Universidade Católica Dom Bosco
Centro de Ciências Exatas e Tecnológicas
Curso de Engenharia de Computação

Utilização de Autômatos Adaptativos para Tradução Texto-Voz

Felipe Augusto Zuffo

Prof. Orientador: Dr. Hemerson Pistori

Relatório Final submetido como um dos requisitos para a obtenção do grau de Engenheiro de Computação.

Campo Grande - MS - DEZEMBRO/2004

Resumo

Nos últimos anos, a síntese de voz vem apresentando-se como solução para interface homen/máquina em sistemas computadorizados. Com isso, várias técnicas para síntese de voz vem sendo desenvolvidas, no intuito de que os sinais sonoros gerados pelos sintetizadores sejam o mais semelhante possível aos da fala humana. Este trabalho é um estudo sobre a utilização de autômatos adaptativos para construção de tradutor texto-voz, beneficiando-se da possibilidade de alteração na sua estrutura em tempo de execução ao reconhecer cadeias. Essa característica oferece a possibilidade de distinção de fonemas, quando sílabas ou palavras possuem mesma sintaxe e geram pronúncia diferentes, dependentemente do contexto.

Abstract

In the last years, voice synthesis is coming as a solution for interface man/machine in computerized systems. So, several techniques for voice synthesis have been developed, in the intention that the sound signs generated by the synthesizers are more similar to human speech. This work is a study about the use of adaptive automata for construction of text-voice translator, benefitting from the structure's alteration possibility in time of execution when recognizing strings. That characteristic offers the possibility of distinction of phonemes, when syllables or words possess same syntax and they generate different speech, dependent on the context.

Conteúdo

Capítulo 1

Introdução

O crescimento da utilização de sistemas computadorizados aumenta a necessidade de comunicação ágil entre máquinas e seres humanos. A síntese de voz representa uma solução prática em diversas aplicações como, por exemplo, em sistemas onde a comunicação visual não é possível, ou ainda, em aplicações que exigem agilidade e automação na interação com o usuário.

Síntese de voz é a geração de sinais sonoros que reproduzem as palavras equivalentes em uma linguagem determinada[1]. As aplicações que sintetizam a voz são projetadas para imitar a fala, assim como na natureza humana. Um bom exemplo, são sistemas que possibilitam a utilização de computadores por deficientes visuais, onde a interação do software com o usuário é feita através do sistema de áudio. Aplicativos com síntese de voz são utilizados também em telefonia, onde a automação do atendimento facilita a navegação dos usuários pelo sistema.

Um sistema de síntese de voz pode utilizar-se de várias técnicas para gerar pronúncia. Porém, de modo geral, podemos dividir o processo de tradução de texto para voz em duas etapas: o reconhecimento de fonemas, partindo de um texto de entrada, e a execução da pronúncia. A estrutura utilizada para reconhececimento dos fonemas deve ser dinâmica, pois a pronúncia de palavras, na maioria das línguas, depende da palavra como um todo, não só da sílaba. Este trabalho apresenta os estudos sobre autômatos adaptativos como proposta de solução para esse caso.

Um autômato adaptativo consiste em um autômato de estados finitos, unido a um dispositivo adaptativo. Esse dispositivo age sobre o autômato de forma a alterar sua estrutura dinamicamente, podendo inserir e remover estados, ou ainda modificar as transições entre os mesmos, conforme a cadeia de entrada [2]. Devido a essa característica, essas máquinas têm um grande poder de representação, contribuindo no desenvolvimento de projetos de aplicações complexas que precisam tomar decisões para solução de proble-

mas [3, 4].

Nos últimos anos, a comunidade científica começou a apresentar trabalhos formais relacionados a tecnologias adaptativas [5, 6]. A partir de então, várias vertentes se abrem para a utilização dessas tecnologias comercialmente e, sobretudo, no meio científico. De um modo especial, esse trabalho contempla estudos sobre autômatos adaptativos e, posteriormente, a aplicação dos mesmos em tradução texto-voz.

O próximo capítulo descreve os fundamentos sobre os quais esse trabalho será executado. A primeira seção descreve o funcionamento e aplicações para sistemas de síntese de voz. A seção 2.2 descreve autômatos adaptativos, bem como suas aplicações. Já a seção 2.3, relata sobre a ferramenta que será utilizada para aplicar experimentos referentes a utilização de Autômatos Adaptativos para tradução texto-voz. O capítulo 3 descreve o contexto para o qual esse trabalho vem apresentar contribuições. O capítulo 4 relata sobre fonética, área intimamente ligada a este trabalho, pois dela vem as regras para a qualidade almejada para a fala gerada. No capítulo 5 está descrita a pesquisa realizada. Por fim, o capítulo 6 consta dos resultados obtidos com esse trabalho, bem como as contribuições deixadas para trabalhos futuros.

1.1 Justificativa

Ultimamente, a síntese de voz vem sendo utilizada em uma vasta gama de aplicações. E de modo mais específico, a tradução texto-voz possibilita que a entrada para os sintetizadores de voz seja um padrão conhecido e de fácil entendimento. Outro sim, aplicações que utilizam sintetizadores de voz a partir de texto visam oferecer maior flexibilidade na formação de mensagens, sendo limitada apenas pelas regras da própria linguagem.

A implementação desse tipo de sintetizador pode ter várias abordagens, implementando técnicas diferentes. Entre essas técnicas existem alguns problemas comuns, como pré-processamento de texto (normalização), representação gráfica diferente para sons iguais, exceções às regras de pronúncia, tratamento de erros, etc. A rapidez da geração da fala nas aplicações está diretamente ligada ao tempo gasto na solução desses problemas. Tendo isso em vista, a utilização de autômatos adaptativos como estrutura para pronúncia de textos mostra-se como solução para alguns desses problemas.

A característica fundamental que define os autômatos adaptavivos está no fato de poderem alterar sua estrutura em tempo de execução, ou seja, alterar os estados e transições no momento em que reconhece a entrada. Algumas sílabas, em palavras da lingua portuguêsa, podem gerar diferentes fonemas, dependendo de outras sílabas na mesma palavra. Por exemplo, nas

palavras *bela* e *beleza* a sílaba *be* assume diferentes pronúncias, dependentemente da presença da sílaba *le*. Nesse caso, o autômato adaptativo assumiria como padrão a pronúncia aberta (som do acento agudo) até que a próxima sílaba fosse reconhecida. Então, através das ações adaptativas, a estrutura do autômato é alterada para que a sílaba *be* gere a pronúncia fechada.

Da mesma forma, pode-se tratar outras particularidades da língua, onde a possibilidade de alterar a estrutura do autômato apresenta uma solução para esses casos. Assim, este trabalho mostra alguns exemplos que justificam a idéia proposta.

1.2 Objetivos

Esse trabalho tem como objetivo contribuir no desenvolvimento de tradutores texto-voz, estudando a viabilidade da utilização de autômatos adaptativos para esse fim. Para isso, se fez necessário um estudo sobre autômatos adaptativos, formando uma perspectiva sobre as contribuições que essa tecnologia pode trazer para sociedade, conciliada a outras tecnologias, no tocante a resolução de problemas específicos. E é na resolução de um desses problemas que está o foco desse trabalho, em implementações de tradutores texto-voz.

Tradução texto-voz vem sendo tema de pesquisas há algum tempo, onde as aplicações englobam desde telefonia à utilização da internet por deficientes visuais. Porém, a implementação desses sistemas com autômatos adaptativos ainda é algo a ser estudado e ponderado. Este projeto, além de verificar a viabilidade da utilização de autômatos adaptativos em sintetizadores de voz, têm como objetivo investigar sobre padrões para representação de fonemas, definindo qual o mais adequado para ser utilizado na ferramenta AdapTools, favorecendo o trabalho cooperativo nessa área.

A seguir, de modo sucinto, estão relacionados os objetivos específicos desse trabalho:

1. Investigar a possibilidade de utilizar autômatos adaptativos para síntese de voz;
2. Definir padrão para representação de fonemas no AdapTools;
3. Implementar protótipo do sintetizador utilizando o padrão escolhido;
4. Construção de um banco de fonemas para sílabas do português brasileiro;
5. Prover melhorias na pronúncia gerada pela ferramenta AdapTools.

Capítulo 2

Fundamentação Teórica

Ao desenvolver o projeto de um sistema de síntese de voz, alguns conceitos devem estar bem claros. É importante se ter o domínio sobre a realidade do problema, e o conhecimento sobre o estado da arte em que se encontram as tecnologias referentes ao problema abordado.

Para fundamentar o desenvolvimento desse projeto, foram investigados os conceitos de síntese de voz, autômatos adaptativos e a ferramenta AdapTools, descritos a seguir. A última seção desse capítulo trata dos estudos referentes a utilização de autômatos adaptativos como solução para síntese de voz.

2.1 Síntese de voz

Nos últimos anos, diferentes técnicas vêm sendo aplicadas no intuito de otimizar a fala produzida por sistemas sintetizadores de voz [7, 8, 9, 10]. Para isso, a adequação de fatores como ritmo e entonação são fundamentais para que a saída sonora soe o mais natural possível [11, 9, 12]. Os métodos de síntese de voz, normalmente, são classificados em três grupos: síntese articulatória, síntese fonética e síntese por concatenação [1].

Os métodos para síntese articulatória baseiam-se em um modelo do aparelho vocal humano. Mantendo um conjunto de regras que simulam a língua, os lábios e as cordas vocais, o sistema imita a voz humana ao produzir ressonância e articulação. Atualmente, esse tipo de síntese é pouco utilizado, pois se mostra muito complexo e exige recursos computacionais de alto custo.

