: a manifestação acústica da prosódia (frequência fundamental, amplitude e duração) pode ser medida diretamente, usando hardware especializado ou algoritmos (tais como algoritmos de determinação da *pitch*).
- Segundo, o nível de percepção representa os eventos prosódicos como ouvido pelos ouvintes. Como para propriedades espectrais dos sons da fala, as características acústicas que podem ser medidas não são sempre perceptíveis.
- Finalmente, o nível lingüístico representa a prosódia de uma locução como uma seqüência de unidades abstratas (sinais, símbolos), algumas das quais tem uma função comunicativa na fala, enquanto outras podem apenas representar requerimentos sintáticos. A estrutura lingüística da prosódia não é algum código escondido que simplesmente pode ser revelado usando alguns procedimentos padrões. O modelo lingüístico é uma interpretação estrutural dos dados, que resultam da análise distribucional ou funcional dos dados prosódicos por um lingüista.

TABELA 3.1 – Níveis de representação

Acústico	Perceptivo	Lingüístico
Frequência Fundamental(F0)	<i>Pitch</i>	Tom, entonação, aspectos de ênfase
Amplitude, energia, intensidade	Volume	Aspectos de ênfase
Duração	Comprimento	Aspectos de ênfase
Amplitude dinâmica	força	Aspectos de ênfase

Fonte: [DUT 97]

3.2.5.3.3 Principais componentes da prosódia

Ênfase é uma propriedade prosódica que tem sido descrita desde o primeiro trabalho em prosódia na fonética. Esta está relacionada com o ruído e a força fonatória. Ambas as caracterizações referem-se à forma perceptual da prosódia: a sílaba carregando ênfase é proeminente com respeito às sílabas que cercam, ou devido ao ruído, ou as propriedades dinâmicas.

O comprimento silábico possui uma função na indicação dos limites e da ênfase. Embora a duração absoluta de um evento seja facilmente medida, este é apenas um aspecto trivial da medida do comprimento. Primeiro de tudo, não é óbvio que os eventos deveriam ser medidos: sílabas inteiras, partes vozeadas, núcleo silábico, e assim por diante. Segundo, a duração das sílabas e sons da fala depende de vários fatores tais como: taxa de fala, ritmo, natureza fonética, etc.

Os parâmetros acústicos relacionados à prosódia (principalmente F0, intensidade e duração) exibem variações de curto prazo que podem ser medidas, mas que não podem ser percebidas e claramente não tem função na entonação. Estas variações são referidas como microprosódia.

3.2.5.3.4 Os propósitos da prosódia

As características prosódicas têm funções específicas na comunicação de fala. O efeito da prosódia mais visível é o foco. Por exemplo, certos eventos da *pitch* fazem

uma sílaba sobressair dentro de uma locução, e indiretamente a palavra ou grupo sintático a que este pertence será destacado como um importante ou novo componente no significado desta locução. A presença de uma marca de foco pode ter vários efeitos, tais como contraste, dependendo do lugar onde este ocorre, ou o contexto semântico da locução.

As características prosódicas criam uma segmentação da fala que encadeia em grupos de sílabas, ou elas dão surgimento a um grupo de sílabas e palavras em maiores pedaços.

Além disso, há características prosódicas que indicam relacionamentos entre tais grupos, indicando que dois ou mais grupos de sílabas são ligadas de alguma maneira. Este efeito de agrupamento é hierárquico, embora não necessariamente idêntico à estrutura sintática da expressão[DUT 97].

3.2.5.3.5 Relacionamentos entre prosódia e outros aspectos da fala

Depois de ter estabelecido os modelos acústicos, perceptuais e lingüísticos da prosódia, é importante examinar a interação com o léxico, com a sintaxe e com a semântica e a pragmática, desde que a tarefa principal do componente de gerador prosódico de um sistema *TPF* seja derivada de uma representação acústica da prosódia destes níveis lingüísticos.

Este relacionamento é mais complexo do que parece em uma primeira vista. De fato, é freqüentemente injusto assumir que a entonação que nós aplicamos nas sentenças que nós lemos é apenas uma função da pontuação delas. Marcas de pontuação, entretanto, são meramente o sinal visível da estruturação forte de uma linguagem, para as quais entonação age como o eco, o traço acústico que deveria ajudar o ouvinte a perceber tudo isto com sutileza. Prosódia, cria demanda em cada e todo nível de competência lingüística do leitor[DUT 97].

3.2.5.3.6 *Parsing* prosódico-sintático

Os sistemas *TPF* estão condenados a uma mera entonação como robô até que um brilhante lingüista computacional anuncie um funcional analisador pragmático-semântico para texto irrestrito. Há várias razões para não pensar assim, desde que aceita uma importante restrição na naturalidade da voz sintética – que a entonação é mantida aceitavelmente neutra: entonação aceitável deve ser plausível, mas não necessita ser a mais apropriada entonação para uma expressão particular: nenhuma suposição de entendimento ou geração pela máquina necessita ser feita. Entonação neutra não expressa rara ênfase, ênfase contrastante ou efeitos estilizados: é a entonação *default* que pode ser usada para uma expressão do contexto. Esta aproximação remove a necessidade para referenciar o contexto ou conhecimento mundial enquanto retendo ambiciosas metas lingüísticas.

A idéia chave é que a estrutura sintática correta, a que precisamente requer algumas percepções semânticas e pragmáticas, não é essencial para produzir tal prosódia. Os leitores humanos basicamente variam a prosódia por duas razões:

1. Particionar a expressão em partes menores.
2. Assinalar certos significados, ou por usando ênfase comparativa para colocar o foco na informação que está no centro do interesse do leitor, ou para enfatizar

palavras que servem novas informações, prosodicamente expressando alguns relacionamentos particulares entre palavras, ou simplesmente usando altos ou baixos tons para indicar se um grupo de entonação é suposto para ser seguido por outro ou não.

A segunda é a mais complexa. Em contraste, no caso simples de um sistema *TPF* que tentaria alcançar a meta 1, através disso concentraria seus esforços para a produção da prosódia neutra, e é largamente aceito que a organização de segmentos de texto curtos que ajem como grupos prosódicos (também chamados de domínios prosódicos) é muito mais simples do que as relacionadas dependências sintáticas. Em outras palavras, a estrutura prosódica de uma sentença é mais favorecida do que a sintática.

3.2.5.3.7 Tarefa de ênfase sentencial

Embora ênfase seja uma noção comum, é bastante difícil caracterizá-la em termos acústicos. Afirma-se que as sílabas enfatizadas são de forma perceptiva relacionadas a algumas proeminências, no senso que elas são tipicamente produzidas mais ruidosas ou mais longas do que as não acentuadas, e que elas são menos receptivas a redução.

Ênfase é de fato um assunto para vários fatores lingüísticos, que contribuem para classificar em duas categorias amplas – léxica e sentencial. A ênfase léxica refere-se à posição de sílabas proeminentes em palavras faladas em isolamento. Entretanto, quando as palavras são agrupadas em sentenças, elas não necessariamente mantêm a ênfase léxica delas. Também, acontece que as sílabas não estressadas lexicalmente tornam-se estressadas.

Freqüentemente distingui-se ênfase gramatical de ênfase enfática e contrastante. O anterior é diretamente relacionado com a estrutura prosódico-sintática da sentença. Em contraste, ênfase enfática e contrastante origina-se em considerações semânticas e pragmáticas. Elas permitem a um locutor atrair atenção especial em algumas palavras para enfatizá-las ou contrastá-las.

Ênfase é também uma característica dependente da linguagem da fala. A tarefa de ênfase léxica, por exemplo, é um problema importante para as chamadas linguagens de ênfase livre, enquanto esta é apenas uma menor preocupação para as chamadas linguagens de ênfase fixada em que a ênfase léxica é determinada em uma posição silábica fixa. Linguagens de ênfase livre permitem à ênfase estar a priori determinada para várias posições silábicas, embora uma dada palavra é geralmente associada com um padrão de ênfase único, ocasionalmente dois.

A tarefa da ênfase em palavras de entrada é de primeira importância em um sistema *TPF*, por muitas razões:

- Um incorreto padrão de ênfase é desordenado em si mesmo. Isto faz um sistema *TPF* possuir pouca naturalidade.
- Em um número de sistemas *TPF*, ênfase é usada como uma informação contextual para *MLRRs*. As qualidades das vogais possuem características dependentes da ênfase. Falhar em determinar os corretos padrões de ênfase então leva a erros fonéticos, os quais implicam em uma perda da inteligibilidade.

- Acentos prosódicos em fala contínua devidamente coincidem com a característica da frequência fundamental relacionada à sílaba de ênfase primária das palavras.

Como pode ser esperado, sistemas *TPF* mais frequentemente focam na ênfase léxica e gramatical, deixando a ênfase enfática e contrastante para considerações posteriores desde que este é duramente mais previsível na base da informação léxica, morfológica ou sintática. Modernos sistemas *TPF* fazem uso de muitas fontes de conhecimentos simultaneamente:

- Silabificação: sílabas são marcadas como pesadas ou leves como uma função dos conteúdos fonéticos delas e o que cerca (tal como comprimento da vogal e o número de consoantes pós-vocálicas), e a missão de ênfase é submetida a limitações contextuais.

- A informação de parte da fala é usada não apenas para desambiguar homógrafos heterofônicos, mas de um modo geral como uma dica para a tarefa de ênfase.

- Morfologia pode também contribuir, por descrevendo a ênfase em termos de raízes enfatizadas que combinadas com afixos e/ou outras raízes enfatizadas, em combinação com regras que consideram deslocamentos de ênfase no caso de inflexão e derivação, e para ordem da ênfase (primária/secundária) em compostos.

- Finalmente, etimologia pode ser feita responsável por muitas irregularidades na tarefa de ênfase, especialmente em nomes próprios[DUT 97].

3.2.5.3.8 Da representação simbólica para a acústica da prosódia

Assumindo que se tenha adequadamente analisado uma sentença e derivado em uma organização prosódica-sintática, na forma de grupos relacionados hierarquicamente, de sílabas marcadas com vários níveis de ênfase, o último passo para fornecer a expressão com uma prosódia consistente está situado em produzir uma representação acústica da entonação e da duração dos fonemas e pausas[DUT 97].

3.2.5.3.8.1 Gerando tempo

Fala exhibe padrões temporais, que devem ser imitados pelos sistemas *TPF* em ordem para soar natural. Mais precisamente, sistemas *TPF* deveriam ser capazes de produzir durações dos segmentos e pausas que não difiram do natural.

Derivar durações naturais do texto basicamente implica em responder a duas questões:

- Que tipo de unidades de destino serão consideradas
- Como deveriam ser estabelecidos os relacionamentos entre informação prosódica-sintática e estas unidades

3.2.5.3.8.2 Geração da frequência fundamental

Como a curva da *F0* (frequência fundamental) é gerada, depende do tipo de modelo prosódico escolhido para a representação. De fato, ligar lacuna entre uma descrição prosódica-sintática da locução em termos de grupos relacionados hierarquicamente de sílabas e a melodia atual envolve dois passos:

1. Estabelecer o relacionamento entre os parâmetros do modelo e a descrição prosódica-sintática
2. Gerar a curva melódica final da representação paramétrica no modelo.

Claramente, o passo dois reduz o mais trivial processamento matemático no caso dos modelos acústico e perceptivo, que são intrinsicamente mais fechados para os dados atuais do que os lingüísticos. Este, entretanto, é balanceado pela complexidade do passo um. Em contraste, os modelos lingüísticos são mais fechados para a descrição prosódica-lingüística e então facilitam o passo um, enquanto não aumenta a complexidade do passo dois.

3.2.5.4 Da transcrição fonética para a fala

Uma vez o módulo PLN de um sistema *TPF* tenha completado sua tarefa, a máquina de leitura está quase no mesmo estado como uma pessoa que sabe o que vai dizer, mas ainda fica pensando como dizer[DUT 97].

3.2.5.4.1 Estratégias de síntese

O diagrama de bloco introduzido na figura 3.5 e a definição de uma interface geral entre o PLN e os blocos PSD apresenta a vantagem de permitir um estudo separado de ambos os processos, se para as necessidades deles, para as funcionalidades deles, para a descrição da arquitetura deles, ou para o resultado deles. Conseqüentemente, nós agora assumiremos que a informação de qualidade é entregue na entrada do módulo de processamento do sinal digital, como o que seria copiado de um leitor humano.

Intuitivamente, as operações envolvidas são a analogia do computador de dinamicamente controlar os músculos articulatórios e a freqüência de vibração das cordas vocais tal que o sinal de saída adapta-se as necessidades de entrada. Para fazer isso apropriadamente, o módulo PSD deveria obviamente, de alguma maneira, levar as limitações articulatórias em conta. Este pode ser basicamente alcançado de duas maneiras:

- Explicitamente, na forma de uma série de regras que formalmente descrevem a influência dos fonemas uns com os outros.
- Implicitamente, armazenando exemplos de transições fonéticas e coarticulações em um banco de dados de segmentos da fala, e usando-os apenas como eles são, como últimas unidades acústicas.

Duas classes principais de sistemas *TPF* têm surgido destas duas alternativas, que rapidamente transformam-se em filosofias de síntese dada a divergência presente nos significados delas e objetivos: síntese por regra e síntese por concatenação[DUT 97].

3.2.5.4.1.1 Sintetizadores baseados em regras

Sintetizadores baseados em regras (Figura 3.11) são na maioria das vezes em favor com os fonéticos e fonologistas, como eles constituem um cognitivo, aproximação produtiva do mecanismo de fonação. O uso freqüente do sintetizador de Klatt, por exemplo, é principalmente devido à assistência inestimável no estudo das características da fala natural, pela audição analítica da fala sintetizada por regras. Além disso, a existência de relacionamentos entre parâmetros articulatórios e as entradas do modelo de Klatt fazem deste uma ferramenta prática para investigar coações fisiológicas.

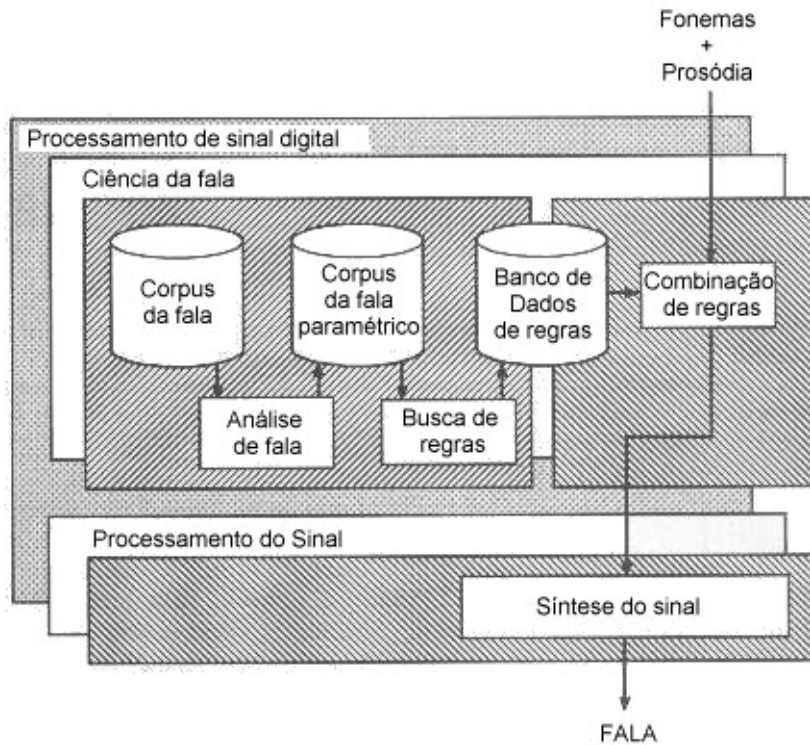


FIGURA 3.11 - Um típico sintetizador baseado em regras [DUT 97]

Em uma fase preliminar, um grande número de palavras, geralmente seqüências consoante-vogal-consoante (CVC), são lidas por um locutor profissional, digitalmente gravado e armazenado. Elas são escolhidas para constituir um *corpus* da fala, representativo de muitas transições e coarticulações a estudar. Um estágio de busca da regra é então aplicado nos dados, que é geralmente realizado por especialistas humanos. Uma primeira inspeção dos dados da fala de todos os locutores é feita para conseguir a forma da regra. Os valores atuais dos parâmetros deles (duração ou valores de frequência formante, por exemplo) são então ajustados para um locutor.

Quando um número suficiente de regras está pronto, a síntese pode começar. As regras são combinadas com as entradas fonéticas, e um sinal de fala paramétrico é produzido. Este é finalmente transformado em fala digital por um sintetizador que implementa o modelo escolhido no estágio de análise.

Por razões históricas e práticas (principalmente a necessidade para uma interpretabilidade física do modelo), os sintetizadores de regras sempre aparecem na forma de sintetizadores de formantes. Estes descrevem a fala como a evolução dinâmica de mais de 60 parâmetros, principalmente relacionados às frequências formantes e anti-formantes e larguras de banda juntos com as formas de onda glotais. Conseqüentemente, eles podem ser feitos quase livres de erros de modelagem intrínsecos, como mostrado por uma recente produção da fala de alta qualidade baseada em uma análise ou resíntese de formantes. Pelo contrário, o grande número de parâmetros complica o estágio de análise e tende a produzir erros extrínsecos. Além disso, frequências formantes e larguras de banda são inerentemente difíceis de estimar dos dados da fala.

3.2.5.4.1.2 Sintetizadores baseados em concatenação

Os sintetizadores baseados em concatenação possuem um conhecimento muito limitado dos dados que eles mantêm: muitos destes são incorporados nos segmentos a serem unidos. Isto claramente é mostrado na figura 3.12, onde todas as operações que poderiam indiferentemente serem usadas no contexto de um sintetizador musical têm sido agrupadas em um bloco de processamento do sinal, como oposto ao bloco ciência da fala do qual o plano requer ao menos algum entendimento de fonética.

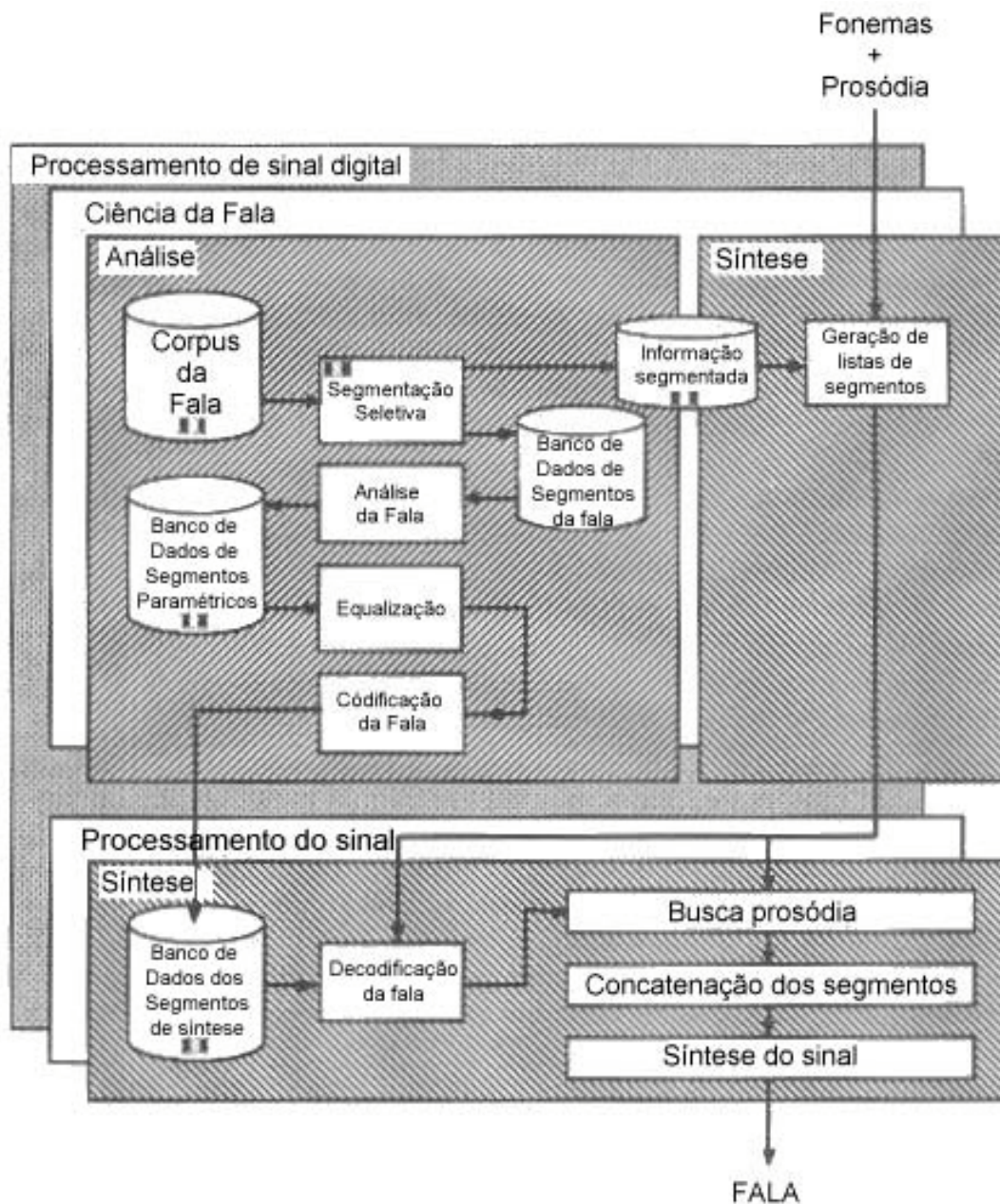


FIGURA 3.12 - Um sintetizador geral baseado em concatenação [DUT 97]

3.2.5.4.1.2.1 Preparação do banco de dados

Como antes, uma série de estágios preliminares têm de ser preenchidos antes que o sintetizador possa produzir a primeira expressão. Primeiro, os segmentos são escolhidos para minimizar os problemas de concatenação futuros. Uma combinação de difones, meia sílaba, e trifones são escolhidos como unidades de entrada, desde que eles envolvem muitas das transições e coarticulações enquanto requerendo uma quantidade permitida de memória.

Quando uma lista completa de segmentos surge, uma lista correspondente de palavras é cuidadosamente completada, de tal maneira que cada segmento apareça ao menos uma vez. Posições desfavoráveis, como dentro de sílabas enfatizadas ou em contextos reduzidos fortemente, são excluídas. Um *corpus* é então digitalmente gravado e armazenado, e os segmentos eleitos são marcados, ou manualmente com a ajuda de uma ferramenta de visualização do sinal, ou automaticamente com algoritmos de segmentação, as decisões dos quais são checadas e corrigidas interativamente. Formas de onda dos segmentos são então colecionadas em um banco de dados de segmentos da fala e a correspondente informação segmental (nomes dos segmentos, durações, etc) é armazenada em um arquivo. No caso dos difones, por exemplo, a posição da fronteira entre fones deveria ser armazenada, tal como ser capaz de modificar a duração de um meio-fone sem afetar o comprimento do outro.

Segmentos são então freqüentemente dados em uma forma paramétrica, na forma de uma seqüência temporal dos vetores de parâmetros colecionados na saída de um analisador de fala, e armazenado em um banco de dados de segmentos paramétricos. O uso do modelo de fala, é freqüentemente mantido por duas razões:

- Modelos bem escolhidos permitem a redução do tamanho dos dados
- Um número de modelos explicitamente separam as contribuições da fonte e do trato vocal.

De fato, a tarefa atual do sintetizador é produzir, em tempo real, uma seqüência adequada de segmentos concatenados, extraídos de um banco de dados de segmentos paramétrico e a prosódia, dos quais tem sido ajustados dos valores armazenados deles – isto é, a entonação e a duração que aparecem dentro do *corpus* de fala original – que é imposto pelo módulo de processamento da linguagem.

Desde que os segmentos a serem concatenados têm geralmente sido extraídos de diferentes palavras – isto é, em diferentes contextos fonéticos – eles freqüentemente apresentam amplitude e timbre diferentes. Mesmo no caso de sons vocálicos estacionários, por exemplo, uma seqüência dura de parâmetros tipicamente leva a descontinuidades audíveis.

3.2.5.4.1.2.2 Síntese de fala

Uma seqüência de segmentos é primeiro deduzida da entrada fonética do sintetizador, em um bloco denominado como geração da lista de segmentos na figura 3.12, que interliga os módulos PLN e PSD. Esta pesquisa o banco de dados de segmentos da fala para informações globais nas unidades que este contém. As durações dos segmentos são escolhidas de acordo, dados os comprimentos desejados dos fonemas incorporados. Em um caso mais simples, os limites dos fonemas são distribuídos proporcionalmente às contribuições respectivas iniciais dos segmentos envolvidos.

Uma vez os eventos prosódicos tenham sido corretamente determinados para segmentos individuais, o decodificador de fala busca o banco de dados de segmentos da síntese para os parâmetros atuais dos sons elementares a serem usados e decodifica-os. O módulo de determinação da prosódia então adapta-os um a um a prosódia exigida.

O bloco de concatenação de segmento está encarregado de dinamicamente encontrar os segmentos um para o outro, por suavizando as discontinuidades.

Dada a naturalidade das proximidades acústicas dos sons a serem encadeados, uma suavização linear simples é frequentemente aplicada, como é mostrado na figura 3.13.

O resultante conjunto de parâmetros é finalmente apresentado na entrada de um bloco de síntese, a exata contraparte da análise. A tarefa é produzir a fala.

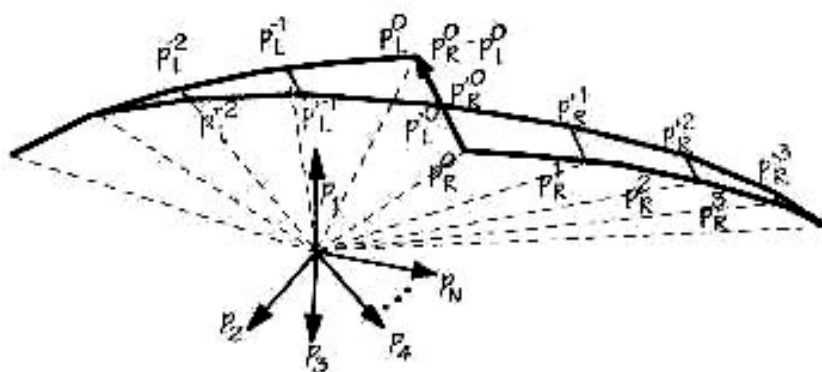


FIGURA 3.13 - Suavização linear paramétrica na borda de segmentos sucessivos [DUT 97]

3.2.5.4.1.2.3 Qualidade segmental

A figura 3.12 leva a uma descrição dos sintetizadores baseados em concatenação em termos da qualidade segmental. Uma vez a frequência de amostra do sintetizador tenha sido escolhida, a eficiência para produzir fala de alta qualidade é sujeita a:

- Tipo de segmentos escolhidos
- De que *corpus* eles foram extraídos
- Qualidade da segmentação
- Modelo do sinal de fala
- Quantidade de degradação introduzida pela fase de código de fala
- Eficiência de determinação da prosódia, que é fortemente relacionada ao modelo
- Capacidades do algoritmo de concatenação.

4 Processamento da Fala

O primeiro dispositivo de síntese elétrico foi introduzido por Stewart em 1922. O primeiro dispositivo a ser considerado um sintetizador de fala foi o VODER (Voice Operating Demonstrator) introduzido por Homer Dudley em Nova Iorque na Feira Mundial em 1939. O primeiro sintetizador de formantes, PAT (Parametric Artificial Talker), foi introduzido por Walter Lawrence em 1953. Em 1979, Allen, Hunnicutt e Klatt demonstraram o MITalk sistema texto para fala desenvolvido em MIT[LEM 90].

No contexto da UFRGS, recentemente foi desenvolvido no projeto, intitulado **Spoltech**, o qual é um projeto científico de desenvolvimento em tecnologias de processamento de linguagem falada humana. As áreas de pesquisa relacionadas envolvem reconhecimento de fala, síntese de fala e fonética. O **Spoltech**, financiado pelo CNPq, foi realizado de forma colaborativa entre o Brasil e os Estados Unidos. As instituições envolvidas foram a Universidade Federal do Rio Grande do Sul (UFRGS), através de seus Institutos de Informática e de Letras, a Universidade de Caxias do Sul (UCS), através de seu Departamento de Informática, o Oregon Graduate Institute (OGI) e a Universidade do Colorado: Center for Spoken Language Understanding (CSLU).

Os principais objetivos do **Spoltech** eram:

- Criar infra-estrutura para pesquisa colaborativa em tecnologias de processamento de linguagem humana entre o Brasil e os Estados Unidos, estabelecendo um ambiente de pesquisa comum entre as instituições participantes.
- Colaborar em pesquisa básica em várias áreas, incluindo reconhecimento de fala em Português, síntese de texto para fala e fonética, com o objetivo de obter-se um sistema de linguagem falada em Português e Inglês mais preciso, robusto e portátil.
- Desenvolver recursos de língua portuguesa falada no Brasil, tais como bancos de dados vocais (corpora) e de palavras.
- Incorporar estes avanços de pesquisa em softwares e disponibilizá-los sem encargos para a comunidade acadêmica, através dos mecanismos disponíveis na Internet.
- Colaborar no desenvolvimento de cursos de laboratório para educar estudantes e pesquisadores no desenvolvimento e pesquisa de sistemas de linguagem falada e suas tecnologias.

Estes objetivos gerais foram consubstanciados em uma primeira fase através da adequação de ferramentas desenvolvidas no CSLU/OGI para o idioma Inglês, já que este ambiente é suficientemente poderoso para o Português Brasileiro. A portabilidade dessas ferramentas de reconhecimento e síntese de fala para Português consistiu no enfoque principal deste projeto Spoltech. Este ambiente, onde estão inseridas diferentes ferramentas de síntese e reconhecimento de voz conhecido como CSLU Toolkit.

4.1 As ferramentas utilizadas

4.1.1 CSLU Toolkit

O CSLU Toolkit consiste de uma ferramenta de desenvolvimento para aplicações de processamento de linguagem falada humana. Ele é composto por um

conjunto de aplicativos que dão suporte, principalmente, às tarefas de rotulamento (separação dos fonemas), reconhecimento e síntese de fala. Os principais softwares que compõem o CSLU Toolkit são o RAD, o Festival e o Speech View, os quais serão descritos sucintamente a seguir. As funções e níveis do CSLU Toolkit são apresentados nas figuras 4.1 e 4.2.

