

Universidade Federal do Rio de Janeiro

Escola Politécnica

Departamento de Eletrônica e de Computação

**Documentação do SASPRO – Sistema de Análise e Síntese
de Prosódia**

Autor:

Leonardo Santos Barbosa

Orientador:

Prof. Sergio Lima Netto, Ph. D.

Examinador:

Prof. Luiz Wagner Pereira Biscainho, D. Sc.

Examinador:

Prof. Tadeu Nagashima Ferreira, D. Sc.

DEL

22 de maio de 2014

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

DEDICATÓRIA

Dedico este trabalho à minha família que me apoiou e sempre me deu todo o suporte para que eu pudesse trilhar este caminho, em especial aos meus pais, que investiram na minha formação, e à minha esposa, que me deu todo o suporte necessário para que tudo isso fosse possível.

AGRADECIMENTO

Agradeço a Deus por me capacitar, pois sem Ele nada é possível, e por não me deixar desistir do meu objetivo ou me desviar do caminho que decidi seguir.

Agradeço à minha esposa Viviane, por me dar todo o suporte e apoios necessários, por cuidar da minha casa e de nossa filha Sofia. Ela foi a base e a estrutura necessária para que eu pudesse fazer meu trabalho.

Agradeço à minha família, em especial aos meus pais, por terem acreditado em mim e investido na minha formação desde o início. Isto foi o que, sem dúvida, propiciou a minha chegada até a Universidade e poder trabalhar neste projeto.

Agradeço ao meu orientador, Sergio Lima Netto, que me orientou, confiou em mim para a realização deste projeto e teve paciência para levar a orientação até o fim.

Agradeço também às demais pessoas que participaram direta ou indiretamente dando conselhos e incentivos para que eu continuasse trilhando meu caminho até o fim.

RESUMO

O trabalho em questão trata basicamente da apresentação e suporte de um sistema de análise e síntese de prosódia chamado de SASPRO, que realiza conversão texto-fala. Sua importância reside no fato de fornecer todas as ferramentas necessárias para que qualquer usuário, que podemos chamar de “comum”, possa usar um sistema de edição de prosódia e, caso tenha habilidades na área de programação, possa desenvolvê-lo ainda mais. O SASPRO, em si, possui larga aplicação em sistemas de navegação, sistemas de segurança, pode permitir acessibilidade a deficientes visuais e ajudar na educação, como por exemplo, sendo aplicado no ensino de idiomas, entre outras aplicações.

O trabalho se inicia com uma detalhada apresentação geral do sistema com suas telas, botões, menus e submenus, para que seja feita uma primeira passagem. Em seguida, são divididas as funcionalidades por grupo, a saber, entrada/saída de dados, análise e síntese de dados. Dentre estas funcionalidades podemos citar a transcrição simples de uma entrada de texto, podendo salvá-la para uma posterior consulta; a gravação de um sinal de voz dentro do próprio sistema, com a possibilidade de visualizarmos o sinal no tempo; a edição da prosódia com relação ao tempo, à intensidade ou à duração do sinal, entre outras. Por fim mostramos o uso do sistema SASPRO em uma aplicação prática de montagem (expansão) do banco de unidades do seu sistema de conversão texto-fala.

Palavras-Chave: prosódia, edição, voz, síntese, sinal, suporte, apresentação.

ABSTRACT

The work includes the presentation and support of a system called SASPRO of analysis and synthesis of prosody that performs text-to-speech conversion. Its importance lies in the fact that it provides all the necessary tools so that any common user, can perform advanced speech processing without special skills. The SASPRO has broad application in navigation systems, security systems, may provide accessibility to the visually impaired, and assist in education, for example, being applied in language teaching, among other applications.

The work begins with a detailed overview of the system with their screens, buttons, menus and submenus. Then, the features are divided into groups, namely, input/output data, analysis and synthesis of data. Among such features we find the simple trascription of a text entry, which can be saved in a directory to posterior view; the direct recording of a speech sinal directly through the system, with respective graphic presentation; prosody edition in time, duration or intensity of sinal speech, and many others. Finally we describe a case study where the SASPRO system is employed to expand the unit database of its own text-to-speech converter.

Keywords: prosody, editing, voice synthesis, sign, support, presentation.

SIGLAS

EGG – Eletroglotógrafo

GCI – *Glottal Closure Instants*

SASPRO – Sistema de Análise e Síntese de Prosódia

TTS – *Text-To-Speech*

UFRJ – Universidade Federal do Rio de Janeiro

Sumário

Lista de Figuras	xi
Lista de Tabelas	xv
1 Introdução	1
1.1 Motivação	1
1.2 Organização do Trabalho	2
2 Apresentação do SASPRO	4
2.1 Introdução	4
2.2 Botão Iniciar	5
2.2.1 Banco de Sentenças	7
2.2.2 Barra de Ferramentas	10
2.2.3 Janela de Manipulação Prosódica	11
2.3 Ambiente Transcritor	14
2.4 Ambiente TTS	15
2.4.1 Janela de Manipulação Prosódica TTS	17
2.5 Ambiente Transplante	19
2.6 Janelas	22
2.7 Análises	22
2.8 Considerações Finais	23
3 Funcionalidades	25
3.1 Introdução	25
3.2 Entrada/Saída de Dados	25
3.2.1 Ferramenta de Busca	26
3.2.2 Botão Iniciar	27

3.2.3	Botão Transcritor	29
3.2.4	Botão TTS	31
3.2.5	Botão Transplante	32
3.3	Análise de Dados	36
3.3.1	Tela Inicial	36
3.3.2	Botão Iniciar	38
3.3.3	Botão Transcritor	39
3.3.4	Botão TTS	39
3.3.5	Botão Transplante	42
3.3.6	Botão Janelas	44
3.4	Síntese de Dados	45
3.4.1	Tela Inicial	45
3.4.2	Botão Iniciar	46
3.4.3	Botão Transcritor	47
3.4.4	Botão TTS	47
3.4.5	Botão Transplante	47
3.5	Conclusão	47
4	Estudo de Caso	49
4.1	Introdução	49
4.2	Usando o SASPRO para Aumentar o Banco de Unidades	49
4.3	Varredura do Banco	50
4.3.1	Observações Gerais	50
4.4	Gravações	51
4.5	Correção das Marcas de EGG	53
4.6	Etiquetagem do Sinal	55
4.6.1	Observações Sobre a Etiquetagem	56
4.7	Incorporação de Novas Unidades ao Banco	57
4.7.1	Arquivo de Exceções	58
4.8	Conclusão	58
5	Conclusão	60
5.1	Contribuições Deste Trabalho	60
5.2	Propostas de Trabalhos Futuros	61

Bibliografia	62
A Exceções e Curiosidades Sobre o Sistema	63
A.1 Introdução	63
A.2 Numerais	63
A.3 Onomatopeias	63
A.4 Dígrafos e Encontros Consonantais	64
A.5 Expressões Comuns e Siglas	64
A.6 Interjeições	65
A.7 Observações Sobre a Língua Inglesa	65
B Exemplo de Tabela Para Avaliação das Unidades	66
B.1 Introdução	66
B.2 Tabela Consoante-Vogal	66
B.3 Tabela Vogal-Consoante-Vogal	67
B.4 Tabela Encontros Consonantais	67
B.5 Encontros Vocálicos	68

Lista de Figuras

2.1	Tela inicial do Editor de Prosódia.	5
2.2	Detalhe do botão Iniciar da Barra de Menus: o modo Lista e o modo Database.	6
2.3	Modo Lista selecionado com o arquivo <code>unitsdef.txt</code> aberto.	6
2.4	Iniciar, no modo Database selecionado, com um dos arquivos <code>unitsdef.txt</code> em destaque. Do lado direito um recorte da janela de análise do arquivo selecionado.	7
2.5	Frase de exemplo: “O grêmio ganhou a quadra de esportes.”. Do lado esquerdo temos um banco de sentenças pré gravadas e, do lado direito, uma janela de análise da sentença.	8
2.6	Detalhe da janela de análise ao selecionar a sentença “O grêmio ganhou a quadra de esportes.”.	8
2.7	Recorte do ambiente de análise, mostrando o resultado do processamento da sentença selecionada e o que significa cada uma das linhas.	9
2.8	Destaque do sinal de fala, mostrando as etiquetas de fronteira e o resultado do processamento da transcrição fonética da sentença selecionada.	9
2.9	Destaque do espectrograma do sinal, juntamente com a energia, as etiquetas de fronteira e a transcrição fonética da sentença selecionada.	10
2.10	Destaque da Barra de Ferramentas do ambiente de análise.	10
2.11	Destaque da tela de análise quando o modo GCI (<i>Glottal Closure Instants</i>) é selecionado na Barra de Ferramentas.	11
2.12	Janela de manipulação prosódica que se abre ao clicar no botão Manipulação Prosódica.	12

2.13	Janela de manipulação prosódica dando destaque à três apresentações do sinal: tempo, segmento e sílaba.	12
2.14	Janela de manipulação prosódica com a aba de edição do <i>pitch</i> selecionada.	13
2.15	Janela de Manipulação Prosódica com a aba de edição da intensidade selecionada.	14
2.16	Tela Inicial do Transcritor com as sentenças “Socorro” e “Problema” já transcritas.	15
2.17	Tela inicial do TTS.	16
2.18	Destaque do ambiente exclusivo do TTS.	16
2.19	Destaque da janela de manipulação prosódica do ambiente TTS. . . .	18
2.20	Destaque da janela de estrutura de performance, aberta a partir da janela de manipulação prosódica do ambiente TTS.	18
2.21	Aba Duração da Janela de Manipulação Prosódica do TTS aberta. . .	19
2.22	Aba <i>Pitch</i> da janela de manipulação prosódica do TTS aberta.	20
2.23	Aba Intensidade da janela de manipulação prosódica do TTS aberta. .	20
2.24	Tela inicial do ambiente de Transplante de Prosódia.	21
2.25	Ambiente de Transplante com a opção <i>Label</i> do botão <i>Edit</i> da Barra de Ferramentas ativado.	21
2.26	Destaque ao submenu do botão <i>ToolStripSplitButton1</i> da Barra de Ferramentas do Ambiente de Transplante.	22
2.27	Destaque da tela gerada pelo botão <i>ButtonModel</i> da Barra de Ferramentas do ambiente de Transplante com os parâmetros do modelo de prosódia.	23
2.28	Destaque do botão Janelas da Barra de Menus do SASPRO.	23
2.29	Destaque do botão Análises da Barra de Menus do SASPRO.	24
3.1	Exemplo de utilização da caixa de busca.	26
3.2	Exemplo de utilização do botão <i>Record</i> da Barra de Ferramentas. A sentença utilizada como exemplo foi “O grêmio ganhou a quadra de esportes.”.	28
3.3	Exemplo de utilização do ambiente Transcritor. A sentença utilizada como exemplo foi “Socorro, preciso resolver isso!”.	30

3.4	Salvando um arquivo usando o ambiente Transcritor.	30
3.5	Entrada de dados no ambiente TTS por meio do teclado. A sentença usada foi “Estou quase terminando o projeto...”.	32
3.6	Entrada de dados no ambiente Transplante. A sentença usada foi “Pode dizer as horas, por favor?”.	33
3.7	Entrada de dados no ambiente Transplante com sentença previamente selecionada. A sentença usada foi “A casa foi vendida sem pressa.”.	34
3.8	Sentença sintetizada a partir de outra previamente selecionada.	34
3.9	Sentença sintetizada a partir de uma entrada de dados via teclado.	35
3.10	Sentença sintetizada a partir da lista suspensa.	35
3.11	Tela inicial do SASPRO com a sentença “O menu incluía pratos bem saborosos.”, dando destaque a Barra de Ferramentas.	36
3.12	Utilização do <i>mouse</i> para selecionar uma parte do sinal.	38
3.13	Detalhe da Barra de Ferramentas da tela de manipulação prosódica do TTS.	40
3.14	Utilização do botão Transcreve com a sentença “O capítulo demorou para terminar.”.	41
3.15	Utilização do botão Concatena com a sentença “O capítulo demorou para terminar.”.	42
3.16	Utilização do botão Modela com a sentença “O capítulo demorou para terminar.”.	43
3.17	Destaque da janela aberta a partir do do botão FormDTW da Barra de Ferramentas.	44
3.18	Destaque da Barra de Ferramentas do ambiente Transplante.	44
3.19	Exemplo de utilização do ambiente de manipulação prosódica dentro do TTS.	48
3.20	Exemplo de utilização do ambiente de manipulação prosódica dentro do TTS. Agora modificamos a duração no domínio do tempo.	48
4.1	Exemplo de teste de sílaba para verificação de unidades existentes no banco. No caso, a sílaba testada é BA.	51

4.2	Destaque do arquivo <code>.txt</code> mostrando como indicar o endereço da pasta contendo as sentenças a serem incorporadas no banco de sentenças.	53
4.3	Destaque da tela de edição das marcas de EGG de um sinal qualquer. O exemplo usado aqui foi a sentença “O exame de lógica foi considerado básico, fácil e abstrato.”.	54
4.4	Destaque de um trecho específico de uma sentença para edição de uma marca de EGG.	54
4.5	Imagem do eletroglotógrafo (EGG) em funcionamento.	55
4.6	Destaque do procedimento de correção de etiquetas de fronteira.	56
4.7	Edição da etiqueta de fronteira em um trecho específico por meio dos botões <i>Plot Selection</i> e <i>InsertMark</i>	56
4.8	Destaque do arquivo <code>lexico.txt</code> para que possa ser visto qual o formato de inserção de palavras no banco de exceções.	59

Lista de Tabelas

B.1	Exemplo de tabela resumida para testar a prosódia e/ou transcrição das unidades do tipo consoante-vogal.	67
B.2	Exemplo de tabela resumida para testar a prosódia e/ou transcrição das unidades do tipo vogal-consoante-vogal.	67
B.3	Exemplo de tabela resumida para testar a prosódia e/ou transcrição das unidades em que ocorre encontro consonantal.	68

Capítulo 1

Introdução

1.1 Motivação

Sistemas de conversão texto-fala ou TTS (*text-to-speech*) possuem diversas aplicações práticas, tais como: interfaces de voz, sistemas de navegação, sistemas de segurança ativados por voz, sistemas de leitura automática para deficientes visuais e auxílio a pessoas com disfunções da fala.

Já existem sistemas de conversão texto-fala com ótima inteligibilidade, mas que ainda precisam melhorar com relação à reprodução natural da voz humana. Isto ocorre porque a prosódia carrega consigo um grande número de informações linguísticas (energia do sinal, frequência fundamental e outras características) que colocam em xeque a simplicidade dos modelos de prosódia.

O tema deste trabalho é a análise/síntese de voz, mais especificamente um sistema dotado de um amplo conjunto de funcionalidades para processamento de sinais de voz, e o apoio ao desenvolvimento de um sintetizador de voz.

O objeto de estudo é a análise/síntese de voz por meio de uma ferramenta de suporte à análise e síntese de voz intitulada Sistema de Análise e Síntese de Prosódia (SASPRO).

O sistema SASPRO tem sua importância justificada por sua ampla aplicação no cotidiano, como citado anteriormente. Assim, sua análise e apresentação justificam-

se pelas possibilidades de expansão, desenvolvimento e conseqüentemente aperfeiçoamento do mesmo. Por isso torna-se necessário o estudo do sistema através do seu uso exaustivo para deixar claras todas as funcionalidades presentes, criando exemplos de uso, explicitando suas vantagens e também suas limitações.

Há ainda de se considerar que uma apresentação geral do SASPRO permite a facilidade de comparação entre este sistema e os já existentes no mercado, mais uma vez evidenciando as possibilidades de desenvolvimento do próprio SASPRO, uma vez que este é um sistema de código aberto e esta análise/comparação permite indiretamente um esclarecimento para programadores que pretendam desenvolvê-lo.

O objetivo geral inclui a apresentação, a análise e o possível desenvolvimento de novas funcionalidades de um conjunto de ferramentas para o apoio do desenvolvimento de um sistema de síntese de voz. O projeto considera ainda a comparação destas ferramentas a outras similares, apresentando suas vantagens e desvantagens em relação às demais. Então, têm-se como objetivos específicos:

- (i) Apresentação inicial do sistema explicitando suas telas, botões, menus e sub-menus;
- (ii) Descrição das funcionalidades do sistema como um todo;
- (iii) Criação de exemplos práticos de utilização do sistema;
- (iv) Explicitação as limitações e exceções do sistema;
- (v) Criação da documentação do sistema;
- (vi) Criação do manual do usuário e do programador.

1.2 Organização do Trabalho

A metodologia utilizada neste projeto consiste em um amplo estudo sobre a teoria de análise/síntese de voz, em uma pesquisa na *internet* sobre alguns sistemas similares já existentes no mercado, na utilização do sistema sob o ponto de vista de um usuário dito “comum”, portanto observando as necessidades que devem ser ilustradas e facilitando a utilização do ambiente, e finalmente a utilização na visão

de um programador. Para esta última, a ideia é observar e analisar as classes que compõem o sistema, deixando claras as possibilidades de melhoria e/ou expansão do mesmo.

Para atingirmos os objetivos anteriormente listados, este trabalho está organizado da maneira descrita a seguir.

No Capítulo 2, temos uma apresentação do sistema SASPRO como um todo, listando seus menus, exibindo sempre que necessário as telas e botões de todo o sistema através de figuras e, se for estritamente necessário, fazendo uma prévia apresentação da funcionalidade.

No Capítulo 3, descrevem-se as funcionalidades como um todo, agrupando-as por finalidade. A ideia é percorrer o sistema novamente, mas agora explicitando as funcionalidades dividindo-as em três “grandes grupos”: entrada/saída de dados, análise e síntese de dados. Assim é possível explicitar tudo o que o sistema é capaz de fazer.

No Capítulo 4, é apresentada uma aplicação prática do sistema, mostrando quais os procedimentos necessários para expandir sua base de dados. Uma vez que já descrevemos as funcionalidades do sistema anteriormente, fica fácil entender esta espécie de “estudo de caso”.

Por fim, este trabalho é concluído no Capítulo 5, em que fazemos alguns comentários, procuramos explicitar as exceções dos sistema e, em casos bastante específicos, damos sugestões de melhoria para o mesmo.

Capítulo 2

Apresentação do SASPRO

2.1 Introdução

No capítulo anterior começamos ressaltando a importância dos sistemas de processamento de voz como um todo, em particular dando foco à sua aplicabilidade no cotidiano por meio dos conversores de texto-fala. Em seguida, fizemos uma breve introdução do conteúdo de cada capítulo, dando ao leitor uma visão macro deste documento.

No presente capítulo, teremos uma apresentação geral do SASPRO (Sistema de Análise e Síntese de Prosódia), fazendo um *tour* pelas telas, menus e botões do sistema para que qualquer usuário possa, mesmo sem nunca tê-lo visto *a priori*, sentir-se “confortável” ao utilizá-lo pela primeira vez. Como dito a ideia não é tratar das funcionalidades neste momento, uma vez que isto será feito em capítulos posteriores, mas se necessário será dada uma mínima explicação, desde que isso ajude a tornar a apresentação mais clara.

Sendo assim, esta Seção 2.1 trata quase que exclusivamente da Barra de Menus e de suas telas, botões e menus, como veremos a seguir.

A ideia deste capítulo, como vimos na introdução, é dar uma visão global do sistema sem entrar em muitos detalhes sobre cada uma das funcionalidades.

Ao abrir o sistema vê-se uma tela inicial do editor de prosódia cujo recorte aparece na figura 2.1. Na parte superior esquerda encontramos a Barra de Menus com os seguintes botões: Iniciar, Transcritor, TTS, Transplante, Janelas e Análises. Abaixo da Barra de Menus temos uma caixa de pesquisa e, abaixo desta, uma lista de sentenças pré-gravadas. Em um primeiro momento, vamos a uma breve apresentação de cada um destes botões.

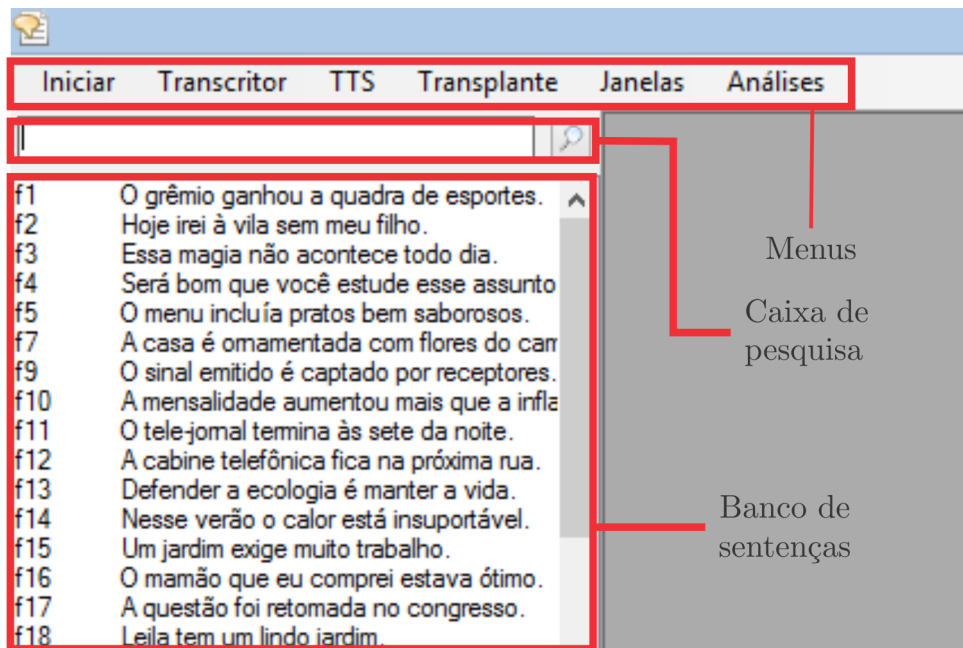


Figura 2.1: Tela inicial do Editor de Prosódia.

2.2 Botão Iniciar

Dentro do botão Iniciar temos duas opções de escolha: o modo Lista e o modo Database, conforme detalhe na figura 2.2. Dentro do modo Database ainda temos as seguintes opções: Abrir, Salvar, Importar Database e Compilar Database.

O modo Lista serve basicamente para a análise da prosódia de sentenças, enquanto que o modo Database serve para a criação de uma base de dados. Voltaremos a estes detalhes no capítulo 3. As figuras 2.3 e 2.4 mostram as telas quando abrimos o modo Lista e o modo Database, respectivamente, para que possamos fazer uma comparação entre os dois modos.

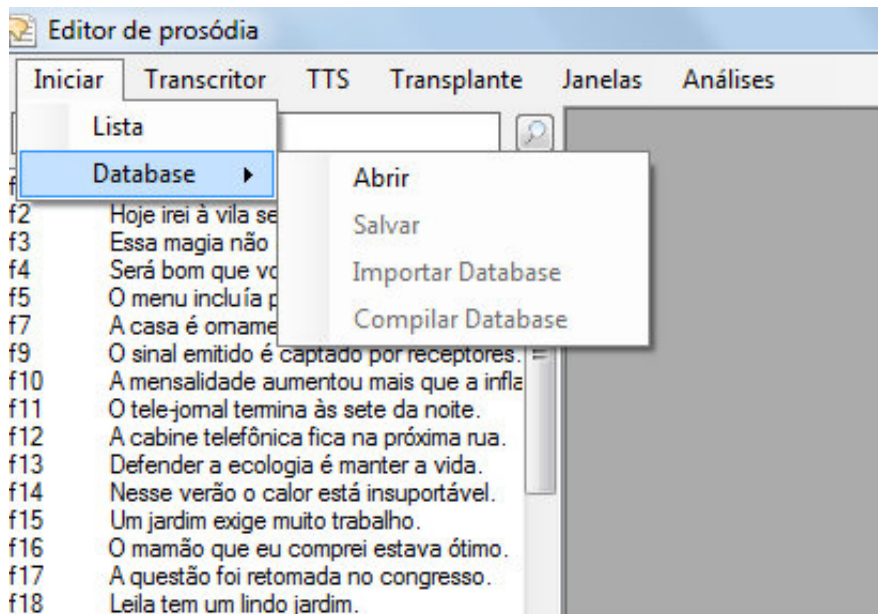


Figura 2.2: Detalhe do botão Iniciar da Barra de Menus: o modo Lista e o modo Database.