As técnicas de síntese fonética são estruturadas sobre um sistema de combinação de freqüências que resulta em uma expressão vocal determinada. A entrada é processada por um conjunto de ressonadores, onde cada um gera a saída equivalente a um som vocal. No final, a saída gerada por todos os ressonadores é somada, produzindo um fonema. A disposição dos res-

sonadores na estrutura de processamento pode variar a cada sintetizador de voz.

Já os métodos de síntese por concatenação, trabalham com uma base de arquivos de som previamente gravados, onde cada fonema é associado a um arquivo. Conforme os fonemas são reconhecidos a partir de uma cadeia de entrada, os arquivos são concatenados, respectivamente, de forma a produzir a pronúncia [13]. Com essa abordagem torna-se mais fácil atingir naturalidade e inteligibilidade nos resultados, porém utiliza-se mais espaço em memória do que nos demais métodos.

De maneira geral, o processo de geração de pronúncia a partir de uma entrada de texto se dá em duas etapas: reconhecimento de texto e geração de som [14, 9]. As subseções 2.1.1 e 2.1.2 comentam essas duas etapas.

2.1.1 Reconhecimento de texto

A primeira etapa de execução em um tradutor texto-voz é o reconhecimento de texto. Nessa fase, o programa analisa o texto de entrada para reconhecer cada seqüência de letras relativa a fonemas. Porém, nem todo conteúdo de um texto se encontra no formato adequado para a execução desse processo, muitas vezes exigindo que a cadeia de entrada seja normalizada. Essa normalização consiste de uma pré-formatação, onde siglas, abreviações, numerais, etc, são identificados e analisados lexicamente.

Primeiramente, é importante levar em consideração a linguagem em que o texto de origem foi redigido. Existem algoritmos que abrangem mais de uma linguagem [15], porém esse aspecto influi diretamente no reconhecimento dos fonemas. Outro ponto importante nessa análise são as abreviações e símbolos. A presença desses elementos no texto implica na transformação automática dos mesmos em palavras, para que depois sejam reconhecidos como fonemas (e.g. 123 = *cento e vinte e três*).

2.1.2 Geração de pronúncia

A semelhança entre a saída vocal gerada por um sintetizador e a voz de uma pessoa está diretamente ligada à transição entre a pronúncia das sílabas, entonação e ritmo. Uma técnica bastante usada é a contatenação dos arquivos de som referentes a execução da pronúncia de cada fonema [16]. Essa concatenação é feita no intuito de diminuir a saliência na transição entre fonemas. Outro parâmetro importante para a naturalidade da fala gerada é o tratamento da duração de cada fonema e o ritmo da pronúncia. Um texto pronunciado com um ritmo uniforme e com tempo de duração dos fonemas

contínuo, se torna desagradável de ser ouvido, e prejudica a inteligibilidade na comunicação.

A dependência de contexto é um fator importante para a pronúncia correta de algumas sílabas, podendo alterar o sentido da palavra ou da sentença. Um exemplo disso é a palavra *gosto*, que assume significado e pronúncia diferentes, dependentemente do contexto em que está inserida. Nesse caso, o contexto indicará se a palavra faz referencia a um verbo ou a um substantivo. Nas frases "Eu gosto de suco"e "Suco de laranja agrada mais ao meu gosto"a palavra *gosto* assume diferentes pronúncias. No primeiro caso, a sílaba tônica assume a pronúncia fechada, já no segundo, a tônica é aberta.

2.2 Autômatos adaptativos

Um autômato adaptativo é um dispositivo guiado por regras em que a camada subjacente consiste de um autômato de pilha estruturado e as ações adaptativas da camada adaptativa são implementadas através de *funções adaptativas* [5].

Nas funções adaptativas está contido o tratamento adequado que deve ser executado na camada subjacente do dispositivo, assim que uma ação adaptativa é invocada. O acionamento de uma ação adaptativa pode ser compreendido como simplesmente uma chamada de função, podendo inclusive, receber parâmetros de entrada. Essa função recebe a denominação de função adaptativa. As funções adaptativas possuem um núcleo constituido de um conjunto de *ações adaptativas elementares*, e elas se enquadram em três tipos: as *ações elementares de consulta*, que tratam da busca de padrões na estrutura definida pelas regras da camada subjacente; e as *ações elementares de inserção* e de *remoção*, que determinam, respectivamente, as regras que deverão ser inseridas ou removidas do conjunto de regras corrente da camada subjacente.

As ações elementares apresentam-se basicamente como regras da camada subjacente, podendo conter nomes de variáveis e de geradores no lugar de alguns dos elementos da regra. Ao executar ações elementares de consulta, atribuição de valores a estas variáveis fica a cargo de um mecanismo que busca por padrões, tendo como base o conjunto de regras do mecanismo subjacente. Uma vez atribuídos, os valores de cada variável não podem mais ser alterados durante a execução da função adaptativa. Quando o mecanismo de busca por padrões não encontra qualquer regra, na camada subjacente, que satisfaça o formato determinado pela ação elementar de consulta, as variáveis permanecem indefinidas.

Na execução de ações elementares de remoção, inicialmente é feita uma

busca de padrões para atribuir valores a variáveis (que também podem ser encontradas nas ações elementares de remoção), assim como nas ações elementares de consulta. Após isso, se todos os valores de variáveis estiverem definidos, a regra correspondente deve ser eliminada da camada subjacente (caso contrário, a ação elementar é simplesmente ignorada). Existe uma proposta para que consultas não sejam realizadas em ações de remoção, restringindo as operações de consulta às ações elementares de consulta [5].

No que se refere às ações elementares de inserção, todas as variáveis devem ter sido previamente instânciadas, ou marcadas como indefinidas, ocorrendo sempre depois da execução das ações de consulta e remoção. Quando todas as variáveis contidas na ação elementar de inserção estão definidas, é feita a inserção das regras correspondentes na camada subjacente. Ações elementares de inserção podem também conter nomes de geradores, ao invés de variáveis. Neste caso, antes da inserção da nova regra na camada subjacente, todos os geradores são substituídos por novos símbolos, diferentes de qualquer símbolo utilizado na camada subjacente.

A seguir está descrita a utilização de autômatos adaptativos como solução para o problema de balanceamento de parênteses.

2.2.1 Exemplo de implementação de autômatos adaptativos

O problema de balanceamento de parênteses consta da certificação de que determinada cadeia de caracteres possua o mesmo número de "("(abre parênteses) e ")"(fecha parênteses). E ainda, ao percorrer a cadeia, da esquerda para a direita, em nenhum momento o número de ")"pode ser maior que o número de "("(não pode fechar parênteses antes de ter aberto).

A solução para o problema de balanceamento de parênteses pode ter várias abordagens, com variadas estruturas de dados, inclusive. A figura 2.1 mostra a solução para esse problema utilizando autômatos adaptativos. Esse autômato, apresenta um *"loop"* com a transição)[.$A1$] no estado 0, e uma transição vazia do estado 0 para o estado 1. Ao ler um "("da cadeia, a ação adaptativa A1 é invocada, inserindo transições como indicado na figura 2.1.

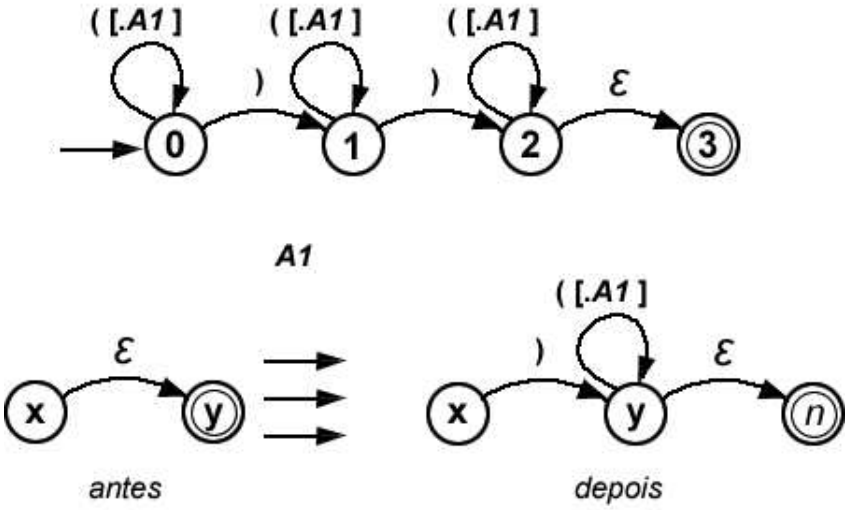

Figura 2.1: Autômato adaptativo apresentado como solução para balanceamento de parênteses

2.3 AdapTools

O AdapTools é uma ferramenta livre que oferece um ambiente para experiências, implementação e depuração de autômatos adaptativos [5, 17]. O núcleo desse software é uma máquina virtual que executa autômatos adaptativos representados em forma de tabela. Os autômatos e as alterações por ele realizadas são visualizados utilizando-se do pacote OpenJGraph, que é uma ferramenta livre, somado de algoritmos de balanceamento de layout baseados nas leis de tensão e repulsão da física. Uma característica importante é a execução de múltiplos dispositivos simultaneamente, possibilitados pelos recursos de *multi-threading* da linguagem Java.

Também como forma de representação, o AdapTools gera animações referentes aos autômatos adaptativos. A figura 2.2 mostra a animação gerada pelo AdapTools para os primeiros estágios de um autômato adaptativo que reconhece $a^n b^n c^n$, cadeia a qual um autômato de estados finitos não consegue reconhecer.