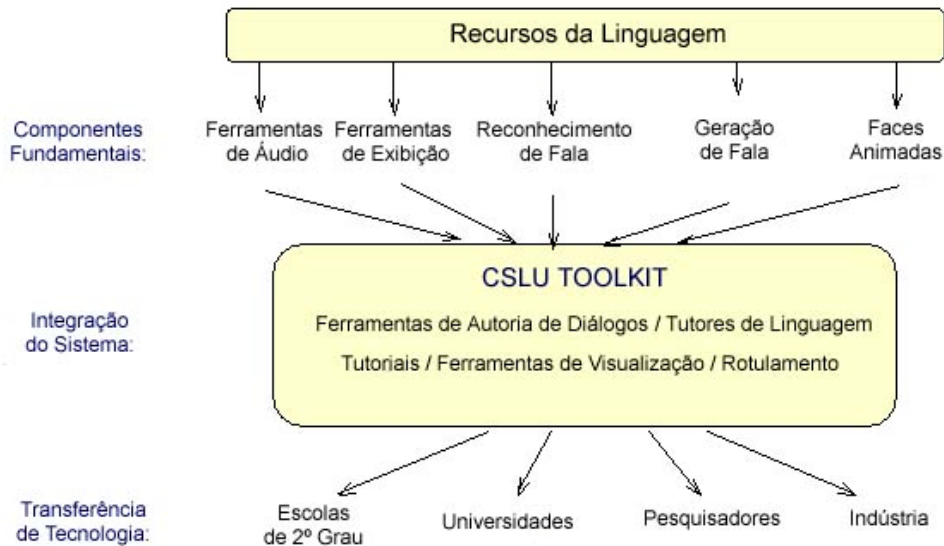


FIGURA 4.1 – CSLU Toolkit[HOS 2000]

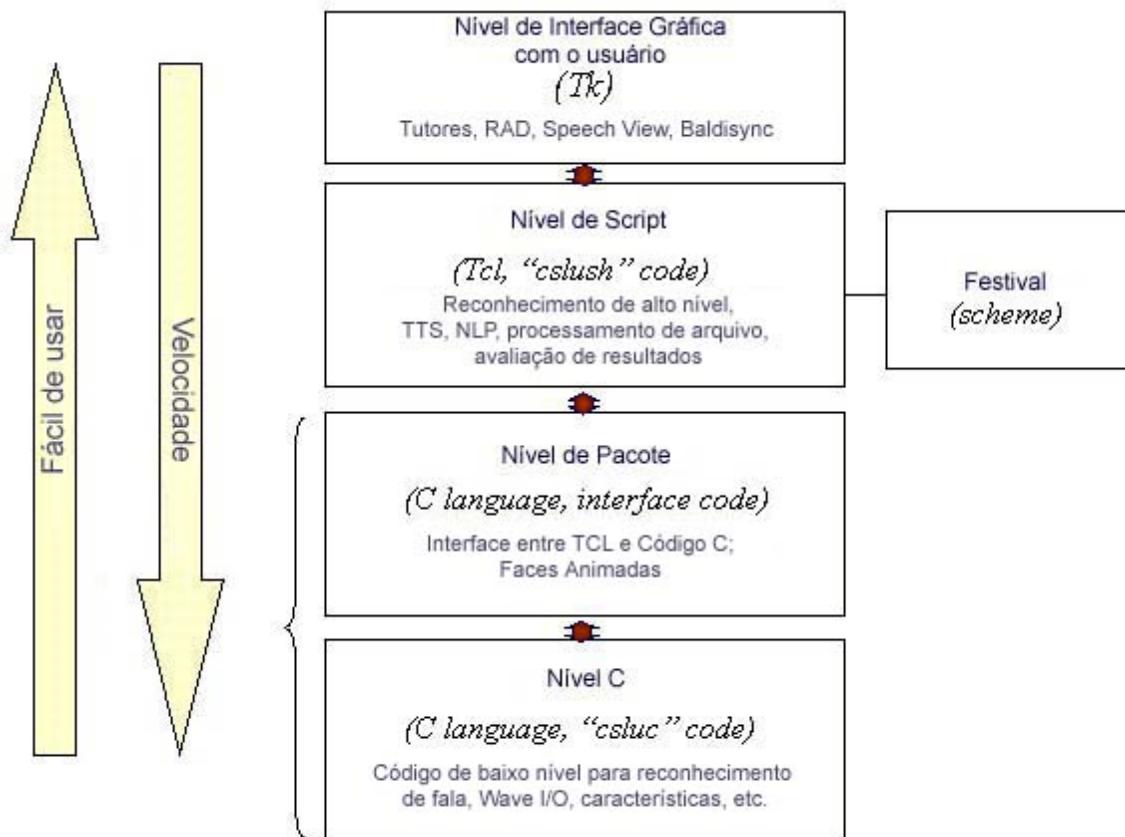


FIGURA 4.2 – Níveis do CSLU Toolkit[HOS 2000]

4.1.1.1 RAD

O CSLU Toolkit apresenta um ambiente para o desenvolvimento rápido de aplicações, o *Rapid Application Developer* – RAD. O mesmo permite a utilização dos recursos de processamento de linguagem falada do CSLU Toolkit, de um modo iterativo e intuitivo, para a implementação de interfaces para usuários. No RAD são disponibilizados uma série de objetos, a partir dos quais são montadas as aplicações. Estes objetos estão classificados em três grupos principais:

- *Objetos base.* São os objetos básicos que fazem parte do CSLU Toolkit. São os mais numerosos dentro do RAD e podem ser utilizados em aplicações de telefonia. A Figura 4.1 mostra os ícones dos objetos base.



FIGURA 4.3 – Ícones dos objetos base do RAD

- *Objetos Tucker-Maxon.* Estes objetos são parte do *plug-in* Tucker-Maxon. Eles habilitam algumas aplicações de multimídia. Na Figura 4.2 são mostrados os ícones dos objetos Tucker-Maxon.



FIGURA 4.4 – Ícones dos objetos Tucker Maxon

- *Objetos PSL.* Obejtos desenvolvidos pela PSL. São utilizados para conduzir experimentos com o RAD. A Figura 4.3 mostra os ícones dos objetos PSL.



FIGURA 4.5 – Ícones dos objetos PSL

4.1.1.2 BaldySinc

É uma ferramenta que pode ser usada para ver e criar animações faciais que estão alinhadas com o som gravado. Pode ser usada uma voz gravada ou texto para fala como fonte de som. Pode ler e escrever arquivos wave, tanto quanto arquivos especiais chamados de objetos de som. A interface do programa é a seguinte:

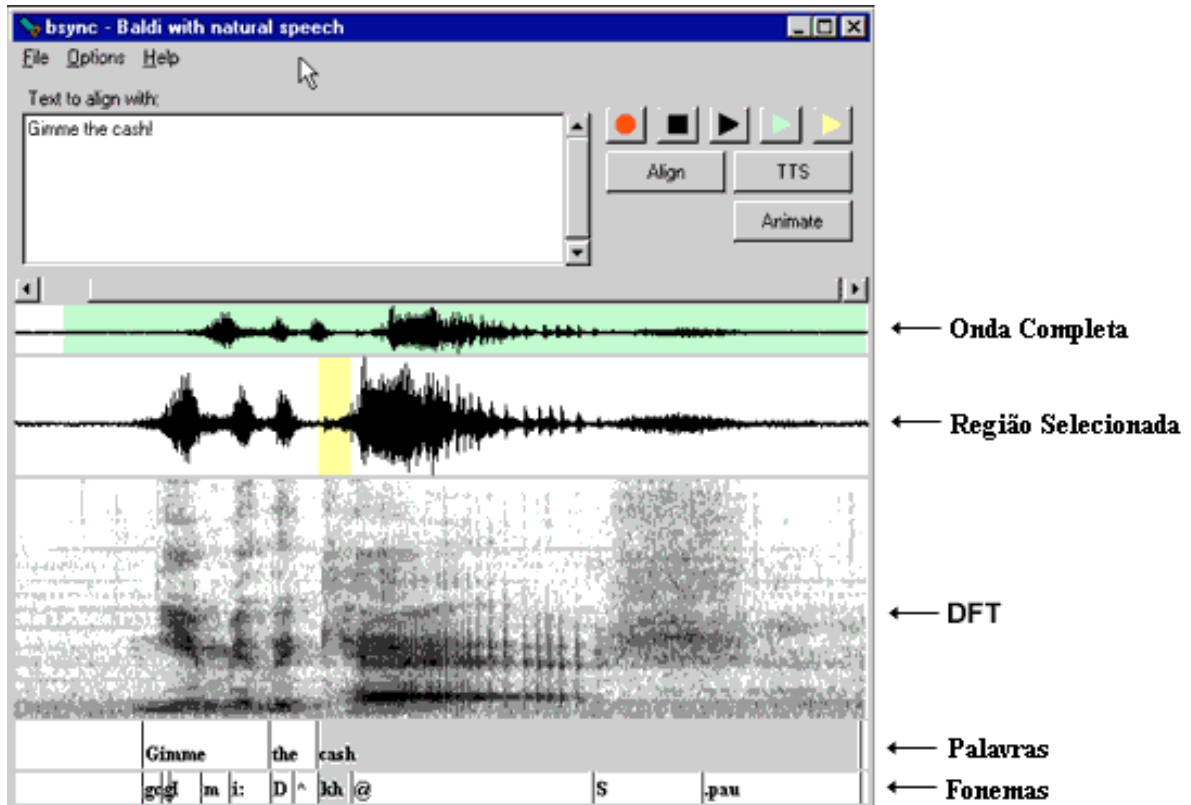


FIGURA 4.6 – Interface do BaldySinc

4.1.1.3 Festival

O Festival, desenvolvido na Universidade Edinburgo, Escócia, oferece um modelo geral para construir sistemas de síntese de fala bem como inclui exemplos de vários modelos. Como um todo oferece texto completo para fala através de várias APIs: do nível shell, através de um interpretador, como uma biblioteca C++, e uma interface Emacs. O Festival é multi-línguas, tendo sido feitas utilizações do mesmo para vários idiomas, incluindo Inglês (Reino Unido e Estados Unidos), Espanhol, italiano, etc.

O sistema está escrito em C++ e usa as ferramentas de fala de Edinburgh para arquitetura de nível baixo e tem um esquema baseado em um interpretador de comandos para controle.

Festival é planejado como um sistema de síntese de fala para ao menos três níveis de usuário. Primeiro, aqueles que simplesmente querem fala de alta qualidade de um texto arbitrário com o mínimo de esforço. Segundo, aqueles que estão desenvolvendo sistemas de linguagem. Neste caso, uma certa quantia de customização é desejada, tais como vozes diferentes, fraseamento específico, tipos de diálogo, etc. O terceiro nível está em desenvolvimento e teste de novos métodos de síntese.

Festival suporta os seguintes sistemas de síntese:

- Sintetizador unisyn
- Sintetizador de difones
- Sintetizador difone LPC
- MBROLA

Um dos grandes problemas no desenvolvimento de síntese da fala, e outras áreas de fala e sistemas de processamento de linguagem, é que há uma quantidade de técnicas bem conhecidas que podem ajudar a realizar este objetivo. Mas em ordem para melhorar algumas partes do sistema é necessário dispor-se um sistema em que se possa trabalhar em uma pequena parte para melhorar o todo. Sem um sistema como o Festival, para iniciar um módulo seria necessário gastar um significativo esforço para construir um sistema, ou adaptar um existente.

Festival é especificamente planejado para permitir a adição de novos módulos, facilmente e eficientemente.

Há outro aspecto do Festival que faz dele ser mais útil do que simplesmente um ambiente para pesquisa em novas técnicas de síntese. É um sistema texto para fala adaptável para incorporar em outros projetos que requerem saída de fala. Um trabalho completo de um sintetizador de fala de fácil uso em adição a apenas um ambiente de teste é bom por duas razões específicas. Primeiro, oferece um caminho para a pesquisa, em que os experimentos podem rapidamente e diretamente beneficiar os usuários do sistema de síntese. E segundo, assegurando que tem-se um sistema completo de trabalho pode-se imediatamente ver quais problemas existem e onde a pesquisa deveria ser direcionada.

Festival oferece um modelo comum onde técnicas múltiplas podem ser implementadas tal que possam ser testadas bastante no mesmo ambiente. O Festival trabalha em dois modos fundamentais: modo comando e o modo texto para fala. No modo comando, a informação (em arquivos ou através de uma entrada padrão) é tratada como comandos e é interpretada por um interpretador Scheme. Em modo TPF, (em arquivos ou através de uma entrada padrão) é tratada como um texto a ser sintetizado como fala. O modo default é o modo comando[BLA 99]. Scheme tem alguns fundamentos que são:

- **Sintaxe:** uma expressão é um átomo ou uma lista. Uma lista consiste de um parênteses esquerdo, um número de expressões e um parênteses direito. Átomos podem ser símbolos, números, strings ou outros tipos especiais como funções, tabelas hash, vetores, etc.
- **Semântica:** todas as expressões podem ser avaliadas. Listas são avaliadas como chamadas de função. Quando avaliando uma lista, todos os membros da lista são avaliados primeiro, então o primeiro item (função) é chamada com os itens restantes na lista como argumentos. Átomos são avaliados dependendo de seu tipo: símbolos são avaliados como variáveis retornando os valores deles. Comentários são iniciados por um ponto-e-vírgula e executados até o fim da linha.

A seguir, lista-se alguns exemplos de comandos do scheme utilizados no Festival.

```

festival> (+ 2 3)
5
festival> (set! a 4)
4
festival> (* 3 a)

```

```

12
festival> (define (add a b) (+ a b))
#<CLOSURE (a b) (+ a b)>
festival> (add 3 4)
7
festival> (set! alist '(apples pears bananas))
(apples pears bananas)
festival> (car alist)
apples
festival> (cdr alist)
(pears bananas)
festival> (set! blist (cons 'oranges alist))
(oranges apples pears bananas)
festival> (append alist blist)
(apples pears bananas oranges apples pears bananas)
festival> (cons alist blist)
((apples pears bananas) oranges apples pears bananas)
festival> (length alist)
3
festival> (length (append alist blist))
7

```

Festival suporta múltiplos conjuntos de fones simultaneamente e permite o mapeamento entre conjuntos quando necessário. Os léxicos, regras letra para som, sintetizadores de forma de onda, etc. todos requerem a definição de um conjunto de fones antes de eles operarem.

Um conjunto de fones é um conjunto de símbolos que podem ser definidos em termos de características, tais como vogal/consoante, local de articulação para consoantes, tipo de vogal, etc. O conjunto de características e os valores dos fones devem ser definidos com um conjunto de fones. A definição é usada para assegurar compatibilidade entre subsistemas tão bem quanto permitir grupos de fones em vários sistemas de predição.

A definição de um conjunto de fones tem a seguinte forma:

```

(defPhoneSet
  NAME
  FEATUREDEFS
  PHONEDEFS
)

```

O NAME é um símbolo único usado, por exemplo, mrpa, darpa, etc. FEATUREDEFS é uma lista de definições cada uma consistindo de um nome de característica e os valores possíveis. A terceira seção é uma lista de definições de fones. Cada definição de fone consiste de um nome de fone e os valores para cada característica na ordem que os valores foram definidos na seção anterior.

Note que o conjunto de fones também incluiria uma definição para qualquer fone silencioso. Em adição a definição do conjunto de fones silenciosos deve também

ser identificado para o sistema. Isto é feito através do comando `PhoneSet.silences`. No conjunto `mrpa` este é feito pelo comando

```
(PhoneSet.silences '(#))
```

Pode haver mais do que um fone silencioso (respiração, silêncio inicial, etc.) em qualquer definição de conjuntos de fones. Entretanto, o primeiro fone neste conjunto é tratado especialmente e deveria ser um silêncio canônico. Entre outras coisas este fone é inserido pelo módulo de predição da pausa.

Em adição a declaração de conjuntos de fones, conjuntos alternados podem ser selecionados pelo comando `PhoneSet.select`.

Fones em conjuntos diferentes podem ser automaticamente mapeados entre si usando características deles. Este mapeamento não é ainda tão geral quanto poderia ser, mas é útil quando mapeando entre vários conjuntos de fones da mesma linguagem. Quando um fone necessita ser mapeado de um conjunto para outro, o fone com características de busca é selecionado. Este permite, ao menos em alguma extensão, léxicos, sintetizadores de forma de onda, módulos de duração, etc a usar conjuntos de fones diferentes (embora em geral não é aconselhado).

Uma lista de conjunto de fones definidos correntemente é retornado pela função `(PhoneSet.list)`

Note-se que os conjuntos de fones não são freqüentemente definidos até que uma voz seja atualmente carregada, então esta lista não é a lista dos conjuntos que são distribuídos, mas a lista dos conjuntos que são usados pelas vozes correntemente carregadas. O nome, os fones, as características e os silêncios do conjunto de fones atual podem ser acessados com a função `(PhoneSet.description nil)`. Se o argumento para esta função é uma lista, apenas aquelas partes da descrição do conjunto de fones são retornados. Por exemplo:

```
(PhoneSet.description '(silences))
(PhoneSet.description '(silences phones))
```

Um léxico em Festival é um subsistema que fornece pronúncias para palavras. Pode consistir de três partes distintas: um adendo, tipicamente consistindo de palavras adicionadas a mão; um léxico compilado e um método para tratar palavras que estejam na lista ou não.

As entradas léxicas consistem de três partes básicas, uma palavra cabeçalho, uma parte da fala e uma da pronúncia. O campo da fala, atualmente, consiste de um átomo simples. Certamente, há muitas partes dos conjuntos de tags da fala e o que quer que você marque em seu léxico deve ser compatível com os subsistemas que usam aquela informação. Pode-se opcionalmente setar uma parte da fala para cada léxico. Como mencionado, anteriormente, os léxicos consistem de três partes básicas (forma compilada, adendo e método de palavra desconhecido) mais algumas outras declarações.

Cada léxico no sistema tem um nome que permite léxicos diferentes serem selecionados eficientemente quando trocando entre vozes durante a síntese. Os passos básicos envolvidos em uma definição léxica são como segue.

Festival oferece um ambiente onde novas vozes e linguagens podem facilmente ser inseridas no sistema. Cada uma é selecionada pela função ‘voice_ *’ que configura o sintetizador de forma de onda, conjunto de fones, léxico, duração e modelos de entonação para aquele falante.

As funções de voz atuais são:

- Voice_rab_diphone
- Voice_ked_diphone
- Voice_kal_diphone
- Voice_don_diphone
- Voice_el_diphone
- Voice_gsw_diphone
- Voice_en1_mbrola
- Voice_us1_mbrola
- Voice_us2_mbrola
- Voice_us3_mbrola

Um método geral para construir uma nova voz é definir os parâmetros para todas as sub-partes, por exemplo, conjunto de fones, parâmetro de duração, parâmetros de entonação etc., então definida uma função da forma voice_NAME que quando chamada atualmente, selecionará a voz.

Para muitas novas linguagens e frequentemente para novos dialetos, um novo conjunto de fones é necessário. É realmente, o bloco de construção básico de uma voz e muitas outras partes são definidas em termos deste conjunto.

4.1.1.4 Speech View

O Speech View é o componente do CSLU Toolkit usado para visualizar, reproduzir e editar fala digitalizada. Ele também é capaz de gerar espectrogramas e outros tipos de dados relacionados com fala, tais como *pitch* e contornos de energia, saídas da rede neural, e rótulos fonéticos. O Speech View é capaz de ler arquivos de voz previamente gravados, bem como gravar novas vozes e salvá-las em disco. As funções básicas do Speech View são utilizadas por outros componentes do CSLU Toolkit, como o BaldySinc e o RAD, citados anteriormente. Na figura 4.7 pode-se ver a interface do Speech View.

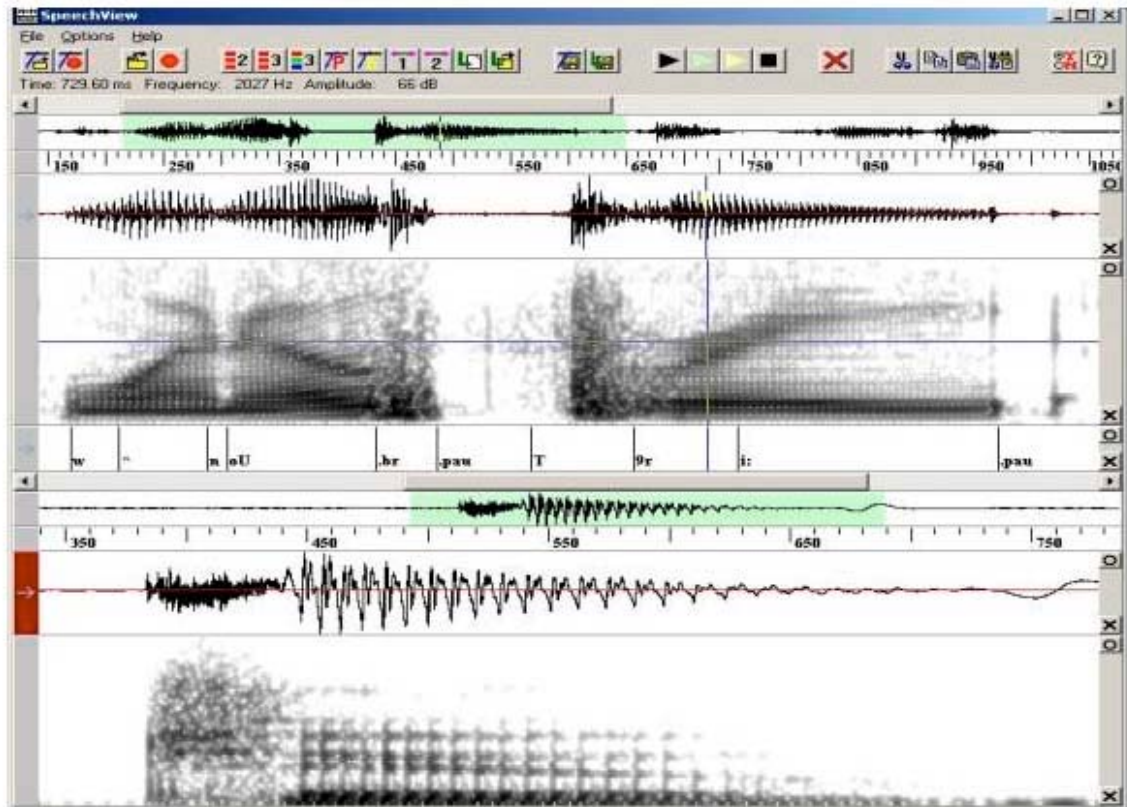


FIGURA 4.7 – Interface do Speech View

4.2 O processo de reconhecimento

No projeto **Spoltech**, o processo de reconhecimento de fala divide-se em quatro etapas principais ilustradas na Figura 4.4 e explicadas a seguir.

Primeiramente, a fala a ser reconhecida é digitalizada. O sinal de voz analógico da fala é captado por um microfone e depois convertido para um arquivo no formato wave. Para tanto, são utilizadas taxas de amostragem que variam de 8000 a 44100 amostras por segundo.

Na segunda etapa são computados as características que representam o domínio espectral contido na voz. A forma de onda digitalizada é convertida em uma representação espectral, na qual diferentes tipos de características podem ser utilizadas, dependendo do reconhecedor. Para os reconhecedores de propósito geral atuais, são utilizados coeficientes mel-cepstrais (MFCC) e coeficientes delta mel-cepstrais (delta MFCC). Estas características são computadas para cada 10 milissegundos de voz gravada. Cada uma dessas seções de 10 milissegundos é chamada de *frame*. Para cada *frame* são computadas 26 características.

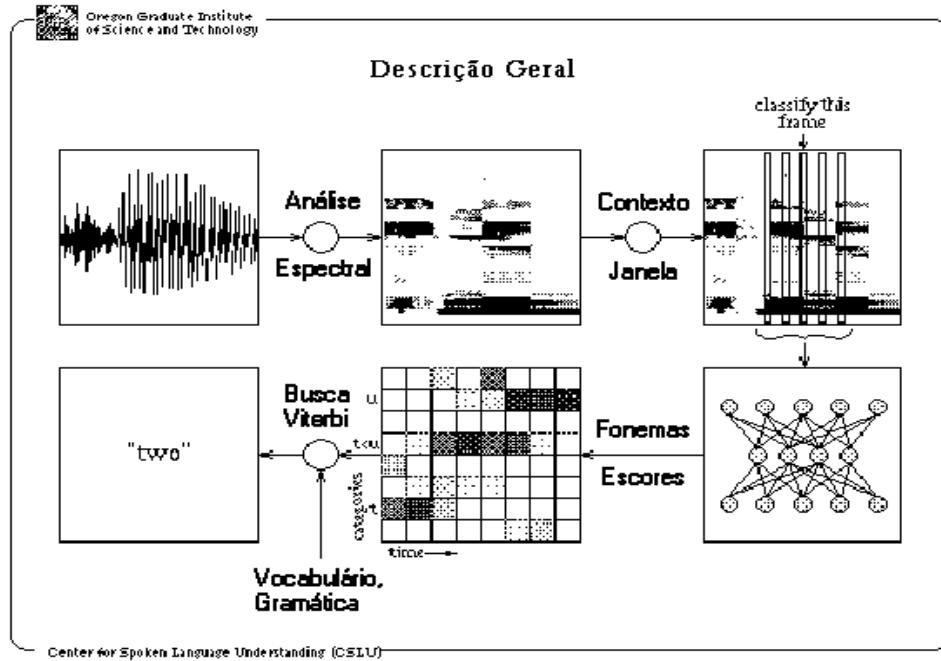


FIGURA 4.8 – Visão geral do processo de reconhecimento de fala[HOS 2000]

Uma rede neural é utilizada na seqüência para classificar os *frames* gerados na etapa anterior em categorias fonéticas. Uma categoria fonética representa, a grosso modo, uma parte de um fone em um determinado contexto, ou seja, sofrendo influências da pronúncia dos fones anteriores e posteriores ao mesmo. Para a classificação dos *frames* em categorias fonéticas é utilizada uma rede neural MLP (*multi-layer perceptron*). As saídas dessa rede são as estimativas das probabilidades, para cada categoria fonética, que os cinco *frames* correntes representem aquela categoria. É importante ressaltar que é a rede neural a responsável pelo mapeamento da fala do nível acústico para o nível fonético. A Figura 4.5 ilustra o uso da rede neural.

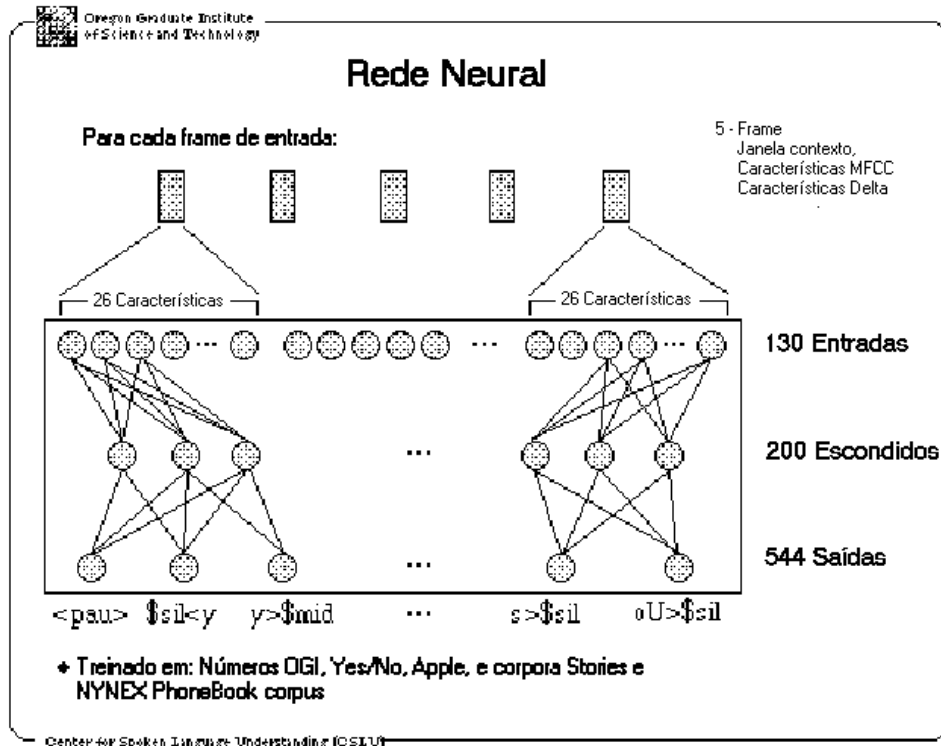


FIGURA 4.9– Cálculo das propabilidades de cada categoria fonética pela rede neural[HOS 2000]

Finalmente, o algoritmo de Viterbi é usado para relacionar as saídas da rede neural (estimativas para cada categoria fonética) com as palavras do vocabulário pré-estabelecido. O objetivo desta etapa é determinar a palavra, ou o comando, que mais se aproxima do que foi falado. Para isso é utilizada uma matriz de probabilidades e um conjunto de modelos de pronúncia. A Figura 4.6 ilustra um exemplo do uso do algoritmo de Viterbi para o cálculo das probabilidades que um conjunto de saídas da rede neural (matriz de probabilidades) representem a palavra “yes” e “no”.

