

**UNIVERSIDADE DE SÃO PAULO**

**ESCOLA POLITÉCNICA**  
**100 ANOS**

DIMAS TREVIZAN CHBANE

**DESENVOLVIMENTO DE SISTEMA PARA  
CONVERSÃO DE TEXTOS EM FONEMAS  
NO IDIOMA PORTUGUÊS**

Dissertação apresentada à Escola  
Politécnica da Universidade de  
São Paulo para obtenção do  
título de Mestre em Engenharia.

**São Paulo**  
**Novembro 1994**

DIMAS TREVIZAN CHBANE

**DESENVOLVIMENTO DE SISTEMA PARA  
CONVERSÃO DE TEXTOS EM FONEMAS  
NO IDIOMA PORTUGUÊS**

Dissertação apresentada à Escola  
Politécnica da Universidade de  
São Paulo para obtenção do  
título de Mestre em Engenharia.

Área de Concentração:  
Engenharia de Eletricidade

Orientador:  
Prof. Dr. Geraldo Lino de Campos

1994

## **AGRADECIMENTOS**

Ao orientador Prof. Dr. Geraldo Lino de Campos pela oportunidade da realização deste estudo, pelo profissionalismo e amizade com que sempre se prontificou para a orientação desta dissertação e pelo apoio indispensável para o desenvolvimento deste trabalho.

À Escola Politécnica e especialmente ao Departamento de Eletrônica Digital, que através do estímulo de vários Professores e Colegas, e da dedicação de diversos Funcionários cooperaram imensamente para a realização desta pesquisa.

Aos prezados Professores e amigos Edgard Casaes, João José Neto, Maria Céu Viana, Suzan Hertz e Yandeci, pela atenção e inestimáveis colaborações.

Aos Engenheiros e Diretores da Spike Eletrônica pela amizade e consideração.

E a todos aqueles que direta ou indiretamente contribuíram para a realização desta dissertação.

# **DESENVOLVIMENTO DE SISTEMA PARA CONVERSÃO DE TEXTOS EM FONEMAS NO IDIOMA PORTUGUÊS**

## **Sumário**

Lista de Figuras

Lista de Tabelas

Resumo

Abstract

## **CAPÍTULO I INTRODUÇÃO**

1.1 Considerações Iniciais

1.2 Objetivo

1.3 Conteúdo

1.4 Justificativa

1.5 Pesquisas na Área

## **CAPÍTULO II CONSIDERAÇÕES LINGÜÍSTICAS**

2.1 Características da Voz Humana

2.1.1 Principais Aspectos

2.1.2 Sons da Fala

2.2 Conceituação de Fonema e Relação Letra-Fonema

2.3 Características do Alfabeto Fonético

2.4 Classificação dos Sons Lingüísticos

2.4.1 Classificação das Vogais

2.4.2 Classificação das Consoantes

## **CAPÍTULO III**

### **CONVERSÃO DE TEXTO PARA FONEMAS**

3.1 Aspectos Conceituais sobre o Processo de Síntese de Voz a Partir de Texto

3.2 Conversão de Texto para Fonemas ao Nível de Palavras

3.2.1 Conversão de Letra para Fonema

3.2.2 Dicionário de Exceções

3.2.3 Compiladores de Regras

3.3 Conversão de Texto para Fonemas para o Idioma Português

## **CAPÍTULO IV**

### **DESENVOLVIMENTO DO SISTEMA**

4.1 Ambiente e Interface do Sistema

4.2 Principais Características do Dicionário

4.2.1 Conteúdo do Dicionário

4.2.2 Estrutura de Armazenamento do Dicionário

4.2.3 Montagem e Estrutura de Dados do Dicionário

4.2.4 Busca de Palavras no Dicionário

4.3 Regras de Conversão de Letras para Fonemas

4.3.1 Separação de Sílabas

4.3.2 Determinação da Sílabas Tônicas

4.3.3 Aplicação das Regras Adotadas

4.3.3.1 Arquivo REGRAS.CON

4.3.3.2 Aplicação de Regras na Transcrição de Palavras

4.3.3.3 Aplicação de Regras na Montagem do Dicionário

4.4 Análise dos Resultados

4.4.1 Análise dos Resultados da Aplicação do Conjunto de Regras para Conversão de Letras para Fonemas

4.4.2 Análise dos Resultados dos Testes de Estrutura de Armazenamento do Dicionário

## **CAPÍTULO V**

### **CONSIDERAÇÕES FINAIS**

5.1 Conclusões

5.2 Sugestões para Futuros Estudos

**ANEXO A.** Regras de Conversão de Letras em Fonemas no Arquivo  
REGRAS.CON

**ANEXO B.** Palavras Submetidas aos Algoritmos de Separação de Sílabas e  
Determinação da Sílabas Tônicas

**ANEXO C.** Textos Submetidos à Transcrição Fonológica

## **REFERÊNCIAS BIBLIOGRÁFICAS**

## **LISTA DE FIGURAS**

### **CAPÍTULO II CONSIDERAÇÕES LINGÜÍSTICAS**

FIGURA 2.1 - Diagrama esquemático do aparelho fonador humano. (p.)

FIGURA 2.2 - Diagrama esquemático dos componentes funcionais do trato vocálico.  
(FLANAGAN, 1972) (p.)

FIGURA 2.3 - Espectrograma da frase "Noon is the sleepy time of the day", com  
indicação das três primeiras formantes. (FLANAGAN et al., 1970) (p.)

### **CAPÍTULO III CONVERSÃO DE TEXTO PARA FONEMA**

FIGURA 3.1 - Elementos de um sistema de síntese de voz a partir de texto.  
(CROCHIERE e FLANAGAN, 1986) (p.)

FIGURA 3.2 - Freqüência acumulada de entradas em um corpus da Língua Inglesa  
em função do número de diferentes palavras, ordenadas por sua freqüência.  
(KLATT, 1989) (p.)

FIGURA 3.3 - Sistema de conversão de letra para fonema. (ALLEN, 1976) (p.)

### **CAPÍTULO IV DESENVOLVIMENTO DO SISTEMA**

FIGURA 4.1 - Esquema de conversão de palavras para seus respectivos fonemas  
utilizado. (p.)

FIGURA 4.2 - Autômato de reconhecimento das palavras *retorno* (substantivo), *retorno* (verbo, 1ª. pessoa do Presente do Indicativo); *contorno* (substantivo), *contorno* (verbo, 1ª. pessoa do Presente do Indicativo). (p.)

FIGURA 4.3 - Autômato de reconhecimento das palavras *chuva*, *casa* e *cebola*. (p.)

FIGURA 4.4 - Estrutura inicial do reconhecedor das palavras *canto*, *canta*, *cantei*, *nado*, *nada*, *nadei*, *pulo*, *pula*, *pulei*. (p.)

FIGURA 4.5 - Estrutura final do reconhecedor das palavras *canto*, *canta*, *cantei*, *nado*, *nada*, *nadei*, *pulo*, *pula*, *pulei*. (p.)

FIGURA 4.6 - Estrutura de dados final do autômato. (p.)

FIGURA 4.7 - Estrutura de dados do reconhecedor das palavras *aclive*, *acabar*, *acerola* e *achar*. (p.)

## **LISTA DE TABELAS**

### **CAPÍTULO II CONSIDERAÇÕES LINGÜÍSTICAS**

TABELA 2.1 - Alfabetos fonéticos para a Língua Portuguesa.

TABELA 2.2 - Classificação das vogais.

TABELA 2.3 - Classificação das consoantes.

### **CAPÍTULO III CONVERSÃO DE TEXTO PARA FONEMAS**

TABELA 3.1 - Regras de conversão de letras para fonemas para o Inglês Britânico com alfabeto IPA. (AINSWORTH, 1973)

TABELA 3.3 - Algumas regras para conversão fonética da letra "x". (ALIANDRO, 1974)

### **CAPÍTULO IV DESENVOLVIMENTO DO SISTEMA**

TABELA 4.1 - Transcrição fonológica da palavra *achar* a partir do reconhecedor de palavras.

TABELA 4.2 - Ações a serem tomadas durante a execução do algoritmo de separação de sílabas.

TABELA 4.3 - Transição de estados para separação de sílabas.

TABELA 4.4 - Seqüência de passos para separação de sílabas da palavra *pneumático*.

# **Desenvolvimento de Sistema para Conversão de Textos em Fonemas no Idioma Português**

## **Resumo**

Esta dissertação tem como propósito o desenvolvimento de um sistema de conversão de textos em fonemas para o Idioma Português, a ser utilizado em projetos de conversão de texto para voz.

São abordados alguns aspectos no que se refere aos fonemas, sua classificação quanto a parâmetros de base acústica e articulatória, e também a relação entre fonemas e a linguagem na forma escrita. Também são apresentadas algumas considerações referentes à produção da voz humana.

Discute-se aspectos relevantes com relação aos sistemas de síntese de voz a partir de texto, suas vantagens e aplicações. Focaliza-se principalmente os aspectos referentes à transcrição fonológica de palavras a partir de regras e a importância de um dicionário de exceções. Comenta-se as tendências mais modernas, como os compiladores de regras.

Detalhes de implementação do sistema de conversão de texto para fonemas na Língua Portuguesa são apresentados, destacando-se a estrutura de armazenamento do dicionário de exceções e os algoritmos usados na conversão a partir de regras, sendo indicados também os resultados alcançados pelo sistema desenvolvido.

Finalmente, são apresentadas as conclusões gerais decorrentes deste trabalho e algumas sugestões para novos estudos.

**Development of Text-to-Phoneme Translator  
System for Portuguese Language**

**Abstract**

# **CAPÍTULO I**

## **INTRODUÇÃO**

### **1.1 - CONSIDERAÇÕES INICIAIS**

O uso da linguagem escrita tem predominado em sistemas de computação desde o surgimento dos primeiros computadores, por ser a maneira mais eficiente, econômica e confiável para transmitir e armazenar informações. Porém, com o avanço da tecnologia e com a crescente tendência de interfaces homem-máquina amigáveis, o uso da linguagem falada em computadores vem tornando-se cada vez mais interessante, na medida em que a fala é uma forma mais natural de comunicação.

Dados estatísticos de Chaponis, citados por YOUNG e FALLSIDE (1979), mostraram que a comunicação da informação através da voz em situações de interação homem-máquina é em média duas vezes mais eficiente do que qualquer outra forma de comunicação. A maior eficiência da comunicação verbal, apoiada na crescente evolução das técnicas de processamento de sinais digitais, tem feito com que sistemas de compreensão da fala e síntese de voz difundam-se cada vez mais como meios de entrada e saída de informações em computadores.

Atualmente os sistemas de compreensão da fala estão restritos a algumas aplicações especiais, enquanto que os sistemas de síntese de voz vêm sendo largamente utilizados.

Dentre esses sistemas, aqueles capazes de sintetizar voz a partir de texto apresentam particular importância pois são capazes de produzir voz a partir de um texto de entrada, unindo a eficiência do armazenamento de dados na forma escrita com a comunicação através da fala.

Essa possibilidade de armazenar informações na forma escrita, fornecendo saídas através de voz, amplia o uso de computadores para diversas aplicações, dispensando a necessidade de treinamento ou equipamentos especiais para os usuários. Tais sistemas permitem, por exemplo, acesso a bancos de dados através de telefone, como os sistemas de consulta de saldo bancário, atualmente bastante difundidos, e abrem perspectivas para transformar os atuais sistemas de correio eletrônico em correio por voz.

Merecem destaque ainda, as facilidades advindas do uso desses sistemas para pessoas portadoras de deficiências vocais e visuais, através de máquinas de auxílio à fala e máquinas de leitura para cegos (ALLEN, 1973).

Ainda podem ser citados outros exemplos, como brinquedos e relógios, disponíveis no mercado há algum tempo e capazes de sintetizar um número limitado de sentenças a partir de um texto fixo. A tecnologia necessária na fabricação desses produtos normalmente é independente do idioma, ou seja, os equipamentos podem ser programados em vários idiomas, mas para um pequeno número de sentenças (LEE, TSENG e YOUNG, 1989).

Porém, os sistemas de síntese de voz a partir de texto irrestrito apresentam maior interesse na medida em que são capazes de sintetizar um número ilimitado de sentenças a partir de um texto de entrada, e vêm sendo desenvolvidos por diferentes grupos de pesquisa em diversos idiomas, especialmente o Inglês.

Tais sistemas são muito mais atrativos, pois apresentam maior flexibilidade e potencial utilização em uma vasta área de aplicações. No entanto, a tecnologia necessária para o desenvolvimento desses sistemas não somente é muito mais complexa e avançada, como também depende do idioma. Assim, apesar do método de desenvolvimento de sistemas de conversão de texto para voz ser muito semelhante para os vários idiomas, os aspectos fonéticos devem ser adaptados especificamente para cada um deles.

ALLEN (1976) divide o processo de conversão de texto para voz em duas etapas distintas. Primeiro o texto de entrada deve ser analisado, determinando-se sua estrutura lingüística, constituída basicamente por fonemas formadores das palavras que serão faladas. Os fonemas são as unidades mais simples da Língua, aos quais correspondem os sons elementares que entram na constituição das sílabas e dos vocábulos. Nesta etapa inicial, é importante a análise das pausas, da acentuação das palavras e sinais de pontuação para simular as características da prosódia da fala,

que são o acento, o ritmo e a entonação. Finalmente, as informações obtidas na etapa anterior são combinadas para produzir voz.

Percebe-se assim que a primeira etapa na conversão de texto para voz depende fundamentalmente de características lingüísticas de cada idioma, sendo a conversão de palavras para fonemas uma etapa básica para o desenvolvimento do processo. Nesse sentido, procurando viabilizar futuros projetos de conversão de texto irrestrito para voz, este trabalho trata especificamente da conversão de palavras para fonemas na Língua Portuguesa.

## **1.2 - OBJETIVO**

O objetivo básico deste trabalho é apresentar a implementação de uma estrutura de "software" capaz de permitir o desenvolvimento de um sistema de transcrição de palavras da Língua Portuguesa para sua correspondente representação fonológica.

Além do desenvolvimento de um conjunto de regras de transcrição de letras para fonemas, que permite efetuar a transcrição fonológica de palavras com base nas letras que a constituem, foi implementada uma estrutura de Dicionário, capaz de armazenar a transcrição fonológica e a classe gramatical de palavras que constituem exceções às regras.

## **1.3 - CONTEÚDO**

Esta dissertação está dividida em cinco capítulos. No Capítulo I são apresentadas além das vantagens e aplicações dos sistemas de síntese de voz a partir de texto, uma breve discussão sobre a importância da conversão de textos em fonemas e uma evolução cronológica das pesquisas na área.

No Capítulo II, apresenta-se inicialmente alguns aspectos relativos à produção da voz humana, como o aparelho fonador e a caracterização dos sons da fala, abordando também os parâmetros físicos relacionados com a voz humana. Em seguida, é apresentada a conceituação de fonema, como são representados na escrita e algumas considerações entre a linguagem falada e a escrita. E posteriormente, trata-se da classificação dos fonemas, segundo critérios de base articulatória e acústica.

No Capítulo III, faz-se inicialmente uma breve introdução com relação a aspectos conceituais sobre síntese de voz a partir de texto, apresentando as várias etapas desse processo. Descreve-se os componentes usualmente presentes em sistemas de conversão de texto para fonema, os quais correspondem às regras de conversão de letras para fonemas e ao dicionário de exceções. São apresentados alguns aspectos sobre compiladores de regras. E a seguir, passa-se ao estudo de questões referentes à conversão de textos para fonemas no Idioma Português.

No Capítulo IV, descreve-se os componentes do sistema desenvolvido e os principais algoritmos utilizados, sendo apresentados também os resultados alcançados com esse sistema.

Finalmente, no Capítulo V, são feitas algumas considerações, sugerindo-se alternativas para o prosseguimento dos estudos, além das conclusões resultantes desta pesquisa.

Constam ainda deste trabalho, como Anexos, o conjunto de regras utilizado na conversão de letras para fonemas, uma lista de palavras submetidas aos algoritmos de separação de sílabas e de determinação de sílaba tônica, e cinco textos submetidos à transcrição fonológica.

## 1.4 - JUSTIFICATIVA

Atualmente há uma vasta área de aplicações para voz a partir de computadores, e dependendo da finalidade, a técnica utilizada em sua produção é diferente. Para casos simples, nos quais é necessário um pequeno vocabulário em mensagens de formato rígido, a voz humana pode ser gravada tanto na forma digital como analógica e apropriadamente acessada para a formação da mensagem desejada. Desta forma, pode-se obter voz de alta qualidade, porém com uma faixa de saída limitada.

A tecnologia necessária para esse processo exige simplesmente uma técnica econômica de armazenamento e de rápido acesso, não envolvendo nenhum uso de ligüística ou fonética. Tem-se assim pouco processamento de voz às custas de substanciais requisitos de armazenamento. No entanto, quando é necessário produzir uma variedade maior de frases e sentenças, atinge-se um ponto a partir do qual não mais é possível o armazenamento de todas as mensagens, sendo então necessário a utilização de outras técnicas (ALLEN, 1976).

Os sistemas de síntese de voz a partir de texto visam atender essas aplicações, oferecendo grande flexibilidade na formação de mensagens, com capacidade praticamente infinita de frases de saída, produzindo voz a partir de um extenso processamento lingüístico. E sendo a conversão de textos para fonemas o ponto de partida para sistemas de síntese de voz a partir de texto, é fundamental o desenvolvimento de estudos nessa área.

Bastante aprimorada em outros idiomas, especialmente o Inglês, a conversão de texto para fonemas normalmente é feita através de um conjunto de regras de

transcrição de letras para fonemas, e de um dicionário contendo transcrições fonéticas de palavras que falham à aplicação das regras.

Para a Língua Portuguesa, assim como na Língua Inglesa, há ainda a necessidade do dicionário armazenar a classe gramatical das palavras, pois existem palavras que possuem diferentes pronúncias dependendo da classe gramatical a qual pertencem. Estas diferenças de pronúncia ocorrem mais frequentemente com verbos e substantivos, como por exemplo, f[o]rça, substantivo e f[ó]rça, verbo.

Por essa razão, além de um conjunto de regras que possam ser facilmente alteradas, também é necessário a elaboração de um dicionário de rápido acesso, que além da transcrição fonológica das palavras, possibilite futuras implementações de análises sintáticas. Assim, seguindo estas diretrizes, neste trabalho foi desenvolvido um conjunto de regras de conversão de letras para fonemas e uma estrutura de dados para armazenamento de um dicionário, capaz de conter informações fonológicas e gramaticais de palavras.

## **1.5 - PESQUISAS NA ÁREA**

A partir da década de 60, principalmente após 1965, o estudo de sistemas de síntese de voz a partir de texto teve um grande impulso, caracterizado por intensas pesquisas na área (ENDRES, 1983). Tratava-se de uma época de pioneirismo, marcada por dúvidas sobre viabilidade dos sistemas, e na qual iniciou-se o desenvolvimento das primeiras regras de conversão de texto para fonemas.

O início dos anos 70 marcou um novo período, com o aperfeiçoamento dos sintetizadores e dos algoritmos de conversão de texto para fonemas. Dois centros de

pesquisas sobressaíram nessa área, o "Bell Laboratories" pertencente à "American Telegraph & Telephony - AT&T", e o "Massachusetts Institute of Technology - MIT" (UMEDA, 1976).

Já no final da década de 70 e início dos anos 80, foram apresentados os primeiros protótipos de sistemas de síntese de voz a partir de texto com vocabulário ilimitado, com destaque para o "MITalk", desenvolvido no MIT durante a década de 70, sob a supervisão de Jonathan Allen (KAPLAN e LERNER, 1985).

Ainda nessa época surgiram os primeiros sistemas comerciais com vocabulário ilimitado, os quais vêm sendo continuamente aprimorados desde então.

KLATT (1987) fez uma ampla descrição cronológica sobre a evolução de equipamentos vendidos comercialmente. O primeiro sistema comercial com vocabulário ilimitado foi uma máquina de leitura para deficientes visuais, lançado em 1976, pela "Kurzweil", baseado no "chip Votrax SC-01", que era capaz de produzir voz a partir de material impresso. Dois anos depois, em 1978, surgiu o conversor de texto para voz "Type-n-Talk", da Votrax, baseado também no mesmo "chip". Entretanto, havia falta de inteligibilidade da voz gerada por esses dois sistemas.

Em 1982, a "Speech Plus Inc." lançou o "Prose-2000", sistema de conversão de texto para voz baseado no "MITalk". No mesmo ano, a "Street Electronics" lançou o "Echo", sistema de baixo custo baseado no "chip TMS-5220", da "Texas Instruments".

Em 1983, a "Digital Equipment Corporation - DEC" lançou o "DECtalk", que teve origem no conversor de texto para voz "Klattalk", também desenvolvido no MIT por Dennis Klatt.

Ainda em 1983, a "Infovox" colocou no mercado o sistema "SA 201/PC", capaz de sintetizar voz a partir de textos em Inglês, Francês, Espanhol, Alemão, Italiano, Suécio e Norueguês, desenvolvido a partir das pesquisas de Rolf Carlson no "Royal Institute of Technology of Stockholm".

Na segunda metade da década de 80, a "Berkeley Speech Technologies" apresentou o sistema "Text-to-Speech - T-T-S", originário das pesquisas de O'MALLEY (1990) na Universidade da Califórnia. Ainda nessa época, a AT&T lançou o sistema "Conversant", capaz de sintetizar voz a partir de textos em Inglês, Francês e Espanhol.\*

Desde então, esses sistemas vêm sofrendo constantes atualizações com a finalidade de aumentar a inteligibilidade e a naturalidade da voz produzida, e também, suportar novos idiomas. Por exemplo, um ano após o lançamento do "Prose 2000", já havia uma implementação para aceitar textos em espanhol (OLABE et al., 1983), e em 1989, o "MITalk" produzia saída de voz a partir de textos em japonês e chinês (JAVKIN et al., 1989).

Diversas outras pesquisas vêm sendo realizadas em âmbito mundial, como o sistema de síntese de voz a partir de textos em Chinês desenvolvido por LEE, TSENG e OUH-YOUNG (1989) e o sistema para textos em Árabe produzido por EL-IMAN (1989).

Além desses estudos, outros sistemas de síntese de voz de baixo custo têm surgido no mercado, como as placas para microcomputador padrão PC-AT "Sound Blaster PRO" da "Creative Labs" e a "Mwave LS2000" da "IBM", acompanhadas do programa "Monolog" da "Creative Labs", capaz de fazer a conversão de textos em Inglês para voz, com qualidade bastante aceitável.

---

\* AT & T do Brasil, São Paulo. **Comunicação Pessoal**, 1994.

No Brasil, estudos de síntese voz tiveram início na Escola Politécnica da Universidade de São Paulo - EPUSP, através das pesquisas de CAMPOS (1980), sobre um sintetizador de voz para o Idioma Português, capaz de aceitar entradas na forma fonética do Português.

ESQUIVEL, em 1984, apresentou um sistema de síntese de voz em tempo real a partir de texto, no qual sinais adicionais eram acrescentados ao texto para a correta pronúncia de determinados sons.

Na Universidade de Campinas - UNICAMP, estudos sobre síntese de voz a partir de texto irrestrito foram realizados posteriormente (EGASHIRA, 1992). Foram feitos trabalhos sobre pré-processamento de texto com a finalidade de permitir a correta elocução de números, abreviaturas e caracteres não alfabéticos.

Além dessa pesquisa, no Instituto de Estudos da Linguagem da UNICAMP está sendo desenvolvido um mini léxico com cerca de 25.000 palavras, com informações sobre transcrição fonética, separação de sílabas e acentuação das palavras, para que possa ser utilizado em investigações e testes de regras de conversão de letras para fonemas (VIOLARO, 1993).

Mais recentemente, vem sendo comercializado um produto de baixo custo para auxílio à deficientes visuais, o "Dosvox", desenvolvido na Universidade Federal do Rio de Janeiro. Esse produto é formado por um conjunto de programas, tais como editor de texto e calculadora, e é capaz de sintetizar voz a partir de texto utilizando um conjunto de regras de conversão de texto para fonemas \*.

---

\* BORGES, J.A. (Universidade Federal do Rio de Janeiro, Rio de Janeiro). **Comunicação Pessoal**, 1994.