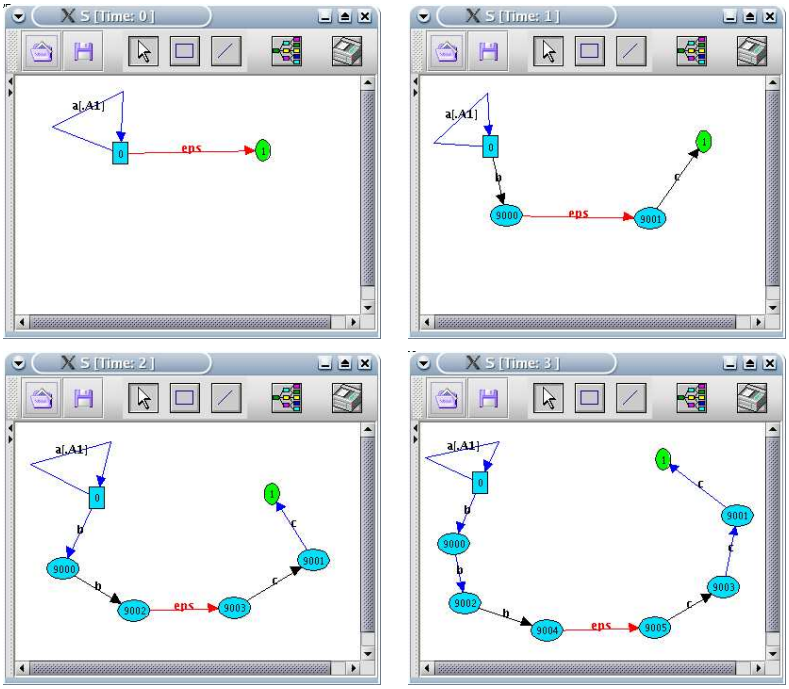

Figura 2.2: Animação gerada pelo Adaptools

Um outro exemplo, também contido no pacote AdapTools, é um protótipo de um conversor texto-voz utilizando autômatos de estados finitos. Este protótipo, embora bastante simples e sem recursos sofisticados de geração automática de entonação ou de ligação suave entre fonemas, já se apresenta como um núcleo importante sobre o qual poderão ser desenvolvidas soluções mais sofisticadas. No campo pedagógico, ele mostra, de uma maneira bastante atraente, onde as técnicas de compilação não se restringem ao domínio das linguagens de programação. No caso específico, temos a tradução (mesmo que ainda pouco sofisticada) de um código-fonte, escrito em linguagem natural, para uma seqüência de sinais sonoros reproduzidos por alto-falantes.

Para armazenar as informações referentes aos autômatos, o AdapTools possui formatos próprios de arquivos [5]. Existem duas versões para os arquivos do AdapTools, e em ambas a primeira linha contém a informação que as distingue ("$[Version]$1"ou "$[Version]$2"), e em seguida, cada linha representa uma transição do autômato ou uma ação adaptativa. Para cada transição existem sete parâmetros de configuração. Na versão 1 os parâmetros das transições são separados em colunas por "espaços em branco". Já na versão 2 esses parâmetros são separados pelo caracter ";"(ponto e vírgula).

O primeiro parâmetro consta de um tipo de cabeçalho para a transição. Nele pode ser indicado o tipo de função adaptativa será executada, caso a linha esteja definindo uma ação (busca, remoção ou inserção). O segundo, terceiro e quarto parâmetros constam respectivamente de estado de origem, transição e estado destino. O quinto campo representa uma flag que pode indicar características do estado de destino daquela transição como, por exemplo, indicar que é um estado final. O sexto parâmetro indica a saída que a transição gera. Em outras palavras, quando esse campo é setado o AdapTools age como um transdutor. Esse campo é utilizado para indicar o fonema relacionado com dada transição, quando o AdapTools é utilizado como tradutor texto-voz. Por fim, o último parâmetro indica o nome da função adaptativa que será invocada quando a transição for executada. A seguir, um exemplo de arquivo do AdapTools para reconhecimento da cadeia $a^n b^n c^n$, cuja representação gráfica se encontra na figura 2.2.

```
[Version] 2
?A1;?x;eps;?y;?z;nop;nop;
-A1;?x;eps;?y;?z;nop;nop;
+A1;?x;b;*n1;nop;nop;nop;
+A1;*n1;eps;*n2;nop;nop;nop;
+A1;*n2;c;?y;?z;nop;nop;
S;0;a;0;nop;nop;.A1;
S;0;eps;1;fin;nop;nop;
```

Capítulo 3

Descrição do Problema

Um sistema de síntese de voz deve oferecer condições mínimas para que o processo de comunicação com o usuário seja eficiente, gerando pronúncias o mais próximo possível da fala humana. Neste intuito, a proposta de utilização de autômatos adaptativos na implementação de tradutores texto-voz faz parte de estudos para melhoria das técnicas utilizadas.

Esse tipo de sintetizador de voz possibilitará a distinção de pronúncia para casos onde a mesma sílaba pode gerar som diferente, de acordo com a palavra em que ela está inserida. Isso é possível pois os autômatos adaptativos são capazes de diferenciar fonemas para uma mesma palavra que é pronunciada de maneira diferente, dependentemente do contexto. Um exemplo disso é a letra x nas palavras "saxofone"e "mexer", pois gera som diferente para cada caso.

Após os fonemas terem sido identificados, o maior problema é que a pronúncia tenha uma fluência sonora agradável aos auvidos. Como primeiro passo para solução desse problema, esse trabalho abordará a concatenação de fonemas, utilizando arquivos de som préviamente gravados em disco. Esse processo tem início com o estudo do funcionamento do protótipo contido na ferramenta AdapTools. E então, aplicar o que foi abstraído, para obter uma melhoria na qualidade da pronúncia gerada.

Capítulo 4

Fonética

A fala humana é um meio de comunicação que possui um grande poder de expressão, pois é segmentável em pequenas unidades linearmente dispostas, podendo sofrer alterações para modificar o significados das palavras. Existe um conjunto de unidades que formam a base de cada linguagem, cada uma dessas unidades recebe o nome de fonema. Esses podem alterar o sentido de uma palavra quando substituídos. Por exemplo, nas palavras pau, mau, e nau, a mudança no sentido da palavra se dá pela substituição da primeira letra da palavra, caracterizando unidades sonoras diferentes. Os fonemas são gerados a partir de alguns parâmetros controlados por cada parte componente do aparelho fonador humano [18], tais como: pulmões, laringe, lábios, fossa nasal, etc.

A fonética engloba estudos sobre a fala humana e pode ser dividida em três vertentes de estudo: fonética fonológica, fonética acústica e fonética auditiva. Fonética fonológica estuda como os sons são produzidos, sobre uma abordagem anatômica, considerando o funcionamento de cada órgão envolvido na fala humana. Ao conjunto desses órgãos chamamos de aparelho fonador, constituído por língua, lábios, cordas vocais, etc. Alguns pesquisadores implementam produção de fala baseando-se em modelos de parâmetros que imitam cada característica existente na anatomia humana, visando aproximar o máximo possível os sons gerados da naturalidade[12].

A fonética acústica analisa características físicas dos sons da fala. Em outras palavras, estuda as ondas mecânicas produzidas pelo aparelho fonador. A partir dessa análise pode-se extrair alguns padrões que a fala humana obedece ao ser gerada, e com isso fundamentar os princípios dessa produção sonora.

Já a fonética auditiva, tem como objetivo o estudo da fala humana sob o aspecto perceptivo, ou seja, a forma como os sons produzidos pelo aparelho fonador são percebidos. Nessa vertente de estudos os pesquisadores

tomam como base a constituição do aparelho auditivo, bem como, suas características sensoriais.

A fala é capaz de reproduzir diversos fones, que são unidades básicas para a fonética. Em 1886 pesquisadores britânicos e franceses, sob os auspícios da Associação Fonética Internacional, desenvolveram uma notação padrão para representar fonemas de todas as linguagens faladas, o Alfabeto Fonético Internacional[1] (IPA - International Phonetic Alphabet). A seção 4.3 contém maiores informações sobre esse alfabeto.

Alguns fones são semelhantes auditivamente e muitas vezes indistinguíveis. Por exemplo, na palavra "rato", o som do *r* no português do Brasil é realizado foneticamente pela consoante fricativa velar (R no padrão SAMPA, descrito na seção 4.3.1), como ocorre também em "rabo". Mas ao substituir pelo fonema da consoante fricativa glotal (r no padrão SAMPA), a palavra ainda poderá ser reconhecida. Os fones que possuem essa características são denominados alófonos. O conjunto formado por dado fone e os seus alófonos no mesmo idioma é denominado fonema. É importante salientar que a alofonia está diretamente ligada as regras de pronúncia das linguagens.

4.1 Aparelho Fonador

Todos os órgãos (pulmões, laringe, lábios, etc) que compõe o aparelho fonador podem influenciar no som gerado na fala de modo geométrico, mecânico e acústico. Fatores como timbre, entonação, volume e ressonância são diretamente ligados a atuação de partes específicas do aparelho fonador. Contudo, as características dos fonemas que são relevantes nesse trabalho são definidas no trato vocal. O trato vocal, representado na figura 4.1, é basicamente formado por glote, língua, dentes, lábios e músculos do interior da boca, pode alterar a som da voz conforme muda a tensão ou distância entre esses elementos[19].