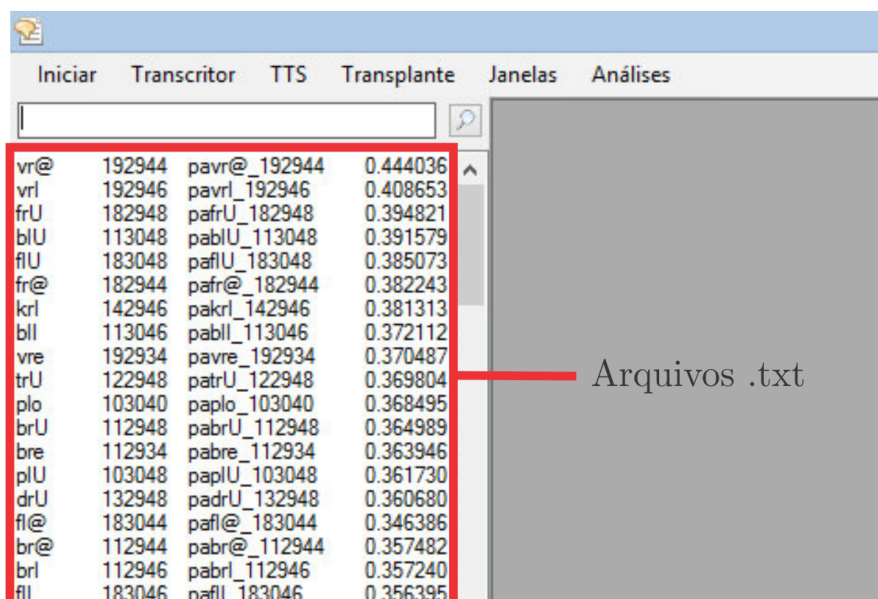


Figura 2.3: Modo Lista selecionado com o arquivo unitsdef.txt aberto.

Ao clicar no botão Iniciar e depois na opção Lista uma janela é aberta para que se selecione o arquivo de origem. Na figura 2.3 abrimos o arquivo unitsdef.txt dentro da pasta SASPRO/Projetos/db/logotomas.

No botão Iniciar, clicamos em Database e depois em Abrir. Feito isso, uma janela de busca se abre para que o arquivo .txt procurado seja selecionado. No exemplo dado, usamos o arquivo unitsdef.txt dentro da pasta SASPRO/Projetos/db/lo-

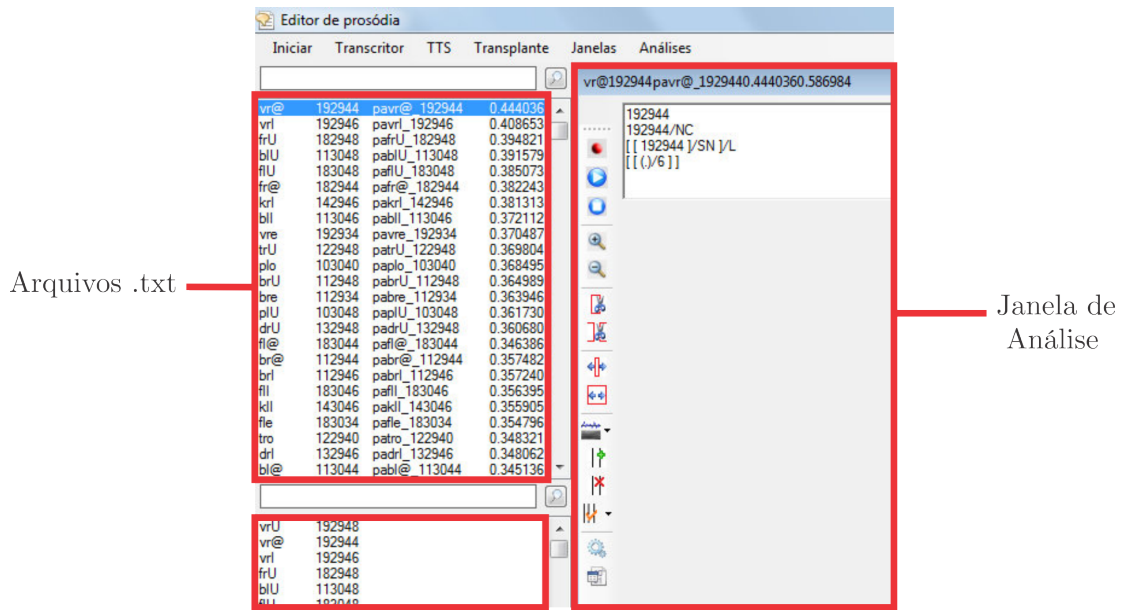


Figura 2.4: Iniciar, no modo Database selecionado, com um dos arquivos `unitsdef.txt` em destaque. Do lado direito um recorte da janela de análise do arquivo selecionado.

gotomas.

2.2.1 Banco de Sentenças

Voltando à exposição da tela inicial (figura 2.1), logo abaixo da Barra de Menus aparece uma caixa de pesquisa. Abaixo desta caixa, temos uma lista de sentenças pré-gravadas que aparece em destaque na mesma figura. Selecionando-se qualquer uma destas sentenças, abre-se uma janela de análise da respectiva sentença no ambiente principal com as características do arquivo (ou sentença) selecionado(a). Na figura 2.5 selecionamos a oração “O grêmio ganhou a quadra de esportes.” como exemplo. Vamos agora dar um destaque à janela de análise, mostrando seus botões e menus, como se vê na figura 2.6.

Na parte esquerda da figura 2.6 temos a Barra de Ferramentas da Janela de Análise. Listados, de cima para baixo, temos os botões com as seguintes funcionalidades deste ambiente: Gravar (*Record*), *Play*, *Stop*, *Zoom In*, *Zoom Out*, *Delete*, *Delete Inverse/Selection*, *Plot All*, *Plot Selection*, *Edit*, *Insert mark*, *Delete Marks*, *Detect Marks*, *Manipulation* e *Add Unit Button*. Como, por enquanto, nosso intuito é de apenas apresentar o sistema como um todo, não daremos maiores detalhes sobre

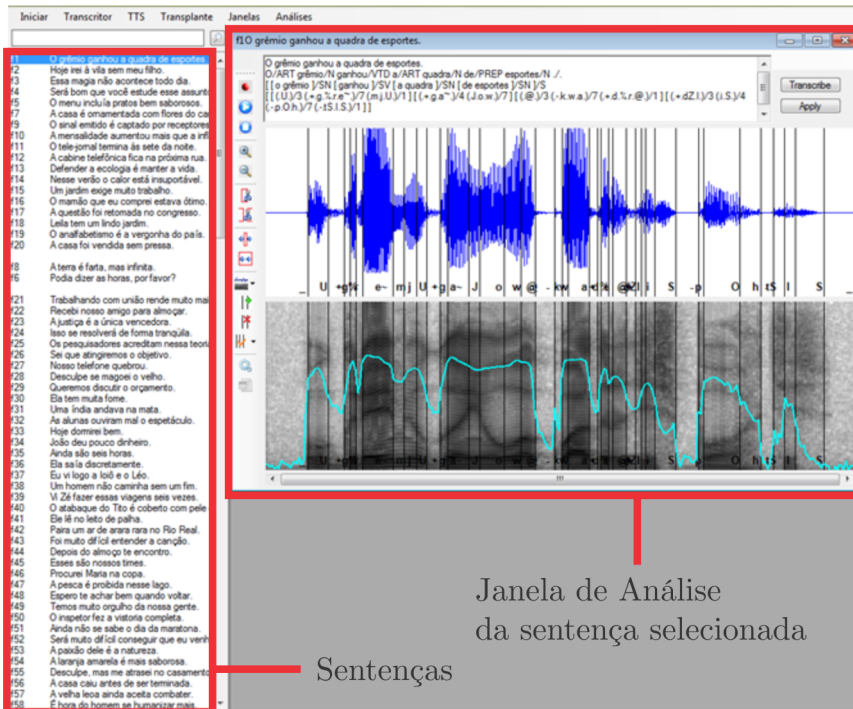


Figura 2.5: Frase de exemplo: “O grêmio ganhou a quadra de esportes.”. Do lado esquerdo temos um banco de sentenças pré gravadas e, do lado direito, uma janela de análise da sentença.

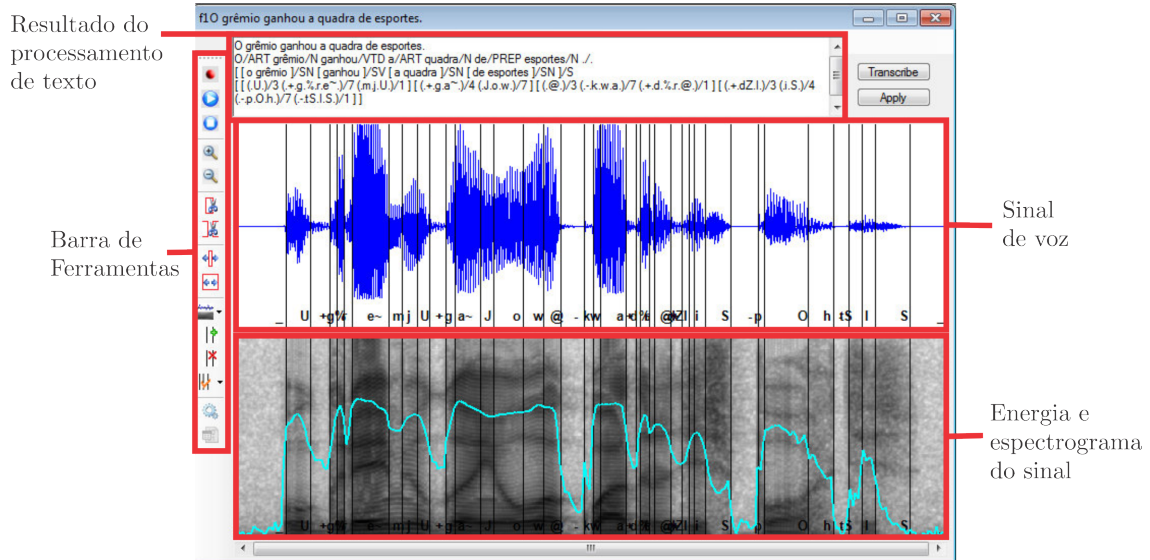


Figura 2.6: Detalhe da janela de análise ao selecionar a sentença “O grêmio ganhou a quadra de esportes.”.

as funções de cada botão agora, postergando isso para o Capítulo 3.

A primeira janela dentro deste subambiente (em destaque figura 2.7) mostra a sentença previamente selecionada e o resultado do processamento do texto. A janela imediatamente abaixo desta (realçada na figura 2.8) mostra o sinal de fala com o que chamamos de etiqueta de fronteira. Na parte inferior desta mesma janela vemos a mesma transcrição fonética destacada na última linha da janela anterior.

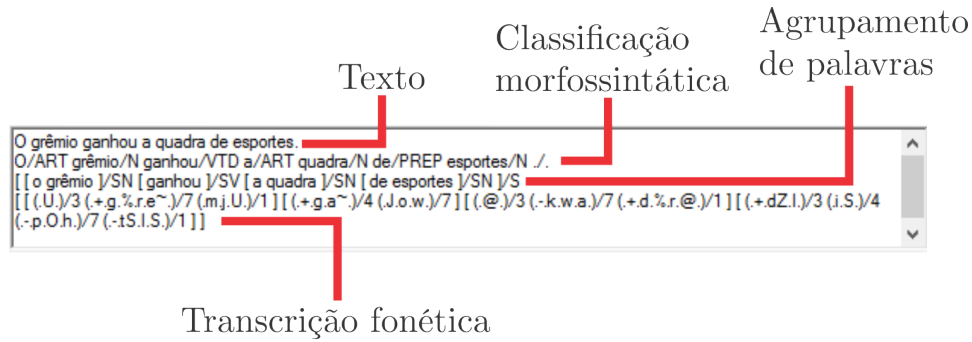


Figura 2.7: Recorte do ambiente de análise, mostrando o resultado do processamento da sentença selecionada e o que significa cada uma das linhas.

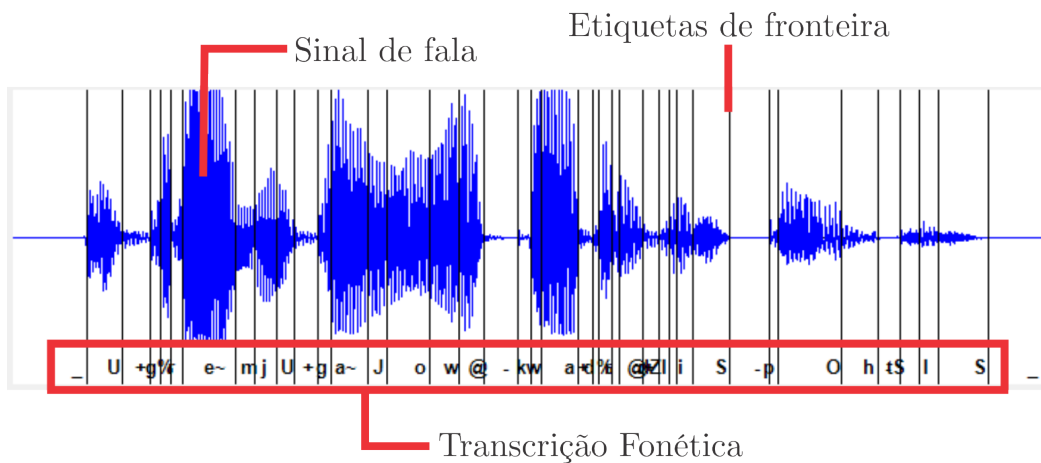


Figura 2.8: Destaque do sinal de fala, mostrando as etiquetas de fronteira e o resultado do processamento da transcrição fonética da sentença selecionada.

A terceira janela da figura 2.6 (em destaque na figura 2.9) mostra o espectrograma do sinal também com as etiquetas de fronteira. A curva em azul-claro mostra a intensidade do sinal ao longo do tempo.

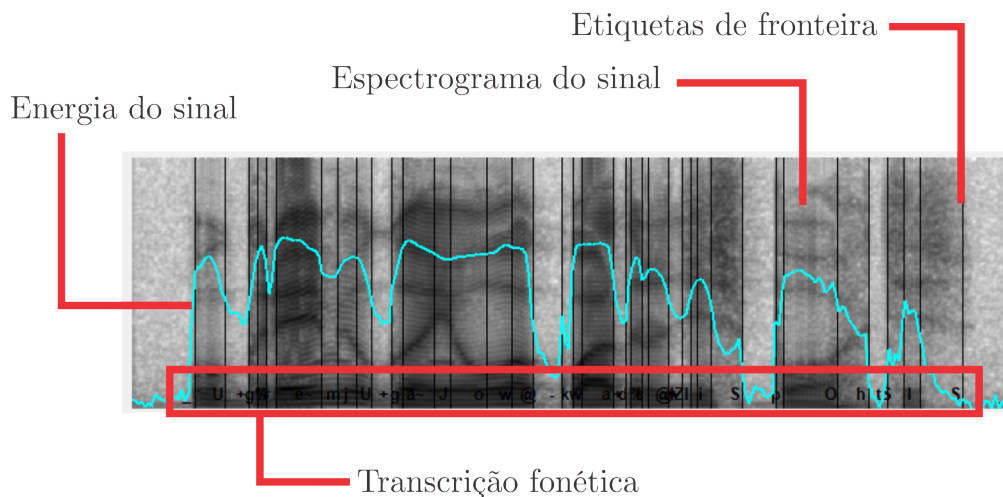


Figura 2.9: Destaque do espectrograma do sinal, juntamente com a energia, as etiquetas de fronteira e a transcrição fonética da sentença selecionada.

2.2.2 Barra de Ferramentas

Como vimos na figura 2.6, do lado esquerdo da janela de análise, há uma Barra de Ferramentas. Trataremos de cada funcionalidade posteriormente, mas agora queremos dar destaque a dois botões que geram novas telas. Estes botões aparecem na figura 2.10.

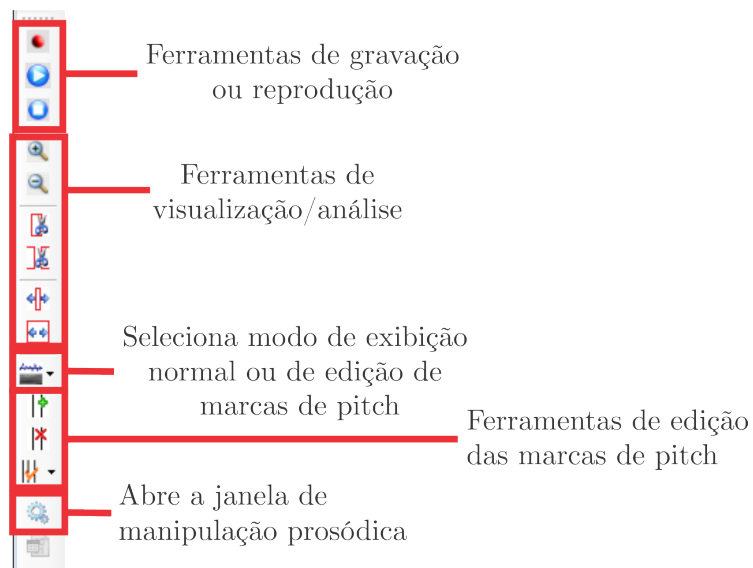


Figura 2.10: Destaque da Barra de Ferramentas do ambiente de análise.

O primeiro botão a ter sua utilidade destacada é o botão que faz a troca entre o modo de exibição normal e o modo de edição de marcas de *pitch*. Aparecem duas

opções quando clicamos neste botão. O modo *Label* (abreviação de *Label Marks*) é como visto na figura 2.6 e aqui chamamos de modo Normal. O modo *GCI*s (sigla de *Glottal Closure Instants*) é o modo de edição de marcas de *pitch*; A figura 2.11 mostra a janela de análise (figura 2.6) no modo de edição de marcas de *pitch* (modo *GCI*s).

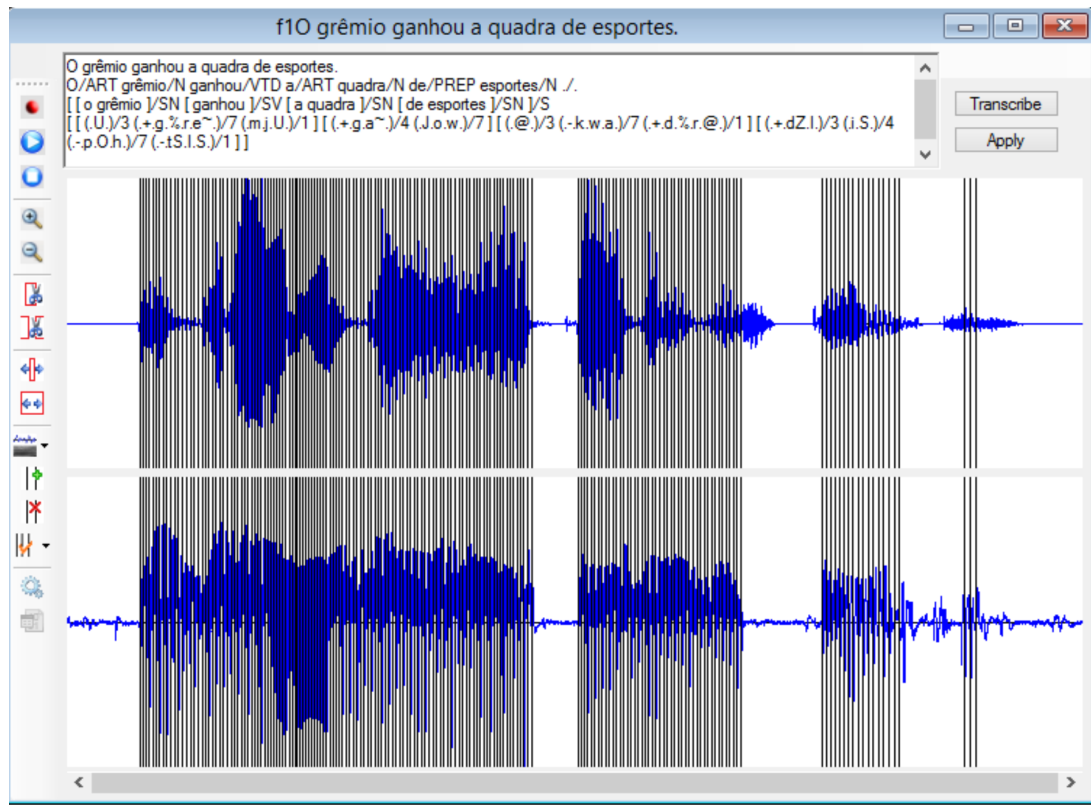


Figura 2.11: Destaque da tela de análise quando o modo GCI (*Glottal Closure Instants*) é selecionado na Barra de Ferramentas.

O segundo botão a ter destaque é o que inicia a janela de manipulação prosódica, descrita em maiores detalhes a seguir.

2.2.3 Janela de Manipulação Prosódica

A janela de manipulação prosódica é ilustrada na figura 2.12.

Essa janela mostra na parte inferior, em três abas, as maneiras possíveis de editar o sinal de fala através das variáveis prosódicas: duração, *pitch* e intensidade.

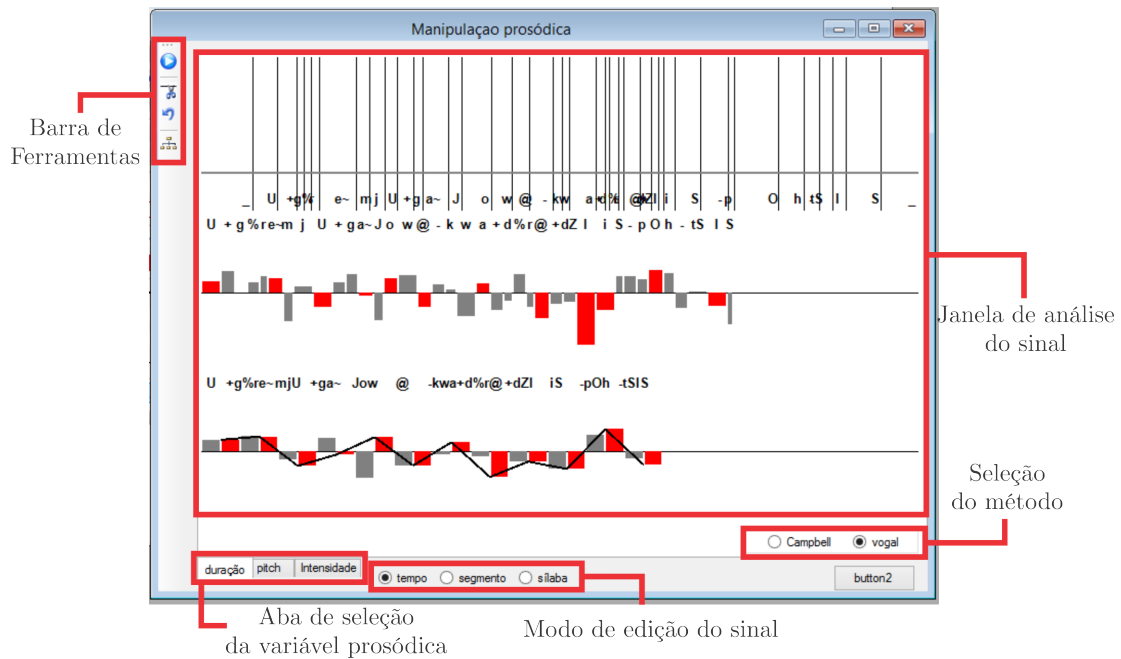


Figura 2.12: Janela de manipulação prosódica que se abre ao clicar no botão Manipulação Prosódica.