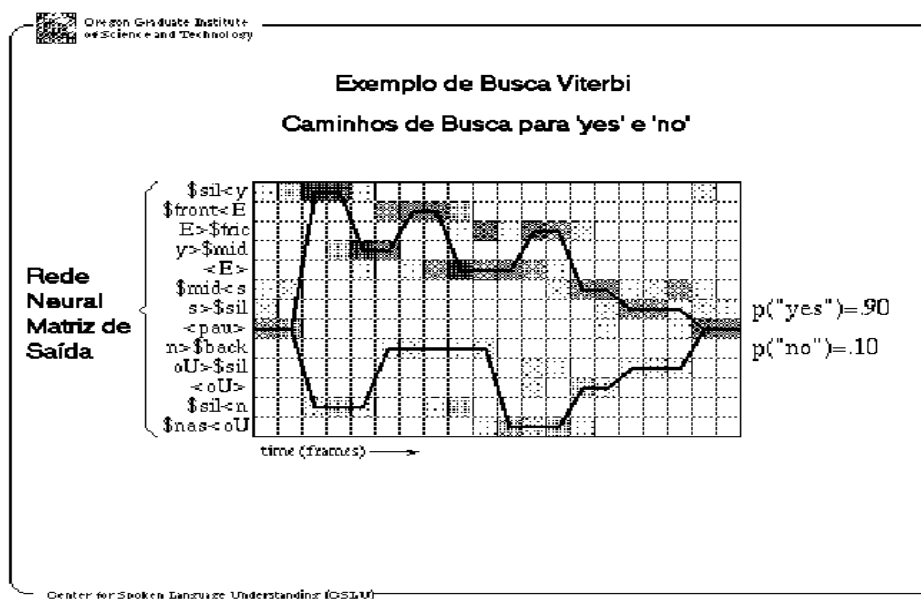


FIGURA 4.10 – Exemplo do uso do algoritmo de Viterbi[HOS 2000]

4.2.1 Treinamento dos reconhecedores de fala

Os reconhecedores de fala desenvolvidos no projeto **Spoltech** utilizam as técnicas descritas na seção anterior para o reconhecimento automático de palavras. A seguir, estão explicados os métodos utilizados para o treinamento destes reconhecedores que envolvem quatro etapas principais. São elas:

- Especificação das categorias fonéticas que o reconhecedor irá reconhecer.
- Procura de amostras de cada categoria utilizada no banco de dados de vozes.
- Treinamento da rede neural para o reconhecimento das categorias.
- Avaliação do desempenho da rede através de testes.

Para a especificação das categorias fonéticas que serão reconhecidas pela rede neural é necessário, primeiramente, definir o vocabulário a ser reconhecido e, para cada palavra deste vocabulário, determinar a pronúncia que será considerada como a correta. Assim, cada palavra é transcrita em seus fones componentes. Por exemplo, para as palavras “zero”, “presidente” e “república”, são definidas as seguintes transcrições fonéticas:

TABELA 4.1 – Transcrições fonéticas

Palavra	Pronúncia
Zero	z E r u
Presidente	pc p r e z i dc d e~ tS I
República	x e pc p u bc b l i kc k a

Fonte: Projeto Spoltech

Em seguida é utilizado um modelo de fones dependente de contexto, o que significa que um fone pode ser classificado diferentemente, dependendo dos fones que aparecem antes e depois do mesmo. Quando é feito esse modelamento, cada fone é dividido em uma, duas ou três partes. Cada uma destas partes corresponde a uma categoria fonética a ser treinada. Por exemplo, caso um fone seja dividido em três partes, a primeira parte sofrerá a influência do fone anterior, a segunda não sofrerá influência e a terceira sofrerá a influência do fone posterior. A Figura 4.7 ilustra o modelamento dependente de contexto.

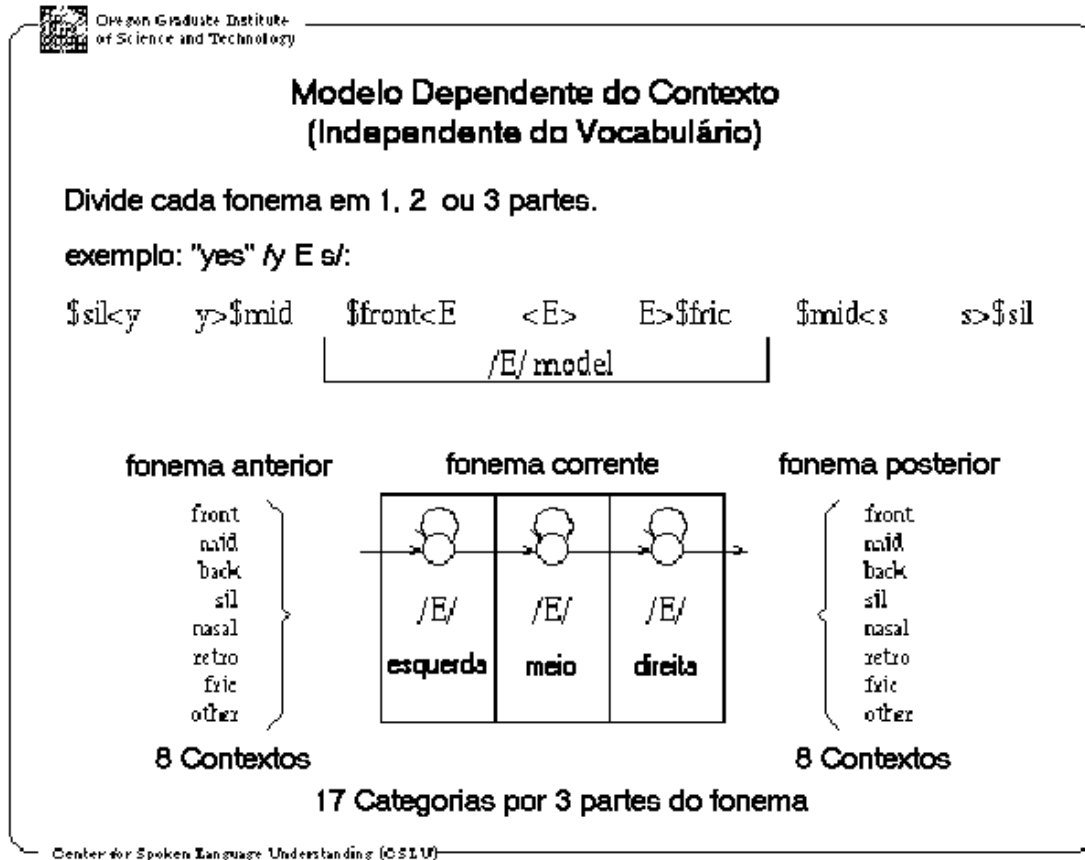


FIGURA 4.11 – Modelamento dependente de contexto[HOS 2000]

Como exemplo de modelamento dependente de contexto, para as palavras da tabela anterior, tem-se:

TABELA 4.2 – Exemplo de modelamento dependente de contexto

Fone	Partes	Fone	Partes
Z	2	e~	3
E	3	TS	1
R	2	I	3
U	3	X	2
Pc	1	Bc	1
P	2	B	2
E	3	L	2
I	3	Kc	1
Dc	1	K	2
D	2	A	3

Fonte: Projeto Spoltech

Na tabela anterior, “1” significa que o fone será independente de contexto, “2” significa que o fone será dividido em duas partes e “3” significa que o fone será dividido em três partes.

Feita a devida especificação das categorias fonéticas, a etapa seguinte consiste em procurar amostras das mesmas no banco de vozes (corpora). Para tanto são utilizados processos automáticos que utilizam programas desenvolvidos em TCL. As amostras encontradas são divididas em três grupos: um grupo para treinamento, um para desenvolvimento e um para teste. O grupo de treinamento, que é o maior dos grupos, é utilizado para treinar a rede neural. O grupo de desenvolvimento serve para medir o desempenho da rede treinada para que possam ser efetuados ajustes e melhorias. E finalmente, o grupo de teste serve para medir o desempenho da rede neural, depois de feitos os ajustes necessários, com dados ainda não utilizados.

Tendo sido encontrado um número razoável de amostras de cada categoria a ser treinada, esses dados são coletados e armazenados em um arquivo. Este arquivo é composto por vetores, os quais contêm, para cada amostra treinada, as características que serão colocadas como entradas na rede neural e a categoria alvo.

O arquivo de vetores é utilizado então para o treinamento da rede neural. As entradas da rede são as características extraídas dos *frames* do sinal de voz (coeficientes mel-cepstrais) e as saídas são as categorias fonéticas. Cada categoria treinada possui um nó correspondente na última camada da rede neural. Assim, para um conjunto de frames (cinco frames) é calculada pela rede uma estimativa de probabilidade que estes frames representem cada uma das categorias fonéticas.

Finalmente, após o treinamento, é feita uma avaliação, utilizando-se o grupo de dados de teste, do desempenho do reconhecedor. Esta avaliação pode ser realizada de duas formas: ao nível de palavras ou ao nível de fones. A avaliação ao nível de palavras compara, palavra por palavra, os resultados do reconhecedor, avaliando a porcentagem de acerto. Ao nível de fones, essa comparação é efetuada fone a fone.

4.2.2 Reconhedores desenvolvidos

Utilizando-se as técnicas descritas na seção anterior é possível a criação de dois tipos de reconhedores de fala: específicos ou de propósito geral. Os reconhedores específicos são aqueles treinados para reconhecer um vocabulário restrito de palavras, como por exemplo, comandos vocais. Os reconhedores de propósito geral são aqueles independentes de vocabulário, para os quais, é possível definir-se em tempo de execução, as palavras a serem reconhecidas.

A escolha de qual tipo de reconhecedor a ser utilizado depende em muito da natureza da aplicação alvo. Os reconhedores específicos, como poderá ser comprovado a seguir, apresentam melhores resultados, com uma taxa de acerto bem superior aos de propósito geral. Por outro lado, esses últimos, são muito mais flexíveis, permitindo o treinamento do reconhecedor sem uma definição prévia do vocabulário a ser utilizado. A equipe do Projeto **Spoltech** desenvolveu dois reconhedores de fala, um específico e um de propósito geral.

O reconhecedor específico criado reconhece dígitos, ou seja, seu vocabulário são os números em português (um, dois, três, quatro, cinco, seis, sete, oito, nove, zero e meia). A taxa de acerto de palavras obtida para este reconhecedor, através do uso dos arquivos de teste, foi de 98,47%, atingindo o mesmo nível de desempenho do reconhecedor de dígitos para a língua inglesa desenvolvido pelo CSLU que foi de 96,32%.

O reconhecedor de propósito geral foi criado a partir das sete frases balanceadas, descritas abaixo, que contém todos os fones da língua portuguesa. Como esse tipo reconhecedor é independente de vocabulário, o teste de seu desempenho é medido através da taxa de acerto de fones, ao invés de palavras, nos arquivos de teste. O resultado obtido para esse reconhecedor atingiu a taxa de acerto de fones de 59,91%, ainda inferior ao nível obtido pelo CSLU para o Inglês, que é em torno de 71%. Acredita-se que este resultado pode ser significativamente melhorado com a introdução de mais dados no corpora, que não possui um número suficiente de informações por ainda estar em fase de coleta.

As frases balanceadas são:

- a) O presidente da República faz advertência ao ministro da justiça.
/u p r e z i ' d e ~ n t S i d a x e ' p u b l i k a f a z a d v e r ' t e j ~ s i a w
m i ' n i s t r u d a Z u s ' t S i s a /
- b) O jovem atacante convenceu fácil na partida contra o México.
/u ' Z > v e j ~ n ~ a t a k a x ~ t S i k o ~ v e ~ ' s e w ' f a s i w n a p a r t S i d a
' k o ~ t r a u ' m E S i k u /
- c) O Zorro é outro dos filmes muito procurados nas locadoras atualmente.
/u ' z o x u E ' o w t r u d o s ' f i w m i s ' m u j ~ t u p r o k u ' r a d u z n a z
l o k a ' d o r a z a t u a w ' m e j ~ t S i /
- d) A primeira maior guerra de todas foi entre o bem e o mal, o céu e a terra.
/a p r i ' m ' e j r a m a j ' > r ' g E x a d e ' t o d a s f o j ' e ~ n t r i w ' b e j ~ n ~
i w ' m a w u s E w i a t E x a /
- e) É melhor nunca engomar os lençóis azuis debaixo de sol.
/E m e ' L > r ' n u ~ k a ' e j ~ g o m a r o s l e ~ ' s > j z a ' z u j z d e ' b a j S u d Z i
' s > w /

Entre barras está a transcrição fonética, feita arbitrariamente, para que houvesse toda essa cobertura; uma vez que conforme a variação que se nota de um dialeto para outro, o inventário de fones pode diminuir. Ou seja, supondo um falante ideal, que fale exatamente como nas transcrições, irá abranger todos os fones. Ainda assim, se quisermos prever uma margem de erro por causa de um falante que não é ideal, sugerem-se mais duas frases apenas a partir dos seguintes dados:

- A maioria das consoantes não muda, não importando o dialeto.
- Podem sofrer alteração:
 - t /tS/ e d /dZ/ antes de /i/, passando a /t/ e /d/;
 - s gráfico do final de palavra soando como /S/ (chiado), ao invés de /s/ ou /z/;
 - r gráfico do final de palavra soando como /x/ (velar) ao invés de /r/ ou /rr/;
 - r intervocálico (ou mesmo o dígrafo rr) soando como /r/.
- Ainda assim os exemplos devem cobrir todos os fones.
- As vogais do português brasileiro são em número de sete, suas respectivas nasais, duas semivogais e a possibilidade destas também nasais, quando parte de um ditongo nasal.

- Todos os dialetos provavelmente usam todas essas vogais e demais possibilidades. No entanto, pode-se verificar num suposto dialeto do Nordeste, é muito maior a ocorrência das vogais abertas /E/ e />/('ó' aberto), o que, dependendo das frases, poderia eliminar a ocorrência das fechadas /e/ e /o/.

Pensando nisso é que foram coletadas duas frases (na verdade uma cortada em duas, extraída de uma música de Raul Seixas) que cobre todas as possibilidades de vogais, abertas e fechadas.

f) Eu prefiro ser essa metamorfose ambulante.

/ew 'prEfiru 'ser E sa mEtam > x'f > zi ax~ bu'lax~ tSi/

g) Do que ter aquela velha opinião formada sobre tudo.

/du ki'ter a'kEl a 'vEL a > pini'ax~w f > x'ma da 'sobri 'tudu/

Superada esta fase de elaboração das frases, iniciou-se a coleta de vozes.

4.3 O processo de síntese da fala

O CSLU Toolkit se utiliza do Festival Speech Synthesis System como sintetizador de fala, o qual foi descrito anteriormente.

Durante dois anos de trabalho no Projeto Spoltech, duas vozes do Português Brasileiro foram desenvolvidas. A principal diferença entre elas é o banco de dados difone. O banco de dados para a segunda voz é mais completo. Ambas as vozes são muito compreensíveis, mas a prosódia não está muito boa ainda.

4.3.1.1 Criação de uma lista de fonemas do português brasileiro

Este foi o primeiro grande objetivo do Projeto Spoltech. Agora, nós temos uma lista dos fonemas do Português Brasileiro. A criação da lista torna possível o desenvolvimento de ferramentas e módulos para a síntese e reconhecimento. A lista de fonemas é mostrada a seguir.

4.3.1.2 Criação de uma lista de difones

Usando todas as combinações possíveis entre fonemas brasileiros que podem ser pronunciados por alguém, nós fizemos uma lista de difones que foram usados para gravar e desenvolver o banco de dados difone para as vozes brasileiras. Atualmente, quase todos os problemas das primeiras versões foram fixados, e foi possível a gravação de um novo e completo conjunto de difones. Basicamente, esta nova lista de difones foi gerada de maneira automática, e então todos os difones mal pronunciados foram removidos manualmente.

TABELA 4.3 – Lista de Fones

Tipo do Fone	Fone
Vogais	a E e i > o u ax~ e~ i~ o~ u~
Semi-vogais	j w
Consoantes	n~ p b t d k g f v s z S Z m n l L r
Silêncio	pau

Fonte: Projeto Spoltech

4.3.1.3 Criação do banco de dados difone para a língua portuguesa falada no Brasil

Todas as vozes necessárias para os bancos de dados difone foram gravadas no OGI, porque não havia equipamento adequado para gravar na UFRGS.

No primeiro banco de dados, o alinhamento dos difones foi forçado. Teve-se alguns problemas, porque vários arquivos de áudio do laringográfico foram cortados, então alguns difones soaram erroneamente. Além disso, houve vários difones perdidos, logo a primeira lista de fones não estava completa.

Na Segunda, nós temos uma lista completa de todos os difones que nós realmente necessitamos, e todas os arquivos de voz foram gravados sem nenhum corte. Os difones foram rotulados, processo que levou duas semanas. Depois disso um novo banco de dados difone foi criado, fornecendo uma voz melhorada.

4.3.1.4 Criação de regras LTS

A criação de um conjunto inicial de regras LTS foi feito no Brasil, mas a primeira versão do trabalho foi completada durante a visita dos pesquisadores do Projeto Spoltech ao OGI em 1999. Esta versão sofreu muitas modificações, gerando um conjunto preciso de regras. Embora seja preciso, há alguns problemas na pronúnciação.

Em comparação com versões anteriores, a qualidade alcançada agora é melhor. Em versões anteriores nós encaramos o problema das vogais “e” e “o” – foi-se incapaz de determinar a pronúnciação correta delas a menos que elas fossem acentuadas. – porque, não há regra na língua portuguesa para determinar isto. Para resolver este problema fez-se uma análise estatística sobre um dicionário de pronúnciação com quase 20.000 palavras. Procurou-se todas as ocorrências de “e” e “o”, a vizinhança e a pronúnciação deles. Então conseguiu-se as pronúnciações mais comuns para cada vizinhança e criou-se regras LTS para estes casos. Todos os outros casos deveriam estar no dicionário de pronúnciação do Festival.

As atuais regras LTS estão trabalhando bem para muitas palavras do Português. Mas, apesar dos melhoramentos, a síntese para algumas palavras encontra problemas, os quais são:

- Todos os homógrafos. Este é um ponto muito difícil de ser resolvido ainda. Uma maneira de alcançar melhores resultados é um bom módulo parte da fala e regras LTS que usam informações de parte da fala para determinar a pronúnciação, mas não foi desenvolvido um bom módulo parte da fala ainda, apenas foi adaptado à versão em Inglês. Nossas regras LTS não usam parte da fala ainda.
- As regras LTS para a pronúnciação de letras “e” e “o” estão trabalhando bem, mas um bom dicionário com casos de exceção não foi desenvolvido.
- Há ainda um problema com a letra “x”, porque esta letra pode assumir várias pronúnciações diferentes, e não há regras na língua portuguesa para definir o correto. A única solução parece ser um bom dicionário de pronúnciação.

4.3.1.5 Criação de um módulo para silabificação das palavras

O módulo atual é quase o mesmo desenvolvido no início do projeto. É muito preciso e trabalha bem em quase 100% das palavras. O algoritmo usado neste módulo foi desenvolvido baseado em algumas regras lingüísticas básicas. Este algoritmo não

trabalha sobre palavras ortográficas, mas sobre as transcrições fonéticas delas. Então, se o módulo recebe uma palavra mal grafada foneticamente (isto pode ser causado por um erro nas regras LTS), isto pode resultar em um erro na silabificação da palavra.

4.3.1.6 Criação de um módulo para setar a acentuação

Em português, toda a palavra tem uma das últimas três sílabas acentuadas, há um conjunto muito bem definido de regras lingüísticas para determinar a posição de acentuação correta, baseada nas últimas letras da palavra. Estas regras onde adaptaram-se as funções sobre os fones, ao invés das letras, usou-se para desenvolver este módulo. Este módulo trabalha muito bem, mas depende dos resultados de módulos anteriores. Se recebe uma palavra mal pronunciada foneticamente e ou fortemente silabificada, pode resultar em uma palavra fortemente enfatizada.

4.3.1.7 Criação de um dicionário de pronúncia

Este dicionário é uma parte muito importante do sintetizador, mas a versão atual é pequena. Tem-se um dicionário de pronúncia, com quase 20.000 palavras, que usam outro conjunto de símbolos fonéticos, ao invés dos Worldbet. Também, a criação de um dicionário para o festival é uma questão simples de converter o formato.

4.3.1.8 Criação de um módulo de tokens para palavras

A versão atual é uma adaptação inicial do módulo para Espanhol, restando ainda vários melhoramentos a fazer[BAR 2001].

5 Implementação

5.1 O problema

É cada vez maior a necessidade do homem interagir de forma mais natural com o computador, ou genericamente, com as máquinas. Modos de entrada/saída naturais vêm sendo vislumbrados para prover esta comunicação, que acarreta em várias facilidades e vantagens[HUG 95].

O uso da fala como uma forma de saída do computador tem crescido de importância como parte das interfaces homem-máquina amigáveis e para aplicações específicas[CAM 90].

Converter um texto irrestrito em fala requer a capacidade de converter qualquer sentença na linguagem; portanto, toda a palavra da linguagem deve ser reconhecida. Isto requer a disponibilidade ou de um algoritmo geral capaz de converter qualquer palavra em uma seqüência de fonemas, ou um vocabulário compreendendo todas as palavras da linguagem[CAM 90].

Em Português, o problema de conversão de texto para fonemas é mais simples do que em Inglês, desde que as letras podem ser de forma não ambígua convertidas para fonemas, mas há exceções que requerem o uso de um dicionário[CAM 90].

Na síntese de fala para o português falado no Brasil, há vários problemas com relação a esta como, por exemplo: separação de sílabas, determinação da sílaba tônica, ambigüidade, abreviações, “e” e “o” aberto e fechado:

- Letras E e O: estas letras podem ser pronunciadas em aberto /E/, />/ ou fechado /e/, /o/;
- Letra X: esta letra representa um dos quatro fonemas: /s/, /ks/, /s/ e /z/;
- Nasalização: pode ocorrer com a vogal /a/, que torna-se /ã/ nas seguintes situações:
 - Quando acentuada e antes de uma consoante nasal /m/, /n/;
 - Sempre antes de uma seqüência de uma consoante nasal ou outra consoante;
 - Sempre antes de uma consoante nasal no final de uma palavra[CAM 90].

Com base nestes problemas apresentados de transcrição fonética da língua portuguesa falada no Brasil, resolveu-se focalizar neste trabalho uma melhor pronúncia automática das palavras com “e” e “o”, ou seja, como determinar que a pronúncia correta da primeira e segunda ocorrência da letra ‘e’ da palavra “febre”, por exemplo, seja um ‘e’ aberto (/E/) ou fechado (/e/). A solução desenvolvida para resolver o problema será apresentada a seguir.

5.2 A solução

Sabendo-se que o Festival utiliza arquivos onde constam comandos ou regras letra para som, as quais são lidas por um interpretador Scheme, criar-se-á novas regras que serão utilizadas pelo Festival para solucionar o problema apresentado

anteriormente. Se estas regras não fossem criadas, mesmo tendo uma palavra a pronúncia do fonema “e”, por exemplo, como aberto (/E/), o Festival geraria a sua pronúncia como um fonema fechado. Sabendo-se disso, utilizar-se-á apenas as regras que geram a pronúncia dos fonemas como abertos (/E/ e />/).

Os fonemas tanto aberto como fechado para as vogais “e” e “o” são listados na tabela a seguir (representação fonética).

TABELA 5.1 – Fonemas respectivos das vogais “e” e “o”

Vogal	Tipo do Fonema	Fonema
E	Fechado	/e/
E	Aberto	/E/
O	Fechado	/o/
O	Aberto	/>/

O método utilizado para gerar uma regra letra para som a ser incorporada ao Festival é a análise da letra anterior e posterior a um ‘e’ ou ‘o’ de uma palavra, e após isto verifica-se a sua pronúncia na transcrição fonética da mesma para determinar se o ‘e’ ou ‘o’ encontrados em uma palavra em processamento é um fonema aberto(/E/, />/) ou fechado(/e/, /o/). O fluxograma do processo pode ser visto na figura a seguir.

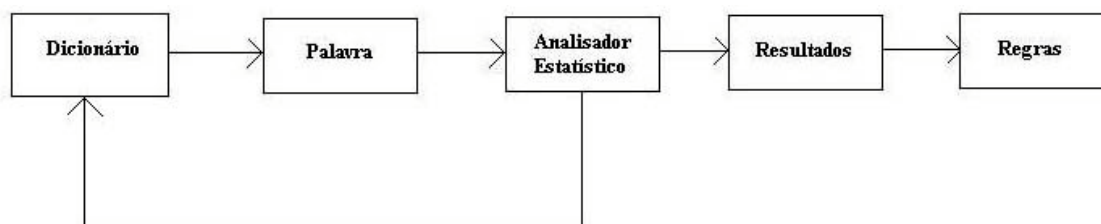


FIGURA 5.1 – Fluxograma da análise estatística

Visando consistir a busca de regras para a pronúncia correta das palavras com ‘e’ ou ‘o’ aberto/fechado, utilizou-se um dicionário eletrônico que contém 19.156 palavras do português falado no Brasil, onde consta a palavra e a sua transcrição fonética (como pode ser visto na tabela 5.2), o qual será utilizado por um analisador estatístico que aplicará o método explicitado anteriormente. O dicionário utilizado para capturar estas palavras e as suas respectivas transcrições fonéticas foi o DIC Prático Michaelis versão 5.1 de julho de 1998. Um exemplo dos verbetes encontrados para a letra “a” do dicionário encontra-se no Anexo 1 deste trabalho.

TABELA 5.2 – Exemplo de palavras encontradas no dicionário

Palavra	Transcrição Fonética
(abafação)	(abafa'säw)
(abafamento)	(abafam'ëtu)
(abalroar)	(abawro'ar)
(babado)	(babadu)
(babador)	(babador)

Exemplo de palavras encontradas no dicionário
Continuação

Palavra	Transcrição Fonética
(babão)	(babäw)
(cabeçalho)	(kabesaLu)
(cabecear)	(kabesear)
(cabeceira)	(kabesejrA)
(daninho)	(daniñu)
(dano)	(dânu)
(danoso)	(danozu)
(éban)	(Ebanu)
(ébrio)	(Ebrju)
(ebulição)	(ebulisäw)

Baseado neste dicionário, com o intuito de determinar se a palavra possui um “e” ou “o” aberto ou fechado em sua representação fonêmica, separou-se as palavras com “e” e “o” em seus radicais com as suas respectivas transcrições fonêmicas, ou seja, criou-se um arquivo com as palavras com “e” em seus radicais (tabela 5.3) e um outro arquivo com as palavras com “o” em seus radicais (tabela 5.4). Para esta separação das palavras utilizou-se um programa em linguagem C, o qual analisou as palavras para efetuar a divisão.

TABELA 5.3 – Palavras com “e” em seus radicais

Palavra	Transcrição Fonética
(abastecedor)	(abastesedor)
(abastecer)	(abasteser)
(abastecimento)	(abastesime~tu)
(abater)	(abater)
(abdômen)	(abdomej~)
(abecedário)	(abesedarju)
(abeirar)	(abejrar)
(abelha)	(abeLa)
(abelhão)	(abeLaxw~)
(abelhudice)	(abeLudZ isi)
(abelhudo)	(abeLudu)
(aberração)	(abexasaxw~)
(aberto)	(abErtu)
(abertura)	(abertura)
(abespinhar)	(abespin~ar)

TABELA 5.4 – Palavras com “o” em seus radicais

Palavra	Transcrição Fonética
(abafador)	(abafador)
(abalroar)	(abawroar)
(abanador)	(abanador)
(abandonado)	(abax~donadu)

Palavras com “o” em seus radicais
Continuação

Palavra	Transcrição Fonética
(abandonar)	(abax~donar)
(abandono)	(abax~donu)
(abarrotado)	(abaxotadu)
(abarrotar)	(abaxotar)
(abastecedor)	(abastesedor)
(abdome)	(abdomi)
(abdômen)	(abdomej~)
(abdominal)	(abdominaw)
(abençoar)	(abe~soar)
(abobadado)	(abobadadu)
(abobadar)	(abobadar)

Após esta divisão, utilizou-se outra rotina em linguagem C (analisador estatístico) para fazer os levantamentos estatísticos das palavras com ‘e’ e ‘o’, os quais serão descritos a seguir.

No arquivo com as palavras com “e” em seus radicais fez-se um levantamento estatístico baseado no método apresentado, e, com isso, criou-se uma tabela de ocorrências de “e” aberto (/E/) ou fechado (/e/) para as letras antes e depois do “e” na palavra. Nas ocorrências que possuíam 80% (limiar) ou mais de pronúncia do “e” como aberto foram consideradas como uma regra para o problema (tabela 5.5). Na tabela estão ressaltadas em negrito as regras encontradas para o fonema “e” aberto (/E/).

TABELA 5.5 – Exemplo da Análise estatística da ocorrência de ‘e’ fechado (/e/) e ‘E’ aberto (/E/) (onde ‘#’ significa nada)

Letra Anterior	Letra Posterior	Casos - e	Casos – E	Total de Casos	%e	%E
#	b	2	0	2	100.00%	0.00%
#	c	20	1	21	95.24%	4.76%
#	d	15	0	15	100.00%	0.00%
#	f	17	0	17	100.00%	0.00%
#	g	4	0	4	100.00%	0.00%
#	i	3	0	3	100.00%	0.00%
#	j	2	0	2	100.00%	0.00%
#	l	50	3	53	94.34%	5.66%
#	m	33	0	33	100.00%	0.00%
#	n	31	0	31	100.00%	0.00%
#	p	18	0	18	100.00%	0.00%
#	q	23	0	23	100.00%	0.00%
#	r	24	2	26	92.31%	7.69%
#	s	754	4	758	99.47%	0.53%
#	t	13	0	13	100.00%	0.00%
#	u	8	0	8	100.00%	0.00%
#	v	27	0	27	100.00%	0.00%

Exemplo da Análise estatística da ocorrência de ‘e’ fechado (/e/) e ‘E’ aberto (/E/)
(onde ‘#’ significa nada)

Continuação

Letra Anterior	Letra Posterior	Casos - e	Casos – E	Total de Casos	%e	%E
#	x	197	2	199	98.99%	1.01%
-	d	3	0	3	100.00%	0.00%
-	s	2	2	4	50.00%	50.00%
a	c	5	0	5	100.00%	0.00%
a	l	1	0	1	100.00%	0.00%
a	s	0	1	1	0.00%	100.00%
a	z	1	0	1	100.00%	0.00%
b	a	10	0	10	100.00%	0.00%
b	b	10	0	10	100.00%	0.00%
b	c	6	3	9	66.67%	33.33%
b	d	17	0	17	100.00%	0.00%
b	f	0	1	1	0.00%	100.00%
b	i	29	0	29	100.00%	0.00%
b	j	3	0	3	100.00%	0.00%
b	l	40	10	50	80.00%	20.00%
b	n	16	0	16	100.00%	0.00%
b	q	1	2	3	33.33%	66.67%
b	r	64	0	64	100.00%	0.00%
b	s	21	0	21	100.00%	0.00%
b	t	13	0	13	100.00%	0.00%
b	u	1	0	1	100.00%	0.00%

O mesmo procedimento foi aplicado no arquivo com as palavras com o “o” em seus radicais, ou seja um levantamento estatístico para determinar as regras para a ocorrência do “o” aberto(/>/) ou fechado(/o/). Com o mesmo objetivo do problema do “e”, as ocorrências que possuíam 80% (limiar) ou mais de aparição na transcrição fonética com o “o” aberto (>/) foram consideradas como uma regra de solução para o problema (Tabela 5.6). Na tabela estão ressaltadas em negrito as regras para a solução do problema.