[1] http://www.arts.gla.ac.uk/IPA/ipa.html

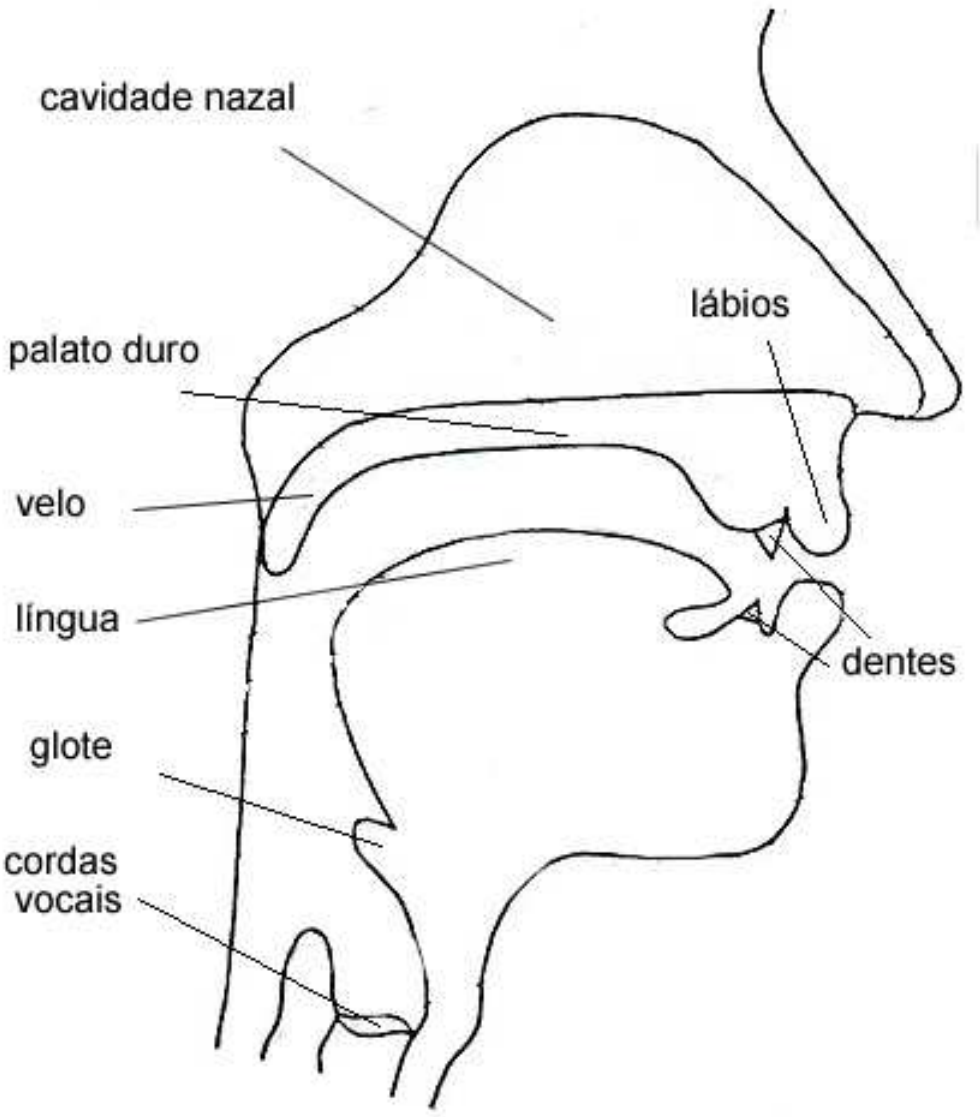

Figura 4.1: Trato vocal

4.2 Classificação Fonética

De modo geral, os segmentos fonéticos são divididos em classes, de acordo com o modelo de articulação pelo qual ele é produzido. As classes são: vogais, glides, oclusivas, fricativas, nasais e líquidas. Ainda, em cada classe, os segmentos fonéticos são distintos pelo ponto de articulação no trato vocal. Para representação dos segmentos fonéticos são utilizados alfabetos fonéticos, no propósito de representar todos os fonemas de determinada lingua. As seguintes subseções classificam os fonemas da língua portuguêsa, utilizando-se a notação SAMPA, descrita na seção 4.3.1.

4.2.1 Vogais

Vogais são segmentos fonéticos vozeados, cujo som normalmente é produzido com o trato vocal em um formato fixo, ou seja, o som é gerado sem obstáculos chegando livremente ao exterior do aparelho fonador. Na língua portuguesa são 13 vogais (na tabela SAMPA, figura 4.8: /6/, /a/, /e/, /E/, /i/, /o/, /O/, /u/, /6~/, /e~/, /i~/, /o~/, /u~/). As vogais tem maior duração que as consoantes e glides e uma melhor definição de frequência.

4.2.2 Glides

Glides, ou semi-vogais, são ditongos compostos pelas letras a com i(pai) e a com u(mau), e seus respectivos sons nasalados na junção de $\tilde{a}$ com e(mãe) e $\tilde{a}$ com o(mão). Respectivamente representados pelos símbolos /w/, /i/, /w~/, e /i~/ (figura 4.8). Pode-se dizer que as glides são vogais com maior constrição e menor duração que vogais comuns.

4.2.3 Oclusivas

As oclusivas são geradas a partir da produção de som pela constrição total do trato vocal (zona de oclusão), e em seguida a rápida liberação da tensão acumulada (zona de explosão). O que diferencia as oclusivas é a presença ou ausência de vozeamento no momento da liberação da pressão e o ponto de ocusão. Por exemplo, as oclusivas /p/, /t/ e /k/ apresentam quase um silêncio total no ponto de oclusão, diferente das oclusivas /b/, /d/ e /g/ (figura 4.9).

4.2.4 Fricativas

Assim como as oclusivas, um dos aspectos que distingue as fricativas é a presença de vozeamento. Além disso o ponto de constrição também é uma característica para distinção dos fonemas produzidos. As fricativas são geradas a partir da tubulência originada pela constrição do trato vocal. São exemplos de fricativas com presença de vozeamento: /v/, /z/ e /Z/, e com ausência de vozeamento /f/, /s/ e /S/. Esses fonemas estão relacionados na tabela da figura 4.9.

4.2.5 Nasais

As nasais são produzidas com a vibração das cordas vocais, mas com a boca fechada, o que ocasiona circulação de ar pelas narinas. Também chamados de oclusivos nasais, por serem geradas com a oclusão da boca, as nasais têm freqüência inversamente proporcional as dimensões do trato vocal. Onde o /m/ tem freqüência menor que o /n/, que por sua vez tem freqüência menor que o /J/ (figura 4.9). É comum o som anasalado de uma vogal que precede uma oclusiva nasal.

4.2.6 Líquidas

As líquidas se dividem em líquidas laterais (/l/, /l~/ e /L/) e líquidas vibrantes (/r/ e /R/). As laterais são produzidas pela passagem de ar nas laterais da lingua, que obstrui o centro do trato vocal. As líquidas vibrantes apresentam uma grande variabilidade, podendo ou não apresentar vozeamento. No caso de /R/ (r múltiplo), é produzido com a vibração da lingua, atingindo o meio "céu da boca". Já o /r/(r simples) é produzido com apenas um toque da lingua nos aovéolos dentários. As fricativas laterais /l/ e /l~/ tem o mesmo ponto de articulação, mas /l~/ ocorre apenas em finais de sílabas. As líquidas estão representadas na figura 4.9.

4.3 Alfabeto Fonético Internacional

Os símbolos utilizados para a representação fonética no Alfabeto Fonético Internacional são, em sua maioria, letras originárias do alfabeto romano, ou derivadas dele. Entretanto, também foram utilizadas algumas letras do alfabeto grego e símbolos que não pertencem a alfabeto nenhum.

Existem duas formas de transcrição para representar os caracteres no alfabeto: transcrição fonológica e transcrição fonética. Respectivamente, utiliza-

se os caracteres entre barras para representar os fonemas e entre colchetes para representar os sons dos fonemas.

Nesse alfabeto buscou-se utilizar um único símbolo para cada som, de modo que não ocorram casos como na língua inglesa, em que *th* e o *sh* são combinados para um som. Alguns sons não representam o mesmo som que aparenta na língua portuguesa, por exemplo, o [j] que equivale ao *j* pronunciado no alemão ou neerlandês. Em português, esse som é obtido através da pronúncia de ditongos com *i* como em *saia*.

As figuras apresentadas a seguir representam os fonemas e seus respectivos símbolos de acordo com o IPA [2]. A figura 4.2 representa as 28 vogais, onde o som representado pelo símbolo, ao ser pronunciado, tem uma projeção frontal (boca e nariz) na esquerda da tabela. No lado direito da tabela estão representados os sons que são projetados no palato, e entre esses dois extremos estão os sons intermediários. Além disso, a figura possui uma mudança gradativa na representação dos sons verticalmente, ou seja, de cima para baixo na tabela. Os sons do topo da figura são de projeção fechada (ressonância), e abrem a projeção conforme posicionados mais a baixo na figura.

As figuras 4.3, 4.4, 4.5, 4.6, e 4.7 representam os fonemas para as consoantes, de acordo com o IPA.