## **CAPÍTULO II**

### **CONSIDERAÇÕES LINGUÍSTICAS**

#### **2.1 - CARACTERÍSTICAS DA VOZ HUMANA**

##### **2.1.1 - Principais Aspectos**

A fala humana distingue-se de outros sistemas simbólicos, como os gestos por exemplo, por ser segmentável em unidades menores que se apresentam em número finito para cada idioma e possibilitam recombinação de modo a expressar idéias diferentes. O contínuo sonoro pode ser dividido em segmentos linearmente dispostos cuja presença ou ausência, assim como sua ordem, tem função distintiva, capaz de ocasionar alterações no significado de uma palavra. Essas unidades componentes do

contínuo sonoro são produzidas por um mecanismo fisiológico denominado aparelho fonador (CALLOU e LEITE, 1990).

A compreensão do funcionamento do aparelho fonador é importante para entender os parâmetros envolvidos na produção da voz, e por esse motivo ainda hoje é um tópico de ativas pesquisas na área de fonética acústica e articulatória (KLATT e KLATT, 1987).

O aparelho fonador humano é constituído pelas seguintes partes, indicadas na Figura 2.1:

- os pulmões, os brônquios e a traquéia, que são os órgãos respiratórios responsáveis pelo fornecimento da corrente de ar, que corresponde à "matéria-prima" da fonação;

- a laringe, na qual se localizam as cordas vocais, que produzem a energia sonora utilizada na fala,

- e as cavidades supralaríngeas (faringe, boca e fossas nasais), que funcionam como uma caixa de ressonância. A cavidade bucal pode variar profundamente de forma e volume, graças aos movimentos dos órgãos ativos, sobretudo da língua. Através da movimentação do palato mole (vélum), a cavidade nasal pode ser acoplada à cavidade bucal.

Estas duas últimas partes, a laringe e as cavidades supralaríngeas, são também conhecidas como trato vocálico.

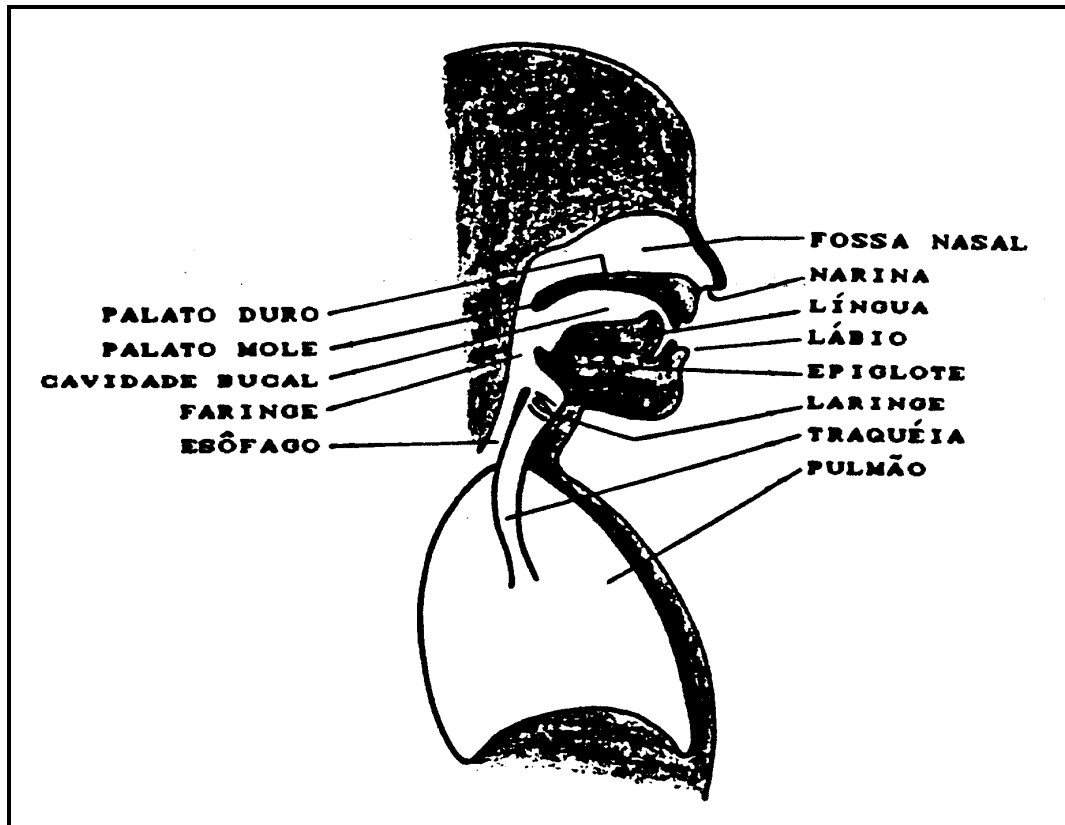


FIGURA 2.1 - Diagrama esquemático do aparelho fonador humano.

O trato vocálico pode ser considerado como um tubo acústico de seção variável, com início nas cordas vocais e que termina nos lábios e narinas, conforme ilustra o esquema da Figura 2.2. Em um adulto do sexo masculino apresenta aproximadamente 17 cm de comprimento, sendo a área seccional determinada pela posição dos lábios, maxilares, língua e vélum, e pode variar de zero (no caso de lábios fechados) até aproximadamente 20 cm<sup>2</sup>. A cavidade nasal tem em média 12 cm de comprimento e volume aproximado de 60 cm<sup>3</sup> (FLANAGAN, 1972).

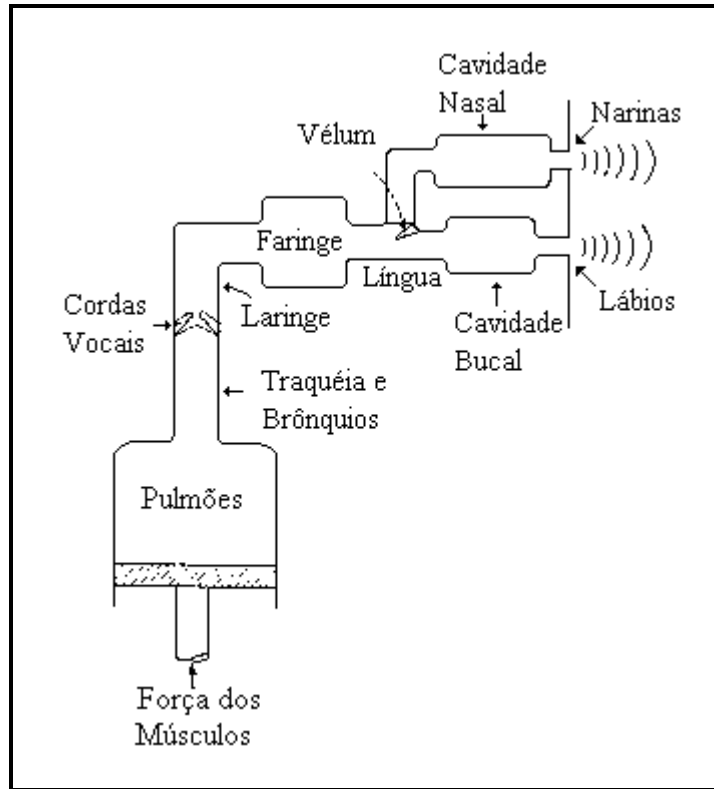


FIGURA 2.2 - Diagrama esquemático dos componentes funcionais do trato vocálico. (FLANAGAN, 1972)

Um órgão essencial na fonação é a laringe, que corresponde a um tubo de paredes cartilagosas semi-rígidas, contendo dois pares sobrepostos de membranas, denominadas cordas vocais, que delimitam uma fenda chamada glote. Quando pretende-se emitir um som, utilizando-se as cordas vocais, a glote é fechada, e sob a ação de um esforço expiratório, o ar afasta ligeiramente as bordas das cordas vocais e escoar pela glote. Simultaneamente, as cordas vocais começam a vibrar, permitindo a passagem de pulsos de ar, que excitam o sistema acústico localizado imediatamente acima das cordas vocais (SANCHES, 1989).

### 2.1.2 - Sons da Fala

A voz, produzida pela passagem do ar fornecido pelos pulmões no trato vocálico, pode ser gerada de três maneiras distintas originando *sons sonoros* ou *vocálicos*, *sons fricativos* e *sons plosivos*. O modo como esses sons são produzidos foi descrito detalhadamente por vários autores como FLANAGAN (1972), CAMPOS (1980), SANCHES (1989), CASAES (1990) e O'MALLEY (1990).

Os *sons sonoros* ou *vocálicos* são produzidos pela elevação da pressão de ar nos pulmões, forçando a sua passagem através do orifício das cordas vocais (glote) e causando sua vibração. Essa vibração obstrue a passagem de ar de maneira periódica, causando a interrupção do fluxo de ar, que excita o trato vocálico. O período dessa interrupção é chamado de "pitch" e seu inverso é a "frequência fundamental ( $f_0$ )".

Os *sons fricativos* são gerados pela formação de uma constrição em algum ponto do trato vocálico, normalmente nos lábios, forçando a passagem de ar através dessa constrição com velocidade suficiente para produzir turbulência, criando assim, uma fonte de "ruído branco". Podem ser produzidos com ou sem vibração das cordas vocais, condição em que serão chamados respectivamente de *fricativos sonoros* ou *fricativos surdos*.

Os *sons plosivos* resultam da constrição completa do trato vocálico em alguma parte, com acumulação de pressão e liberação abrupta em seguida. O ponto de completo fechamento pode ser efetuado em várias zonas de articulação e a excitação pode ou não causar vibração das cordas vocais, como no caso dos sons fricativos.

Segundo CAMPOS (1980) à medida que os sons, gerados por qualquer uma das formas acima descritas, propagam-se pelo trato vocálico, apresentam alteração em seu espectro de frequências e com ressonância em determinadas frequências. Estas frequências são denominadas *frequências formantes* do som, ou simplesmente *formantes*, sendo o número de formantes variável conforme o som.

Ainda de acordo com Campos, um som pode ser caracterizado pelas suas três frequências formantes mais baixas, que são comumente designadas por  $F_1$ ,  $F_2$  e  $F_3$ . As frequências formantes dependem da forma do trato vocálico e conseqüentemente as propriedades espectrais do som produzido variam em decorrência da geometria do trato vocálico.

Juntamente com a frequência fundamental, as formantes constituem os principais parâmetros acústicos da voz. Tipicamente, para uma voz masculina a frequência fundamental varia entre 60 e 240 Hz, enquanto que as três formantes variam em torno de 500 Hz, 1500 Hz e 2500 Hz. Para uma voz feminina, a frequência fundamental tem valores entre 100 e 400 Hz, enquanto que as formantes estão aproximadamente 10% acima das formantes masculinas (O'MALLEY, 1991).

A estrutura das formantes é comumente representada através de espectrogramas sonoros, conforme exemplifica a Figura 2.3, na qual está indicado o espectrograma sonoro da frase "Noon is the sleepy time of the day", obtido por FLANAGAN et al. (1970), com suas três frequências formantes representadas por linhas tracejadas. O eixo das abscissas corresponde ao tempo de elocução da frase, enquanto que o eixo das ordenadas corresponde às frequências, sendo que os padrões escuros ocorrem nas frequências com intensidade sonora.

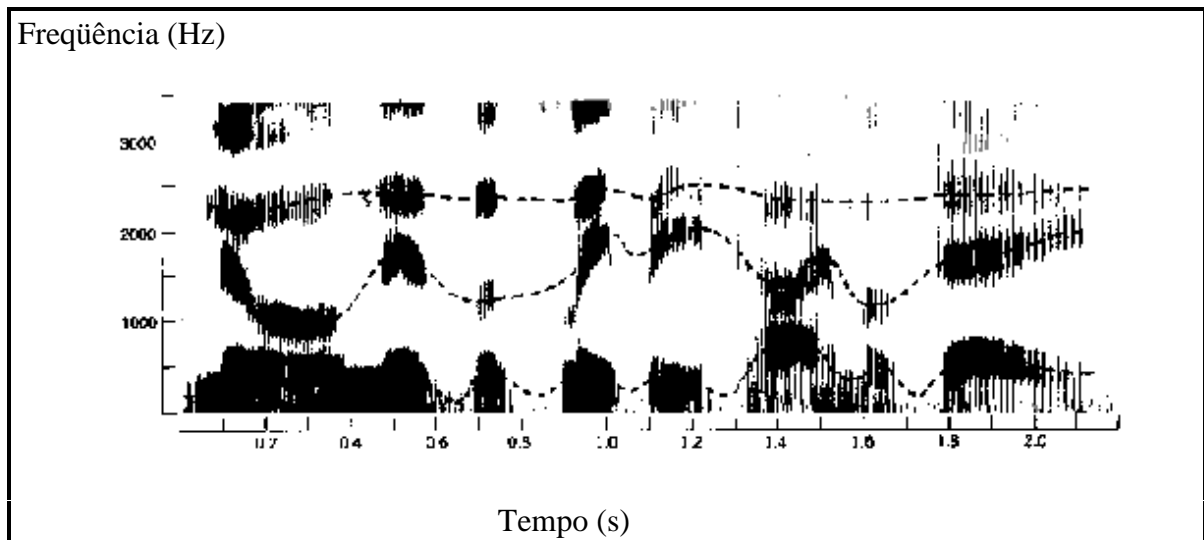


FIGURA 2.3 - Espectrograma da frase "Noon is the sleepy time of the day", com indicação das três primeiras formantes. (FLANAGAN et al., 1970)

## 2.2 - CONCEITUAÇÃO DE FONEMA E RELAÇÃO LETRA-FONEMA

Os *fonemas* são as unidades básicas de uma Língua, e têm a propriedade de mudar o sentido de uma palavra quando uma unidade é substituída por outra (FLANAGAN, 1972). Por exemplo, na série de palavras *dia, fia, mia, pia, tia* e *via*, a distinção entre as palavras ocorre apenas pelo elemento consonântico inicial, que caracterizam unidades sonoras distintas, correspondendo cada uma delas a um fonema diferente.

Entendidos como uma unidade de som no início do século XIX, os fonemas são hoje considerados como unidades mentais, abstratas, das quais o som é a sua realização física. O fonema é uma unidade da Língua e os sons ou fones são unidades da fala (CALLOU e LEITE, 1990).

Os fonemas são comuns a todos os indivíduos que falam a mesma Língua, enquanto que os sons que o representam variam não apenas de um indivíduo para outro, como também, para um mesmo indivíduo de um ato para outro (PAIS, 1986). Aos vários sons que realizam o mesmo fonema dá-se a denominação *variantes* ou *alofones*. Por exemplo, os fonemas /d/ e /t/ apresentam em determinados dialetos do Português uma realização palatal diante do /i/, como nas palavras *tia* e *dia* e uma realização alveolar ou dental diante das outras vogais como nas palavras *dado*, *docas*, *tela*, *tua* (CALLOU e LEITE, 1990).

Para distinguí-los dos sons realmente produzidos, os fonemas são normalmente representados entre barras oblíquas (/ /), enquanto que os sons são representados entre colchetes ([ ]). No caso da representação entre barras, a transcrição é dita fonológica e no caso da representação entre colchetes, a transcrição é fonética. A palavra *dia* por exemplo, é representada pelos fonemas /dia/ e pode ser pronunciada como [djia] (EGASHIRA, 1992).

Cada idioma tem seus próprios fonemas, que são elementos fônicos dotados de função representativa no sistema. De acordo com CALLOU e LEITE (1990), a Língua Portuguesa tem 26 fonemas segmentais (19 consoantes e 7 vogais) e um fonema supra-segmental, o *acento*, que não é um segmento e sim uma qualidade que se superpõe a certos segmentos. Formas como *dívida* e *divida*; *sábia*, *sabia* e *sabiá* opõem-se entre si apenas pela posição do acento tônico.

Para que as sequências fônicas de uma Língua sejam reproduzidas na escrita, utilizam-se sinais gráficos representativos desses sons, que são as *letras* ou *grafemas*. No entanto, não há uma correspondência exata entre número de letras e o número de fonemas nos idiomas. Alguns exemplos foram fornecidos por CEGALLA (1977):

- na Língua Portuguesa pode-se observar que uma mesma letra pode representar mais de um fonema, como por exemplo na seqüência de palavras *exame, xale* e *próximo*;

- um mesmo fonema pode ser figurado por mais de uma letra, como nas palavras *casa, exílio, cozinha* ou representado por um grupo de duas letras, os dígrafos, como na palavras *machado, mulher, unha, missa* e *carro*;

- há ainda letras que por vezes não representam fonemas, funcionando somente como notações léxicas, como nas palavras *campo [cãpo]* e *regue*, na qual o **u** é insonoro, para não seja proferido *reje*;

- e também são utilizadas letras simplesmente decorativas, na medida em que não representam fonemas e não funcionam como notações léxicas, como em *discípulo [dicipulo]*, *hotel [otél]* e *exceção [esesão]*; além de fonemas que, em certos casos, não são representados graficamente como em *eram [érãu]*, *falam [fálãu]*.

Há um sistema ortográfico que rege essa representação na linguagem escrita, sendo a ortografia vigente até hoje no Brasil, a oficialmente adotada nas normas do Vocabulário Ortográfico de 1943, com as alterações determinadas pela Lei nº. 5.765 de 18 de dezembro de 1971 (FERREIRA, 1986).

Segundo CALLOU e LEITE (1990), recentemente tem-se discutido a possibilidade de uma reforma ortográfica que leve em consideração as relações entre a pronúncia e a ortografia portuguesa do Brasil e de Portugal e que também procure aproximar o sistema de fonemas ao sistema de letras, como a substituição da letra "s" por "z" em palavras nas quais a letra "s" representa o som [z] (*casa, mesa*) e de "ss", "c", "ç" e "x" por "s" para representarem o som [s] (*posso, cedo, laço, próximo*).

No entanto, ainda segundo esses autores, esse sistema integrado letra-fonema parece ser inviável, pois em um País com a dimensão do Brasil qualquer tentativa de aproximação seria precária e deixaria a desejar, já que teriam de ser levados em consideração todas as diferenças regionais, sócio-culturais e até mesmo individuais.

Citam também que, quanto mais um idioma desenvolve-se, mais o sistema ortográfico afasta-se do sistema fonológico, como ocorre com os idiomas Inglês e Francês. Ainda com relação à simplificação abordada anteriormente, a representação do som [s] sempre pela letra "s" e do som [z] sempre pela letra "z" esbarra na questão das palavras homófonas como *coser/cozer*, *expiar/espiar*, *cessão/sessão/seção*, além de palavras como *aterrisar* e *subsídios*, para as quais existem normalmente duas pronúncias, *aterri[s]ar* e *aterri[z]ar*, *sub[s]ídios* e *sub[z]ídios*.

Assim, considerando-se todos esses argumentos, a convivência com o sistema ortográfico atual parece inevitável, pelo menos a curto e médio prazo.

### **2.3 - CARACTERÍSTICAS DO ALFABETO FONÉTICO**

Para simbolizar na escrita a pronúncia real de um som utiliza-se um alfabeto especial, conhecido como *alfabeto fonético*. A finalidade da transcrição fonética e portanto, do alfabeto fonético é justamente a transcrição e a leitura de um som em qualquer Idioma por uma pessoa treinada. Assim, esse alfabeto deve apresentar convenções inequívocas e de maneira explícita. Algumas dessas convenções tornaram-se bastante difundidas, como por exemplo, as propostas no "International Phonetic Alphabet - IPA" pela Sociedade Internacional de Fonética. Esse alfabeto,

no entanto, emprega caracteres pouco comuns em máquinas de escrever e computadores, o que dificulta sua utilização (CALLOU e LEITE, 1990).

A Tabela 2.1 a seguir apresenta o alfabeto fonético baseado nos símbolos IPA, e outros dois possíveis alfabetos para a Língua Portuguesa, sendo um deles baseado em letras maiúsculas, utilizando até dois caracteres e outro, que será adotado neste trabalho, utilizando apenas um único caracter.

TABELA 2.1 - Alfabetos fonéticos para a Língua Portuguesa.

Símbolos IPA  (CUNHA e CINTRA, 1985)	Símbolos com 1 ou 2 caractéres  (CAMPOS,1980)	Símbolos com 1 caracter	Exemplos
a	A	a	<b>pá</b> , gato
e	E	e	<b>vê</b> , medo
ɛ	EH	é	<b>pé</b> , ferro
i	I	i	<b>vir</b> , bico
o	O	o	<b>avô</b> , morro
ɔ	OH	ó	<b>avó</b> , cola
u	U	u	<b>tu</b> , bambu
ã	AN	ã	<b>lã</b> , cama
m	M	m	<b>mar</b> , amigo
n	N	n	<b>nada</b> , cano
ɲ	NH	ñ	<b>vinha</b> , caminho
b	B	b	<b>bravo</b> , ambos
p	P	p	<b>pai</b> , caprino
d	D	d	<b>dar</b> , andar
t	T	t	<b>tu</b> , canto
g	G	g	frango, agrado
k	C	k	<b>casa</b> , <b>que</b>
f	F	f	<b>filho</b> , afiar
v	V	v	<b>vinho</b> , uva
s	S	s	<b>saber</b> , posso
z	Z	z	<b>bazar</b> , casa
ʃ	X	x	<b>chover</b> , xarope
ʒ	J	j	<b>já</b> , jarra
l	L	l	<b>lado</b> , veludo
l	L		<b>alto</b> , fuzil
λ	LH	L	<b>filho</b> , pilha
r	R	r	<b>caro</b> , cores
r	R	h	<b>mar</b> , carta
R	RR	R	<b>carro</b> , roda

Na Língua Portuguesa os fonemas /i/ e /u/, quando formam sílaba com outra vogal, são chamados *semi-vogais* e normalmente transcritos como [j] e [w], como em [rej] e [mew] (CUNHA e CINTRA, 1985).

No entanto, CAMPOS (1980) mostrou que um ditongo pode ser considerado como junção de duas vogais de menor duração com transições suaves entre as suas frequências formantes. Portanto, ao longo desse trabalho não será feita distinção entre o /i/ e o /u/ vogais ou semi-vogais.

Em posição final de sílaba ou palavra, a consoante "l" pode ser pronunciada como [ɫ] ou [w], como em *alto* e *Brasil* (CUNHA e CINTRA, 1985). Por essa razão, quando encontrado nessas posições o "l" será associado ao símbolo [ɫ]. E nessa mesma situação a letra "r" será associada ao símbolo [h], conforme indicado anteriormente na Tabela 2.1.

E é comum utilizar um apóstrofe (') para indicar a sílaba ou vogal tônica na transcrição, com em [ 'bo-la] ou [b'ola].

## **2.4 - CLASSIFICAÇÃO DOS SONS LINGÜÍSTICOS**

A classificação dos sons da Língua Portuguesa foi discutida em detalhes por diversos autores, como CUNHA e CINTRA (1985), PAIS (1986), CALLOU e LEITE (1990) e CASAES (1990).

Apesar de algumas divergências, é de consenso entre os autores a existência de duas classes de sons, as vogais e as consoantes, as quais serão destacadas a seguir.

### **2.4.1 - Classificação das Vogais**

As vogais são normalmente classificadas segundo quatro critérios: quanto à região de articulação, quanto ao timbre, quanto ao papel das cavidades bucal e nasal e quanto à intensidade. Os três primeiros critérios são fundamentalmente de base articulatória e o último de base acústica.

#### **- Classificação quanto à Região de Articulação**

Diz respeito ao ponto ou parte em que se dá o contato ou aproximação dos órgãos que cooperam para a produção dos fonemas, no caso das vogais, a língua e o palato. Produz-se a *vogal média* [a] mantendo-se a língua baixa, quase em posição de descanso, e a boca entreaberta.

Para passar da vogal *a* para as *vogais anteriores* ([e], [é], [i]) levanta-se gradualmente a parte anterior da língua em direção ao palato duro, ao mesmo tempo em que diminui-se a abertura da boca. Para emitir as *vogais posteriores* ([o], [ó], [u]), eleva-se a parte posterior da língua em direção ao véu palatino, arredondando progressivamente os lábios.

#### **- Classificação quanto ao Timbre**

Refere-se ao maior ou menor grau de abertura dos lábios. Essa abertura é máxima para a vogal [a] e mínima para as vogais [i] e [u].

#### **- Classificação quanto ao Papel das Cavidades Bucal e Nasal**

Depende da posição da úvula durante a passagem de ar pelo trato vocálico.

Quando a corrente sonora é impedida de ressoar na cavidade nasal devido à posição levantada da úvula, tem-se a produção das *vogais orais* ([a], [e], [é], [i], [o], [ó], [u]).

Quando as fossas nasais são acopladas à cavidade bucal através do abaixamento da úvula, parte da corrente sonora ressoa na cavidade nasal, produzindo as *vogais nasais* ([ã], [ẽ], [ĩ], [õ], [ũ]).

### - Classificação quanto à Intensidade

É uma qualidade física da vogal que depende da força expiratória e da amplitude da vibração das cordas vocais.