TABELA 5.6 – Exemplo da Análise estatística da ocorrência de ‘o’ fechado (/o/) e ‘O’ aberto (>/) (onde ‘#’ significa nada)

Letra Anterior	Letra Posterior	Casos - o	Casos - O	Total de Casos	%o	%O
#	b	59	0	59	100.00%	0.00%
#	c	25	0	25	100.00%	0.00%
#	d	5	0	5	100.00%	0.00%
#	e	1	0	1	100.00%	0.00%
#	f	22	0	22	100.00%	0.00%
#	i	6	0	6	100.00%	0.00%
#	j	1	0	1	100.00%	0.00%
#	l	21	0	21	100.00%	0.00%

Exemplo da Análise estatística da ocorrência de ‘o’ fechado (/o/) e ‘O’ aberto (/>/) (onde ‘#’ significa nada)

Continuação

Letra Anterior	Letra Posterior	Casos - o	Casos - O	Total de Casos	%o	%O
#	m	5	0	5	100.00%	0.00%
#	n	4	0	4	100.00%	0.00%
#	p	30	0	30	100.00%	0.00%
#	r	57	0	57	100.00%	0.00%
#	s	18	0	18	100.00%	0.00%
#	t	5	0	5	100.00%	0.00%
#	u	20	0	20	100.00%	0.00%
#	v	12	0	12	100.00%	0.00%
#	x	6	0	6	100.00%	0.00%
#	z	1	0	1	100.00%	0.00%
-	u	1	0	1	100.00%	0.00%
a	l	1	0	1	100.00%	0.00%
a	r	1	0	1	100.00%	0.00%
b	a	8	0	8	100.00%	0.00%
b	b	12	0	12	100.00%	0.00%
b	c	25	1	26	96.15%	3.85%
b	d	6	0	6	100.00%	0.00%
b	e	1	0	1	100.00%	0.00%
b	f	3	0	3	100.00%	0.00%
b	g	1	0	1	100.00%	0.00%
b	i	11	0	11	100.00%	0.00%
b	j	2	0	2	100.00%	0.00%
d	q	0	1	1	0.00%	100.00%
d	r	340	1	341	99.70%	0.30%
d	á	1	0	1	100.00%	0.00%
d	s	26	2	28	92.85%	7.15%
D	t	5	0	5	100.00%	0.00%
D	u	19	0	19	100.00%	0.00%
D	v	2	0	2	100.00%	0.00%
D	x	0	1	1	0.00%	100.00%

Nas 19.156 palavras do dicionário ocorreram 8.910 palavras que possuíam a vogal “e” com uma ou mais ocorrências na mesma palavra, sendo o fonema para esta aberto (/E/) e/ou fechado (/e/) e 5.290 palavras que possuíam a vogal “o” com uma ou mais ocorrências na mesma palavra, sendo o respectivo fonema aberto (/>/) e/ou fechado (/o/). Ou seja, 46.51% do dicionário contém a letra “e” e 27.62% do dicionário contém a letra “o” nos radicais das palavras, conforme pode-se ver na figura 5.2, demonstrando a relevância do presente trabalho na tentativa de melhorar a produção sonora dessas palavras.

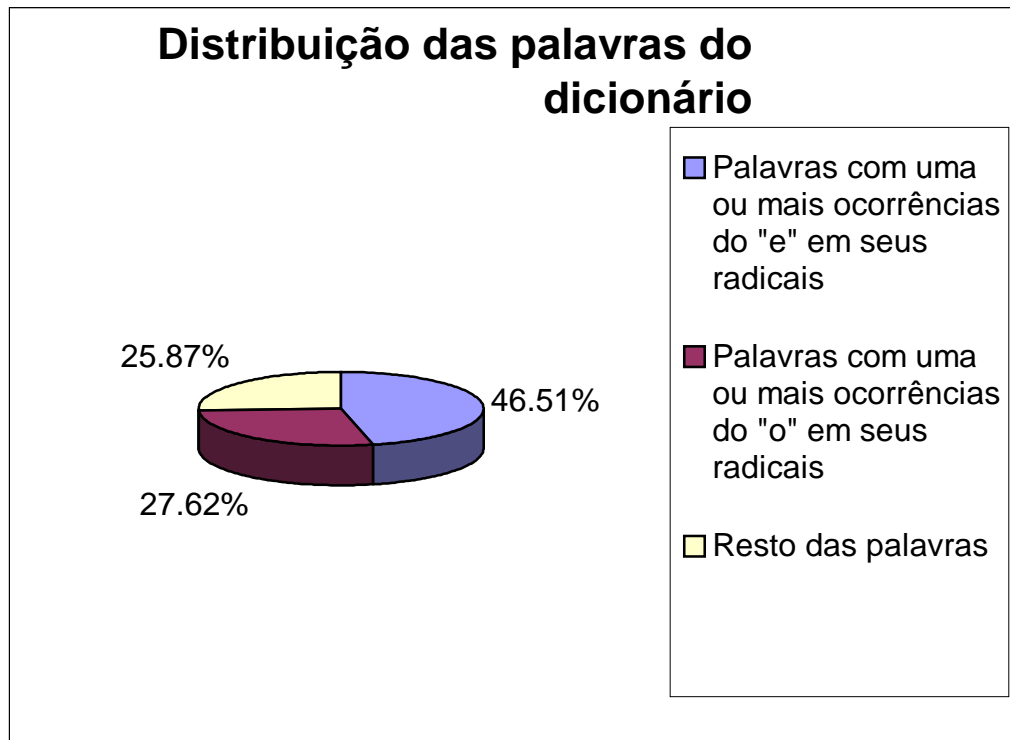


FIGURA 5.2 – Distribuição das palavras do dicionário

Dentro das 8.910 palavras com a letra “e”, ocorreram 9.870 casos com “e” fechado e 510 casos com “e” pronunciado como aberto, visto que pode ocorrer mais de uma ocorrência. Ou seja, 95.09% das palavras com “e” fechado e 4.91% das palavras com “e” aberto, conforme pode-se ver na figura 5.3.

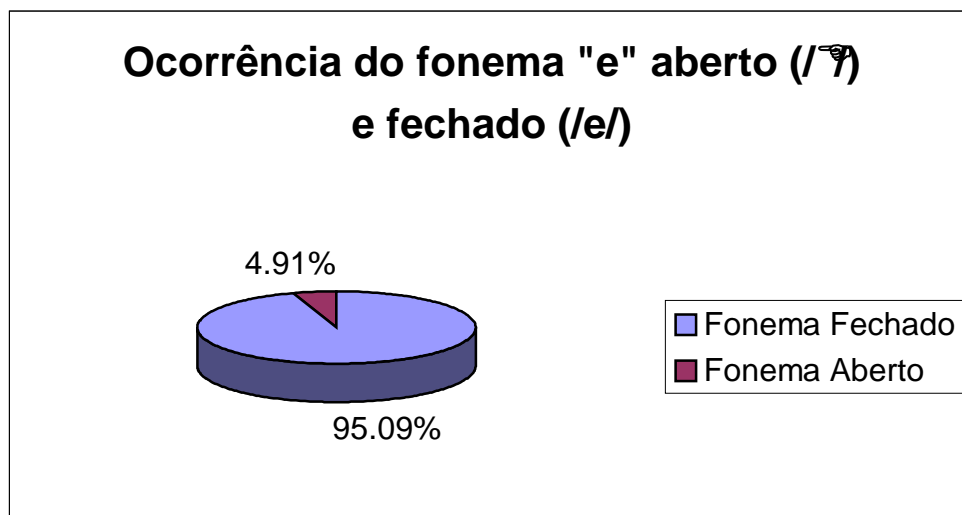


FIGURA 5.3 – Ocorrência do fonema “e” aberto (/E/) e fechado (/e/)

Dentro das 5.290 palavras com a letra “o” ocorreram 6.168 casos com “o” fechado e 36 casos com “o” aberto, visto que pode ocorrer mais de uma ocorrência. Ou seja, 99.42% das palavras com “o” fechado e 0.58% das palavras com “o” aberto, conforme pode-se ver na figura 5.4.

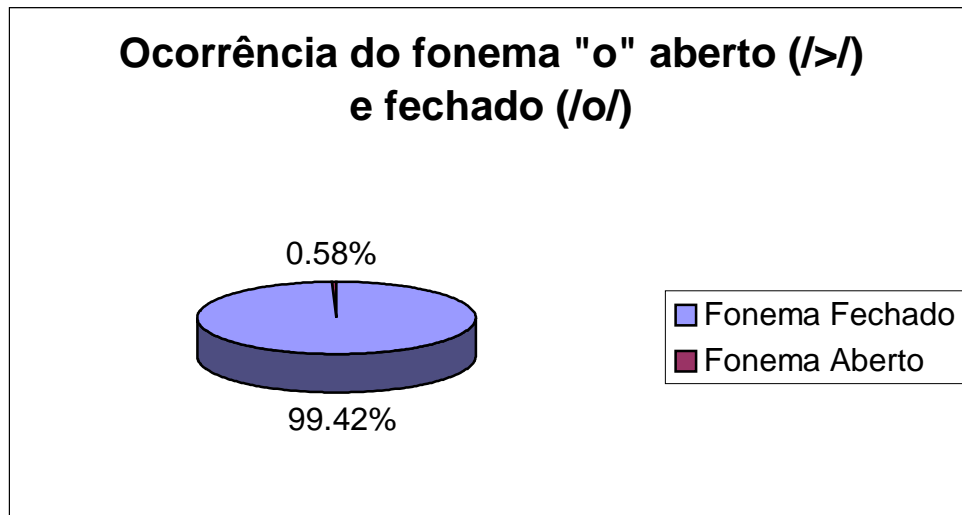


FIGURA 5.4 – Ocorrência do fonema “o” aberto (/>/) e fechado (/o/)

As regras estabelecidas por esta análise estatística para a correta pronúncia do fonema aberto das palavras com as vogais “e” e “o” por parte do Festival são apresentadas na tabela a seguir.

TABELA 5.7 – Regras estabelecidas pela análise estatística

Regra Nº	Regra	Total de Casos e/o fechado	Total de Casos e/o aberto
1	AES	0	1
2	BEF	0	1
3	HEQ	0	1
4	IEL	1	4
5	JEC	0	1
6	JEL	0	1
7	KEL	0	1
8	LEL	0	3
9	XEL	0	1
10	XEQ	0	1
11	DOQ	0	1
12	DOX	0	2

As palavras que geraram estas regras com a representação do fonema aberto estão listadas na tabela seguinte.

TABELA 5.8 – Palavras que geraram as regras

Regra Nº	Regra	Palavra
1	AES	Mae(/E/) stro
2	BEF	Tabe(/E/)fe
3	HEQ	Che(/E/)que
4	IEL	Bie(/E/)la
4	IEL	Fie(/E/)l

Palavras que geraram as regras
Continuação

Regra Nº	Regra	Palavra
4	IEL	Infie(/E/)l
4	IEL	Vie(/E/)la
5	JEC	Je(/E/)ca
6	JEL	Berinja(/E/)la
7	KEL	Ke(/E/)lvin
8	LEL	Bale(/E/)la
8	LEL	Parale(/E/)la
8	LEL	Parale(/E/)lo
9	XEL	Baixa(/E/)la
10	XEQ	Xe(/E/)que
11	DOQ	Bodo(>/)que
12	DOX	Ortodo(>/)xo
12	DOX	Parado(>/)xo

A regra número 7 (KEL) não será considerada na aplicação dos testes, pois a palavra que a gerou (kelvin) é de origem estrangeira; logo, não é um radical da língua portuguesa.

Para verificar a eficiência das regras criadas a partir da análise estatística, fez-se um levantamento das palavras do dicionário [SIL75] que possuíam as letras das regras em seus radicais. As palavras encontradas e suas transcrições fonéticas do “e” e “o” para cada regra no dicionário estão listadas nas tabelas a seguir, onde está assinalada a seqüência de letras (conforme regras) que formam os fonemas abertos ou fechados.

TABELA 5.9 – Lista de palavras usadas para o teste da regra 1 (AES)

Regra Nº	Palavra	Fonema
1	Maestria	[e]
1	Maestrina	[e]
1	Paraestatal	[e]
1	Maestroso	[e]
Total de Palavras	4	
Total de Fonemas Abertos	0	
Total de Fonemas Fechados	4	

TABELA 5.10 – Lista de palavras usadas para o teste da regra 2 (BEF)

Regra Nº	Palavra	Fonema
2	<i>Hebefrenia</i>	[e]
Total de Palavras	1	
Total de Fonemas Abertos	0	
Total de Fonemas Fechados	1	

TABELA 5.11 – Lista de palavras usadas para o teste da regra 3 (HEQ)

Regra Nº	Palavra	Fonema
3	<i>Pachequice</i>	[e]
3	<i>Tcheque</i>	[E]
Total de Palavras	2	
Total de Fonemas Abertos	1	
Total de Fonemas Fechados	1	

TABELA 5.12 – Lista de palavras usadas para o teste da regra 4 (IEL)

Regra Nº	Palavra	Fonema	Regra Nº	Palavra	Fonema
4	<i>Amielencefalia</i>	[e]	4	<i>mielocelo</i>	[E]
4	<i>Amielia</i>	[e]	4	<i>Mielocitemia</i>	[E]
4	<i>amieloneuria</i>	[e]	4	<i>Mielocitose</i>	[E]
4	<i>amielotrofia</i>	[e]	4	<i>mielodisplasia</i>	[E]
4	<i>Anielado</i>	[e]	4	<i>mieloganglíte</i>	[E]
4	<i>Anielagem</i>	[e]	4	<i>mielografia</i>	[E]
4	<i>Anielar</i>	[e]	4	<i>mielopatia</i>	[E]
4	<i>Coxiela</i>	[E]	4	<i>mieloplacoma</i>	[E]
4	<i>Crapiela</i>	[E]	4	<i>Mieloplasto</i>	[E]
4	<i>Desmielinar</i>	[e]	4	<i>mieloplaxe</i>	[E]
4	<i>Arriel</i>	[E]	4	<i>mieloplaxoma</i>	[E]
4	<i>Bertiela</i>	[E]	4	<i>mieloplegia</i>	[E]
4	<i>Bielo-russo</i>	[E]	4	<i>Mielorrafia</i>	[E]

Lista de palavras usadas para o teste da regra 4 (IEL)

Continuação

Regra Nº	Palavra	Fonema	Regra Nº	Palavra	Fonema
-----------------	----------------	---------------	-----------------	----------------	---------------

4	Cistopielite	[e]	4	Mielorragia	[E]
4	Cistopielografia	[e]	4	mielosclerose	[E]
4	Cistopielonefrite	[e]	4	mielose	[e]
4	Desmielinizar	[e]	4	Mielospasmo	[E]
4	Dieletrina	[e]	4	mielossarcoma	[E]
4	Espondilomielite	[e]	4	mielotoxina	[E]
4	Furriel	[E]	4	Nefropielite	[e]
4	Hematomielia	[e]	4	Nielo	[E]
4	hematomielite	[e]	4	Notomielite	[e]
4	hematomieloporose	[e]	4	Orquiomieloma	[e]
4	Hidromielia	[e]	4	Ostomielialgia	[e]
4	Idieletricidade	[e]	4	Ostomielite	[e]
4	Infieldade	[E]	4	Perielise	[e]
4	Ixomielite	[e]	4	Piela	[E]
4	Mielograma	[e]	4	Pielite	[e]
4	Mieloma	[e]	4	Pielocistite	[E]
4	mielomalacia	[E]	4	Pielonefrite	[E]
4	mielomatose	[E]	4	pielonefroze	[E]
4	Mielomenia	[E]	4	pielostomia	[E]
4	mielomeningite	[E]	4	Pielotrombose	[E]
4	Mieloneurite	[E]	4	pielotomia	[E]
4	Mieloparalisia	[E]	4	labielo	[E]
4	meningomielite	[e]	4	Platielminte	[e]
4	meningomielocele	[e]	4	platielminto	[e]
4	Mielalgia	[e]	4	radielemento	[e]
4	Mielanalose	[e]	4	Radieletricidade	[e]
4	Mielastenia	[e]	4	leucomielite	[e]
4	mielatelia	[e]	4	leucomielopatia	[e]
4	mielatrofia	[e]	4	samiel	[E]
4	mielauxesia	[e]	4	sediela	[E]
4	mielencefalite	[e]	4	Siringomielia	[e]
4	mielina	[e]	4	Trielco	[E]
4	mielinizado	[e]	4	Ureteropielite	[e]
4	mielite	[e]	4	Ureteropielografia	[e]
4	mieloblastemia	[e]	4	Ureteropielonefrite	[e]
4	mieloblasto	[E]	4	Ureteropieloneostomia	[e]
4	mieloblastoma	[E]	4	ureteropielostomia	[e]
4	mielocele	[E]	4		
Total de Palavras	101				
Total de Fonemas Abertos	46				
Total de Fonemas Fechados	55				

TABELA 5.13 – Lista de palavras usadas para o teste da regra 5 (JEC)

Regra Nº	Palavra	Fonema
5	Interjecional	[e]
5	Jecoral	[e]
5	Jecorina	[e]
5	Lojeca	[E]
Total de Palavras	4	
Total de Fonemas Abertos	1	
Total de Fonemas Fechados	3	

TABELA 5.14 – Lista de palavras usadas para o teste da regra 6 (JEL)

Regra Nº	Palavra	Fonema
6	Pajelança	[e]
6	Sarjel	[E]
Total de Palavras	2	
Total de Fonemas Abertos	1	
Total de Fonemas Fechados	1	

TABELA 5.15 – Lista de palavras usadas para o teste da regra 8 (LEL)

Regra Nº	Palavra	Fonema	Regra Nº	Palavra	Fonema
8	Paralelismo	[e]	8	Colelitotripsia	[e]
8	paralelogramo	[E]	8	Desparalelo	[E]
8	alelarga	[e]	8	Dialelo	[E]
8	Aleli	[E]	8	Filele	[E]

Lista de palavras usadas para o teste da regra 8 (LEL)

Continuação

Regra Nº	Palavra	Fonema	Regra Nº	Palavra	Fonema
8	Alelo	[E]	8	Filelenismo	[e]
8	alelognatia	[E]	8	Fileleno	[e]
8	alelomaquia	[E]	8	Metabolelogia	[e]
8	alelomorfia	[E]	8	Paralelepipedal	[e]
8	alelomorfismo	[E]	8	Paraleli	[E]
8	alelomorfo	[E]	8	Paralelismo	[e]

8	<i>alelotaxe</i>	[E]	8	<i>Paralelivenoso</i>	[e]
8	<i>alelotaxia</i>	[E]	8	<i>Paralelizar</i>	[e]
8	<i>aleluia</i>	[e]	8	<i>paralelocinesia</i>	[E]
8	<i>aleluiado</i>	[e]	8	<i>paralelografia</i>	[E]
8	<i>Aleluiar</i>	[e]	8	<i>paralelogramo</i>	[E]
8	<i>Aleluquia</i>	[e]	8	<i>Polela</i>	[E]
8	<i>Colelitolonia</i>	[e]	8	<i>Lele</i>	[E]
8	<i>Colelogia</i>	[e]	8	<i>vilela</i>	[E]
Total de Palavras	36				
Total de Fonemas Abertos	20				
Total de Fonemas Fechados	16				

TABELA 5.16 – Lista de palavras usadas para o teste da regra 9 (XEL)

Regra N°	Palavra	Fonema
9	<i>Afrouxelar</i>	[e]
9	<i>baixel</i>	[E]
9	<i>Frouxel</i>	[E]
9	<i>Frouxelado</i>	[e]
9	<i>Frouxeleiro</i>	[e]
9	<i>Mexelhar</i>	[e]
9	<i>Peixelim</i>	[e]
9	<i>Trouxel</i>	[E]
9	<i>Trouxelo</i>	[E]
9	<i>Vaxelo</i>	[E]
9	<i>Xelim</i>	[e]
9	<i>Xelma</i>	[E]
Total de Palavras	12	
Total de Fonemas Abertos	6	
Total de Fonemas Fechados	6	

TABELA 5.17 – Lista de palavras usadas para o teste da regra 10 (XEQ)

Regra N°	Palavra	Fonema
10	<i>Enxequetado</i>	[e]
10	<i>Enxequetar</i>	[e]
10	<i>Exequatur</i>	[e]

10	<i>Exeqüenda</i>	[e]
10	<i>Exeqüente</i>	[e]
10	<i>Exequial</i>	[e]
10	<i>Exeqüibilidade</i>	[e]
10	<i>Exeqüível</i>	[e]
10	<i>Xequear</i>	[e]
Total de Palavras	9	
Total de Fonemas Abertos	0	
Total de Fonemas Fechados	9	

TABELA 5.18 – Lista de palavras usadas para o teste da regra 11 (DOQ)

Regra Nº	Palavra	Fonema
11	<i>Adoquim</i>	[e]
11	<i>adoquina</i>	[e]
11	<i>bodoqueiro</i>	[e]
11	<i>doqueiro</i>	[e]
11	<i>rabadoquim</i>	[e]
11	<i>sialodoquite</i>	[e]
11	<i>xenodoquia</i>	[e]
Total de Palavras	7	
Total de Fonemas Abertos	0	
Total de Fonemas Fechados	0	

TABELA 5.19 – Lista de palavras usadas para o teste da regra 12 (DOX)

Regra Nº	Palavra	Fonema
12	<i>Ortodoxia</i>	[e]
12	<i>Paradoxal</i>	[e]
12	<i>Amidoxima</i>	[e]
12	<i>Adoxo</i>	[E]
12	<i>adoxografia</i>	[E]
12	<i>cacodoxia</i>	[e]
12	<i>cacodoxo</i>	[E]
12	<i>heterodoxia</i>	[e]
12	<i>doxografia</i>	[e]
12	<i>doxologia</i>	[e]
12	<i>Doxomania</i>	[e]

12	Filodoxia	[e]
12	Filodoxo	[E]
12	Heterodoxia	[e]
12	Heterodoxo	[E]
12	Inortodoxo	[E]
12	Paradoxal	[e]
12	Paradoxalidade	[e]
12	Paradoxar	[e]
12	Paradoxista	[e]
12	Paradoxita	[e]
12	Ortodoxia	[e]
12	Ortodoxografia	[e]
12	Paradoxuro	[e]
12	Pseudoxantina	[e]
12	Quelidoxatina	[e]
12	Rodoxantina	[e]
12	ultraparadoxal	[e]
Total de Palavras	28	
Total de Fonemas Abertos	6	
Total de Fonemas Fechados	22	

Usou-se a pronúncia da Região Sul para estabelecer a transcrição fonética das palavras encontradas no dicionário. Aplicando-se cada regra sobre o seu conjunto de palavras, pode-se verificar que as regras 3 (HEQ), 6 (JEL), 8 (LEL) e 9 (XEL) obtiveram resultados razoáveis, conforme pode ser visto na tabela a seguir.

TABELA 5.20 – Número de acertos e respectivo percentual

Regra N°	Regra	ACERTOS	%
3	HEQ	1	50
6	JEL	1	50
8	LEL	20	55.55
9	XEL	6	50

Conforme o resultado apresentado para a regra 3 (HEQ) na tabela 5.20, pode-se notar que o conjunto de palavras encontradas é pequeno, talvez devido ao fato desta seqüência de letras não possuir uma grande utilização nos radicais da língua portuguesa, o que torna a regra de certo modo eficiente, pois dos dois vocábulos encontrados no dicionário [SIL 75], de acordo com a tabela 5.11, obteve-se acerto de um dos verbetes (*tcheque*), já que a primeira ocorrência da vogal “e” na palavra é um fonema aberto (/E/). A segunda palavra (*pachequice*) possui fonema fechado (/e/) tanto na primeira quanto na segunda ocorrência da vogal “e” na palavra.

Com relação ao que foi dito sobre a regra 3, pode-se dizer o mesmo para o resultado da regra 6 (JEL), já que também foram encontrados dois verbetes no dicionário [SIL 75] para a mesma, conforme tabela 5.14. A regra foi eficiente para uma das palavras (*Sarjel*), já que a pronúncia da vogal “e” é um fonema aberto (/E/). Para a outra palavra (*Pajelança*) a vogal “e” é pronunciada como um fonema fechado (/e/).

De acordo com o resultado apresentado para a regra 8 (LEL) na tabela 5.20, pode-se notar que foi a que alcançou maior porcentagem de acertos de todas as regras testadas e, por conseqüência, a mais eficiente para a solução do problema enfrentado. Foram encontrados trinta e seis verbetes para a regra no dicionário [SIL 75], conforme a tabela 5.15. Dessas palavras, a regra foi eficiente para vinte palavras, as quais possuem a pronúncia da vogal “e” para a seqüência (regra) como um fonema aberto (/E/).

Baseado no resultado apresentado para a regra 9 (XEL) na tabela 5.20, pode-se notar que conseguiu-se o acerto de metade das palavras encontradas no dicionário [SIL 75], ou seja, dos 12 verbetes encontradas (tabela 5.16), apenas 6 possuíam fonema aberto para a vogal “e” da seqüência (regra).

Dos resultados apresentados para as regras do fonema “e” aberto (/E/), pode-se afirmar que apenas 40% das regras tornaram-se eficientes para a resolução da pronúncia automática do fonema em questão.

As regras encontradas para a solução do fonema “o” aberto (/>/) não obtiveram índices razoáveis de eficiência, devido ao fato de a ocorrência do mesmo fonema ser muito pequena dentro do dicionário utilizado, conforme pode-se verificar na figura 5.4.

6 Conclusão

É cada vez maior a necessidade do homem interagir de forma mais natural com o computador, ou genericamente, com as máquinas. Modos de entrada/saída naturais vêm sendo vislumbrados para prover esta comunicação, que acarreta em várias facilidades e vantagens. A fala irá certamente abrir muito mais as portas do mundo da computação e da tecnologia aos leigos, facilitando o acesso e manuseio de vários tipos de equipamentos computacionais[HUG 95].

Pensando em contribuir para o desenvolvimento da fala no âmbito da Pós-Graduação em Computação do Instituto de Informática da Universidade Federal do Rio Grande do Sul, resolveu-se explorar os conhecimentos da síntese de fala e desenvolver uma solução para um dos problemas enfrentados pela respectiva área na transcrição de um texto em língua portuguesa falada no Brasil para a fala. Mais especificamente, esta dissertação tem o intuito de contribuir para o sucesso do Projeto Spoltech - um projeto de cooperação entre os Estados Unidos e o Brasil para o avanço da tecnologia da língua falada no Brasil (Português Brasileiro).

O objetivo principal do trabalho foi criar regras de pronúncia automática para o Festival – ferramenta para a construção de sistemas de síntese de fala desenvolvida pelo *Centre for Speech Technology Research* (University of Edinburgh, Reino Unido) – com o intuito de solucionar o problema de pronúncia do “e” e “o”. O trabalho também inclui um estudo sobre a forma cognitiva de produção da fala até a sua representação fonêmica, uma descrição das principais técnicas de síntese de fala servindo assim como um suporte para futuros trabalhos na área e uma descrição geral do Projeto Spoltech, um projeto de cooperação entre os Estados Unidos e o Brasil para o avanço da tecnologia da língua falada no Brasil (Português Brasileiro).

O problema do “e” e do “o” resume-se na pronúncia correta dos mesmos, ou seja, determinar quando esses devem ser pronunciados com um fonema fechado (Por exemplo, Ge[*e*]/ntil, No[*o*]/rma) ou com um fonema aberto (Por exemplo, Fe[*E*]/bre, Bo[*>*]/la). Sabendo-se que o Festival gera fonemas fechados para todas palavras, exceto nos casos em que há uma regra que informe o contrário, resolveu-se criar novas regras para melhorar a pronúncia das palavras com fonema “e” e “o” aberto, /*E*/ e /*>*/ respectivamente.

Esta nova regra seria formada pelas letra anterior e a posterior ao “e” ou “o” de uma palavra. Por exemplo, considerando a palavra “tabefe” e sabendo-se que a primeira ocorrência da vogal “e” é um fonema aberto (/*E*/), pode-se dizer que uma nova regra para o fonema “e” aberto (/*E*/) seria “BEF”, ou seja, para o Festival todas as palavras com esta seqüência de letras seriam geradas com o fonema “e” aberto (/*E*/).

Para ser considerada uma regra válida teria que atingir 80% ou mais de ocorrência de casos dentro do dicionário (limiar). Conforme dito anteriormente, o Festival gera os fonemas para o “e” e “o” como fechados, se não há uma regra que imponha o contrário; logo, considerou-se para a solução do problema do “e” e “o” apenas as regras encontradas para os fonemas abertos.

Baseado no método descrito anteriormente, necessitava-se de um certo número de palavras para obter as regras. Logo, utilizou-se o dicionário eletrônico DIC Prático Michaelis versão 5.1 de julho de 1998, onde constavam 19.156 palavras e suas respectivas transcrições fonéticas para o Português Brasileiro. Separando-se por intermédio de um programa as palavras com “e” e com “o” em seus respectivos arquivos, aplicou-se um analisador estatístico sobre os mesmos para encontrar as regras para o problema do “e” e “o”.

Após a análise obteve-se 12 regras para tentar solucionar o problema de pronúncia do “e” e “o”. Para testar a eficiência destas regras fez-se um levantamento de palavras que continham as seqüências de letras no dicionário [SIL 75]. Utilizou-se a pronúncia da Região Sul para estabelecer a transcrição fonética do “e” e “o” das palavras do mesmo. Obtidas as palavras aplicou-se as regras sobre as mesmas e verificou-se que as regras 3 (HEQ), 6 (JEL), 8 (LEL) e 9 (XEL) foram as que alcançaram os melhores resultados para o fonema “e” aberto (/E/).

Pode-se notar que apenas 40% das regras criadas para o fonema “e” aberto (/E/) estão em níveis razoáveis de eficiência para o problema em questão. Já para o fonema “o” aberto (/>/), as regras estabelecidas pelo analisador estatístico não atingiram um percentual de acerto favorável.