Combinada à edição das variáveis de prosódia é possível, ainda, realizar esta edição do sinal por tempo, por segmento ou por sílaba. Estas três possibilidades correspondem às três apresentações do sinal, como vemos na figura 2.13.

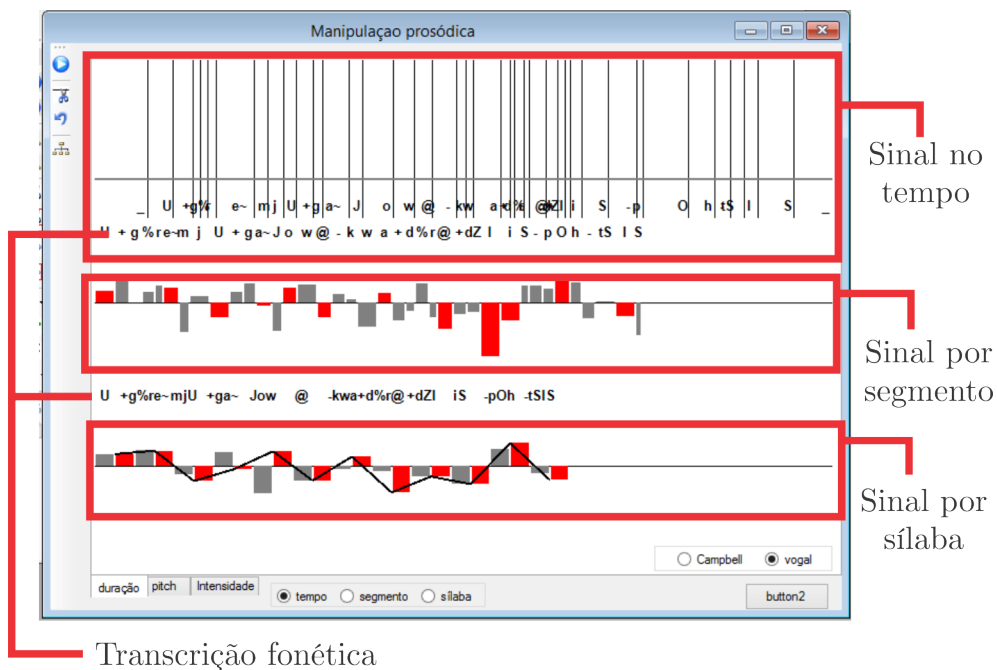


Figura 2.13: Janela de manipulação prosódica dando destaque à três apresentações do sinal: tempo, segmento e sílaba.

Uma vez selecionada a forma de apresentação do sinal, esta é a que será editada pelo usuário de acordo com a variável prosódica pré-escolhida. Para realizar esta edição basta clicar duas vezes acima ou abaixo da linha cinza e utilizar o *mouse* para alterar o sinal correspondente. Em particular, na figura 2.13, vemos a janela de manipulação prosódica para a opção de alteração da duração.

Na figura 2.14 vemos a janela de Manipulação Prosódica com a aba *pitch* selecionada. Nesta aba só é possível ver o sinal no tempo e por sílaba, como aparece em destaque na figura.

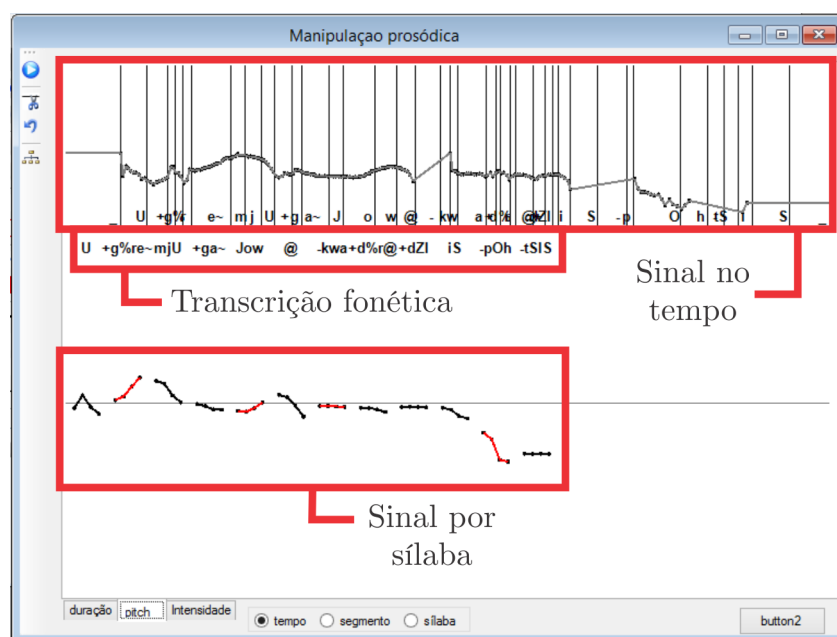


Figura 2.14: Janela de manipulação prosódica com a aba de edição do *pitch* selecionada.

A última aba da janela de Manipulação Prosódica é a de intensidade. Na figura 2.15 vemos um destaque desta janela.

Com isto, passamos por todas as telas, opções e menus do botão Iniciar do sistema, além de uma visão geral da tela inicial. Dando prosseguimento a esta apresentação, passamos ao botão Transcritor da Barra de Menus vista na figura 2.1.



Figura 2.15: Janela de Manipulação Prosódica com a aba de edição da intensidade selecionada.

2.3 Ambiente Transcritor

Para abrir o ambiente Transcritor vamos até a Barra de Menus vista na tela inicial do SASPRO (figura 2.1). O ambiente tem a tela que aparece abaixo na figura 2.16. Na parte superior do Transcritor temos a Barra de Menus com os seguintes botões: Abrir, Salvar, Executar e Limpar. Há ainda duas caixas retangulares logo abaixo dos menus, como vemos na figura. Nestes locais aparecem a sentença a ser transcrita e sua transcrição, respectivamente. Apesar de este capítulo ser apenas uma apresentação rápida, como este ambiente possui apenas uma tela vamos então a uma breve descrição das funções de cada botão da Barra de Menus.

- Abrir: abre uma sentença já existente em seu computador. É possível abrir qualquer arquivo com a extensão `.txt` e editá-lo na própria janela do Transcritor. Também é possível escrever a sentença desejada na caixa retangular logo abaixo do menus. Basta para isso clicar com o botão esquerdo do *mouse* nesta caixa e o cursor já aparecerá dentro da dela. Também é possível editar a transcrição feita, clicando na caixa inferior também com o botão esquerdo do *mouse*.
- Salvar: permite salvar a transcrição da sentença criada e/ou modificada. Por padrão a extensão do arquivo é `.txt`, mas se omitida a extensão o arquivo

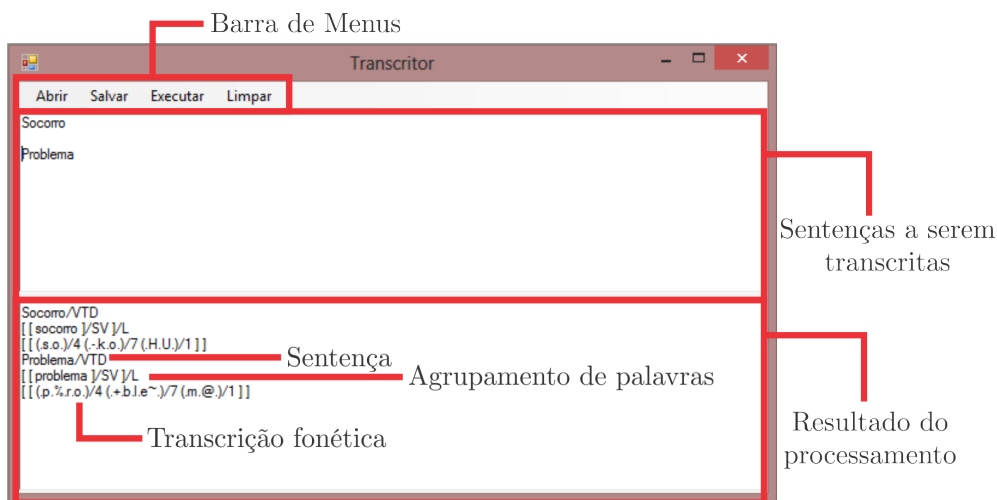


Figura 2.16: Tela Inicial do Transcritor com as sentenças “Socorro” e “Problema” já transcritas.

será salvo como `.phn`. Só é salva a transcrição, não o arquivo original que foi transcrito.

- Executar: executa a transcrição fonética propriamente dita de todas as sentenças contidas na primeira caixa retangular, como podemos ver no exemplo da figura 2.16. Um detalhe importante aqui é que o ambiente ignora linhas em branco. Na figura, pulamos uma linha propositalmente para ilustrar tal fato. Perceba que, na caixa com a transcrição, as sentenças aparecem juntas.
- Limpar: apaga todas as transcrições presentes na tela do transcritor.

Passaremos então agora a uma breve apresentação do ambiente TTS (*Text-To-Speech*).

2.4 Ambiente TTS

Na tela inicial do SASPRO (figura 2.1), na Barra de Menus, clicando no botão TTS do editor de prosódia executamos o sistema protótipo *Text-To-Speech* ou TTS. Ao abri-lo pela primeira vez, vemos a tela que se encontra na figura 2.17.

A figura 2.17 mostra o sistema aberto no modo Lista, mas, independentemente de o SASPRO estar no modo Lista ou no modo Database, o ambiente continua com

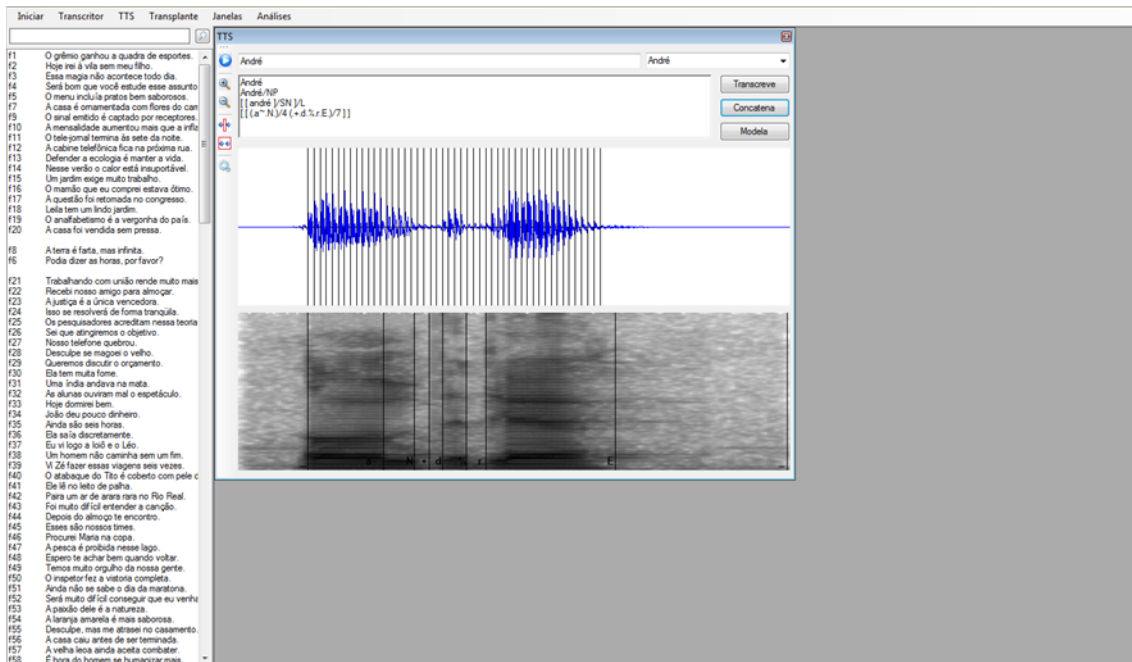


Figura 2.17: Tela inicial do TTS.

as sentenças – no respectivo modo – em destaque na parte esquerda. Vamos então agora a um detalhamento do ambiente.

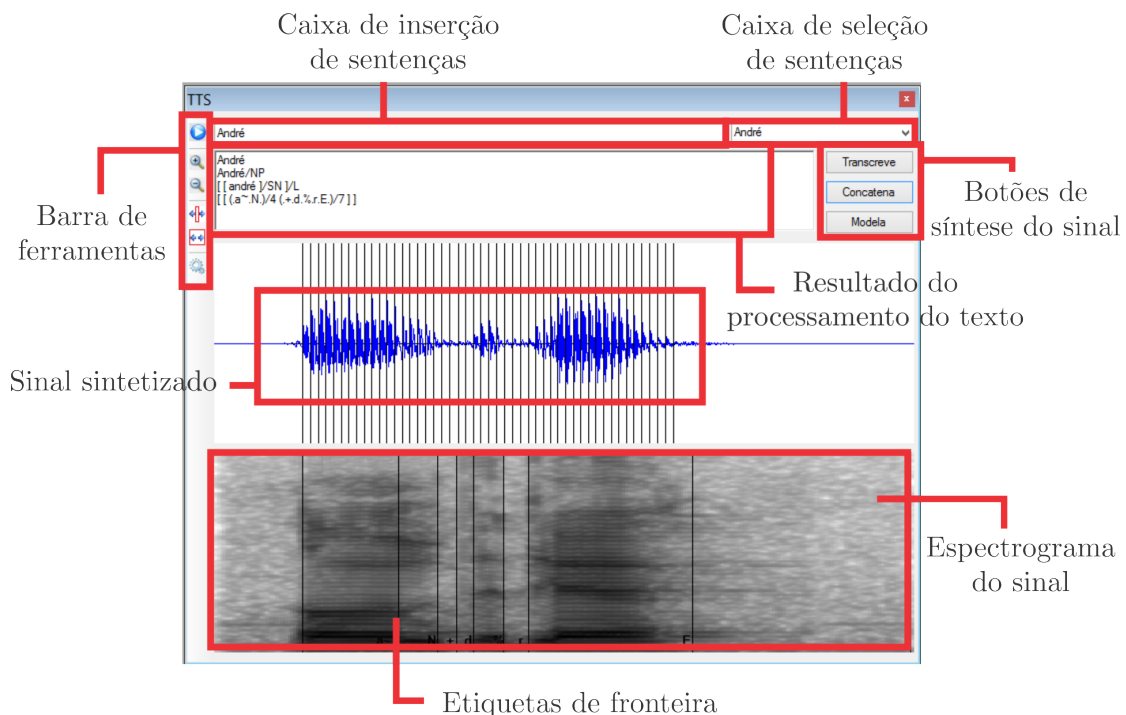


Figura 2.18: Destaque do ambiente exclusivo do TTS.

Na figura 2.18 damos destaque à tela principal do TTS, em que podemos identificar alguns elementos importantes. Na parte esquerda da janela exclusiva do TTS (destacada na figura 2.18) encontramos uma Barra de Ferramentas semelhante ao menu da tela inicial do SASPRO com os seguintes botões (listados de cima para baixo): *Play*, *Zoom In*, *Zoom Out*, *Plot All*, *Plot Selection* e Manipulação Prosódica. As funcionalidades de cada botão serão tratadas no Capítulo 3, mas, apesar disso, trataremos mais adiante nesta seção do botão intitulado Manipulação Prosódica, pois ele dá origem a uma nova janela.

Outros elementos que aparecem nesta janela do TTS também merecem destaque. Na parte superior temos uma caixa de inserção de sentenças e, à direita desta caixa, temos uma lista suspensa de sentenças que podemos selecionar. Abaixo da caixa de inserção de sentenças temos o resultado do processamento da sentença inserida. Dentro desta caixa há quatro linhas; na primeira linha temos a sentença propriamente; na segunda, a classificação morfossintática; na terceira, o agrupamento da palavras; e na quarta, a transcrição fonética.

Do lado direito da caixa que contém o resultado do processamento do texto temos três botões referentes ao processamento e à síntese do sinal: *Transcreve*, *Concatena*, e *Modela*.

Nas duas caixas retangulares da parte inferior temos o sinal sintetizado com as respectivas marcas de *pitch* (caixa superior) e o espectrograma do sinal, juntamente com as etiquetas de fronteira (caixa inferior).

2.4.1 Janela de Manipulação Prosódica TTS

Como vimos, o ambiente TTS possui uma Barra de Ferramentas semelhante à da tela inicial do SASPRO, quando selecionamos alguma sentença (figura 2.6). O último botão desta Barra de Ferramentas abre uma nova janela: a de Manipulação Prosódica, ilustrada na figura 2.19.

Na parte superior do lado esquerdo temos a Barra de Ferramentas cujos botões são (listados de cima para baixo): *Play*, *Delete*, *Reset* e *Toolstripbutton3*. Este

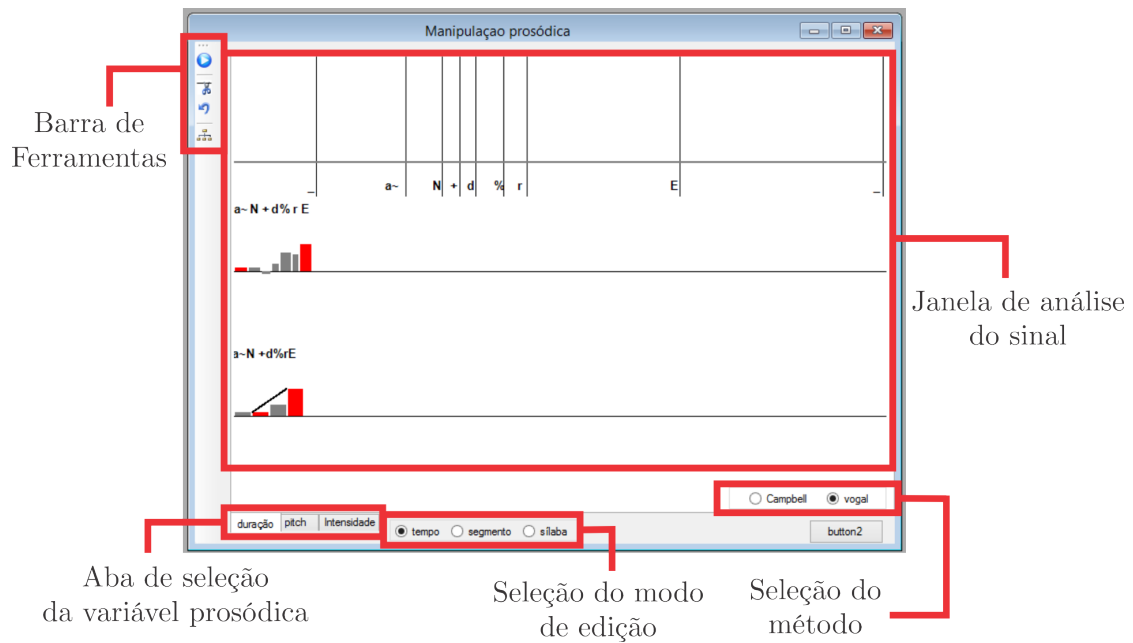


Figura 2.19: Destaque da janela de manipulação prosódica do ambiente TTS.

último botão dá origem a uma nova janela intitulada Estrutura de Performance, que é mostrada na figura 2.20.

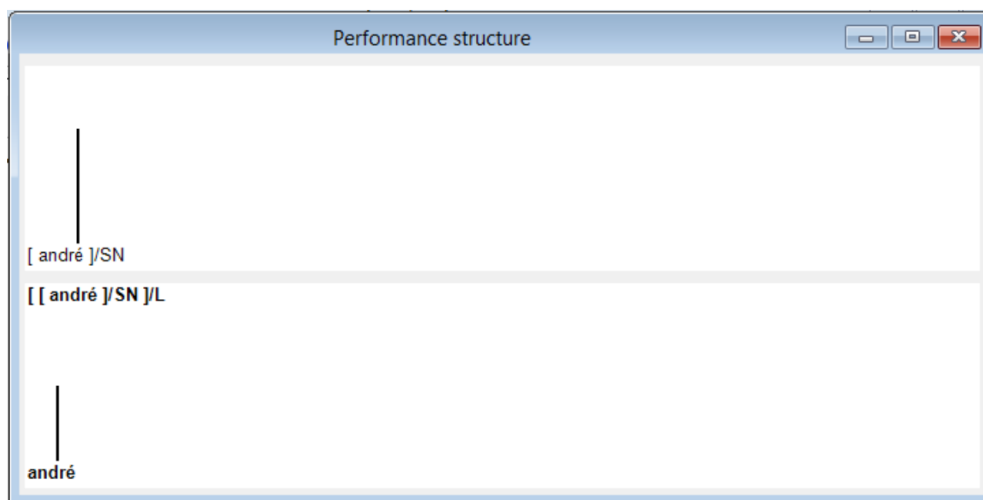


Figura 2.20: Destaque da janela de estrutura de performance, aberta a partir da janela de manipulação prosódica do ambiente TTS.

Voltando à Janela de Manipulação Prosódica (figura 2.19), vemos que na parte inferior há três abas que permitem selecionar a variável prosódica que se quer editar; são elas: duração, *pitch* e intensidade. Quando a janela de Manipulação Prosódica é aberta, a guia “duração” é a que está selecionada; sendo assim, daremos destaque

à ela primeiro e, em seguida, passaremos às demais guias.

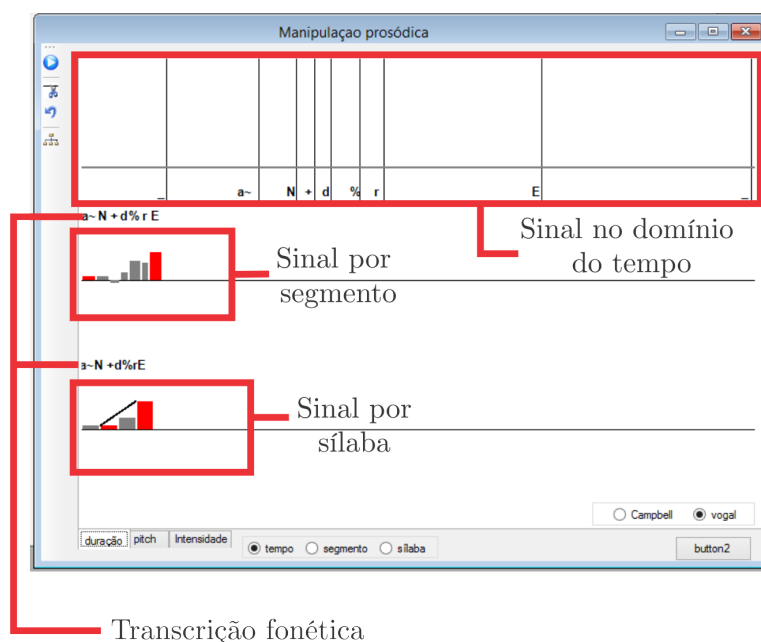


Figura 2.21: Aba Duração da Janela de Manipulação Prosódica do TTS aberta.

Na figura 2.21 vemos que o sinal se apresenta de três formas: no domínio do tempo (parte superior), por segmento (parte central) e por sílaba (parte inferior). Esta janela se comporta de maneira semelhante à janela vista na figura 2.13. Isto quer dizer que uma vez selecionada a forma de edição – tempo, segmento ou sílaba – só conseguimos modificar a variável prosódica – duração, *pitch* e intensidade – naquele modo. As próximas duas figuras (2.22 e 2.23) dão destaque às abas *pitch* e intensidade, respectivamente.

Com isso encerramos a descrição do ambiente do TTS. Passaremos então ao próximo botão na Barra de Menus da tela inicial: Ambiente Transplante (figura 2.1).