[2] tabelas encontradas em http://www.arts.gla.ac.uk/IPA/ipachart.html

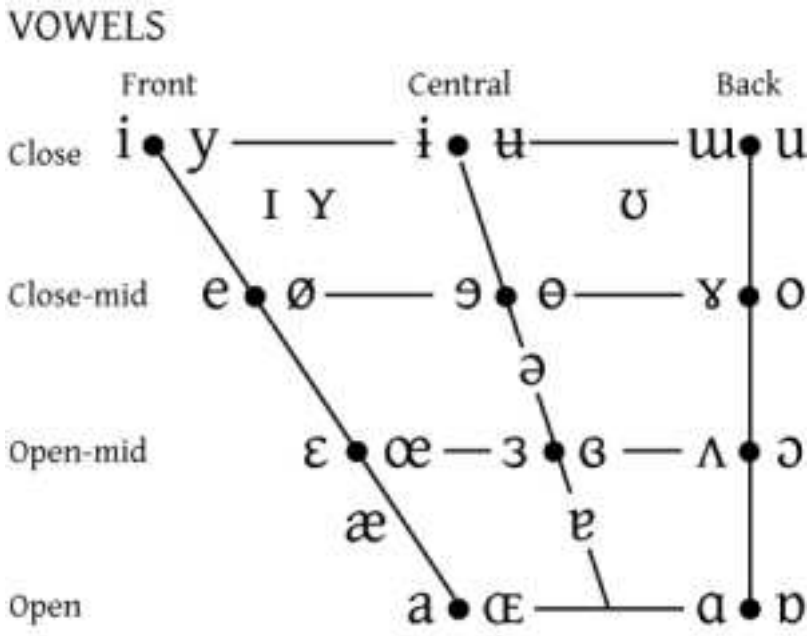

Figura 4.2: Vogais de acordo com IPA

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Figura 4.3: Consoantes Pulmonicas de acordo com IPA

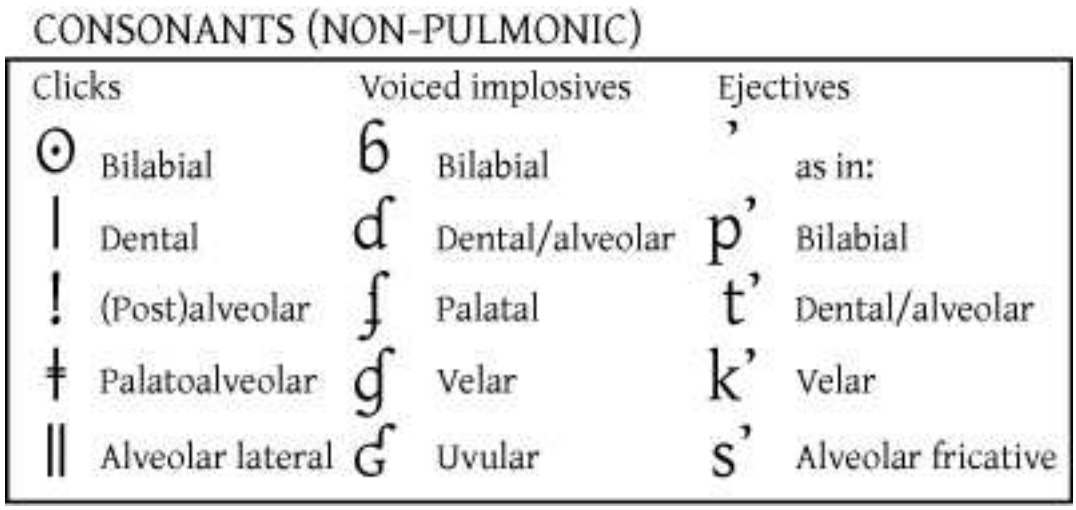

Figura 4.4: Consoantes Não-Pulmonicas de acordo com IPA

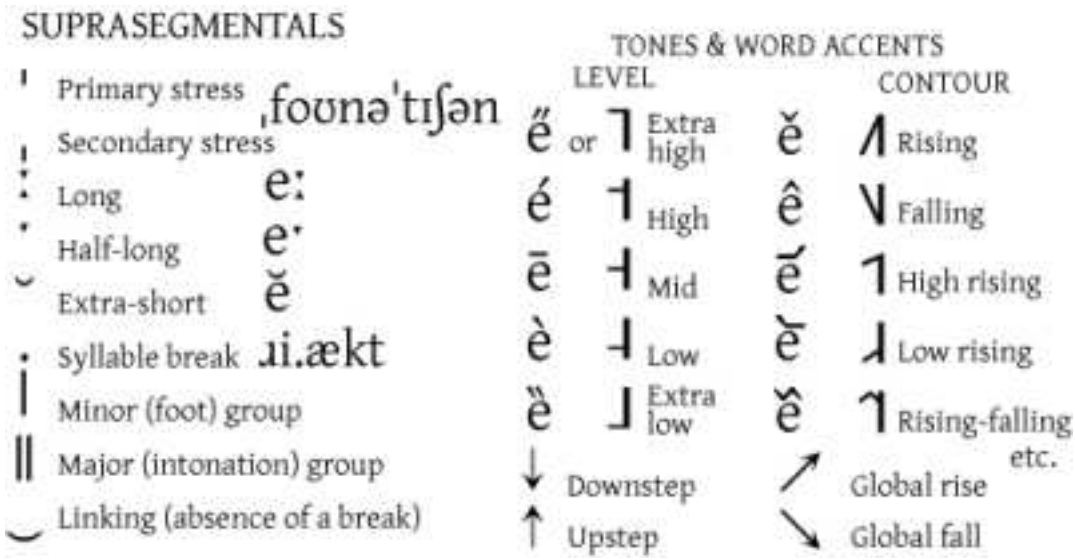

Figura 4.5: Fonemas supra-segmentais de acordo com o IPA

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. n̊ŋ̊

◌̥ Voiceless n̥ d̥	◌̤ Breathy voiced b̤ a̤	◌̪ Dental t̪ d̪
◌̬ Voiced s̬ t̬	◌̰ Creaky voiced b̰ a̰	◌̺ Apital t̺ d̺
◌ʰ Aspirated tʰ dʰ	◌̼ Linguolabial t̼ d̼	◌̻ Laminal t̻ d̻
◌̹ More rounded ɔ̹	◌ʷ Labialized tʷ dʷ	◌̃ Nasalized ẽ
◌̜ Less rounded ɔ̜	◌ʲ Palatalized tʲ dʲ	◌ⁿ Nasal release dⁿ
◌̟ Advanced u̟	◌ˠ Velarized tˠ dˠ	◌ˡ Lateral release dˡ
◌̠ Retracted i̠	◌ˤ Pharyngealized tˤ dˤ	◌̚ No audible release d̚
◌̈ Centralized ë	~ Velarized or pharyngealized ɬ	
◌̽ Mid-centralized e̽	◌̝ Raised e̝ (ɹ̝ = voiced alveolar fricative)	
◌̩ Syllabic ɹ̩	◌̞ Lowered e̞ (β̞ = voiced bilabial approximant)	
◌̯ Non-syllabic e̯	◌̘ Advanced Tongue Root e̘	
◌˞ Rhoticity ɚ	◌̙ Retracted Tongue Root e̙	

Figura 4.6: *Diacritics* de acordo com IPA

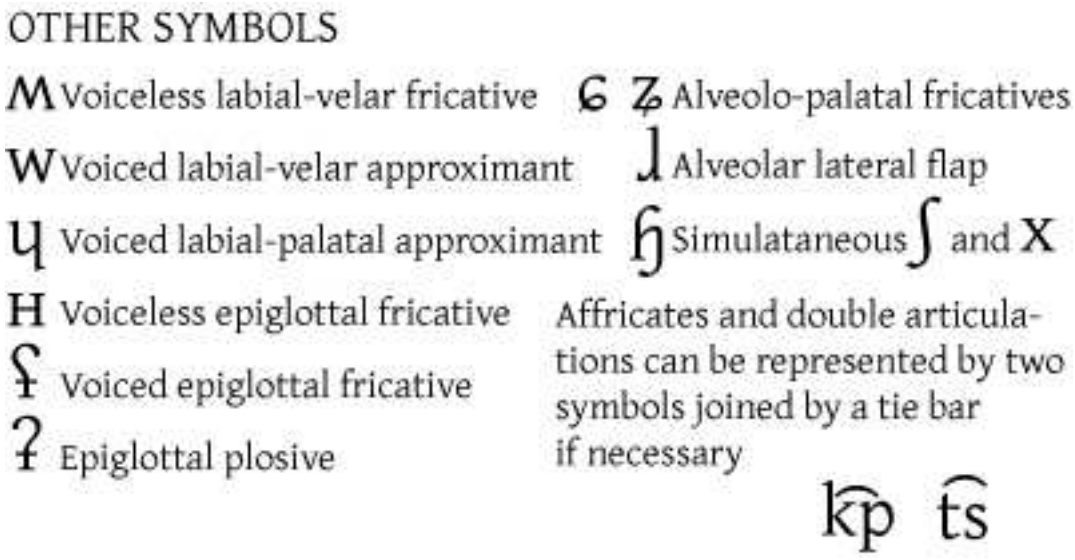

Figura 4.7: Simbolos de acordo com IPA

4.3.1 Notação SAMPA

Como o IPA faz uso de símbolos não comuns aos utilizados na lingua portuguesa, em 1980 foi criado o padrão SAMPA (Speech Assessment Methods Phonetic Alphabet - Alfabeto fonético dos métodos de avaliação da fala) [22]. SAMPA é um alfabeto fonético inicialmente desenvolvido para o português europeu, é um sistema de escrita fonético, legível por computadores que usa o conjunto de caracteres ASCII de 7 bits, baseado no IPA. Apesar de ter sido desenvolvida utilizando-se o maior número de caracteres do IPA possível, a tabela SAMPA só é válida para os idiomas para os quais ela foi adaptada, não representando todo o IPA.

A figura 4.8 mostra a representação dos fonemas para vogais e glides de acordo com a tabela SAMPA[3]. Já a tabela da figura 4.9 contém a representação de consoantes. Ambas as tabelas estão codificadas para o português europeu, porém a diferença para o português brasileiro consta apenas de que o europeu contém alguns fonemas a mais.