As vogais que se encontram nas sílabas pronunciadas com maior intensidade chamam-se *tônicas* e caracterizam-se no idioma Português por um reforço da energia expiratória.

As vogais que se encontram em sílabas não acentuadas denominam-se *átonas*. No idioma Português normal do Brasil, as vogais [é] e [ó] não aparecem em posição átona, assim como as vogais nasais.

A Tabela 2.2 resume a classificação das vogais da Língua Portuguesa segundo os critérios acima descritos.

TABELA 2.2 - Classificação das vogais.

Intensidade	Papel das Cavidades Bucal e Nasal	Timbre	Região de Articulação		
			Anteriores ou Palatais	Médias ou Centrais	Posteriores ou Velares
Tônicas	Orais	Fechadas	[i]		[u]
		Semi-fechadas	[e]		[o]

		Semi-abertas	[é]		[ó]
		Abertas		[a]	
	Nasais	Fechadas	[ĩ]		[ũ]
		Semi-fechadas	[ẽ]	[ã]	[õ]
Átonas	Orais	Fechadas	[i]		[u]
		Semi-fechadas	[e]		[o]
		Abertas		[a]	

CAMPOS (1980) mostrou que do ponto de vista da fonética acústica não há razão para considerar [ẽ], [ĩ], [õ], e [ũ] fonemas distintos, pois seus espectros apresentam uma parte inicial idêntica ao das vogais que o iniciam, seguidos de uma parte muito semelhante a todos eles, que caracteriza o [m] final desses sons. Por isso, podem ser tratados como o encontro de dois fonemas, com transições suaves entre duas configurações do trato vocálico.

No Brasil, nas sílabas átonas ocorre a chamada "neutralização", na qual as vogais anteriores "e" e "i", quando em posição final absoluta, são reduzidas a uma única vogal [i], como na palavra *tarde* → [tardi] e as vogais posteriores "o" e "u", quando nessa situação também são reduzidas a uma única vogal [u], como no caso da palavra *povo* → [povu].

## 2.4.2 - Classificação das Consoantes

São dezenove as consoantes da Língua Portuguesa e tradicionalmente classificadas em função de quatro critérios de base articatória, ou seja, quanto ao modo de articulação, quanto ao ponto de articulação, quanto à função das cordas vocais e quanto ao papel das cavidades bucal e nasal.

### - Classificação quanto ao Modo de Articulação

Refere-se à maneira pela qual os fonemas consonantais são articulados. Vindo da laringe, a corrente de ar chega à boca, onde encontra obstáculos totais ou parciais da parte dos órgãos bucais. Se o fechamento dos lábios ou a interrupção da corrente de ar é total, tem-se as *consoantes oclusivas* ([p], [t], [k], [b], [d], [g]); se o fechamento for parcial, produz-se as *consoantes constritivas*.

No segundo caso, dependendo de como a corrente expiratória escapa, as consoantes podem ser:

- *fricativas*: são produzidas quando o trato vocálico é excitado por um fluxo de ar turbulento, que se forma quando a corrente expiratória passa pela constrição ([f], [s], [x], [v], [z], [j]).

- *vibrantes*: são caracterizadas pelo movimento vibratório rápido da língua ([r]) ou da úvula ([R]), que provocam breves interrupções da passagem da corrente expiratória.

- *laterais*: caracterizam-se pela passagem da corrente expiratória pelos dois lados da cavidade bucal, em virtude de um obstáculo formado no centro desta pelo contato da língua com os alvéolos dos dentes ([l]) ou com o pálato ([L]).

### **- Classificação quanto ao Ponto de Articulação**

Diz respeito ao lugar onde os órgãos fonadores entram em contato para a emissão do som, podendo ser bilabiais ([p], [b], [m]), labiodentais ([f], [v]), lingüodentais ([t], [d], [s], [z]), alveolares ([l], [r], [n]), palatais ([x], [j], [L], [ñ]) ou velares ([k], [g], [R]).

### **- Classificação quanto à Função das Cordas Vocais**

Se durante a produção das consoantes a corrente de ar produzir vibração das cordas vocais tem-se uma *consoante sonora*; caso contrário, a consoante será *surda*.

### **- Classificação quanto ao Papel das Cavidades Bucal e Nasal**

Quando o ar sai exclusivamente pela boca, as consoantes são ditas *orais*. Quando o ar penetra nas fossas nasais pelo abaixamento da úvula, as consoantes são ditas *nasais* ([m], [n], [ñ]).

A Tabela 2.3 resume a classificação das consoantes para a Língua Portuguesa com base nos critérios acima descritos.

TABELA 2.3 - Classificação das consoantes.

Função das cavidades bucal e nasal		Orais						Nasais
Modo de articulação		Oclusivas		Constritivas				Oclusivas
				Fricativas		Vibrantes	Laterais	
Função das cordas vocais		Surdas	Sonoras	Surdas	Sonoras	Surdas	Sonoras	Sonoras
Ponto de Articulação	Bilabiais	[p]	[b]					[m]
	Labiodentais			[f]	[v]			
	Lingüodontais	[t]	[d]	[s]	[z]			
	Alveolares					[l]	[r]	[n]
	Palatais			[x]	[j]	[L]		[ñ]
	Velares	[k]	[g]				[R]	

## **CAPÍTULO III**

### **CONVERSÃO DE TEXTO PARA FONEMAS**

#### **3.1 - ASPECTOS CONCEITUAIS SOBRE O PROCESSO DE SÍNTESE DE VOZ A PARTIR DE TEXTO**

O processo de síntese de voz a partir de texto inicia-se com o processamento lingüístico para a determinação da estrutura fonológica das sentenças dos textos de entrada. Essa estrutura, constituída essencialmente pelos fonemas que formam as palavras, é analisada para a determinação dos parâmetros acústicos que serão utilizados posteriormente no controle do sintetizador de voz, o qual produzirá a fala.

Este processo constitui uma tarefa ampla e complexa, e foi dividido em várias etapas por CROCHIERE e FLANAGAN (1990), conforme indica a Figura 3.1.

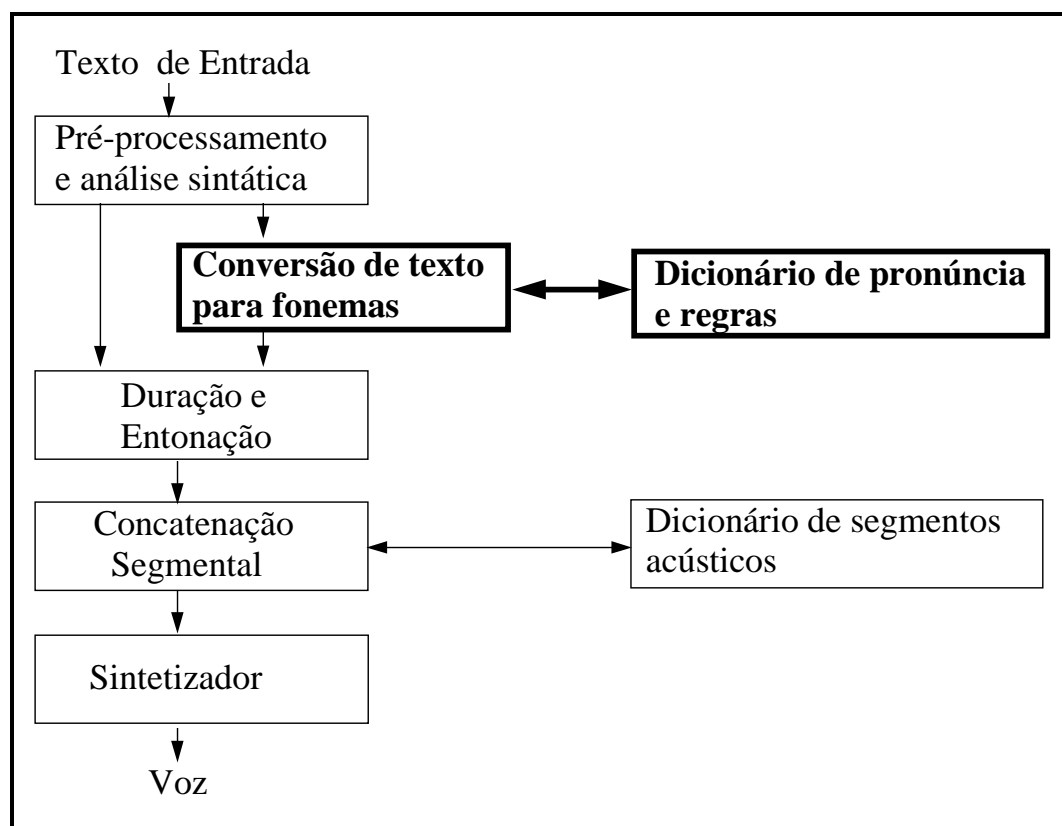


FIGURA 3.1 -Elementos de um sistema de síntese de voz a partir de texto. (CROCHIERE e FLANAGAN, 1986)

Inicialmente, o texto de entrada contendo abreviações, acrônimos, datas, caracteres não alfabéticos e sinais de pontuação, deve sofrer um pré-processamento para que seja utilizado nas fases seguintes.

HIRSCHBERG et al. (1990) comentam à respeito das dificuldades relativas a esse processo, na medida em que, podem ocorrer muitas ambigüidades, como por

exemplo, um "ponto" pode ser usado em uma abreviação ou no final de uma sentença.

Além disso, números constituem um séria dificuldade; 1/2 por exemplo, pode ser uma data ou uma fração e também, muitas abreviaturas comuns podem ter múltiplos sentidos, como nas seguintes sentenças, "Dr. Jones lives on Jones Dr. (Doctor Jones lives on Jones Drive); St. James St. (Saint James Street)", citadas por O'MALLEY (1990).

Uma análise sintática simples é necessária para a pronúncia de determinadas palavras, como "object" e "desert", as quais em Inglês são acentuadas na primeira sílaba se forem um substantivo ou na segunda sílaba, se forem um verbo.

A análise sintática é importante também para determinar a correta entonação de uma sentença, como por exemplo, para fazer com que uma palavra seja pronunciada com maior destaque em relação à outra, conforme comentam HIRSCHBERG et al. (1990).

Citam ainda que, na maioria dos sistemas as palavras são divididas em *palavras de função*, como as preposições e os artigos, e *palavras de conteúdo*, como substantivos e verbos. Normalmente, as palavras de conteúdo são pronunciadas com maior destaque, como na frase "Cat likes to eat mice.", cujas palavras grifadas são pronunciadas com maior ênfase em relação à palavra "to".

A etapa seguinte à análise sintática consiste em determinar os fonemas correspondentes ao texto de entrada, tarefa executada normalmente através de um dicionário e de um conjunto de regras de conversão de letras para fonemas (ATAL e RABINER, 1986).

A conversão de uma palavra para a sua correspondente forma fonética inicia-se com a sua busca no dicionário, o qual contém informações fonéticas sobre

algumas palavras. Caso não seja encontrada, a palavra deve ser submetida à aplicação de um conjunto de regras para a obtenção dos fonemas a partir de suas letras.

VIEIRA et al. (1991) realizaram essa etapa da transcrição fonética através de dois conjuntos de procedimentos independentes porém capazes de operarem sobre os mesmos níveis de estruturas de dados. A cadeia segmental foi percorrida da esquerda para a direita e a transcrição fonética realizada palavra por palavra, ao nível da forma de citação. Seguindo o Alfabeto SAM\_PA, a frase "*Ele disse à Antónia que vinha amanhã*" foi inicialmente transcrita como "*el@ d''is@ a 6~t''Onj6 k@ v''inh6 am6j''6~*" e depois foram tratados os encontros vocálicos e consonantais em fronteira de palavras obtendo-se *el~dis a~tOnj6 k viJ am6J6~*.

Cabe ressaltar nesse exemplo, algumas diferenças quanto à pronúncia do Português falado em Portugal e no Brasil, como no caso da palavra *ele*, que em Portugal pronuncia-se [el] e no Brasil [eli] ou na palavra *que*, a qual é pronunciada como [k] em Portugal e [kê] no Brasil, o que evidencia mais uma vez a necessidade de regras fonéticas e fonológicas próprias para cada país.

As informações fonéticas e sintáticas são utilizadas posteriormente para determinação da prosódia da mensagem, caracterizada por LAPORTE (1989) através das características de ritmo, acentuação, entonação e expressas como resultado combinado de parâmetros de duração, intensidade e "pitch". Cada sílaba em uma sentença tem uma duração específica, geralmente diferente das sílabas vizinhas, que determinam o ritmo da fala. A intensidade caracteriza o volume do som da fala, que em uma sentença varia entre valores baixos e altos. O mesmo acontece com os valores de "pitch", os quais definem a entonação de uma sentença.

A partir dessas características, são determinados os parâmetros acústicos da sentença a ser pronunciada, eventualmente com base em um dicionário de segmentos acústicos, e finalmente a voz é produzida por um sintetizador.

Este trabalho concentra-se nas etapas do processo de síntese de voz a partir de texto indicadas em negrito na Figura 3.1 apresentada anteriormente. Alguns aspectos relevantes sobre esse processo serão destacados a seguir.

## **3.2 CONVERSÃO DE TEXTO PARA FONEMAS AO NÍVEL DE PALAVRAS**

### **3.2.1 - Conversão de Letra para Fonema**

Várias tentativas iniciais foram feitas com a finalidade de prever a pronúncia de palavras a partir das letras que as compunham, com base na hipótese de que uma letra ou par de letras poderia ser convertido para o fonema apropriado caso fosse examinado o contexto, ou vizinhança, na qual a palavra estava inserida. Para cada letra, as regras deveriam ser ordenadas de modo que as primeiras tratariam dos casos mais complexos, e o último caso corresponderia à tradução fonética "default" (KLATT, 1987).

Partindo deste princípio, AINSWORTH (1973) apresentou um algoritmo para conversão de letras para fonemas para o Inglês Britânico a partir de um conjunto de 159 regras, que estão apresentadas na Tabela 3.1.

**TABELA 3.1 - Regras de conversão de letras para fonemas para o Inglês Britânico com alfabeto IPA. (AINSWORTH, 1973)**

Letras	Fonemas	Letras	Fonemas	Letras	Fonemas	Letras	Fonemas
-(a)-	/ə/	(ough)	/ə u/	(g)et	/g/	-C(ie)	/ɑi/
-(are)	/ɑ/	(oul)d	/u/	c(ow)	/ɑ u/	VC(ie)	/i/
(a)E	/ei/	(ou)	/ɑ u/	h(ow)	/ɑ u/	(i)	/I/
(ar)	/ɑ/	(oor)	/ɔ/	n(ow)	/ɑ u/	(j)	/dz/
(a)sk	/ɑ/	(oo)k	/u/	v(ow)	/ɑ u/	-(k)n	//
(a)st	/ɑ/	f(oo)d	/u/	r(ow)	/ɑ u/	(k)	/k/
(a)th	/ɑ/	(oo)d	/u/	(ow)	/ə u/	(le)-	/əl/
(a)ft	/ɑ/	f(oo)t	/u/	g(o)-	/ə u/	(l)	/l/
(ai)	/ei/	s(oo)t	/u/	n(o)-	/ə u/	(m)	/m/
(ay)	/ei/	w(oo)	/u/	s(o)-	/ə u/	(n)g	/ŋ/
(aw)	/ɔ/	(oo)	/u/	(o)-	/u/	(n)	/n/
(au)	/ɔ/	sh(oe)	/u/	(o)	/o/	(or)	/ɔ/
(al)l	/ɔ/	(oe)	/ə u/	(ph)	/f/	(o)E	/ə u/
(a)ble	/ei/	VCd(e)d-	/ə/	(psy)	/sai/	(oa)	/ə u/
(a)ngSUF	/ei/	VCt(e)d-	/ə/	(p)	/p/	(their)	/ð e ə/
(a)	/æ/	VC(e)d-	//	(q)	/kw/	(th)r	/θ/
(b)	/b/	(e)r-	/ə/	(r)-	//	(th)	/ð/
(ch)	/tʃ/	wh(ere)	/e ə/	(rho)	/r ə u/	(t)ion	/ʃ/
(ck)	/k/	h(ere)	/i ə/	(r)	/r/	(t)	/t/
(c)y	/s/	w(ere)	/ɜ/	(sh)	/ʃ/	(u)pon	/ʌ/
(c)e	/s/	(ere)	/ir/	(ss)	/s/	(u)V	/u/
(c)i	/s/	(ee)	/i/	(sch)	/sk/	(u)C-	/ʌ/
(c)	/k/	(ear)	/ir/	Xv(s)	/z/	r(u)	/u/
(d)	/d/	(ea)	/i/	V(s)-	/z/	l(u)	/u/
VC(e)-	//	(e)ver	/e/	(s)	/s/	(u)	/ju/
th(e)-	/ə/	(eye)	/ɑi/	(there)	/ð e ə/	(v)	/v/
-C(e)-	/i/	(e)E	/i/	(g)e	/dz/	(w)r	//
-C(e)d-	/e/	c(ei)	/i/	(gh)	/g/	(wh)o	/h/
(o)ld	/ə u/	(ei)	/ɑi/	(g)	/g/	(wha)t	/wo/
(oy)	/ɔi/	(e)r	/ɜ/	w(h)	//	(wa)	/wo/
(o)ing	/ə u/	(eo)	/i/	(ha)v	/hæ/	(wo)r	/wɜ/
(oi)	/ɔi/	(ew)	/ju/	(h)	/h/	(w)	/w/
y(ou)	/u/	(e)u	//	-(i)-	/ɑi/	(x)	/ks/
(ou)s	/ʌ/	(e)	/ɜ/	(i)ty	/I/	-(y)	/j/
(ough)t	/ɔ/	(f)-	/v/	(i)E	/ɑi/	VC(y)	/I/
b(ough)	/ɑ u/	(f)	/f/	(ir)	/ɜ/	-C(y)	/ɑi/
t(ough)	/ʌf/	(g)e-	/dz/	(igh)	/ɑi/	(y)E	/ɑi/
c(ough)	/of/	(g)es-	/dz/	t(io)n	/ʌ/	(y)	/I/
-r(ough)	/ʌf/	(g)SUF	(g)	(i)nd	/ɑi/	(z)	/z/
r(ough)	/u/	(g)i	/dz/	(i)ld	/ɑi/		

Para esse algoritmo, nos casos onde a tradução de uma letra era ambígua e não houvesse um contexto que pudesse orientar a aplicação de determinada regra, o fonema mais comum era substituído. Algumas palavras foram consideradas exceções e incluídas em regras individuais.

Esse conjunto de regras foi testado em três textos com mil palavras cada um, sendo uma passagem de um livro sobre fonética, um trecho de um romance e um artigo sobre política, nos quais foram detectadas respectivamente 8%, 11% e 11% das palavras com erros de transcrição fonética, sendo a maioria dos erros devido à conversão das vogais.

Baseado neste sistema, ELOVITZ et al. (1976) implementaram um conjunto de regras para o Inglês Americano, que foi utilizado no "Naval Research Laboratory System - NRL". O sistema desenvolvido em três versões respectivamente com 182, 264 e 319 regras, foi testado em um conjunto de palavras mais freqüentes de um corpus da Língua Inglesa.

Os resultados mostraram que o desempenho do conjunto de regras era sensível ao nível do vocabulário ao qual era aplicado, sendo que a taxa de erros na transcrição fonética de palavras aumentava com o decréscimo de sua freqüência de utilização. Para o conjunto das mil primeiras palavras mais freqüentes, a última versão de regras apresentou 3,9% de palavras transcritas erroneamente, enquanto que para as cinco mil primeiras palavras de maior ocorrência esse índice aumentou para 7,2%. A esses dois conjuntos de palavras corresponderam respectivamente 1,4% e 2,4% de acertos na transcrição de fonemas, visto que em alguns casos o mesmo erro de transcrição repetiu-se em mais de uma palavra.

No entanto, os critérios usados para aceitação da transcrição de uma palavra nesses sistemas eram bastante subjetivos, levando-se em consideração que se

procurava obter uma transcrição "aceitável", sem especificação de onde o nível "aceitável" situava-se entre o "perfeito" e o "compreensível". Por exemplo, no "NRL System" a troca de uma vogal acentuada por uma não acentuada não foi considerada um erro, mas, um "problema de acentuação" (ELOVITZ et al., 1976).

Segundo KLATT (1987), o melhor algoritmo de conversão de letra para fonema desenvolvido na década de 70 foi o algoritmo de Hunnicut, utilizado no "MITalk" e no "DECtalk".

Esse algoritmo era bem mais complexo e executava em uma primeira fase a eliminação dos afixos (prefixo e sufixo) da palavra. Em seguida, fazia a conversão das consoantes e finalmente as vogais eram transcritas. Aproximadamente 15 prefixos e 50 sufixos eram detectados e posteriormente, aproximadamente 500 regras eram aplicadas.

Em testes realizados com palavras extraídas aleatoriamente de um dicionário, esse algoritmo atingia em média um índice de 65% de palavras transcritas corretamente .

ALLEN (1976) descreveu vários exemplos de transcrição fonética de palavras usando as regras de Hunnicut, com alfabeto IPA. Na palavra "table" por exemplo, não há afixos e as consoantes [t] e [b] são traduzidas diretamente. A letra [l] é seguida por um [e] final e é precedida por um [b]; nesse caso, ela é silábica, sendo indicada por /l/. Nesse ponto começa o processo de conversão de vogais.

A seqüência [ble] forma o contexto à direita de [a]. Todas as vogais, com exceção do [e], se estiverem na primeira sílaba de uma palavra, são longas quando seguidas por uma seqüência de *consoante não vocálica* + [l] + [e] em final de palavra. Essa regra é suficiente para converter [a] para /e/. Finalmente, a letra [e] por estar no final da palavra é muda, resultando a transcrição final /tebl/.

A palavra "subversion", por sua vez é submetida primeiro à remoção dos afixos *sub-* e *-ion*. Em seguida são convertidas as consoantes e ao mesmo tempo em que é feita a conversão das vogais, adicionam-se os afixos novamente, os quais são transcritos diretamente resultando em /səb- vɜrʒ -ən/.

Vários outros algoritmos foram então apresentados, com destaque para o de Berstein, citado por KLATT (1987), que examinava as palavras da direita para a esquerda, e atingia um índice de 85% de acertos com palavras aleatórias extraídas de um dicionário.

KLATT (1987) aponta como uma das principais razões para as falhas desses algoritmos o fato de que a conversão de uma vogal depende em parte da acentuação, como por exemplo, o "e" de "permit", que é realizado como ə: (*p'ə:rmite*, como em "her") ou como ə (*pərm'it*, como em "dinner").

No primeiro caso o acento está na primeira sílaba e a palavra é um substantivo significando *autorização, permissão*; no segundo caso, o acento está na segunda sílaba e a palavra é um verbo significando *permitir*. Esse mesmo autor comenta ainda que freqüentemente a análise correta de uma palavra requer a detecção dos limites dos seus morfemas, que são as menores unidades de uma Língua com significado. De fato, qualquer algoritmo de conversão baseado somente em regras terá problemas em palavras compostas, como o "th" em "hothouse" ou o "e mudo" em "houseboat".

### **3.2.2 - Dicionário de Exceções**

Levando-se em consideração que apenas a conversão de letras para fonemas não é suficiente para uma transcrição fonética absolutamente correta de todas as palavras de uma Língua, a alternativa é a utilização de um *dicionário de exceções*, que contenha palavras que falhem a essas regras. Por exemplo, ALLEN (1976) refere que a letra "f" sempre é pronunciada como /f/, exceto em *of* (/ɔ:v/).

Para KLATT (1987), a vantagem da elaboração de um *dicionário de exceções* advém do fato de que um pequeno número de palavras repetem-se inúmeras vezes em um texto aleatório. HIRSCHBERG et al. (1990) citam que um dicionário com apenas 150 palavras chega a cobrir 50% das palavras de um texto.

No entanto, a utilização de um dicionário contendo todas as palavras de uma Língua é inviável. O gráfico da Figura 3.2, construído a partir de um corpus da Língua Inglesa com mais de um milhão de palavras, contendo aproximadamente 50.000 palavras diferentes, mostra a frequência acumulada de palavras no corpus em função do número de diferentes palavras, ordenadas pela sua frequência no corpus (KLATT, 1987).

Pode-se observar que um dicionário de duas mil palavras abrange mais de 70% das palavras dos textos, enquanto que com dez mil palavras atinge 90%. Entretanto, uma extrapolação aproximada do número de palavras necessárias para aumentar essa abrangência de 90% para 93% em um texto aleatório, conforme indica a linha tracejada na Figura 3.2 mostra que seriam necessárias aproximadamente 40.000 palavras adicionais.