2.5 Ambiente Transplante

Vamos agora a uma breve descrição do Ambiente Transplante. Antes de abrir o ambiente de transplante, é preciso selecionar uma sentença na tela inicial do SASPRO (figura 2.1). Feito isso, ao clicar no botão Transplante no menu principal do SASPRO (figura 2.1) vemos a tela que aparece na figura 2.24. Nesta figura, logo

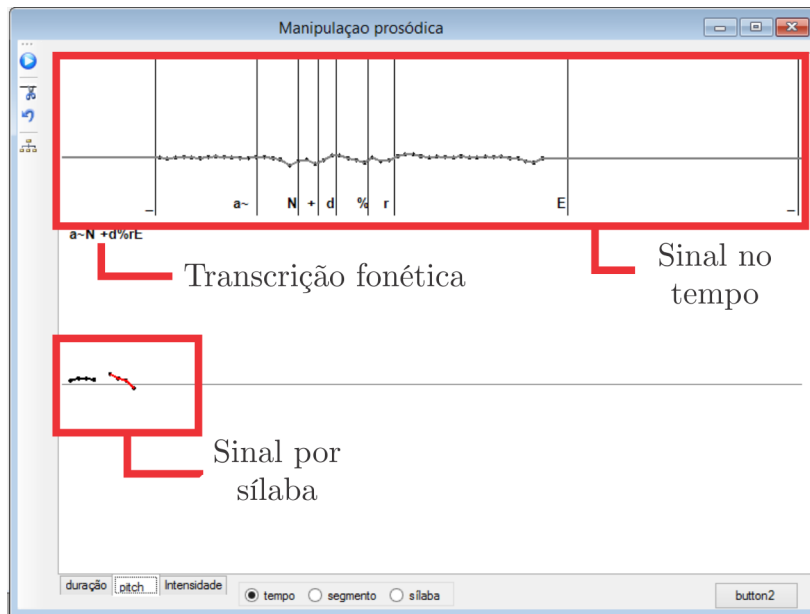


Figura 2.22: Aba *Pitch* da janela de manipulação prosódica do TTS aberta.

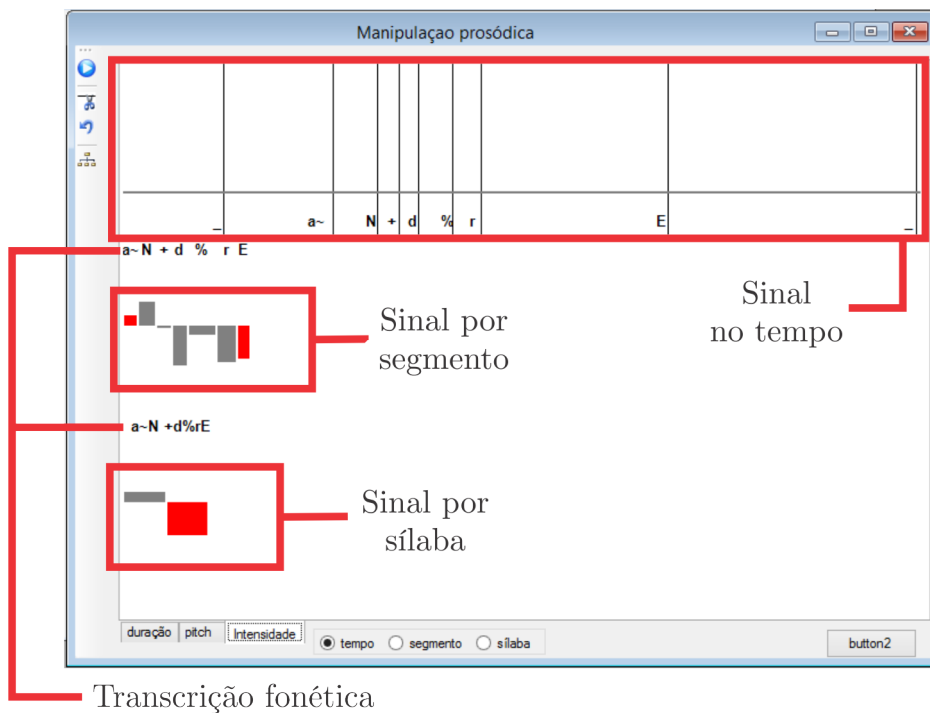


Figura 2.23: Aba *Intensidade* da janela de manipulação prosódica do TTS aberta.

abaixo da Barra de Ferramentas temos duas caixas retangulares separadas por uma lista suspensa de sentenças. A caixa superior mostra o sinal de teste juntamente com a energia do sinal. A caixa inferior mostra o sinal de referência também com a respectiva energia. A lista suspensa entre as duas caixas mostra o sinal de referência selecionado.

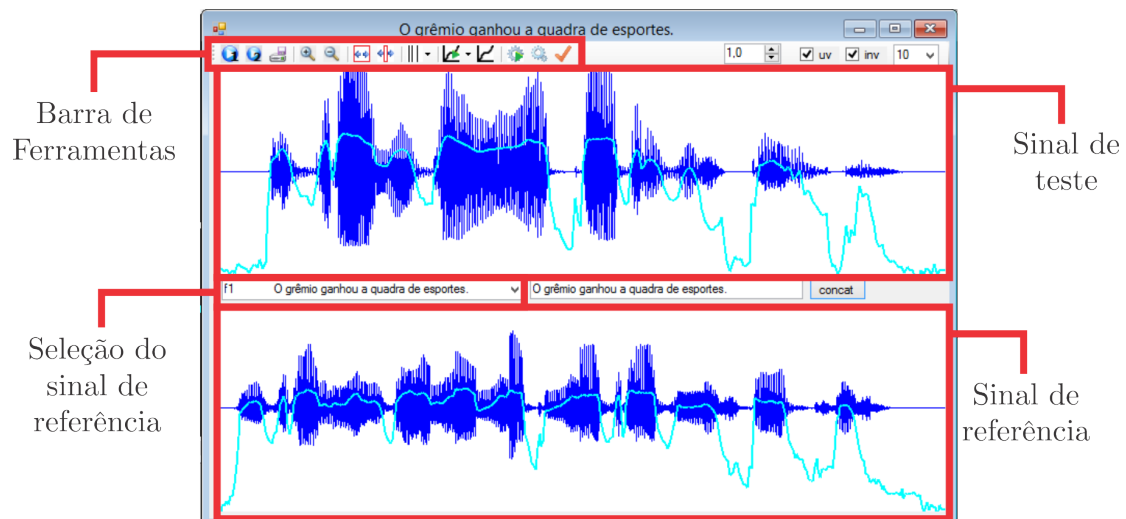


Figura 2.24: Tela inicial do ambiente de Transplante de Prosódia.

Na parte superior temos a Barra de Ferramentas com os botões (listados da esquerda para a direita) *Play*, *Play Sint*, *Save*, *Zoom In*, *Zoom Out*, *Plot All*, *Plot Selection*, *Edit* – que possui as opções *GCI* e *Label*, *ToolStripSplitButton1*, *Plot DTW*, *ToolStripButton2*, *Manipulation*, *Button Model*. Alguns destes botões dão origem a novas telas. O botão *Edit* quando selecionado na opção *Label* dá origem à tela que aparece na figura 2.25.

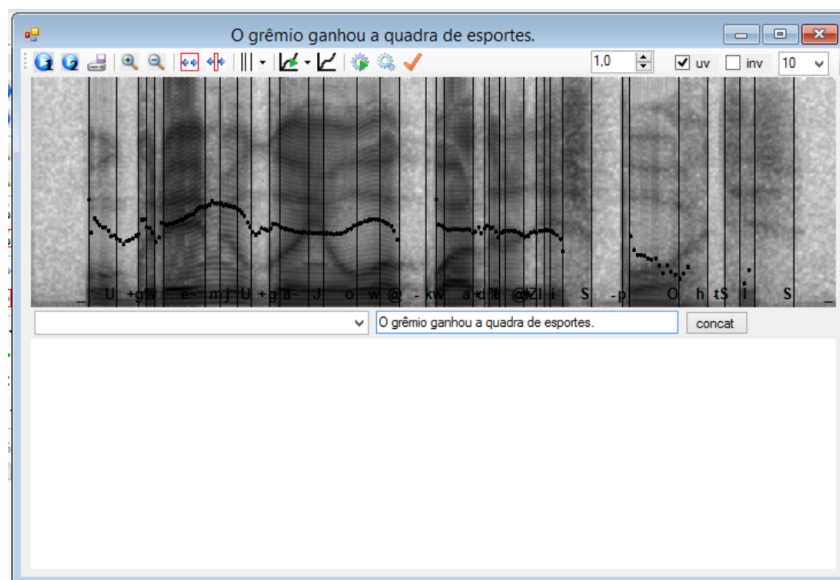


Figura 2.25: Ambiente de Transplante com a opção *Label* do botão *Edit* da Barra de Ferramentas ativado.

Na figura 2.26 damos destaque ao submenu que aparece do botão *ToolStripSplitButton1*.

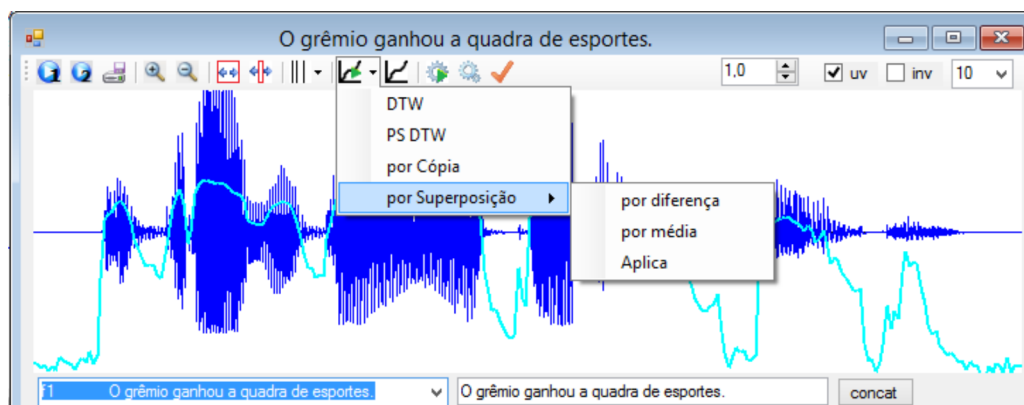


Figura 2.26: Destaque ao submenu do botão *ToolStripSplitButton1* da Barra de Ferramentas do Ambiente de Transplante.

O botão *Manipulation* abre uma tela muito semelhante às vistas anteriormente com o acionamento do botão de Manipulação Prosódica no ambiente inicial do SAS-PRO (figura 2.13) e também no ambiente TTS (figura 2.19). O último botão da Barra de Ferramentas é o *ButtonModel*, que dá origem à tela que aparece na figura 2.27 e mostra os parâmetros do modelo de prosódia.

2.6 Janelas

Este é o penúltimo botão da Barra de Menus do sistema. Ele dá origem a um submenu que é mostrado na figura 2.28. Basicamente ele tem a função de mostrar/organizar/fechar as janelas abertas. A janela que está selecionada é marcada com um símbolo do lado esquerdo do nome da janela.

2.7 Análises

O último botão da Barra de Menus é o botão Análises. Ao clicar neste botão, temos o submenu que se apresenta na figura 2.29.

Por uma questão de organização, as funcionalidades serão tratadas em capítulos posteriores.

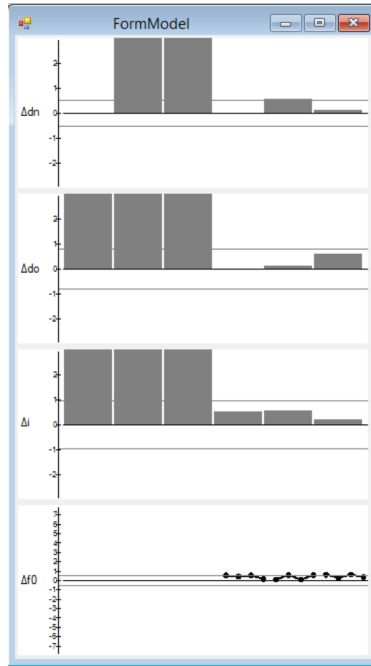


Figura 2.27: Destaque da tela gerada pelo botão *ButtonModel* da Barra de Ferramentas do ambiente de Transplante com os parâmetros do modelo de prosódia.

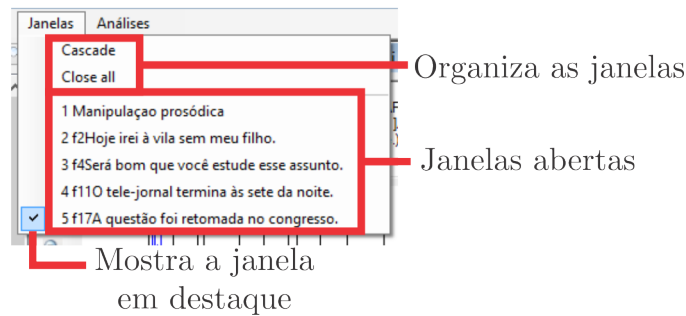


Figura 2.28: Destaque do botão Janelas da Barra de Menus do SASPRO.

2.8 Considerações Finais

Neste capítulo procuramos descrever com o máximo de detalhamento e clareza ao nosso alcance os ambientes, menus e telas do Editor de Prosódia visando a dar uma visão que fosse a mais global possível do sistema.

No próximo capítulo descrevemos, de maneira mais incisiva, as funcionalidades para dar ao usuário uma facilidade no uso deste aplicativo. Para que haja maior clareza no entendimento destas funcionalidades, optamos por agrupá-las em três grandes grupos: Entrada/Saída de Dados, Análise de Dados e Síntese de Dados. Como o SASPRO possui, muitas vezes, funções semelhantes em suas diversas partes,

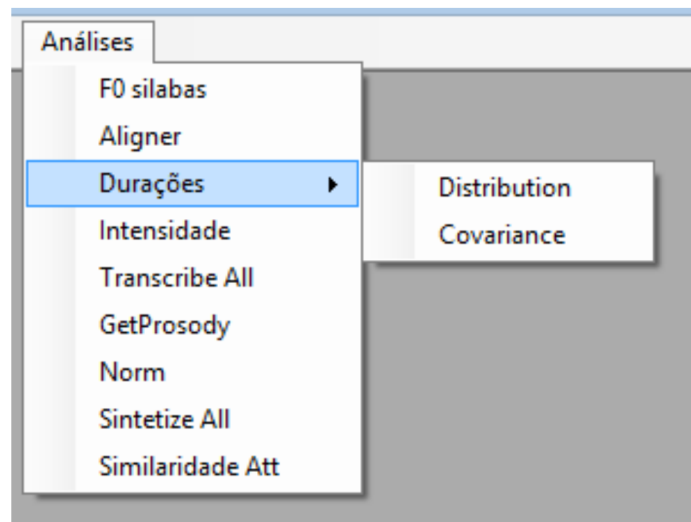


Figura 2.29: Destaque do botão Análises da Barra de Menus do SASPRO.

parece razoável agrupar estas funções por finalidade. É o que veremos no capítulo a seguir.

Capítulo 3

Funcionalidades

3.1 Introdução

No Capítulo 2 vimos quais são, de maneira geral, as funcionalidades que o SAS-PRO oferece e conhecemos sua interface através de suas telas, botões, menus e submenus. Neste capítulo procuramos agrupar estas funcionalidades por finalidade, dividindo-as em três grupos: Entrada/saída de dados, análise de dados e síntese de dados. Desta forma é possível perceber, dependendo da finalidade de qual(is) desta(s) funcionalidade(s) o usuário necessita e, usando o Capítulo 2, onde, no SAS-PRO ela se encontra.

Agora, mais do que apresentar, queremos olhar cada funcionalidade, explicando como opera e, sempre que possível, dando exemplos de utilização para facilitar o entendimento.

Em termos de sequência de apresentação, o que faremos agora é percorrer todo o sistema na mesma ordem do Capítulo 2, mas dando ênfase a cada grupo separadamente.

3.2 Entrada/Saída de Dados

Antes de tudo vamos tentar definir o que estamos chamando de entrada de dados. Aqui, entendemos por entrada de dados toda intervenção feita pelo usuário, digitando um texto, ou simplesmente copiando-o e colando-o em algum lugar do

sistema. Além disso, entendemos também como entradas de dados todo arquivo que tenha sido aberto e usado no SASPRO, tal como arquivos de texto ou de áudio.

Por saída de dados entendemos qualquer arquivo gerado pelo SASPRO, possa ou não este arquivo ser salvo pelo usuário.

3.2.1 Ferramenta de Busca

Na tela inicial do SASPRO (figura 2.1) a primeira maneira de realizar uma entrada de dados é através da ferramenta de busca. Ela aparece logo abaixo da Barra de Menus. Ao clicar nela com o botão esquerdo do *mouse* um cursor de texto aparece dentro da caixa de busca e é possível digitar o termo ou sentença que se quer buscar.

Após inserido o texto, para utilizar a ferramenta de busca é preciso clicar no botão com uma lupa do lado direito. Na figura 3.1 buscamos pelo termo “magia”.

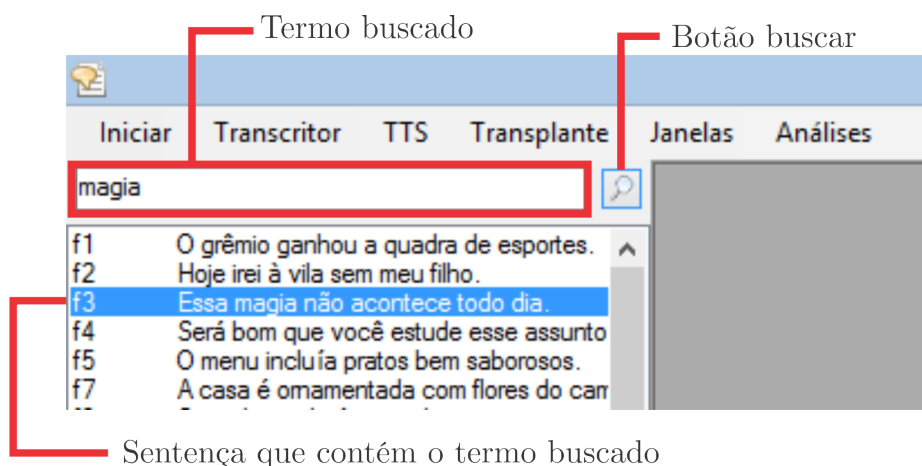


Figura 3.1: Exemplo de utilização da caixa de busca.

A ferramenta de busca percorre as sentenças começando da sentença selecionada e indo até a última. Então, o ideal é sempre selecionar a primeira sentença antes de iniciar uma busca, pois o sistema não retorna para a primeira sentença (O grêmio ganhou a quadra de esportes) ao atingir a última pré-gravada. Ao se encontrar o termo buscado, a sentença em que o termo se encontra fica em destaque.

Algumas observações sobre o funcionamento e utilização da caixa de busca:

- A caixa não funciona apertando-se o botão *Enter* do teclado. É preciso clicar no botão buscar (figura 3.1).
- A busca diferencia minúsculas e maiúsculas, desta forma as buscas por “menu”, “Menu” e “MENU” terão resultados diferentes.
- Palavras com acento e cedilha devem ser buscadas exatamente como a grafia está no sistema. Portanto “mamão” e “mamao” têm resultados de busca diferentes.
- Lembrar sempre que a caixa de busca não retorna ao início após atingir o final do banco de sentenças. Assim, se a palavra ou expressão buscada está na terceira sentença do banco e a busca começou na quinta o sistema não passará pela primeira. Isto é remediado começando-se a busca clicando na primeira sentença.
- Se for utilizado mais de um termo para a busca, ele deve estar na mesma sequência que a sentença existente. Por exemplo, “próxima rua” e “rua próxima” têm resultados distintos.
- A busca não identifica mais de um termo separado por outro termo. Por exemplo, existe a sentença “Essa magia não acontece todo dia.”, mas se digitarmos “magia dia” a busca não retorna nenhum resultado.

3.2.2 Botão Iniciar

Ao abrir o SASPRO vemos a tela figura 2.1. Vimos que, ao clicar em alguma sentença da lista de sentenças do lado esquerdo, uma janela de análise se abre no ambiente principal (figuras 2.5 e 2.6). Do lado esquerdo vimos que há uma Barra de Ferramentas. O primeiro botão desta barra é o botão *Record* (Gravar). Ao pressionar este botão o sistema passa a capturar o áudio vindo do microfone interno do computador (caso exista, como em um *notebook*, por exemplo) ou de um microfone externo. Na figura 3.2 damos o exemplo da sentença “O grêmio ganhou a quadra de esportes.” gravada via microfone externo por meio do botão *Record*. Vale a pena comparar a figura 3.2 com o sinal da figura 2.6.

Para terminar a gravação, basta pressionar o botão *Stop*. Automaticamente o sinal gravado aparecerá na caixa retangular superior como visto na figura 3.2. Para ouvir o sinal gravado basta apertar o botão *Play* da Barra de Ferramentas. Não é possível apertar o botão *Play* durante a gravação.

Ao fechar a janela aparece uma caixa de diálogo perguntando se o usuário deseja salvar o arquivo gravado por meio do botão *Record*. Ao clicar em “Sim” o sistema sobrescreverá a sentença selecionada no banco de sentenças à esquerda; portanto, é necessário certo cuidado quando se utiliza esta funcionalidade, para não sobrescrever indevidamente algum arquivo.

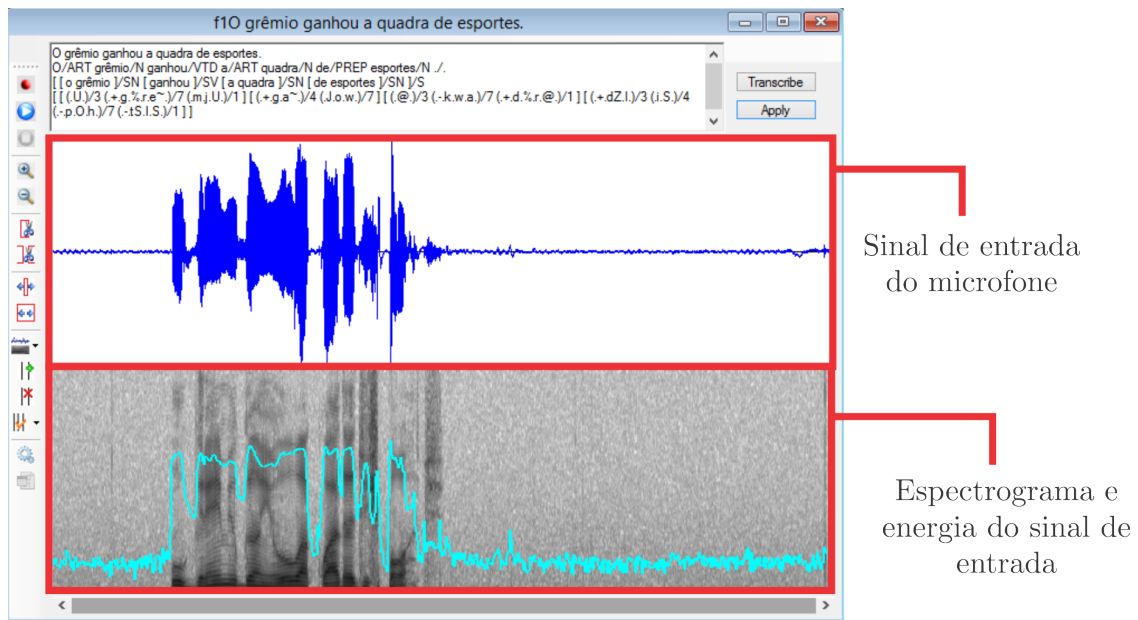


Figura 3.2: Exemplo de utilização do botão *Record* da Barra de Ferramentas. A sentença utilizada como exemplo foi “O grêmio ganhou a quadra de esportes.”.