Classe	símbolo IPA	símbolo SAMPA	Altura da elevação da língua	Posição da língua na cavidade bocal	palavra	transcrição SAMPA
Vogais	ɐ	6	média	média	cama	k6m6
	a	a	baixa	média	cara	kar6
	e	e	média	anterior	pêra	per6
	ɛ	E	baixa	anterior	sete	sEt@
	ɨ	@	alta	média	que	k@
	i	i	alta	anterior	fita	fit6
	o	o	média	posterior	dou	do
	ɔ	O	baixa	posterior	corda	kOrd6
	u	u	alta	posterior	mudo	mudu
	ɐ̃	6~	média	média	manta	m6~t6
	ẽ	e~	média	anterior	menta	me~t6
	ĩ	i~	alta	anterior	pinta	pi~t6
	õ	o~	média	posterior	ponta	po~ta
	ũ	u~	alta	posterior	mundo	mu~du
Glides	w	w	alta	posterior	pau	paw
	j	j	alta	anterior	pai	paj
	w̃	w~	alta	posterior	cão	k6~w~
	j̃	j~	alta	anterior	mãe	m6~j~

Figura 4.8: Vogais e Glides de acordo com alfabeto SAMPA

[3]http://www.phon.ucl.ac.uk/home/sampa/home.htm

Classe	símbolo IPA	símbolo SAMPA	Presença de Vozeamento	Ponto de articulação	palavra	transcrição SAMPA
Oclusivas	p	p0,p	não	bilabial	pai	p0paj
	t	t0,t	não	apicodental	tia	t0ti6
	k	k0,k	não	velar	casa	k0k6za
	b	b0,b	sim	bilabial	bar	b0bar
	d	d0,d	sim	apicodental	data	d0dat6
	g	g0,g	sim	velar	gato	g0gatu
Fricativas	f	f	não	labiodental	férias	fEri6S
	s	s	não	apicodental	selo	selu
	ʃ	S	não	palatal	chave	Sav@
	v	v	sim	labiodental	vaca	vak6
	z	z	sim	apicodental	azul	6zul~
	ʒ	Z	sim	palatal	agir	6Zir
Nasais	m	m	sim	bilabial	meta	mEt6
	n	n	sim	apicodental	neta	nEt6
	ɳ	J	sim	palatal	senha	s6J6
		N	sim			
Líquidas	l	l	sim	apicodental	lado	ladu
	ɫ	l~	sim	apicodental	sal	sal~
	ʎ	L	sim	palatal	folha	foL6
	R	R		velar	carro	kaRu
	ɾ	r		apicodental	caro	karu
Silêncio		sil				

Figura 4.9: Consoantes de acordo com alfabeto SAMPA

Capítulo 5

Desenvolvimento da Pesquisa

Os estudos referentes a autômatos adaptativos indicaram que eles são uma poderosa ferramenta para reconhecimento de gramáticas dependentes de contexto. Outro sim, o conflito das informações estudadas fortalece a possibilidade de implementação de sintetizadores de voz baseados em autômatos adaptativos. A seção a seguir descreve detalhes dessa implementação.

5.1 Síntese de voz e autômatos adaptativos

A síntese de voz implementada no AdapTools utiliza métodos de síntese por concatenação, porém faz apenas a "união dos sons"executando os sons seqüencialmente, sem nenhum tipo de otimização para que a pronúncia soe mais naturalmente. A entrada para a tradução texto-voz é uma seqüência de caracteres. Para isso, deve-se ter um banco de fonemas previamente gravado em arquivos de som. Depois que é definido o texto de entrada, o AdapTools faz o reconhecimento do mesmo usando como estrutura de base um autômato adaptativo. Em seguida, mapeia cada fonema reconhecido com o arquivo de som respectivo, reproduzindo-o.

A figura 5.1 mostra algumas das transições do AEF que traduz texto em voz. Estão indicados nela legenda no formato s/f, onde s é o símbolo de entrada, e f é a saída respectiva. Normalmente, f representa a união de dois fonemas, o da consoante e o da vogal. Isso se deve ao fato de que os sons gerados pelas consoantes só existem se gerados juntamente com vogais.

A semântica deste dispositivo foi implementada de forma tal que cada símbolo de saída f seja mapeado em um arquivo de som, no formato *wave* (.wav), que é automaticamente enviado para a saída de som do computador assim que a transição f é executada. As transições $(6, a, 7)$ e $(10, a, 7)$, por exemplo, produzirão ambas o mesmo som, pois apesar de serem reconhecidas

distintamente, elas são mapeadas para os mesmos fonemas *za*.

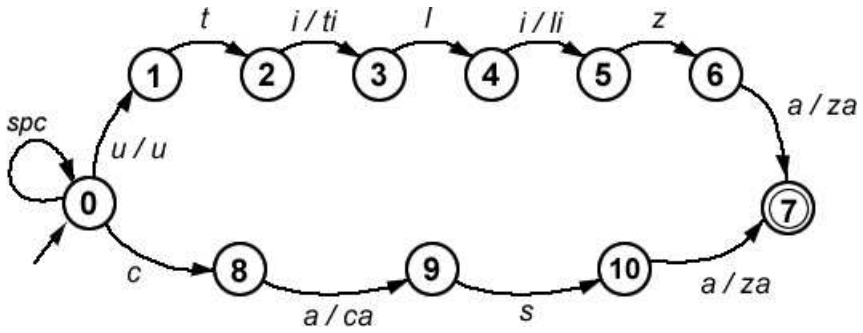

Figura 5.1: Exemplos de transições em autômato de tradução texto-voz

Por fim, a simples reprodução dos fonemas na seqüência que eles são identificados, não garante bons resultados na pronúncia gerada. As sílabas ficam desconexas, sem fazer a suavização entre a pronúncia de dois fonemas, assim como fazemos ao falar. Para resolver esse problema, propõe-se um estudo sobre manipulação do arquivo de som utilizado (*wave*). E com isso, fazer a edição desses arquivos unindo-os na seqüência em que forem reconhecidas as palavras [20, 21].

5.2 Áudio

5.2.1 Bibliotecas utilizadas

A execução do som é feita utilizando classes livres disponíveis no pacote *SoundPlayer*[1], criado por Martin Stepp, fazendo com que a programação seja simples. Abaixo, segue um trecho de código que utiliza o *SoundPlayer* para reproduzir um arquivo de som.

```
public SoundPlayer player = new SoundPlayer();

private Token say(String nomeFonema) {
   player.playAndWait(diretorioBase+nomeFonema+".wav");
}
```

Para que a fala gerada seja o mais natural possível, os arquivos wave contendo os fonemas devem ser submetidos a um processamento antes de serem executados. Esse processamento envolve concatenação de fonemas, sobreposição de vogais, ajuste de entonação, etc[22, 13]. Nesse intuito foram estudadas classes para manipulação de arquivos *wave*, de forma a implementar concatenação sequencial dos fonemas em um único arquivo para depois ser executado. Para essa tarefa foi utilizada a classe *AudioConcat*[2].

Essa classe implementa a leitura e escrita de arquivos de áudio, gerando um arquivo a partir da leitura de outros. Esse processo pode ser executado de duas formas: a concatenação (utilizando -c como parâmetro) e o *mix* (utilizando -m como parâmetro). A rotina para o tratamento dos arquivos de som, no intuito de melhorar a fala no AdapTools, utiliza o método de concatenação (com opção "-c"), seguida da opção "-o"(indicando o nome do arquivo destino) e o nome do arquivo onde serão escritos os dois fonemas concatenados. Essa rotina é chamada por um objeto da classe *Semantics*

[1]http://www.cs.arizona.edu/ stepp/java.html
[2]http://www.jsresources.org/examples/AudioConcat.html

para cada fonema reconhecido, concatenando-o e gerando um arquivo de saída. Enquanto o autômato não reconhecer um "espaço"(*pausa*), o objeto concatena o arquivo de saída com o arquivo referente ao fonema em questão. Quando um espaço é reconhecido o objeto executa o arquivo de saída, nessa instância, contendo a pronúncia de toda a palavra.

5.2.2 Banco de fonemas

Para realizar os experimentos de tradução texto-voz no AdapTools, e assim, constatar a viabilidade de utilização de autômatos adaptativos para esse fim, é preciso ter os fonemas previamente gravados em arquivos de áudio. E para aplicar testes com uma maior variedade de palavras é necessário um banco de fonemas que tenha todos os sons que os testes exijam. Tendo isso em vista, uma das contribuições deste trabalho é disponibilizar um banco de arquivos no formato *wave*, contendo um conjunto de fonemas para experimentos no AdapTools.

Devido a dificuldade de gravar e manipular o som das consoantes separadamente, cada arquivo contém a pronúncia de pares (consoante + vogal) de fonemas, onde cada consoante é combinada com cada vogal em um arquivo. O nome do arquivo é definiddo a partir do padrão SAMPA, utilizado na chamada de execução dos fonemas, sendo que este consta da sílaba em questão. Por exemplo, para a palavra *bola* teríamos os arquivos *bO.wav* e *la.wav*. A seguir está representado, no formato de arquivo do AdapTools, um exemplo de autômato que reconhece a palavra *bola* e utiliza o banco de fonemas criado nesse projeto.

```
[Version] 2
B;0;b;1;nop;nop;nop;
B;1;o;0;nop;bO;nop;
L;0;l;2;nop;nop;nop;
L;2;a;0;nop;la;nop;
Fin;0;eps;pop;fin;nop;nop;
```

5.3 Ação adaptativa e pronúncia

Um dos maiores desafios no projeto de um tradutor texto-voz para a língua portuguesa é o tratamento das exceções das regras de pronúncia. Normalmente, os mecanismos de tradução texto-voz, fazem reconhecimento das cadeias de entrada de forma sequencial. Nesse caso, ao mapear sílabas individualmente para arquivos de som, um autômato adaptativo oferece a possibilidade de alterar o link feito para o fonema, mesmo depois que o mecanismo já tenha analisado certo trecho da cadeia de entreda.