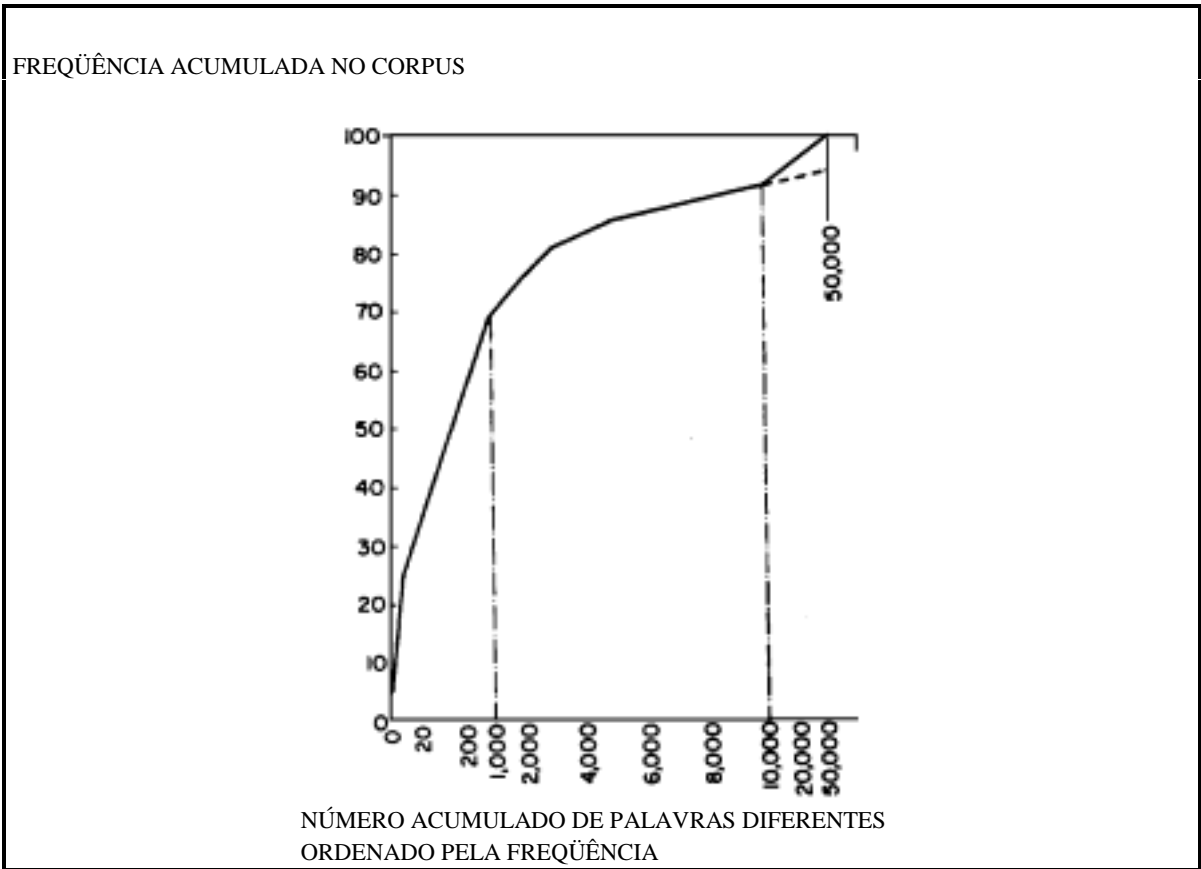


FIGURA 3.2 -Frequência acumulada de entradas em um corpus da Língua Inglesa em função do número de diferentes palavras, ordenadas por sua frequência. (KLATT, 1989).

Essas considerações sugerem que um sistema híbrido, contendo um conjunto de regras de conversão de letras para fonemas e um dicionário de exceções, é uma solução adequada à conversão de textos em fonemas, pois um dicionário de exceções de tamanho moderado pode reduzir deficiências de um conjunto de regras de conversão de letras para fonemas.

Para KLATT (1987), o tamanho de um dicionário necessário para obter um certo nível de acerto na conversão de um texto está extremamente relacionado ao

desempenho do conjunto de regras de conversão. Esse autor cita como exemplo, o sistema "Prose-2000", com um dicionário de exceções de 3.000 palavras e um conjunto de regras de conversão de letras para fonemas de Bernstein, com nível de acerto de 85%, resultando em um sistema com um nível de acerto total superior a 97%, apenas 1 palavra em 33 em um texto aleatório conteria um erro de conversão de fonemas ou de acentuação.

Por outro lado, a primeira versão do "DECtalk", empregando as regras de Hunnicutt com desempenho de 65% e um dicionário de exceções com 6.000 palavras, apresentou um desempenho que raramente alcançava 95%, ou seja, 1 erro a cada 20 palavras.

Uma alternativa para aumentar a abrangência de um dicionário de exceções é a utilização de um dicionário de *morfemas*, que correspondem às menores unidades de um idioma com significado. Nesse caso, é necessário também a utilização de um conjunto de regras para dividir as palavras nos morfemas que a compõe.

KLATT (1987) comenta que uma das principais vantagens de utilização de dicionário de morfemas é a possibilidade de a partir de um conjunto de 12.000 morfemas representar mais de 100.000 palavras da Língua Inglesa, sendo portanto, uma maneira econômica de armazenar um grande número de palavras. Além disso, os morfemas são estáveis com o tempo, podendo ser formadores de novas palavras.

Segundo ALLEN (1976), na Língua Inglesa existem em média menos que dois morfemas por palavra. Esses morfemas incluem os *prefixos* (*con-*, *be-*, *mini-*), *sufixos derivacionais*, que afetam o significado da palavra (*-dom*, *-ness*, *-ship*, *-al*), e *sufixos inflexionais*, que afetam o conteúdo gramatical da palavra (*-s*, *-ed*, *-ing*). Além disso, existem dois tipos de radicais, os *radicais livres*, que podem ser empregados isoladamente como em "snow", "boat" e "house", e os *radicais*

*combinados*, que necessariamente são empregados ligados a um outro morfema adjacente (*-turb, -ceive, crimin-*).

Todas as palavras em Inglês são formadas por combinações dessas unidades, através do processo de *composição*, no qual dois radicais são concatenados, como nas palavras "houseboat" e "snowplow"); e através do processo de *afixação*, no qual prefixos e sufixos são adicionados aos radicais, como nas palavras "enable" e "receive".

As regras de decomposição de palavras em morfemas devem dividir as palavras em seus respectivos morfemas, e também eleger dentre as decomposições possíveis, a mais adequada, pois existem palavras que podem ser separadas em morfemas de modos diferentes. Por exemplo, a palavra "scarcity" pode ser separada como *scarce + ity* ou *scar + city*, preferindo-se neste caso a primeira opção por ser a derivação afixional mais comum do que a composição (KLATT, 1987).

ALLEN (1976) utilizou a divisão de palavras em morfemas no algoritmo de conversão de texto para fonemas utilizado no "MITalk", apresentado na Figura 3.3.

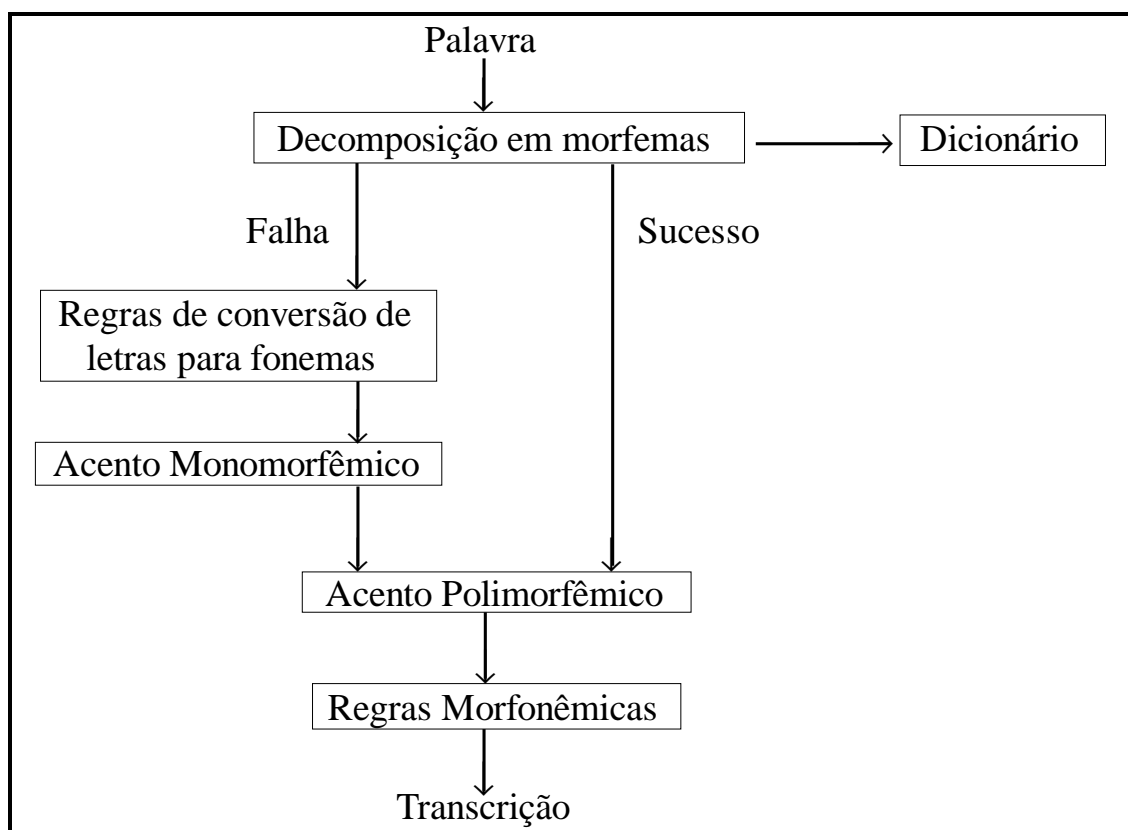


FIGURA 3.3 - Sistema de conversão de letra para fonema. (ALLEN, 1976)

Inicialmente a palavra a ser transcrita é procurada no dicionário ou decomposta em uma seqüência de morfemas do dicionário. Caso não haja sucesso nessas tentativas, a palavra é submetida à um conjunto de regras de conversão de letras para fonemas. No caso do "MITalk" esse conjunto é composto por uma implementação das regras propostas por Hunnicutt, determinando-se a tonicidade de cada morfema separadamente (acento monomorfêmico).

A etapa seguinte consiste em determinar a tonicidade da palavra, considerando o efeito conjunto de todos os morfemas (acento léxico polimorfêmico). Por exemplo, as palavras "photograph", "photography" e "photographic", embora

comecem com o radical *photo*, têm acento respectivamente na primeira, segunda e terceira sílabas (/ˈfɒtə græf/, /fə ˈtɒgrə fi/ e /fɒtə ˈgræfɪk/).

Finalmente, são feitos os acertos finais para correção dos fonemas devido aos efeitos de concatenação dos morfemas. Para o caso de palavras terminadas em *-ed*, a conversão do sufixo depende da última letra do morfema que está sendo modificado. Caso o sufixo *-ed* seja precedido por /t/ ou /d/, será transcrito como /əd/, como nas palavras "persuaded" e "stated"; se estiver precedido por /p/, /k/, /s/, /f/, /θ/, /ʃ/ ou /tʃ/, será transcrito como /t/, como em "stoped" e "walked". Nos demais casos, será transcrito apenas como /d/, como nas palavras "measured" e "doomed".

Outro algoritmo baseado na decomposição em morfemas foi desenvolvido por COKER (1985), com 43.000 morfemas da Língua Inglesa, sendo usado posteriormente no sistema "Conversant" de conversão de texto para voz (HIRSCHBERG et al., 1990).

### **3.2.3 - Compiladores de Regras**

O crescente interesse em sistemas de síntese de voz a partir de texto tem sido acompanhado por uma tendência de aumento na flexibilidade de sistemas de síntese por regras. Esses sistemas têm evoluído de programas específicos para determinada Língua, para sistemas onde linguagens especiais de programação permitem expressar regras em qualquer Idioma, as quais podem ser modificadas facilmente. São os chamados *compiladores de regras* (HERTZ, 1982).

Essa técnica foi utilizada por Carlson no desenvolvimento do "Infovox", capaz de sintetizar voz em sete idiomas (KLATT, 1987).

HERTZ (1982) desenvolveu, na Universidade de Cornell, o "Speech Research System - SRS", com uma linguagem especial de programação que permitia a construção de regras de modificação de texto, regras de conversão de texto para fonemas, regras de modificação de seqüências de fonemas e regras de conversão de fonemas para parâmetros acústicos a serem utilizados por sintetizadores.

As regras de modificação de texto eram as primeiras a serem aplicadas, sendo responsáveis por modificar o texto de entrada, acrescentando informações necessárias às fases seguintes.

A etapa seguinte correspondia à aplicação das regras de conversão de letras para fonemas nas palavras isoladas do texto, que produziam uma seqüência de fonemas correspondentes ao texto de entrada. A seqüência de fonemas podia ser alterada posteriormente, através da aplicação das regras de coarticulação.

E finalmente, eram gerados os parâmetros acústicos para a produção da voz, a partir das entradas da etapa anterior.

A partir da experiência adquirida com o "SRS", HERTZ (1985, 1994) desenvolveu o "Delta System", com capacidade para combinar regras com um dicionário fonético, no qual eram armazenadas palavras que constituíam exceções às regras.

A Figura 3.4 mostra o diagrama de blocos dos componentes do sistema, que inclui um conjunto de regras aplicáveis ciclicamente ou não, seqüencialmente ou simultaneamente, da esquerda para direita ou da direita para esquerda, em sílabas ou morfemas, entre outras características.

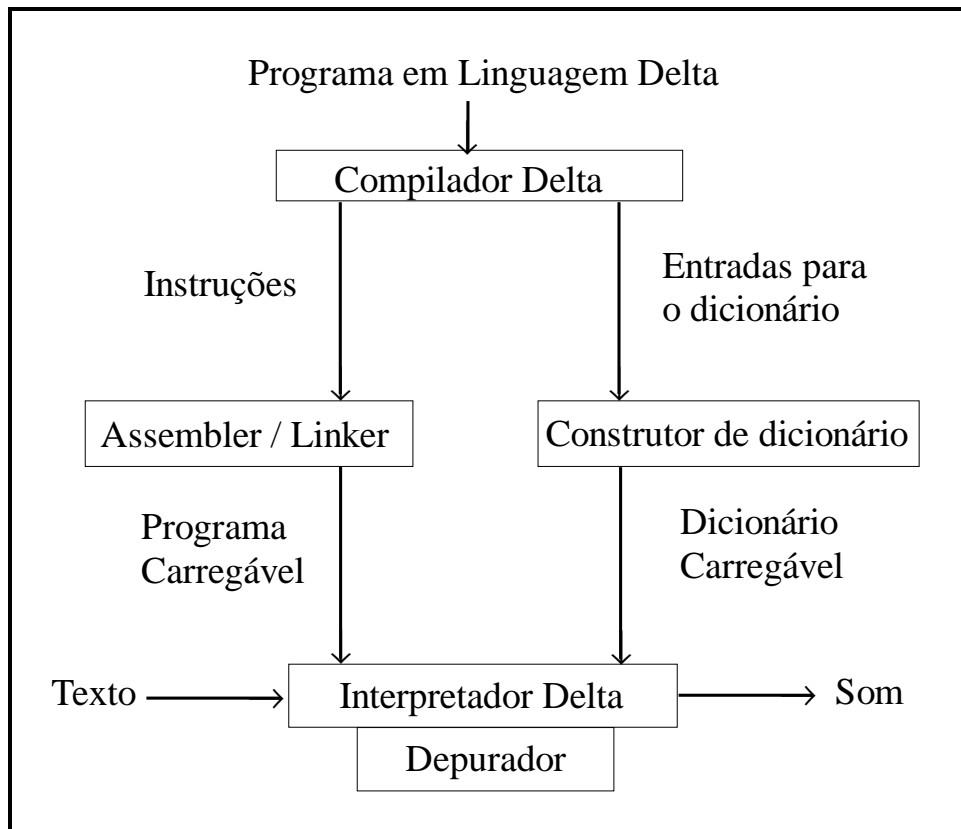


FIGURA 3.4 - Diagrama de blocos do "Delta System". (HERTZ, 1985)

O sistema pode ser executado em microcomputadores PC ou MacIntosh, foi desenvolvido em linguagem C, sendo constituído por cinco módulos. O compilador opera sobre um programa escrito na "linguagem Delta", contendo as regras de conversão e entradas para o dicionário. As regras são convertidas em instruções para o "interpretador Delta" e são transformadas em um programa carregável pelo módulo "assembler/linker". Em seguida, o módulo "construtor de dicionário" produz um dicionário carregável, que juntamente com o programa anteriormente gerado podem ser carregados e executados pelo "interpretador Delta", que utilizando essas

informações, produz a transcrição fonética do texto de entrada, a qual pode ser usada para a síntese de voz.

Vários outros sistemas semelhantes surgiram posteriormente, como o "Development Environment for Pronunciation Expert Systems - DEPES", desenvolvido por VAN COILE (1989) e utilizado na criação de um conjunto de regras para o idioma Holandês e o "Speech Compiler for your Language - SCYLA", adotado por VIANA et al.(1991) para o desenvolvimento de regras para o idioma Português, em estudos realizados no Centro de Linguística da Universidade de Lisboa.

### **3.3 - CONVERSÃO DE TEXTO PARA FONEMAS PARA O IDIOMA PORTUGUÊS**

A conversão de texto para fonemas no idioma Português, de maneira análoga a que ocorre na Língua Inglesa, deve ser executada em vários passos para que possa ser bem sucedida.

Partindo-se do texto de entrada, é necessário a execução de pré-processamento para eliminar abreviaturas, siglas, números e caracteres não alfabéticos, expandindo-os para as correspondentes palavras. Da mesma forma que acontece na língua inglesa, podem ocorrer ambiguidades, como no caso dos caracteres "1" e "2", que podem ser escritos respectivamente como *um* ou *uma*, e *dois* ou *duas* (EGASHIRA, 1992).

OLIVEIRA, VIANA E TRANCOSO (1994) sugerem ainda para essa fase de normalização que os caracteres acentuados sejam codificados em dois caracteres, como por exemplo, o *á* seria transformado em *a'* e o *ç* tornar-se-ia *c,*. Esse

procedimento possibilita que textos escritos em computadores de fabricantes diferentes, que utilizam extensões de códigos ASCII distintos, possam ser tratados nas etapas seguintes alterando-se apenas o módulo de pré-processamento de texto.

A análise sintática é necessária não apenas para determinar a correta entonação do "pitch", mas também, para determinar a correta pronúncia de palavras como *g[ó]sto*, verbo, e *g[o]sto*, substantivo.

Em estudos realizados em um corpus da Língua Portuguesa, VIANA et al. (1991) verificaram que essas palavras, denominadas homógrafas heterófonas, ou seja, com a mesma grafia porém com pronúncias diferentes, contituíam cerca de 3% do corpus de teste.

A conversão de texto para fonemas mais especificamente na Língua Portuguesa não oferece tantas dificuldades quanto na Língua Inglesa, visto que a maioria das consoantes podem ser transcritas diretamente a partir de um conjunto de regras.

Na conversão de letras para fonemas na Língua Portuguesa um dos principais problemas encontrados pelos diversos autores, refere-se à transcrição das letras "e" e "o", quando pertencem à sílaba tônica de uma palavra podendo ser pronunciadas respectivamente como [e] ou [é] e [o] ou [ó].

Poder-se-ia tentar formular algumas regras que mesmo não sendo adequadas a todas as situações, abrangeriam a maioria dos casos. No entanto, EGASHIRA (1992) comenta que seria necessário o acesso a medidas estatísticas de associação ortográfico-fonética. Além disso, deve-se considerar ainda o caso das palavras homógrafas heterófonas, que somente podem ter suas pronúncias determinadas após análise sintática. Por esse motivo, é indispensável um dicionário de exceções, que contenha a transcrição de palavras que falhem à aplicação das regras.

O mesmo ocorre para a letra "x", que pode ser associada a quatro fonemas diferentes, [x] na palavra *xale*; [ks] na palavra *fixo*; [s] na palavra *texto* e [z] na palavra *exame*. A Tabela 3.3 apresenta algumas regras para a conversão fonética dos fonemas [x], [z] e [s], porém, deve-se notar mais uma vez que essas regras não serão suficiente para atender a todos os casos, devendo-se utilizar um dicionário de exceções.

TABELA 3.3 - Algumas regras para conversão fonética da letra "x". (ALIANDRO, 1974)

Fonema	Ocorrência	Exemplo
[x]	- Início de palavra - Depois de "n" - Depois de "ai", "ei" e "ou"	xícara, xarope enxame, enxofre caixa, eixo, frouxo
[z]	- Palavras iniciadas com "ex" seguido de vogal	exame, exercício, exótico
[s]	- Seguido de consoante	texto, sexto

Alguns aspectos referentes a nasalidade das vogais foram destacados por CALLOU e LEITE (1990), ressaltando dificuldades na transcrição de vogais nasais quando estas não estão marcadas com til. Segundo Perrini, a nasalização das vogais ocorre em três situações distintas, quando a vogal é acentuada seguida de consoante nasal ([m] ou [n]); sempre que a vogal for seguida por consoante nasal e outra consoante, e quando a vogal estiver antes de consoante nasal em final de vocábulo.

Entretanto, existem exceções a essa regra, que não é capaz de distinguir *c[ã]minha*, substantivo, de *c[a]minha*, verbo. Nesses casos, as duas alternativas devem estar presentes no dicionário de exceções e a seleção da transcrição adequada depende de análise sintática do texto de entrada.

Cabe ressaltar ainda a questão do "acento secundário" de algumas palavras derivadas por sufixação. Os sufixos *-(z)inha* e *-(z)inho* ocasionam nos radicais a que são acrescentados um comportamento diferente da maioria dos outros sufixos, na medida em que permitem a manutenção da qualidade da vogal do radical, como na palavra *av[ó]zinha*. Esse aspecto contraria a generalização de que em posição pré-tônica neutraliza-se o contraste entre /e:/ /é/ e /o:/ /ó/. O mesmo comportamento é apresentado também pelos sufixos *-mente*, *-íssimo* e *-íssima* (CALLOU e LEITE, 1990).

No caso de palavras com o sufixo *-mente* a determinação do "acento secundário" torna-se mais simples uma vez que esse sufixo não altera a palavra à qual está sendo acrescentado, e portanto, procede-se à verificação da palavra obtida através da eliminação do sufixo, como por exemplo no caso da palavra *fortemente* → *forte+mente* → *f[ó]rte+mente*.

No entanto, a determinação do "acento secundário" torna-se um processo mais complexo no caso de palavras com os sufixos *-íssimo(a)* e *-inho(a)*, pois é comum a supressão ou a mudança da vogal final, como nas palavras *belo* → *belíssimo*, *excelente* → *excelentíssimo*, *amarelo* → *amarelinho*, *verde* → *verdinho*.

## **CAPÍTULO IV**

### **DESENVOLVIMENTO DO SISTEMA**

#### **4.1 - AMBIENTE E INTERFACE DO SISTEMA**

O sistema desenvolvido para conversão de texto para fonemas no idioma Português compõe-se basicamente por dois elementos, um dicionário com informações fonológicas e gramaticais e um conjunto de regras de conversão de letras para fonemas.

Cada palavra a ser convertida é procurada no dicionário e sua transcrição ou possíveis transcrições fonológicas serão fornecidas caso seja encontrada no dicionário. Caso contrário, não sendo encontrada no dicionário, a palavra deverá ser

submetida a um conjunto de regras para sua conversão. A Figura 4.1 ilustra o processo.