3.2.2.1 Modo Lista

Continuando com as maneiras de entrada de dados, temos o botão Iniciar da Barra de Menus dos SASPRO. Nele temos duas opções: Lista e Database (figura 2.2). Ao clicarmos em Lista, uma caixa de seleção se abre para que selecionemos o arquivo *.txt* desejado. Este arquivo *.txt* segue padrões pré-estabelecidos de que trataremos em capítulos posteriores.

3.2.2.2 Modo Database

Ainda no botão Iniciar da Barra de Menus temos a opção Database, que, como vimos no capítulo anterior tem algumas subopções (Abrir, Salvar, Importar Database e Compilar Database). Ao clicar em Abrir uma janela de seleção se abre para que seja escolhido o arquivo `.txt` desejado. As demais funções serão tratadas posteriormente.

3.2.3 Botão Transcritor

Ao clicar no botão Transcritor da Barra de Menus acessamos o ambiente Transcritor, como visto na figura 2.16. A maneira mais simples de entrar com dados neste ambiente é clicando na caixa retangular superior com o botão esquerdo do *mouse* e digitando a sentença desejada. Fizemos isso no capítulo anterior com as palavras “Socorro” e “Problema”. É possível inserir sentenças longas ou apenas palavras. Além disso, o Transcritor identifica as quebras de linha, mas ignora linhas em branco em termos de processamento (figura 2.16).

Para abrir uma sentença usando o ambiente, vá até a Barra de Menus e clique em Abrir. Uma janela de seleção se abrirá e basta escolher o arquivo desejado.

Depois de realizada uma transcrição usando o ambiente transcritor, é possível salvar um arquivo no formato `.txt`. O arquivo a ser salvo já vem com um nome padrão `textout.txt`, mas caso queira basta renomear e salvar na pasta do interesse do usuário. Uma observação é que o SASPRO não salva a sentença original, somente a transcrição da sentença. Segue então um exemplo de utilização do ambiente Transcritor.

Abra o SASPRO e vá até Barra de menus (figura 2.1). Clique em Transcritor. Ao abrir a janela do ambiente clique com o botão esquerdo do *mouse* no retângulo superior e insira uma sentença. Aqui utilizamos a sentença “Socorro, preciso resolver isso!”. Clique em Executar na Barra de Menus do ambiente. A tela parecerá com a figura 3.3.

Para salvar, vá ao botão Salvar da Barra de Menus. Uma janela de seleção aparecerá para que se escolha o local e o nome do arquivo a ser salvo. Na figura 3.4

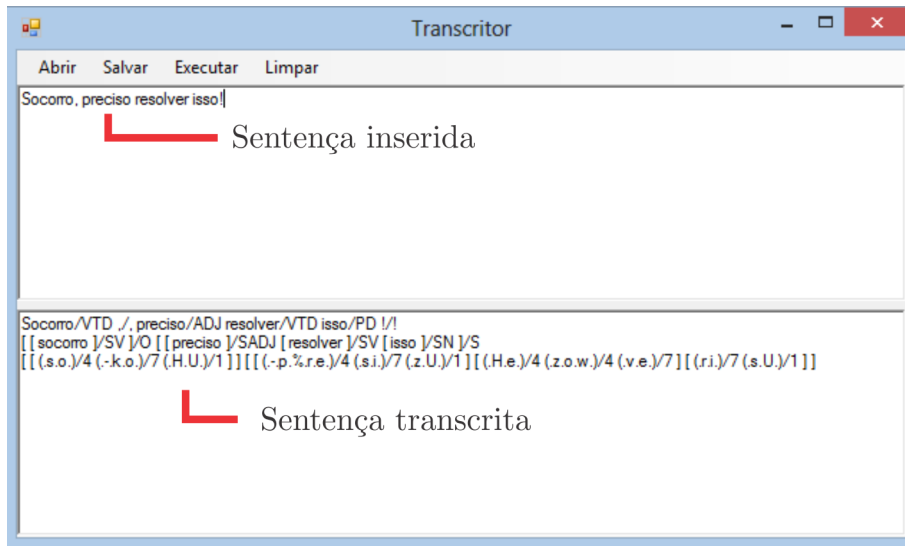


Figura 3.3: Exemplo de utilização do ambiente Transcritor. A sentença utilizada como exemplo foi “Socorro, preciso resolver isso!”.

usamos o nome `Socorro.txt`.

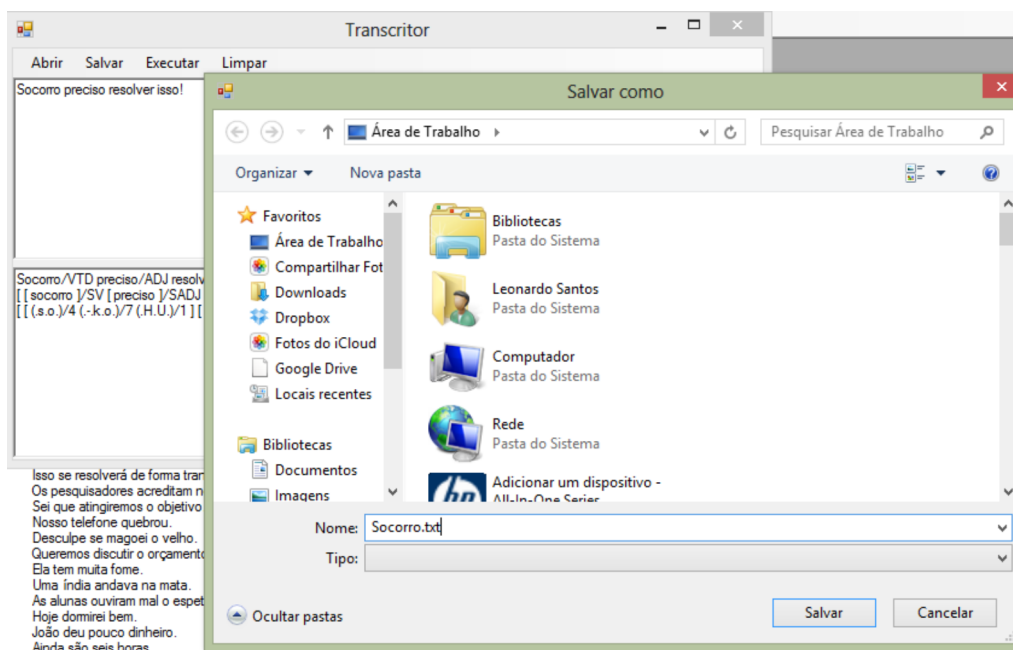


Figura 3.4: Salvando um arquivo usando o ambiente Transcritor.

Seguem algumas observações sobre o uso do Transcritor:

- Caso não seja escolhida a extensão para o arquivo, o SASPRO dá como padrão a extensão `.phn`.

- Os arquivos salvos a partir do Transcritor podem ser abertos no bloco de notas do *Windows*TM, tenham eles extensão `.txt` ou `.phn`.
- Algumas sentenças que possam conter símbolos, como *emails* ou números de telefones, por exemplo, podem gerar problemas e fazer com que o SASPRO apresente uma mensagem de erro.
- Atalhos de teclado como `Ctrl+X`, `Ctrl+V`, etc. funcionam normalmente na caixa retangular de entrada de sentença do Transcritor.

3.2.4 Botão TTS

No capítulo anterior, vimos que ao clicar no botão TTS da Barra de Menus temos a tela que aparece na figura 2.17. Nesta tela há uma caixa retangular na parte superior onde é possível inserir uma sentença clicando com o botão esquerdo do *mouse* e digitando a sentença desejada.

Na figura 3.5 utilizamos a sentença “Estou quase terminando o projeto...”. A entrada foi feita via teclado e, para sintetizar, utilizamos os botões Transcreve, Concatena e Modela (ferramentas de síntese), que estão à direita do retângulo que contém o resultado do processamento do texto. Depois de digitada a sentença, basta clicar em Transcreve e já será possível visualizar o sinal (em azul) com marcas de *pitch* e também ouvi-lo. No retângulo inferior temos o espectrograma do sinal com as etiquetas de fronteira (em cinza).

Ao clicar em Modela, é inserida a prosódia no sinal. Mais uma vez será possível visualizar o sinal (em azul) e seu espectrograma (no retângulo inferior em cinza); no entanto, para ouvi-lo é preciso clicar no botão *Play* da Barra de Ferramentas do lado esquerdo da figura 3.5.

Além da entrada via teclado, existe a possibilidade de inserir sentenças por meio da lista suspensa do lado direito na figura 3.5. Neste caso, também é possível seguir a sequência Transcreve→Concatena→Modela, mas deste modo a única etapa que surtirá efeito é a de concatenação, pois a sentença da lista já foi previamente transcrita e concatenada, diferentemente do caso de uma sentença entrada via teclado.



Figura 3.5: Entrada de dados no ambiente TTS por meio do teclado. A sentença usada foi “Estou quase terminando o projeto. . .”.

3.2.5 Botão Transplante

Para abrir o ambiente Transplante corretamente, primeiro devemos selecionar uma sentença do lado esquerdo da tela inicial (figura 2.1). Feito isso devemos ir até a Barra de Menus e clicar no botão Transplante. Na figura 3.6, selecionamos a sentença “Podia dizer as horas, por favor?”

A entrada de dados pode ser feita de duas maneiras. Primeiramente, por meio da caixa retangular ao lado do botão *concat*; para isto, basta clicar com o botão esquerdo do *mouse* e digitar a sentença desejada; Alternativamente, a entrada pode ser feita através da lista suspensa destacada na figura 3.6; nesse caso, ao selecioná-la, observamos que aparece na caixa retangular da direita ao lado do botão *concat*. Um exemplo de uso destas duas maneiras esclarece melhor.

Abra o SASPRO e selecione qualquer sentença do banco de sentenças do lado esquerdo (figura 2.1). Vá até a Barra de Menus e clique no botão Transplante. Com o ambiente Transplante aberto, teremos uma tela semelhante à da figura 3.7, porém

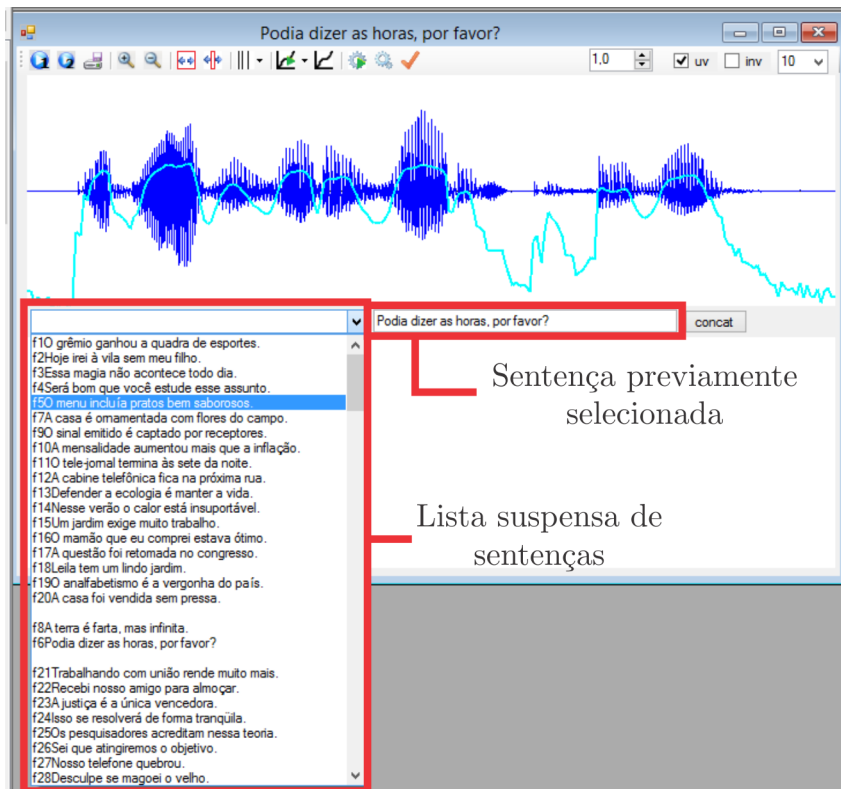


Figura 3.6: Entrada de dados no ambiente Transplante. A sentença usada foi “Pode dizer as horas, por favor?”.

com a sentença escolhida pelo usuário. Aqui escolhemos previamente a sentença “A casa foi vendida sem pressa.”.

Ao clicar no botão com um símbolo de *Play* com o número 1 na Barra de Ferramentas do ambiente, é possível ouvir a sentença original. Basta agora clicar no botão *concat* e teremos uma tela semelhante à da figura 3.8 em que o retângulo inferior contém o sinal sintetizado, e é possível ouvi-lo clicando no botão com um símbolo de *Play* com o número 2 na Barra de Ferramentas do ambiente.

Partindo da tela mostrada na figura 3.8, faremos a entrada de dados via teclado. Para isto, basta clicar com o botão esquerdo do *mouse* na caixa retangular ao lado do botão *concat*. Na figura 3.9 inserimos a sentença “O Botafogo quase chegou lá!”. Depois basta clicar no botão *concat* e o sinal sintetizado aparecerá na caixa retangular inferior.

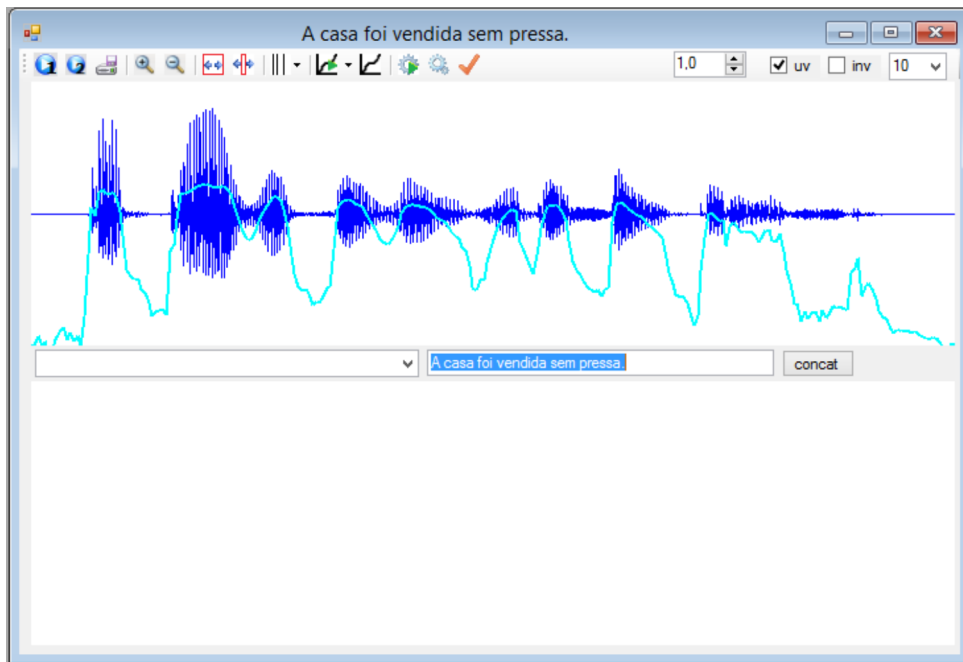


Figura 3.7: Entrada de dados no ambiente Transplante com sentença previamente selecionada. A sentença usada foi “A casa foi vendida sem pressa.”.

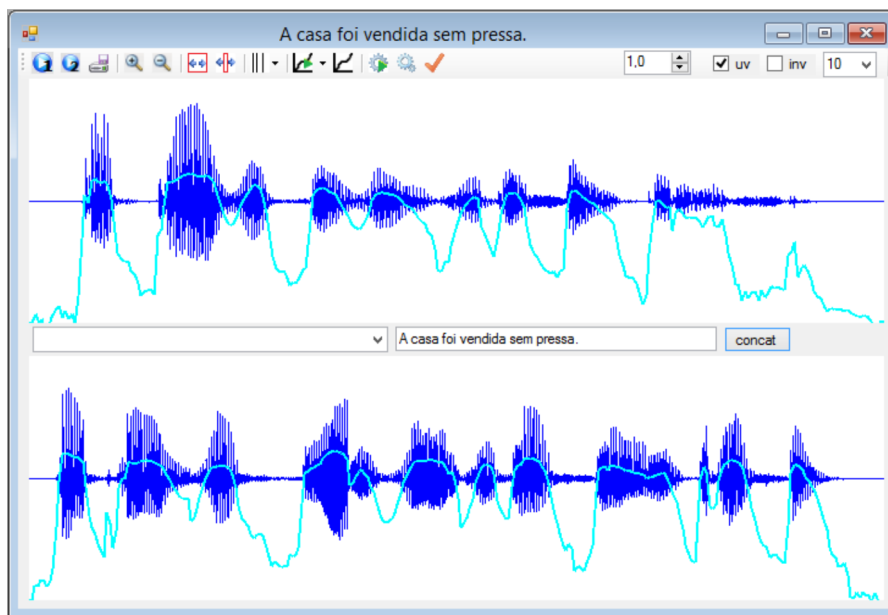


Figura 3.8: Sentença sintetizada a partir de outra previamente selecionada.

Ainda na mesma tela, vejamos a última forma de entrar com uma sentença, que é por meio da lista suspensa mostrada na figura 3.6. Basta clicar na seta ao lado da caixa retangular e selecionar uma das sentenças. Na figura 3.10 selecionamos a sentença “Sei que atingiremos o objetivo.”.

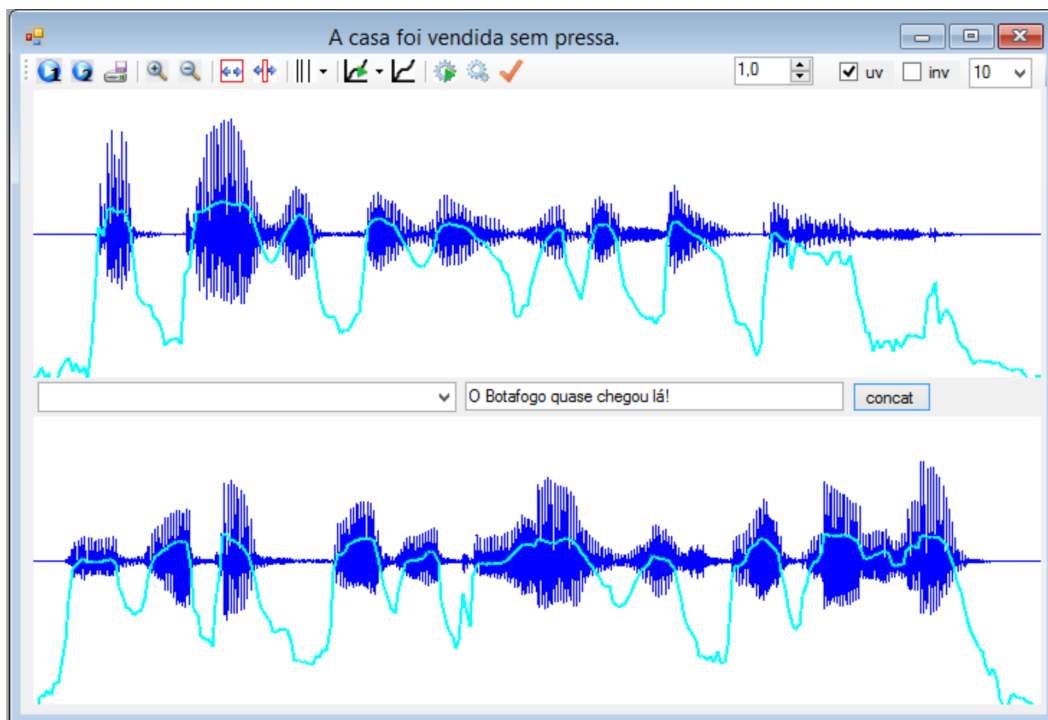


Figura 3.9: Sentença sintetizada a partir de uma entrada de dados via teclado.

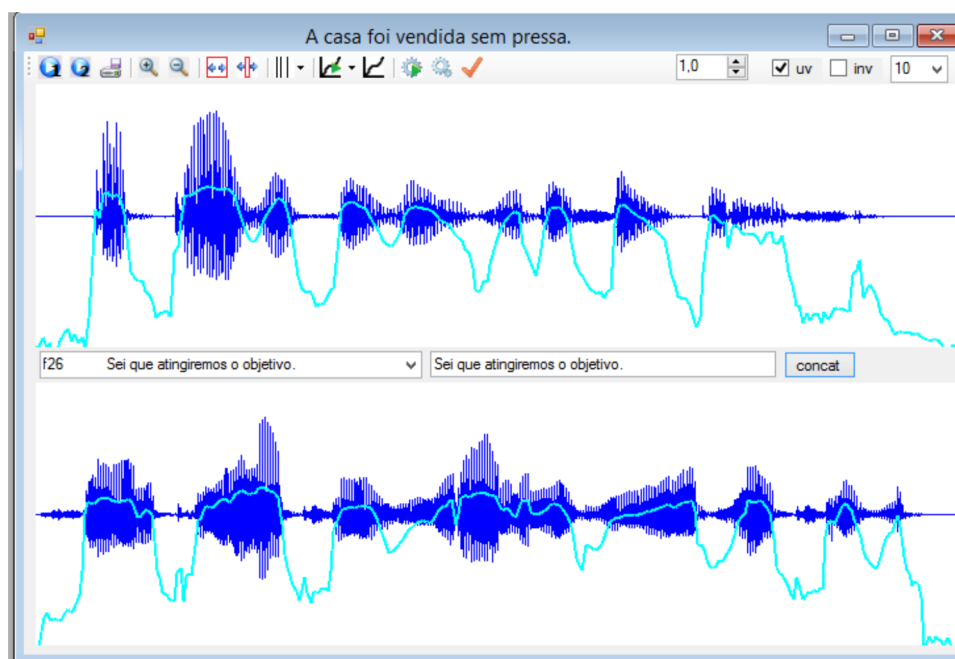


Figura 3.10: Sentença sintetizada a partir da lista suspensa.

Isto finaliza a seção de entrada de dados, passaremos agora a ver as formas de análise de dados presentes no SASPRO. Este é o tema da próxima seção deste capítulo.

3.3 Análise de Dados

Antes de começar a listar as maneiras de analisarmos dados no SASPRO, é importante tentar esclarecer o que chamamos aqui de Análise de Dados. Estamos nos referindo apenas à observação de resultados, isto é, uma vez realizada a operação de síntese/edição não há mudança em parâmetros do sinal/sentença, apenas análise/observação de suas características, seja por meio de gráficos, tabelas e/ou dados numéricos. Sendo assim, são exemplos de funcionalidades de Análise de Dados presentes no SASPRO: visualização do processamento do texto e do sinal propriamente dito, visualização da energia do sinal, das etiquetas de fronteira, do espectrograma do sinal e *zoom*. Há outras ferramentas de análise dentro do ambiente de manipulação prosódica e assim por diante; e como dito inicialmente, o objetivo desta seção é “percorrer” estas ferramentas.

3.3.1 Tela Inicial

Na tela inicial do SASPRO (figura 2.1) temos uma tela semelhante à da figura 3.11. Vamos voltar a esta tela agora, porém dando ênfase às ferramentas de análise.