Um fato que colaborou para as regras encontradas através da análise estatística não serem tão eficientes foi o tamanho do dicionário ou número de palavras utilizado (19.156 palavras) e, por conseqüência, no seu conjunto o número de palavras com “e” e “o” em seus radicais de 46.51% e 25.87%, respectivamente. Do percentual de palavras com “e” em seus radicais, apenas 4.91% são fonemas abertos. Já do total de palavras com “o” em seus radicais, somente 0.58% são fonemas abertos.

Em vista dos resultados alcançados, para perspectivas futuras, uma alternativa que poderia ser aplicada para tentar melhorar os resultados é determinar a classe gramatical ou gênero da palavra analisada (substantivo, verbo, etc), para após fazer uma análise da transcrição fonética do fonema “e” e “o”, utilizando-se para isto de um dicionário com um conjunto de verbetes significativamente maior.

Anexo 1 Exemplo do dicionário eletrônico

A seguir está relacionada a parte do dicionário referente a letra “a” utilizada na análise estatística deste trabalho e sua respectiva pronúncia, para dar uma amostra do formato do dicionário utilizado.

(a) (a)	(abismo) (abizmu)	(abrutalhado) (abrutaLadu)
(à) (a)	(abisso) (abisu)	(abrutado) (abrutadu)
(A, a) (a)	(abjeção) (abZesaxw~)	(abscesso) (absEsu)
(aba) (ab<)	(abjeto) (abZetu)	(abscissa) (absisa)
(abacate) (abakatS i)	(abjudicar) (abZudZ ikar)	(absolutamente) (absolutame~tS i)
(abacaxi) (abakaSi)	(abjuração) (abZurasaxw~)	(absoluto) (absolutu)
(abade) (abadZ i)	(abjurar) (abZurar)	(absolver) (absowver)
(abadia) (abadZ ia)	(ablação) (ablasaxw~)	(absorção) (absowvisaxw~)
(abafação) (abafasaxw~)	(ablativo) (ablatS ivu)	(absorção) (absorsaxw~)
(abafado) (abafadu)	(ablução) (ablusaxw~)	(absorto) (absortu)
(abafador) (abafador)	(abnegação) (abnegasaxw~)	(absorvente) (absorve~tS i)
(abafamento) (abafame~tu)	(abnegado) (abnegadu)	(absorver) (absorver)
(abafar) (abafar)	(abnegar) (abnegar)	(abstêmio) (abstemju)
(abaixar) (abajSar)	(abóbada) (ab>bada)	(abstenção) (abste~saxw~)
(abaixo) (abajSu)	(abobadado) (abobadadu)	(abstencionismo) (abste~sjonizmu)
(abajur) (abaZur)	(abobadar) (abobadar)	(abster) (abster)
(abalado) (abaladu)	(abobado) (abobadu)	(abstinência) (abstS ine~sja)
(abalamento) (abalame~tu)	(abóbora) (ab>bora)	(abstração) (abstrasaxw~)
(abalar) (abalar)	(aboboreira) (aboborejra)	(abstrair) (abstrair)
(abalizado) (abalizadu)	(abobrinha) (abobrin~a)	(abstrato) (abstratu)
(abalizar) (abalizar)	(abocanhar) (abokan~ar)	(absurdo) (absurdu)
(abalo) (abalu)	(aboletar) (aboletar)	(abundância) (abu~dax~sja)
(abalroar) (abawroar)	(abolição) (abolisaxw~)	(abundante) (abu~dax~tS i)
(abanador) (abanador)	(abolicionista) (abolisjonista)	(abundar) (abundar)
(abanar) (abanar)	(abolir) (abolir)	(abusado) (abuzadu)
(abandonado) (abax~donadu)	(abominar) (abominar)	(abusar) (abuzar)
(abandonar) (abax~donar)	(abominável) (abominavew)	(abusivo) (abuzivu)
(abandono) (abax~donu)	(abonado) (abonadu)	(abuso) (abuzu)
(abano) (abax~nu)	(abonançar) (abonax~sar)	(abutre) (abutri)
(abarcamento) (abarkame~tu)	(abonar) (abonar)	(acabado) (akabadu)
(abarcar) (abarkar)	(abono) (abonu)	(acabamento) (akabame~tu)
(abarracar) (abaxakar)	(abordagem) (abordaZej~)	(acabar) (akabar)
(abarroado) (abaxotadu)	(abordar) (abordar)	(acabrunhado) (akabrun~adu)
(abarroar) (abaxotar)	(aborígenes) (aboriZenis)	(acabrunhar) (akabrun~ar)
(abastado) (abastadu)	(aborrecer) (aboxeser)	(academia) (akademia)
(abastança) (abastax~sa)	(aborrecido) (aboxesidu)	(acadêmico) (akademiku)
(abastar) (abastar)	(aborrecimento) (aboxesime~tu)	(acafajestado) (akafaZestadu)
(abastardar) (abastardar)	(abortar) (abortar)	(açafração) (asafraxw~)
(abastecedor) (abastesedor)	(abortivo) (abortS ivu)	(acaipirado) (akajpiradu)
(abastecer) (abasteser)	(aborto) (abortu)	(acaju) (akaZu)
(abastecimento) (abastesime~tu)	(abotoadura) (abotoadura)	(acalcanhar) (akawkan~ar)
(abate) (abatS i)	(abotoar) (abotoar)	(acalantar) (akale~tar)
(abater) (abater)	(abraçar) (abrasar)	(acalmar) (akawmar)
(abatido) (abatS idu)	(abraço) (abrasu)	(acalorar) (akalorar)
(abatimento) (abatS ime~tu)	(abrandar) (abrax~dar)	(acamado) (akmadu)
(abaular) (abawlar)	(abranger) (abrax~Zer)	(acamar) (akamar)
(abaulamento) (abawlame~tu)	(abrasador) (abrazador)	(açambarcador) (asax~barkador)
(abdicação) (abdZ ikasaxw~)	(abrasante) (abrazax~tS i)	(açambarcar) (asax~barkar)
(abdicar) (abdZ ikar)	(abrasar) (abrazar)	(acampamento) (akax~pame~tu)
(abdome) (abdomi)	(abrasileirar) (abrazilejrar)	(acampar) (akax~par)
(abdômen) (abdomej~)	(abrasivo) (abrazivu)	(acanhado) (akan~adu)
(abdominal) (abdominaw)	(abreviação) (abreviousaxw~)	(acanhamento) (akan~ame~tu)
(abeceário) (abesedarju)	(abreviado) (abreviadu)	(acanhhar) (akan~ar)
(abeirar) (abejrar)	(abreviar) (abreviar)	(acanto) (akax~tu)
(abelha) (abeLa)	(abreviatura) (abrevjatura)	(ação) (asaxw~)
(abelhão) (abeLaxw~)	(abricó) (abrik>)	(acareação) (akareasaxw~)
(abelhudice) (abeLudZ isi)	(abridor) (abridor)	(acarear) (akarear)
(abelhudo) (abeLudu)	(abrigado) (abrigadu)	(acariciar) (akarisiar)
(abençoar) (abe~soar)	(abrigar) (abrigar)	(acarinhar) (akarín~ar)
(aberração) (abexasaxw~)	(abrigo) (abrigu)	(acarretar) (akaxetar)
(aberto) (abErtu)	(abril) (abriw)	(acasalar) (akazalar)
(abertura) (abertura)	(abrilhantar) (abriLax~tar)	(acaso) (akazu)
(abespinhar) (abespin~ar)	(abrir) (abrir)	(acatamento) (akatame~tu)
(abeto) (abetu)	(ab~rogar) (abxogar)	(acatar) (akatar)
(abismado) (abizmadu)	(ab~rogativo) (abxogatS ivu)	(acautelado) (akawteladu)
(abismal) (abizmaw)	(ab~rogatório) (abxogat>rju)	(acautelar) (akawtelar)
(abismar) (abizmar)	(abrolho) (abroLu)	(aceder) (aseder)

(aceitação) (asejtasaxw~)	(aconchegar) (ako~Segar)	(adestrar) (adestrar)
(aceitar) (asejtar)	(aconchego) (ako~Segu)	(adeus) (adews)
(aceitável) (asejtavew)	(acondicionado) (ako~dZ isjonadu)	(adiamento) (adjame~tu)
(aceito) (asejtu)	(acondicionar) (ako~dZ isjonar)	(adiantado) (adjax~tadu)
(aceleração) (aselerasaxw~)	(aconselhar) (ako~seLar)	(adiantamento) (adjax~tame~tu)
(acelerador) (aselerador)	(aconselhável) (ako~seLavew)	(adiantar) (adjax~tar)
(acelerar) (aselerar)	(acontecer) (ako~teser)	(adiante) (adjax~tS i)
(acenar) (asenar)	(acontecido) (ako~tesidu)	(adiar) (adZ iar)
(acendedor) (ase~dedor)	(acontecimento) (ako~tesime~tu)	(adição) (adZ isaxw~)
(acender) (ase~der)	(acordado) (akordadu)	(adicional) (adZ isjonaw)
(aceno) (asenu)	(acórdão) (ak>rdaxw~)	(adicionar) (adZ isjonar)
(acento) (ase~tu)	(acordar) (akordar)	(adido) (adZ idu)
(acentuação) (ase~twasaxw~)	(acorde) (ak>rdZ i)	(aditamento) (adZ itame~tu)
(acentuar) (ase~tuar)	(acordeão) (akordeaxw~)	(adivinhação) (adZ ivin~asaxw~)
(acepção) (asepsaxw~)	(acordo) (akordu)	(adivinhar) (adZ ivin~ar)
(acepipe) (asepipi)	(acorrentar) (akoxe~tar)	(adjacência) (adZase~sja)
(acerbar) (aserbar)	(acorrer) (akoxer)	(adjacente) (adZase~tS i)
(acerbo) (aserbu)	(acortinar) (akortS inar)	(adjetivo) (adZetS ivu)
(acerca) (aserka)	(acossar) (akosar)	(adjudicar) (adZudZ ikar)
(acercar) (aserkar)	(acostumado) (akostumadu)	(adjunto) (adZu~tu)
(acertar) (asertar)	(acostumar) (akostumar)	(administração) (administrasaxw~)
(acerto) (asertu)	(acotovelar) (akotovelar)	(administrador) (administrador)
(acervo) (aservu)	(açougue) (asowgi)	(administrar) (administrar)
(aceso) (asezu)	(açougueiro) (asowgejru)	(admiração) (admirasaxw~)
(acessível) (asesivew)	(acre) (akri)	(admirador) (admirador)
(acesso) (asEsu)	(acreditar) (akredZ itar)	(admirar) (admirar)
(acessório) (ases>rju)	(acrescentar) (akrese~tar)	(admirável) (admiravew)
(acetinado) (asetS inadu)	(acrescer) (akreser)	(admissão) (admisaxw~)
(acetona) (asetona)	(acrésimo) (akrEsimu)	(admitir) (admitir)
(acha) (aSa)	(acrimônia) (akrimonja)	(admoestação) (admoestasaxw~)
(achacar) (aSakar)	(acrobacia) (akrobasia)	(admoestar) (admoestar)
(achado) (aSadu)	(acrobata) (akrobata)	(admoestação) (admonisaxw~)
(achaque) (aSaki)	(acróbata) (akr>bata)	(adoção) (adosaxw~)
(achar) (aSar)	(acrobático) (akrobatS iku)	(adoçar) (adosar)
(achatado) (aSatadu)	(acrônimo) (akronimu)	(adocicado) (adosikadu)
(achatar) (aSatar)	(acróstico) (akr>stS iku)	(adocicar) (adosikar)
(achegar) (aSegar)	(acuado) (akuadu)	(adoecer) (adoeser)
(achincalhar) (aSi~kaLar)	(acuar) (akuar)	(adoentado) (adoe~tadu)
(acidentado) (aside~tadu)	(açúcar) (asukar)	(adolescência) (adolese~sja)
(acidental) (aside~taw)	(açucarado) (asukaradu)	(adolescente) (adolese~tS i)
(acidental) (aside~tar)	(açucarar) (asukarar)	(adoração) (adorasaxw~)
(acidente) (aside~tS i)	(açucareiro) (asukarejru)	(adorador) (adorador)
(acidez) (asides)	(açude) (asudZ i)	(adorável) (adoravew)
(acidificar) (asidZ ifikar)	(acudir) (akudZ ir)	(adormecer) (adormeser)
(ácido) (asidu)	(acuidade) (akujdadZ i)	(adornar) (adornar)
(acima) (asima)	(açular) (asular)	(adorno) (adornu)
(acinte) (asi~tS i)	(acumulação) (akumulaxw~)	(adotar) (adotar)
(acintoso) (asi~tozu)	(acumulador) (akumulador)	(adquirir) (adquirir)
(acionar) (asjonar)	(acumular) (akumular)	(aduação) (aduax~na)
(acionista) (asjonista)	(acúmulo) (akumulu)	(aduaneiro) (aduanejru)
(aclamação) (aklamasaxw~)	(acurado) (akuradu)	(adubação) (adubasaxw~)
(aclamar) (aklamar)	(acurar) (akurar)	(adubar) (adubar)
(aclamar) (aklarar)	(acurralar) (akuxalar)	(adubo) (adubu)
(aclimação) (aklimasaxw~)	(acusação) (akuzasaxw~)	(adulação) (adulasaxw~)
(aclimar) (aklimar)	(acusado) (akuzadu)	(adulador) (adulador)
(aclimatar) (aklimatar)	(acusar) (akuzar)	(adular) (adular)
(aclive) (aklivi)	(acusativo) (akuzatS ivu)	(adúltera) (aduwtera)
(aço) (asu)	(acústica) (akustS ika)	(adulterador) (aduwterador)
(acobertado) (akobertadu)	(acústico) (akustS iku)	(adulterar) (aduwterar)
(acobertar) (akobertar)	(adaga) (adaga)	(adultério) (aduwtErju)
(acocorar) (akokorar)	(adágio) (adaZju)	(adúltero) (aduwteru)
(açodar) (asodar)	(adaptabilidade) (adaptabilidadZ i)	(adulto) (aduwtu)
(açoitamento) (asojtame~tu)	(adaptação) (adaptasaxw~)	(adunco) (adu~ku)
(acoitar) (akojtar)	(adaptar) (adaptar)	(adutora) (adutora)
(açoiar) (asojtar)	(adaptável) (adaptavew)	(aduzir) (aduzir)
(açuite) (asojtS i)	(adega) (adEga)	(advento) (adve~tu)
(acolá) (akola)	(adelgaçar) (adewgasar)	(advérbio) (advErbju)
(acolchoar) (akowSoar)	(adentro) (ade~tru)	(adversário) (adversarju)
(acolhedor) (akoLedor)	(adepito) (adEptu)	(adversidade) (adversidadZ i)
(acolher) (akoLer)	(adequado) (adekwadu)	(adverso) (advErsu)
(acolhida) (akoLida)	(adequar) (adekwar)	(advertência) (adverte~sja)
(acolhimento) (akoLime~tu)	(adereçar) (aderesar)	(advertir) (advertS ir)
(acometer) (akometer)	(aderência) (adere~sja)	(advir) (advir)
(acomodação) (akomodasaxw~)	(aderente) (adere~tS i)	(advocacia) (advokasia)
(acomodado) (akomodadu)	(aderir) (aderir)	(advogado) (advogadu)
(acomodar) (akomodar)	(adernar) (adernar)	(advogar) (advogar)
(acompanhamento) (ako~pan~ame~tu)	(adesão) (adezaxw~)	(aéreo) (aErju)
(acompanhante) (ako~pan~ax~tS i)	(adesivo) (adezivu)	(aerodinâmico) (aErodZ inax~miku)
(acompanhar) (ako~pan~ar)	(adestrado) (adestradu)	(aeromoço) (aEromosu)

(aeronauta) (aEronawta)	(afugentar) (afuZe~tar)	(agudeza) (agudeza)
(aeronáutica) (aEronawtS ika)	(afundamento) (afu~dame~tu)	(agudo) (agudu)
(aeronáutico) (aEronawtS iku)	(afundar) (afu~dar)	(agüentar) (agwe~tar)
(aeronave) (aEronavi)	(afunilado) (afuniladu)	(água) (agia)
(aeroplano) (aEroplax~nu)	(agachar-se) (agaSarsi)	(agulhão) (agiLaxw~)
(aerporto) (aEroportu)	(ágape) (agapi)	(aguilhoada) (agiLoada)
(aeroviário) (aErov iarju)	(agarrado) (agaxadu)	(aguilhoar) (agiLoar)
(afã) (afax~)	(agarramento) (agaxame~tu)	(agulha) (aguLa)
(afabilidade) (afabilidadZ i)	(agarrar) (agaxar)	(ai) (aj)
(afagar) (afagar)	(agasalhar) (agazaLar)	(aí) (ai)
(afago) (afagu)	(agasalho) (agazaLu)	(aidético) (ajdEtS iku)
(afamado) (afamadu)	(agastar) (agastar)	(aids) (ajdz)
(afasia) (afazia)	(ágata) (agata)	(ainda) (ai~da)
(afastado) (afastadu)	(agência) (aZe~sja)	(aipim) (ajpi~)
(afastamento) (afastame~tu)	(agenciar) (aZe~siar)	(aipo) (ajpu)
(afastar) (afastar)	(agenda) (aZe~da)	(airoso) (ajrozu)
(afável) (afavew)	(agente) (aZe~tS i)	(ajaezar) (aZaezar)
(afazeres) (afazeris)	(agigantar) (aZigax~tar)	(ajantarado) (aZax~taradu)
(afeição) (afejsaxw~)	(ágil) (aZiw)	(ajardinado) (aZardZ inadu)
(afeiçoar) (afejsoar)	(agilidade) (aZilidaz i)	(ajardinar) (aZardZ inar)
(afeminado) (afeminadu)	(agilizar) (aZilizar)	(ajeitar) (aZejtar)
(aferição) (aferisaxw~)	(ágio) (aZiw)	(ajoelhar) (aZoeLar)
(aferido) (aferidu)	(agiota) (aZi>ta)	(ajuda) (aZuda)
(aferir) (aferir)	(agiotagem) (aZjotaZej~)	(ajudante) (aZudax~tS i)
(aferrolhar) (afexoLar)	(agiotar) (aZjotar)	(ajudar) (aZudar)
(afetação) (afetasaxw~)	(agir) (aZir)	(ajuizado) (aZuizadu)
(afetado) (afetadu)	(agitação) (aZitasaxw~)	(ajuizar) (aZuizar)
(afetar) (afetar)	(agitador) (aZitador)	(ajuntamento) (aZu~tame~tu)
(afetivo) (afetS ivu)	(agitar) (aZitar)	(ajuntar) (aZu~tar)
(afeto) (afEtU)	(aglomeração) (aglomeraxw~)	(ajuramentar) (aZurame~tar)
(afetuoso) (afetuoZu)	(aglomerar) (aglomerar)	(ajustagem) (aZustazej~)
(afiado) (afiadu)	(aglutinar) (aglutS inar)	(ajustar) (aZustar)
(afiador) (afiador)	(agonia) (agonia)	(ajustável) (aZustavew)
(afiançar) (afi ax~sar)	(agoniar) (agoniar)	(ajuste) (aZustS i)
(afiar) (afiar)	(agonizante) (agonizax~tS i)	(ala) (ala)
(afigurar) (afigurar)	(agonizar) (agonizar)	(alabastro) (alabastru)
(afilhada) (afilada)	(agora) (ag>ra)	(alacridade) (alakridadZ i)
(afilhado) (afiladu)	(agosto) (agostu)	(alado) (aladu)
(afiliação) (afiliasaxw~)	(agourar) (agowrar)	(alagadiço) (alagadZ isu)
(afiliado) (afiliar)	(agourento) (agowre~tu)	(alagado) (alagadu)
(afim) (afi~)	(agouro) (agowru)	(alagar) (alagar)
(afinado) (afinadu)	(agradar) (agradar)	(alambique) (alax~biki)
(afinal) (afinaw)	(agradável) (agradavew)	(alambrar) (alax~brar)
(afinar) (afinar)	(agradecer) (agradeser)	(alameda) (alameda)
(afinco) (afi~ku)	(agradecido) (agradesidu)	(álamo) (alamu)
(afinidade) (afinidaz i)	(agradecimento) (agradesime~tu)	(alar) (alar)
(afirmação) (afirmasaxw~)	(agrado) (agradu)	(alaranjado) (alarax~Zadu)
(afirmar) (afirmar)	(agrário) (agrarju)	(alarde) (alardZ i)
(afirmativa) (afirmatS iva)	(agravação) (agrasaxw~)	(alardeador) (alardeador)
(afirmativo) (afirmatS ivu)	(agravar) (agravar)	(alardear) (alardear)
(afivelar) (afivelar)	(agravo) (agravu)	(alargamento) (alargame~tu)
(afixação) (afiksasaxw~)	(agredir) (agredZ ir)	(alargar) (alargar)
(afixar) (afiksar)	(agregação) (agregasaxw~)	(alarido) (alaridu)
(afixo) (afiksu)	(agregado) (agregadu)	(alarmante) (alarmax~tS i)
(aflicção) (aflisaxw~)	(agregar) (agregar)	(alarmar) (alarmar)
(afligir) (afliZir)	(agremiação) (agremjasaxw~)	(alarme) (alarmi)
(aflito) (aflitu)	(agressão) (agresaxw~)	(alastramento) (alastrame~tu)
(aflorar) (aflorar)	(agressivo) (agresivu)	(alastrar) (alastrar)
(afluência) (aflue~sja)	(agressor) (agresor)	(alaúde) (alaudZ i)
(afluente) (aflue~tS i)	(agreste) (agrEstS i)	(alavanca) (alavax~ka)
(afluir) (afluir)	(agrião) (agriaxw~)	(alazão) (alazaxw~)
(afobação) (afobasaxw~)	(agrícola) (agrikola)	(albatroz) (awbatr>s)
(afobado) (afobadu)	(agricultor) (agrikuwtor)	(albergue) (awbErgi)
(afobar) (afobar)	(agricultura) (agrikuwtura)	(álbum) (awbu~)
(afogado) (afogadu)	(agridoce) (agridosi)	(alça) (awsa)
(afogador) (afogador)	(agrilhoar) (agriLoar)	(alcachofra) (awkaSofra)
(afogar) (afogar)	(agrimensor) (agrime~sor)	(alcaçuz) (awkasus)
(afoiteza) (afojteza)	(agronomia) (agronomia)	(alçada) (awsada)
(afoito) (afojtu)	(agrônomo) (agronomu)	(álcali) (awkali)
(afora) (af>ra)	(agrupamento) (agrupame~tu)	(alcalóide) (awkal>jdZ i)
(afortunado) (afortunadu)	(agrupar) (agrupar)	(alcançar) (awkax~sar)
(afrancesar) (afrax~sezar)	(água) (agwa)	(alcance) (awkax~si)
(africano) (afrikax~nu)	(aguaceiro) (agwasejru)	(alcano) (awkax~nu)
(afronta) (afro~ta)	(aguado) (agwadu)	(alcantiloso) (awkax~tS ilozu)
(afrontado) (afro~tadu)	(aguar) (agwar)	(alçapão) (awsapaxw~)
(afrontamento) (afro~tame~tu)	(aguardar) (agwardar)	(alcaparra) (awkapara)
(afrontar) (afro~tar)	(aguardente) (agwarde~tS i)	(alçar) (awsar)
(afrouxamento) (afrowSame~tu)	(aguarrás) (agwaxas)	(alcatéia) (awkatEja)
(afrouxar) (afrowSar)	(aguçar) (agusar)	(alcatrão) (awkatraxw~)

(alcatraz) (awkatras)	(álibi) (alibi)	(altitude) (awtS itudZ i)
(alcatroar) (awkatroar)	(alicate) (alikatS i)	(altivez) (awtS ives)
(alce) (awsi)	(alicerçar) (alisersar)	(altivo) (awtS ivu)
(álcool) (awko>w)	(alicerce) (alisersi)	(alto) (awtu)
(alcoólatra) (awko>latra)	(aliciação) (alijasaxw~)	(altruísmo) (awtruizmu)
(alcoólico) (awko>liku)	(aliciar) (alisiar)	(altruísta) (awtruísta)
(alcooolismo) (awkolizmu)	(alienação) (aljenasaxw~)	(altura) (awtura)
(alcooolizar) (awkolizar)	(alienado) (aljenadu)	(aluado) (aluadu)
(alcorão) (awkoraxw~)	(alienar) (aljenar)	(alucinação) (alusinasaxw~)
(alcova) (awkova)	(alienígena) (aljeniZena)	(alucinar) (alusinar)
(alcovitar) (awkovitar)	(alijamento) (aliZame~tu)	(aludir) (aludZ ir)
(alcoviteiro) (awkovitejru)	(alijar) (alizar)	(alugar) (alugar)
(alcunha) (awkun~a)	(alimentação) (alime~tasaxw~)	(aluguel) (alugEw)
(alcunhar) (awkun~ar)	(alimentar) (alime~tar)	(alumiar) (alumiar)
(aldeão) (awdeaxw~)	(alimentício) (alime~tS isju)	(alumínio) (aluminju)
(aldeia) (awdeja)	(alimento) (alime~tu)	(aluna) (aluna)
(aldrava) (awdrava)	(alínea) (alinja)	(aluno) (alunu)
(aldravar) (awdravar)	(alinhado) (alin~adu)	(alusão) (aluzaxw~)
(aleatório) (aleat>rju)	(alinhamento) (alin~ame~tu)	(alva) (awva)
(alecrim) (alekri~)	(alinhar) (alin~ar)	(alvaiade) (awvjadZ i)
(alegação) (alegasaxw~)	(alinhar) (alin~ar)	(alvar) (awvar)
(alegar) (alegar)	(alinhar) (alin~ar)	(alvará) (awvara)
(alegoria) (alegoria)	(alinhar) (alin~ar)	(alvejar) (awveZar)
(alegrar) (alegrar)	(alinhar) (alin~ar)	(alvenaria) (awvenaria)
(alegre) (alEgri)	(alinhar) (alin~ar)	(alvéolo) (awvEolu)
(alegria) (alegria)	(alinhar) (alin~ar)	(alvissareiro) (awvisarejru)
(aleijado) (alejZadu)	(alinhar) (alin~ar)	(alvitre) (awvitri)
(aleijão) (alejZaxw~)	(alinhar) (alin~ar)	(alvo) (awvu)
(aleijar) (alejZar)	(aliviar) (aliviar)	(alvorada) (awvorada)
(aleitar) (alejtar)	(alívio) (alivju)	(alvorecer) (awvoreser)
(além) (alej~)	(alma) (awma)	(alvorçado) (awvorosadu)
(alemão) (alemawxw~)	(almaço) (awmasu)	(alvorçamento) (awvorosame~tu)
(alentar) (ale~tar)	(almanaque) (awmanaki)	(alvorçoar) (awvorosar)
(alento) (ale~tu)	(almeirão) (awmejraxw~)	(alvorço) (awvorosu)
(alergia) (alerZia)	(almejar) (awmeZar)	(ama) (ax~ma)
(alérgico) (alErZiku)	(almirante) (awmirax~tS i)	(amabilidade) (amabilidadZ i)
(alerta) (alErta)	(almíscar) (awmiskar)	(amaciar) (amasiar)
(alertar) (alertar)	(almoçar) (awmosar)	(amada) (amada)
(aletria) (aletria)	(almoço) (awmosu)	(amado) (amadu)
(alfabetar) (awfabetar)	(almofada) (awmofada)	(amador) (amador)
(alfabetizar) (awfabetS izar)	(almofariz) (awmofaris)	(amadurecer) (amadureser)
(alfabeto) (awfabEtu)	(almôndega) (awmo~dega)	(amadurecimento) (amaduresime~tu)
(alface) (awfasi)	(almoxarife) (awmoSarifi)	(âmagu) (ax~magu)
(alfaiataria) (awfajataria)	(alô) (alo)	(amainar) (amajnar)
(alfaiate) (awfajats i)	(alocar) (alokar)	(amaldiçoar) (amawdZ isoar)
(alfândega) (awfax~dega)	(alocução) (alokusaxw~)	(amálgama) (amawgama)
(alfandegário) (awfax~degarju)	(alojamento) (aloZame~tu)	(amalgamar) (amawgamar)
(alfarrábio) (awfaxabju)	(alojar) (aloZar)	(amalucado) (amalukadu)
(alfazema) (awfazema)	(alongamento) (alo~game~tu)	(amamentação) (amame~tasaxw~)
(alferes) (awfEris)	(alongar) (alo~gar)	(amamentar) (amame~tar)
(alfinetada) (awfinetada)	(alopatia) (alopatS ia)	(amanhã) (aman~ax~)
(alfinete) (awfinetS i)	(alocado) (alokadu)	(amanhecer) (aman~eser)
(alforje) (awf>rZi)	(alourar) (alowrar)	(amansar) (amax~sar)
(alforria) (awfoxia)	(alpaca) (awpaka)	(amante) (amax~tS i)
(alga) (awga)	(alpargata) (awpargata)	(amar) (amar)
(algarismo) (awgarizmu)	(alpendre) (awpe~dri)	(amarelão) (amarelaxw~)
(algazarra) (awgazaxa)	(alpinista) (awpinista)	(amarelar) (amarelar)
(álgebra) (awZebra)	(alpiste) (awpistS i)	(amarelecer) (amarelecer)
(algema) (awZema)	(alquebramento) (awkebrame~tu)	(amarelo) (amarElu)
(algemar) (awZemar)	(alquebrar) (awkebrar)	(amarfanhar) (amarfan~ar)
(algibeira) (awZibejra)	(alquimista) (awkimista)	(amargar) (amargar)
(álgido) (awZidu)	(alta) (awta)	(amargo) (amargu)
(algo) (awgu)	(altaneiro) (awtanejru)	(amargor) (amargor)
(algodão) (awgodaxw~)	(altar) (awtar)	(amargura) (amargura)
(algoritmo) (awgoritmu)	(altear) (awtear)	(amargarur) (amargarur)
(algez) (awgos)	(alteração) (awterasaxw~)	(amarra) (amara)
(alguém) (awgej~)	(alterado) (awteradu)	(amarracão) (amaxasaw)
(algum) (awgu~w)	(alterar) (awterar)	(amarrado) (amaxadu)
(algueres) (awguris)	(alteração) (awterkasaxw~)	(amarrar) (amaxar)
(alheamento) (aLeame~tu)	(altercar) (awterkar)	(amarrotar) (amaxotar)
(alhear) (aLear)	(alternação) (awternasaxw~)	(amásia) (amazja)
(alheio) (aLeju)	(alternado) (awternadu)	(amásio) (amazju)
(alho) (aLu)	(alternar) (awternar)	(amassar) (amasar)
(alhures) (aLuris)	(alternativa) (awternatS iva)	(amável) (amavew)
(ali) (ali)	(alturoso) (awterozu)	(âmbar) (ax~bar)
(aliado) (aliadu)	(alteza) (awteza)	(ambição) (ax~bisaxw~)
(aliança) (alix~sa)	(altibaixos) (awtS ibajSus)	(ambicionar) (ax~bisjonar)
(aliar) (aliar)	(altiplano) (awtS iplax~nu)	(ambicioso) (ax~bisjozu)
(aliás) (alias)	(Altíssimo) (awtS isimu)	(ambidestro) (ax~bidestru)
	(altista) (awtS ista)	