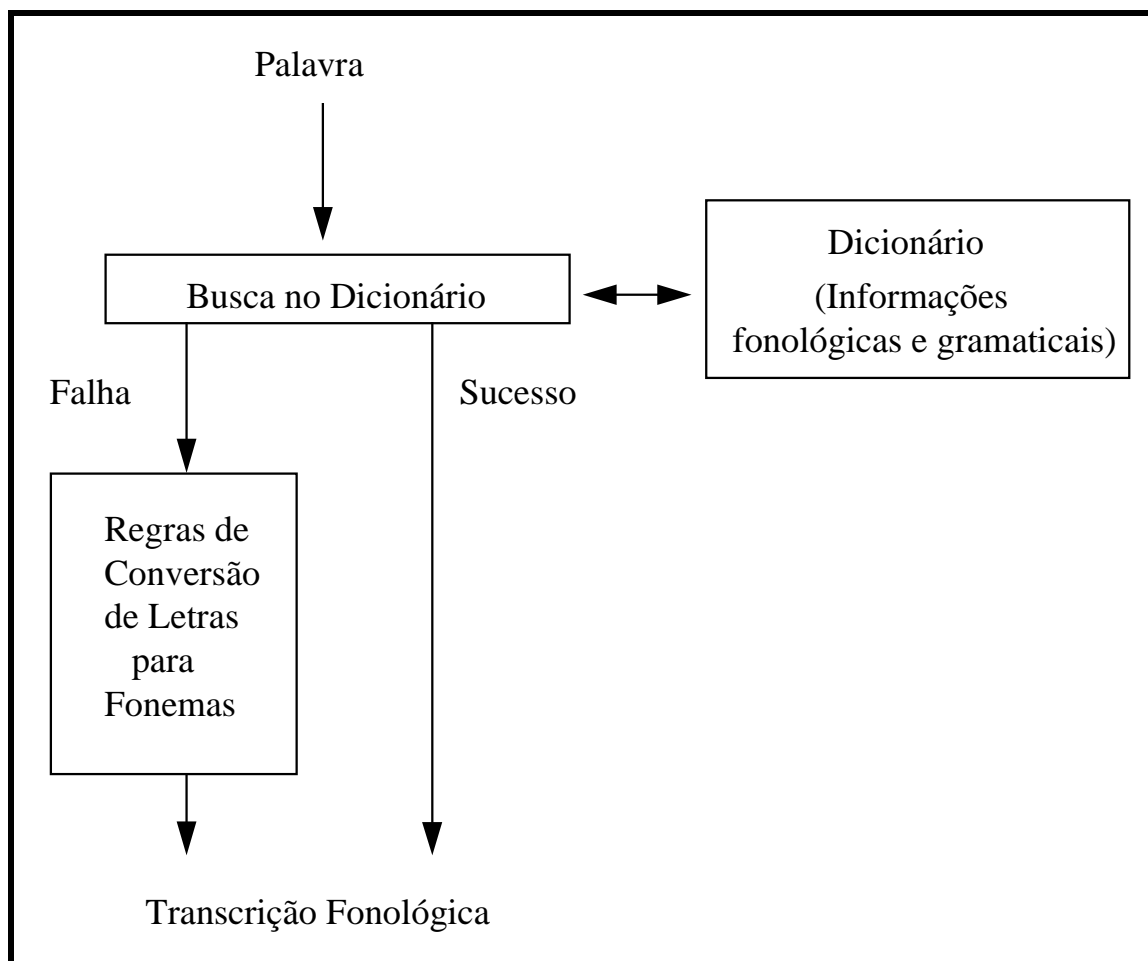


FIGURA 4.1 - Esquema de conversão de palavras para seus respectivos fonemas utilizado.

O dicionário de exceções, além de informações fonológicas armazena também a classe gramatical das palavras, possibilitando a utilização desta estrutura em implementações de análise sintática.

As principais atividades associadas ao dicionário correspondem à montagem de sua estrutura e à busca de palavras para conversão. A primeira atividade resultou na montagem de uma estrutura para suportar o dicionário, podendo ser acrescentadas ao Dicionário palavras que estejam em um arquivo de formato pré-determinado ou podendo ser introduzidas uma a uma.

A busca de palavras no Dicionário corresponde à atividade que ocorre durante o processo de conversão de uma palavra, o que pode resultar na transcrição fonológica da palavra (ou possíveis transcrições) ou na necessidade da aplicação das regras de conversão.

O conjunto de regras de conversão é utilizado em duas situações, na montagem do Dicionário, devendo ser capaz de absorver informações adicionais do arquivo de entrada, e durante a conversão de uma palavra não encontrada no Dicionário, quando a transcrição fonológica é feita apenas com as informações da própria palavra.

Para executar a tarefa da transcrição fonológica das palavras foi desenvolvido o programa CONTEXFO, que pode ser executado em micros da linha PC-AT, com DOS 5.0 ou superior, tendo sido implementado na linguagem PASCAL com a utilização do Turbo Pascal 5.0 da Borland.

Os caracteres acentuados da Língua Portuguesa devem estar codificados de acordo com a Tabela Code Page 850. Para a geração desses caracteres deve-se executar o programa KEYB, que acompanha o DOS a partir da versão 5.0 (KEYB BR,850, C:\DOS\KEYBOARD.SYS), permitindo que caracteres acentuados sejam gerados pressionando-se o acento seguido da correspondente letra.

Caracteres com acento agudo são obtidos digitando-se primeiro o acento depois o caracter (Ex: 'a → á). Os caracteres com acento circunflexo, til ou crase são

obtidos dessa mesma maneira. O *cê cedilha* é obtido digitando-se acento agudo + "c" ('c → ç), e o *u com trema* é obtido digitando-se aspas duplas + "u" ("u → ü).

O programa CONTEXFO utiliza dois outros arquivos, que serão descritos com maiores detalhes posteriormente. Um dos arquivos contém o Dicionário de exceções, denominado DIC.CON, e o outro arquivo de configuração, denominado REGRAS.CON, contém as regras de conversão a serem utilizadas.

Ao ser executado, o programa CONTEXFO oferece inicialmente três opções, que devem ser selecionadas de acordo com o número.

- 1 - Transcrição fonológica de palavras
- 2 - Montagem do Dicionário.
- 3 - Fim.

A primeira dessas opções permite a conversão de palavras para fonemas, a segunda encarrega-se da construção do Dicionário a ser utilizado e a última opção encerra o programa.

Quando a primeira opção for selecionada, deve-se digitar isoladamente cada palavra a ser convertida, seguida de <ENTER>, para que sua transcrição seja fornecida. Quando for pedida uma palavra, teclando-se apenas <ENTER>, serão novamente oferecidas as três opções iniciais.

Quando a segunda opção for selecionada, o programa oferece um novo conjunto de opções, que devem ser selecionadas de acordo com o número.

- 1 - Inclui palavras isoladas.
- 2 - Inclui palavras de um arquivo.
- 3 - Minimiza Dicionário.
- 4 - Procura palavras no Dicionário.
- 5 - Mostra Dicionário gerado.

6 - Salva Dicionário em disco.

7 - Recuperar Dicionário gravado.

8 - Retorna.

As duas primeiras opções incluem palavras em uma estrutura de "árvore de letras", isoladamente ou a partir de um arquivo, contendo palavras que devem seguir o formato do Vocabulário Ortográfico proposto pela Academia Brasileira de Letras, conforme item 4.2.1. Quando as palavras forem introduzidas isoladamente, poderão ser fornecidas algumas características fonológicas adicionais e a classe gramatical a que pertencem, seguindo o mesmo formato dos arquivos.

A terceira opção permite minimizar a estrutura de árvore de letras gerada pelas duas primeiras opções, a quarta opção permite a busca de palavras no Dicionário, e a quinta opção mostra a estrutura do dicionário gerado na tela ou na impressora.

A sexta e a sétima opções permitem respectivamente salvar e recuperar o Dicionário gerado em um arquivo cujo nome é pedido. E a última opção retorna para o "menu" inicial.

## **4.2 - PRINCIPAIS CARACTERÍSTICAS DO DICIONÁRIO**

### **4.2.1 - Conteúdo do Dicionário**

O Dicionário deve conter palavras que falham à aplicação das regras de conversão de letras para fonemas. As entradas para o Dicionário seguem o formato

do Vocabulário Ortográfico, que contém informações sobre a pronúncia de algumas letras ("e", "o", "x"), e sobre classes gramaticais, origem e sinônimos das palavras, com mais de 320.000 palavras, sem contar as conjunções verbais que não estão expandidas. Uma entrada típica desse Dicionário corresponde à:

*palavra (transc. fon. de uma letra) classe gramatical, comentários: sinônimo*

Seguindo as características da entrada esquematizada, cita-se como exemplos algumas entradas para palavras iniciadas com a letra "a":

*a s.m.; f. sing. do art. e do pron. pess. ou dem. o; prep.*

*aacheniano adj. s.m.:aaqueniano*

*ababelador (ô) adj.*

*abajur s.m., do fr. abat-jour*

*abastar v.*

Algumas palavras possuem informações parentéticas sobre transcrições fonológicas, como o (ô) de *ababelador* no exemplo acima.

Durante a montagem do dicionário, as seguintes informações adicionais sobre transcrições de letras também são consideradas:

- (ô) e (ó), para a letra "o", corresponde a [o] e [ó]
- (ê) e (é), para a letra "e", corresponde a [e] e [é]
- (cs), (ss) e (z), para a letra "x", corresponde a [ks], [s] e [z], não existindo porém, indicação para [x].

Quando houver mais de uma transcrição possível, como na palavra *exostose* (z ou cs) somente a primeira será considerada por questões de simplificação, embora a estrutura montada permita que ambas sejam consideradas.

São tratadas também informações relativas à separação de vogais. Sempre que houver uma indicação de hiato, do tipo (*vogal-vogal*), essa informação será

utilizada em conjunto com o algoritmo de separação de sílabas para a correta separação da palavra. Cita-se como exemplo a entrada da palavra *abairense (a-i) adj. s.2g.*

Quando houver indicação de hiato e informação fonológica de uma letra, estas deverão estar separadas por vírgulas, como em na palavra *abstraidor (a-i, ô) adj. s.m.*

O Vocabulário Ortográfico fornece ainda indicações de acentuação secundária, como em *abatezinhense (tè) adj.s.2g.*, palavra na qual o acento secundário é marcado com acento grave. Para que essa informação seja tratada corretamente, a letra acentuada deve estar acompanhada da letra que a antecede ou da letra que a sucede.

Foram utilizadas para a etapa da seleção de palavras, na preparação do Dicionário, partes do Vocabulário Ortográfico, dada a inviabilidade da sua utilização integral. Para realização dos testes trabalhou-se com partes desse vocabulário, contendo até 2.000 palavras. A grande dificuldade para a utilização dos arquivos desse vocabulário está na falta de uniformidade dos dados apresentados, principalmente no que se refere à irregularidade na pontuação, a qual necessitará de correção em trabalhos posteriores.

#### **4.2.2 - Estrutura de Armazenamento do Dicionário**

O Dicionário é armazenado através da montagem de um reconhecedor de estados finitos (autômato), cuja linguagem por ele reconhecida é o conjunto de palavras aceitas. Segundo autores como GROS (1989), LUCCHESI e

KOWALTOWSKI (1991) o armazenamento de palavras através de um autômato finito é uma forma bastante eficiente de armazenamento de palavras nas Línguas Românicas, pois oferece uma maneira implícita de compartilhar prefixos e sufixos.

Ao reconhecedor são acrescentadas informações fonológicas e gramaticais adicionais, de modo que a transcrição fonológica é realizada a medida que o autômato está sendo percorrido.

Antes da palavra ser inserida no Dicionário, realiza-se a sua conversão fonológica com base no conjunto de regras e nas informações adicionais fornecidas no formato do Vocabulário Ortográfico.

Ao estado final (estado de aceitação) de cada palavra será acrescentada a classe gramatical à qual pertence, e a cada transição será associada além da letra, seu correspondente fonema.

Nestas condições, podem ocorrer duas situações:

- a existência de estados que possuam mais de uma transição com a mesma letra, representando fonemas diferentes. Esta situação ocorre para o autômato do exemplo a seguir, que reconhece as palavras *retorno* (substantivo), *retorno* (verbo, 1ª. pessoa do Presente do Indicativo); *contorno* (substantivo), *contorno* (verbo, 1ª. pessoa do Presente do Indicativo). No estado 4 ocorrem duas transições com a letra "o", que caracterizam suas pronúncias aberta e fechada, representadas por [o] e [ó];

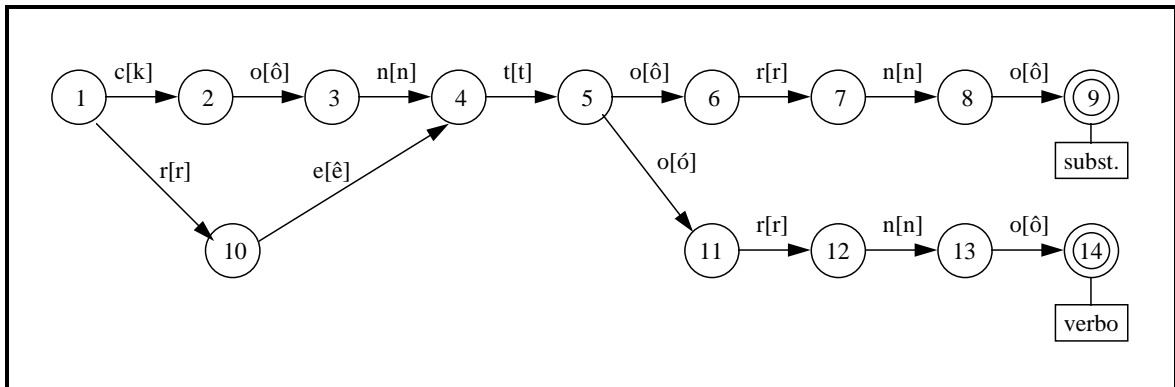


FIGURA 4.2 - Autômato de reconhecimento das palavras *retorno* (substantivo), *retorno* (verbo, 1ª. pessoa do Presente do Indicativo); *contorno* (substantivo), *contorno* (verbo, 1ª. pessoa do Presente do Indicativo).

- a existência de transições nas quais não há fonemas associados. É o caso dos dígrafos, aos quais não se associam fonemas às segundas letras, como na palavra *chuva*, conforme ilustra o autômato da Figura 4.3.

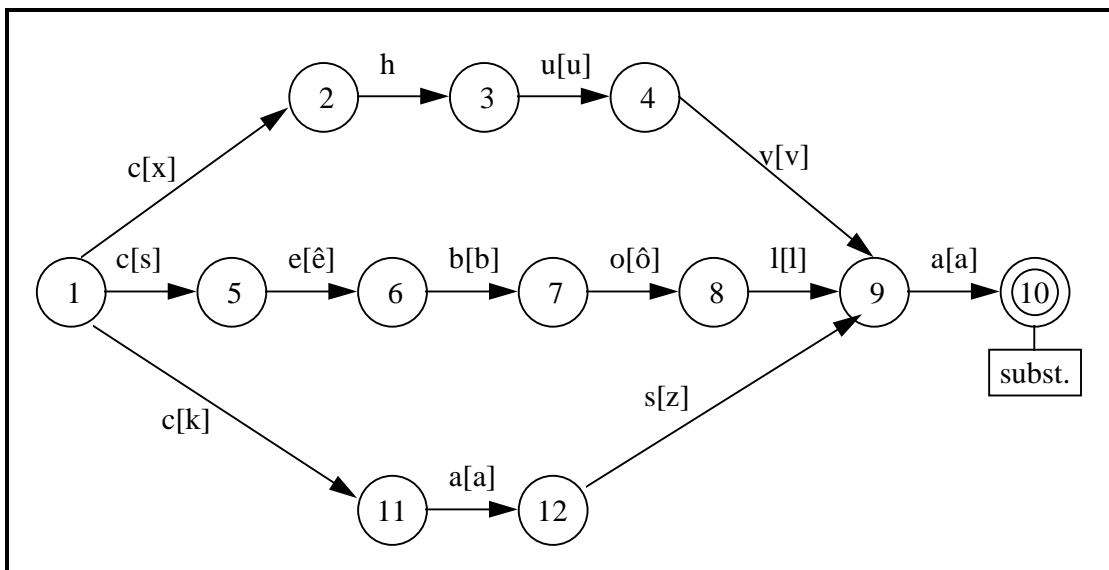


FIGURA 4.3 - Autômato de reconhecimento das palavras *chuva*, *casa* e *cebola*.

#### 4.2.3 - Montagem e Estrutura de Dados do Dicionário

O reconhecedor de palavras do Dicionário cria inicialmente uma estrutura conhecida como "trie", termo derivado de "retrieval" (AL-SUWAOYEL e HOROWITZ, 1984).

Para essa estrutura, que se assemelha a uma "árvore de letras", quando duas palavras começam com a mesma letra elas compartilham os estados iniciais de seus caminhos. A primeira letra que diferir nas duas palavras gera transições para estados diferentes (APPEL e JACOBSON, 1988).

A Figura 4.4 mostra a estrutura do reconhecedor inicial das palavras *canto*, *canta*, *cantei*, *nado*, *nada*, *nadei*, *pula*, *pulo*, *pulei*, sem as informações gramaticais. Nesse exemplo, o estado 1 é o estado inicial e os estados marcados com duas circunferências correspondem aos estados finais.

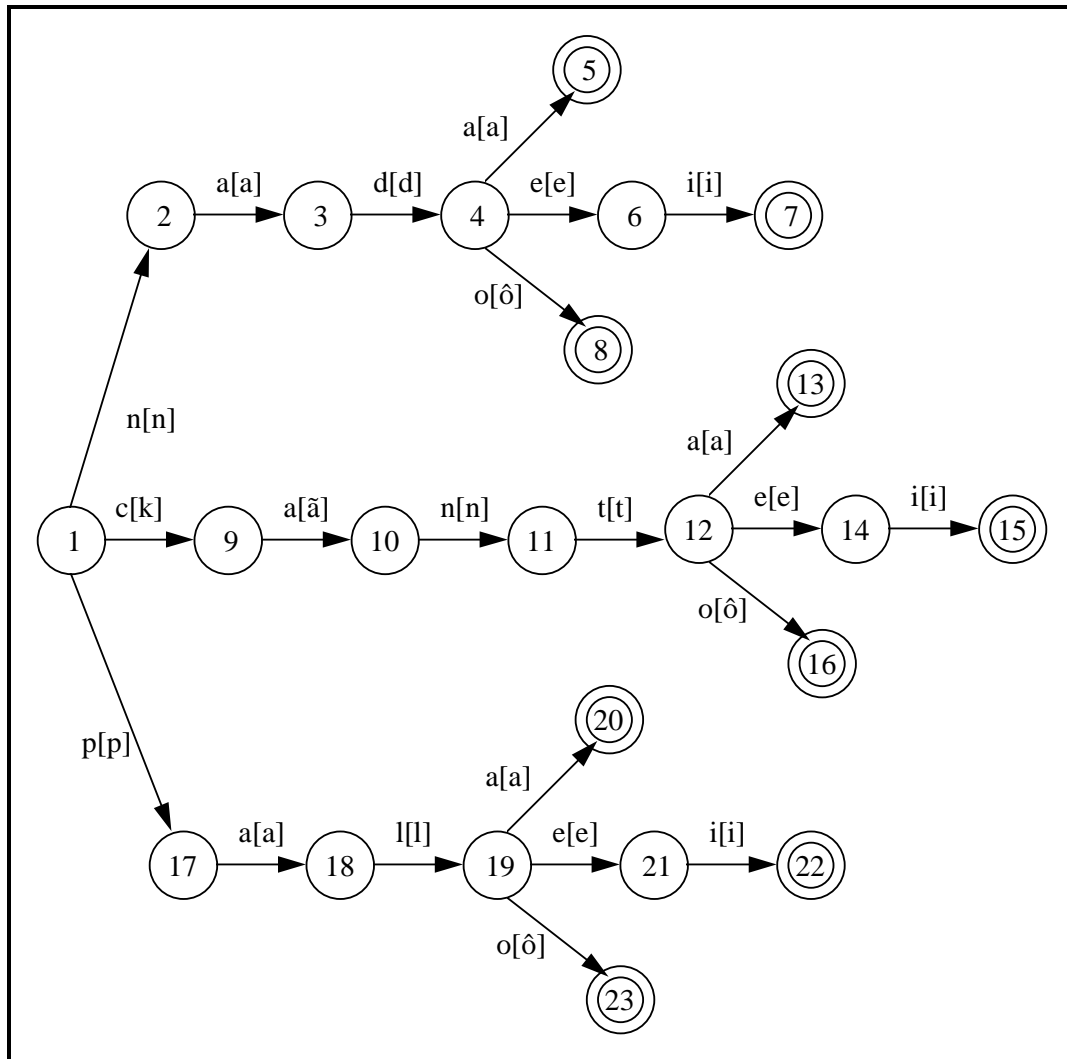


FIGURA 4.4 - Estrutura inicial do reconhecedor das palavras *canto*, *canta*, *cantei*, *nado*, *nada*, *nadei*, *pulo*, *pula*, *pulei*.

Com a minimização da "árvore de letras" obter-se-á o reconhecedor mínimo, permitindo que palavras que terminem com as mesmas letras compartilhem os estados finais. A estrutura final do reconhecedor das palavras *canto*, *canta*, *cantei*, *nado*, *nada*, *nadei*, *pula*, *pulo*, *pulei* está indicada na Figura 4.5, que possui dez estados a menos do que o reconhecedor inicial.

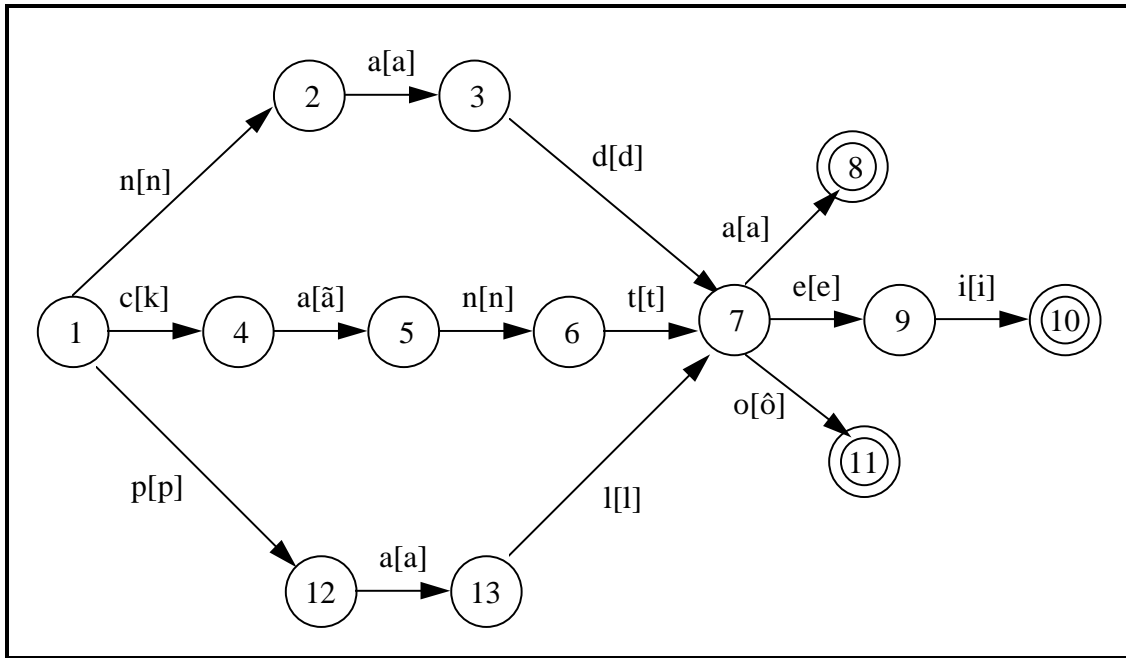


FIGURA 4.5 - Estrutura final do reconhecedor das palavras *canto*, *canta*, *cantei*, *nado*, *nada*, *nadei*, *pulo*, *pula*, *pulei*.

A minimização do reconhecedor de palavras é feita através de algoritmo clássico descrito por JOSÉ NETO (1987), baseado na identificação de estados equivalentes através de refinamentos sucessivos. Esse algoritmo leva em consideração que um estado final não pode ser equivalente a um estado não-final; e que estados que transitam consumindo letras diferentes também não podem ser equivalentes.

Na estrutura de dados implementada, armazenada no arquivo DIC.CON, cada transição é representada pela tripla (*letra*, *fonema*, *próximo\_estado*). Cada estado está associado à outra tripla (*número\_de\_transições*, *classe\_gramatical*, *ponteiro\_para\_transições*). Os estados finais são indicados com a presença de

informação no campo *classe\_gramatical*, que corresponde à classe gramatical das palavras que terminam naquele estado. Para cada estado, as triplas correspondendo às transições são colocadas juntas, conforme indica a Figura 4.6, sendo alocadas dinamicamente durante a montagem do Dicionário. Transições com a mesma letra representando fonemas diferentes são colocadas lado a lado, de modo que para saber se há mais de uma transição com a mesma letra basta verificar a transição seguinte.