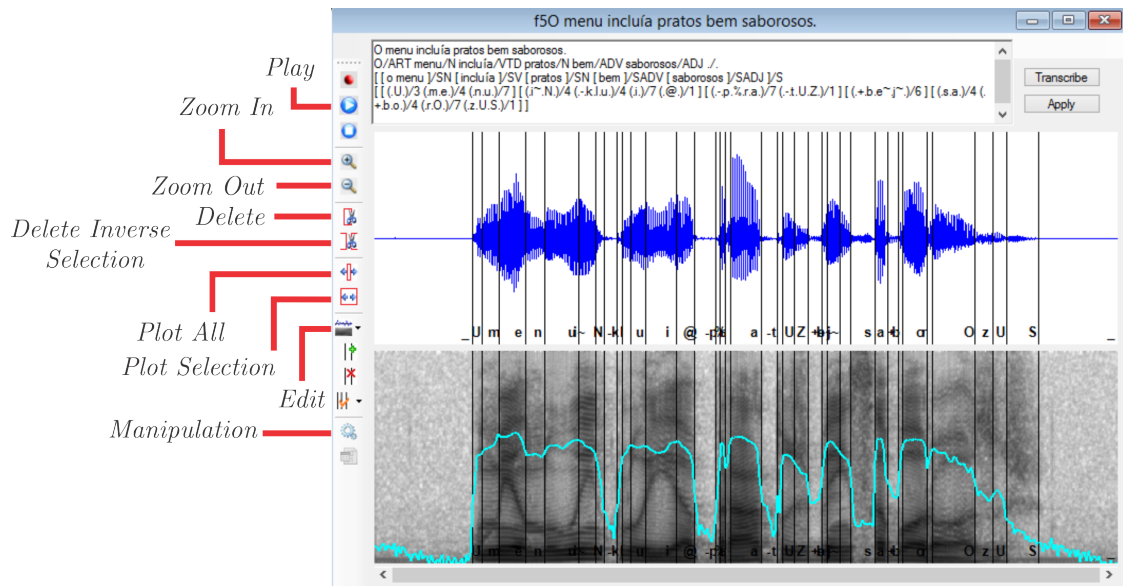


Figura 3.11: Tela inicial do SASPRO com a sentença “O menu incluía pratos bem saborosos.”, dando destaque a Barra de Ferramentas.

Como vimos no Capítulo 2, o primeiro retângulo contém a classificação morfosintática, o agrupamento de palavras e a transcrição fonética (figura 2.7); o retângulo central contém o sinal, as etiquetas de fronteira e a transcrição fonética na parte inferior (figura 2.8); e na parte inferior, temos o espectrograma e a energia do sinal (figura 2.9). Na parte esquerda da figura 3.11, vemos uma Barra de Ferramentas vertical. Vamos agora a uma descrição mais detalhada das funções de cada botão.

- *Play*: Reproduz o sinal mostrado na caixa retangular superior. Caso tenha sido feito algum recorte no sinal (veja os botões *Delete* e *Delete Inverse Selection* desta lista), será reproduzida apenas a parte vista no retângulo superior.
- *Zoom In/Out*: Amplia/diminui o sinal sem modificar a janela de visualização, ou seja, diminui/aumenta a escala de modo que seja possível ver uma parte menor/maior do sinal, porém com mais/menos detalhes.
- *Delete*: Apaga a parte selecionada do sinal dentro do retângulo superior. Para selecionar uma parte do sinal basta clicar e segurar com o botão esquerdo do *mouse* arrastando até o ponto desejado (figura 3.12). Depois basta clicar no botão. Caso nada seja selecionado, o botão *Delete* apagará todo o sinal.
- *Delete Inverse Selection*: Possui a mesma função que o *Delete*, porém apaga a parte não selecionada do sinal. No caso de nenhuma parte do sinal estar selecionada, o acionamento do botão apaga as etiquetas de fronteira.
- *Plot All*: Mostra no retângulo de visualização todo o sinal em questão. Caso tenha sido usado o botão *Zoom* anteriormente, este comando somente retornará ao ponto de última utilização do mesmo. Funciona como um *reset* do botão *Plot Selection*.
- *Plot Selection*: Funciona como uma espécie de *zoom* local, ampliando somente a área selecionada. Para usá-lo basta selecionar com o *mouse*, clicando e arrastando pela área que deseja visualizar. Um retângulo vermelho será desenhado (figura 3.12), e depois basta clicar no botão.
- *Edit*: Permite mudar modo de visualização para a janela que mostra as marcas de *pitch* (modo GCI). Esta tela já foi mostrada na figura 2.11.

- *Manipulation*: Abre a janela de manipulação prosódica. Esta janela é vista na figura 2.12. Em relação à análise, não há o que acrescentar ao que já foi visto. Vamos então falar apenas sobre os botões que estão na Barra de Ferramentas na parte superior esquerda desta janela.
 - *playModif*: reproduz o sinal com todas as modificações feitas dentro desta janela.
 - *delete*: neste módulo este botão fica desativado.
 - *reset*: retorna o sinal às suas condições iniciais.
 - *Performance Structure*: abre a janela de análise da estrutura de performance (figura 2.20).

A figura 3.12 mostra uma sentença com parte dela selecionada pelo *mouse*.

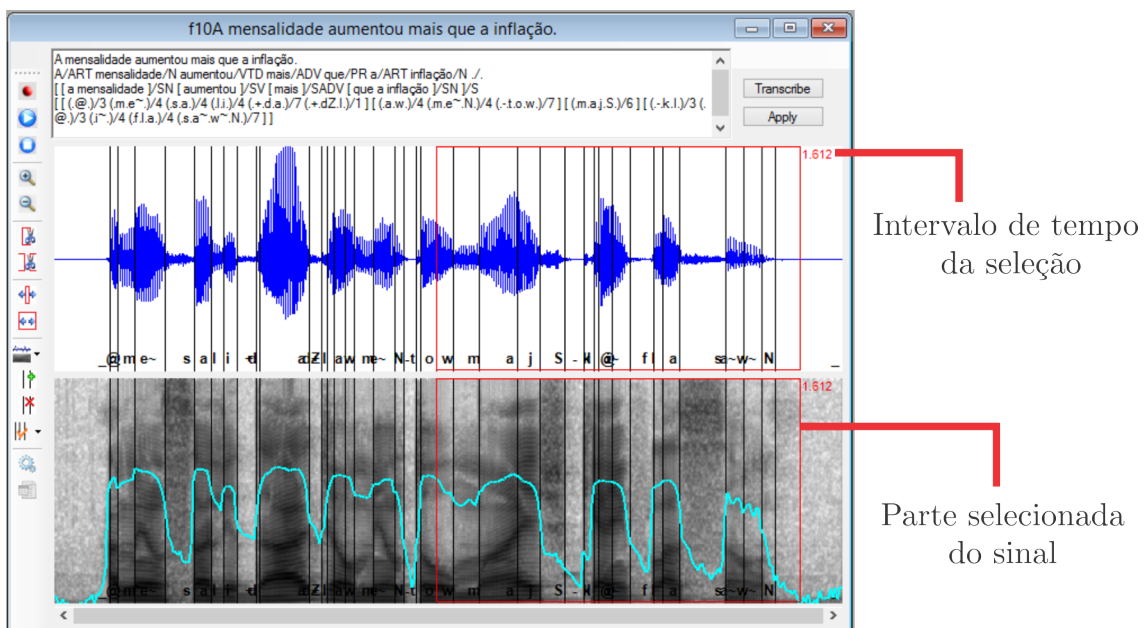


Figura 3.12: Utilização do *mouse* para selecionar uma parte do sinal.

3.3.2 Botão Iniciar

Como vimos no Capítulo 2, o botão Iniciar oferece duas opções: Modo Lista e Modo Data Base (figura 2.2). Basicamente, nos dois modos as telas são praticamente idênticas às mostradas em 3.3.1, portanto não nos aprofundaremos mais neste botão agora. Passaremos ao botão Transcritor.

3.3.3 Botão Transcritor

O botão Transcritor abre a tela mostrada na figura 2.16. Com relação ao ambiente do transcritor, toda a parte de análise é vista na própria figura 2.16. Maiores aprofundamentos aparecerão em exemplos de uso e exceções do sistema tratados em capítulos posteriores. Passaremos então ao ambiente do botão TTS.

3.3.4 Botão TTS

Ao clicar no botão TTS da Barra de Menus, temos a tela das figuras 2.17 e 3.5. Nestas telas vemos uma Barra de Ferramentas do lado esquerdo. Vamos agora descrever cada uma das funções.

- *Play*: Como o próprio nome sugere, reproduz o sinal mostrado na caixa retangular superior, com o sinal em azul. Como observação, veja que é possível escrever na caixa de inserção de sentenças, clicando com o botão esquerdo do *mouse*, mas só depois de clicar no botão Concatena é que o sinal será reproduzido a partir do botão *Play*.
- *Zoom In/Out*: Faz o que o próprio nome diz: dá *zoom* no sinal mostrado na caixa retangular superior.
- *Plot All*: Plota todo o sinal, funcionando como uma espécie de *reset* para o botão *Plot Selection*.
- *Plot Selection*: Plota a parte do sinal que foi selecionada com o botão esquerdo do *mouse* na caixa retangular superior.
- Manipulação Prosódica: este botão dá origem a uma nova tela, como vimos na figura 2.19. Vamos então às funções da Barra de Ferramentas desta tela.
 - *Play*: Reproduz o sinal como ele se apresenta na janela da figura 2.19. Sendo assim, se for feita alguma alteração no sinal nesta janela, o botão de *Play* reproduzirá o sinal alterado.
 - *Delete*: Aqui a função deste botão é remover as alterações feitas na parte selecionada (com o botão esquerdo do *mouse*). Um exemplo é mostrado na figura 3.13.

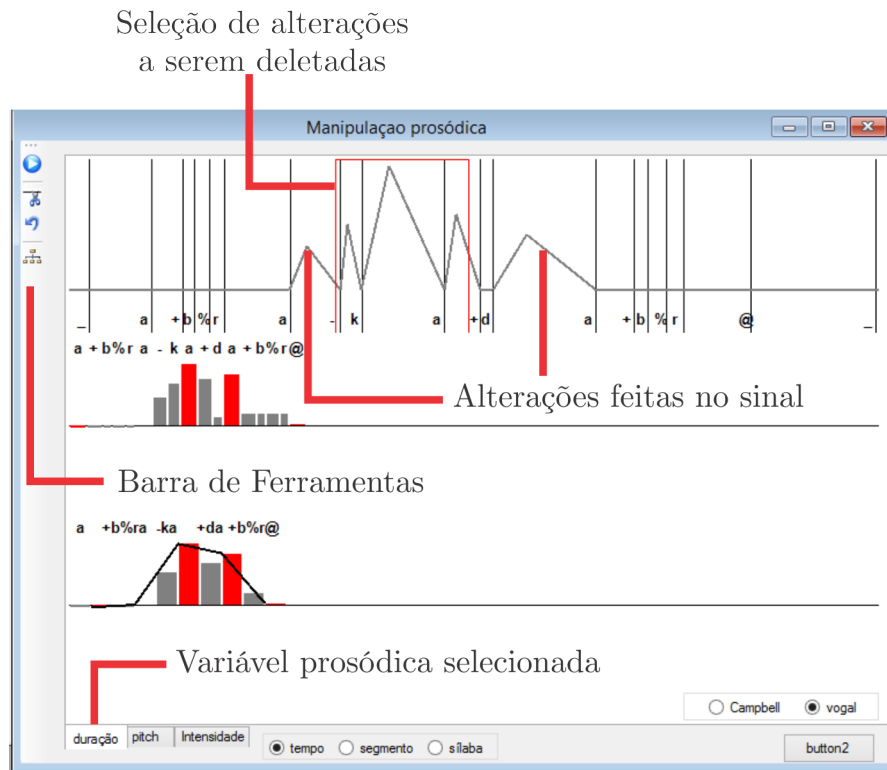


Figura 3.13: Detalhe da Barra de Ferramentas da tela de manipulação prosódica do TTS.

- *Reset*: Faz o que o próprio nome diz, volta o sinal às configurações iniciais.
- *toolStripButton3*: Abre a janela de estrutura de performance já vista na figura 2.20.

Ao lado da caixa de inserção de sentenças vemos uma lista suspensa (figura 2.18). Ao selecionar qualquer sentença desta lista, automaticamente o sinal é reproduzido e mostrado nas caixas retangulares maiores como o sinal sintetizado (em azul) e o espectrograma do sinal (caixa inferior). Na figura 2.18 a sentença escolhida foi “André”. Abaixo da lista suspensa, temos os botões Transcreve, Concatena e Modela. Vamos a uma descrição do seu funcionamento.

- **Transcreve**: Este botão transcreve a sentença mostrada na caixa de inserção de sentenças. Na figura 3.14 inserimos a sentença “O capítulo demorou para terminar.”. Seguem algumas observações sobre o botão.
 - Repare que a sentença na lista suspensa é Sergio Lima Netto. O sinal em azul corresponderá a esta sentença até que o usuário clique em Concatena.

- Também é importante perceber que o resultado do processamento do texto já é mostrado logo abaixo da caixa de inserção de sentenças, apesar de o sinal em azul ainda não ter sido atualizado.

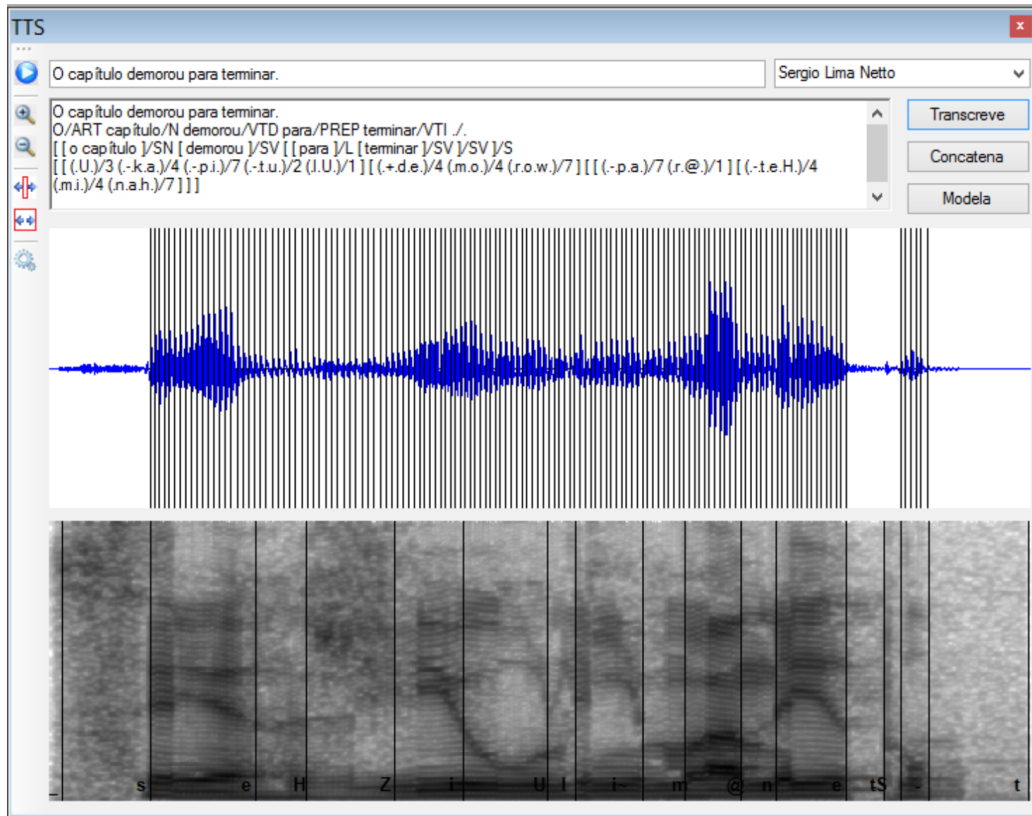


Figura 3.14: Utilização do botão Transcreve com a sentença “O capítulo demorou para terminar.”.

- Concatena: Ao clicar neste botão, as unidades transcritas pelo botão Transcreve são concatenadas e o resultado é mostrado nas caixas retangulares inferiores como é visto na figura 3.15. Além disso, automaticamente o sinal é reproduzido.
- Modela: este botão é que dá a prosódia (entonação) ao sinal transcrito e concatenado anteriormente. Na figura 3.16 é mostrado o resultado com a mesma sentença anteriormente usada.

Isto fecha a parte de análise do ambiente TTS. Passaremos ao ambiente Transplante.

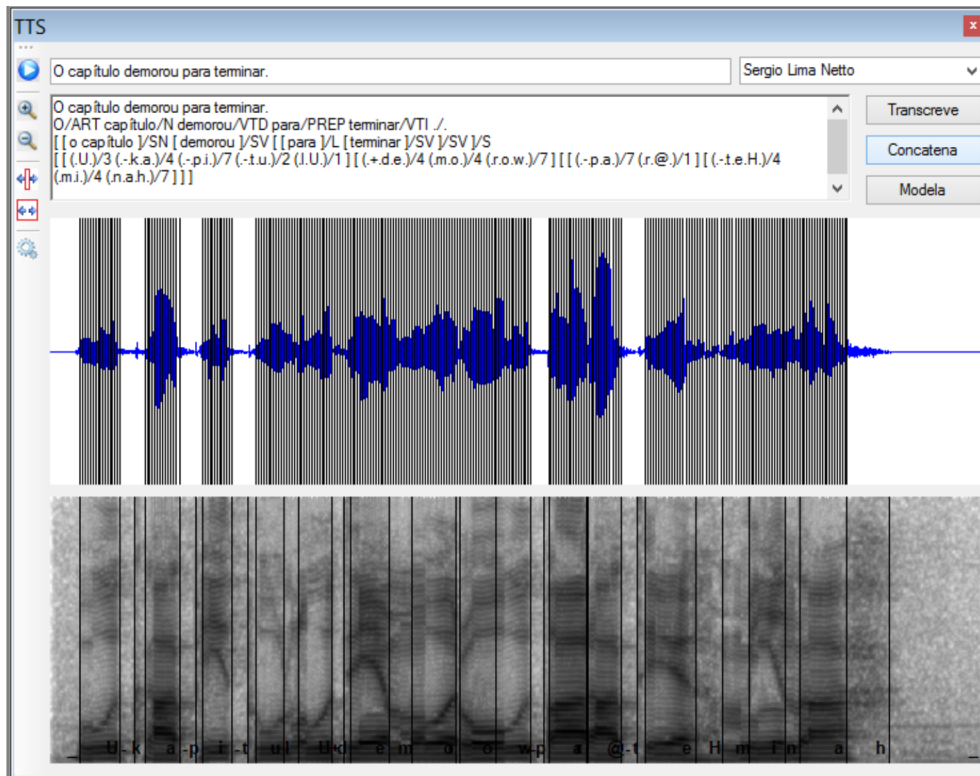


Figura 3.15: Utilização do botão Concatena com a sentença “O capítulo demorou para terminar.”.

3.3.5 Botão Transplante

O botão Transplante da Barra de Menus dá origem a tela vista na figura 2.24. Como nosso intuito agora é observar as ferramentas de análise, vamos falar de forma mais completa da Barra de Ferramentas deste ambiente que vemos em destaque na figura 3.18.

- *Play*: Reproduz o sinal de referência presente na caixa retangular superior.
- *play_sint*: Reproduz o sinal sintetizado por transplante de prosódia. Seguem algumas observações:
 - Este botão só funciona se houver um sinal sintetizado.
 - Caso uma sentença seja escolhida na lista suspensa, este botão reproduzirá esta sentença.
- *Save*: Este botão não funciona neste módulo.
- *Zoom In/Out*: Este botão não funciona neste módulo.

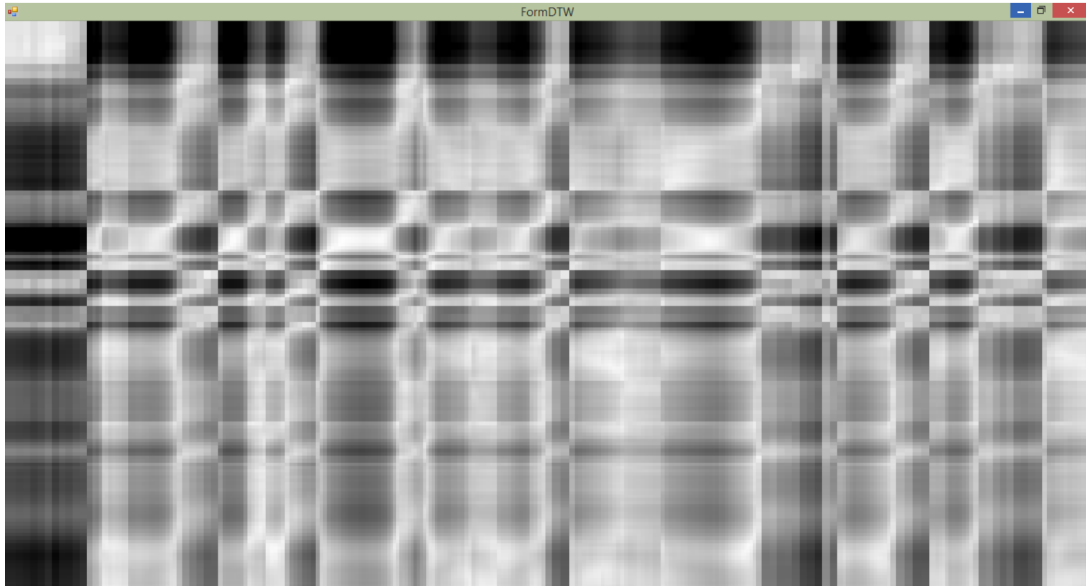


Figura 3.17: Destaque da janela aberta a partir do do botão FormDTW da Barra de Ferramentas.

- *ButtonModel*: Dá origem à tela mostrada na figura 2.27.



Figura 3.18: Destaque da Barra de Ferramentas do ambiente Transplante.

3.3.6 Botão Janelas

Como já visto no Capítulo 2, este botão organiza as janelas abertas.

3.4 Síntese de Dados

Agora que já falamos de toda a parte de entrada/saída de dados e passamos por todas as ferramentas de análise contidas no SASPRO, falaremos das ferramentas de síntese. Mais uma vez, para que fique claro, estamos chamando de síntese de dados a toda mudança feita nos parâmetros do sinal, seja ela feita pelo usuário propriamente ou automaticamente, pelo sistema.

Dentro da parte de síntese de dados temos, como exemplo para edição dos sinais, as funcionalidades de manipulação da prosódia com possibilidades de manipulação da duração do *pitch* e da intensidade, sendo que cada uma destas pode ser ainda editada tendo como foco o tempo, o segmento ou a sílaba; temos ainda a inserção/retirada de marcas de *pitch* entre outras.

3.4.1 Tela Inicial

Como vimos na figura 2.5, ao selecionar uma sentença na tela inicial (figura 2.1), temos uma janela de análise com uma Barra de Ferramentas vista na figura 2.6. Desta Barra de Ferramentas temos algumas funcionalidades para explicitar aqui. Vamos a elas:

- *Delete*: Remove a parte do sinal que foi selecionada.
- *Delete Inverse Selection*: Remove a parte do sinal que não foi selecionada.
- *InsertMark*: Serve para inserir marcas de *pitch* (no modo GCI) ou etiquetas de fronteira (no modo *Label*). Para utilizar esta ferramenta, primeiro é preciso selecionar o lugar onde se vai inserir a marca/etiqueta com o botão esquerdo do *mouse* e só depois clicar no botão.
- *Delete Marks*: Apaga as marcas de *pitch* ou etiquetas de fronteira. Para utilizar esta ferramenta, primeiro devemos selecionar a área do sinal da qual queremos apagar as marcas, usando o botão esquerdo do *mouse*, e depois clicar no botão.
- *Manipulation*: Abre a janela de manipulação prosódica já vista anteriormente da figura 2.12 até a figura 2.15 e da qual falaremos agora em mais detalhes

em relação à análise de dados. Na figura 2.12 vemos que na parte superior esquerda há uma Barra de Ferramentas. Vamos a cada função:

- *Play*: Reproduz o sinal mostrado na parte superior contendo o sinal ao longo do tempo.
- *Delete*: Remove qualquer modificação previamente selecionada pelo *mouse* com o botão esquerdo. Primeiro selecionamos com o *mouse*, clicando e arrastando para desenhar uma caixa retangular em vermelho.
- *reset*: Volta o sinal às condições originais, isto é, antes de ser modificado.
- *toolStripButton3*: Mostra a estrutura de performance do sinal.