(ambientar) (ax~bje~tar)	(amplificar) (ax~plifikar)	(angu) (ax~gu)
(ambiente) (ax~bie~tS i)	(amplitude) (ax~plitudZ i)	(angular) (ax~gular)
(ambigüidade) (ax~bigwidadZ i)	(amplo) (ax~plu)	(ângulo) (ax~gulu)
(ambíguo) (ax~bigwu)	(ampola) (ax~pola)	(anguloso) (ax~gulozu)
(âmbito) (ax~bitu)	(ampluheta) (ax~puLeta)	(angústia) (ax~gustja)
(ambos) (ax~bus)	(amputação) (ax~putasaxw~)	(angustiado) (ax~gustS iadu)
(ambrósia) (ax~br>zja)	(amputar) (ax~putar)	(angustiante) (ax~gustS iax~tS i)
(ambulância) (ax~mbulax~sja)	(amuar) (amuar)	(angustiar) (ax~gustS iar)
(ambulatório) (ax~bulat>rju)	(amuleto) (amuletu)	(aniagem) (aniaZej~)
(ameaça) (ameasa)	(amuio) (amuio)	(anil) (aniw)
(ameaçador) (ameasador)	(amurada) (amurada)	(animação) (animasaxw~)
(ameaçar) (ameasar)	(anacronismo) (anakronizmu)	(animado) (animadu)
(ameaço) (ameasu)	(anáguas) (anagwa)	(animal) (animaw)
(amedrontado) (amedro~tadu)	(anais) (anajs)	(animalizar) (animalizar)
(amedrontar) (amedro~tar)	(anal) (anaw)	(animar) (animar)
(ameixa) (amejSa)	(analfabetismo) (anawfabetS izmu)	(ânimo) (ax~nimu)
(amém) (amej~)	(analfabeto) (anawfabEtu)	(animosidade) (animozidadZ i)
(amêndoa) (ame~dwa)	(analgésico) (anawZEziku)	(animoso) (animozu)
(amendoim) (ame~doi~)	(analizador) (analizador)	(aninhar) (anin~ar)
(amenidade) (amenidadZ i)	(analisar) (anализar)	(aniquilação) (anikilasaxw~)
(amenizar) (amenizar)	(análise) (analizi)	(aniquilado) (anikiladu)
(ameno) (amenu)	(analista) (analista)	(aniquilar) (anikilar)
(americanizar) (amerikanizar)	(analítico) (analitS iku)	(anis) (anis)
(americano) (amerikax~nu)	(analogia) (analoZia)	(anistia) (anistS ia)
(amerissar) (amerisar)	(analógico) (anal>Ziku)	(anistiar) (anistS iar)
(amesquinhar) (ameskin~ar)	(análogo) (analogu)	(aniversariante) (aniversariax~tS i)
(amestrar) (amestrar)	(ananás) (ananas)	(aniversariar) (aniversariar)
(ametista) (ametS ista)	(anão) (anaxw~)	(aniversário) (aniversarju)
(amianto) (amiax~tu)	(anarquia) (anarkia)	(anjo) (ax~Zu)
(amido) (amidu)	(anarquismo) (anarkizmu)	(ano) (ax~nu)
(amiga) (amiga)	(anarquista) (anarkista)	(anódio) (an>dju)
(amigaço) (amigasaxw~)	(anarquizar) (anarkizar)	(ânodo) (ax~nodu)
(amigado) (amigadu)	(anátoma) (anatema)	(anoitecer) (anojtesser)
(amigar) (amigar)	(anatomia) (anatomia)	(anomalia) (anomalia)
(amigável) (amigavew)	(anatomista) (anatomista)	(anonimato) (anonimatu)
(amígdala) (amigdala)	(anatomizar) (anatomizar)	(anônimo) (anonimu)
(amigo) (amigu)	(anavalhar) (anavaLar)	(anormal) (anormaw)
(amistoso) (amistozu)	(anca) (ax~ka)	(anormalidade) (anormalidadZ i)
(amiudado) (amiudadu)	(ancestral) (ax~sestraw)	(anotação) (anotasaxw~)
(amiúde) (amiudZ i)	(ancestre) (ax~sEstri)	(anotar) (anotar)
(amizade) (amizadZ i)	(ancião) (ax~siaxw~)	(anseio) (ax~seju)
(amo) (ax~mu)	(ancinho) (ax~sin~u)	(ânsia) (ax~sja)
(amodorrar) (amodoxar)	(âncora) (ax~kora)	(ansiar) (ax~siar)
(amofinação) (amofinasaxw~)	(ancoradouro) (ax~koradowru)	(ansiedade) (ax~sjedadZ i)
(amofinar) (amofinar)	(ancorar) (ax~korar)	(ansioso) (ax~siozu)
(amolação) (amolasaxw~)	(andaime) (ax~dax~jmi)	(anta) (ax~ta)
(amolado) (amoladu)	(andamento) (ax~dame~tu)	(antagonismo) (ax~tagonizmu)
(amolador) (amolador)	(andante) (ax~dax~tS i)	(antagonista) (ax~tagonista)
(amolar) (amolar)	(andar) (ax~dar)	(antagonizar) (ax~tagonizar)
(amoldar) (amowdar)	(andarilho) (ax~dariLu)	(antanho) (ax~tax~n~u)
(amolecer) (amoleser)	(andor) (ax~dor)	(antártico) (ax~tartS iku)
(amolecimento) (amolesime~tu)	(andorinha) (ax~dorin~a)	(ante) (ax~tS i)
(amônia) (amonja)	(andrajo) (ax~draZu)	(antebraço) (ax~tebrasu)
(amontoado) (amo~toadu)	(andrajoso) (ax~draZozu)	(antecâmara) (ax~tekax~mara)
(amontoamento) (amo~toame~tu)	(anedota) (aned>ta)	(antecedência) (ax~tesede~sja)
(amontoar) (amo~toar)	(anel) (anEw)	(antecedente) (ax~tesede~tS i)
(amor) (amor)	(anelar) (anelar)	(anteceder) (ax~teseder)
(amora) (am>ra)	(anelo) (anElu)	(antecessor) (ax~tesesor)
(amordçar) (amordasar)	(anemia) (anemia)	(antecipação) (ax~tesipasaxw~)
(amornar) (amornar)	(anêmico) (anemiku)	(antecipar) (ax~tesipar)
(amoroso) (amorozu)	(anêmona) (anemona)	(antegozar) (ax~tegozar)
(amortalhar) (amortalar)	(anestesia) (anestezia)	(antemão) (ax~temaxw~)
(amortecedor) (amortecedor)	(anestésiar) (anesteziar)	(antemuro) (ax~temuru)
(amortecer) (amortecer)	(anestésico) (anestEziku)	(antena) (ax~tena)
(amortecido) (amortesidu)	(anexação) (aneksasaxw~)	(anteontem) (ax~teo~tej~)
(amortecimento) (amortesime~tu)	(anexar) (aneksar)	(anteparar) (ax~teparar)
(amortização) (amortS izasaxw~)	(anexo) (anEksu)	(anteparo) (ax~teparu)
(amortizar) (amortS izar)	(anfíbio) (ax~fibju)	(antepassado) (ax~tepasadu)
(amostra) (am>stra)	(anfiteatro) (ax~fiteatru)	(antepasto) (ax~tepastu)
(amotinar) (amotS inar)	(anfitrião) (ax~fitriaxw~)	(antepenúltimo) (ax~tepenuwtS imu)
(amparar) (ax~parar)	(ânfora) (ax~fora)	(anteprojeto) (ax~tS iprozEtu)
(amparo) (ax~paru)	(angariador) (ax~garjador)	(anterior) (ax~terior)
(amperagem) (ax~peraZej~)	(angariar) (ax~gariar)	(antes) (ax~tS is)
(ampliação) (ax~pljasaxw~)	(angelical) (ax~Zelikaw)	(antever) (ax~tever)
(amplificador) (ax~pljador)	(angélico) (ax~ZEliku)	(antevéspera) (ax~tevEspera)
(ampliar) (ax~pliar)	(angina) (ax~Zina)	(antiaéreo) (ax~tS iaErju)
(amplidão) (ax~plidaxw~)	(anglicano) (ax~glikax~nu)	(antibiótico) (ax~tS ibi>tS iku)
(amplificação) (ax~plifikasaxw~)	(anglo~saxão) (ax~glusaksaxw~)	(anticoncepcional) (ax~tS iko~sepsjonaw)
(amplificador) (ax~plifikador)	(angra) (ax~gra)	

(anticonstitucional)	(ax~tS iko~stS((apartar)	(apartar)	(apresentável)	(aprez'tavew)
itusjonaw)		(aparte)	(apartS i)	(apressar)	(apresar)
(anticorpo)	(ax~tS i	(apartear)	(apartear)	(aprimorar)	(aprimorar)
(anticiclone)	(ax~tS isikloni)	(apasentar)	(apasentar)	(aprisonar)	(aprizjonar)
(antiderrapante)	(ax~tS idexapax~tS i)	(apatetar)	(apatetar)	(aprofundar)	(aprofundar)
(antídoto)	(ax~tS idotu)	(apatia)	(apatS ia)	(aprontar)	(aprotar)
(antiestético)	(ax~tS iestEtS iku)	(apavorado)	(apavoradu)	(apropriação)	(aproprijasaw)
(antifebril)	(ax~tS ifebriw)	(apavorar)	(apavorar)	(apropriado)	(apropriadu)
(antígeno)	(ax~tS izenu)	(apaziguar)	(apazigwar)	(apropriar)	(apropriar)
(antigo)	(ax~tS igu)	(apedrejar)	(apedreZar)	(aprovação)	(aprovasaw)
(antiguidade)	(ax~tS igidadZ i)	(apegar)	(apegar)	(aprovar)	(aprovar)
(anti-higiênico)	(ax~tjizieniku)	(apego)	(apegu)	(aproveitamento)	(provejtam'tu)
(anti-histamínico)	(ax~tjistaminiku)	(apelar)	(apelar)	(aproveitar)	(provejtar)
(anti-horário)	(ax~tjorarju)	(apelidar)	(apelidar)	(aprovisionamento)	(aprovizjonam'tu)
(antílope)	(ax~tS ilopi)	(apelido)	(apelidu)	(aprovisionar)	(aprovizjonar)
(antiofídico)	(ax~tjofidZ iku)	(apelo)	(apelu)	(aproximação)	(aprosimasaw)
(antipatia)	(ax~tS ipatS ia)	(apenas)	(apenas)	(aproximado)	(aprosimadu)
(antipático)	(ax~tS ipatS iku)	(apêndice)	(ap'dZ isi)	(aproximar)	(aprosimar)
(antipatizar)	(ax~tS ipatS izar)	(apendicite)	(ap'dZ isitS i)	(aprumar)	(aprumar)
(antipatriótico)	(ax~tS ipatri>tS iku)	(apensar)	(ap'sar)	(aptidão)	(aptS idaw)
(antiquado)	(ax~tS ikwadu)	(aperfeiçoado)	(aperfejsoadu)	(apto)	(aptu)
(antiquar)	(ax~tS ikwar)	(aperfeiçoamento)	(aperfejsoam'tu)	(apunhalar)	(apun~alar)
(antiquário)	(ax~tS ikwarju)	(aperfeiçoar)	(aperfejsoar)	(apurar)	(apurar)
(anti-sepsia)	(ax~tS isepsia)	(aperitivo)	(aperitS ivu)	(apurado)	(apuradu)
(anti-séptico)	(ax~tS isEptS iku)	(aperrear)	(apexear)	(apurar)	(apurar)
(anti-social)	(ax~tS isociaw)	(apertado)	(apertadu)	(apuro)	(apuru)
(antítese)	(ax~tS itezi)	(apertar)	(apertar)	(aquarela)	(akwarEla)
(antitoxina)	(ax~tS itoksina)	(apesar de)	(apezardZ i)	(aquário)	(akwarju)
(antolhos)	(ax~t>Lus)	(apetecer)	(apeteser)	(aquartelar)	(akwartelar)
(antologia)	(ax~toloZia)	(apetite)	(apetS itS i)	(aquático)	(akwatS iku)
(antologista)	(ax~toloZista)	(apetrecho)	(apetreSu)	(aquecedor)	(akesedor)
(antônimo)	(ax~tonimu)	(ápice)	(apisi)	(aquecer)	(akeser)
(antracite)	(ax~trasitS i)	(apiedar)	(apiedar)	(aquecimento)	(akesim'tu)
(antro)	(ax~tru)	(apimentar)	(apim'tar)	(aqueduto)	(akedutu)
(antropófago)	(ax~trop>fagu)	(apinhar)	(apin~ar)	(aquela)	(akEla)
(antropologia)	(ax~tropolozia)	(apitar)	(apitar)	(aquele)	(akeli)
(antropologista)	(ax~tropolozista)	(apilacar)	(aplakar)	(aquém)	(ak'j)
(antropólogo)	(ax~trop>logu)	(aplainar)	(aplainar)	(aqui)	(aki)
(anual)	(anuaw)	(aplanar)	(aplanar)	(aquietar)	(akjetar)
(anuário)	(anuarju)	(aplaudir)	(aplawdZ ir)	(aquilatar)	(akilatar)
(anuência)	(anue~sja)	(aplauso)	(aplawzu)	(aquilo)	(akilu)
(anuidade)	(anujdadZ i)	(aplicar)	(aplikar)	(aquisição)	(akizisaw)
(anuir)	(anuir)	(aplicável)	(aplikavew)	(aquosidade)	(akwozidadZ i)
(anulação)	(anulasaxw~)	(apoderar-se)	(apoderarsi)	(aquoso)	(akwozu)
(anular)	(anular)	(apodrecer)	(apodreser)	(ar)	(ar)
(anunciação)	(anu~sjasaxw~)	(apogeu)	(apoZew)	(árabe)	(arabi)
(anunciante)	(anu~siax~tS i)	(apojar)	(apojar)	(arado)	(aradu)
(anunciar)	(anu~siar)	(apoio)	(apoju)	(aragem)	(araZ'j)
(anúncio)	(anu~sju)	(apólice)	(ap'lisi)	(arame)	(arax~mi)
(anuviar)	(anuviar)	(apologia)	(apoloZia)	(aranha)	(arax~n~a)
(anzol)	(ax~z>w)	(apontado)	(apötadu)	(arar)	(arar)
(ao)	(aw)	(apontador)	(apötador)	(arauto)	(arawtu)
(aonde)	(aödZ i)	(apontamento)	(apötam'tu)	(arbitrar)	(arbitrar)
(apadrinhar)	(apadrin~ar)	(apontar)	(apötar)	(arbitrariedade)	(arbitrarjedadZ i)
(apagar)	(apagar)	(apoplexia)	(apopleksia)	(arbitrário)	(arbitrarju)
(apaixonado)	(apajSonadu)	(apoquentar)	(apok'tar)	(árbitro)	(arbitrju)
(apaixonar)	(apajSonar)	(após)	(ap>s)	(árbitro)	(arbitru)
(apalpar)	(apawpar)	(apostado)	(apoz'tadu)	(arborizar)	(arborizar)
(apanhado)	(apax~n~adu)	(aposentado)	(apoz'tadu)	(arbusto)	(arbustu)
(apanhador)	(apax~n~ador)	(aposentadoria)	(apoz'tadoria)	(arca)	(arka)
(apanhar)	(apax~n~ar)	(aposentar)	(apoz'tar)	(arcada)	(arkada)
(apara)	(apara)	(aposento)	(apoz'tu)	(arcaico)	(arkajku)
(aparafusar)	(aparafuzar)	(apossar)	(aposar)	(arcar)	(arkar)
(aparar)	(aparar)	(aposta)	(ap>sta)	(arcebispo)	(arsebispu)
(aparato)	(aparatu)	(apostar)	(apostar)	(arco)	(arku)
(aparatoso)	(aparatozu)	(apóstolo)	(ap>stulu)	(arco-íris)	(arkuiris)
(aparecer)	(apareser)	(apóstrofo)	(ap>strofu)	(ardente)	(ard'tS i)
(aparecimento)	(aparesim'tu)	(apoteose)	(apote>zi)	(arder)	(arder)
(aparelhagem)	(apareLaZ'j)	(aprazível)	(aprazivew)	(ardido)	(ardZ idu)
(aparelhamento)	(apareLam'tu)	(apreciação)	(apresjasaw)	(ardil)	(ardZ iw)
(aparelhar)	(apareLar)	(apreciar)	(apresiar)	(ardiloso)	(ardZ ilozu)
(aparelho)	(apareLu)	(apreender)	(apre'der)	(ardor)	(ardor)
(aparência)	(apar'sja)	(apreensão)	(apre'saw)	(ardósia)	(ard>zja)
(aparentado)	(apar'tadu)	(apreensivo)	(apre'sivu)	(árcado)	(ardwu)
(aparentar)	(apar'tar)	(apregoar)	(apregoar)	(área)	(arja)
(aparente)	(apar'tS i)	(aprender)	(apr'der)	(areia)	(areja)
(aparição)	(aparisaw)	(aprendiz)	(apr'dZ is)	(arejar)	(areZar)
(apartado)	(apartadu)	(apresar)	(aprezar)	(arenoso)	(arenozu)
(apartamento)	(apartam'tu)	(apresentação)	(aprez'tasaw)	(arenque)	(ar'ki)
		(apresentar)	(aprez'tar)		

(aresta) (arEsta)	(arreio) (axeju)	(ascensor) (ase~sor)
(arfar) (arfar)	(arreliar) (axeliar)	(ascensorista) (ase~sorista)
(argamassa) (argamasa)	(arrematar) (axematar)	(asceta) (asEta)
(argentino) (arZ'tS inu)	(arremate) (axemateS i)	(ascético) (asEtS iku)
(argila) (arZila)	(arremedar) (axemedar)	(asco) (asku)
(argola) (arg>la)	(arremessar) (axemesar)	(ascórbico) (ask>rbiku)
(argúcia) (argusja)	(arremesso) (axemesu)	(asfaltar) (asfawtar)
(argüir) (argwir)	(arremeter) (axemeter)	(asfalto) (asfawtu)
(argumentar) (argum'tar)	(arrendador) (ax'dador)	(asfixia) (asfiksia)
(argumento) (argum'tu)	(arrendamento) (ax'dam'tu)	(asfixiar) (asfiksiar)
(argutez) (argutu)	(arrendar) (ax'dar)	(asiático) (aziatS iku)
(aridez) (arides)	(arrendatário) (ax'datarju)	(asilar) (azilar)
(aristocracia) (aristokrasia)	(arrender-se) (axep'dersi)	(asilo) (azilu)
(aritmética) (aritmEtS ika)	(arrendimento) (axep'dZ im'tu)	(asma) (azma)
(arlequim) (arlekõ)	(arrepisar) (axepiar)	(asneira) (aznejra)
(arma) (arma)	(arrevesar) (axevezar)	(asno) (aznu)
(armação) (armasãw)	(arriar) (axiar)	(aspa) (aspa)
(armada) (armada)	(arribar) (axibar)	(aspargo) (aspargu)
(armadilha) (armadZ iLe)	(arrimar) (aximar)	(aspecto) (aspEktu)
(armado) (armadu)	(arrimo) (aximu)	(aspereza) (aspereza)
(armador) (armador)	(arriscar) (axiskar)	(aspergir) (asperZir)
(armadura) (armadura)	(arrivista) (axivista)	(áspero) (asperu)
(armamento) (armam'tu)	(arroba) (axoba)	(aspersão) (aspersaxw~)
(armar) (armar)	(arrochar) (axoSar)	(aspiração) (aspirasaxw~)
(armarinho) (armarin-u)	(arrocho) (axoSü)	(aspirador) (aspirador)
(armário) (armarju)	(arrogância) (axogãsjja)	(aspirante) (aspirax~tS i)
(armazém) (armaz'j)	(arrogante) (axogãtS i)	(aspirar) (aspirar)
(armazenagem) (armazenaZ'j)	(arrogar) (axogar)	(aspirina) (aspirina)
(armazenar) (armazenaar)	(arroio) (axoju)	(asqueroso) (askerozu)
(armênio) (armenju)	(arrojado) (axoZadu)	(assacar) (asakar)
(armistício) (armistS isju)	(arrojar) (axoZar)	(assadeira) (asadejra)
(aro) (aru)	(arrolar) (axolar)	(assado) (asadu)
(aroma) (aroma)	(arrolhar) (axoLár)	(assador) (asador)
(arpão) (arpãw)	(arrombador) (axõbador)	(assadura) (asadura)
(arpoar) (arpoar)	(arrombar) (axõbar)	(assalarlar) (asalarlar)
(arquear) (arkear)	(arrostar) (axostar)	(assaltante) (asawtax~tS i)
(arqueiro) (arkejru)	(arrotar) (axotar)	(assaltar) (asawtar)
(arquejar) (arkeZar)	(arroto) (axotu)	(assalto) (asawtu)
(arqueologia) (arkeoloZia)	(arroubamento) (axowbam'tu)	(assanhado) (asan~adu)
(arquétipo) (arkEtS ipu)	(arroubo) (axowbu)	(assanhamento) (asan~ame~tu)
(arquitetada) (arkibãkada)	(arroz) (axos)	(assanhar) (asan~ar)
(arquitetar) (arkitetar)	(arruaça) (axuasa)	(assar) (asar)
(arquiteto) (arkitEtu)	(arruela) (axuEla)	(assassinar) (asasinar)
(arquivar) (arkivar)	(arrufar) (axufar)	(assassinato) (asasinatu)
(arquivo) (arkivu)	(arruinar) (axwinar)	(assassínio) (asasinju)
(arrabalde) (axabawdZ i)	(arrulhar) (axuLar)	(assassino) (asasinu)
(arraia) (arxaja)	(arrulho) (axuLu)	(assaz) (asas)
(arraigar) (axajgar)	(arrumadeira) (axumadejra)	(asseado) (aseadu)
(arrancar) (axãkar)	(arrumar) (axumar)	(asear) (asear)
(arranco) (axãku)	(arsenal) (arsenaw)	(assediar) (asedZ iar)
(arranha-céu) (axax~n~asEw)	(arte) (artS i)	(assédio) (asEdju)
(arranhão) (axax~n~ãw)	(artefato) (artefatu)	(assegurar) (asegurar)
(arranhar) (axax~n~ar)	(arteiro) (artejru)	(asseio) (aseju)
(arranjar) (axãZar)	(artelho) (arteLu)	(asselvajar) (asewvaZar)
(arranjo) (axãZu)	(artéria) (artErja)	(assembléia) (ase~blEja)
(arranque) (axãki)	(arterial) (arteriaw)	(assemelhar) (asemeLar)
(arrasar) (axazar)	(artesanato) (artezanatu)	(assenhorear-se) (asen~orearsi)
(arrastar) (axastar)	(artesão) (artezãw)	(assentado) (ase~tadu)
(arrazoar) (axazoar)	(ártico) (artS iku)	(assentar) (ase~tar)
(arre) (axi)	(articulação) (artS ikulasaxw~)	(assente) (ase~tS i)
(arrear) (axear)	(articular) (artS ikular)	(assentimento) (ase~tS ime~tu)
(arrebancar) (axebax~n~ar)	(artificial) (artS ifisiaw)	(assentir) (ase~tS ir)
(arreatador) (axebatador)	(artifício) (artS ifisiw)	(assento) (ase~tu)
(arreatamento) (axebatam'tu)	(artigo) (artS igu)	(assepsia) (asepsia)
(arreatar) (axebatar)	(artilharia) (artS iLaria)	(asserção) (asersaxw~)
(arreatação) (axeb'tasãw)	(artimãna) (artS imax~n~a)	(assessor) (asesor)
(arreatar) (axeb'tar)	(artista) (artS ista)	(assessoramento) (asesorame~tu)
(arreatado) (axebitadu)	(artístico) (artS istS iku)	(assessorar) (asesorar)
(arreatar) (axebitar)	(artrite) (artritS i)	(assíduo) (asidwu)
(arreatadar) (axekadar)	(arvorar) (arvorar)	(assim) (asi~)
(arreatar) (axedar)	(árvore) (arvori)	(assimilação) (asimilasaxw~)
(arreato) (axedZ iu)	(arvoredo) (arvoredu)	(assimilar) (asimilar)
(arreatar) (axedõdar)	(as) (as)	(assinado) (asinadu)
(arreator) (axed>r)	(ás) (as)	(assinalar) (asinalar)
(arreatores) (axed>ris)	(às) (as)	(assinante) (asinax~tS i)
(arreatecer) (axefeser)	(asa) (aza)	(assinar) (asinar)
(arreatçar) (axegasar)	(ascendência) (ase~de~sjja)	(assinatura) (asinatura)
(arreatalar) (axegalar)	(ascender) (ase~der)	(assistência) (asiste~sjja)
(arreatanhar) (axegax~n~ar)	(ascensão) (ase~saxw~)	(assistente) (asiste~tS i)

(assistir) (asistS ir)
 (assoalhar) (asoaLar)
 (assoalho) (asoaLu)
 (assoar) (asoar)
 (assoberbar) (asoberbar)
 (assobiar) (asobiar)
 (assobio) (asobiu)
 (assobradado) (asobradadu)
 (associação) (asosjasaxw~)
 (associado) (asosiadu)
 (associar) (asosiar)
 (assolação) (asolasaxw~)
 (assolar) (asolar)
 (assomar) (asomar)
 (assombração) (aso~brasaxw~)
 (assombrado) (aso~bradu)
 (assombrar) (aso~brar)
 (assombro) (aso~bru)
 (assombroso) (aso~brozu)
 (assomo) (asomu)
 (assovio) (asoviu)
 (assumir) (asumir)
 (assunção) (asu~saxw~)
 (assuntar) (asu~tar)
 (assunto) (asu~tu)
 (assustadiço) (asustadZ isu)
 (assustado) (asutadu)
 (assustar) (asustar)
 (asterisco) (asterisku)
 (asteróide) (aster>jdZ i)
 (astro) (astru)
 (astrologia) (astroloZia)
 (astronauta) (astronawta)
 (astronáutica) (astronawtS ika)
 (astronomia) (astronomia)
 (astrônomo) (astronomu)
 (astúcia) (astusia)
 (astucioso) (astusiozu)
 (astuto) (astuto)
 (ata) (ata)
 (atabalhoar) (atabaLoar)
 (atacadista) (atakadZ ista)
 (atacado) (atakadu)
 (atacador) (atakador)
 (atacante) (atakax~tS i)
 (atacar) (atakar)
 (atado) (atadu)
 (atadura) (atadura)
 (atalaia) (atalaja)
 (atalho) (ataLu)
 (atamancar) (atamax~kar)
 (atapetar) (atapetar)
 (ataque) (ataki)
 (atar) (atar)
 (atarantar) (atarax~tar)
 (atarefado) (atarefadu)
 (atarefar) (atarefar)
 (atarracado) (ataxakadu)
 (atarraxar) (ataxaSar)
 (ataúde) (ataudZ i)
 (ataviar) (ataviar)
 (atavio) (ataviu)
 (atavismo) (atavizmu)
 (atazanar) (atazanar)
 (até) (atE)
 (atear) (atear)
 (ateísmo) (ateizmu)
 (ateísta) (ateista)
 (atemorizante) (atemorizax~tS i)
 (atemorizar) (atemorizar)
 (atenção) (ate~saxw~)
 (atencioso) (ate~siozu)
 (atender) (ate~der)
 (ateneu) (atenu)
 (atentado) (ate~tadu)
 (atentar) (ate~tar)
 (atenuação) (atenwasaxw~)
 (atenuante) (atenuax~tS i)
 (atenuar) (atenuar)
 (aterrado) (atexadu)
 (aterrador) (atexador)
 (aterrar) (atexar)
 (aterrissagem) (atexisaZej~)
 (aterrissar) (atexisar)
 (aterro) (atexu)
 (aterrorizar) (atexorizar)
 (ater-se) (atersi)
 (atestado) (atestadu)
 (atestar) (atestar)
 (ateu) (atew)
 (atiçador) (atS isador)
 (atiçamento) (atS isame~tu)
 (atiçar) (atS isar)
 (ático) (atS iku)
 (atilado) (atS iladu)
 (atinado) (atS inadu)
 (atinar) (atS inar)
 (atinente) (atS ine~tS i)
 (atingir) (ati~Zir)
 (atiradeira) (atS iradejra)
 (atirado) (atS iradu)
 (atirador) (atS irador)
 (atirar) (atS irar)
 (atitude) (atS itudZ i)
 (ativa) (atS iva)
 (ativar) (atS ivar)
 (atividade) (atS ividadZ i)
 (ativo) (atS ivu)
 (atlântico) (atIax~tS iku)
 (atlas) (atlas)
 (atleta) (atlEta)
 (atletico) (atlEtS iku)
 (atletismo) (atletS izmu)
 (atmosfera) (atmosfEra)
 (atmosférico) (atmosfEriku)
 (ato) (atu)
 (atocaiar) (atokajar)
 (atochar) (atoSar)
 (atol) (at>w)
 (atoladiço) (atoladZ isu)
 (atolar) (atolar)
 (atoleiro) (atolejru)
 (atômico) (atomiku)
 (átomo) (atomu)
 (atônito) (atonitu)
 (ator) (ator)
 (atordoado) (atordoadu)
 (atordoamento) (atordoame~tu)
 (atordoante) (atordoax~tS i)
 (atordoar) (atordoar)
 (atormentado) (atorme~tadu)
 (atormentar) (atorme~tar)
 (atracção) (atrasaxw~)
 (atracar) (atrakar)
 (atraente) (atrae~tS i)
 (atraiçoado) (atrajsoadu)
 (atraiçoador) (atrajsoador)
 (atraiçoar) (atrajsoar)
 (atrair) (atrair)
 (atrapalhação) (atrapaLasaxw~)
 (atrapalhar) (atrapaLar)
 (atrás) (atras)
 (atrasado) (atrazadu)
 (atrasar) (atrazar)
 (atraso) (atrazu)
 (atrativo) (atratS ivu)
 (atravancamento) (atravax~kame~tu)
 (atravancar) (atravax~kar)
 (através) (atravEs)
 (atravessado) (atravesadu)
 (atravessar) (atravesar)
 (atrelar) (atrelar)
 (atrever-se) (atreversi)
 (atrevido) (atrevidu)
 (atrevimento) (atrevime~tu)
 (atribuição) (atribujsaxw~)
 (atribuir) (atribuir)
 (atribulação) (atribulasaxw~)
 (atribular) (atribular)
 (atributo) (atributu)
 (átrio) (atrju)
 (atrito) (atritu)
 (atriz) (atris)
 (atroar) (atroar)
 (atrocidade) (atrosidadZ i)
 (atrofia) (atrofia)
 (atrofiar) (atrofiar)
 (atropelar) (atropelar)
 (atroz) (atr>s)
 (atuação) (atwasaxw~)
 (atual) (atuaw)
 (atualidade) (atwalidadZ i)
 (atualizado) (atwalizadu)
 (atualizar) (atwalizar)
 (atuar) (atuar)
 (atulhar) (atuLar)
 (atum) (atu~)
 (aturar) (aturar)
 (aturdido) (aturdZ idu)
 (aturdimento) (aturdZ ime~tu)
 (aturdir) (aturdZ ir)
 (audácia) (awdasja)
 (audacioso) (awdasiozu)
 (audaz) (awdas)
 (audição) (awdZ isaxw~)
 (audiência) (awdZ ie~sja)
 (auditivo) (awdZ itS ivu)
 (auditor) (awdZ itor)
 (auditoria) (awdZ itoria)
 (auditório) (awdZ it>rju)
 (audível) (awdZ ivew)
 (auferir) (awferir)
 (auge) (awZi)
 (augurar) (awgurar)
 (augúrio) (awgurju)
 (augusto) (awgustu)
 (aula) (awla)
 (aumentar) (awme~tar)
 (aumentar) (awme~tu)
 (aura) (awra)
 (áureo) (awrju)
 (auréola) (awrEola)
 (aurífice) (awrifisi)
 (aurora) (awr>ra)
 (auscultar) (awskuwtar)
 (ausência) (awze~sja)
 (ausentar-se) (awze~tarsi)
 (ausente) (awze~tS i)
 (auspício) (awspisju)
 (auspicioso) (awspisiozu)
 (austeridade) (awsteridadZ i)
 (austero) (awstEru)
 (austríaco) (awstriaku)
 (autarquia) (awtarkia)
 (autenticar) (awte~tS ikar)
 (autenticidade) (awte~tS isidadZ i)
 (autêntico) (awte~tS iku)
 (auto) (awtu)
 (autocracia) (awtokrasia)
 (autocrata) (awtokrata)
 (autocrítica) (awtokritS ika)
 (autóctone) (awt>ktoni)
 (autodomínio) (awtodominju)
 (autografar) (awtografar)
 (autógrafo) (awt>grafu)
 (autolotação) (awtolotasaxw~)
 (automação) (awtomasaxw~)
 (automático) (awtomatS iku)
 (automatização) (awtomatS izasaxw~)
 (automatizar) (awtomatS izar)
 (autômato) (awtomatu)
 (automobilismo) (awtomobilizmu)
 (automotriz) (awtomotris)
 (automóvel) (awtom>vew)
 (autonomia) (awtonomia)
 (autônomo) (awtonomu)
 (autópsia) (awt>psja)