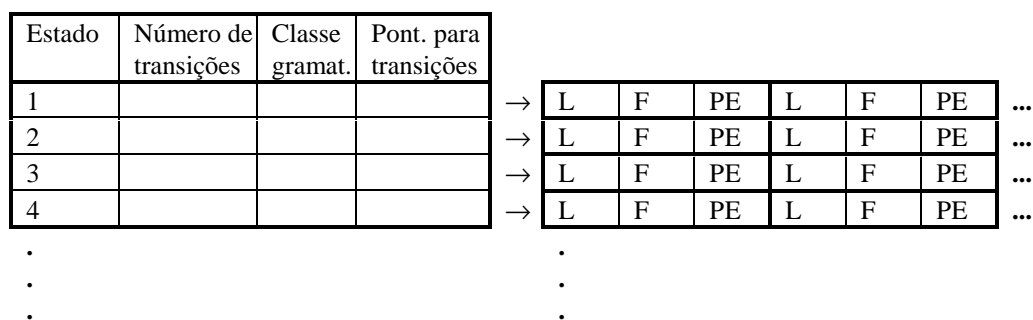


FIGURA 4.6 - Estrutura de dados final do autômato.

A Figura 4.7 ilustra a estrutura de dados do reconhecedor das palavras *active*, *acabar*, *acerola* e *achar*.

Estado	Número de Transições	Classe Gramat.	Ponteiro p/ Transições	Transições
1	1			→ a, [a], 2
2	3			→ c, [k], 3    c, [s], 12    c, [x], 18
3	2			→ l, [l], 4    a, [a], 8
4	1			→ i, [i], 5
5	1			→ v, [v], 6
6	1			→ e, [e], 7
7	0	s.m., adj.		
8	1			→ b, [b], 9
9	1			→ a, [a], 10
10	1			→ r, [r], 11
11	0	v.		
12	1			→ e, [e], 13
13	1			→ r, [r], 14
14	1			→ o, [ó], 15
15	1			→ l, [l], 16
16	1			→ a, [a], 17
17	0	s.f.		
18	1			→ h, [ ], 19
19	1			→ a, [a], 20
20	1			→ r, [r], 21
21	0	v.		

FIGURA 4.7 - Estrutura de dados do reconhecedor das palavras *active*, *acabar*, *acerola* e *achar*.

#### 4.2.4 - Busca de Palavras no Dicionário

A busca de palavras no Dicionário corresponde à atividade que ocorre durante a transcrição fonológica das palavras, assim, à medida que o reconhecedor vai sendo percorrido, obtém-se a transcrição fonológica das palavras.

Para os casos nos quais há estados transitando com a mesma letra representando fonemas diferentes é necessário a execução de um algoritmo de "backtracking". Para o caso das palavras homógrafas heterófonas as possíveis transcrições são fornecidas, juntamente com as respectivas classes gramaticais.

O exemplo da transcrição da palavra *achar* a partir do reconhecedor das palavras *active*, *acabar*, *acerola* e *achar* (Figura 4.6) ilustra esse processo. Partindo-se do estado 1, tem-se a seqüência de análise indicada na Tabela 4.2.

TABELA 4.2 - Transcrição fonológica da palavra *achar* a partir do reconhecedor de palavras.

Passo	Estado	Letra de entrada	Próxima transição com a mesma letra?	Pilha para backtracking			Próximo estado	Saída	Comentário
				Est.	Trans.	Letra			
1	1	a					2	[a]	
2	2	c	sim	2	2a.	2a.	3	[ak]	
3	3	h					2	[a]	Executa backtracking
4	2	c	sim	2	3a.	2a.	12	[as]	
5	12	h					2	[a]	Executa backtracking
6	2	c					18	[ax]	
7	18	h					19	[ax]	
8	19	a					20	[axa]	
9	20	r					21	[axar]	
10	21							[axar], verbo	

Nos passos 2 e 4, a transição seguinte também é realizada com a letra "c", por essa razão foram armazenadas determinadas informações:

- o estado atual, que possui mais de uma transição com a letra de entrada,
- a posição da próxima transição com a correspondente letra dentro das transições do referido estado, e
- a posição da letra na palavra de entrada.

Não havendo sucesso na busca, como nos passos 3 e 5, retorna-se para a última condição armazenada na pilha de "backtracking". Assim, no passo 4 e 6, retorna-se para o estado 2, examinando a palavra a partir da letra "c", partindo-se respectivamente da 2ª e 3ª transições desse estado.

Caso não haja sucesso na busca de uma palavra no dicionário, a palavra é submetida ao conjunto de regras descritas a seguir.

### **4.3 - REGRAS DE CONVERSÃO DE LETRAS PARA FONEMAS**

As regras de conversão de letras para fonemas são aplicadas examinando-se o contexto próximo ao caracter que estiver sendo analisado. Essas regras são necessárias durante a montagem do Dicionário, etapa na qual podem estar presentes informações adicionais a respeito da transcrição fonológica de algumas letras, ou durante a transcrição fonológica de uma palavra, quando a mesma não estiver presente no Dicionário. Esse processo de aplicação das regras de conversão pode ser dividido em três etapas, na separação de sílabas, na determinação da sílaba tônica e finalmente na aplicação das regras propriamente ditas.

A necessidade da determinação da sílaba tônica deve-se ao fato de existirem regras que exigem essa informação. Assim por exemplo, na palavra *banana* ([banãna]), as duas primeiras letras "a" são seguidas de "n"; no entanto, apenas a segunda torna-se nasal por pertencer à sílaba tônica.

Além disso, apesar do Vocabulário Ortográfico apresentar as transcrições fonológicas das letras "e" e "o" quando tônicas não acentuadas, existem entradas como nas palavras *ababosador* (*ô*) *adj.* e *abafarete* (*ê*) *s.m.* nas quais o (*ô*) e o (*ê*) referem-se respectivamente ao segundo "o" de *ababosador* e ao primeiro "e" de *abafarete*, que pertencem às sílabas tônicas dessas palavras. Assim, quando uma palavra do Vocabulário tiver mais de uma letra "e" ou "o" e tiver uma indicação de (*ê*), (*é*), (*ô*) ou (*ó*) em seguida, é necessário determinar a qual das letras corresponde essa indicação, relacionada à letra pertencente à sílaba tônica.

Embora para a divisão silábica de uma palavra seja necessário a utilização de informações relativas à posição da sílaba tônica, a separação das sílabas será efetuada de início, a partir de um algoritmo baseado apenas nas letras de entrada. Correções posteriores podem ser feitas por meio de informações adicionais obtidas no Vocabulário Ortográfico.

Os algoritmos utilizados abrangem a maioria dos casos, porém, as exceções devem ser incluídas no dicionário.

#### **4.3.1 - Separação de Sílabas**

A separação de sílabas é realizada através da análise das letras das palavras, que são examinadas da esquerda para a direita. O algoritmo proposto não separa

encontros vocálicos (ditongos, tritongos e hiatos), e foi implementado como um autômato finito, cuja transição de estados está representada na Tabela 4.4.

Na Tabela 4.4 cada linha corresponde a um estado e as colunas possuem duas informações para cada "átomo de entrada", uma referente ao próximo estado e a outra referente à ação a ser tomada. O estado inicial corresponde ao 1 e o estado final ao 15. A notação utilizada para os "átomos de entrada" é a seguinte:

V	- a, e, i, o, u, y, ã, õ
VA	- á, é, í, ó, ú, â, ê, ô
CO	- ç, c, b, d, f, j, k, v, w, x, q, z,
G	- g
P	- p
T	- t
S	- s
R	- r
L	- l
N	- n
M	- m
H	- h
#	- delimitador de palavra (toda palavra é precedida e sucedida por #)

As ações a serem executadas nas transições de estado, são designadas por números (ação1...ação11) e estão expressas na Tabela 4.4, na qual foram usadas as seguintes convenções:

p - "string" com a palavra iniciada e terminada por #

p [i] - i-ésima letra da palavra

s - vetor de mesmo tamanho de p, onde ao final da separação, cada posição indica o número da sílaba da correspondente letra de p. Por exemplo:

p = #caderno# → s = #1122233#

ns = número da sílaba atual (iniciado com 1)

$i$  = indexador de letras (iniciado com 2, e incrementado após cada ação)

TABELA 4.3 - Ações executadas pelo do algoritmo de separação de sílabas.

Número da Ação	Ações Executadas
1	$s[i] = ns$
2	$ns = ns + 1$ $s[i] = ns$
3	Se $p[i] = p[i-1]$ então executar ação 2 senão executar ação 1
4	$s[i-2] = ns - 1$ $s[i-1] = ns$ $s[i] = ns$
5	$s[i-2] = ns - 1$ $s[i-1] = ns - 1$ $s[i] = ns$
6	$s[i-1] = ns - 1$ $s[i] = ns$
7	$s[i-1] = ns - 1$
8	$s[i-2] = ns - 2$ $s[i-1] = ns - 1$
9	Nada
10	Se $(p[i-2] = V \text{ e } p[i-3] = V) \text{ e } (p[i-4] = g \text{ ou } q)$ então $s[i-2] = ns$ $s[i-1] = ns$ senão $s[i-1] = ns - 1$
11	Se $p[i-2] = V \text{ ou } VA$ então $s[i-2] = ns - 1$ $s[i-1] = ns - 1$

TABELA 4.4 - Transição de estados para separação de sílabas.

	V		VA		CO		G		P		T		S		R		L		N		M		H		#	
	PE	A	PE	A	PE	A	PE	A	PE	A	PE	A	PE	A	PE	A	PE	A	PE	A	PE	A	PE	A	PE	A
1	2	1	3	1	4	1	5	1	6	1	7	1	8	1	9	1	10	1	10	1	11	1	4	1	15	9
2	2	3	3	2	4	2	5	2	6	2	7	2	8	2	9	2	10	2	10	2	11	2	4	2	15	9
3	2	2	15	9	4	2	5	2	6	2	7	2	8	2	9	2	10	2	10	2	11	2	4	2	15	9
4	2	1	3	1	4	6	5	6	6	6	7	6	13	9	12	1	12	1	10	6	11	6	12	1	15	7
5	2	1	3	1	13	9	13	9	13	9	13	9	13	9	12	1	12	1	14	1	13	9	13	9	15	7
6	2	1	3	1	13	9	13	9	13	9	14	1	14	1	12	1	12	1	14	1	13	9	12	1	15	7
7	2	1	3	1	4	6	5	6	6	6	7	6	8	6	12	1	12	1	7	6	14	1	4	6	15	7
8	2	1	3	1	4	6	5	6	6	6	7	6	8	6	9	6	10	6	10	6	11	6	4	6	15	11
9	2	1	3	1	4	6	5	6	6	6	7	6	8	6	9	6	10	6	10	6	11	6	4	6	15	10
10	2	1	3	1	4	6	5	6	6	6	7	6	8	6	9	6	10	6	10	6	11	6	12	1	15	10
11	2	1	3	1	4	6	5	6	6	6	7	6	8	6	9	6	10	6	14	4	11	7	4	7	15	10
12	2	1	3	1	4	6	5	6	6	6	7	6	8	6	9	6	10	6	14	4	11	7	4	7	15	8
13	2	4	3	4	4	5	5	5	6	5	7	5	8	5	9	5	10	5	10	5	11	5	4	5	15	8
14	2	1	3	1	4	2	5	2	6	2	7	2	8	2	9	2	10	2	10	2	11	2	4	2	15	11
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Como exemplo, a palavra *pneumático* foi separada segundo o algoritmo descrito acima. Na Tabela 4.5 estão indicados os passos para a separação silábica.

TABELA 4.5 - Seqüência de passos para separação de sílabas da palavra *pneumático*.

Passo	Estado	Letra	Classe	Ação	Próximo Estado	Resultado
1	1	p	P	1	6	pneumático
2	6	n	N	1	14	pneumático
3	14	e	V	1	2	pneumático
4	2	u	V	3	2	pneumático
5	2	m	M	2	11	pneu-mático
6	11	á	VA	1	3	pneu-mático
7	3	t	T	2	7	pneu-má-tico
8	7	i	V	1	2	pneu-má-tico
9	2	c	CO	2	4	pneu-má-ti-co
10	4	o	V	1	2	pneu-má-ti-co
11	2	#	FIM	9	15	pneu-má-ti-co
12	15					pneu-má-ti-co

Quando houver um encontro vocálico com hiato não acentuado, o algoritmo descrito não executa a separação, como na palavra *rai-nha*. Nesse caso, durante a montagem do dicionário as informações entre parênteses devem ser utilizadas para correção. Entretanto, se esse tipo de encontro vocálico ocorrer na transcrição fonológica de uma palavra que não conste do dicionário, não será factível qualquer correção.

Um teste exaustivo desse algoritmo seria extremamente trabalhoso, dado o elevado número de palavras da Língua Portuguesa. Assim, o algoritmo de separação de sílabas foi testado para um conjunto de 200 palavras previamente selecionadas, e o seu desenvolvimento sofreu refinamentos sucessivos até que produzisse 100% de acerto na determinação da sílaba tônica das palavras que não apresentam hiato. O conjunto de palavras utilizadas e a separação gerada pelo programa estão indicados no Anexo B.

#### **4.3.2 - Determinação da Sílaba Tônica**

A importância da determinação da sílaba tônica relaciona-se ao papel de destaque que desempenha no estudo da prosódia, e especialmente à correta aplicação de determinadas regras.

A posição da sílaba tônica é uma informação importante a ser considerada na formulação de modelos que controlem a variação dos parâmetros prosódicos a fim de imprimir à fala sintetizada uma maior naturalidade. A variação dos parâmetros prosódicos da fala tais como duração segmental, frequência fundamental e amplitude são fortemente dependentes da posição da sílaba tônica (EGASHIRA, 1992).

E para a determinação da sílaba tônica são utilizados principalmente aspectos como a presença ou a ausência de acentos ortográficos.

Na Língua Portuguesa, como todas as palavras proparoxítonas são acentuadas, não oferecem obstáculos ao processo de determinação da sílaba tônica. No entanto, as dificuldades surgem quando há a necessidade de determinar quando

uma palavra não acentuada é oxítona ou paroxítona, principalmente quando ocorre encontro vocálico na última ou penúltima sílaba.

Durante a montagem do Dicionário, havendo informações a respeito da separação de encontros vocálicos, a determinação da sílaba tônica tornar-se-á simples. No entanto, durante a transcrição de uma palavra inexistente no Dicionário, surge a possibilidade do aparecimento de incorreções. Assim por exemplo, se o Dicionário não contiver a palavra *raínha*, sua divisão corresponderá à *raí-nha* e a sílaba tônica erroneamente determinada como *raí*. E embora essa incorreção nem sempre leve a erros de transcrição fonológica, pode acarretar problemas subsequentes de implementações de prosódia.

As regras utilizadas para a determinação da sílaba tônica funcionam corretamente na grande maioria dos casos, e devendo ser aplicadas na seguinte ordem:

1 - Se a palavra contiver alguma letra acentuada com acento agudo (') ou circunflexo (^), então a sílaba tônica corresponde à que contém essa letra.

2 - Se a palavra contiver *ã* ou *õ*, então a sílaba tônica corresponde à que contém essa letra (CUNHA e CINTRA, 1985).

3 - Palavras não acentuadas terminadas em *-i, -is, -u, -us, -um, -uns, -l, -n, -ns, -r, -x, -ps* são oxítonas. Essa regra fundamenta-se no fato das palavras paroxítonas com essas terminações serem necessariamente acentuadas (CEGALLA, 1977; CUNHA e CINTRA, 1985).

4 - Palavras não acentuadas terminadas em ditongos orais são oxítonas (CUNHA e CINTRA, 1985).

5 - Palavras não acentuadas terminadas em vogal seguida de *z* são oxítonas (EGASHIRA, 1992).

6 - Se a palavra não se encaixar em nenhuma das regras anteriores, então será considerada paroxítona.

A determinação de acentuação secundária será possível apenas quando houver indicação adicional no Vocabulário Ortográfico, conforme descrito no item 4.2.1. Essa informação será simplesmente superposta à transcrição após a aplicação das regras.

### **4.3.3 - Aplicação das Regras Adotadas**

As regras para transcrição fonológicas são aplicadas em função do contexto, ou vizinhança da letra considerada, e são armazenadas no arquivo REGRAS.CON, que é um arquivo ASCII e que pode ser facilmente alterado através da maioria dos editores de texto comerciais.

O arquivo REGRAS.CON é lido pelo programa CONTEXFO no início de sua execução e suas informações são utilizadas para obter os fonemas formadores das palavras. O formato do arquivo REGRAS.CON e a sua utilização pelo programa CONTEXFO estão descritos a seguir.



#### 4.3.3.1 - Arquivo REGRAS.CON

Para a descrição das regras do arquivo REGRAS.CON foram utilizados alguns caracteres especiais com a finalidade de simplificar as regras a serem aplicadas. Assim, o símbolo # serve como *delimitador de palavras*; o símbolo \$ como *indicador de qualquer uma das vogais*; o símbolo \* como *indicador de qualquer uma das consoantes* e o símbolo % como *indicador das consoantes nasais (m ou n)*.

As linhas iniciadas por ponto-e-vírgula (;) são consideradas comentários, e linhas em branco não são consideradas. Para cada letra deve existir um conjunto de regras separadas por barra (/), no seguinte formato:

```
/letra, fone_default, incremento  
regra 1  
...  
regra n  
/
```

sendo que:

- *letra* corresponde a letra a ser convertida;
- *fone\_default* é o fonema que deve ser substituído caso o contexto não se encaixe em nenhuma das regras descritas abaixo, e
- *incremento* é a quantidade que deve ser adicionada ao indexador de letras da palavra. Quando *incremento* = 1, significa que se deve analisar a próxima letra da palavra; *incremento* = 2, significa que se deve "pular" a próxima letra e considerar a seguinte.

No caso da letra "b", por exemplo, não há regras especiais de conversão, tornando essa definição como simplesmente:

/b,b,1  
/

Caso haja alguma regra de conversão, sua especificação deve ser feita após a primeira linha contendo barra (/), no seguinte formato:

*carac\_anterior, carac\_posterior, fone, incremento, tônica*

onde:

- *carac\_anterior* são os caracteres (ou caracter) que precedem a letra considerada;
- *carac\_posterior* são os caracteres (ou caracter) que sucedem a letra considerada;
- *fone* é transcrição fonológica, constituída por apenas um caracter, a ser aplicada caso os caracteres que precedem e sucedem a letra sejam os mesmos de *carac-anterior* e *carac\_posterior*;
- *incremento* é definido de maneira análoga à do *fone\_default*, e
- *tônica* indica a necessidade da aplicação da regra, no caso de letra pertencente à sílaba tônica.

Para a letra "c", por exemplo, essas regras seriam:

/c,k,1  
,e,s,1,  
,i,s,1,  
,a,k,1,  
,o,k,1,  
,u,k,1,  
,h,x,2,  
/

Neste caso não há contexto anterior à letra "c". Havendo as letras "e" ou "i" após a letra "c", sua transcrição corresponderá à /s/; caso seja seguida por "a", "o" ou "u", a transcrição corresponderá à /k/, e se for seguida por "h", será transcrita para /x/, e a letra "h" será "pulada" (incremento = 2). Não há exigências para que essas letras façam parte da sílaba tônica, em nenhuma dessas regras.

Para a letra "a", a terceira regra indicada abaixo apenas deverá ser aplicada quando esta letra pertencer à sílaba tônica:

/a,a,1  
,%\*,ã,2,  
,%#,Ã,2,  
,%,ã,2,t  
/

Conforme essa regra, o "a" seguido de consoante nasal deve ser transcrito para /ã/ quando pertencer à sílaba tônica. Deverá igualmente ser transcrito para /ã/ quando seguido de consoante nasal e depois por consoante (1ª. regra), e para /Ã/, representando /ãu/, quando seguido por consoante nasal em final de palavra (2ª. regra).

As demais regras utilizadas estão descritas no Anexo A, utilizando-se o alfabeto fonético de um carácter apresentado na Tabela 2.1, com os seguintes caracteres adicionais:

- a letra "E", para designar o encontro vocálico "ei" que ocorre em palavras terminadas por "em", como na palavra *sem* (/seim/), transcrita como /sEm/.
- a letra "Ã", para designar o encontro vocálico "ãu" que ocorre em palavras terminadas por "am", como na palavra *falam* (/falãu/), transcrita como /falÃ/.
- a letra "K", para designar a transcrição /ks/ para a letra "x", que ocorre em palavras como *fixo* (/fikso/), transcrita como /fiKo/.



#### **4.3.3.2. - Aplicação de Regras na Transcrição de Palavras**

A aplicação de regras na transcrição fonológica das palavras deverá ocorrer sempre que a palavra não for encontrada no Dicionário. Nesse caso, não havendo nenhuma informação adicional a ser utilizada, a transcrição fonológica da palavra será feita exclusivamente através das regras do arquivo REGRAS.CON, e sendo portanto, o índice de acerto dependente da qualidade do conjunto de regras desenvolvido.

#### **4.2.3.3 - Aplicação de Regras na Montagem do Dicionário**

Na montagem do Dicionário, as regras do arquivo REGRAS.CON devem ser utilizadas conjuntamente às informações presentes no arquivo de entrada, que pode fornecer transcrições de algumas letras. Assim, palavras que constituem exceções às regras necessariamente devem ser acrescentadas ao Dicionário.

Primeiro o programa CONTEXFO procede a transcrição fonológica das palavras através das regras e depois superpõem as informações adicionais do arquivo de entrada.

#### **4.4 - ANÁLISE DOS RESULTADOS**

Para a análise dos resultados, foram abordados separadamente dois elementos do sistema de conversão de palavras para fonemas, o conjunto de regras de

conversão de letras para fonemas e a estrutura de armazenamento de dados do Dicionário.

#### 4.4.1 - Análise dos Resultados da Aplicação do Conjunto de Regras para Conversão de Letras para Fonemas

Foram escolhidos cinco trechos extraídos de fontes com diferentes gêneros literários, de forma a submeter palavras diversas à aplicação do conjunto de regras para conversão de letras para fonemas. Os textos selecionados e a transcrição gerada pelo programa CONTEXFO estão indicados no Anexo B. Convém lembrar que nesses textos foram verificadas as transcrições ao nível de palavra, não sendo considerada a questão da coarticulação entre palavras. A análise das transcrições fonológicas geradas está indicada na Tabela 4.6.

TABELA 4.6 - Análise das transcrições fonológicas.

Texto	Total de Palavras	Palavras Diferentes	Palavras com Erros de Transcrição Fonológica	Palavras Diferentes com Erros de Transcrição Fonológica	Acertos na Transcrição Fonológica de Palavras Diferentes (%)
1	114	75	6	5	93,3
2	125	76	10	7	90,9
3	147	103	11	11	89,3
4	152	98	12	8	91,8
5	173	122	14	12	90,3

Conforme indica a Tabela 4.6 o índice de acertos nas transcrições fonológicas variou na faixa de 89,3% a 93,3%.

Ocorreram para os textos 1, 2, 3 e 4, respectivamente, 5, 8, 6 e 5 erros na transcrição fonológica da letra "e" em posição tônica não acentuada, e 6, 2, 5 e 7 erros na transcrição da letra "o" para essa mesma posição.

No Texto 5, dos 14 erros que ocorreram, 8 referiam-se a transcrição das letras "e" em posição tônica não acentuada, 6 erros referiam-se à letra "o" nessa posição, 1 referia-se à transcrição da letra "x", e 1 à transcrição incorreta de um nome próprio (Drummond).

Para a eliminação desses erros poderiam ser incluídas no Dicionário de exceções, as palavras que falharam à aplicação das regras. Outra opção possível seria a criação de regras específicas para algumas palavras, conforme sugere AINSWORTH (1973). Por exemplo, para a palavra *ela* poderia ser criada uma regra, na qual a letra "e" em início de palavra, seguida por (la#), seria transcrita como /é/. Essa regra poderia ser estendida para outras palavras terminandas em *ela*, como *panela*, *amarela*, *sardela*. No entanto, seria necessário a realização de um estudo estatístico sobre a ocorrência dessa terminação com a finalidade de verificar a possibilidade de tal extensão.

Nos Textos 2 e 5 ocorreram, respectivamente, 1 e 2 erros na determinação da sílaba tônica de algumas palavras, porém, não chegaram a interferir na sua transcrição fonológica. Esses erros foram causados pela não separação de hiatos, nas palavras *poe-mas*, *co-ti-dia-no* e *toa-lha*.

#### 4.4.2 - Análise dos Resultados dos Testes de Estrutura de Armazenamento do Dicionário

A estrutura utilizada para o armazenamento do Dicionário foi testada sob dois aspectos, a quantidade de memória necessária para o armazenamento e a eficiência na busca de palavras.