Por exemplo, ao reconhecer as palavras *bela* e *beleza* não se pode saber se os fonemas relativos à primeira sílaba serão /bE/ ou /be/ até que se tenha lido a segunda sílaba, ou ainda a palavra inteira. Ao utilizar autômatos adaptativos como mecanismo de reconhecimento de texto, pode-se alterar o fonema que será pronunciado, mesmo após esse processo já ter sido executado em outro momento do reconhecimento da cadeia.

No anexo B está o arquivo onde define-se as transições e ações adaptativas que possibilitam a pronúncia das palavras *caça*, *casa*, *casto*, *casório* e *cassino*. Esse exemplo apresenta uma solução para tradutores texto-voz utilizando autômatos adaptativos, assim como o proposto nesse trabalho. Esse autômato é composto por dois autômatos de funções distintas: o de reconhecimento e o de pronúncia. O autômato de reconhecimento possui transições para identificar as palavras, enquanto o autômato de pronúncia é composto por uma sequência de transições vazias, com o campo transdutor indicando o arquivo de áudio a ser executado (fonemas). O autômato de reconhecimento identifica as letras de cada palavra, e então, quando formam uma sílaba, uma ação adaptativa insere no autômato de pronúncia uma transição referente ao respectivo fonema. Após o reconhecimento da cadeia de entrada, o último estado para cada palavra tem uma transição vazia para o primeiro estado do autômato de pronúncia, executando assim a sequência de fonemas.

Capítulo 6

Considerações Finais

As atividades executadas durante a primeira fase desse projeto consistiram de estudos sobre síntese de voz e autômatos adaptativos. Foi feito o levantamento sobre o estado atual das aplicações para tradução texto-voz, constando que são justificáveis as pesquisas para aprimoramento dos sintetizadores atuais, visando sempre uma interface mais amigável entre computadores e humanos.

Inserido neste contexto foi estudada a ferramenta AdapTools, que implementa um protótipo de tradutor texto-voz baseado em autômatos adaptativos. Essa ferramenta gera pronúncia a partir de execução seqüencial de arquivos previamente gravados com fonemas para as sílabas respectivas. Apesar de não possuir nenhum mecanismo de suavização entre sílabas, nem controle de entonação ou ritmo, o AdapTools oferece uma base para o desenvolvimento de pesquisas sobre a utilização de autômatos adaptativos para síntese de voz, podendo ser aprimorado.

Parte das primeiras pesquisas desse projeto compuseram um artigo submetido e aceito no "V Workshop sobre Software Livre", que ocorreu em junho de 2004. O artigo traz como título "Tecnologia Adaptativa e Síntese de Voz: Primeiros Experimentos"[23] e descreve os primeiros estudos sobre esse projeto, propondo uma nova metodologia para implementação de sintetizadores de voz, utilizando autômatos adaptativos.

Para gerar pronúncia a partir de arquivos de áudio previamente gravados, existem diversas características importantes a serem tratadas. São parâmetros específicos, diretamente ligados a linguagem utilizada na fala e a anatomia do sistema fonador humano. No intuito de tornar a pronúncia o mais natural possível, pode-se ajustar esses parâmetros em uma sílaba de acordo com o contexto que ela está inserida.

É possível melhorar a pronúncia das palavras no AdapTools processando os arquivos de áudio antes de serem executados. Com a análise de casos es-

pecíficos de dependência de contexto na lingua portuguêsa, notou-se que automatos adaptativos apresentam uma solução eficaz para resolução de problemas de entonação, pronúncia diferente para sílabas iguais, pronúncia igual para sílabas diferentes e exceções. Entre as técnicas estudadas para atenuar a transição entre fonemas, este trabalho contribui com a concatenação dos arquivos de aúdio, fazendo com que as palavras sejam "pronunciadas"de uma só vez, com um único arquivo de áudio. Em trabalhos futuros, o tratamento de cada fonema antes da concatenação, pode vir a melhorar ainda mais a qualidade da fala gerada, bem como configurar a entonação das palavras e destacar a sílaba tônica.

No intuito de estudar a viabilidade de implementação de tradutores texto-voz, tendo como estrutura base autômatos adaptativos, pôde-se salientar o poder que esse mecanismo oferece, adequando-se à solução para o problema proposto. E ainda, aliado a rotinas de tratamento semântico, firmou-se a possibilidade de obter resultados satisfatórios, no sentido de gerar fala com qualidade próxima à gerada pelo aparelho fonador humano.

Anexo A

Autômato Adaptativo para execução de fonemas

A seguir, o conteúdo do arquivo (no formato AdapTools) que representa o autômato para pronúncia de sílabas. O campo transdutor utiliza o padrão (SAMPA) adotado nesse trabalho para nomenclatura de arquivos de áudio (wav). Esse autômato utiliza o banco de fonemas desenvolvido nesse projeto.

```
[Version] 2


Paus;0;spc;0;nop;pausa;nop;


Voga;0;a;0;nop;a;nop;
Voga;0;e;0;nop;e;nop;
Voga;0;é;0;nop;E;nop;
Voga;0;i;0;nop;i;nop;
Voga;0;o;0;nop;o;nop;
Voga;0;ó;0;nop;O;nop;
Voga;0;u;0;nop;u;nop;


B;0;b;1;nop;nop;nop;
B;1;a;0;nop;ba;nop;
B;1;e;0;nop;be;nop;
B;1;ê;0;nop;be;nop;
B;1;é;0;nop;bE;nop;
```

```
B;1;i;0;nop;bi;nop;
B;1;o;0;nop;bo;nop;
B;1;ô;0;nop;bo;nop;
B;1;ó;0;nop;bO;nop;
B;1;u;0;nop;bu;nop;


C;0;c;2;nop;nop;nop;
C;2;a;0;nop;ka;nop;
C;2;o;0;nop;ko;nop;
C;2;ô;0;nop;ko;nop;
C;2;ó;0;nop;kO;nop;
C;2;u;0;nop;ku;nop;


D;0;d;3;nop;nop;nop;
D;3;a;0;nop;da;nop;
D;3;e;0;nop;de;nop;
D;3;ê;0;nop;de;nop;
D;3;é;0;nop;dE;nop;
D;3;i;0;nop;di;nop;
D;3;o;0;nop;do;nop;
D;3;ó;0;nop;do;nop;
D;3;ô;0;nop;dO;nop;
D;3;u;0;nop;du;nop;


F;0;f;4;nop;nop;nop;
F;4;a;0;nop;fa;nop;
F;4;e;0;nop;fe;nop;
F;4;ê;0;nop;fe;nop;
F;4;é;0;nop;fE;nop;
F;4;i;0;nop;fi;nop;
F;4;o;0;nop;fo;nop;
F;4;o;0;nop;fo;nop;
F;4;ó;0;nop;fO;nop;
F;4;u;0;nop;fu;nop;


G;0;g;5;nop;nop;nop;
G;5;a;0;nop;ga;nop;
```

```
G;5;o;0;nop;go;nop;
G;5;ô;0;nop;go;nop;
G;5;ó;0;nop;gO;nop;
G;5;u;0;nop;gu;nop;


H;0;h;6;nop;nop;nop;
H;6;a;0;nop;Ra;nop;
H;6;e;0;nop;Re;nop;
H;6;ê;0;nop;Re;nop;
H;6;é;0;nop;RE;nop;
H;6;i;0;nop;Ri;nop;
H;6;o;0;nop;Ro;nop;
H;6;ô;0;nop;Ro;nop;
H;6;ó;0;nop;RO;nop;
H;6;u;0;nop;Ru;nop;


J;0;j;7;nop;nop;nop;
J;7;a;0;nop;Za;nop;
J;7;e;0;nop;Ze;nop;
J;7;ê;0;nop;Ze;nop;
J;7;é;0;nop;ZE;nop;
J;7;i;0;nop;Zi;nop;
J;7;o;0;nop;Zo;nop;
J;7;ô;0;nop;Zo;nop;
J;7;ó;0;nop;ZO;nop;
J;7;u;0;nop;Zu;nop;


L;0;l;8;nop;nop;nop;
L;8;a;0;nop;la;nop;
L;8;e;0;nop;le;nop;
L;8;ê;0;nop;le;nop;
L;8;é;0;nop;lE;nop;
L;8;i;0;nop;li;nop;
L;8;o;0;nop;lo;nop;
L;8;ô;0;nop;lo;nop;
L;8;ó;0;nop;lO;nop;
L;8;u;0;nop;lu;nop;
```

```
M;0;m;9;nop;nop;nop;
M;9;a;0;nop;ma;nop;
M;9;e;0;nop;me;nop;
M;9;ê;0;nop;me;nop;
M;9;é;0;nop;mE;nop;
M;9;i;0;nop;mi;nop;
M;9;o;0;nop;mo;nop;
M;9;ô;0;nop;mo;nop;
M;9;ó;0;nop;mO;nop;
M;9;u;0;nop;mu;nop;


N;0;n;10;nop;nop;nop;
N;10;a;0;nop;na;nop;
N;10;e;0;nop;ne;nop;
N;10;ê;0;nop;ne;nop;
N;10;é;0;nop;nE;nop;
N;10;i;0;nop;ni;nop;
N;10;o;0;nop;no;nop;
N;10;ô;0;nop;no;nop;
N;10;ó;0;nop;nO;nop;
N;10;u;0;nop;nu;nop;


P;0;p;11;nop;nop;nop;
P;11;a;0;nop;pa;nop;
P;11;e;0;nop;pe;nop;
P;11;ê;0;nop;pe;nop;
P;11;é;0;nop;pE;nop;
P;11;i;0;nop;pi;nop;
P;11;o;0;nop;po;nop;
P;11;ô;0;nop;po;nop;
P;11;ó;0;nop;pO;nop;
P;11;u;0;nop;pu;nop;


R;0;r;13;nop;nop;nop;
R;13;a;0;nop;ra;nop;
R;13;e;0;nop;re;nop;
R;13;ê;0;nop;re;nop;
```

```
R;13;é;0;nop;rE;nop;
R;13;i;0;nop;ri;nop;
R;13;o;0;nop;ro;nop;
R;13;ô;0;nop;ro;nop;
R;13;ó;0;nop;rO;nop;
R;13;u;0;nop;ru;nop;


S;0;s;14;nop;s;nop;
S;14;a;0;nop;sa;nop;
S;14;e;0;nop;se;nop;
S;14;ê;0;nop;se;nop;
S;14;é;0;nop;sE;nop;
S;14;i;0;nop;si;nop;
S;14;o;0;nop;so;nop;
S;14;ô;0;nop;so;nop;
S;14;ó;0;nop;sO;nop;
S;14;u;0;nop;su;nop;


T;0;t;15;nop;nop;nop;
T;15;a;0;nop;ta;nop;
T;15;e;0;nop;te;nop;
T;15;ê;0;nop;te;nop;
T;15;é;0;nop;tE;nop;
T;15;i;0;nop;ti;nop;
T;15;o;0;nop;to;nop;
T;15;ô;0;nop;to;nop;
T;15;ó;0;nop;tO;nop;
T;15;u;0;nop;tu;nop;

V;0;v;16;nop;nop;nop;
V;16;a;0;nop;va;nop;
V;16;e;0;nop;ve;nop;
V;16;ê;0;nop;ve;nop;
V;16;é;0;nop;vE;nop;
V;16;i;0;nop;vi;nop;
V;16;o;0;nop;vo;nop;
V;16;ô;0;nop;vo;nop;
V;16;ó;0;nop;vO;nop;
V;16;u;0;nop;vu;nop;
```

```
X;0;x;17;nop;nop;nop;
X;17;a;0;nop;Sa;nop;
X;17;e;0;nop;Se;nop;
X;17;ê;0;nop;Se;nop;
X;17;é;0;nop;SE;nop;
X;17;i;0;nop;Si;nop;
X;17;o;0;nop;So;nop;
X;17;ô;0;nop;So;nop;
X;17;ó;0;nop;SO;nop;
X;17;u;0;nop;Su;nop;

Z;0;z;18;nop;nop;nop;
Z;18;a;0;nop;za;nop;
Z;18;e;0;nop;ze;nop;
Z;18;ê;0;nop;ze;nop;
Z;18;é;0;nop;zE;nop;
Z;18;i;0;nop;zi;nop;
Z;18;o;0;nop;zo;nop;
Z;18;ô;0;nop;zo;nop;
Z;18;ó;0;nop;zO;nop;
Z;18;u;0;nop;zu;nop;

Fin;0;eps;pop;fin;nop;nop;
```