Vamos, agora, a algumas observações sobre o uso da janela de manipulação prosódica:

- Todas as observações feitas aqui valerão para as janelas de manipulação prosódica dos demais ambientes.
- Uma vez selecionada a variável prosódica (duração, *pitch*, intensidade) o modo de edição (tempo, segmento, sílaba) definirá como o sinal será editado.
- As variáveis prosódicas não interferem umas nas outras, ou seja, alterar uma não significa alterar a outra.
- O botão *reset* volta todas as variáveis prosódicas às suas configurações iniciais.

Isto encerra a parte inicial do sistema. Passaremos agora aos botões da Barra de Menus.

3.4.2 Botão Iniciar

Como já vimos inúmeras vezes até aqui, o botão Iniciar nos permite escolher entre o modo Lista e o modo Database. Cada um destes modos tem suas “particularidades”, como já vimos e, no que diz respeito à síntese de dados, nada há para acrescentar, a não ser que tudo o que foi visto em 3.4.1 é válido aqui também. Passaremos então ao ambiente Transcritor.

3.4.3 Botão Transcritor

Este é um ambiente que oferece pouca opção em termos de funcionalidade. Praticamente tudo já foi visto em 2.3. As funcionalidades também estão descritas nesta seção.

3.4.4 Botão TTS

Ao clicarmos no Botão TTS da Barra de Menus, temos a tela já vista anteriormente na figura 2.18. No que diz respeito a esta tela nada há a acrescentar.

O botão de manipulação prosódica abre uma janela idêntica à da figura 2.12. Todas as funcionalidades já citadas anteriormente, incluindo observações, são válidas aqui. Em relação à síntese, vamos a alguns comentários sobre esta tela.

A figura 3.19 mostra a tela de manipulação prosódica do ambiente, com a sentença “Isso já aconteceu antes!”. O sinal nesta figura foi transcrito, concatenado e modelado. Repare que a guia duração é a que está selecionada na parte inferior. Também o modo tempo é que está selecionado. Sendo assim, com um duplo clique com o botão esquerdo do *mouse* no sinal, representado na parte superior, podemos editá-lo com respeito à duração no domínio do tempo. Na figura 3.20 aumentamos a duração do meio para o final do mesmo.

Para alterar, por exemplo, o *pitch* no modo de edição do segmento, basta selecionar a aba correspondente ao *pitch*, mudar para o modo de edição por segmento e clicar no sinal correspondente, como visto na figura 2.14.

3.4.5 Botão Transplante

O botão Transplante da Barra de Menus dá origem à tela 2.24. Já falamos sobre a Barra de Ferramentas e como utilizá-la.

3.5 Conclusão

Basicamente, neste capítulo, voltamos a explicar sobre as funcionalidades presentes no SASPRO, porém tentando encaixá-las em três grandes grupos: Entrada/Saída

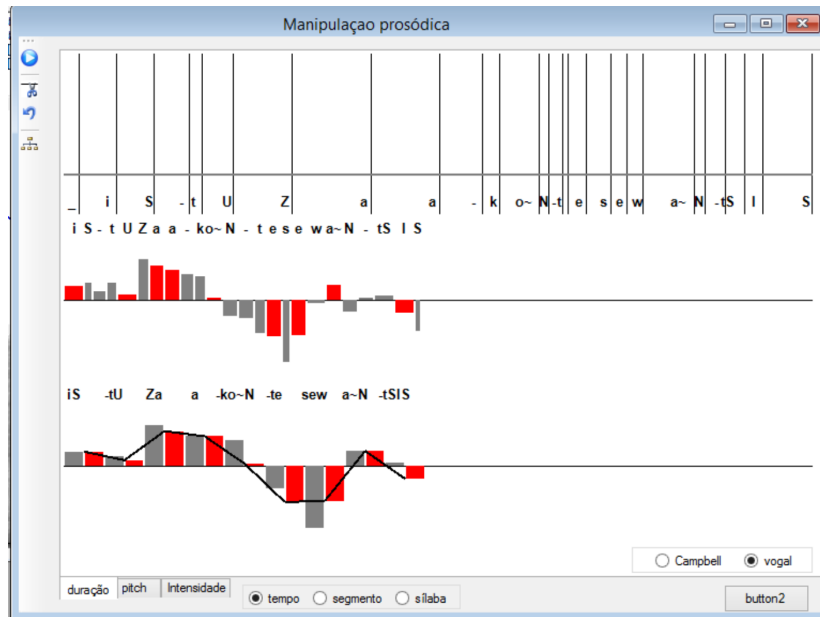


Figura 3.19: Exemplo de utilização do ambiente de manipulação prosódica dentro do TTS.

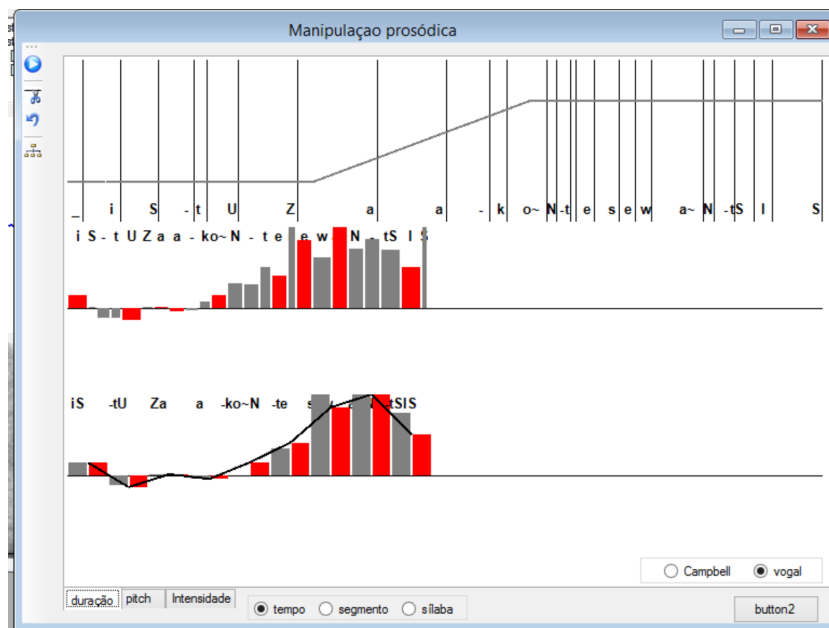


Figura 3.20: Exemplo de utilização do ambiente de manipulação prosódica dentro do TTS. Agora modificamos a duração no domínio do tempo.

de Dados, Análise de Dados e Síntese de Dados. Como ficou claro neste capítulo, nem sempre isso é possível, pois dependendo da finalidade as ideias de análise e síntese podem ficar dentro da mesma tela ou menu.

Capítulo 4

Estudo de Caso

4.1 Introdução

Nos Capítulos 2 e 3 demos uma visão geral do sistema, sendo que no primeiro apresentamos as telas, os botões, os menus e as funcionalidades presentes no SASPRO; no segundo, procuramos agrupar e apresentar estas funcionalidades por finalidade de maneira geral, separando-as em entrada/saída de dados, análise de dados e síntese de dados. Nosso objetivo no presente capítulo é explorar um caso real de uso, mais especificamente, verificar como podemos utilizar as ferramentas existentes dentro do SASPRO para criar ou aumentar um banco de unidades já sintetizadas. Assim, nossa ideia é mostrar como identificar que palavras ou sentenças não possuem transcrição ou possuem uma transcrição inadequada e ir melhorando sua transcrição através de uma nova síntese.

4.2 Usando o SASPRO para Aumentar o Banco de Unidades

Como vimos, a ideia do presente capítulo é mostrar como aumentar o banco de unidades. Uma das maiores dificuldades para se obter uma boa qualidade no sinal de saída é o fato de que junto às ferramentas computacionais esteja também, aliado, um banco de unidades robusto, ou seja, que contenha um número significativo de unidades. Para criar este “banco” dividimos o trabalho em etapas que permitem realizar esta tarefa:

1. “Varredura” do banco de unidades existente;
2. Gravações de sentenças que contenham as unidades problemáticas;
3. Inserção e/ou correção das marcas de EGG;
4. Etiquetagem do sinal; e
5. Incorporação das novas unidades ao banco.

Aqui queremos dar ênfase às etapas (1), (4) e (5). Estas são descritas a seguir, com maiores detalhes, utilizando-se o SASPRO.

4.3 Varredura do Banco

Em um primeiro momento é preciso identificar, através de alguns testes, possíveis unidades faltantes no acervo. Este é um trabalho tedioso e demorado, uma vez que é feito manualmente e caso a caso. Uma estratégia possível para realizar esta etapa é a seguinte:

1. Faz-se uma tabela com as principais sílabas/difones do nosso idioma (considerando sílabas com acentos gráficos distintos como unidades distintas). Pode-se pensar inicialmente em consoantes-vogais, depois vogais-consoantes e alguns encontros consonantais, tais como FR, BL etc. e, por fim, os encontros vocálicos como, por exemplo, AE, AI, etc. É possível ver um resumo destas tabelas no apêndice B.
2. Escreve-se a sílaba que se deseja testar, avalia-se a saída do programa e se classifica como existente ou ausente. Caso seja ausente, o próximo passo é testar a unidade dentro de uma palavra e classificá-la (a palavra) como boa ou ruim. Para fazer o teste, abra o SASPRO, vá em TTS na Barra de Menus digite na primeira linha a sílaba a ser testada. Na figura 4.1 usamos a sílaba BA. A seguir, clique em Transcreve e depois em Concatena.

4.3.1 Observações Gerais

Durante as verificações, é comum encontrar alguns equívocos, que se devem justamente ao número limitado de unidades presentes. Um destes, diz respeito à acen-

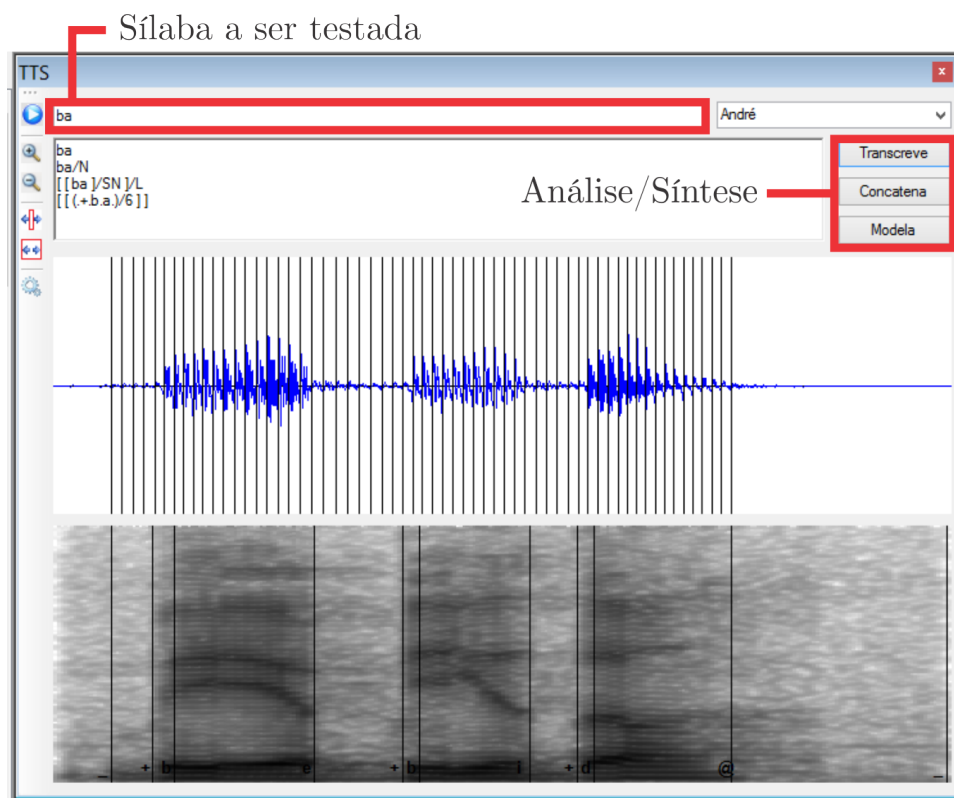


Figura 4.1: Exemplo de teste de sílaba para verificação de unidades existentes no banco. No caso, a sílaba testada é BA.

tuação das palavras, mais precisamente a sua sílaba tônica. Na saída do sintetizador o SASPRO muitas vezes não dá ênfase necessária nas sílabas acentuadas. Por exemplo, a palavra “fácil” não é pronunciada corretamente, dando ênfase na última sílaba e não na primeira. A ideia é identificar estes problemas e classificá-los, usando a tabela sugerida anteriormente, de maneira a corrigi-los com inserção das unidades faltantes.

4.4 Gravações

Passada a etapa inicial de busca e identificação/classificação dos erros, por meio da varredura do banco, são formuladas sentenças com as unidades fonéticas desejadas, e que devem ser gravadas. No exemplo, dado aqui, foram gravadas 10 sentenças pelo criador do SASPRO¹. Foi necessária a sua presença para que pudéssemos manter as mesmas características do acervo do banco já existente. Esta é uma dificuldade que

¹Vagner Latsch.

a maioria dos sistemas de síntese possui, uma grande dependência do locutor.

Após esta fase há dois procedimentos a serem realizados nos sinais gravados: a correção das marcas de EGG e a etiquetagem, pois o sinal enviado pelo locutor não vem com nenhuma das etiquetas de fronteira. O arquivo enviado será um arquivo mono (com apenas um canal) que conterà a extensão `.wav` e será semelhante ao da figura 3.2. Veja que nesta figura o sinal de entrada oriundo do microfone não possui nenhuma etiqueta de fronteira como, por exemplo, o sinal da figura 3.11.

As sentenças já gravadas e que fazem parte do sistema que são vistas na figura 2.1 estão na pasta cujo endereço é `SASPRO/projetos/db/frases_dur`. Há dois caminhos possíveis a serem seguidos: (1) salvar as novas sentenças em uma nova pasta (recomendado); ou (2) salvar junto com as sentenças existentes (requer um procedimento maior). Vamos aos detalhes de cada uma.

Se o usuário optar por criar uma nova pasta ele deve ir ao endereço em que já estão as sentenças iniciais `SASPRO/projetos/db` e criar uma pasta em que irá salvar os arquivos que ele vai gerar a partir dos procedimentos descritos nas seções 4.5 e 4.6. Sendo assim, o novo endereço passará a ser, por exemplo, `SASPRO/projetos/db/frases_novas`. Para que este procedimento funcione é preciso, ainda, colocar nesta mesma pasta um arquivo `.txt` com o mesmo nome da pasta em que estão as novas sentenças; no nosso exemplo, o nome deverá ser `frases_novas.txt`. Na primeira linha deste arquivo deve constar o endereço da pasta com os arquivos `.wav` com as novas sentenças. Veja na figura 4.2 o destaque do arquivo de texto contendo o endereço que indica a pasta contendo as sentenças.

Se o usuário optar por acrescentar as novas sentenças na pasta contendo as sentenças já existentes, ele deve salvar os arquivos `.wav` novos na pasta já indicada antes (`SASPRO/projetos/db/frases_dur`) e, depois, deve procurar o arquivo `.txt` correspondente (neste caso, `frases_dur.txt`) e acrescentar lá as novas sentenças no final do arquivo.

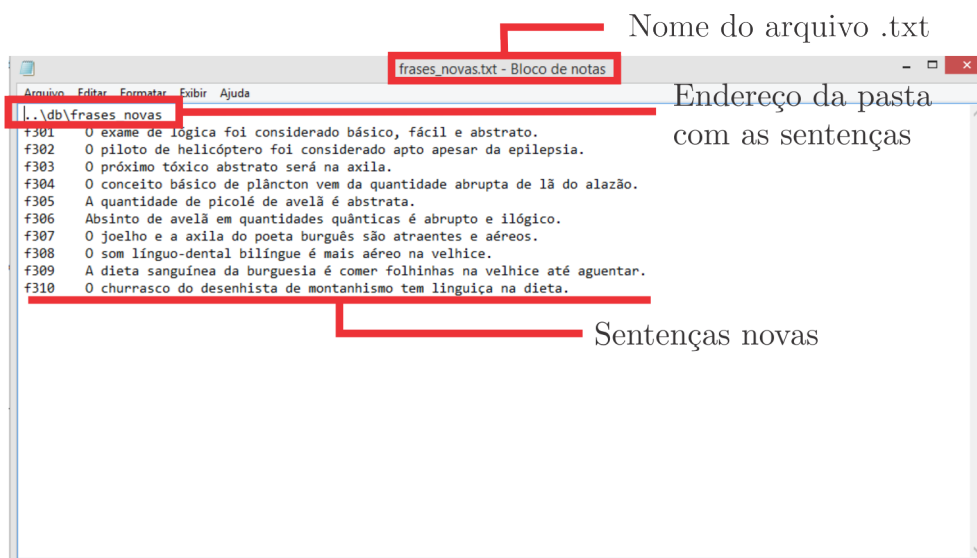


Figura 4.2: Destaque do arquivo `.txt` mostrando como indicar o endereço da pasta contendo as sentenças a serem incorporadas no banco de sentenças.

4.5 Correção das Marcas de EGG

O primeiro procedimento consiste em fazer o “aperfeiçoamento” do sinal de EGG gerado paralelamente às gravações da voz. A tela que o usuário verá será semelhante à da figura 2.11 após clicar na Barra de Ferramentas (figura 2.10) no botão *Edit* e depois em GCIs. Esse sinal é obtido através de alguns eletrodos que são colocados ao redor do pescoço do locutor e que captam a impedância elétrica da laringe; após alguma manipulação matemática, este sinal nos fornece os instantes de fechamento da glote. A figura 4.5 mostra o eletroglotógrafo em uso.

Pode ser que o usuário queira inserir/corrigir uma marca de EGG entre dois instantes muitos “próximos”. Para fazer isso, selecione com o botão esquerdo do *mouse* o intervalo em que quer-se corrigir/inserir a marca de EGG. Depois, basta ir à Barra de Ferramentas já mostrada na figura 2.10, clicar em *Plot Selection*. A figura 4.3 mostra o sinal correspondente a sentença “O exame de lógica foi considerado básico fácil e abstrato.” com um trecho curto selecionado.

Já na figura 4.4 mostramos o trecho em destaque citado na figura 4.3, no qual queremos editar a marca de EGG. Depois de destacado o trecho em questão, basta clicar no botão *InsertMark* (para inserir) ou no botão *Delete Marks* (para remover).

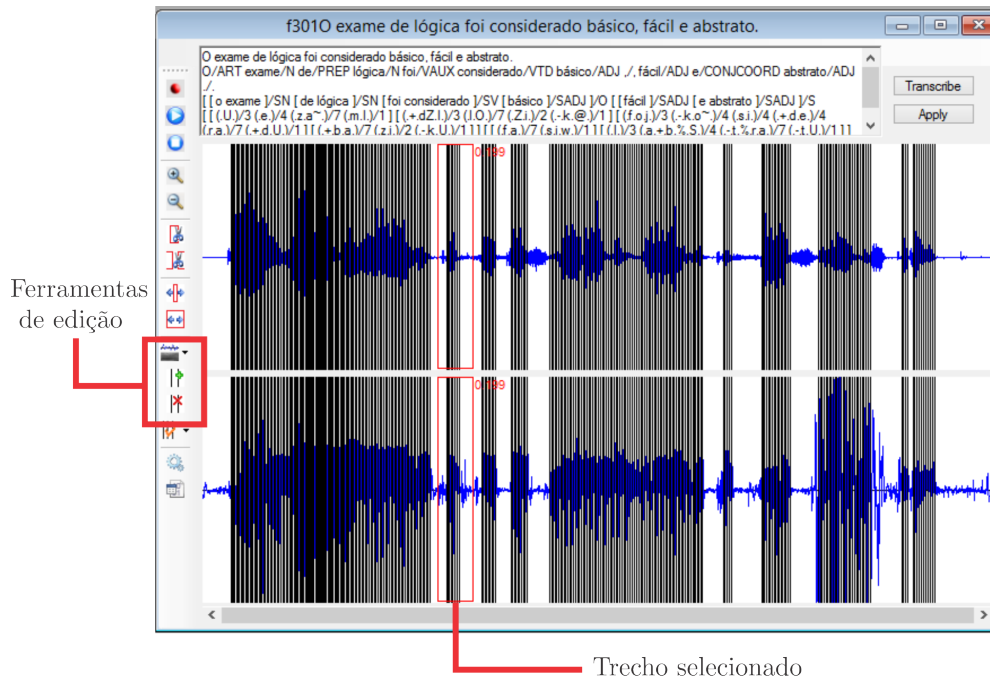


Figura 4.3: Destaque da tela de edição da marcas de EGG de um sinal qualquer. O exemplo usado aqui foi a sentença “O exame de lógica foi considerado básico, fácil e abstrato.”.

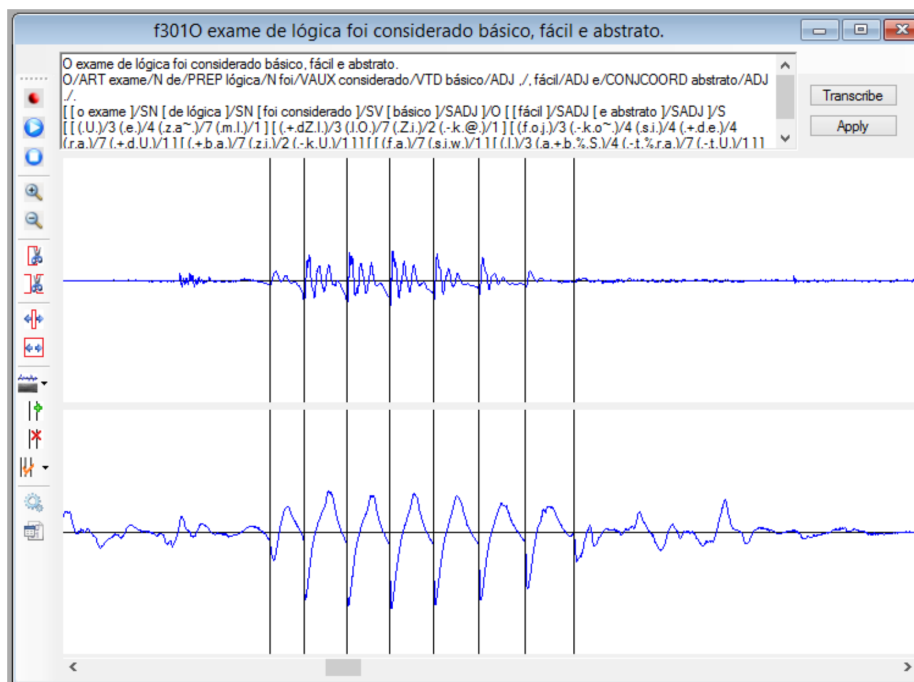


Figura 4.4: Destaque de um trecho específico de uma sentença para edição de uma marca de EGG.

Este sinal é importante para a estimação do *pitch* do sinal, que é de suma importância para a prosódia do sinal de saída, algo que pode ser entendido também

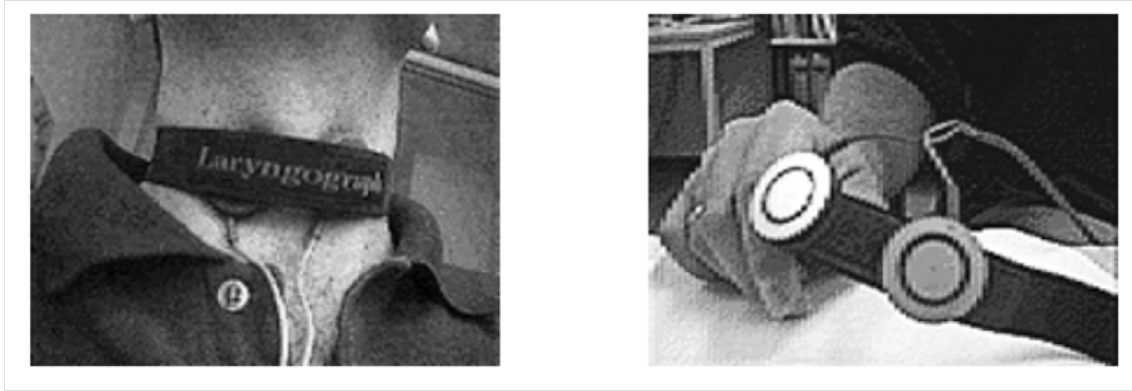


Figura 4.5: Imagem do eletroglotógrafo (EGG) em funcionamento.

como a entonação da voz. Na edição do sinal de EGG modificado nós apenas retiramos ou colocamos algumas marcas que o processo automático pôs ou deixou de por, baseando-se em nossa experiência e observando a figura 2.11.