Para a análise da quantidade de memória necessária para o armazenamento, foram criados cinco arquivos com 50, 100, 500, 1000 e 2000 palavras, contendo apenas palavras iniciadas com a letra "a" do Vocabulário Ortográfico. A partir desses arquivos, foram criadas as estruturas para 5 dicionários com as palavras de cada um dos arquivos, através da opção 2 - *Inclui Palavras de um Arquivo* do Menu de *Montagem do Dicionário* do programa CONTEXFO

O número de estados e de transições e a quantidade de memória necessária em cada um dos reconhecedores gerados, estão indicadas na Tabela 4.7. Os valores iniciais referem-se aos reconhecedores gerados na forma de "árvores de letras", e os valores finais referem-se aos valores obtidos após a minimização. A Tabela 4.7 apresenta também, o tamanho do arquivo original, no formato do Vocabulário Ortográfico.

TABELA 4.7 - Características dos Arquivos.

Arquivo	Número de Palavras	Tamanho do Arquivo (bytes)	Número de Estados Iniciais	Número de Transições Iniciais	Memória Inicial Necessária (bytes)	Número de Estados Finais	Número de Transições Finais	Memória Final Necessária (bytes)
va50	50	896	160	159	1.596	97	132	1.110
va100	100	1.806	312	311	3.116	154	232	1.852

va500	500	8.887	1.492	1.491	14.916	457	820	6.022
va1000	1.000	18.297	3.131	3.130	31.306	912	1.625	11.972
va2000	2.000	36.858	6.395	6394	63.946	1.696	3.195	22.956

Comparando o tamanho de cada arquivo com a quantidade de memória necessária ao armazenamento da estrutura de dados inicial, pode-se observar que o reconhecedor inicial necessitou de uma quantidade maior de memória em relação ao arquivo de entrada. A estrutura final de armazenamento do reconhecedor, por sua vez apresentou significativa economia de memória para os arquivos com mais de 500 palavras, conforme verifica-se na Figura 4.8.

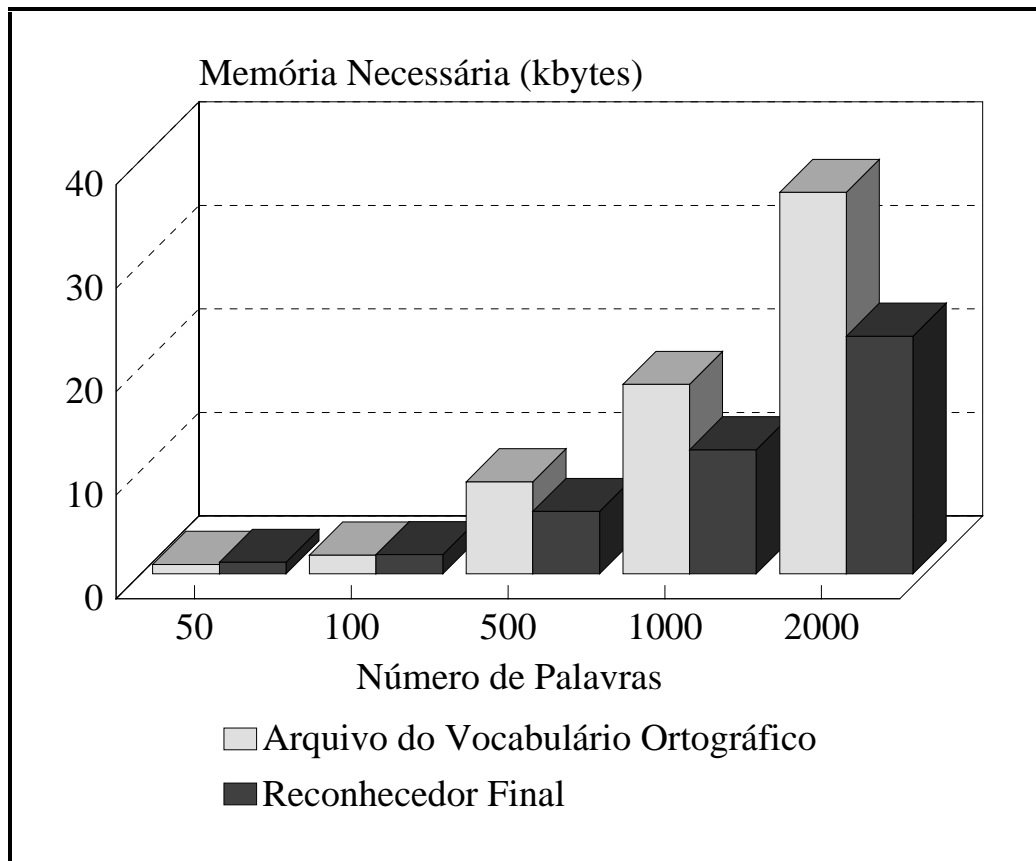


FIGURA 4.8 - Comparação da quantidade de memória necessária ao armazenamento de um arquivo no formato do vocabulário ortográfico e da quantidade de memória para armazenar a mesma informação na forma final do Dicionário de exceções.

Através da Figura 4.9 observa-se ainda que o número médio de bytes por palavra, em cada um dos reconhecedores gerados, obtidos a partir da divisão da quantidade de memória necessária para o reconhecedor final pelo número de palavras armazenadas, diminui com o aumento de palavras no Dicionário. A tendência assintótica da curva em torno de 11 bytes por palavra, incluindo a transcrição fonológica e a classe gramatical das mesmas, mostra que esta estrutura é adequada para o armazenamento dessas informações.

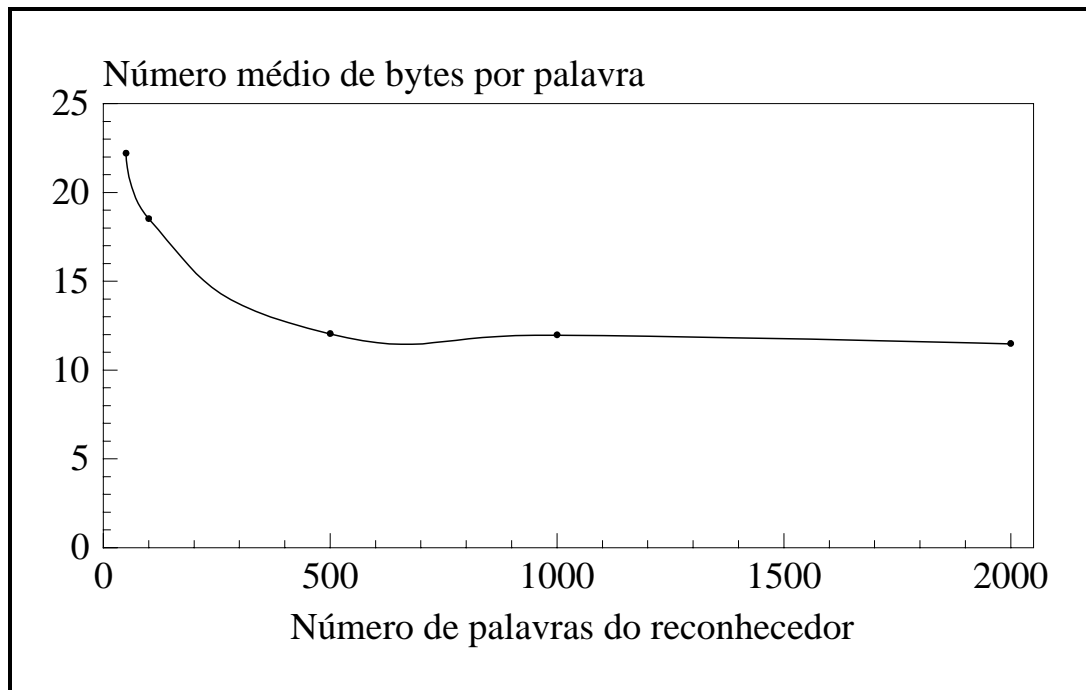


FIGURA 4.9 - Número médio de bytes por palavra, em função do número de palavras incluídas no reconhecedor.

Na Figura 4.10 observa-se crescimento do número de estados e transições praticamente linear com a inclusão de palavras no reconhecedor, quando inclusas em ordem alfabética. Os coeficientes de correlação linear do número de estados e transições em função do número de palavras do dicionário foram respectivamente de 0,9980 e 0,9992.

Esse aspecto também foi relatado por LUCCHESI e KOWALTOWSKI (1991), que observaram crescimento quase linear na quantidade de memória necessária para o armazenamento do autômato usado na construção de dicionários para as Línguas Inglesa e Portuguesa, quando as palavras eram incluídas em ordem alfabética.

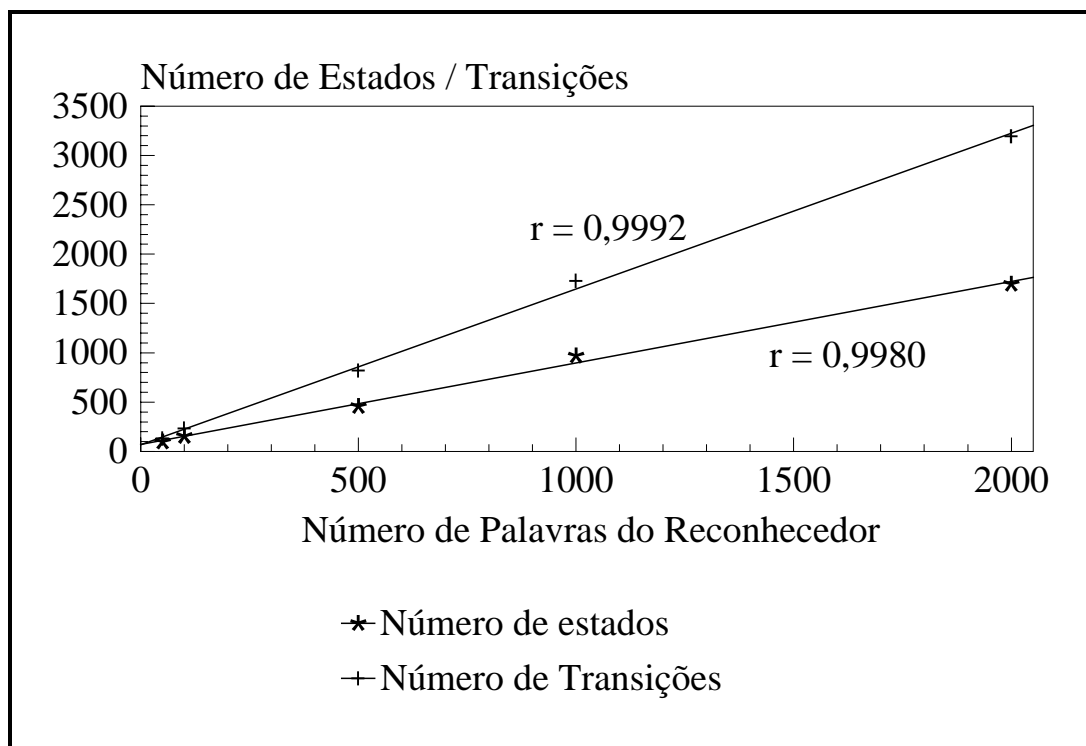


FIGURA 4.10 - Número de estados e transições em função do número de palavras incluídas no reconhecedor.

Com a finalidade de verificar a eficiência da obtenção de transcrições de palavras pertencentes ao Dicionário, foi desenvolvida uma rotina especial de busca de palavras contidas em um arquivo. Essa rotina foi utilizada para procurar todas as palavras do arquivo va2000.doc na estrutura previamente obtida a partir desse mesmo arquivo. O tempo necessário para obter a transcrição das 2.000 palavras em um microcomputador PC-AT 386 DX-40 MHz foi de 6,22 segundos, resultando uma média de 321 palavras por segundo. Levando-se em conta que, o DECTalk pode "falar" a uma taxa máxima de 550 palavras por minuto, o que resultaria numa "fala" com inteligibilidade bastante prejudicada, a taxa média obtida de 321 palavras por segundo para busca de palavras é bastante aceitável.

## **CAPÍTULO V**

### **CONSIDERAÇÕES FINAIS**

#### **5.1 - CONCLUSÕES**

Em sistemas de síntese de voz a partir de texto, a conversão de textos em fonemas desempenha papel fundamental, considerando-se que as frases de saída são construídas a partir das menores unidades sonoras formadoras das palavras. A cada uma dessas unidades sonoras mínimas corresponde uma unidade mental, ou seja, o fonema.

De certo, outras variáveis envolvidas no processo da conversão de textos para voz influenciam a qualidade final da voz gerada. No entanto, a correta determinação

dos fonemas formadores das palavras corresponde a um ponto de grande interesse, tendo em vista a sua essencialidade na transmissão de determinada mensagem.

Assim sendo, espera-se com este trabalho ter contribuído para o desenvolvimento de um sistema de conversão de textos em fonemas na Língua Portuguesa, que venha a ser utilizado em processos de síntese de voz a partir de textos, como parte integrante de um conjunto de etapas mais amplo.

A conversão de texto para fonemas na Língua Portuguesa, da mesma forma que na Língua Inglesa, não pode ser obtida apenas através de um conjunto de regras, necessitando, portanto, de um Dicionário de exceções para aquelas palavras que falham à aplicação das regras. Em muitos casos, há ainda a necessidade uma análise sintática para a determinação da correta transcrição fonológica das palavras.

O conjunto de regras desenvolvido neste trabalho, alcançou um índice de acertos da ordem de 90%, quando aplicado a palavras de textos com diversos estilos literários e extraídos de fontes diferentes.

A estrutura elaborada para o armazenamento do Dicionário apresentou-se igualmente adequada, sendo necessário em média, um pouco mais que 11 bytes por palavra para o armazenamento de um Dicionário de 2.000 entradas, e com uma velocidade de busca superior a 300 palavras por segundo em um micro PC-AT 386 DX-40 MHz.

Cabe ressaltar, como conclusão final, alguns pontos referentes ao sistema desenvolvido. Assim, além do bom desempenho do sistema, a possibilidade de aprimoramento das regras utilizadas sem a necessidade da alteração do programa, e a facilidade de inclusão de palavras no Dicionário de exceções constituem são aspectos de destaque.

## **5.2 - SUGESTÕES PARA FUTUROS ESTUDOS**

Levando-se em consideração alguns aspectos apontados neste trabalho de pesquisa, e temas ainda não explorados na Literatura, conclui-se que alguns pontos merecem consideração especial, sobre os quais serão feitas algumas sugestões para futuros estudos:

- Realização de um estudo estatístico para a preparação de um conjunto de regras com maior abrangência, propiciando melhores resultados na conversão fonética de palavras através de regras.
- Implementação de novas regras que possibilitem tratar fenômenos de coarticulação existentes nas frases, como as reduções sofridas pelas vogais "e" e "o", quando átonas, e em final de palavras.
- Realização de um levantamento, a partir de um corpus da Língua Portuguesa, das palavras mais utilizadas, para a elaboração de um dicionário de exceções completo.
- Utilização da estrutura proposta para o armazenamento do dicionário para a execução de procedimentos de análise sintática, permitindo a correta pronúncia de palavras homógrafas heterófonas presentes nos textos.
- Desenvolvimento de um pré-processador de texto, para o tratamento de siglas, números e abreviaturas.

## ANEXO A. REGRAS DE CONVERSÃO DE LETRAS EM FONEMAS DO ARQUIVO REGRAS.CON

; Simbolos utilizados

;

; # - delimitador de palavra

; \$ - vogais

; \* - consoante

; % - m ou n

;

/a,a,1

,%\*,ã,2,

,%#,Ã,2,

,%,ã,2,t

/

/b,b,1

/

/c,k,1

,e,s,1,

,ê,s,1,

,é,s,1,

,i,s,1,

,í,s,1,

,a,k,1,

,o,k,1,

,u,k,1,

,h,x,2,

/

/d,d,1

/

/e,e,1

,%#,E,1,

/

/f,f,1

/

/g,g,1

,ue,g,2,

,uê,g,2,

,ui,g,2,  
,i,j,1,  
,e,j,1,  
,ü,g,1,  
/

/h,,1  
#,,,1,  
#,,1,  
/

/i,i,1  
/

/j,j,1  
/

/k,k,1  
/

/l,l,1  
,h,L,2,  
,\*,|,1,  
,#,|,1,  
/

/m,m,1  
/

/n,n,1  
,h,N,2,  
/

/o,o,1  
,so#,o,1,  
,sa#,ó,1,  
/

/p,p,1  
/

/q,k,1  
,ua,k,1,  
,uâ,k,1  
,ue,k,2,  
,uê,k,2  
,ui,k,2,

,uo,k,1,  
,uô,k,1  
,ü,k,1  
/

/r,r,1  
#,,R,1,  
,r,,1,  
r,,R,1,  
,\*,h,1,  
#,h,1,  
/

/s,s,1  
#,,s,1,  
#,s,1,  
,ce,,1,  
,ci,,1,  
,ç,,1,  
\$,z,z,1,  
,s,,1,  
s,,s,1,  
\$,\*,s,1,  
#ex,\$,z,1,  
\*,\$,s,1,  
/

/t,t,1  
,h,t,2,  
/

/u,u,1  
/

/v,v,1  
/

/w,w,1  
/

/x,x,1  
#,,x,1,  
,ce,,1,  
,ci,,1,  
n,,x,1,  
ai,,x,1,  
ei,,x,1,

ou,,x,1,  
#e,\$,z,1,  
#e,s\$,,1,  
\*,s,1,  
/

/y,i,1  
/

/z,z,1  
#,,z,1,  
,#,s,1,  
\*,s,1,  
/

/ã,ã,1  
/

/õ,õ,1  
/

/á,a,1  
/

/é,é,1  
/

/í,i,1  
/

/ó,ó,1  
/

/ú,u,1  
/

/â,a,1  
,%ã,1,t  
/

/ê,e,1  
/

/î,i,1  
/

/ô,o,1

/

/û,u,1

/

/ç,s,1

/

/ü,u,1

/

## ANEXO B. PALAVRAS SUBMETIDAS AOS ALGORITMOS DE SEPARAÇÃO DE SÍLABAS E DETERMINAÇÃO DA SÍLABA TÔNICA

- As sílabas tônicas determinadas pelo programa utilizado estão indicadas com apóstrofe.

- Palavras indicadas com <sup>(1)</sup> apresentaram incorreção na separação silábica devido à não separação de vogais, com conseqüente erro na determinação da sílaba tônica.

- Segundo CUNHA e CINTRA (1985), alguns encontros consonantais como *gn*, *mn*, *pn*, *ps*, *pt*, *tm* entre outros, são naturalmente inseparáveis quando iniciais (*gno*-mo, *mne*-mô-ni-co, *pneu*-má-ti-co, *psi*-có-lo-go, *pti*-a-li-na, *tme*-se). No entanto, quando mediais, podem ser articulados em uma só sílaba ou em sílabas separadas (*a-pto* ou *ap-to*, *di-gno* ou *dig-no*, *ri-tmo* ou *rit-mo*). Assim, neste trabalho por questões de simplificação, esses encontros consonantais foram considerados como inseparáveis em ambos os casos, e palavras que os apresentam tem a indicação <sup>(2)</sup>.

aaquênio	a-a-'quê-nio
ababelação	a-ba-be-la-'ção
abdicar	ab-di-'car
abluir	a-blu-'ir
abrolhos	a-'bro-lhos
abscesso	abs-'ces-so
abscissa	abs-'cis-sa
absoluto	ab-so-'lu-to
abstração	abs-tra-'ção
abstrair	abs-tra-'ir
acre	'a-cre
aditivo	a-di-'ti-vo
adjetivo	ad-je-'ti-vo
adnominal	ad-no-mi-'nal
adstringente	ads-trin-'gen-te
advir	ad-'vir
afta	af-ta
aglutinar	a-glu-ti-'nar
agüem	'a-güem
agüentar	a-güen-'tar
álbum	'ál-bum
albuns	'ál-buns

alfaiate	al-ˈfaia-te (1)
algum	al-ˈgum
alguns	al-ˈguns
apto	ˈa-pto (2)
arfar	ar-ˈfar
arlequim	ar-le-ˈquim
arquivo	ar-ˈqui-vo
arsênio	ar-ˈsê-nio
atlas	ˈa-tlas
atrás	a-ˈtrás
aurora	au-ˈro-ra
banana	ba-ˈna-na
banco	ˈban-co
bíceps	bí-ceps
bisneto	bis-ˈne-to
bloco	ˈblo-co
boba	ˈbo-ba
branco	ˈbran-co
caíeis	ca-ˈí-eis
cama	ˈca-ma
cárie	ˈcá-rie
carro	ˈcar-ro
casa	ˈca-sa
castanha	cas-ˈtan-ha
chinelo	chi-ˈne-lo
claro	ˈcla-ro
cognome	co-ˈgno-me (2)
concreto	con-ˈcre-to
conseqüência	con-se-ˈqüên-cia
contíguo	con-ˈtí-guo
conversões	con-ver-ˈsões
coordenar	co-or-de-ˈnar
copo	ˈco-po
corrente	cor-ˈren-te
cortina	cor-ˈti-na
cravo	ˈcra-vo
cresça	ˈcres-ça
demais	de-ˈmais
descer	des-ˈcer
dicionário	di-cio-ˈná-rio
digno	ˈdi-gno (2)
diskete	dis-ˈke-te
disquete	dis-ˈque-te
dragão	dra-ˈgão
ealmar	e-al-ˈmar
eambose	e-am-ˈbo-se
eautognosia	eau-to-ˈgno-sia (1),(2)

eberthemia	e-ber-the-'mia (1)
ebriez	e-bri-'ez
ebuliente	e-bu-li-'en-te
ebulioscopia	e-bu-li-os-co-'pia (1)
écbase	'éc-ba-se
echião	e-chi-'ão
eclesiarca	e-cle-si-'ar-ca
ecoencefalograma	e-co-en-ce-fa-lo-'gra-ma
editar	e-dí-'tar
egípcio	e-'gíp-cio
empresa	em-'pre-sa
esclarecer	es-cla-re-'cer
eucaliptocarpo	eu-ca-li-ptro-'car-po (2)
exceção	ex-ce-'ção
exceder	ex-ce-der
exibir	e-xi-'bir
exsudar	ex-su-'dar
facção	fac-'ção
feira	'fei-ra
feldspato	felds-'pa-to
fenolftaleína	fe-nolf-ta-le-'í-na
fiel	fi-'el
filho	'fi-lho
flor	'flor
fonema	fo-'ne-ma
fonte	'fon-te
fórceps	'fór-ceps
formatar	for-ma-'tar
francês	fran-'cês
fuga	'fu-ga
funga	'fun-ga
glória	'gló-ria
glutão	glu-'tão
gnomo	'gno-mo (2)
grampeador	gram-pea-'dor
grande	'gran-de
gringo	'grin-go
helicóptero	he-li-'có-pte-ro (2)
heptatlo	he-'pta-tlo (2)
hotel	ho-'tel
ignição	i-gni-'ção (2)
imprensa	im-'pren-sa
informativo	in-for-ma-'ti-vo
injeção	in-je-'ção
inserir	in-se-'rir
janela	ja-'ne-la
jargão	jar-'gão

jarro	ˈjar-ro
justificar	jus-ti-fi-ˈcar
lâmpada	ˈlâm-pa-da
lenda	ˈlen-da
lhama	ˈlha-ma
lida	ˈli-da
linda	ˈlin-da
magna	ˈma-gna (2)
mágoa	ˈmá-go-a
manhã	ma-ˈnhã
manta	ˈman-ta
mário	ˈmá-ri-o
mata	ˈma-ta
mesa	ˈme-sa
micro	ˈmi-cro
miúdo	mi-ˈú-do
mnemônico	mne-ˈmô-ni-co (2)
monitor	mo-ni-ˈtor
monta	ˈmon-ta
motor	mo-ˈtor
muito	ˈmui-to
nacional	na-cio-ˈnal (1)
normal	nor-ˈmal
nostalgia	nos-tal-ˈgia (1)
occipital	oc-ci-pi-ˈtal
olhar	o-ˈlhar
palavra	pa-ˈla-vra
papel	pa-ˈpel
paraguai	pa-ra-ˈguai
partiu	par-ˈtiu
passo	ˈpas-so
perspicácia	pers-pi-ˈcá-cia
plano	ˈpla-no
pneumático	pneu-ˈmá-ti-co (2)
prato	ˈpra-to
precaução	pre-cau-ˈçã-o
precauções	pre-cau-ˈçõ-es
professor	pro-fes-ˈsor
psicólogo	psi-ˈcó-lo-go (2)
ptialina	ptia-ˈli-na (1),(2)
qüinqüelíngüe	qüin-qüe-ˈlín-güe
qüinqüênio	qüin-ˈqüê-nio
rachar	ra-ˈchar
rainha	ˈrai-nha (1)
recepção	re-cep-ˈçã-o
refrão	re-ˈfrã-o
regra	ˈre-gra

regredir	re-gre-`dir
régua	`ré-gua
remalina	re-ma-`li-na
resumo	re-`su-mo
ritmo	`ri-tmo (2)
rubro	`ru-bro
ruflar	ru-`flar
saúde	sa-`ú-de
seda	`se-da
silepse	si-`le-pse (2)
solstício	sols-`tí-cio
sopro	`so-pro
subjulgar	sub-jul-`gar
submeter	sub-me-`ter
subpor	sub-`por
subscrever	subs-cre-`ver
subtrair	sub-tra-`ir
superfluidificante	su-per-flui-di-fi-`can-te
tabela	ta-`be-la
teclado	te-`cla-do
tensão	ten-`são
tentativa	ten-ta-`ti-va
tênuê	`tê-nue
terpsícore	ter-`psí-co-re (2)
terra	`ter-ra
tetracampeão	te-tra-cam-pe-`ão
tmese	`tme-se (2)
tomada	to-`ma-da
transatlântico	tran-sa-`tlân-ti-co
transcrição	trans-cri-`ção
tribo	`tri-bo
triplo	`tri-plo
tungstênio	tungs-`tê-nio
utilitários	u-ti-li-`tá-rios
vidro	`vi-dro
voltagem	vol-`ta-gem
zangado	zan-`ga-do

## ANEXO C. TEXTOS SUBMETIDOS À TRANSCRIÇÃO FONOLÓGICA

Cinco textos foram submetidos à transcrição fonológica, sendo que as palavras que apresentaram erros de transcrição estão sublinhadas, e as palavras com erros na determinação da sílaba tônica estão em indicadas por *letras itálicas*. As sílabas tônicas das diversas palavras estão indicadas por apóstrofes, com exceção das palavras monossílabas.