Anexo B

Exemplo de solução para tradutor texto voz.

Exemplo de implementação autômato adaptativo a ser utilizado em tradutor texto-voz. Abaixo, está o conteúdo do arquivo do AdapTools que pronuncia as palavras *caca*, *casa*, *casto*, *casório* e *cassino*.

```
[Version] 2

C;0;c;1;nop;nop;nop;
C;1;a;2;nop;nop;.insere(ka);
C;2;eps;1000;nop;pausa;nop;

S;2;s;3;nop;nop;.insere(J);
S;3;eps;1000;nop;pausa;nop;

A;3;a;4;nop;nop;.troca(J,za);
A;4;eps;1000;nop;pausa;nop;

O;3;o;41;nop;nop;.troca(J,zo);
O;41;eps;1000;nop;pausa;nop;

Ó;3;ó;441;nop;nop;.troca(s,zO);
Ó;441;eps;1000;nop;pausa;nop

R;441;r;442;nop;nop;nop;
R;442;i;443;nop;nop;.insere(ri);
R;443;o;444;nop;nop;.insere(o);
```

```
R;444;eps;1000;nop;pausa;nop;

T;3;t;4;nop;nop;nop;
T;4;o;5;nop;nop;.insere(to);
T;5;eps;1000;nop;pausa;nop;

Ç;2;ç;31;nop;nop;nop;
Ç;31;a;32;nop;nop;.insere(Ja);
Ç;32;eps;1000;nop;pausa;nop;

SS;3;s;4441;nop;nop;nop;
SS;4441;i;4442;nop;nop;.insere(Ji);
SS;4442;n;4443;nop;nop;nop;
SS;4443;o;4444;nop;nop;.insere(no);
SS;4444;eps;1000;nop;pausa;nop;


?insere;?x1;eps;?x2;nop;nop;nop;
-insere;?x1;eps;?x2;nop;nop;nop;
+insere;?x1;eps;*n1;nop;?p1;nop;
+insere;*n1;eps;?x2;nop;nop;nop;

?troca;?x1;eps;?x2;nop;?p1;nop;
-troca;?x1;eps;?x2;nop;?p1;nop;
+troca;?x1;eps;?x2;nop;?p2;nop;

arauto;1000;eps;1001;nop;nop;nop;
arauto;1001;eps;0;fin;pausa;nop;
```

Referências Bibliográficas

[1] S. Lemmetty. *Review of Speech Synthesis Technology.* PhD thesis, Helsinki University of Technology - Department of Electrical and Communications Engineering, 1999.

[2] J. J. Taniwaki, C. Y. O. e Neto. Autômatos adaptativos no tratamento sintático de linguagem natural boletim técnico pbt/pcs. Technical report, Escola Politécnica, São Paulo, Brasil, 2001.

[3] Rocha R. L. de A. and Neto J. J., editors. *Uma proposta de método adaptativo para a seleção automática de soluções*, Buenos Aires, 2000. Proceedings of ICIE Y2K - International Congress on Informatics Engineering.

[4] A. V. Neto, J. J.; Freitas. Um método de escolha automática de soluções usando tecnologia adaptativa. Technical report, Escola Politécnica, São Paulo, Brasil, 2000.

[5] H. Pistori. *Tecnologia Adaptativa em Engenharia de Computação: Estado da Arte e Aplicações.* PhD thesis, Escola Politécnica da Universidade de São Paulo, 2003.

[6] A. V. Neto, J. J.; Freitas. Using adaptive automata in a multi-paradigm programming environment. *In: ASM2001 - IASTED International Conference on Applied Simulation and Modelling. Marbella, Espanha: [s.n.]*, 2001.

[7] B. L. Goff and C. Benoît. A text-to-audiovisual-speech synthesizer for french. In *Proc. ICSLP '96*, volume 4, pages 2163–2166, Philadelphia, PA, 1996.

[8] T. Burrows. Speech processing with linear and neural network models, 1996.

[9] S. Lemmetty. Review of speech synthesis technology. Master's thesis, Helsinki University of Technology, 1999.

[10] M. Cohen and D. Massaro. Modeling coarticulation in synthetic visual speech, 1993.

[11] Mark Tatham and Eric Lewis. Improving text-to-speech synthesis. *Proceedings of the Institute of Acoustics*, 18(9):35–42, 1996.

[12] Across Some Swedish. Distribution of dental and retroflex l-sounds.

[13] P. Carvalho, L. Oliveira, I. Trancoso, and M. Viana. Concatenative speech synthesis for european portuguese, 1998.

[14] Michael H. O'Malley. Text-to-speech conversion technology. *IEEE Computer*, 23(8):17–23, aug 1990.

[15] Richard Sproat. Multilingual text analysis for text-to-speech synthesis. Technical report, Speech Synthesis Research Department - Bell Laboratories, 1996.

[16] Y. Stylianou. Concatenative speech synthesis using a harmonic plus noise model, 1998. In: The 3 rd ESCA/COCOSDA Workshop on Speech Synthesis, Jenolan Caves, NSW, Australia, Nov. 1998, Paper H.1.

[17] H. Pistori and J. J. Neto. A free software for the development of adaptive automata. *In Proceedings of the IV Workshop on Free Software - WSL (IV International Forum on Free Software)*, June 2003.

[18] D. T. Chbane. Desevolvimento de sistemas para conversão de textos em fonemas no idioma português. Master's thesis, Escola Politécnica da Universidade de São Paulo, 1994.

[19] Carlos A. Silva, Isabel Trancoso, and Samir Chennoukh.

[20] Dutoit T. High-quality text-to-speech synthesis: an overview. *Journal of Electrical Electronics Engineering, Australia: Special Issue on Speech Recognition and Synthesis*, 1997.

[21] Kanthak and Ney. Multilingual acoustic modeling using graphemes.

[22] J. Wells. Computer-coding the ipa: a proposed extension of sampa, 1995.

[23] F. A. Zuffo and H. Pistori. Tecnologia adaptativa e síntese de voz: Primeiros experimentos. *V Workshop sobre Software Livre*, 2004.