4.6 Etiquetagem do Sinal

O segundo procedimento é uma técnica de dividir a gravação em várias unidades fonéticas com a ajuda do SASPRO para posterior inserção dessas unidades no banco. Aqui damos ênfase à etiquetagem do sinal, que será mais detalhada a seguir.

Para inserir as etiquetas de fronteira, basta usar a Barra de Ferramentas vista na figura 2.10 por meio dos botões de ferramentas de edição das marcas de *pitch*. Esses são os mesmos botões que servem para editar as marcas incorretas (ou faltantes) decorrentes das marcas de EGG feitas no procedimento anterior. Basta selecionar, no sinal, o instante em que se quer inserir a etiqueta de fronteira, clicando com o botão esquerdo do *mouse* e depois disso basta clicar no botão *InsertMark*. Da mesma maneira que na edição das marcas de EGG, também é possível selecionar um pequeno trecho do sinal e “ampliá-lo” para que se possa inserir uma etiqueta de fronteira de maneira “mais precisa”. As figuras 4.6 e 4.7 dão destaque a esta propriedade do sistema.

Embora este procedimento seja relativamente simples, seguem algumas observações importantes para que a inserção das etiquetas tenha um efeito eficaz.

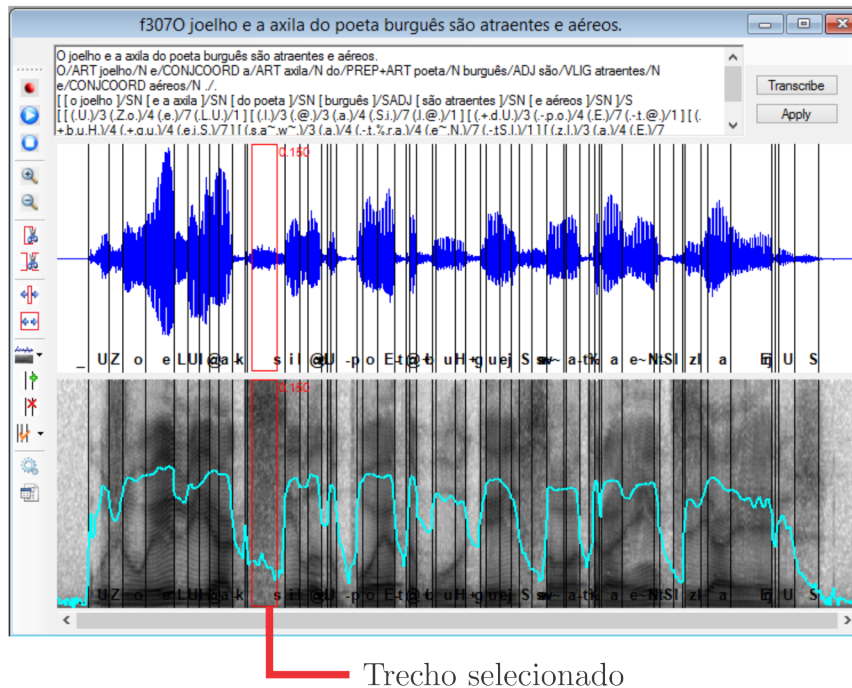


Figura 4.6: Destaque do procedimento de correção de etiquetas de fronteira.

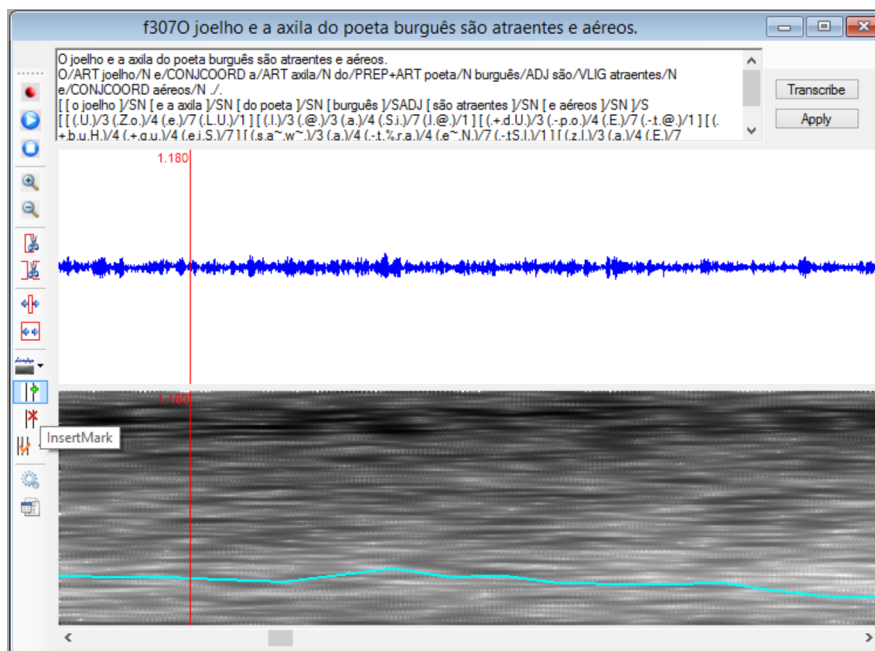


Figura 4.7: Edição da etiqueta de fronteira em um trecho específico por meio do botões *Plot Selection* e *InsertMark*.

4.6.1 Observações Sobre a Etiquetagem

- Uma situação comum durante a segmentação é quando as unidades contém consoantes como “p”, “t”, “k”, “d”. Essas letras são classificadas como plosivas, sempre são precedidas por um “-”, que representa o ar que é expirado e

bloqueado por um obstáculo bucal antes de pronunciar a sílaba propriamente dita. Geralmente é bem difícil distinguir a fronteira entre esses dois instantes; neste caso, o uso do espectrograma auxilia o usuário nesta identificação.

- O espectrograma também é muito útil para diferenciar algumas vogais, já que estas possuem “franjas” de energia em seus espectrograma, e as consoantes não. Um exemplo de sílaba em que esta característica pode ser usada é “si”.
- O método de diferenciação anterior não ajuda quando se trata de duas vogais seguidas como, por exemplo, na fronteira entre os dois “a’s” no trecho “na axila”. Uma alternativa é dividir o intervalo na metade se as duas vogais forem iguais, como no caso apresentado anteriormente; ou, tentar fazer uma aproximação pelo som quando ocorre um encontro fonético de duas vogais distintas como “ea” em “de avelã”. Essa forma de tratamento é variável, pois depende diretamente do indivíduo que esteja fazendo a segmentação e sua percepção.
- Bem como os encontros vocálicos, os encontros consonantais também podem gerar dificuldades. Palavras com os dígrafos (o dígrafo pressupõe um fonema só) “lh” e “nh” são mais difíceis de serem etiquetadas do que outras justamente pela dificuldade em definir a fronteira entre as duas consoantes.
- Um entrave comum à todas as sentenças é que nas últimas palavras, no fim da sentença, o sinal já possui baixa energia, e por isso fica ainda mais difícil distinguir cada letra.

Passemos então, agora, à última etapa, que é a incorporação das unidades anteriormente identificadas, gravadas e segmentadas ao banco de unidades.

4.7 Incorporação de Novas Unidades ao Banco

Depois de segmentadas, as novas unidades precisam ser acrescentadas ao banco de unidades do SASPRO.

Primeiro vá ao menu Iniciar na Barra de Menus (figura 2.1) e, depois, em Database. Em seguida escolha, no submenu, a opção Banco Automático. Para que

a opção Banco Automático apareça é preciso abrir o SASPRO por meio do Visual Studio™. Caso você não esteja visualizando esta opção deve fechar o SASPRO e reabri-lo como dito anteriormente. Para fazer isso, siga o caminho dado por SASPRO/Projetos/ProsodyEditor e procure pelo arquivo `ProsodyEditor.sln`. Depois de aberto vá até a opção *Build*. Depois de feito isso a opção estará ativada.

Ao clicar na opção Banco Automático, abre-se então uma janela para que possamos escolher qual arquivo contém as sentenças “novas” a serem anexadas. Aqui utilizamos, como exemplo, o procedimento de salvá-las em uma nova pasta chamada de `frases_novas`. Seguindo este caminho, deve-se escolher o arquivo `frases_novas.txt`, que conterá, como já foi visto na figura 4.2, o caminho para os arquivos `.wav` a serem adicionados.

Ao fim do processo o SASPRO pergunta se o usuário deseja salvar as alterações em `DataDef`. Deve-se confirmar, e no final é exibida uma mensagem que diz que o arquivo foi alterado com sucesso.

4.7.1 Arquivo de Exceções

Quando mesmo depois de selecionadas as unidades de alguma forma defeituosas, terem sido gravadas e etiquetadas ainda continuam com problema, há ainda uma outra possibilidade no tratamento dos erros: a utilização do arquivo de exceções. O arquivo `lexico.txt` contém a transcrição fonética de algumas palavras; veja em destaque na figura 4.8. Ao utilizar essa alternativa, estamos forçando o transcritor a usar nossa transcrição no lugar de utilizar a transcrição via TTS. O caminho para o arquivo `lexico.txt` é SASPRO/Projetos/ProsodyEditor/`lexico.txt`.

4.8 Conclusão

Neste capítulo vimos como o SASPRO pode ser usado para gerar um banco de unidades de transcrições. Enumeramos as etapas e, em seguida, mostramos como estudar as unidades que precisam ser melhoradas. Vimos também como realizar estas tarefas através da correção das marcas de EGG inseridas pelo eletroglotógrafo

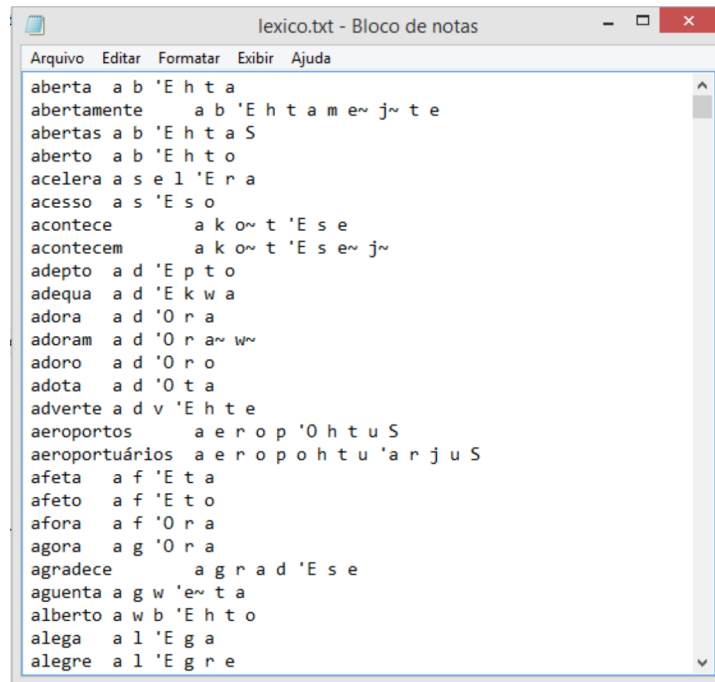


Figura 4.8: Destaque do arquivo `lexico.txt` para que possa ser visto qual o formato de inserção de palavras no banco de exceções.

e também como colocar as etiquetas de fronteira. Por fim, vimos como adicionar de maneira simples unidades ao banco de exceções.

Capítulo 5

Conclusão

5.1 Contribuições Deste Trabalho

Este projeto teve como principal objetivo descrever, apresentar e expandir sob certo ponto de vista um sistema de análise e síntese de prosódia, intitulado SAS-PRO, que é importante e funcional e pode ser aplicado tanto para dar acessibilidade a portadores de necessidades especiais, como por exemplo em relação a visão, denotando um cunho social, quanto para finalidades mais comerciais, como por exemplo em sistemas de segurança.

O Capítulo 2 apresentou todo o sistema com suas telas, botões, menus e submenus, de modo que funcionasse com uma visão/apresentação geral do sistema.

O Capítulo 3 descreveu e apresentou as funcionalidades presentes no sistema dividindo-as em três grandes grupos: entrada/saída de dados, síntese de dados e análise de dados, explicando e dando exemplos de uso da cada funcionalidade sempre que necessário para um melhor entendimento.

O Capítulo 4 teve o papel de dar ao leitor uma espécie de estudo de caso, mostrando como pode ser expandida e/ou melhorada a base de unidades já existentes no SASPRO, a fim de se obter resultados melhores por meio do uso do sistema.

5.2 Propostas de Trabalhos Futuros

O SASPRO, como ficou claro neste trabalho, é composto de módulos menores. Algumas sugestões de trabalhos futuros são:

- Ampliação do banco de unidades já existentes, fazendo-se uma ampla varredura do banco, verificando e corrigindo problemas de transcrição.
- Aprimoramento da interface gráfica, tornando-a mais amigável e intuitiva para o usuário;
- Tratamento de erros de transcrição. A maioria dos erros não tem nenhum apontamento de causa. Pode-se criar uma forma de informar ao usuário a causa do erro;
- Criação de uma primeira versão com adaptação para funcionamento via web; e
- Inclusão de um *help* interno ao sistema para facilitar a ajuda ao usuário.

Referências Bibliográficas

- [1] Latsch, Vagner Luis. *Desenvolvimento de um Sistema de Conversão Texto–Fala com Modelagem de Prosódia*. Rio de Janeiro : Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Tese de Doutorado, 2011.
- [2] Silva, Kersey Wirleide Anacleto Xavier da. *Sistema de Conversão Texto–Fala com Busca Otimizada de Unidades Acústicas em Banco de Voz*. Rio de Janeiro : Escola Politécnica – Universidade Federal do Rio de Janeiro, Projeto de Graduação, 2011.
- [3] Netto, Sergio Lima. *Codificação de Voz Para Sistemas de Telecomunicações*. Rio de Janeiro : EPOLI–COPPE/UFRJ, Apostila.
- [4] —. *Introdução A MATLABTM*. Rio de Janeiro : Apostila, s.n., 2003.
- [5] Alcaim, Abraham e Oliveira, Carlos Alexandre dos Santos. *Fundamentos do Processamento de Sinais de Voz e Imagem*. Rio de Janeiro : PUC Rio/Interciência, 2011. pp. 19–41. 978-85-7193-267-8.
- [6] Diniz, Paulo Sergio Ramirez, Silva, Eduardo Antônio Barros da e Netto, Sergio Lima. *Processamento Digital de Sinais – Projeto e Análise de Sistemas*. s.l. : Bookman, 2004. pp. 160–176. 978-85-363-0418-2.
- [7] Dalcastagnê, André Luís. *Apostila Básica do Software MATLABTM*. Florianópolis : Centro Federal de Educação de Santa Catarina CEFET–SC- Departamento Acadêmico de Eletrônica , 2008.
- [8] Latsch, Vagner Luis. *Construção de Banco de Unidades Para Síntese de Fala Por Concatenação no Domínio Temporal*. Rio de Janeiro : COPPE/UFRJ, Tese de Mestrado, 2005.

Apêndice A

Exceções e Curiosidades Sobre o Sistema

A.1 Introdução

Este apêndice trata basicamente de exceções não tratadas pelo sistema, seja porque não são do interesse real do mesmo ou porque o sistema ainda possa, de certa maneira, evoluir e incluir as características aqui citadas.

A.2 Numerais

O sistema não funciona bem para numerais. Expressões como “Meu telefone é 35866476.” causam uma mensagem de erro. Isto acaba por não ser uma limitação do sistema, uma vez que é possível escrever “Meu telefone é três cinco oito meia meia quatro sete meia.”. A mesma ideia vale para exemplos do tipo “Moro na Rua dos Sistemas, número 20.”. O ideal é substituir o número pelo sua descrição por extenso. Outros exemplos são “Ganhei um milhão de reais!” ou “Ele tinha seis anos quando veio pra cá.”

A.3 Onomatopeias

O sistema tem algumas dificuldades com certas onomatopeias. Palavras como “hahahaha” funcionam com o “h” com som de “r”. Porém quando escrevemos “ha ha ha” ele interpreta como o verbo “haver” na terceira pessoa do singular (há) e

não consegue concatenar a sentença. Isto acaba por não ser bom em palavras tais como “bahia”, uma vez que o h do meio não deve ser pronunciado.

Foram testadas outras onomatopeias tais como “toc toc”, “au au”, “atchim!”, “cocoricó”. Para a maioria delas o TTS não enfrenta problemas. Algumas como “tchibum!” não são modeladas, mas são concatenadas sem problemas.

A.4 Dígrafos e Encontros Consonantais

O sistema mostra dificuldade em transcrever e modelar palavras tais como “ani-nhar”, “companhia”, por causa do dígrafo nh, mas nada que comprometa o entendimento. Para o dígrafo lh o som sai bem natural, bem como para ch como em “chave”. Apesar de não se tratar de dígrafos, mencionamos aqui palavras que começam com h, tais como hindu e hotel, que são muito bem pronunciadas.

Em encontros consonantais de vogais repetidas tais como siglas e/ou palavras com tt, pp, bb, cc como em “cóccix” ou em “occipital”, dd, gg (como em *Goggle*), há certa dificuldade em entender algumas palavras, mas pelo seu uso pouco comum isso não parece ser um grande problema. Outros dígrafos, como sc em “nascer”, não mostram problema, mas em palavras como em “mnemônico” o primeiro m não é pronunciado. Mais uma vez, apesar da limitação não há comprometimento significativo do uso do sistema.

A.5 Expressões Comuns e Siglas

Na linguagem falada são comuns algumas expressões do tipo “Nos Capítulos 3 e 4.” Como vimos neste apêndice, é melhor substituir os numerais pelas palavras “três” e “quatro”, respectivamente. Expressões como “três x quatro” e “três p/ quatro” são pronunciadas pelo sistema ignorando o “x” e o “p/”.

Siglas como “UFRJ” tentam ser lidas pelo TTS como uma palavra. Para contornar, basta escrever “uefe erre jota” e os resultados são apreciáveis. Porém, acrônimos como “UERJ” ou “COPPE” funcionam bem. No entanto, ao modelar a “palavra” “COPPE” há um travamento do programa ele é encerrado.

A.6 Interjeições

Repetições seguidas de vogais como “eeeeee” ou “aaaahhhh” também travam e fecham o programa quando clicamos no botão Modela.

A.7 Observações Sobre a Língua Inglesa

O sistema não foi feito para funcionar com a língua inglesa, mas, apesar disso, fizemos alguns testes e certas palavras são lidas normalmente. Outras dão algum tipo de problema. Como exemplo, a palavra *through* é transcrita e concatenada e com pronúncia razoável, mas não é modelada. Os sons de “dog” e “dogue” são praticamente idênticos, se não idênticos.

Palavras como “delete”, do verbo deletar, ou *internet* são lidas perfeitamente. Como era de se esperar o sistema não reconhece palavras como *play*, pronunciando exatamente como se escreve, ou seja, “plai”.

Apêndice B

Exemplo de Tabela Para Avaliação das Unidades

B.1 Introdução

O objetivo deste apêndice é mostrar exemplos de tabelas a serem usadas para testar as unidades já existentes no banco do SASPRO. Uma sugestão de utilização é escrever no cruzamento da vogal com a consoante uma palavra – que exista ou não na língua portuguesa – que contenha a sílaba em questão ou a própria sílaba. Para casos em que não temos certeza se a unidade existe ou não colocamos a unidade a ser testada entre colchetes.

Feito isso, basta testar a sílaba (ou a palavra), como visto em capítulos anteriores, e criar uma metodologia para simbolizar se a transcrição está boa ou ruim. Pode-se, por exemplo, colocar (OK) ao lado daquelas que tiveram avaliação positiva e (R) ao lado das que estão com transcrições inexistentes e/ou inadequadas. Aqui não colocamos nenhuma simbologia, pois nosso intuito é tão somente exemplificar uma metodologia para testar unidades. Sendo assim, é isto que se encontra nas tabelas B.1, B.2 e B.3.

B.2 Tabela Consoante-Vogal

Temos, a seguir (tabela B.1), um exemplo de tabela para testar as unidades do tipo consoante-vogal (CV). Obviamente ela está resumida às consoantes B, C, D e

Consoante-Vogal (CV)									
(V)	A				E			I	
(C)	a	á	â	ã	e	é	ê	i	í
B-	bacana	básico	[bânxi]	talibã	belo	[béto]	bebê	bilu	bíblia
C-	casa	cálice	câmbio	cão	cebola	céu	[cê]	ci	cílio
D-	dado	dá	dâ	dã	dedo	[déli]	dedê	dida	dízimo
F-	faca	fácil	[fâ]	fã	feto	fé	fê	fitá	físico

Tabela B.1: Exemplo de tabela resumida para testar a prosódia e/ou transcrição das unidades do tipo consoante-vogal.

Vogal-Consoante-Vogal (VCV)									
(V)	A				E			I	
(CV)	a	á	â	ã	e	é	ê	i	í
-BA	aba	ábaco	âba	[ãba]	eba	[éba]	[êba]	iba	[íba]
-CA	academia	ácaro	[âca]	[ãca]	eca	[éca]	[êca]	tica	íca
-DA	dada	[áda]	[âda]	[ãda]	seda	[péda]	[lêda]	dida	saída
-FA	afazer	[páfa]	[âfa]	[ãfa]	cefa	[éfa]	[pêfa]	cifa	[fífato]

Tabela B.2: Exemplo de tabela resumida para testar a prosódia e/ou transcrição das unidades do tipo vogal-consoante-vogal.

F e as vogais A, E e I. Basta apenas expandi-la completando as demais consoantes e vogais e testando as palavras correspondentes.

B.3 Tabela Vogal-Consoante-Vogal

Temos, a seguir (tabela B.2), um exemplo de tabela para testar unidades do tipo vogal-consoante-vogal. Mais uma vez, de maneira análoga à tabela B.1, ressalta-se que ela está resumida, bastando fazer os complementamentos necessários.

B.4 Tabela Encontros Consonantais

Temos, a seguir (tabela B.3), um exemplo de tabela para testar unidades em que ocorre encontro consonantal. Mais uma vez, de maneira análoga às tabelas B.1 e B.2

Encontros Consonantais (EC)									
(V)	A				E			I	
(EC)	a	á	â	ã	e	é	ê	i	í
BR-	brabo	brá	brâ	brã	bre	bré	brê	brito	brí
CR-	cravo	crá	crânio	[crã]	[creca]	[créca]	[crê]	critica	[crí]
DR-	dragar	drá	[drâ]	[drã]	dre	dré	drê	dri	drí
FR-	fraco	[fráfa]	[frâfa]	[frã]	freta	[fré]	[frê]	frita	[frí]

Tabela B.3: Exemplo de tabela resumida para testar a prosódia e/ou transcrição das unidades em que ocorre encontro consonantal.

ressalta-se que ela está resumida, bastando fazer os complementamentos necessários.

B.5 Encontros Vocálicos

Para encontros vocálicos pode também ser feita uma tabela nos moldes das mostradas neste apêndice. Mas pode-se simplesmente listar as possibilidades, uma vez que encontros vocálicos são em menor número que encontros consonantais.