### Texto 1\*

*"Mas ninguém poderia adivinhar o que ela pensava. E para aqueles que junto da porta ainda a olharam uma vez, a aniversariante era apenas o que parecia ser: sentada à cabeceira da mesa imunda, com a mão fechada sobre a toalha como encerrando um cetro, e com aquela mudez que era a sua última palavra. Com um punho fechado sobre a mesa, nunca mais ela seria apenas o que ela pensasse. Sua aparência afinal a ultrapassara e, superando-a, se agigantava serena. Cordélia olhou-a espantada. O punho mudo e severo sobre a mesa dizia para a infeliz nora que sem remédio amava talvez pela última vez. É preciso que se saiba. É preciso que se saiba. Que a vida é curta. Que a vida é curta."*

\* Trecho extraído do Conto "Feliz Aniversário" de Clarice Lispector, selecionado por Alfredo Bosi, para integrar o Livro "O Conto Brasileiro Contemporâneo", da Editora Clutrix, São Paulo.

### Texto 1 após transcrição fonológica:

"mas nin'guEm pode'ria adivi'Nah o ke 'ela pen'sava. e 'para a'keles ke 'junto da 'pohta a'inda a o'LarÃ 'uma ves, a anivehsari'ãte 'era a'penas o 'ke pare'sia seh: sen'tada a kabe'seira da 'meza i'munda, kom a mão fe'xada 'sobre a 'toaLa 'komo ense'Rãdo um 'setro, e kom a'kela mu'des ke 'era a sua 'u|tima pa'lavra. kom um 'puNo fe'xado 'sobre a 'meza, 'nunka ma'is 'ela 'seria a'penas o ke 'ela pen'sase. sua apa'rensia afi'nal| a u|trapa'sara e, supe'rãdo-a, se ajigã'tava se'rena. koh'délia o'Lou-a espã'tada. o 'puNo 'mudo e se'vero 'sobre a 'meza 'dizia 'para a infe'lis 'nora ke sEm Re'médio a'mava ta|'ves 'pela 'u|tima ves. é pre'sizo ke se 'saiba. é pre'sizo ke se 'saiba. ke a 'vida é 'kuhta. ke a 'vida é 'kuhta."

Texto 2 \*

*"As armas e os barões assinalados  
Que, da ocidental praia lusitana,  
Por mares nunca de antes navegados  
Passaram ainda além da Taprobana,  
Em perigos e guerras esforçados,  
Mais do que prometia a força humana,  
E entre gente remota edificaram  
Novo reino, que tanto sublimaram.*

*E também as memórias gloriósas  
Daqueles reis que foram dilatando  
A fé, o império, e as terras viciosas  
Da África a Ásia andaram devastando,  
E aqueles que por obras valerosas  
Se vão da lei da morte libertando:  
Cantando espalharei por toda parte,  
Se a tanto me ajudar o engenho e arte.*

*Cessem do sábio grego e do troiano  
As navegações grandes que fizeram;  
Cale-se de Alexandro e de Trajano  
A fama das vitórias que tiveram;  
Que eu canto o peito ilustre lusitano,  
A quem Netuno e Marte obedeceram.  
Cesse tudo o que a musa antiga canta,  
Que outro valor mais alto se alevanta."\**

\* O trecho apresentado corresponde à Proposição dos "Lusíadas" de Luís de Camões, analisada por Aires da Mata Machado Filho, em Livro da Série "Nossos Clássicos", da Livraria e Editora Agir, Rio de Janeiro (1978).

Texto 2 após transcrição fonológica:

"as 'ahmas e os ba'rões asina'lados  
ke, da osiden'tal| 'praia luzi'tãna,  
poh 'mares 'nunca de 'ãtes nave'gados  
pa'sarÃ a'inda a'lEm da tapro'bãna,

Em pe'rigos e 'geRas esfoh'sados,  
mais do ke prome'tia a 'fohsa u'mãna,  
e 'entre 'jente Re'mota edifi'karÃ  
'novo 'Reino, ke 'tãto subli'marÃ.

e tã'bEm as me'mórias glori'ózas  
da'keles Re'is ke 'forÃ dila'tãdo  
a fé, o im'pério, e as teRas vi'siozas  
da 'afrika a 'azia ã'darÃ devas'tãdo,  
e a'keles ke poh obras vale'rozas  
se vão da lei da mohte libeh'tãdo:  
kã'tãdo espaLa'rei poh 'toda 'pahte,  
se a 'tãto me aju'dah o en'jeNo e 'ahte.

sesEm do 'sabio 'grego e do 'troiãno  
as navega'sões 'grãdes ke fi'zerÃ;  
'kale-se de ale'xãdro e de tra'jãno  
a 'fãma das vi'tórias ke ti'verÃ;  
ke eu 'kãto o 'peito i'lustre luzi'tãno,  
a kEm ne'tuno e 'mahte obede'serÃ.  
'sese 'tudo o ke a 'muza ã'tiga 'kãta,  
ke 'outro va'loh ma'is 'a|to se ale'vãta."

### Texto 3 \*

*"A geração de voz por técnicas digitais oferece um número bem maior de opções, que se dividem em duas categorias.*

*A primeira categoria é formada pelos sistemas que apenas codificam digitalmente as formas de onda dos sinais da voz. Estes sistemas constituem apenas versões mais sofisticadas dos antigos sistemas de síntese de voz baseados em palavras pré-gravadas em meio magnético, já referido, e padecem das mesmas limitações fundamentais, embora exijam menor espaço de armazenamento (desde que utilizem técnicas mais sofisticadas, como modulação por código de pulso delta adaptativo).*

*A segunda categoria é mais importante; são os sistemas que procuram representar a voz através de um conjunto de parâmetros que resultam da análise do sinal da voz, e posteriormente reconstituem o sinal a partir destes parâmetros. Através destes esquemas temos a possibilidade de representar qualquer palavra desde que se tenha um sistema que preserve a identidade dos fonemas durante o processo de análise." \**

\* Trecho extraído da Tese do Prof. Dr. Geraldo Lino de Campos, apresentada à Escola Politécnica da Universidade de São Paulo, para a obtenção do Grau de Doutor em Engenharia, São Paulo (1980).

Texto 3 após transcrição fonológica:

"a jera'são de vos poh 'téknikas diji'tais ofe'rese um 'número bEm mai'oh de op'sões, ke se di'vidEm Em 'duas katego'rias. a pri'meira kate'goria é foh'mada 'pelos sis'temas ke a'penas kodi'fikÃ dijital'mente as 'fohmas de 'onda dos sina'is da vos. 'estes sis'temas konsti'tuEm a'penas veh'sões ma'is sofisti'kadas dos ã'tigos sis'temas de 'sintese de vos ba'zeados Em pa'lavras 'pré-gra'vadas Em 'meio ma'gnético, ja Refe'rido, e pa'desEm das 'mesmas limita'sões fundamenta'is, em'bora e'zijÃ me'noh es'paso de ahmazena'mento ('desde ke uti'lizEm 'téknikas mais sofisti'kadas, 'komo modula'são poh 'kódigo de 'pu|so delta adapta'tivo). a se'gunda kate'goria é mais impoh'tâte; são os sis'temas ke pro'kurÃ Reprezen'tah a vos atra'vés de um kon'junto de pa'râmetros ke Re'zu|tÃ da a'nalíse do si'nal da vos, e posterioh'mente Rekonsti'tuEm o si'nal a pah'tih 'destes pa'râmetros. atra'vés 'destes es'kemas 'temos a posibili'dade de Reprezen'tah kua|'keh pa'lavra 'desde ke se 'teNa um sis'tema ke pre'zehve a identi'dade dos fo'nemas du'râte o pro'seso de a'nalíse."

Texto 4 \*

*"O comportamento da empresa monopolista difere do comportamento da empresa perfeitamente competitiva à medida que, no monopólio, a empresa tem amplo domínio sobre o mercado em que opera. Não há um preço de mercado ao qual ela deva subordinar-se. Há, apenas, uma curva de procura com que ela se defronta e à qual se aplica o princípio*

*fundamental de que as quantidades procuradas aumentam à medida que os preços diminuem. Além disso, a empresa defronta-se, ainda, com os rigores das estruturas típicas de custo. Afinal, a empresa monopolista, apesar de seu poder sobre o mercado, não é imune às regras fundamentais da demanda nem aos pressupostos das economias e das deseconomias de escala." \**

\* Trecho extraído do Livro "Introdução à Economia" de José Paschoal Rossetti, Editora Atlas S.A., São Paulo (1988).

Texto 4 após transcrição fonológica:

"o kompohta'mento da em'preza monopo'lista di'fere do kompohta'mento da em'preza pehfeita'mente kompeti'tiva a me'dida ke, no mono'pólio, a em'preza tEm 'ãplo do'minio 'sobre o meh'kado Em ke o'pera. não a um 'preso de meh'kado ao 'kua| 'ela 'deva subohdi'nah-se. a, a'penas, 'uma 'kuhva de pro'kura kom ke 'ela se de'fronta e a 'kua| se a'plika o prin'kipio fundamen'ta| de ke as kuãti'dades proku'radas au'mentã a me'dida ke os 'presos dimi'nuem. a'lEm 'diso, a em'preza de'fronta-se, a'inda, kom os Ri'gores das estru'turas 'tipikas de 'kusto. afi'nal|, a em'preza monopo'lista, ape'zah de seu po'deh 'sobre o meh'kado, não é i'mune as 'Regras fundamenta'is da de'mãda nEm aos presu'postos das ekono'mias e das dezekono'mias de es'kala."

Texto 5 \*

*"De certo modo representando a própria evolução da poesia moderna brasileira, a obra poética de Carlos Drummond de Andrade percorreu várias fases ou maneiras. Não sendo possível exemplificá-las todas, convocaram-se para esta coletânea as que mais salientemente lhe marcaram a personalidade e o lugar que ocupa nas literaturas de Língua Portuguesa deste século. A uma análise pormenorizada dos poemas citados, percebe-se que os atravessam duas linhas de força capitais: de um lado, o cotidiano ou/e o humor nele implícito; de outro, a visão dum transcendental para além da superfície opaca da realidade diária. 'No Meio do Caminho', pertencente ao volume Alguma Poesia, expõe de modo flagrante as duas tendências: por detrás da epidérmica atitude de quem se compraz no jogo vocabular e sonoro com as expressões 'no meio*

*do caminho' e 'tinha uma pedra', que atenderia à inclinação cotidianista de seu estro, divisa-se a gravidade tensa de 'retinas tão fatigadas' auscultando a monotonia inexorável, a que se reduz a tragédia da própria condição humana: sempre 'no meio do caminho tinha uma pedra.' " \**

\* Trecho extraído do Livro "A Literatura Brasileira através dos Textos", de Massaud Moisés, Editora Cultrix, São Paulo (1980) sobre Carlos Drummond de Andrade, renomado poeta modernista.

Texto 5 após transcrição fonológica:

"de 'sehto 'modo Reprezen'tãdo a 'própria evolu'são da poe'zia mo'dehna brazi'leira, a 'obra po'étika de 'kahlos 'drummond de ã'drade pehko'Reu 'varias 'fazes ou ma'neiras. não 'sendo po'sive| ezemplifi'ka-las 'todas, konvo'karÃ-se 'para 'esta kole'tãnea as ke mais saliente'mente Le mah'karÃ a pehsonali'dade e o lu'gah ke o'kupa nas litera'turas de 'lingua pohtu'geza 'deste 'sékulo. a 'uma a'nalize pohmenori'zada dos 'poemas si'tados, peh'sebe-se ke os atra'vesÃ 'duas 'liNas de 'fohsa kapi'tais: de um 'lado, o *koti'diãno* ou/e o u'moh 'nele im'plisito; de 'outro, a vi'zão dum trãseden'ta| 'para a'lEm da supeh'fisie o'paka da Reali'dade di'aria. no 'meio do ka'miNo, pehten'sente ao vo'lume a|'guma poe'zia, es'põe de 'modo fla'grãte as 'duas ten'densias: poh de'tras da epi'déhmika ati'tude de kEm se kom'pras no 'jogo vokabu'lah e so'noro kom as espre'sões no 'meio do ka'miNo e 'tiNa 'uma 'pedra, ke aten'deria a inklina'são kotidia'nista de seu 'estro, di'viza-se a gravi'dade 'tensa de 'Re'tinas tão fati'gadas' ausku'tãdo a monoto'nia inexo'rave|, a ke se Re'dus a tra'gédia da 'própria kondi'são u'mãna: 'sempre no 'meio do ka'miNo 'tiNa 'uma 'pedra."

## REFERÊNCIAS BIBLIOGRÁFICAS

- AINSWORTH, W.A. A System for Converting English Text into Speech. **IEEE Transactions on Audio and Electroacoustics**, v. 21, n. 3, p. 288-90, June 1973.
- ALLIANDRO, H. **The Portuguese-English Dictionary**. 10 ed. New York, Pocket Book. 1974.
- ALLEN, J. Reading Machines for the Blind: The Technical Problems and the Methods Adopted for Their Solution. **IEEE Transactions on Audio and Electroacoustics**, v. 21, n. 3, p. 259-64, June 1973.
- ALLEN, J. Synthesis of Speech from Unrestricted Text. **Proceedings of the IEEE**, v. 64, n. 4, p. 433-42, Apr. 1987.
- AL-SUWAIYEL, M.; HOROWITZ, E. Algorithms for Trie Compaction. **ACM Transactions on Database Systems**, v. 9, n. 2, p. 243-63, June 1984.
- APPEL, A.W.; JACOBSON, G.J. The World's Fastest Scrabble Program. **Communications of the ACM**, v. 31, n. 5, p. 572-8, May 1988.
- ATAL, B.S.; RABINER, L.R. Speech Research Directions. **AT & T Technical Journal**, v. 65, n. 5, p. 75-88, Sept./Oct. 1986.
- CALLOU, D.; LEITE, Y. **Iniciação à Fonética e à Fonologia**. 1 ed. Rio de Janeiro, São Paulo, Jorge Zahar Editor Ltda. 1990.
- CAMPOS, G.L. **Síntese de Voz para o Idioma Português**. São Paulo, 1980. Tese (Doutorado) - Escola Politécnica, Universidade de São Paulo.
- CASAES, E.J. **Descrição Acústico-Articulatória dos Sons da Voz para um Modelo dos Sons do Português do Brasil**. São Paulo, 1990. Tese (Doutorado) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.
- CEGALLA, D.P. **Novíssima Gramática da Língua Portuguesa**. 16 ed. São Paulo, Companhia Editora Nacional. 1977.

- COKER,C.H.; UMEDA,N.; BROWMAN,C.P. Automatic Synthesis from Ordinary English Text. **IEEE Transactions on Audio and Electroacoustics**, v. 21, n. 3, p. 293-8, June 1973.
- CROCHIERE,R.E.; FLANAGAN,J.L. Speech Processing: an Evolving Technology. **AT & T Technical Journal**, v. 65, n. 5, p. 2-11, Sept/Oct. 1990.
- CUNHA,C.; CINTRA,L. **Nova Gramática do Português Contemporâneo**. 2 ed. Rio de Janeiro, Editora Nova Fronteira. 1985.
- EGASHIRA,F.; VIOLARO,F. **Síntese de Voz a Partir de Texto**. Campinas, Faculdade de Engenharia Elétrica da Universidade Estadual de Campinas, 1993. (Publicação FEE 01/93).
- EL-IMAN,Y.A. An Unrestricted Vocabulary Arabic Speech Synthesis System. **IEEE Transactions on Acoustics, Speech and Signal Processing**, v. 37, n.12, p.1829-45, Dec. 1989.
- ELOVITZ,H.S.; JOHNSON,R.; McHUGH,A.; SHORE,J.E. Letter-to-Sound Rules for Automatic Translation of English Text to Phonetics. **IEEE Transactions on Acoustics, Speech and Signal Processing**, v. 24, n. 6, p. 446-59, Dec. 1976.
- ENDRES,W.K. Problems of Speech Analysis and Synthesis. In: H.W. Schüssler, ed. **EURASIP Signal Processing: Theories and Applications**, 2., Erlangen, 1983. **Proceedings**. North-Holland, Elsevier Science Publishers B.V., 1983. C 1.0, p. 315-22.
- ESQUIVEL,A.S. Um Sistema de Síntese de Voz. In: Congresso Nacional de Informática, 18., São Paulo, 1985. **Anais**. São Paulo, Sucesu, 1985. p. 776-82.
- FERREIRA, A.B.H. **Novo Dicionário da Língua Portuguesa**. 2 ed. Rio de Janeiro, Editora Nova Fronteira, 1986.
- FLANAGAN,J.L. **Speech Analysis Synthesis and Perception**. 2 ed. New Jersey, Springer-Verlag, 1972.
- FLANAGAN,J.L.; COKER,C.H.; RABINER,L.R.; SCHAFER,R.W.; UMEDA, N. Synthetic Voices for Computers. **IEEE Spectrum**, v. 7, p. 22-45, Jan. 1970.

- GROSS,M. The Use of Finite Automata in the Lexical Representation of Natural Language. In: M. Gross and D. Perrin, ed. **Electronic Dictionaries and Automata in Computational Linguistics**. Berlin, Springer-Verlag, Berlin, 1989. p. 34-50 (Lectures Notes in Computer Science 377)
- HAGGARD,M.P.; MATTINGLY,I.G. A Simple Program for Synthesizing British English. **IEEE Transactions on Audio and Electroacoustics**, v.16, n. 1, p. 95-9, Mar. 1968.
- HERTZ,S.R. From Text to Speech with SRS. **Journal of the Acoustical Society of America**, v. 74, n. 4, p. 1155-70, Oct. 1982.
- \_. **The Delta Programming Language: An Integrated Approach to Non-Linear Phonology, Phonetics, and Speech Synthesis**. Ithaca, Phonetics Laboratory, 1994.
- HERTZ,S.R.; KADIN,J; KARPLUS,K.J. The Delta Rule Development System for Speech Synthesis from Text. **Proceedings of the IEEE**, v. 73, n. 11, p. 1589-601, Nov. 1985.
- HIRSCHBERG,J.B.; RIEDERER, S.A.; ROWLEY,J.E.; SYRDAL,A.K. Voice Response Systems: Technologies and Applications. **AT & T Technical Journal**, v. 65, n. 5, p. 42-51, Sept./Oct. 1990.
- JAVKIN,H. et al., A Multi-Lingual Text-to-Speech System. In: IEEE International Conference on Acoustics, Speech and Signal Processing, \_\_\_\_, \_\_\_\_\_, 1989. **Proceedings**. \_\_\_\_\_, IEEE, 1989. S 5.8. p. 242-5.
- JOSÉ NETO,J. **Introdução à Compilação**. \_\_ ed., Rio de Janeiro, Livros Técnicos e Científicos Editora S.A., 1987.
- KAPLAN,G.; LERNER,E.J. Realism in Synthetic Speech. **IEEE Spectrum**, v. 22, p. 32-7, Apr. 1985.
- KLATT,D.H. Linguistics Uses of Segmental Duration in English: Acoustics and Perceptual Evidence. **Journal of the Acoustical Society of America**, v. 59, n. 5, p. 1208-21, May 1976.

- \_. Structure of a Phonological Rule Component for a Synthesis-by-Rule Program. **IEEE Transactions on Acoustics, Speech and Signal Processing**, v. 24, n. 5, p. 391-8, Oct. 1976.
- \_. Review of Text-to-Speech Conversion for English. **Journal of Acoustical Society of America**, v. 82, n. 3, p. 737-93, Sept. 1987.
- KLATT,D.H.; KLATT,L.C. Analysis, Synthesis, and Perception of Voice Quality Variations Among Female and Male Talkers. **Journal of Acoustical Society of America**, v. 87, n. 2, p. 820-56, Feb. 1990.
- LAPORTE,E. Applications of Phonetic Description. In: M. Gross and D. Perrin, ed. **Electronic Dictionaries and Automata in Computational Linguistics**. Berlin, Springer-Verlag, Berlin, 1989. p. 65-78 (Lectures Notes in Computer Science 377)
- LEE,L.S.; TSENG,C.H.; OUH-YOUNG,M. The Synthesis Rules in a Chinese Text-to-Speech System. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 37, n. 9, p. 1309-20, Sept. 1989.
- LUCCHESI,C.L.; KOWALTOWSKI,T. **Applications of Finite Automata Representing Large Vocabularies**. Campinas, Departamento de Ciência de Computação, 1991.
- OLABE,J.C.; SANTOS,A.; MARTÍNEZ,R.; MUÑOZ,E.; MARTÍNEZ, M.; QUILIS,A.; BERNSTEIN,J. Real Time Text to Speech Conversion System for Spanish. In: H.W. Schüssler, ed. **EURASIP Signal Processing: Theories and Applications**, 2., Erlangen, 1983. **Proceedings**. North-Holland, Elsevier Science Publishers B.V., 1983. C 1.3, p. 331-3.
- OLIVEIRA,L.C.; VIANA,M.C.; TRANCOSO,I.M. **A Rule-Based Text-to-Speech System for Portuguese**. Lisboa, INESC/IST/CLUL, 1994.
- O'MALLEY,M.H. Text-To-Speech Conversion Technology. **IEEE Computer**. v. \_\_\_, n.\_\_\_, p. 17-23, Aug. 1990.
- PAIS,C.D. Elementos de Fonologia Estrutural. In: **Manual de Lingüística**. Petrópolis, Editora Vozes Ltda, 1978, p. 9-80.

- SANCHES,I. **Reconhecedor de Dígitos Isolados Independente do Locutor**. São Paulo, 1989. Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo.
- UMEDA,N. Linguistic Rules for Text-to-Speech Synthesis. **Proceedings of the IEEE**, v. 64, n. 4, p. 443-51, Apr. 1976.
- VAN COILE,B.M.J. The DEPES Development System for Text-to-Speech Synthesis. In: IEEE International Conference on Acoustics, Speech and Signal Processing, \_\_., \_\_\_\_\_, 1989. **Proceedings**. \_\_\_\_\_, IEEE, 1989. S 5.10. p. 250-3, 1989.
- VIANA,M.C.; ANDRADE,E; OLIVEIRA,L.C.;TRANCOSO,I.M. Ler\_PE: Um Utensílio para o Estudo da Ortografia do Português. In: Encontro da Associação Portuguesa Lingüística, 7., Lisboa, 1991. **Anais**. Lisboa, APL, 1991. p. 474-89.
- VIOLARO,F. Panorama de Investigações em Processamento de Fala no Brasil. In: Encontro de Processamento da Língua Portuguesa Escrita e Falada, 1., Lisboa, 1993. **Anais**. Lisboa, \_\_\_\_\_, 1993. p. 183-93.
- YOUNG,S.J.; FALLSIDE,F. Speech Synthesis from Concept: A Method for Speech Output from Information Systems. **Journal of Acoustical Society of America**, v. 66, n. 3, p. 685-95, Sept. 1989.