(autor) (awtor)
 (autora) (awtora)
 (autoria) (awtoria)
 (autoridade) (awtoridadZ i)
 (autoritário) (awtoritarju)
 (autorização) (awtorizasaxw~)
 (autorizado) (awtorizadu)
 (autorizar) (awtorizar)
 (auxiliador) (awsiljador)
 (auxiliar) (awsiliar)
 (auxílio) (awsilju)
 (avacalhação) (avakaLasaxw~)
 (avacalhado) (avakaLadu)
 (avacalhar) (avakaLar)
 (aval) (avaw)
 (avalancha) (avalax~Sa)
 (avalanche) (avalax~Si)
 (avaliação) (avaljasaxw~)
 (avaliado) (avaliadu)
 (avaliador) (avaljador)
 (avaliar) (avaliar)
 (avalista) (avalista)
 (avaliar) (avaliar)
 (avançada) (avax~sada)
 (avançado) (avax~sadu)
 (avançar) (avax~sar)
 (avanço) (avax~su)
 (avantajado) (avax~taZadu)
 (avantajar) (avax~taZar)
 (avante) (avax~tS i)
 (avarento) (avare~tu)
 (avareza) (avareza)
 (avariado) (avariadu)
 (avariar) (avariar)
 (avaro) (avaru)
 (avassalar) (avasalar)
 (ave) (avi)
 (aveia) (aveja)
 (avelã) (avelax~)
 (aveleira) (avelejra)
 (aveludado) (aveludadu)
 (aveludar) (aveludar)
 (avenca) (ave~ka)
 (avenida) (avenida)
 (avental) (ave~taw)
 (aventar) (ave~tar)
 (aventura) (ave~tura)
 (aventurar) (ave~turar)
 (aventureiro) (ave~turejru)
 (averbar) (averbar)
 (averiguação) (averigwasaxw~)
 (averiguar) (averigwar)
 (avermelhado) (avermeLadu)
 (avermelhar) (avermeLar)
 (aversão) (aversaxw~)
 (avessas) (avEsas)
 (avesso) (avesu)
 (avestruz) (avestrus)
 (aviação) (avjasaxw~)
 (aviador) (avjador)
 (aviamento) (avjame~tu)
 (avião) (aviaxw~)
 (aviar) (aviar)
 (avícola) (avikola)
 (avicultura) (avikuwtura)
 (avidez) (avides)
 (ávido) (avidu)
 (aviltamento) (aviwtame~tu)
 (aviltante) (aviwtax~tS i)
 (aviltar) (aviwtar)
 (avinagrado) (avinagradu)
 (avisado) (avizadu)
 (avisar) (avizar)
 (aviso) (avizu)
 (avistar) (avistar)
 (avitaminose) (avitamin>zi)
 (avivar) (avivar)
 (avizinhar) (avizin~ar)
 (avô) (avo)

(avó) (av>)
 (avoado) (avoadu)
 (avoamento) (avoame~tu)
 (avocar) (avokar)
 (avolumar) (avolumar)
 (avós) (av>s)
 (avulso) (avuwsu)
 (avultado) (avuwatadu)
 (avultar) (avuwatar)
 (axila) (aksila)
 (azado) (azadu)
 (azáfama) (azafama)
 (azafamar) (azafamar)
 (azagaia) (azagaja)
 (azálea) (azalja)
 (azar) (azar)
 (azarado) (azaradu)
 (azarar) (azarar)
 (azarento) (azare~tu)
 (azedar) (azedar)
 (azedo) (azedu)
 (azedume) (azedumi)
 (azeitar) (azejtar)
 (azeite) (azejtS i)
 (azeiteira) (azejtejra)
 (azeitona) (azejtona)
 (azeviche) (azeviSi)
 (azevim) (azevi~)
 (azevinho) (azevin~u)
 (azia) (azia)
 (aziago) (aziagu)
 (ázimo) (azimu)
 (azimute) (azimutS i)
 (azo) (azu)
 (azoar) (azoar)
 (azorrague) (azoragi)
 (azotar) (azotar)
 (azoto) (azotu)
 (azougue) (azowgi)
 (azucrinar) (azukrinar)
 (azul) (azuw)
 (azulado) (azuladu)
 (azular) (azular)
 (azulejar) (azuleZar)
 (azulejo) (azuleZu)

Anexo 2 Passos básicos para criar uma definição léxica no Festival

A seguir são apresentados os passos básicos para criar uma definição léxica no Festival.

Primeiro um novo léxico deve ser criado com um novo nome

```
(lex.create "cstrlex")
```

Um conjunto de fones deve ser declarado para o léxico, permitir ambas as checagens nas entradas e permitir o mapeamento de fones entre conjuntos de fones diferentes usados no sistema

```
(lex.set.phoneset "mrpa")
```

O conjunto de fones já deve estar declarado no sistema.

Exemplo de um léxico compilado, a seguir.

```
(lex.set.compile.file "/projects/festival/lib/dicts/cstrlex.out")
```

O método para tratar com palavras desconhecidas, pode ser setar

```
(lex.set.lts.method 'lts_rules)
```

```
(lex.set.lts.ruleset 'nrl)
```

Neste caso, está se especificando o uso de um conjunto de regras letra para som originalmente desenvolvido pelos Laboratórios de Pesquisa da Marinha dos Estados Unidos. O método default é dar um erro, se uma palavra não é encontrada no adendo ou léxico compilado.

Finalmente, itens do adendo podem ser adicionados para palavras que são conhecidas por serem comuns, mas não no léxico e não podem ser, razoavelmente, analisadas pelas regras letra para som.

```
(lex.add.entry
 '( "awb" n ((( ei ) 1) ((d uh) 1) ((b @) 0) ((y uu) 0) ((b ii) 1))))
(lex.add.entry
 '( "cstr" n ((( s ii ) 1) (( e s ) 1) (( t ii ) 1) (( aa ) 1) ))
(lex.add.entry
 '( "Edinburgh" n ((( e m ) 1) (( b r @ ) 0) ))
```

Anexo 3 Forma básica de uma regra para o Festival

A seguir é apresentada a forma básica de uma regra para o Festival.

(LEFTCONTEXT [ITEMS] RIGHTCONTEXT = NEWITEMS)

Esta interpretação é que se ITEMS aparecem no contexto direito especificado e esquerdo então a string de saída contém NEWITEMS. Qualquer de LEFTCONTEXT, RIGHTCONTEXT ou NEWITEMS podem estar vazios. Note que NEWITEMS está escrito para uma fita diferente e daí não pode alimentar mais regras. Como exemplo está (# [c h] C = k).

O caracter especial # significa um limite da palavra, e o símbolo C significa que o conjunto de todas as consoantes, conjuntos são declarados antes das regras. Esta regra afirma que um ch no início de uma palavra seguido por uma consoante é para ser traduzido como um fonema k. Símbolos nos contextos podem ser seguidos pelo símbolo * para zero ou mais ocorrências, ou + para uma ou mais ocorrências.

Os símbolos nas regras são tratados como nomes de conjuntos se eles são declarados como tal ou como símbolos nos alfabetos de entrada/saída. Os símbolos podem ser mais do que um caracter longo e os nomes são sensíveis ao contexto.

A função usada para aplicar um conjunto de regras se dado um átomo explodirá em uma lista de caracteres simples, enquanto se dado uma lista usará como é. Isto reflete o uso comum de desejar a reescrita de letras individuais em uma palavra para fonemas mas sem excluir a possibilidade de usar o sistema para manipulações complexas, tais como sistemas LTS multi-pass e conversão de fonemas.

Anexo 4 Exemplo de construção de um conjunto de fones para o Festival

A seguir é apresentado um exemplo de construção de um conjunto de fones para o Festival.

```
(defPhoneSet
  spanish
  ;; Phone Features
  ;; vowel or consonant
  (vc + -)
  ;; vowel length: short long diphthong schwa
  (vln g s l d a 0)
  ;; vowel height: high mid low
  (vheight 1 2 3 -)
  ;; vowel frontness: front mid back
  (vfront 1 2 3 -)
  ;; lip rounding
  (vrnd + -)
  ;; consonant type: stop fricative affricative nasal liquid
  (ctype s f a n l 0)
  ;; place of articulation: labial alveolar palatal labio-dental
  ;;                          dental velar
  (cplace l a p b d v 0)
  ;; consonant voicing
  (cvox + -)
)
;; Phone set members (features are not! set properly)
(
  (# - 0 - - - 0 0 -)
  (a + 1 3 1 - 0 0 -)
  (e + 1 2 1 - 0 0 -)
  (i + 1 1 1 - 0 0 -)
  (o + 1 3 3 - 0 0 -)
  (u + 1 1 3 + 0 0 -)
  (b - 0 - - + s l +)
  (ch - 0 - - + a a -)
  (d - 0 - - + s a +)
  (f - 0 - - + f b -)
  (g - 0 - - + s p +)
  (j - 0 - - + l a +)
  (k - 0 - - + s p -)
  (l - 0 - - + l d +)
  (ll - 0 - - + l d +)
  (m - 0 - - + n l +)
  (n - 0 - - + n d +)
  (ny - 0 - - + n v +)
  (p - 0 - - + s l -)
  (r - 0 - - + l p +)
)
```



```
(rr - 0 - - + l p +)  
(s - 0 - - + f a +)  
(t - 0 - - + s t +)  
(th - 0 - - + f d +)  
(x - 0 - - + a a -)  
)  
)  
(PhoneSet.silences '#)
```

Anexo 5 Exemplo de regras letra para som

Um sistema de regras letra para som do Espanhol é bastante simples no formato suportado por Festival. O seguinte é um bom início para um conjunto completo.

```
(lts.ruleset
; Name of rule set spanish
; Sets used in the rules
( (LNS l n s )
(AEOU a e o u )
(AEO a e o )
(EI e i )
(BDGLMN b d g l m n )
)
; Rules
( ( [ a ] = a )
( [ e ] = e )
( [ i ] = i )
( [ o ] = o )
( [ u ] = u )
( [ "" a ] = a1 ) ;; stressed vowels
( [ "" e ] = e1 )
( [ "" i ] = i1 )
( [ "" o ] = o1 )
( [ "" u ] = u1 )
( [ b ] = b )
( [ v ] = b )
( [ c ] "" EI = th )
( [ c ] EI = th )
( [ c h ] = ch )
( [ c ] = k )
( [ d ] = d )
( [ f ] = f )
( [ g ] "" EI = x )
( [ g ] EI = x )
( [ g u ] "" EI = g )
( [ g u ] EI = g )
( [ g ] = g )
( [ h u e ] = u e )
( [ h i e ] = i e )
( [ h ] = )
( [ j ] = x )
( [ k ] = k )
( [ l l ] # = l )
( [ l l ] = ll )
( [ l ] = l )
( [ m ] = m )
( [ ~ n ] = ny )
( [ n ] = n )
( [ p ] = p )
```

([q u] = k)
([r r] = rr)
(# [r] = rr)
(LNS [r] = rr)
([r] = r)
([s] BDGLMN = th)
([s] = s)
(# [s] C = e s)
([t] = t)
([w] = u)
([x] = k s)
(AEO [y] = i)
(# [y] # = i)
([y] = ll)
([z] = th))

Anexo 6 Exemplo de uma Cart tree

A seguir um exemplo de CART tree para prever quebras simples da pontuação.

```
(set! spanish_phrase_cart_tree
,
((lisp_token_end_punc in ("?" "." ":"))
((BB)
((lisp_token_end_punc in ("'" "\"" "," ";""))
((B)
((n.name is 0) ;; end of utterance
((BB)
((NB))))))
```

Anexo 7 Resultado da análise estatística para a vogal “e”

A seguir vê-se o resultado da análise estatística da ocorrência de ‘e’ fechado e ‘E’ aberto, onde ‘#’ significa nada (nenhum caracter).

Anterior	Posterior	Casos - e	Casos – E	Total de Casos	%e	%E
#	b	2	0	2	100.00%	0.00%
#	c	20	1	21	95.24%	4.76%
#	d	15	0	15	100.00%	0.00%
#	f	17	0	17	100.00%	0.00%
#	g	4	0	4	100.00%	0.00%
#	i	3	0	3	100.00%	0.00%
#	j	2	0	2	100.00%	0.00%
#	l	50	3	53	94.34%	5.66%
#	m	33	0	33	100.00%	0.00%
#	n	31	0	31	100.00%	0.00%
#	p	18	0	18	100.00%	0.00%
#	q	23	0	23	100.00%	0.00%
#	r	24	2	26	92.31%	7.69%
#	s	754	4	758	99.47%	0.53%
#	t	13	0	13	100.00%	0.00%
#	u	8	0	8	100.00%	0.00%
#	v	27	0	27	100.00%	0.00%
#	x	197	2	199	98.99%	1.01%
-	d	3	0	3	100.00%	0.00%
-	s	2	2	4	50.00%	50.00%
a	c	5	0	5	100.00%	0.00%
a	l	1	0	1	100.00%	0.00%
a	s	0	1	1	0.00%	100.00%
a	z	1	0	1	100.00%	0.00%
b	a	10	0	10	100.00%	0.00%
b	b	10	0	10	100.00%	0.00%
b	c	6	3	9	66.67%	33.33%
b	d	17	0	17	100.00%	0.00%
b	f	0	1	1	0.00%	100.00%
b	i	29	0	29	100.00%	0.00%
b	j	3	0	3	100.00%	0.00%
b	l	40	10	50	80.00%	20.00%
b	n	16	0	16	100.00%	0.00%
b	q	1	2	3	33.33%	66.67%
b	r	64	0	64	100.00%	0.00%
b	s	21	0	21	100.00%	0.00%
b	t	13	0	13	100.00%	0.00%
b	u	1	0	1	100.00%	0.00%
b	v	2	0	2	100.00%	0.00%
b	x	3	0	3	100.00%	0.00%

b	z	2	0	2	100.00%	0.00%
b	ç	6	1	7	85.71%	14.29%
b	ó	1	0	1	100.00%	0.00%
c	a	13	0	13	100.00%	0.00%
c	b	16	0	16	100.00%	0.00%
c	c	2	0	2	100.00%	0.00%
c	d	43	1	44	97.73%	2.27%
c	f	2	0	2	100.00%	0.00%
c	g	6	2	8	75.00%	25.00%
c	i	45	0	45	100.00%	0.00%
c	j	7	0	7	100.00%	0.00%
c	l	39	10	49	79.59%	20.41%
c	m	3	0	3	100.00%	0.00%
c	n	14	0	14	100.00%	0.00%
c	o	1	0	1	100.00%	0.00%
c	p	20	0	20	100.00%	0.00%
c	r	217	9	226	96.02%	3.98%
c	s	32	0	32	100.00%	0.00%
c	t	25	3	28	89.29%	10.71%
c	u	1	0	1	100.00%	0.00%
c	v	4	0	4	100.00%	0.00%
c	ç	1	0	1	100.00%	0.00%
d	#	19	0	19	100.00%	0.00%
d	-	1	0	1	100.00%	0.00%
d	a	30	1	31	96.77%	3.23%
d	b	16	0	16	100.00%	0.00%
d	c	84	0	84	100.00%	0.00%
d	d	18	0	18	100.00%	0.00%
d	e	1	0	1	100.00%	0.00%
d	f	37	0	37	100.00%	0.00%
d	g	15	0	15	100.00%	0.00%
d	i	99	0	99	100.00%	0.00%
d	j	3	0	3	100.00%	0.00%
d	l	44	16	60	73.33%	26.67%
d	m	34	0	34	100.00%	0.00%
d	n	25	0	25	100.00%	0.00%
d	ã	3	0	3	100.00%	0.00%
d	p	30	0	30	100.00%	0.00%
d	q	3	0	3	100.00%	0.00%
d	r	126	6	132	95.45%	4.55%
d	s	563	6	569	98.95%	1.05%
d	t	19	1	20	95.00%	5.00%
D	u	1	0	1	100.00%	0.00%
d	u	6	0	6	100.00%	0.00%
d	v	22	0	22	100.00%	0.00%
d	x	6	1	7	85.71%	14.29%
d	z	51	1	52	98.08%	1.92%
e	d	5	0	5	100.00%	0.00%

e	i	5	0	5	100.00%	0.00%
e	l	2	0	2	100.00%	0.00%
e	m	2	0	2	100.00%	0.00%
e	r	1	0	1	100.00%	0.00%
e	s	1	0	1	100.00%	0.00%
f	a	2	0	2	100.00%	0.00%
f	b	4	1	5	80.00%	20.00%
f	c	19	2	21	90.48%	9.52%
f	d	8	0	8	100.00%	0.00%
f	g	5	0	5	100.00%	0.00%
f	i	61	0	61	100.00%	0.00%
f	j	2	0	2	100.00%	0.00%
f	l	12	1	13	92.31%	7.69%
f	m	7	0	7	100.00%	0.00%
f	n	6	0	6	100.00%	0.00%
f	ã	1	0	1	100.00%	0.00%
f	r	113	12	125	90.40%	9.60%
f	s	23	2	25	92.00%	8.00%
f	t	23	0	23	100.00%	0.00%
f	u	2	0	2	100.00%	0.00%
f	v	1	0	1	100.00%	0.00%
f	í	1	0	1	100.00%	0.00%
f	z	3	2	5	60.00%	40.00%
g	a	4	0	4	100.00%	0.00%
g	b	1	0	1	100.00%	0.00%
g	d	2	0	2	100.00%	0.00%
g	i	7	0	7	100.00%	0.00%
g	l	24	4	28	85.71%	14.29%
g	m	7	0	7	100.00%	0.00%
g	n	33	0	33	100.00%	0.00%
g	o	6	0	6	100.00%	0.00%
g	r	42	1	43	97.67%	2.33%
g	s	15	1	16	93.75%	6.25%
g	t	5	0	5	100.00%	0.00%
g	u	1	0	1	100.00%	0.00%
g	v	1	0	1	100.00%	0.00%
g	ó	1	0	1	100.00%	0.00%
h	a	8	0	8	100.00%	0.00%
h	b	2	0	2	100.00%	0.00%
h	c	17	0	17	100.00%	0.00%
h	d	5	0	5	100.00%	0.00%
h	f	3	1	4	75.00%	25.00%
h	g	7	0	7	100.00%	0.00%
h	i	53	0	53	100.00%	0.00%
h	l	2	0	2	100.00%	0.00%
h	m	7	0	7	100.00%	0.00%
h	p	1	0	1	100.00%	0.00%
h	q	0	1	1	0.00%	100.00%

h	r	31	3	34	91.18%	8.82%
h	s	4	0	4	100.00%	0.00%
h	t	12	1	13	92.31%	7.69%
h	u	1	0	1	100.00%	0.00%
h	v	2	0	2	100.00%	0.00%
h	z	2	0	2	100.00%	0.00%
i	d	5	0	5	100.00%	0.00%
i	g	1	1	2	50.00%	50.00%
i	i	8	0	8	100.00%	0.00%
i	l	1	4	4	0.00%	100.00%
i	n	4	0	4	100.00%	0.00%
i	r	2	0	2	100.00%	0.00%
i	s	3	0	3	100.00%	0.00%
i	t	6	5	11	54.55%	45.45%
i	v	1	0	1	100.00%	0.00%
i	z	4	0	4	100.00%	0.00%
j	a	5	0	5	100.00%	0.00%
j	c	0	1	1	0.00%	100.00%
j	i	23	0	23	100.00%	0.00%
j	j	2	0	2	100.00%	0.00%
j	l	0	1	1	0.00%	100.00%
j	r	4	0	4	100.00%	0.00%
j	s	4	1	5	80.00%	20.00%
j	t	18	0	18	100.00%	0.00%
j	ç	4	0	4	100.00%	0.00%
k	l	0	1	1	0.00%	100.00%
k	t	2	1	3	66.67%	33.33%
l	a	21	0	21	100.00%	0.00%
l	b	8	3	11	72.73%	27.27%
l	c	23	3	26	88.46%	11.54%
l	d	1	0	1	100.00%	0.00%
l	e	2	0	2	100.00%	0.00%
l	f	11	1	12	91.67%	8.33%
l	g	57	0	57	100.00%	0.00%
l	i	87	0	87	100.00%	0.00%
l	j	12	0	12	100.00%	0.00%
l	l	0	3	3	0.00%	100.00%
l	m	17	0	17	100.00%	0.00%
l	n	20	0	20	100.00%	0.00%
l	ã	3	0	3	100.00%	0.00%
l	o	7	0	7	100.00%	0.00%
l	p	5	3	8	62.50%	37.50%
l	q	1	3	4	25.00%	75.00%
l	r	29	0	29	100.00%	0.00%
l	á	1	0	1	100.00%	0.00%
l	s	26	4	30	86.67%	13.33%
l	t	66	17	83	79.52%	20.48%
l	u	5	0	5	100.00%	0.00%

l	v	28	2	30	93.33%	6.67%
l	x	12	6	18	66.67%	33.33%
l	z	10	0	10	100.00%	0.00%
l	ç	4	0	4	100.00%	0.00%
m	#	2	1	3	66.67%	33.33%
m	a	21	0	21	100.00%	0.00%
m	c	8	1	9	88.89%	11.11%
m	d	57	1	58	98.28%	1.72%
m	e	1	0	1	100.00%	0.00%
m	g	5	0	5	100.00%	0.00%
m	i	17	0	17	100.00%	0.00%
m	j	7	0	7	100.00%	0.00%
m	l	46	8	54	85.19%	14.81%
m	m	14	0	14	100.00%	0.00%
m	n	27	0	27	100.00%	0.00%
m	o	3	0	3	100.00%	0.00%
m	r	89	4	93	95.70%	4.30%
m	á	1	0	1	100.00%	0.00%
m	s	28	10	38	73.68%	26.32%
m	t	67	2	69	97.10%	2.90%
m	u	2	0	2	100.00%	0.00%
m	x	10	0	10	100.00%	0.00%
m	z	3	0	3	100.00%	0.00%
m	ç	4	0	4	100.00%	0.00%
n	a	17	0	17	100.00%	0.00%
n	b	8	0	8	100.00%	0.00%
n	c	21	6	27	77.78%	22.22%
n	d	3	0	3	100.00%	0.00%
n	f	11	0	11	100.00%	0.00%
n	g	28	0	28	100.00%	0.00%
n	i	42	0	42	100.00%	0.00%
n	j	6	0	6	100.00%	0.00%
n	l	9	14	23	39.13%	60.87%
n	m	6	0	6	100.00%	0.00%
n	n	7	0	7	100.00%	0.00%
n	o	7	0	7	100.00%	0.00%
n	p	2	0	2	100.00%	0.00%
n	q	2	0	2	100.00%	0.00%
n	r	45	1	46	97.83%	2.17%
n	á	1	0	1	100.00%	0.00%
n	s	24	3	27	88.89%	11.11%
n	t	36	10	46	78.26%	21.74%
n	u	12	0	12	100.00%	0.00%
n	v	8	1	9	88.89%	11.11%
n	x	17	2	19	89.47%	10.53%
n	z	4	0	4	100.00%	0.00%
n	ó	1	0	1	100.00%	0.00%
n	ô	1	0	1	100.00%	0.00%

o	c	2	0	2	100.00%	0.00%
o	d	3	1	4	75.00%	25.00%
o	f	1	0	1	100.00%	0.00%
o	i	17	0	17	100.00%	0.00%
o	j	1	0	1	100.00%	0.00%
o	l	5	2	7	71.43%	28.57%
o	m	4	0	4	100.00%	0.00%
o	r	12	0	12	100.00%	0.00%
o	s	7	3	10	70.00%	30.00%
o	t	2	1	3	66.67%	33.33%
o	x	2	0	2	100.00%	0.00%
o	z	1	0	1	100.00%	0.00%
p	#	1	0	1	100.00%	0.00%
p	a	11	0	11	100.00%	0.00%
p	b	1	0	1	100.00%	0.00%
p	c	44	6	50	88.00%	12.00%
p	d	47	1	48	97.92%	2.08%
p	f	4	0	4	100.00%	0.00%
p	g	12	2	14	85.71%	14.29%
p	i	30	0	30	100.00%	0.00%
p	j	6	0	6	100.00%	0.00%
p	l	48	5	53	90.57%	9.43%
p	n	45	0	45	100.00%	0.00%
p	o	1	0	1	100.00%	0.00%
p	ã	3	0	3	100.00%	0.00%
p	p	2	0	2	100.00%	0.00%
p	q	3	0	3	100.00%	0.00%
p	r	272	9	281	96.80%	3.20%
p	s	38	3	41	92.68%	7.32%
p	t	46	2	48	95.83%	4.17%
p	u	2	0	2	100.00%	0.00%
p	v	2	0	2	100.00%	0.00%
p	z	3	0	3	100.00%	0.00%
p	ç	6	1	7	85.71%	14.29%
r	#	5	0	5	100.00%	0.00%
r	a	90	0	90	100.00%	0.00%
r	b	47	3	50	94.00%	6.00%
r	c	254	10	264	96.21%	3.79%
r	d	89	2	91	97.80%	2.20%
r	e	38	0	38	100.00%	0.00%
r	f	91	1	92	98.91%	1.09%
r	g	117	11	128	91.41%	8.59%
r	i	105	0	105	100.00%	0.00%
r	j	38	0	38	100.00%	0.00%
r	l	85	13	98	86.73%	13.27%
r	m	116	0	116	100.00%	0.00%
r	n	51	0	51	100.00%	0.00%
r	â	1	0	1	100.00%	0.00%

r	o	12	0	12	100.00%	0.00%
r	p	125	1	126	99.21%	0.79%
r	q	23	1	24	95.83%	4.17%
r	r	24	0	24	100.00%	0.00%
r	s	299	0	299	100.00%	0.00%
r	t	157	18	175	89.71%	10.29%
r	u	6	0	6	100.00%	0.00%
r	v	111	4	115	96.52%	3.48%
r	z	30	1	31	96.77%	3.23%
r	ç	9	0	9	100.00%	0.00%
r	ó	1	0	1	100.00%	0.00%
s	#	5	0	5	100.00%	0.00%
s	a	23	0	23	100.00%	0.00%
s	b	6	2	8	75.00%	25.00%
s	c	24	0	24	100.00%	0.00%
s	d	15	1	16	93.75%	6.25%
s	g	31	0	31	100.00%	0.00%
s	i	29	1	30	96.67%	3.33%
s	j	6	0	6	100.00%	0.00%
s	l	20	6	26	76.92%	23.08%
s	m	21	0	21	100.00%	0.00%
s	n	19	0	19	100.00%	0.00%
s	p	12	0	12	100.00%	0.00%
s	q	14	0	14	100.00%	0.00%
s	r	79	7	86	91.86%	8.14%
s	s	14	2	16	87.50%	12.50%
s	t	8	8	16	50.00%	50.00%
s	u	5	0	5	100.00%	0.00%
s	v	6	0	6	100.00%	0.00%
s	x	10	1	11	90.91%	9.09%
s	í	1	0	1	100.00%	0.00%
s	z	1	0	1	100.00%	0.00%
s	ç	2	0	2	100.00%	0.00%
t	#	2	1	3	66.67%	33.33%
t	a	75	0	75	100.00%	0.00%
t	b	5	1	6	83.33%	16.67%
t	c	40	10	50	80.00%	20.00%
t	d	7	0	7	100.00%	0.00%
t	f	1	0	1	100.00%	0.00%
t	g	16	0	16	100.00%	0.00%
t	i	103	0	103	100.00%	0.00%
t	j	14	0	14	100.00%	0.00%
t	l	59	28	87	67.82%	32.18%
t	m	25	0	25	100.00%	0.00%
t	n	22	0	22	100.00%	0.00%
t	ã	1	0	1	100.00%	0.00%
t	o	11	0	11	100.00%	0.00%
t	p	6	1	7	85.71%	14.29%

t	q	3	0	3	100.00%	0.00%
t	r	248	24	272	91.18%	8.82%
t	s	44	4	48	91.67%	8.33%
t	t	17	6	23	73.91%	26.09%
t	u	3	0	3	100.00%	0.00%
t	v	3	0	3	100.00%	0.00%
t	x	3	1	4	75.00%	25.00%
t	í	4	0	4	100.00%	0.00%
t	z	18	0	18	100.00%	0.00%
t	ç	1	0	1	100.00%	0.00%
t	ó	1	0	1	100.00%	0.00%
t	ú	1	0	1	100.00%	0.00%
u	#	1	0	1	100.00%	0.00%
u	a	24	0	24	100.00%	0.00%
u	b	10	7	17	58.82%	41.18%
u	c	19	4	23	82.61%	17.39%
u	d	5	2	7	71.43%	28.57%
u	f	1	0	1	100.00%	0.00%
u	g	1	0	1	100.00%	0.00%
u	i	68	0	68	100.00%	0.00%
u	j	13	0	13	100.00%	0.00%
u	l	11	8	19	57.89%	42.11%
u	m	2	0	2	100.00%	0.00%
u	n	4	0	4	100.00%	0.00%
u	o	1	0	1	100.00%	0.00%
u	p	0	1	1	0.00%	100.00%
u	r	24	4	28	85.71%	14.29%
u	s	12	1	13	92.31%	7.69%
u	t	25	3	28	89.29%	10.71%
u	z	8	0	8	100.00%	0.00%
v	a	1	0	1	100.00%	0.00%
v	c	8	0	8	100.00%	0.00%
v	d	12	0	12	100.00%	0.00%
v	e	2	0	2	100.00%	0.00%
v	g	11	0	11	100.00%	0.00%
v	i	18	0	18	100.00%	0.00%
v	j	11	1	12	91.67%	8.33%
v	l	331	13	344	96.22%	3.78%
v	n	30	0	30	100.00%	0.00%
v	r	196	0	196	100.00%	0.00%
v	s	34	4	38	89.47%	10.53%
v	t	12	2	14	85.71%	14.29%
v	x	5	1	6	83.33%	16.67%
v	í	1	0	1	100.00%	0.00%
v	z	9	0	9	100.00%	0.00%
x	a	3	0	3	100.00%	0.00%
x	c	5	0	5	100.00%	0.00%
x	l	0	1	1	0.00%	100.00%

x	i	5	0	5	100.00%	0.00%
x	m	1	0	1	100.00%	0.00%
x	n	1	0	1	100.00%	0.00%
x	p	1	0	1	100.00%	0.00%
x	q	0	1	1	0.00%	100.00%
x	r	20	0	20	100.00%	0.00%
x	t	1	0	1	100.00%	0.00%
x	z	1	0	1	100.00%	0.00%
y	u	1	0	1	100.00%	0.00%
z	a	3	0	3	100.00%	0.00%
z	b	1	0	1	100.00%	0.00%
z	d	5	0	5	100.00%	0.00%
z	g	1	0	1	100.00%	0.00%
z	i	12	0	12	100.00%	0.00%
z	j	2	0	2	100.00%	0.00%
z	l	11	4	15	73.33%	26.67%
z	m	2	0	2	100.00%	0.00%
z	n	7	0	7	100.00%	0.00%
z	r	32	1	33	96.97%	3.03%
z	s	2	0	2	100.00%	0.00%
z	t	3	0	3	100.00%	0.00%
z	v	3	0	3	100.00%	0.00%
ü	s	1	2	3	33.33%	66.67%
Ü	t	1	0	1	100.00%	0.00%

Anexo 8 Resultado da análise estatística para a vogal “o”

A seguir vê-se o resultado da análise estatística da ocorrência de ‘o’ fechado e ‘O’ aberto, onde ‘#’ significa nada (nenhum caracter).

Letra Anterior	Letra Posterior	Casos - o	Casos - O	Total de Casos	%o	%O
#	b	59	0	59	100.00%	0.00%
#	c	25	0	25	100.00%	0.00%
#	d	5	0	5	100.00%	0.00%
#	e	1	0	1	100.00%	0.00%
#	f	22	0	22	100.00%	0.00%
#	i	6	0	6	100.00%	0.00%
#	j	1	0	1	100.00%	0.00%
#	l	21	0	21	100.00%	0.00%
#	m	5	0	5	100.00%	0.00%
#	n	4	0	4	100.00%	0.00%
#	p	30	0	30	100.00%	0.00%
#	r	57	0	57	100.00%	0.00%
#	s	18	0	18	100.00%	0.00%
#	t	5	0	5	100.00%	0.00%
#	u	20	0	20	100.00%	0.00%
#	v	12	0	12	100.00%	0.00%
#	x	6	0	6	100.00%	0.00%
#	z	1	0	1	100.00%	0.00%
-	u	1	0	1	100.00%	0.00%
a	l	1	0	1	100.00%	0.00%
a	r	1	0	1	100.00%	0.00%
b	a	8	0	8	100.00%	0.00%
b	b	12	0	12	100.00%	0.00%
b	c	25	1	26	96.15%	3.85%
b	d	6	0	6	100.00%	0.00%
b	e	1	0	1	100.00%	0.00%
b	f	3	0	3	100.00%	0.00%
b	g	1	0	1	100.00%	0.00%
b	i	11	0	11	100.00%	0.00%
b	j	2	0	2	100.00%	0.00%
b	l	49	1	50	98.00%	2.00%
b	m	2	0	2	100.00%	0.00%
b	n	29	0	29	100.00%	0.00%
b	q	3	0	3	100.00%	0.00%
b	r	79	0	79	100.00%	0.00%
b	s	9	0	9	100.00%	0.00%
b	t	26	0	26	100.00%	0.00%
b	u	1	0	1	100.00%	0.00%

b	v	1	0	1	100.00%	0.00%
b	x	2	0	2	100.00%	0.00%
b	ç	3	0	3	100.00%	0.00%
c	#	32	0	32	100.00%	0.00%
c	-	4	0	4	100.00%	0.00%
c	a	23	0	23	100.00%	0.00%
c	b	30	0	30	100.00%	0.00%
c	c	22	0	22	100.00%	0.00%
c	d	4	0	4	100.00%	0.00%
c	e	14	0	14	100.00%	0.00%
c	f	4	0	4	100.00%	0.00%
c	g	9	0	9	100.00%	0.00%
c	i	15	0	15	100.00%	0.00%
c	l	115	1	116	99.14%	0.86%
c	m	100	0	100	100.00%	0.00%
c	n	28	0	28	100.00%	0.00%
c	o	10	0	10	100.00%	0.00%
c	p	14	0	14	100.00%	0.00%
c	q	3	0	3	100.00%	0.00%
c	r	205	0	205	100.00%	0.00%
c	á	1	0	1	100.00%	0.00%
c	s	36	1	37	97.29%	2.71%
c	t	35	2	37	94.59%	5.41%
c	u	8	0	8	100.00%	0.00%
c	v	13	1	14	92.85%	7.15%
c	x	3	0	3	100.00%	0.00%
c	z	8	0	8	100.00%	0.00%
c	ç	2	0	2	100.00%	0.00%
c	ó	2	0	2	100.00%	0.00%
d	#	7	0	7	100.00%	0.00%
d	a	8	0	8	100.00%	0.00%
d	b	7	0	7	100.00%	0.00%
d	c	9	0	9	100.00%	0.00%
d	d	2	0	2	100.00%	0.00%
d	e	9	0	9	100.00%	0.00%
d	g	1	0	1	100.00%	0.00%
d	i	6	0	6	100.00%	0.00%
d	l	20	0	20	100.00%	0.00%
d	m	26	0	26	100.00%	0.00%
d	n	10	0	10	100.00%	0.00%
d	p	3	0	3	100.00%	0.00%
d	q	0	1	1	0.00%	100.00%
d	r	340	1	341	99.70%	0.30%
d	á	1	0	1	100.00%	0.00%
d	s	26	2	28	92.85%	7.15%
d	t	5	0	5	100.00%	0.00%
d	u	19	0	19	100.00%	0.00%
d	v	2	0	2	100.00%	0.00%

d	í	1	0	1	100.00%	0.00%
d	x	0	1	1	0.00%	100.00%
d	z	1	0	1	100.00%	0.00%
d	ç	3	0	3	100.00%	0.00%
e	#	2	0	2	100.00%	0.00%
e	-	3	0	3	100.00%	0.00%
e	a	1	0	1	100.00%	0.00%
e	b	1	0	1	100.00%	0.00%
e	c	7	0	7	100.00%	0.00%
e	d	1	0	1	100.00%	0.00%
e	f	1	0	1	100.00%	0.00%
e	g	3	0	3	100.00%	0.00%
e	l	9	0	9	100.00%	0.00%
e	m	2	0	2	100.00%	0.00%
e	n	1	0	1	100.00%	0.00%
e	p	2	0	2	100.00%	0.00%
e	r	8	0	8	100.00%	0.00%
e	s	2	1	3	66.66%	33.34%
e	t	1	0	1	100.00%	0.00%
f	b	7	0	7	100.00%	0.00%
f	c	7	1	8	87.5%	12.5%
f	f	3	0	3	100.00%	0.00%
f	g	17	0	17	100.00%	0.00%
f	i	3	0	3	100.00%	0.00%
f	l	25	0	25	100%	0.00%
f	m	5	0	5	100.00%	0.00%
f	n	17	0	17	100.00%	0.00%
f	q	1	0	1	100.00%	0.00%
f	r	99	0	99	100%	0.00%
f	s	6	1	7	85.71%	14.29%
f	t	7	1	8	87.5%	12.5%
g		6	0	6	100.00%	0.00%
g	a	2	0	2	100.00%	0.00%
g	c	6	0	6	100.00%	0.00%
g	d	5	0	5	100.00%	0.00%
g	e	1	0	1	100.00%	0.00%
g	g	4	0	4	100.00%	0.00%
g	i	3	0	3	100.00%	0.00%
g	l	13	0	13	100.00%	0.00%
g	m	4	0	4	100.00%	0.00%
g	n	21	0	21	100.00%	0.00%
g	r	29	0	29	100.00%	0.00%
g	s	20	0	20	100.00%	0.00%
g	t	12	0	12	100.00%	0.00%
g	u	3	0	3	100.00%	0.00%
g	v	6	0	6	100.00%	0.00%
g	í	2	0	2	100.00%	0.00%
g	z	8	0	8	100.00%	0.00%

h	#	2	0	2	100.00%	0.00%
h	-	1	0	1	100.00%	0.00%
h	a	7	0	7	100.00%	0.00%
h	c	12	0	12	100.00%	0.00%
h	e	2	0	2	100.00%	0.00%
h	f	4	0	4	100.00%	0.00%
h	j	1	0	1	100.00%	0.00%
h	l	2	1	3	66.66%	33.34%
h	m	12	0	12	100.00%	0.00%
h	n	8	0	8	100.00%	0.00%
h	p	1	0	1	100.00%	0.00%
h	r	44	0	44	100.00%	0.00%
h	s	27	0	27	100.00%	0.00%
h	t	9	0	9	100.00%	0.00%
h	u	3	0	3	100.00%	0.00%
h	v	2	0	2	100.00%	0.00%
i	#	18	0	18	100.00%	0.00%
i	c	3	0	3	100.00%	0.00%
i	d	2	0	2	100.00%	0.00%
i	f	2	0	2	100.00%	0.00%
i	g	10	0	10	100.00%	0.00%
i	i	1	0	1	100.00%	0.00%
i	l	33	0	33	100.00%	0.00%
i	m	3	0	3	100.00%	0.00%
i	n	126	0	126	100.00%	0.00%
i	p	3	0	3	100.00%	0.00%
i	q	2	0	2	100.00%	0.00%
i	r	20	0	20	100.00%	0.00%
i	s	72	0	72	100.00%	0.00%
i	t	8	1	9	88.88%	11.12%
i	u	1	0	1	100.00%	0.00%
j	#	2	0	2	100.00%	0.00%
j	a	9	0	9	100.00%	0.00%
j	c	1	0	1	100.00%	0.00%
j	e	3	0	3	100.00%	0.00%
j	g	5	0	5	100.00%	0.00%
j	l	1	0	1	100.00%	0.00%
j	ã	1	0	1	100.00%	0.00%
j	q	1	0	1	100.00%	0.00%
j	r	10	0	10	100.00%	0.00%
j	s	8	0	8	100.00%	0.00%
j	u	1	0	1	100.00%	0.00%
j	v	2	0	2	100.00%	0.00%
l	a	1	0	1	100.00%	0.00%
l	b	10	0	10	100.00%	0.00%
l	c	28	0	28	100.00%	0.00%
l	d	9	0	9	100.00%	0.00%
l	f	3	0	3	100.00%	0.00%

l	g	87	0	87	100.00%	0.00%
l	i	2	0	2	100.00%	0.00%
l	j	6	1	7	85.71%	14.29%
l	l	1	0	1	100.00%	0.00%
l	m	11	0	11	100.00%	0.00%
l	n	17	0	17	100.00%	0.00%
l	p	4	0	4	100.00%	0.00%
l	q	8	1	9	88.88%	11.12%
l	r	76	1	77	98.70%	1.30%
l	s	34	0	34	100.00%	0.00%
l	t	16	2	18	88.89%	11.11%
l	u	20	0	20	100.00%	0.00%
l	w	1	0	1	100.00%	0.00%
m	#	16	0	16	100.00%	0.00%
m	b	12	0	12	100.00%	0.00%
m	c	16	0	16	100.00%	0.00%
m	d	34	0	34	100.00%	0.00%
m	e	8	0	8	100.00%	0.00%
m	f	12	0	12	100.00%	0.00%
m	g	5	0	5	100.00%	0.00%
m	i	4	0	4	100.00%	0.00%
m	k	1	0	1	100.00%	0.00%
m	l	42	0	42	100.00%	0.00%
m	m	6	0	6	100.00%	0.00%
m	n	50	0	50	100.00%	0.00%
m	p	2	0	2	100.00%	0.00%
m	q	1	0	1	100.00%	0.00%
m	r	118	1	119	99.16%	0.84%
m	s	36	1	37	97.29%	2.71%
m	t	21	0	21	100.00%	0.00%
m	u	7	0	7	100.00%	0.00%
m	v	16	0	16	100.00%	0.00%
m	x	1	0	1	100.00%	0.00%
m	í	1	0	1	100.00%	0.00%
m	ç	13	0	13	100.00%	0.00%
n	#	3	0	3	100.00%	0.00%
n	a	2	0	2	100.00%	0.00%
n	b	7	0	7	100.00%	0.00%
n	c	8	0	8	100.00%	0.00%
n	d	3	0	3	100.00%	0.00%
n	f	4	0	4	100.00%	0.00%
n	g	11	0	11	100.00%	0.00%
n	i	10	0	10	100.00%	0.00%
n	j	4	0	4	100.00%	0.00%
n	l	9	0	9	100.00%	0.00%
n	m	32	0	32	100.00%	0.00%
n	n	6	0	6	100.00%	0.00%
n	p	7	0	7	100.00%	0.00%

n	r	26	2	28	92.85%	7.15%
n	s	26	0	26	100.00%	0.00%
n	t	21	1	22	95.45%	4.55%
n	u	2	0	2	100.00%	0.00%
n	v	19	0	19	100.00%	0.00%
n	w	1	0	1	100.00%	0.00%
n	x	1	0	1	100.00%	0.00%
n	z	1	0	1	100.00%	0.00%
n	ç	1	0	1	100.00%	0.00%
ã	#	237	0	237	100.00%	0.00%
ã	-	1	0	1	100.00%	0.00%
o	l	3	1	4	75.00%	25.00%
o	p	4	0	4	100.00%	0.00%
o	r	3	0	3	100.00%	0.00%
p	#	1	0	1	100.00%	0.00%
p	a	1	0	1	100.00%	0.00%
p	b	3	0	3	100.00%	0.00%
p	c	7	0	7	100.00%	0.00%
p	d	10	0	10	100.00%	0.00%
p	e	12	0	12	100.00%	0.00%
p	g	3	0	3	100.00%	0.00%
p	i	7	0	7	100.00%	0.00%
p	j	3	0	3	100.00%	0.00%
p	l	78	0	78	100.00%	0.00%
p	m	4	0	4	100.00%	0.00%
p	n	16	0	16	100.00%	0.00%
p	p	12	0	12	100.00%	0.00%
p	q	2	0	2	100.00%	0.00%
p	r	120	1	121	99.17%	0.83%
p	s	82	1	83	98.79%	1.21%
p	t	22	0	22	100.00%	0.00%
p	é	1	0	1	100.00%	0.00%
p	u	11	0	11	100.00%	0.00%
p	v	8	0	8	100.00%	0.00%
p	ç	4	0	4	100.00%	0.00%
r	#	35	0	35	100.00%	0.00%
r	a	21	0	21	100.00%	0.00%
r	b	21	0	21	100.00%	0.00%
r	c	69	1	70	98.57%	1.43%
r	d	40	0	40	100.00%	0.00%
r	e	9	0	9	100.00%	0.00%
r	f	36	0	36	100.00%	0.00%
r	g	33	0	33	100.00%	0.00%
r	i	6	0	6	100.00%	0.00%
r	j	10	0	10	100.00%	0.00%
r	l	53	1	54	98.15%	1.85%
r	m	48	0	48	100.00%	0.00%
r	n	45	0	45	100.00%	0.00%

r	o	1	0	1	100.00%	0.00%
r	p	59	0	59	100.00%	0.00%
r	q	3	0	3	100.00%	0.00%
r	r	12	0	12	100.00%	0.00%
r	s	122	1	123	99.18%	0.82%
r	t	56	1	57	98.24%	1.76%
r	u	19	0	19	100.00%	0.00%
r	v	56	0	56	100.00%	0.00%
r	x	7	0	7	100.00%	0.00%
r	í	5	0	5	100.00%	0.00%
r	z	1	0	1	100.00%	0.00%
r	ç	19	0	19	100.00%	0.00%
s	#	6	0	6	100.00%	0.00%
s	a	13	0	13	100.00%	0.00%
s	b	57	0	57	100.00%	0.00%
s	c	24	0	24	100.00%	0.00%
s	d	3	0	3	100.00%	0.00%
s	e	1	0	1	100.00%	0.00%
s	f	12	0	12	100.00%	0.00%
s	g	1	0	1	100.00%	0.00%
s	l	81	0	81	100.00%	0.00%
s	m	6	1	7	85.72%	14.28%
s	n	34	0	34	100.00%	0.00%
s	p	11	0	11	100.00%	0.00%
s	q	2	0	2	100.00%	0.00%
s	r	69	0	69	100.00%	0.00%
s	s	9	0	9	100.00%	0.00%
s	t	6	0	6	100.00%	0.00%
s	u	8	0	8	100.00%	0.00%
s	v	8	0	8	100.00%	0.00%
s	z	1	0	1	100.00%	0.00%
s	ç	3	0	3	100.00%	0.00%
t		10	0	10	100.00%	0.00%
t	a	16	0	16	100.00%	0.00%
t	c	26	0	26	100.00%	0.00%
t	d	11	0	11	100.00%	0.00%
t	e	2	0	2	100.00%	0.00%
t	f	3	0	3	100.00%	0.00%
t	g	11	0	11	100.00%	0.00%
t	i	3	0	3	100.00%	0.00%
t	j	1	0	1	100.00%	0.00%
t	l	33	0	33	100.00%	0.00%
t	m	24	0	24	100.00%	0.00%
t	n	26	0	26	100.00%	0.00%
t	p	20	0	20	100.00%	0.00%
t	q	1	0	1	100.00%	0.00%
t	r	193	0	193	100.00%	0.00%
t	s	35	0	35	100.00%	0.00%

t	t	5	0	5	100.00%	0.00%
t	u	15	0	15	100.00%	0.00%
t	v	4	0	4	100.00%	0.00%
t	x	5	0	5	100.00%	0.00%
t	z	1	0	1	100.00%	0.00%
u	#	2	0	2	100.00%	0.00%
u	c	1	0	1	100.00%	0.00%
u	d	1	0	1	100.00%	0.00%
é	l	3	0	3	100.00%	0.00%
u	s	18	0	18	100.00%	0.00%
u	t	3	0	3	100.00%	0.00%
v	a	16	0	16	100.00%	0.00%
v	c	22	0	22	100.00%	0.00%
v	e	3	0	3	100.00%	0.00%
v	g	8	0	8	100.00%	0.00%
v	l	53	0	53	100.00%	0.00%
v	m	1	0	1	100.00%	0.00%
v	n	1	0	1	100.00%	0.00%
v	r	35	0	35	100.00%	0.00%
v	s	6	0	6	100.00%	0.00%
v	t	4	0	4	100.00%	0.00%
v	u	2	0	2	100.00%	0.00%
v	v	2	0	2	100.00%	0.00%
v	z	2	0	2	100.00%	0.00%
v	ç	1	0	1	100.00%	0.00%
x	d	2	0	2	100.00%	0.00%
x	f	2	0	2	100.00%	0.00%
x	n	4	0	4	100.00%	0.00%
x	r	8	0	8	100.00%	0.00%
x	t	6	0	6	100.00%	0.00%
x	v	2	0	2	100.00%	0.00%
í	c	1	0	1	100.00%	0.00%
í	d	1	0	1	100.00%	0.00%
í	l	2	0	2	100.00%	0.00%
í	p	1	0	1	100.00%	0.00%
z	a	4	0	4	100.00%	0.00%
z	e	1	0	1	100.00%	0.00%
z	f	2	0	2	100.00%	0.00%
z	i	1	0	1	100.00%	0.00%
z	n	3	0	3	100.00%	0.00%
z	o	2	0	2	100.00%	0.00%
z	r	1	0	1	100.00%	0.00%
z	á	1	0	1	100.00%	0.00%
z	t	2	0	2	100.00%	0.00%
z	u	1	0	1	100.00%	0.00%
ç	#	4	0	4	100.00%	0.00%
ç	a	14	0	14	100.00%	0.00%
ç	b	3	0	3	100.00%	0.00%

ç	d	1	0	1	100.00%	0.00%
ç	e	1	0	1	100.00%	0.00%
ç	i	3	0	3	100.00%	0.00%
ç	n	4	0	4	100.00%	0.00%
ç	s	10	0	10	100.00%	0.00%
ç	u	2	0	2	100.00%	0.00%
ô		3	0	3	100.00%	0.00%

Referências

- [AND 82] ANDRÉ, Hildebrando A. de. **Gramática Ilustrada**. 3.ed. São Paulo: Moderna, 1982.
- [BAI 90] BAILLY, Gérard; LABOISSIÈRE, Rafael; GALVÁN, Arturo. **Learning to Speak: Speech Production and Sensori-Motor Representations**. Grenoble: [s.n.], 1990
- [BAR 2001] BARONE, Dante A. C.; COLLISCHON, Gisela et al. SPOLTECH – Advancing Human Language Technology in Brazil and the United States Through Collaborative Research on Portuguese Spoken Language Systems. **ProTeM-CC – Projects Evaluation Workshop – International Cooperation: Proceeding**. Rio de Janeiro: [s.n.], 2001. p. 118-142.
- [BLA 99] BLACK, Alan W.; TAYLOR, Paul; CALEY, Richard. **The Festival Speech Synthesis System System documentation Edition 1.4, for Festival Version 1.4.0**. 1999. Disponível em: <http://www.cstr.ed.ac.uk/projects/festival/manual/festival_21.html>. Acesso em: 27 jan. 2000.
- [BOR 94] BORDEN, Gloria J; HARRIS, Katherine; RAPHAEL, Lawrence. **Speech Science Primer: Physiology, Acoustics, and Perception of Speech**. 3rd. ed. Baltimore: Williams & Wilkins, 1994.
- [CAM 90] CAMPOS, Geraldo Lino de; CHBANE, Dimas Trevisan. **Translating Text to Phonemes for the Portuguese Language**. São Paulo: Departamento de Engenharia de Computação e Sistemas Digitais, Escola Politécnica da Universidade de São Paulo, 1990.
- [DUT 97] DUTOIT, Thierry. **An Introduction to Text-to-Speech Synthesis**. Netherland: Kluwer Academic Publishers, 1997
- [FUR 89] FURUI, Sadaoki. **Digital Speech Processing, Synthesis and Recognition**. New York: Marcel Dekker, 1989.
- [HOS 2000] HOSOM, Paul. **The CSLU Toolkit: A Platform for Research and Development of Spoken-Language Systems**. Oregon: Oregon Health & Science University, 2000. Transparências.
- [HUG 95] HUGO, Marcelo. **Uma interface de reconhecimento de voz para o sistema de gerenciamento de central de informação de fretes**. Florianópolis: Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 1995.
- [LEM 90] LEMMETTY, Sami. **Review of speech synthesis technology**. 1990. Disponível em: <www.acoustics.hut.fi/~slemmet/dippa/chap1.html>. Acesso em: 27 jan. 2000..
- [MOS 90] MOSSER, Von. **Introduction to speech synthesis**. 1990. Disponível em: <www.clark.net/pub/mhsmedia/physics/intro.html>. Acesso em: 27 jan. 2000.
- [OLI 96] OLIVEIRA, Elvira de. **Help! Sistema de consulta interativa – Língua Portuguesa**. Porto Alegre: Zero Hora, 1996.
- [OSH 2000] O'SHAUGHNESSY, Douglas. **Speech Communications**:

- Human and Machine. 2nd ed. New York: IEEE Press, 2000.
- [SIL 75] SILVA, Adalberto Prado e et al. **Grande Dicionário Brasileiro Melhoramentos**. São Paulo: Melhoramentos, 1975.
- [SIL 99] SILVA, Thaís Cristófaró. **Fonética e Fonologia do Português**:roteiro de estudos e guia de exercícios. São Paulo:Contexto, 1999.

Obras Consultadas

- [ALL 87] ALLEN, John; HUNNICUT, Sharon; KLATT, Dennis H. **From Text to Speech: The MITalk System**. Massachusetts: MIT Press, 1987.
- [BAR 2000] BARONE, Dante et al. A Brazilian Portuguese Language Corpus Development. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, ICSLP, 6., 2000, Beijing. **Proceedings...**, [S.l.: s.n.], 2000. p. 579-582.
- [DEL 93] DELLER JUNIOR, J. R.; PROAKIS, J. G.; HANSEN, J. H. L. **Discrete-time processing of speech signals**. New York: Macmillan, 1993.
- [FLA 72] FLANAGAN, J. L. **Speech Analysis, Synthesis and Perception**. 2nd ed. Berlin: Springer-Verlag, 1972.
- [FUR 92] FURUI, S.; SONDHI, M. M. **Advances in speech signal processing**. New York: Marcel Dekker, 1992.
- [GOL 2000] GOLD, B.; MORGAN, N. **Speech and Audio Signal Processing**. New York: John Wiley & Sons, 2000.
- [GOM 99] GÓMEZ CIPRIANO, José Luis. **Modelos Ocultos de Markov para Sistemas de Reconhecimento Automático de Voz**. Exame de Qualificação (Doutorado em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul. Porto Alegre, 1999.
- [HAU 99] HAUSSER, Roland R. **Foundations of computational linguistics: man-machine communication in natural language**. Berlin: Verlag, 1999.
- [HER 2000] HERTZ, Susan R.; YOUNES, Rebecca J.; HOSKINS, Steven R. **Space, Speed, Quality, and Flexibility: Advantages of Rule-Based Speech Synthesis**. San Jose: Eloquent Technology, 2000.
- [HOL 88] HOLMES, J. N. **Speech Synthesis and Recognition**. Wokingham, Berkshire, England: Van Nostrand Reinhold (UK), 1988.
- [HOS 2000] HOSOM, John-Paul. **CSLU Toolkit**. 1990. Disponível em: <www.cslu.ogi.edu>. Acesso em: 20 jul. 2000.
- [KLE 95] KLEIJN, W.B; PALIWAL, K.K. **Speech Coding and Synthesis**. Amsterdam: Elsevier, 1995.
- [PAR 87] PARSONS, W. **Voice and speech processing**. New York: McGraw-Hill, 1987.
- [RAB 78] RABINER, L. R.; SCHAFER, R. W. **Digital processing of speech signals**. Englewood Cliffs, N.J.: Prentice-Hall, 1978.
- [RAM 97] RAMAN, T. V. **Auditory user interfaces – Toward the speaking Computer**. Boston: Kluwer Academic Publishers, 1997.
- [SAN 97] SANTEN, J.P.H. van et al. **Progress in Speech Synthesis**. New York: Springer-Verlag, 1997.
- [SAV 88] SAVADOVSKY, Pedro. **Introdução ao projeto de interfaces em linguagem natural**. São Paulo: SID Informática, 1988.
- [THR 91] THRO, Ellen. **The artificial intelligence dictionary**. San Marcos : Microtrend Books, 1991.

- [TRI 94] TRINDADE, Ana Lizete de Oliveira. **Identificação dos Padrões Espectrais de Fonemas da Língua Portuguesa e sua Localização Automática em Amostras de Fala.** 1994. Trabalho Individual. Instituto de Informática, Universidade Federal do Rio Grande do Sul. Porto Alegre.
- [WIT 82] WITTEN, I. H. **Principles of Computer Speech.** London: Academic Press, Inc., 1982.